Received: 28 January 2022 | Revised: 4 March 2022 | Accepted: 15 April 2022

DOI: 10.1002/sta4.467

ORIGINAL ARTICLE

WILEY

Fast estimators for the mean function for functional data with detection limits

Haiyan Liu[1] | Jeanine Houwing-Duistermaat[1,2]

[1]Department of Statistics, University of Leeds, Leeds, LS2 9JT, UK

[2]Department of Statistical Sciences "Paolo Fortunati", University of Bologna, Bologna, 40126, Italy

Correspondence
Haiyan Liu, Department of Statistics, University of Leeds, Leeds LS2 9JT, UK.
Email: h.liu1@leeds.ac.uk

Funding information
Alan Turing Institute; International Cooperation Project of Shaanxi Province, Grant/Award Number: 2021KW-14

In many studies on disease progression, biomarkers are restricted by detection limits, hence informatively missing. Current approaches either ignore the problem by just filling in the value of the detection limit for the missing observations or apply a global approach for estimation of the mean function. The latter is time-consuming for dense data, and the obtained estimate depends on the whole observed interval which might not be realistic. We will propose novel estimators for the mean function for both unbalanced sparse and dense data subject to the detection limit. We will derive the asymptotic properties of the estimators. We will compare our methods to the existing methods via simulations and illustrate the new methods with a data application. Our methods appear to perform well. For dense data, the approximation methods are computationally much faster than existing methods.

KEYWORDS
detection limit, functional data analysis, informative missing, local log-likelihood mean estimation

1 | INTRODUCTION

The availability of dense observations along a continuum has motivated the development of functional data analysis (FDA) methods; see, for example, Ramsay and Silverman (2005), Ferraty and Vieu (2006), Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2017). These methods are also of interest for sparse data because of their flexibility in estimation of the mean function and the covariance structure of the continuum (Li & Hsing, 2010; Peng & Paul, 2009; Wang et al., 2016; Yao et al., 2005; Zhang & Wang, 2016). A drawback of these methods however is that they do not fully model the distribution of the observed data points. Methods which do not model the whole distribution may provide biased results in the case of missing data. Only when the data are missing completely at random (Little & Rubin, 2019), the FDA estimation procedure provides unbiased estimators. For missing at random or for missing not at random, the observed sample is not a random sample of the population; hence, the estimators might be biased. To obtain valid estimators, the missingness has to be modelled. In this paper, we consider a specific case of missing data not at random, namely, missingness due to the presence of a detection limit (DL). We propose a model and computational fast estimation procedures for the model parameters.

When levels of a specific marker in a sample have to be determined in a laboratory, we often deal with DLs. The amount of the marker might be too low to be detected. This results in too many zeros in the dataset, and an observed 'zero' might be true zero or just very small. Also on the other extreme of the distribution, DLs might occur, since measurement techniques are often optimized for a certain range of values, and values above and below a certain threshold cannot be accurately measured. DLs are not restricted to laboratory measurements only. Devices which measure certain characteristics (the number of steps for example) might be out of charge yielding an underestimation of the characteristic per day (e.g., the true number of steps for a day is higher if the device was out of charge). For simplicity in this paper, we only consider DLs on the lower

extreme of the distribution; that is, we do not observe values lower than a specific value, instead we observe this specific value which is also called a DL.

The motivating data example is a cohort of 217 scleroderma patients with hospital visits every 6 months from 2010 to 2015. The aim is to identify biomarkers associated with disease progresses over time; see, for example, Clements et al. (1993), Muangchan et al. (2013), Jaeger et al. (2018) and Khanna et al. (2017). While the patients were advised to visit a hospital once in every 6 months, patients skipped visits, which resulted in a sparsely observed unbalanced dataset with in total 408 visits. Here, we focus on estimation of profiles of two biomarkers which are restricted by a DL at the lower end of the distribution, namely, aldose reductase (AR) with 7.8% values below the DL and alpha fetoprotein (AF) with 75% of the values below the DL.

To obtain more accurate estimators, we need to model the missingness of the data. We propose to use the probability density function $f(Y)$ for the observed values and the probability distribution function $F(Y \le DL)$ for the missing observations. As we will see, the presence of the probability distribution function necessaries numerical approximations in the estimation procedures. Therefore, we propose, in additional to the 'exact' methods, also methods which are based on approximations of the probability distribution function. Such an approximation procedure is expected to be much faster for dense observations. Furthermore, to estimate the mean function around the observed time points, we propose to use the local polynomial kernel method (Beran & Liu, 2014; Fan & Gijbels, 1995, 2018) instead of the global method which was used by Shi et al. (2021). The reason is that the observations close to the target point t contain more information about the mean function at t than observations far away from t. Moreover, in addition to assigning larger weight to the observations in the neighbourhood of t by using a kernel function, we propose two weighing schemes for subjects, namely, the 'SUBJ' scheme and the 'OBS' scheme following the same terminology in Zhang and Wang (2016). While the SUBJ scheme assigns the same weight to each subject, the OBS scheme assigns the same weight to each observation which implies that subjects with more observations will have more weight.

An alternative approach is to replace the missing values by the value of the DL and apply standard methods (Li & Hsing, 2010; Yao et al., 2005; Zhang & Wang, 2016). Such an approach will give biased estimators, especially when there are many observations restricted by a DL (Uh et al., 2008). An approach that addresses the DL problem is the global method of Shi et al. (2021). This estimator might be unbiased but may also give inaccurate estimators in some situations, for example, when the mean changes over time. Moreover, its estimation procedure is time-consuming, especially for dense data.

We propose local constant and local linear estimators with approximation and without approximation ('exact') for dense and sparse functional data with DLs and derive their asymptotic behaviour. While the global mean estimator proposed by Shi et al. (2021) might be unbiased, it is less accurate than our estimator when the mean function changes over time as we observed in our application. Via simulations, we evaluate the performance of our estimators in a sparse and a dense settings under both SUBJ and OBS weighing schemes and compare their performance with the global approach and with using a standard method where the DL is used for the missing values. Our method which uses an appropriate approximation in the likelihood function reduces the computational time considerably compared to the method of Shi et al. (2021), for example, in our simulation more than 100 times compared to the existing global method in the dense setting and around 40 times for the sparse setting. We also investigate the asymptotic behaviour of the estimators via simulations. The asymptotic properties of the existing global method of Shi et al. (2021) have not been investigated. Finally, we apply the method to data on two biomarkers and finish with a conclusion.

2 | METHODOLOGY

2.1 | Functional principal component analysis (FPCA)

Let $\{X(t) : t \in I\}$ be an L^2 stochastic process on interval I. Let $\mu(t) = E[X(t)]$ and $C(s,t) = E[(X(s) - \mu(s))(X(t) - \mu(t))]$ be the mean and covariance functions of $X(t)$, respectively. Then $X(t)$ can be decomposed into

$$X(t) = \mu(t) + U(t)$$

where $U(t)$ is the stochastic part of $X(t)$ which has mean zero, that is, $E[U(t)] = 0$ for $t \in I$, and covariance $C(s,t) = E[U(s)U(t)]$ for all $s,t \in I$. By Karhunen–Loeve expansion and Mercer's Theorem, we have

$$C(s,t) = \sum_{l=1}^{\infty} \lambda_l \psi_l(s) \psi_l(t)$$

and

$$U(t) = \sum_{l=1}^{\infty} \xi_l \psi_l(t)$$

where $\psi_l(t)$ are eigenfunctions of the covariance operator corresponding to $C(s,t)$, the positive real numbers $\lambda_1 > \lambda_2 > \ldots$ are eigenvalues of the covariance operator corresponding to $C(s,t)$, and $var(\xi_l) = \lambda_l$. Notice the FPCs $\{\psi_l(t)\}$ consist of an orthonormal basis for $L^2(I)$.

Let $X_1(t),\ldots,X_n(t)$ be n iid copies of $X(t)$ with $t \in I$. Notice that typically, $X_1(t),\ldots,X_n(t)$ are not observed. Instead we have observations at discrete time points $t_{i1},\ldots,t_{iN_i}$ for subject i, with N_i the number of measurements for subject i. Furthermore, these observations are perturbations by (additive) random errors of the true values. Specifically, let Y_{ij} denote the jth observation for subject i with $j = 1,\ldots,N_i$ and $i = 1,\ldots,n$, then Y_{ij} can be written as,

$$Y_{ij} = X_i(t_{ij}) + \epsilon_{ij} = \mu(t_{ij}) + U_i(t_{ij}) + \epsilon_{ij} = \mu(t_{ij}) + \sum_{l=1}^{\infty} \xi_{il}\psi_l(t_{ij}) + \epsilon_{ij} \tag{1}$$

where ϵ_{ij} is an independent random measurement error term following a distribution in the exponential family with mean zero and variance σ_ϵ^2. We assume further that ϵ_{ij} is independent of $U_i(t)$ (or equivalently ξ_{il}). Often a Gaussian distribution is assumed; that is, we have $\epsilon_{ij} \sim N(0,\sigma_\epsilon^2)$ and $\xi_{il} \sim N(0,\lambda_l)$.

In addition, the data are subjected to informative missingness due to DL, and not all Y_{ij} are observed. Let δ be the missingness indicator; that is, $\delta_{ij} = 0$ if Y_{ij} is observed, and $\delta_{ij} = 1$ if Y_{ij} is not observed. For the data points that are not observed, we assume that their values are constrained by a known threshold value c_{ij} from below which means the actual unobserved Y_{ij} is smaller than a threshold c_{ij}. For the sake of simplicity of notation, we assume the threshold is fixed, that is, $c_{ij} = c$ for all i,j. Therefore, the observations are as follows.

$$\{(t_{ij}, Y_{ij}, \delta_{ij})\}, i = 1,\ldots,n, j = 1,\ldots,N_i$$

We first propose a method to estimate the mean $\mu(t)$ from the observations. The estimation of covariance $C(s,t)$, ξ, and $\psi(t)$ will be discussed in subsequent papers.

2.2 | Local kernel weighted log-likelihood estimation of mean function

For the sake of simplicity of notation, we assume in this section that $U(t) = 0$ and $\epsilon \sim N(0,\sigma_\epsilon^2)$.

The contribution to the log-likelihood from the jth observation of the ith subject, $(t_{ij}, Y_{ij}, \delta_{ij})$, is

$$l_{ij} = (1-\delta_{ij})l(\mu(t_{ij}), Y_{ij}) + \delta_{ij} l(\mu(t_{ij}), c). \tag{2}$$

Under the Gaussian distribution, we have

$$l(\mu(t_{ij}), Y_{ij}) = \log\left(\phi(Y_{ij}; \mu(t_{ij}), \sigma^2\right)$$

and

$$l(\mu(t_{ij}), c) = \log\left(\Phi(c; \mu(t_{ij}), \sigma^2\right),$$

where $\phi(\cdot;\mu,\sigma^2)$ and $\Phi(\cdot;\mu,\sigma^2)$ are the pdf and cdf of $N(\mu,\sigma^2)$, respectively, with $\sigma^2 = \sigma_\epsilon^2$; see in Li and Zhang (2011) and Shi et al. (2021).

Now to estimate $\mu(t)$ at a time point $t \in I$, we approximate $\mu(t_{ij})$ by its Pth order Taylor expansion at t:

$$\begin{aligned}\mu(t_{ij}) &= \mu(t) + \mu'(t)(t_{ij}-t) + \frac{\mu''(t)}{2!}(t_{ij}-t)^2 + \ldots \\ &\approx \begin{pmatrix} 1 \\ (t-t_{ij}) \\ \vdots \\ (t-t_{ij})^P \end{pmatrix}^T \begin{pmatrix} \mu(t) \\ \mu'(t) \\ \vdots \\ \frac{\mu^{(P)}(t)}{P!} \end{pmatrix} =: \boldsymbol{b}^T(t)\boldsymbol{\beta}.\end{aligned}$$

where $\boldsymbol{b}^T(t) = \left(1, (t-t_{ij}), \ldots, (t-t_{ij})^P\right)^T$ and $\boldsymbol{\beta} = (\beta_0,\beta_1,\ldots,\beta_{P-1})^T = \left(\mu(t), \mu'(t), \ldots, \frac{\mu^{(P)}(t)}{P!}\right)^T$

Therefore, for a data point $(t_{ij}, Y_{ij}, \delta_{ij})$ in a neighbourhood of t, we define its contribution to the local kernel-weighted log-likelihood

$$l_{ij}(\boldsymbol{\beta}; h, t) = w_i\left[\delta_{ij} l\left(c, \boldsymbol{b}^T(t)\boldsymbol{\beta}, \sigma^2\right) + (1-\delta_{ij}) l\left(Y_{ij}, \boldsymbol{b}^T(t)\boldsymbol{\beta}, \sigma^2\right)\right] K_h(t_{ij}-t) \tag{3}$$

where $K_h(\cdot)=K(\cdot/h)/h$ with $K(\cdot)$ a one-dimensional kernel function with bandwidth h. Such a kernel regression method is widely used; see Fan and Gijbels (1995), Yao et al. (2005), Beran and Liu (2014) and Beran and Liu (2016). For the weights w_i, we propose to use

$$w_i^{SUBJ}=\frac{1}{nN_i}$$

or

$$w_i^{OBS}=\frac{1}{\sum_{i=1}^{n}N_i}.$$

Notice that $\sum_{i=1}^{n}N_iw_i=1$. In FDA, the SUBJ and OBJ are the most commonly used weighting schemes (see Zhang & Wang, 2016). The SUBJ scheme assigns a higher weight to the observations from subject i than from subject k if $N_i<N_k$, While, the OBS assigns the same weight to each observation independent of the number of observations of a subject.

Therefore, the local (kernel-weighted) log-likelihood is

$$L_P(\boldsymbol{\beta};h,t)=\sum_{i=1}^{n}w_i\sum_{j=1}^{N_i}l_{ij}(\boldsymbol{\beta};h,t).$$

Maximizing the above equation with respect to $\boldsymbol{\beta}$ gives the estimator of $\boldsymbol{\beta}$: $\hat{\boldsymbol{\beta}}=(\hat{\beta}_0,\ldots,\hat{\beta}_{P-1})^T=\left(\hat{\mu}(t),\ldots,\hat{\mu}^{(P)}(t)/P!\right)^T$, and obviously, $\hat{\beta}_0$ is the estimator of $\mu(t)$, and $\hat{\beta}_p$ is the estimator of $p!\mu^{(p)}(t)$ for $p=1,\ldots,P$.

Now for $P=2$, the local log-likelihood is

$$\begin{aligned}L_2(\boldsymbol{\beta};h,t)\;=&\sum_{i=1}^{n}w_i\sum_{j=1}^{N_i}\left[\delta_{ij}\log\left\{\Phi\left(\frac{c-(\beta_0+\beta_1(t_{ij}-t))}{\sigma}\right)\right\}\right.\\&\left.+(1-\delta_{ij})\log\left\{\phi\left(\frac{y_{ij}-(\beta_0+\beta_1(t_{ij}-t))}{\sigma}\right)\right\}\right]K_h(t_{ij}-t).\end{aligned}\tag{4}$$

Optimization techniques can be used to maximize $L_2(\boldsymbol{\beta};h,t)$ (see simulation studies). Maximizing it directly to obtain an analytic form of the estimator is difficult, because the log-likelihood contains the function $\log(\Phi(x))$ for the observations restricted by the DL. However, in most cases when dealing with a DL, the function values of $\log(\Phi(x))$ are only needed for a small range of x values, namely, $x=\sigma^{-1}[c-(\beta_0+\beta_1(t_{ij}-t))]>0$, as $c>\beta_0+\beta_1(t_{ij}-t)$. Furthermore, $\log(\Phi(x))\approx 0$ for $x>2$. Thus, if we take the random error into account, we only need the function values of $\log(\Phi(x))$ for $-1<x=\sigma^{-1}[c-(\beta_0+\beta_1(t_{ij}-t))]<2$. We propose to approximate $\log(\Phi(x))$ for $-1<x<2$ by a quadratic function. The quadratic function is obtained by first dividing $[-1,2]$ into 300 equal subintervals and then computing the function values of $\log(\Phi(x))$ at each dividing point. Then a regression is performed with the function values of $\log(\Phi(x))$ at these dividing points as the response variable and the dividing points x and the square of x as the covariates. Thus, $\log(\Phi(x)), x\in[-1,2]$, is approximated by

$$\log(\Phi(x))\approx -0.7172+0.8194x-0.251x^2, x\in[-1,2].$$

From Figure 1, we see that this approximation is accurate.

This approximation does not only reduce the computational burden but also simplifies the estimation procedure. Moreover, it helps to obtain the asymptotic property of the estimators in a more intuitive way.

Under this quadratic approximation, the estimators $\boldsymbol{\beta}=(\beta_0,\beta_1)^T$ have a closed form. Notice that, based on this approximation, the local log-likelihood (4) reduces to (ignoring a constant term)

$$\begin{aligned}L_2(\boldsymbol{\beta};h,t)\;=&\sum_{i=1}^{n}w_i\sum_{j=1}^{N_i}\left[-0.251\delta_{ij}\left(\frac{c-(\beta_0+\beta_1(t_{ij}-t))}{\sigma}\right)^2\right.\\&+0.8194\delta_{ij}\frac{c-(\beta_0+\beta_1(t_{ij}-t))}{\sigma}\\&\left.+(0.5-0.5\delta_{ij})\left(\frac{y_{ij}-(\beta_0+\beta_1(t_{ij}-t))}{\sigma}\right)^2\right]K_h(t_{ij}-t).\end{aligned}\tag{5}$$

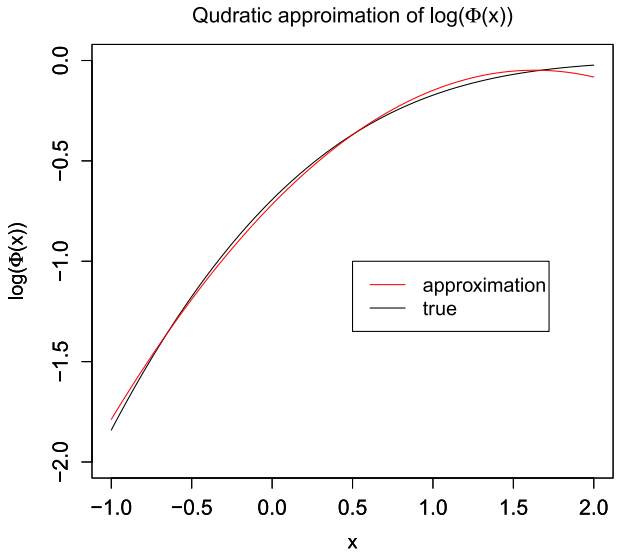

FIGURE 1 Quadratic approximation of $\log(\Phi(x))$ when $x \in [-1,2]$

Maximizing (5) with respect to β_0 and β_1 gives a local linear estimator of mean function, $\hat{\mu}(t)$. The case of $P=1$ gives a local constant estimator of the mean function. Here, we only derive the asymptotic normality of the local linear estimator of mean (see Theorem 1). The asymptotic normality of local constant estimator of mean can be derived in a similar way but less complicated.

Theorem 1. *Asymptotic normality of $\hat{\mu}(t)$:*

For a fixed interior point $t \in I$, under Assumptions 1-3 given in the Appendix, we have

$$(\Gamma_{n,N_i})^{-1/2}\left[\hat{\mu}(t)-\mu(t)-B(t)\sigma-\frac{h^2}{2}\sigma_K^2\mu''(t)+o(h^2)\right]\to N(0,1)$$

where

$$\Gamma_{n,N_i}=\frac{\sum_i N_i(w_{i1}(t)+w_{i4}(t)-B(t)w_{i2}(t))^2+w_{i3}^2(t)}{h}\|K\|^2\frac{\sigma^2}{f(t)}$$

with

$$\begin{aligned}
B(t) &= \frac{\sum_i N_i(w_{i1}(t)+w_{i4}(t))}{\sum_i N_i w_{i2}(t)},\\
w_{i1}(t) &= -0.8194\delta_i(t)w_i,\\
w_{i2}(t) &= (1-0.498\delta_i(t))w_i,\\
w_{i3}(t) &= (1-\delta_i(t))w_i,\\
w_{i4}(t) &= 0.502\delta_i(t)[-1,2]w_i.
\end{aligned}$$

Remark 1. We notice $w_{i4}(t)$ has the expression $0.502\delta_i(t)[-1,2]w_i$, which is the likelihood contribution of the observations that are restricted by the DL. Because we approximate the distribution in the interval $[-1,2]$, this interval appears in the expression. Clearly, if $\delta_i(t)=0$, that is, none of the observations is restricted by the DL, and $w_{i4}(t)$ is omitted.

Remark 2. If all $\delta_{ij}=0(\delta_i(t)=0)$, that is, no DL, then the asymptotic bias reduces to

$$\frac{h^2}{2}\sigma_K^2\mu''(t)+o(h^2)$$

which coincides with existing results in Zhang and Wang (2016).

If all $\delta_{ij} = 1(\delta_i(t) = 1)$, that is, all observations are contaminated with the DL, then the asymptotic bias is

$$(-0.8194 + 0.502[-1,2])\sigma + \frac{h^2}{2}\sigma_K^2\mu''(t) + o(h^2)$$

which of course is influenced by the extent to which the true values are different from threshold c.

Remark 3. If all $\delta_{ij} = 0(\delta_i(t) = 0)$, that is, no DL, then the asymptotic variance reduces to

$$\frac{\sum_i N_i w_i^2}{h}\|K\|^2\frac{\sigma^2}{f(t)}$$

which coincides with existing results in Zhang and Wang (2016).

Remark 4. Theoretically, Assumption 3 in Appendix 3 provides a guide for selecting an optimal bandwidth for the estimation of the mean function; that is, it depends on n, N_i, w_i and δ_{ij}. In practice, we would suggest a data-driven method to select an appropriate bandwidth, for example, GCV which we used in the data analysis.

3 | SIMULATION STUDY

We evaluate the performance of our proposed methods (local constant with approximation, local linear with approximation, local constant without approximation) for various scenarios via simulations. We compare their performance with the global method of Shi et al. (2021) and with a standard method where the missing observations are replaced with the DL value (Yao et al., 2005). We compare the methods in terms of bias, efficiency, asymptotic behaviour and computation time.

We assume that $U(t) = 0$ for simplicity and define the true mean function as follows:

$$\mu(t) = -0.5 + 1.5\sin(10\pi(t + 0.5)) + 4(t-1)^3, t \in [1,2].$$

The observed time points $t_{ij} \sim U[1,2]$ are iid sampled from the continuous uniform distribution in the interval $[1,2]$. Additive errors are sampled from $\epsilon_{ij} \sim N(0,1)$. Then, the response is generated by

$$Y_{ij} = \mu(t_{ij}) + \epsilon_{ij}, i = 1,...,n, j = 1,...,N_i.$$

Finally, we create missing data by replacing observations smaller than zero with zero; that is, we assume a DL of zero.

We consider two settings, namely, a sparse and a dense grid for the observations for each subject i. Specifically,

- Sparse setting: $N_i \sim U\{3,4,5,6,7,8,9,10\}$, that is, N_i are iid from a discrete uniform distribution in the interval $[3,10]$.
- Dense setting: $N_i \sim U\{175,176,...,200\}$, that is, N_i iid from a discrete uniform distribution in the interval $[175,200] = \left[\frac{7}{8}n, n\right]$ where $n = 200$ is the number of subjects.

For each setting, we simulate $Q = 100$ replicates. Each replicate contains information of $n = 200$ subjects.

To estimate the mean functions in the replicates, we consider the following estimation methods:

- local constant approximation: the local constant method ($P = 1$) but using the quadratic approximation.
- local linear approximation: the local linear method ($P = 2$) but using the quadratic approximation.
- local constant exact: the local constant method ($P = 1$) without approximation, using numerical optimization methods to obtain the parameters.
- imFunPCA: with adjustment of DL but not in a local way proposed by Shi et al. (2021).
- PACE: without adjustment of detection limit but in a local linear way proposed by Yao et al. (2005).

The variance of ϵ_{ij} is estimated as the mean squared error based on the least-squared fit using all the data (including the values subject to DL). We use the Gaussian kernel for the estimation procedure. To select the bandwidth h, the integrated squared error (ISE) is computed for a dense grid of values; namely, $h = (3:10)/400$. The ISE is defined as follows:

$$ISE(\hat{\mu}(t), h) = \int_1^2 (\hat{\mu}(t) - \mu(t))^2 dt.$$

The bandwidth which minimizes $ISE(\hat{\mu}(t), h)$ is selected as the optimal bandwidth, and the corresponding ISE is denoted with $ISE_{opt}(\hat{\mu}(t))$.

We then calculate the mean integrated squared error (MISE) and the SD of ISE over 100 replicates:

$$MISE(\hat{\mu}(t)) = \frac{1}{Q}\sum_{i=1}^{Q} ISE_{opt}(\hat{\mu}^{([i])}(t)).$$

$$SD(\hat{\mu}(t)) = \sqrt{\frac{1}{Q-1}\sum_{i=1}^{Q}\left(ISE_{opt}(\hat{\mu}^{([i])}(t)) - MISE(\hat{\mu}(t))\right)^2},$$

where $\hat{\mu}^{([i])}(t)$ is the mean estimation based on the ith replicate.

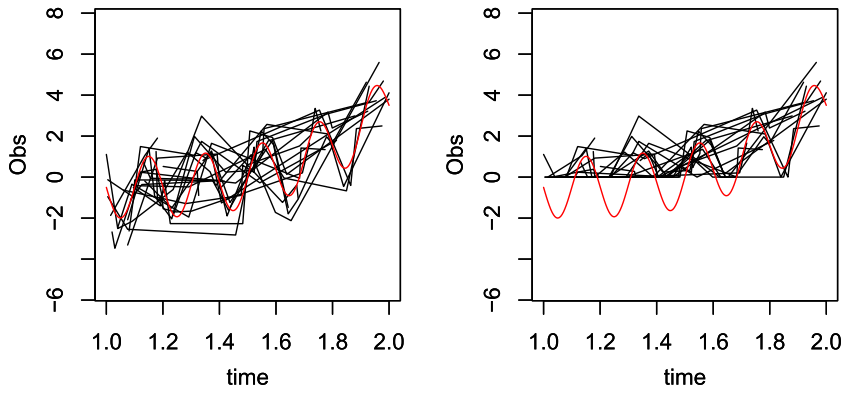

FIGURE 2 The observations in the sparse setting: left without the detection limit, right with the detection limit, the red curve is the true curve

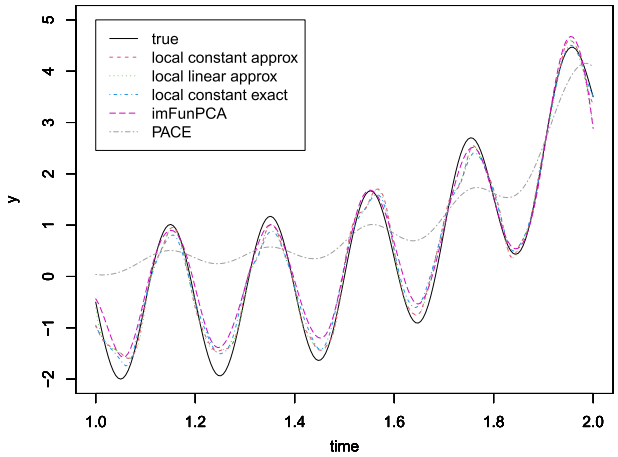

FIGURE 3 The mean estimation in the sparse setting (estimated at 200 equal-distant points in [1,2]: bandwidth for the constant approximation method is 0.015; bandwidth for the linear approximation method is 0.015; for the exact method is 0.015. The number of basis for imFunPCA is 20)

For the sparse setting and the first replicate, Figure 2 depicts the data and the true mean function, and Figure 3 shows the corresponding estimation of mean for different methods, respectively. The proportion of observations subject to DL is 42.17% in this replicate. The mean function is estimated on 200 equal-distant time points in $[1,2]$. For this replicate, the proposed methods (local constant approximation, local linear approximation, local constant exact) appear to perform slightly better than imFunPCA and better than PACE. The time needed for calculating the estimated mean function appeared to vary across the methods. These are 0.03 secs for the local constant approximation, 0.06 seconds for the local linear approximation, 22.51 seconds for the local constant exact, 1.26 seconds for imFunPCA, and 28.90 seconds for PACE. Thus, our proposed local approximation methods are more time efficient than the other methods.

For the dense setting, the data of one replicate and the mean functions estimated by the various methods are given in Figures 4 and 5, respectively. The proportion of observations subject to DL is 42.43% in this replicate. The mean function is estimated on 200 equal-distant time points in $[1,2]$. For this replicate, the proposed methods (local constant approximation, local linear approximation, local constant exact) perform much better than the existing ones: imFunPCA and PACE. The time needed for calculating the estimated mean function appeared again to vary across the methods, namely 0.11 seconds for the local constant approximation, 0.24 seconds for the local linear approximation, 647.67 seconds for the local constant exact, 31.72 seconds for imFunPCA, and 8047.90 seconds for PACE. Thus, our proposed local approximation methods are considerably more time efficient than the other methods for the dense setting.

In Table 1, the MISE*100 and the corresponding standard deviation (SD) for local constant approximation, local linear approximation, local constant exact, imFunPCA and PACE for the two weighing schemes (SUBJ or OBS) are given. Also the mode of optimal tuning parameter (i.e., bandwidth in local constant/linear estimation or the number of basis in imFunPCA) selected for each replicate is provided. The mean functions are estimated on 50 equal-distant time points in $[1,2]$. Due to the numerical optimization, the exact methods are computational intensive especially for the dense setting. Therefore, we did not consider the linear exact method and only used the OBS scheme. Moreover for the dense

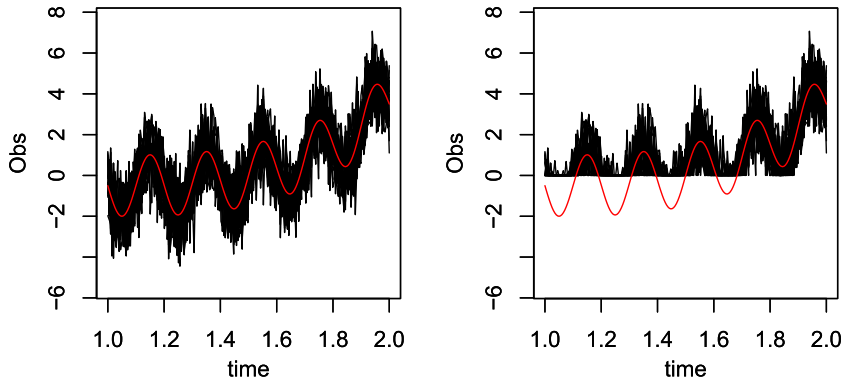

FIGURE 4 The observations in the dense setting: left without the detection limit, right with the detection limit, in red is the true curve

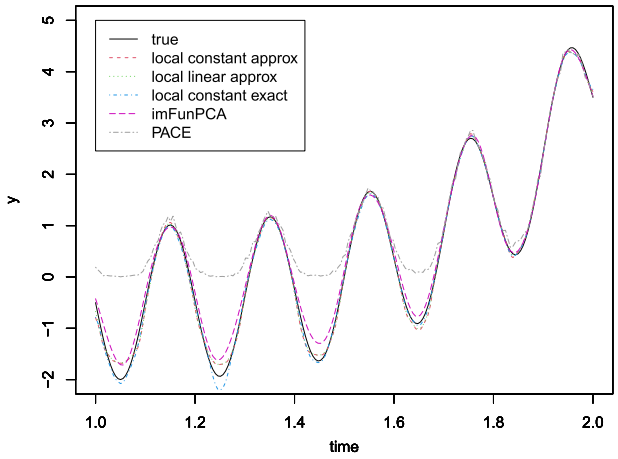

FIGURE 5 The mean estimation in the dense setting (estimated at 200 equal-distant points in $[1,2]$: bandwidth for the constant approximation method is 0.0075; bandwidth for the linear approximation method is 0.01; for the exact method is 0.0075. The number of basis for imFunPCA is 20)

TABLE 1 MISE × 100 and SD(MISE × 100) of local constant approximation, local linear approximation, local constant exact, imFunPCA and PACE for both SUBJ and OBS weighing schemes and both dense and sparse settings and the mode of bandwidth (on the grid (3:10)/400) or the number of basis (on the grid (8:15)*2) selected for each replicate

	Dense		Sparse	
	MISE*100 (SD)	mode(bw/nb)	MISE*100 (SD)	mode(bw/nb)
constant approx(OBS)	1.073337(0.1227)	0.0075	3.89(0.93)	0.0150
constant approx(SUBJ)	1.073083(0.1234)	0.0075	4.34(1.05)	0.0175
linear approx(OBS)	1.024649(0.115)	0.0100	3.84(0.94)	0.0150
linear approx(SUBJ)	1.024568(0.116)	0.0100	4.29(1.05)	0.0175
	Based on 10 replicates			
constant exact(OBS)	0.72(0.137)	0.0075(fixed)	4.24(1.42)	0.0150
imFunPCA(OBS)	2.14(0.229)	20	4.39(1.94)	20
	Based on 1 replicate			
PACE(OBS)	138(NA)		230(3.87)	

Note: Notice that 'fixed' means, for the constant exact method in dense setting, we set bandwidth to be 0.0075 which is the same as that in the approximation approaches to reduce the computation burden. There are in total 100 replicates. The mean function is estimated on 50 equal-distant time points in $[1,2]$. The results of constant exact and imFunPCA in dense setting are based on 10 replicates. The PACE method in dense setting only has one replicate.

setting, we only simulated 10 replicates and did not select an optimal bandwidth. Note that for the dense setting, the approximation methods using the OBS scheme and the SUBJ scheme perform similarly. Here we only simulated 10 replicates and did not select an optimal bandwidth. And the bandwidth was set to be 0.0075 which is the obtained value in the approximation approaches. Also for the imFunPCA method and the dense setting, the computation burden is large. Hence, we only analysed the first 10 replicates. For these replicates, we still select an optimal number of basis over a grid ((8:15)*2), since based on the results of the local methods, the sparse case might need a bigger bandwidth. It appeared however that the same number of basis is selected as in the sparse case, namely, 20. Finally, for the PACE method in the dense setting, we only analysed 1 replicate, as the computation time is huge. Note that PACE method automatically selected an optimal bandwidth by GCV.

It appears that for the dense setting, the SUBJ scheme gives slightly smaller MISE compared to the OBS scheme. This in contrast to the sparse setting, for which the OBS scheme gives smaller MISE than the SUBJ scheme. The local linear approximation gives slightly better results than the local constant approximation. Note that in contrast to the other methods, the local linear approximation also gives an estimate of the first derivative of the mean function. With regard to the existing methods, the MISE of the global method imFunPCA is larger than that of the local methods proposed in this paper, especially for the dense setting. For PACE, its MISE is much larger than that of the local methods proposed in this paper for both dense and sparse settings as could be expected since it does not take into account of the DL.

We also evaluated the asymptotic behaviour of the local constant approximation, local linear approximation and local constant exact methods by computing MISE and SD by increasing the sample size n. The mean function is estimated on 50 equal-distant time points in $[1,2]$. The results are given in Table 2. Clearly as n increases (from 50 to 200), the MISE and the corresponding SD decrease for all the methods. For all considered sample sizes, the local methods have the highest accuracy, although the difference decreases with the sample size. Especially, the local linear approximation method in the sparse setting performed better. For the sparse setting, the OBS scheme performs better than the SUBJ scheme. For the dense setting, the two weighing schemes perform similar, but as n increases, the SUBJ scheme appears to outperform the OBS scheme slightly.

4 | DATA APPLICATION

In total, information on two biomarkers from 217 scleroderma patients with hospital visits from 2010 to 2015 has been retrospectively obtained for this study. Scleroderma has a heterogeneous disease course across patients. For some patients, the disease worsens over time while, for others, the disease is stable. The data were collected according to an ethically approved protocol for the observational study HRA number 15/NE/0211. Typically, scleroderma patients visit the hospital every 6 months to check whether the disease has progressed. However, patients missed their appointments, or their data were not recorded, resulting in a sparse unbalanced dataset. Here, we are interested in estimation of the mean function of two biomarkers for disease progression which are subject to a DL, namely, AR and AF. The DL for AR is 3.1; that is, the values that are less than 3.1 are set to 3.1; while the DL for AF is 0.98; that is, the values that are less than 0.98 are set to 0.98. For AR, 7.8% observations are missing due to the DL, while for AF, this percentage is 75%. After removing observations at time points with no outcome or no

biomarker values, some outlier values (at one time point AR has a value larger than 3 times the SD and patients with only one observation), our final dataset comprises 90 patients with in total 268 observations.

We estimate the mean of AR and AF by using the three novel methods proposed in this paper and two existing methods. Since the data are sparse and the number of patients is relatively small, based on the results of the simulation, we would prefer the local constant approximation and local linear approximation methods for estimation of the mean functions of AR and AF. We only consider the OBS scheme which appeared to be the best option for sparse data in the simulation study. For the local constant approximation, the bandwidth is selected using CV over a fine grid. This selected bandwidth is also used in the local linear approximation and constant exact methods. The number of basis in imFunPCA is also selected based on CV over a grid. The results are shown in Figures 6 and 7 for AR and AF, respectively. For AR, the various methods provide similar estimates of the mean function, which was expected because of the small number of missing values due to DL. For AF, the percentage of observations subject to DL is much larger; hence, methods which do not adjust for DL should not be used. From Figure 7, PACE provides, indeed, a different estimate for the mean function, and this result should not be trusted. Furthermore, the local constant approximation, local linear approximation and local constant exact method give similar results before 40 months. And imFunPCA varies more than the local methods. Note that after 40 months, none of the estimations can be trusted since there is very limited information.

Because of the lack of information after 40 months, we also estimated the curves until 40 months. The results are depicted in Figures 8 and 9 for AR and AF, respectively. We do not expect the local approaches to change much if the bandwidth is not big, while the global approaches might

TABLE 2 MISE $\times$ 100 and and SD(IMSE $\times$ 100) of local constant approximation, local linear approximation, local constant exact, for both OBS and SUBJ weighing schemes and both dense and sparse settings under different sample sizes (50, 100, 150 and 200)

	$n = 50$	$n = 100$	$n = 150$	$n = 200$
	Dense			
constant approx(OBS)	2.3027(0.462)	1.3985(0.253)	1.1425(0.149)	1.0733(0.123)
constant approx(SUBJ)	2.3065(0.461)	1.3993(0.252)	1.1422(0.148)	1.0731(0.123)
linear approx(OBS)	2.3375(0.487)	1.3044(0.229)	1.0791(0.133)	1.0246(0.115)
linear approx(SUBJ)	2.3400(0.488)	1.3050(0.227)	1.0788(0.132)	1.0246(0.116)
	Sparse			
constant approx(OBS)	10.12(2.79)	6.36(1.46)	4.61(1.05)	3.89(0.93)
constant approx(SUBJ)	11.29(3.16)	7.07(1.64)	5.21(1.27)	4.34(1.05)
constant exact(OBS)	12.15(5.60)	8.24(5.36)	5.14(1.49)	4.24(1.42)
linear approx(OBS)	9.86(2.92)	6.34(1.46)	4.54(1.07)	3.84(0.94)
linear approx(SUBJ)	10.89(3.30)	7.07(1.63)	5.15(1.28)	4.29(1.05)

Note: The optimal bandwidth is selected based on the MISE criteria. There are in total 100 replicates. The mean function is estimated on 50 equal-distant time points in $[1,2]$.

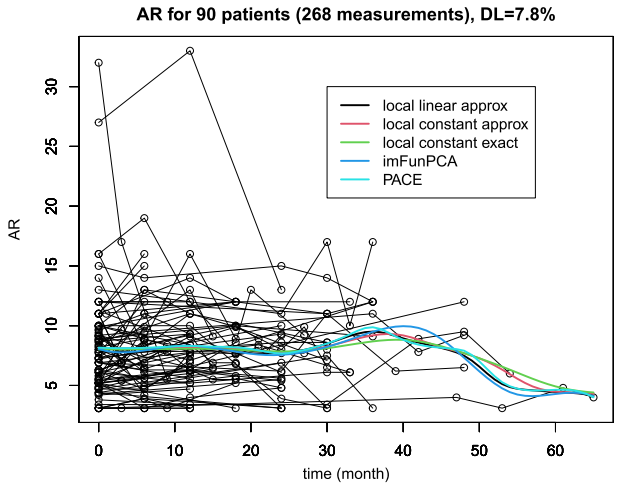

FIGURE 6 The mean estimation for AR by using different methods

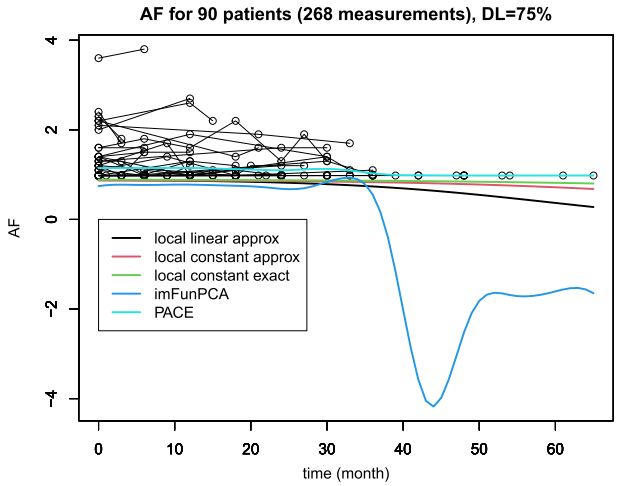

FIGURE 7 The mean estimation for AF by using different methods

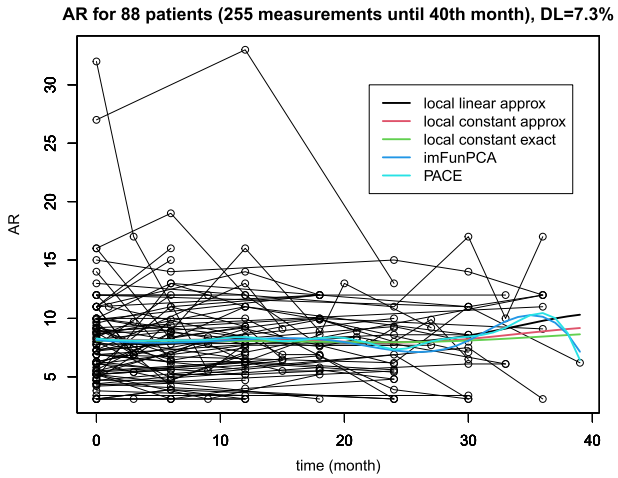

FIGURE 8 The mean estimation for AR by using different methods

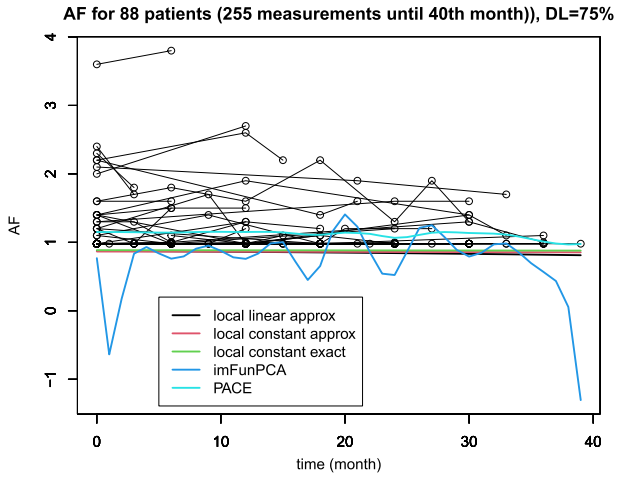

FIGURE 9 The mean estimation for AF by using different methods

be affected by restricting the analysis to the first 40 months. The bandwidth for the local linear approximation is selected based on CV over a fine grid, and this bandwidth is also used for the local constant approximation and the constant exact methods. The number of basis in imFunPCA is selected based on CV over a grid. Since there are not many observations observed below DL for the AR, the various methods gave similar results (Figure 8). While for AF (Figure 9), the local methods give similar results as expected. PACE gives again too high values since it does not take into account of DL. The mean function estimated by imFunPC is quite different and varies across time probably due to undersmoothing. Thus, just as in the simulation study for the dataset, the novel local approaches perform best especially when there are many observations subject to DL and in the sparse setting.

5 | DISCUSSION

We have proposed novel estimators for the mean function using unbalanced sparse and dense data subject to DL. Our method is based on local smoothing of the mean functions using kernel functions. We derived the asymptotic properties of the estimators. We compared our methods to existing methods and showed that our methods performed better in terms of efficiency, bias and computation time. We also considered two weighing schemes for the observations: one based on single observations and the other based on subjects. For sparse data, weighing per observation appeared to perform better. For dense data, the approximation methods are computationally fast.

We applied our methods to sparse and unbalanced functional data on two biomarkers. The two biomarkers have different percentage of observations restricted to DL, namely, 7.8% and 75%, respectively. Our proposed methods appear to give appropriate estimates of the mean function. The global method appears to be affected by reducing the observation period. In contrast to the other methods, its estimate of the mean function fluctuated. Ignoring DL was not an option for the biomarker which had 75% observations subject to DL.

An alternative to our approach might be imputation of the missing observations. For cross sectional data, Uh et al. (2008) studied the performance of imputation methods. They conclude that these methods may give biased estimators or underestimated variances. Moreover, they do not perform well if the percentage of DL is large (say larger than 30%). In this paper, we investigated the DL problem for functional data over time and estimated the mean curves for individuals. Given the results of Uh et al. (2008) and the fact that multiple imputations would increase the computation time, we did not consider this approach for the estimation of the mean curve.

Currently, we are working on an estimator for the covariance function estimator for datasets with observations that are restricted by the DL. Estimators of the mean and covariance functions would enable us to setup the corresponding FPCA. FPCA provides us a dimension reduction method from infinite dimension to a finite dimension. FPCA will also provide smooth individual curves which are often required for sparse datasets. In addition, a functional regression model can be developed to investigate the influence of covariates with DL on the outcomes which might be also subject to DL. The problem of missing data not at random is not limited to laboratory measurement. Similar informative missingness occurs due to the malfunction of devices; examples are pollutant monitoring stations and wearable devices which collect health status data. This is sometimes called partially observed functional data. To conclude, our work is a first step to FPCA and functional regression methods for data partially missing, which is a common situation in real data examples.

ACKNOWLEDGEMENT

We thank Dr. Francesco Del Galdo from the School of Medicine at the University of Leeds for providing the scleroderma dataset and his expertise and assistance throughout all aspects of the scleroderma data analysis. This work is supported by a fellowship of the Alan Turing Institute and by the International Cooperation Project of Shaanxi Province under Grant 2021KW-14.

R CODE

R code for part of the simulation is given on https://github.com/HLiuMath/FDAwithDL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are provided by Dr. FrancescoDel Galdo from the School of Medicine at the University of Leeds.

REFERENCES

Beran, J., & Liu, H. (2014). On estimation of mean and covariance functions in repeated time series with long-memory errors. *Lithuanian Mathematical Journal*, *54*(1), 8–34.

Beran, J., & Liu, H. (2016). Estimation of eigenvalues, eigenvectors and scores in FDA models with dependent errors. *Journal of Multivariate Analysis*, *147*, 218–233.

Clements, P. J., Lachenbruch, P. A., Seibold, J. R., Zee, B., Steen, V. D., Brennan, P., Silman, A. J., Allegar, N., Varga, J., & Massa, M. (1993). Skin thickness score in systemic sclerosis: an assessment of interobserver variability in 3 independent studies. *The Journal of rheumatology*, *20*(11), 1892–1896.

Fan, J., & Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(2), 371–394.

Fan, J., & Gijbels, I. (2018). *Local polynomial modelling and its applications*: Routledge.
Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: Theory and practice*, Vol. 76: Springer.
Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*, Vol. 200: Springer Science & Business Media.
Jaeger, V. K., Distler, O., Maurer, B., Czirják, L., Lóránd, V., Valentini, G., Vettori, S., Del Galdo, F., Abignano, G., Denton, C., Nihtyanova, S., Allanore, Y., Avouac, J., Riemekasten, G., Siegert, E., Huscher, D., Matucci-Cerinic, M., Guiducci, S., Frerix, M., ..., & Walker, U. A. (2018). Functional disability and its predictors in systemic sclerosis: a study from the desscipher project within the eustar group. *Rheumatology*, *57*(3), 441–450.
Khanna, D., Furst, D. E., Clements, P. J., Allanore, Y., Baron, M., Czirjak, L., Distler, O., Foeldvari, I., Kuwana, M., Matucci-Cerinic, M., Mayes, M., Medsger Jr, T., Merkel, P. A., Pope, J. E., Seibold, J. R., Steen, V., Stevens, W., & Denton, C. P. (2017). Standardization of the modified rodnan skin score for use in clinical trials of systemic sclerosis. *Journal of Scleroderma and Related Disorders*, *2*(1), 11–18.
Kokoszka, P., & Reimherr, M. (2017). *Introduction to functional data analysis*: Chapman and Hall/CRC.
Li, J., & Zhang, W. (2011). A semiparametric threshold model for censored longitudinal data analysis. *Journal of the American Statistical Association*, *106*(494), 685–696.
Li, Y., & Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*, *38*(6), 3321–3351.
Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data*, Vol. 793: John Wiley & Sons.
Muangchan, C., Baron, M., Pope, J., & Canadian Scleroderma Research Group (2013). The 15% rule in scleroderma: the frequency of severe organ complications in systemic sclerosis. A systematic review. *The Journal of rheumatology*, *40*(9), 1545–1556.
Peng, J., & Paul, D. (2009). A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics*, *18*(4), 995–1015.
Ramsay, J., & Silverman, B. (2005). *Functional data analysis*: Springer.
Shi, H., Dong, J., Wang, L., & Cao, J. (2021). Functional principal component analysis for longitudinal data with informative dropout. *Statistics in Medicine*, *40*(3), 712–724.
Uh, H.-W., Hartgers, F. C., Yazdanbakhsh, M., & Houwing-Duistermaat, J. J. (2008). Evaluation of regression methods when immunological measurements are constrained by detection limits. *BMC Immunology*, *9*(1), 1–10.
Wang, J.-L., Chiou, J.-M., & Müller, H.-G. (2016). Functional data analysis. *Annual Review of Statistics and Its Application*, *3*, 257–295.
Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, *100*(470), 577–590.
Zhang, X., & Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics*, *44*(5), 2281–2321.

How to cite this article: Liu, H., & Houwing-Duistermaat, J. (2022). Fast estimators for the mean function for functional data with detection limits. *Stat*, *11*(1), e467. https://doi.org/10.1002/sta4.467

APPENDIX A

ASSUMPTIONS

Assumption 1. Assumptions for kernel function:

(A1) The kernel function $K(\cdot)$ is a symmetric probability density function on $[-1,1]$, and

$$\sigma_K^2 = \int u^2 K(u) du <$$

and

$$\|K\|^2 = \int K^2(u) du < .$$

Assumption 2. Assumptions for time points and true functions:

(B1) Time points $\{t_{ij}, i = 1,...,n, j = 1,...,N_i\}$ are iid copies of a random variable T defined on interval I with density $f(\cdot)$:

$$0 < m_f \le \min f(t) \le \max f(t) \le M_f <$$

and $f^{'}(t)$ is bounded.

(B2) Indicator curves $\{\delta_i(t), i=1,...,n\}$ are defined on $[0,1]$ with range $\{0,1\}$ and $\delta_i(t_{ij})=\delta_{ij}$.

(B3) $X(t)$ is independent of T, ϵ is independent of T.

(B4) $\mu^{'}(t)$ is bounded on I.

Assumption 3. Assumptions for deriving the asymptotic distribution of mean estimation:

(C1)

$$h\to 0, \sqrt{\frac{\sum_i N_i w_{i2}^2(t)}{h}}\to 0, \sqrt{\frac{\sum_i N_i (w_{i1}^2(t)+w_{i3}^2(t)+w_{i4}^2(t))}{h}}\to 0.$$

(C2)

$$\min\left\{\frac{h}{\sum_i N_i w_{i2}^2(t)}, \frac{h}{\sum_i N_i (w_{i1}^2(t)+w_{i3}^2(t)+w_{i4}^2(t))}\right\} h^6\to 0.$$

Assumption (A1) is standard in the context of kernel smoothing. Assumptions (B1), (B3) and (B4) are standard in the local polynomial smoothing and the context FDA. Assumption (B2) is the assumption of the DL behaviour of the observations. Assumptions (C1) and (C2) guarantee consistency of the estimators, and (C1) is also used to check the Lyapunov condition for asymptotic normality.

PROOF OF THEOREM 1

Proof. By calculation, we have the following derivatives:

$$\begin{aligned}\frac{\partial L_2}{\partial \beta_0} &= \sigma^{-2}\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i}\big[0.502\delta_{ij}(c-\beta_0-\beta_1(t_{ij}-t))-0.8194\delta_{ij}\sigma \\ &\quad +(1-\delta_{ij})(y_{ij}-\beta_0-\beta_1(t_{ij}-t))\big]K_h(t_{ij}-t)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial L_2}{\partial \beta_1} &= \sigma^{-2}\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i}\big[0.502\delta_{ij}(c-\beta_0-\beta_1(t_{ij}-t))-0.8194\delta_{ij}\sigma \\ &\quad +(1-\delta_{ij})(y_{ij}-\beta_0-\beta_1(t_{ij}-t))\big](t_{ij}-t)K_h(t_{ij}-t).\end{aligned}$$

Setting them to be zero and, and solving β_0 and β_1, we have

$$\hat{\mu}(t)=\hat{\beta}_0=\frac{R_0S_2-R_1S_1}{S_0S_2-S_1^2}$$

where, for $r=0,1,2$,

$$S_r=\sum_{i=1}^{n} w_i \sum_{j=1}^{N_i}(1-0.498\delta_{ij})\left(\frac{t_{ij}-t}{h}\right)^r K_h(t_{ij}-t)$$

and

$$R_r = \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} \left[-0.8194\delta_{ij}\sigma + 0.502\delta_{ij}c + (1-\delta_{ij})y_{ij}\right] \left(\frac{t_{ij}-t}{h}\right)^r K_h(t_{ij}-t).$$

Actually, we can write

$$\hat{\mu}(t) = \hat{\beta}_0 = \frac{R_0}{\widetilde{S}_0} - \frac{\widetilde{S}_1}{\widetilde{S}_0}\hat{\mu}'(t) + t\hat{\mu}'(t)$$

where

$$S_r = \sum_{i=1}^{n} w_i \sum_{j=1}^{N_i} (1 - 0.498\delta_{ij}) t_{ij}^r K_h(t_{ij}-t).$$

Define

$$\tilde{\mu}(t) = \hat{\beta}_0 = \frac{R_0}{\widetilde{S}_0} - \frac{\widetilde{S}_1}{\widetilde{S}_0}\mu'(t) + t\mu'(t).$$

We first prove that we have

$$\sqrt{\min\left\{\frac{h}{\sum N_i w_{i2}^2(t)}, \frac{h}{\sum N_i (w_{i1}^2(t) + w_{i3}^2(t) + w_{i4}^2(t))}\right\}} (\hat{\mu}(t) - \tilde{\mu}(t)) = o_P(1).$$

To see it, we can show that

$$\hat{\mu} - \tilde{\mu} = -\frac{S_1}{S_0} \frac{S_0(R_1 - \mu S_1 - h\mu' S_2) - S_1(R_0 - \mu S_0 - h\mu' S_1)}{S_0 S_2 - S_1^2}.$$

It is straightforward to show that both S_0 and $S_0 S_2 - S_1^2$ are positive and bounded away from 0 with probability tending to one. It is also straightforward to show that

$$\begin{aligned}
S_1 &= O_P\left(h + \sqrt{\frac{\sum N_i w_{i2}^2(t)}{h}}\right), \\
R_1 - \mu S_1 - h\mu' S_2 &= O_P\left(h^2 + \sqrt{\frac{\sum N_i (w_{i1}^2(t) + w_{i3}^2(t) + w_{i4}^2(t))}{h}}\right), \\
R_0 - \mu S_0 - h\mu' S_1 &= O_P\left(h^2 + \sqrt{\frac{\sum N_i (w_{i1}^2(t) + w_{i3}^2(t) + w_{i4}^2(t))}{h}}\right).
\end{aligned}$$

Then we show the asymptotic normality of $\tilde{\mu}(t)$. By the Lyapunov condition and Cramer–Wald device, we can derive the asymptotic joint normality of $(R_0 - E[R_0], \tilde{S}_1 - E[\tilde{S}_1], \tilde{S}_0 - E[\tilde{S}_0])$, and the convergence rate is $\sqrt{\min\left\{\frac{h}{\sum N_i w_{i2}^2(t)}, \frac{h}{\sum N_i (w_{i1}^2(t) + w_{i3}^2(t) + w_{i4}^2(t))}\right\}}$.

In order to calculate the asymptotic variance, we first calculate the asymptotic bias, and we obtain

$$\begin{aligned}
E[\tilde{S}_0] &= \left[f(t)+\frac{h^2}{2}\sigma_K^2 f''(t)+o(h^2)\right]\sum N_i w_{i2}(t),\\
E[\tilde{S}_1] &= \left[tf(t)+\frac{h^2}{2}\sigma_K^2(2f'(t)+tf''(t))+o(h^2)\right]\sum N_i w_{i2}(t),\\
E[R_0] &= \left[f(t)+\frac{h^2}{2}\sigma_K^2 f''(t)+o(h^2)\right]\sigma\sum N_i(w_{i1}(t)+w_{i4}(t))\\
&\quad+\left[\mu(t)f(t)+\frac{h^2}{2}\sigma_K^2(\mu''(t)f(t)+2\mu'(t)f'(t)+\mu(t)f''(t)+o(h^2)\right]\sum N_i w_{i2}(t)
\end{aligned}$$

where $\sigma_K^2=\int u^2K(u)du$. Therefore, by using the delta method, the asymptotic bias is

$$E[\tilde{\mu}(t)]-\mu(t)=\frac{\sum N_i(w_{i1}(t)+w_{i4}(t))\sigma}{\sum N_i w_{i2}(t)}+\frac{h^2}{2}\sigma_K^2\mu''(t)+o(h^2).$$

Then, in order to calculate the asymptotic variance, we calculate:

$$\begin{aligned}
var(\tilde{S}_0) &= \frac{\sum N_i w_{i2}^2(t)}{h}(\|K\|^2 f(t)+o(1)),\\
var(\tilde{S}_1) &= \frac{\sum N_i w_{i2}^2(t)}{h}(\|K\|^2 t^2 f(t)+o(1)),\\
cov(\tilde{S}_0,\tilde{S}_1) &= \frac{\sum N_i w_{i2}^2(t)}{h}(\|K\|^2 tf(t)+o(1)),\\
var(R_0) &= \frac{1}{h}\sum N_i[(w_{i1}^2(t)+w_{i3}^2(t)+w_{i4}^2(t))\sigma^2+w_{i2}^2(t)\mu^2(t)\\
&\quad+2\sigma(w_{i1}(t)+w_{i4}(t))w_{i2}(t)\mu(t)]\|K\|^2 f(t),\\
cov(R_0,\tilde{S}_0) &= \frac{\sum N_i\left[w_{i2}^2(t)\mu(t)+\sigma(w_{i1}(t)+w_{i4}(t))w_{i2}(t)\right]}{h}\|K\|^2 f(t),\\
cov(R_0,\tilde{S}_1) &= \frac{\sum N_i\left[w_{i2}^2(t)\mu(t)+\sigma(w_{i1}(t)+w_{i4}(t))w_{i2}(t)\right]}{h}\|K\|^2 tf(t).
\end{aligned}$$

Therefore, by using the delta method, the asymptotic variance is

$$var(\tilde{\mu}(t))=\frac{\sum N_i(w_{i1}(t)+w_{i4}(t)-Cw_{i2}(t))^2+w_{i3}^2(t)}{h}\|K\|^2\frac{\sigma^2}{f(t)}$$

where $C=\frac{\sum N_i(w_{i1}(t)+w_{i4}(t))}{\sum N_i w_{i2}(t)}$.