Functional Data Analysis: Part I
Overview and Mean/Covariance Estimation

Jane-Ling Wang and Hans-Georg Müller

University of California Davis

August 26-27, 2016
Summer Course for COMPSTAT 2016
Oviedo, Spain

Outline

Outline

What is Functional Data?

- Functional Data: A sample of random functions, with one function per subject.

  - These functions can be curves (1D), images (2D or 3D), or higher dimension object data.

What is Functional Data?

- Functional Data: A sample of random functions,
  with one function per subject.

  - These functions can be curves (1D), images (2D or 3D), or higher dimension object data.

- Characteristics of functional data:

  (i) The atom of functional data is a “function”.
  (ii) They are ∞−dimensional data.

Example: Curve Data

- **Curve data**

  - real-valued functions defined on an interval $I \in \mathbb{R}$.

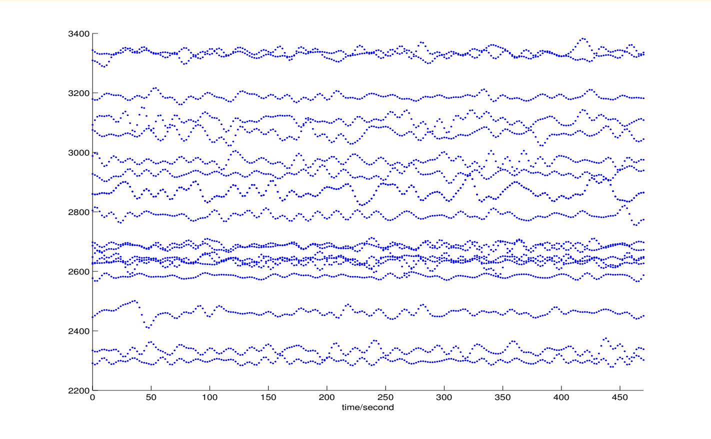
  - one curve per subject

  These curves are usually considered realizations of a stochastic process $X(t)$ in a **Hilbert space**, e.g.

  $L^2(I)$ or RKHS (reproducing kernel Hilbert space).

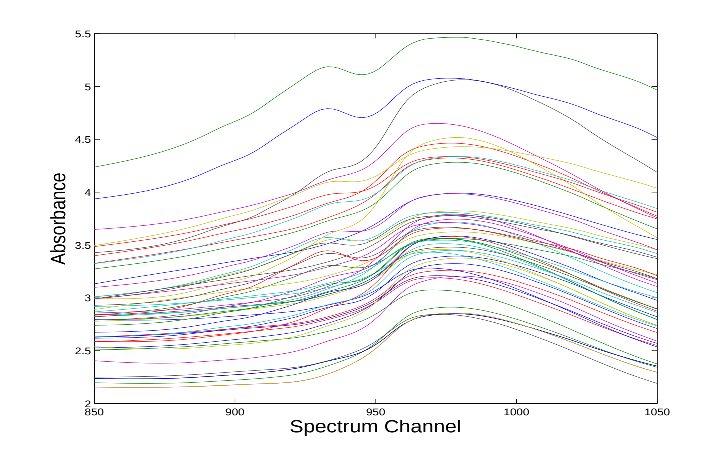
Examples of Functional Data

- fMRI data at a particular voxel for 20 subjects $\Longrightarrow n = 20.$



Examples of Functional Data

- Spectrum data for meat content - here the function is over the spectrum channels of n pieces of meat.



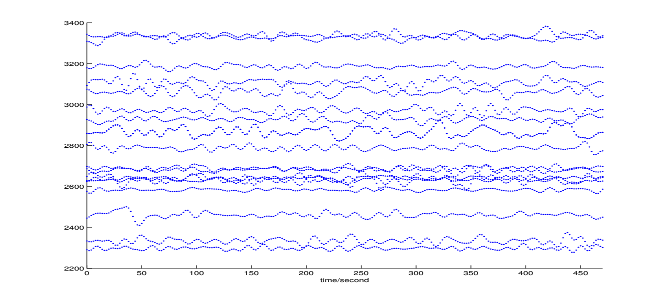
Additional Examples of Functional Data

- 10 minutes EKG recordings of 100 patients.
  $n = 100$
- Daily temperature recording in January at 240 locations.
  $n = 240$
- Daily reproduction (# of eggs) of 1000 female medflies (Mediterranean fruit flies) till death.
  $n = 1000$

Real/Observed Functional Data

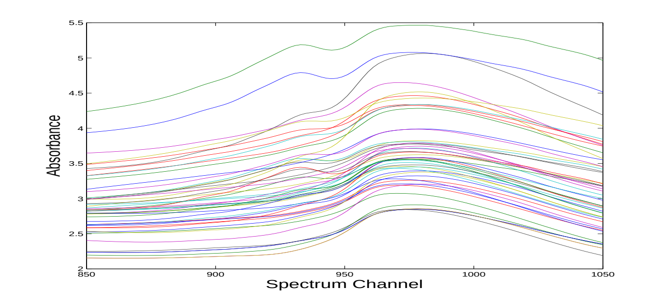
- In reality, functional data are recorded intensely on a time grid $\Longrightarrow$ high-dimensional data.

  * The fMRI data were recorded every two seconds for about 10 minutes (300 time points) $\Longrightarrow$ 300 dimensional data.



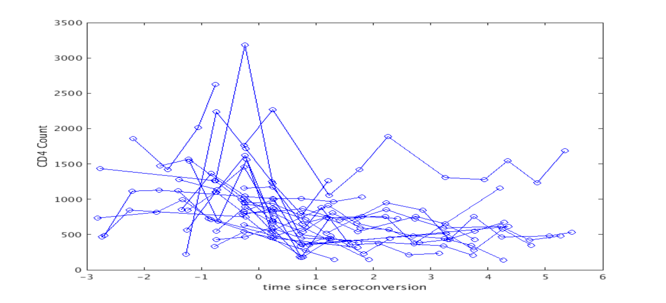
Real/Observed Functional Data

- The spectrum data were recorded at 100 frequency channels (hence 100-dim) and smoothed individually, i.e. pre-smoothed.



Longitudinal Data as Functional Data

- Longitudinal Data - Irregularly sampled functional data.
- They are often only a few measurements per subject, as in medical follow-up or social studies.



Longitudinal AIDS Data

- CD4 counts of 369 patients.

  $n_i = \#$ of repeated measurements for subject i,

  varies with subject.

  - An average of 6.44 measurement per subject.

Longitudinal AIDS Data

- CD4 counts of 369 patients.

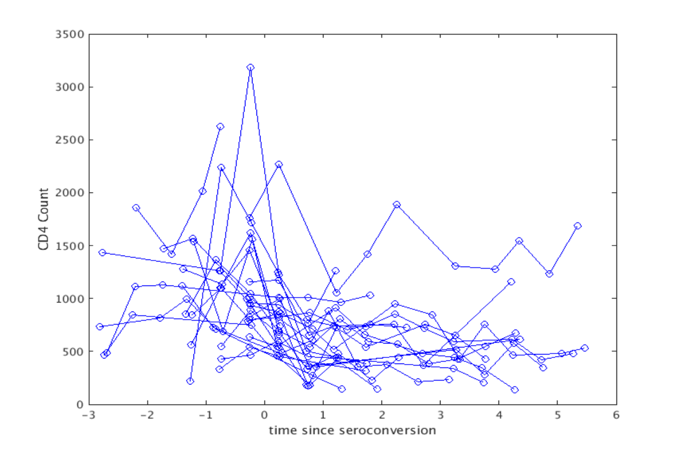
  $n_i = \#$ of repeated measurements for subject i,

  varies with subject.

  - An average of 6.44 measurement per subject.

- This results in longitudinal data with uneven $\#$ of measurements at irregular time-points.

CD4 Counts of First 25 Patients



Longitudinal vs Functional data

- Because of the intense recording, functional data have been customarily modeled with a nonparametric approach.

  - Often smoothness of the functions is assumed.

Longitudinal vs Functional data

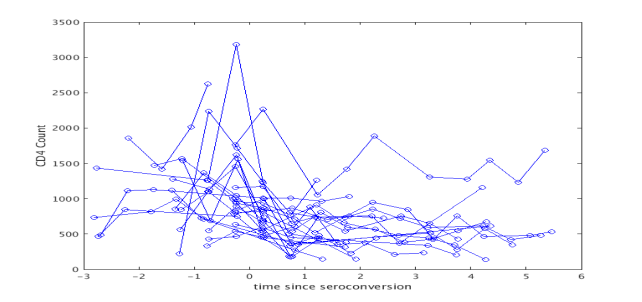
- Because of the intense recording, functional data have been customarily modeled with a nonparametric approach.

  - Often smoothness of the functions is assumed.

- Longitudinal data have traditionally been modeled by a parametric approach, such as a linear mixed-effects model.

  - However, it may not be easy to spot the pattern due to sparsity of and noise in the longitudinal data.

Longitudinal vs Functional data



- This motivates a data oriented nonparametric approach, which luckily is feasible under mild design conditions.

Longitudinal vs Functional data

- The analysis of functional data is termed “Functional Data Analysis” (FDA) by Ramsay (1982).

Longitudinal vs Functional data

- The analysis of functional data is termed "Functional Data Analysis" (FDA) by Ramsay (1982).

- Both longitudinal and functional data may be observed with noise (measurement errors).

  $\Longrightarrow$ the observed data for subject i might be

  $$Y_{ij} = X_i(t_{ij}) + e_{ij}, \quad j = 1, \ldots, n_i,$$

  where $X_i(t)$ is a smooth random function,
  e_{ij} **are independent** $\forall\ i, j$.

Longitudinal vs Functional data

- The analysis of functional data is termed "Functional Data Analysis" (FDA) by Ramsay (1982).

- Both longitudinal and functional data may be observed with noise (measurement errors).

  $\Longrightarrow$ the observed data for subject i might be

  $$Y_{ij} = X_i(t_{ij}) + e_{ij}, \quad j = 1, \ldots, n_i,$$

  where $X_i(t)$ is a smooth random function,
  e_{ij} **are independent** $\forall\ i, j$.

- A strength of the FDA approach is its ability to handle noise.

Summary

- There are the three types of functional data:

  (i) stochastic processes $\Longrightarrow$ ∞-dimensional data
  (ii) intense/dense functional data $\Longrightarrow$ high dimensional data
  (iii) sparse functional/longitudinal data $\Longrightarrow$ irregular dim. data.

Summary

- There are the three types of functional data:

  (i) stochastic processes $\Longrightarrow$ ∞-dimensional data
  (ii) intense/dense functional data $\Longrightarrow$ high dimensional data
  (iii) sparse functional/longitudinal data $\Longrightarrow$ irregular dim. data.

- The stochastic process $X_i(t)$ is assumed to be a continuous function.
  (But the observed data may contain noise, a.k.a. measurement error.)

Summary

- There are the three types of functional data:

  (i) stochastic processes $\Longrightarrow$ ∞-dimensional data
  (ii) intense/dense functional data $\Longrightarrow$ high dimensional data
  (iii) sparse functional/longitudinal data $\Longrightarrow$ irregular dim. data.

- The stochastic process $X_i(t)$ is assumed to be a continuous function.
  (But the observed data may contain noise, a.k.a. measurement error.)

- Nonparametric approaches are typically employed to all three types of functional data.

Questions

- What are the pros and cons for each type of functional data?

  - Which one is the easiest or hardest to handle?
  (We'll discuss how to handle them in the next section.)

Questions

- What are the pros and cons for each type of functional data?

  - Which one is the easiest or hardest to handle?
  (We'll discuss how to handle them in the next section.)

  Answer: More is better!

Questions

- What are the pros and cons for each type of functional data?

  - Which one is the easiest or hardest to handle?
  (We'll discuss how to handle them in the next section.)

  Answer: More is better!

- How did we get away with the curse of high dimensionality?

Questions

- What are the pros and cons for each type of functional data?

  - Which one is the easiest or hardest to handle?
  (We'll discuss how to handle them in the next section.)

  Answer: More is better!

- How did we get away with the curse of high dimensionality?

  Answer: We borrow information by smoothing!

  - We can do so because we have a natural ordering of the data.

  - Chen, Chen, M. and W. (2011) proposed a way to reorder multivariate data and convert high-dim data to functional data. (Stay tuned!)

Questions

- What are the pros and cons for each type of functional data?

  - Which one is the easiest or hardest to handle?
  (We'll discuss how to handle them in the next section.)

  Answer: More is better!

- How did we get away with the curse of high dimensionality?

  Answer: We borrow information by smoothing!

  - We can do so because we have a natural ordering of the data.

  - Chen, Chen, M. and W. (2011) proposed a way to reorder multivariate data and convert high-dim data to functional data. (Stay tuned!)

History of FDA

- The term “functional data” was coined by Ramsey (1982) but the term “curve data” was often used as well, e.g. by Gasser et al. (1984) for growth curves.

History of FDA

- The term "functional data" was coined by Ramsey (1982) but the term "curve data" was often used as well, e.g. by Gasser et al. (1984) for growth curves.

- The first implementation of functional principal component analysis (FPCA) were attributed to Rao (1958), where the growth data were recorded as multivariate data.

History of FDA

- The term "functional data" was coined by Ramsey (1982) but the term "curve data" was often used as well, e.g. by Gasser et al. (1984) for growth curves.

- The first implementation of functional principal component analysis (FPCA) were attributed to Rao (1958), where the growth data were recorded as multivariate data.

- The analysis of stochastic processes went even further back to Grenander (1950), Karhunen (1946), Loève (1946), and later include Kleffe (1973) and Dauxois and Poussé (1976).

History of FDA

- The handling of longitudinal data as sparse functional data was the focus in Yao et al. (2005), but nonparametric approaches for longitudinal data had already been employed by Shi et al. (1996), Staniswalis and Lee (1998), James et al. (2000), and Rice and Wu (2001).

History of FDA

- The handling of longitudinal data as sparse functional data was the focus in Yao et al. (2005), but nonparametric approaches for longitudinal data had already been employed by Shi et al. (1996), Staniswalis and Lee (1998), James et al. (2000), and Rice and Wu (2001).

- First generation functional data typically consist of a random sample, $X_1(t), \ldots, X_n(t)$, of **independent real-valued functions**.

History of FDA

- The handling of longitudinal data as sparse functional data was the focus in Yao et al. (2005), but nonparametric approaches for longitudinal data had already been employed by Shi et al. (1996), Staniswalis and Lee (1998), James et al. (2000), and Rice and Wu (2001).

- First generation functional data typically consist of a random sample, $X_1(t), \ldots, X_n(t)$, of **independent real-valued functions**.

- Next generation functional data refer to functional data that are part of complex data objects,
possibly multivariate, correlated, or involve images and shapes.

History of FDA

- The handling of longitudinal data as sparse functional data was the focus in Yao et al. (2005), but nonparametric approaches for longitudinal data had already been employed by Shi et al. (1996), Staniswalis and Lee (1998), James et al. (2000), and Rice and Wu (2001).

- First generation functional data typically consist of a random sample, $X_1(t), \ldots, X_n(t)$, of **independent real-valued functions**.

- Next generation functional data refer to functional data that are part of complex data objects,
possibly multivariate, correlated, or involve images and shapes.

- Brain and neuroimaging data are examples of next generation functional data.

Questions

- What is the difference between functional and time-series data?

Questions

- What is the difference between functional and time-series data?

- What is the benefit of functional data over multivariate data?

Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- Is there a curse of dimensionality for dense functional data?

Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- Is there a curse of dimensionality for dense functional data?

  - How did we get away with the curse?

Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- Is there a curse of dimensionality for dense functional data?

  - How did we get away with the curse?
- What are the three types of functional data?

Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- Is there a curse of dimensionality for dense functional data?

  - How did we get away with the curse?
- What are the three types of functional data?

  - What does **"dense"** or **"sparse"** mean?

  - We address this in Section 3, Zhang and W. (2016)

Questions

- What is the difference between functional and time-series data?
- What is the benefit of functional data over multivariate data?
- Is there a curse of dimensionality for dense functional data?

  \- How did we get away with the curse?
- What are the three types of functional data?

  \- What does **"dense"** or **"sparse"** mean?

  \- We address this in Section 3, Zhang and W. (2016)
- Review of FDA W., Chiou and Müller (2016)

End of Introduction

Outline

Mean and Covariance Function

Data $\{X_1, \ldots, X_n\}$ are i.i.d. copies of a random function $X(t)$:

- Mean function: $\mu(t) = E(X(t))$
- Covariance function: $\Sigma(s,t) = \text{cov}(X(s), X(t))$,
  where s & $t \in$ interval I.

Mean and Covariance Function

Data $\{X_1, \ldots, X_n\}$ are i.i.d. copies of a random function $X(t)$:

- Mean function: $\mu(t) = E(X(t))$

- Covariance function: $\Sigma(s,t) = \text{cov}(X(s), X(t))$,
  where s & $t \in$ interval I.

- Regular functional data - All subjects are measured at the same time grid, $t_1, \ldots, t_m$, often equally spaced $\Longrightarrow$ multivariate data.

  Irregular functional data - The measurement schedule for subject i is $t_{i1}, \ldots, t_{in_i} \Longrightarrow$ longitudinal data.

Estimation of Mean and Covariance Functions: Dense and Regular Functional Data

- For regular or fully observed functional data, the cross-sectional (sample) mean at the observed time t can be used to estimate the mean function.

Estimation of Mean and Covariance Functions: Dense and Regular Functional Data

- For regular or fully observed functional data, the cross-sectional (sample) mean at the observed time t can be used to estimate the mean function.

  - Such a mean estimate provides a $\sqrt{n}$-consistent (pointwise) estimate of the mean $\mu(t)$ even in the presence of measurement errors. (**WHY?**)

Estimation of Mean and Covariance Functions: Dense and Regular Functional Data

- For regular or fully observed functional data, the cross-sectional (sample) mean at the observed time t can be used to estimate the mean function.

  - Such a mean estimate provides a $\sqrt{n}$-consistent (pointwise) estimate of the mean $\mu(t)$ even in the presence of measurement errors. (**WHY?**)

- The cross-section mean can further be smoothed slightly to obtain a smooth mean estimate (This requires a dense/intense measurement schedule).

  - **Think how you should smooth to retain the $\sqrt{n}$-consistent of the smooth estimate** .

Estimation of Mean and Covariance Functions: Dense and Regular Functional Data

- Likewise, the sample covariance matrix is also $\sqrt{n}$-consistent and can be slightly smoothed to retain the $\sqrt{n}$-rate of consistency.

Estimation of Mean and Covariance Functions: Irregular Functional Data

- The cross-sectional approaches no longer work for irregular data.

Estimation of Mean and Covariance Functions: Irregular Functional Data

- The cross-sectional approaches no longer work for irregular data.

  - Smoothing is needed to estimate the mean function.

Estimation of Mean and Covariance Functions: Irregular Functional Data

- The cross-sectional approaches no longer work for irregular data.

  - Smoothing is needed to estimate the mean function.

  - The mean estimate could be $\sqrt{n}$-consistent (pointwise) for "**dense**" functional data but will have nonparametric rates otherwise.
  ("Dense" functional data $\Longleftrightarrow$ $\sqrt{n}$-rate of convergence is feasible.)

Estimation of Mean and Covariance Functions: Irregular Functional Data

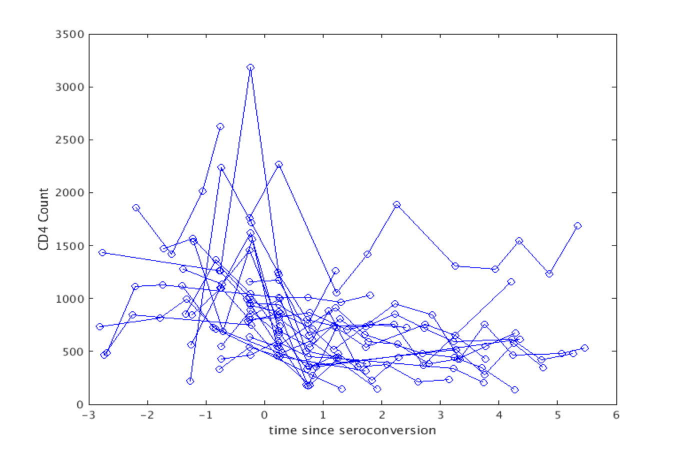
- The cross-sectional approaches no longer work for irregular data.

  - Smoothing is needed to estimate the mean function.

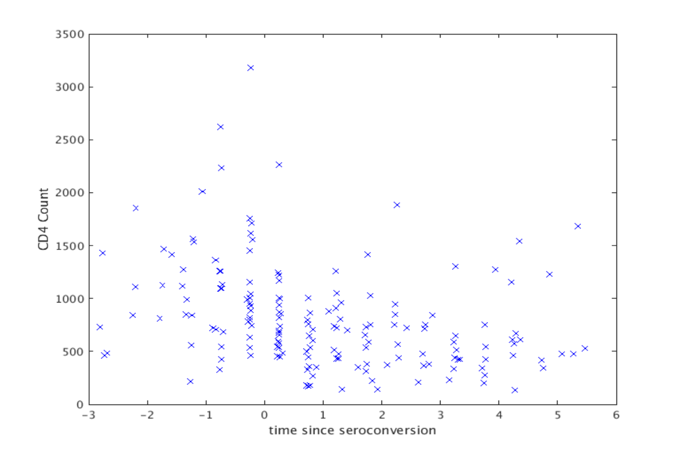
  - The mean estimate could be $\sqrt{n}$-consistent (pointwise) for "**dense**" functional data but will have nonparametric rates otherwise.
  ("Dense" functional data $\Longleftrightarrow$ $\sqrt{n}$-rate of convergence is feasible.)

- Likewise, 2D smoothing is needed to estimate the covariance function. and $\sqrt{n}$-consistency can be achieved for **dense** functional data.

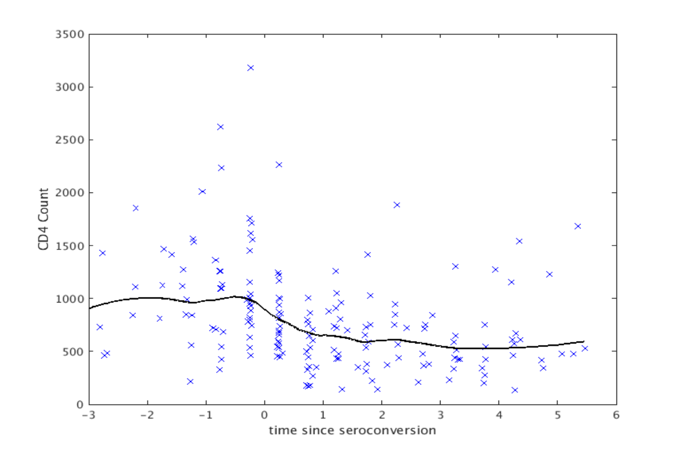
CD4 Counts of First 25 Patients



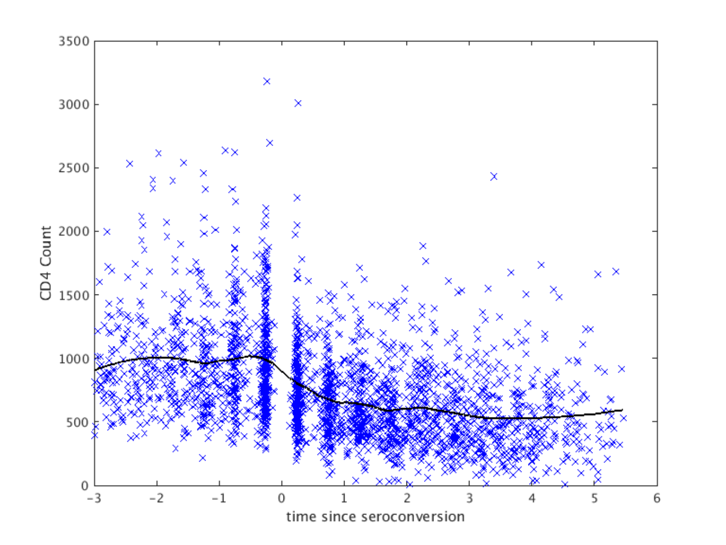
CD4 Counts of first 25 Patients



Mean Curve: CD4 Counts of first 25 Patients



Mean Curve: CD4 counts of all patients



Estimation of Mean Function

If we employ the local linear smoother, the estimate for the mean function is:

$$\hat{\mu}(t) = \hat{\beta}_0, \quad \text{where}$$
$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0,\beta_1}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \sum_{j=1}^{n_i} \Big[Y_{ij} - \beta_0 - \beta_1 (T_{ij} - t) \Big]^2 K_{h_\mu}(T_{ij} - t).$$

Remarks

- In addition to the local linear (or polynomial) smoother, any smoothing method, such as penalized splines, B-splines, Wavelets, and Fourier filtering, can be employed.

Remarks

- In addition to the local linear (or polynomial) smoother, any smoothing method, such as penalized splines, B-splines, Wavelets, and Fourier filtering, can be employed.

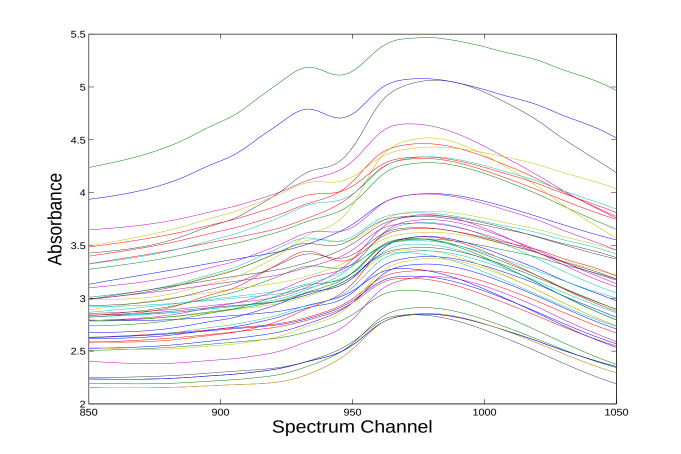
- These smoothing methods (scatter plot smoothers) can also be applied to dense data, whether regular or not, so a unified approach is feasible.

Remarks

- In addition to the local linear (or polynomial) smoother, any smoothing method, such as penalized splines, B-splines, Wavelets, and Fourier filtering, can be employed.
- These smoothing methods (scatter plot smoothers) can also be applied to dense data, whether regular or not, so a unified approach is feasible.
- Another common approach for dense data is to pre-smooth the data from each subject separately, then take the cross-section mean at each time point t.

Remarks

This is the case for the spectrum data.



Remarks

- Such a pre-smoothing typically aims at:

  (i) getting rid of the noise in the data,
  (ii) revealing the latent smooth curve for each subject.

  Hall, M. W. (2006) and Zhang and Chen (2007).

Remarks

- Such a pre-smoothing typically aims at:

  (i) getting rid of the noise in the data,
  (ii) revealing the latent smooth curve for each subject.

  Hall, M. W. (2006) and Zhang and Chen (2007).

- However, both depend on the amount of smoothing and design of the dense data, so whether the mission has been accomplished is unclear.

- Because of this uncertainty and because pre-smoothing alters the data, we prefer not to adopt a pre-smoothing approach.

End of Estimation of Mean Function

Estimation of Covariance Function

- Our target is $\Sigma(s,t) = \text{cov}(X(s), X(t))$, but we do not observe X.

Estimation of Covariance Function

- Our target is $\Sigma(s,t) = \text{cov}(X(s), X(t))$, but we do not observe X.

- Observe $Y_{ij} = Y_i(t_{ij}) = X_{ij} + e_{ij}$, where $\text{var}(e_{ij}) = \sigma^2(t_{ij})$.

  $\Longrightarrow \text{cov}(Y(s), Y(t)) = \text{cov}(X(s), X(t))$, when $s \neq t$,

  $$\text{var}(Y(t)) = \text{cov}(Y(t), Y(t)) = \text{cov}(X(t), X(t)) + \sigma^2(t).$$

Estimation of Covariance Function

- Our target is $\Sigma(s,t) = \text{cov}(X(s), X(t))$, but we do not observe X.

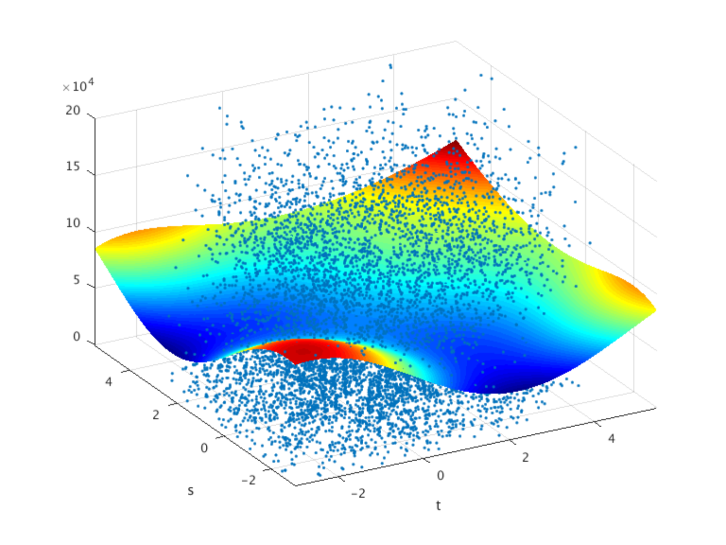
- Observe $Y_{ij} = Y_i(t_{ij}) = X_{ij} + e_{ij}$, where $\text{var}(e_{ij}) = \sigma^2(t_{ij})$.

  $\Longrightarrow \text{cov}(Y(s), Y(t)) = \text{cov}(X(s), X(t))$, when $s \neq t$,

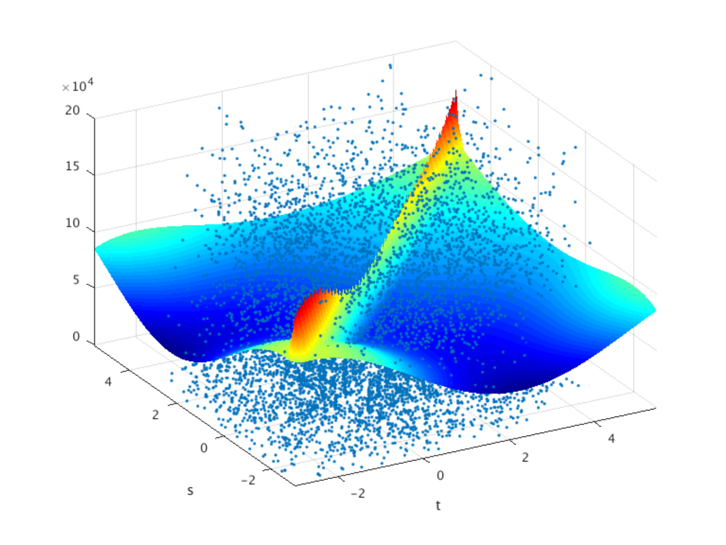
  $$\text{var}(Y(t)) = \text{cov}(Y(t), Y(t)) = \text{cov}(X(t), X(t)) + \sigma^2(t).$$

- This means we need to handle the diagonal part of $\text{cov}(Y)$, which corresponds to the variance of Y, differently from the rest of $\text{cov}(Y)$.

Covariance Surface of $X(t)$

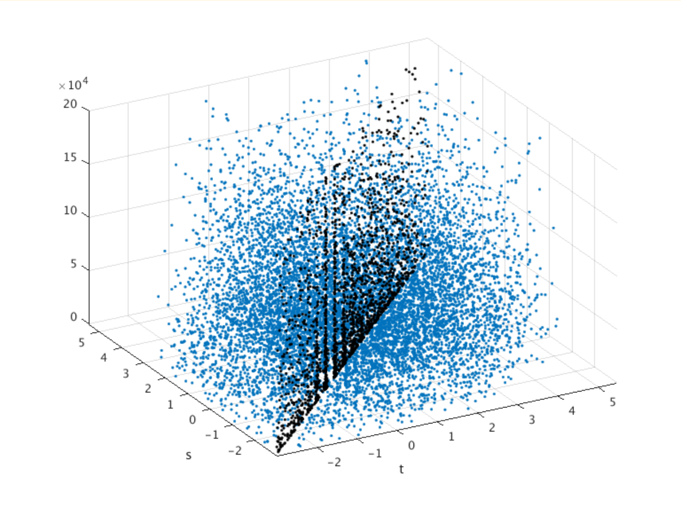


Covariance of $Y(t)$



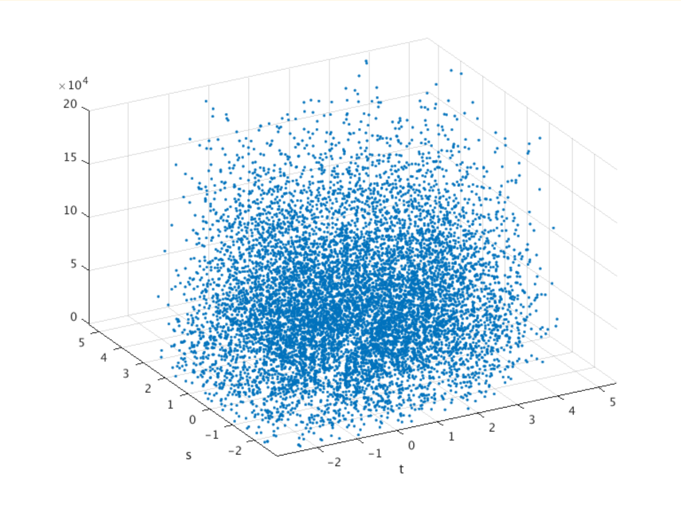
Raw Covariance Plot: Diagonal Data in Black

$$[Y(t_{ij}) - \hat{\mu}(t_{ij})][Y(t_{ik}) - \hat{\mu}(t_{ik})],\ \forall\ i, j, k$$

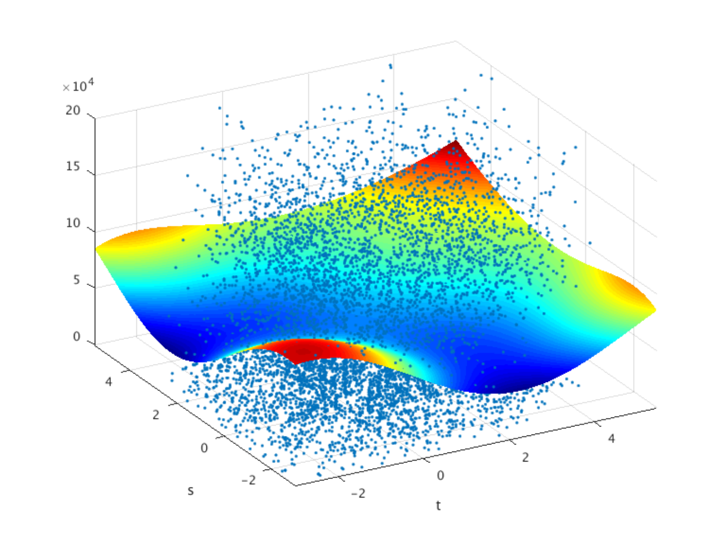


Raw Covariance with Diagonal Data Removed

$$[Y(t_{ij}) - \hat{\mu}(t_{ij)}][Y(t_{ik}) - \hat{\mu}(t_{ik})],\ \forall\ i, j \neq k$$

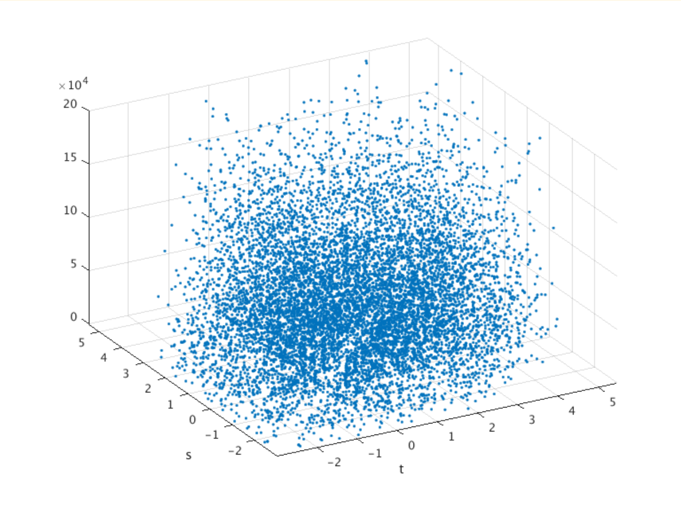


2D Smoother to Estimate the Covariance of $X(t)$

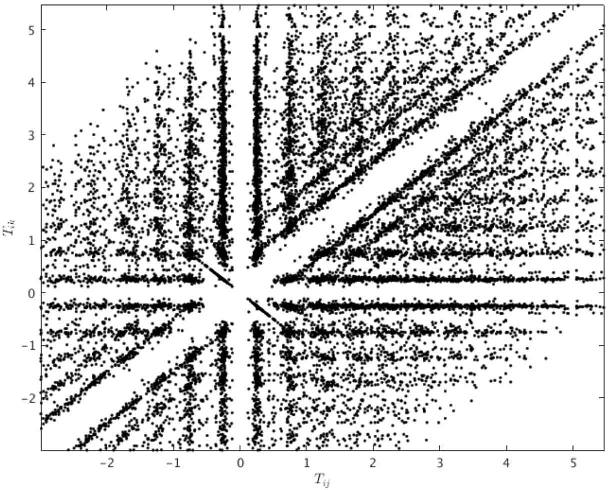


Raw Covariance with Diagonal Data Removed

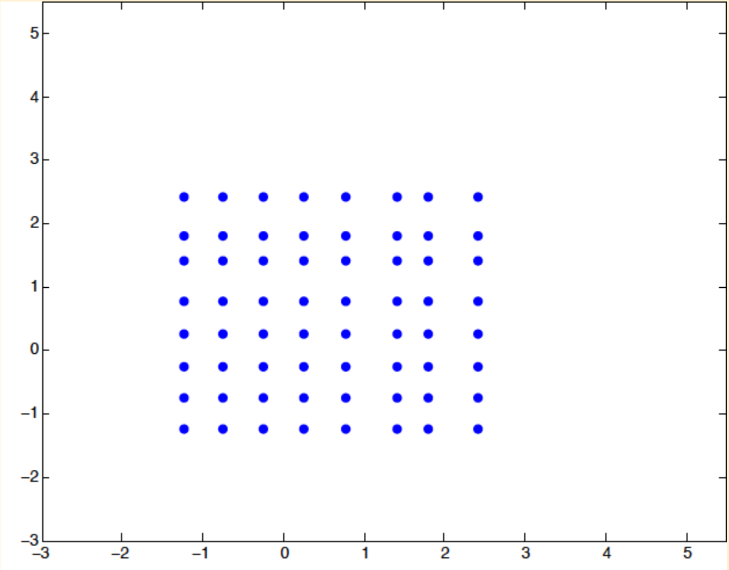
$$[Y(t_{ij}) - \hat{\mu}(t_{ij)}][Y(t_{ik}) - \hat{\mu}(t_{ik})], \ \forall \ i, j \neq k$$



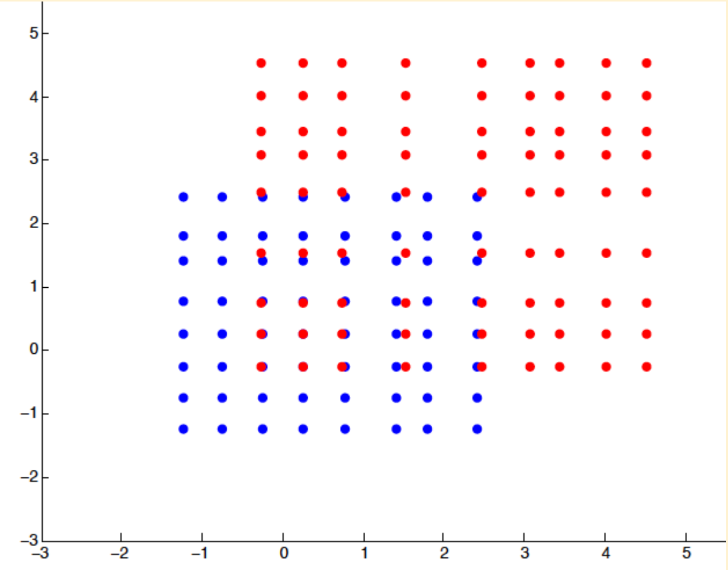
Design Plot for Covariance: $(t_{ij}, t_{ik}),\ \forall\ i, j \neq k$



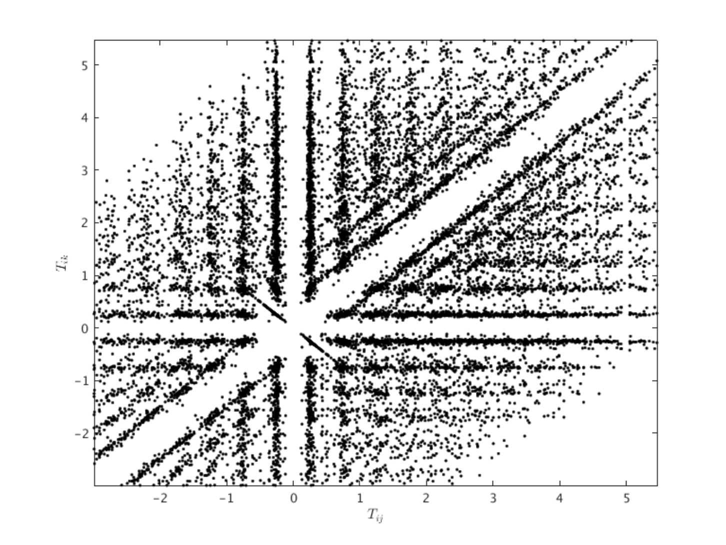
Design Plot for One Subject: $(t_{ij}, t_{ik}), \forall j, k$



Design Plot for Two Subjects: $(t_{ij}, t_{ik}), \forall j, k$



Design Plot for All Subjects: $(t_{ij}, t_{ik}),\ \forall\ i, j, k$



Estimated Covariance Surface of $X(t)$

Let $D_{ijl} = (Y_{ij} - \hat{\mu}(t_{ij}))(Y_{il} - \hat{\mu}(t_{il}))$, be the raw covariances.

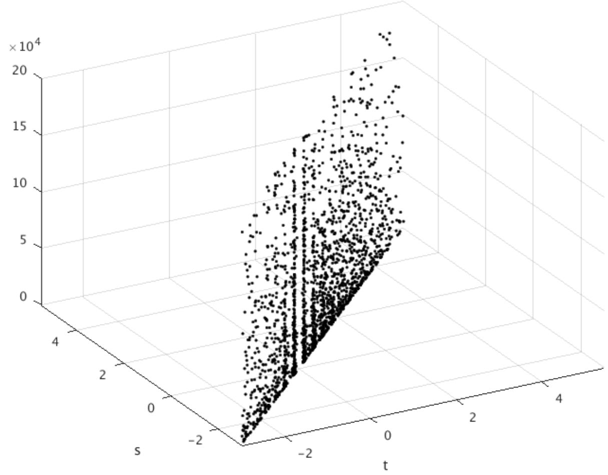
Employing a local linear smoother, the estimate for $\Sigma(s, t)$ is:

$\hat{\Sigma}(s, t) = \hat{\beta}_0$, where

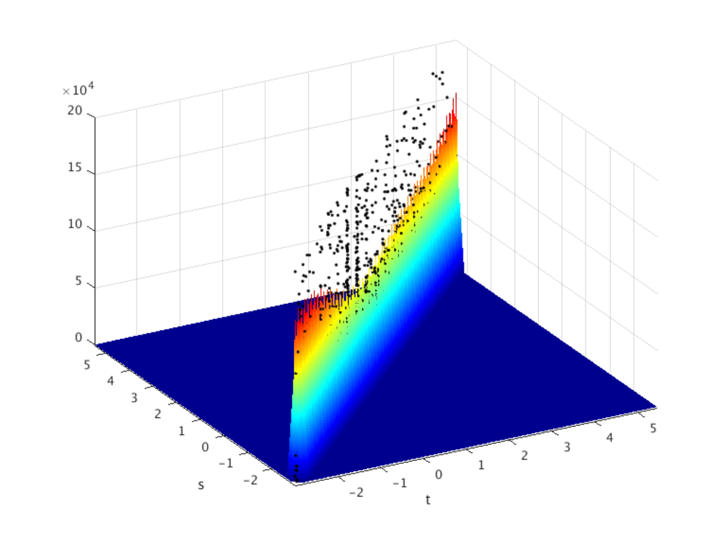
$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \underset{\beta_0, \beta_1, \beta_2}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{1 \le j \ne l \le n_i} \Big[D_{ijl} - \beta_0$$

$$- \beta_1(t_{ij} - s) - \beta_2(t_{il} - t) \Big]^2 K_{h_\Sigma}(t_{ij} - s) K_{h_\Sigma}(t_{il} - t).$$

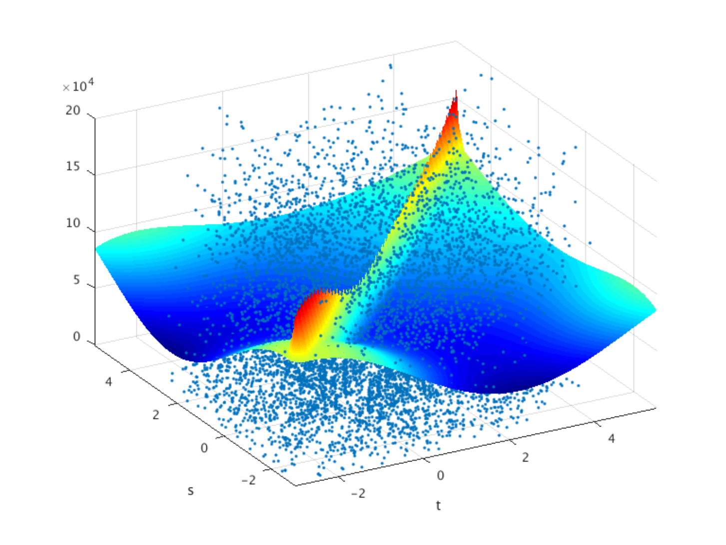
Raw Variance Plot: $[Y(t_{ij}) - \mu(t_{ij})]^2, \ \forall\ i, j$



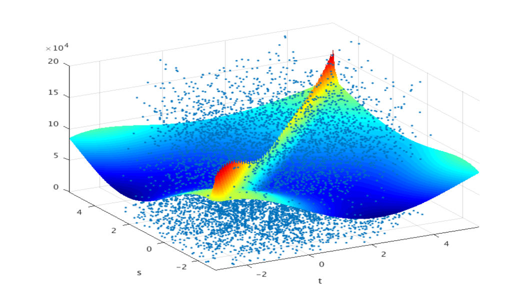
Estimated Variance function of $Y(t)$



Estimated Covariance & Variance of $Y(t)$

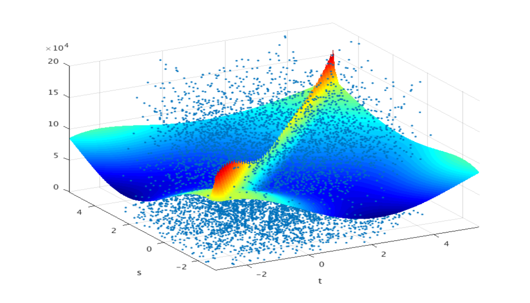


Estimates of $\sigma^2(t)$: Variance of Measurement Errors



$\sigma^2(t) = \mathsf{var}(Y(t)) - \mathsf{var}(X(t)).$

Estimates of $\sigma^2(t)$: Variance of Measurement Errors



$\sigma^2(t) = \mathsf{var}(Y(t)) - \mathsf{var}(X(t))$.

When $\sigma^2(t) = \sigma^2$ for all t, one can estimate σ^2 by $\int_I \hat{\sigma}^2(t)dt$.

Due to boundary effects, PACE replaces I by a sub-interval.

PACE (Matlab) Package for FDA: http://www.stat.ucdavis.edu/PACE/

fdapace: R-package of PACE

- An R-package, **fdapace**, is now available in CRAN.

  - But it only contains a few core functions in PACE.

fdapace: R-package of PACE

- An R-package, **fdapace**, is now available in CRAN.

  - But it only contains a few core functions in PACE.

- fdapace is still under construction - Feedback welcome!

End of Covariance Estimation

Outline

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- So far, we have used "**scatter plot smoothers**" to estimate the mean and covariance functions.

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- So far, we have used "**scatter plot smoothers**" to estimate the mean and covariance functions.

  \- This is the method of PACE (a Matlab package for functional data), first proposed in Yao, Müller and W. (2005).
  (An R package, fdapace, has been released recently.)

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- So far, we have used "**scatter plot smoothers**" to estimate the mean and covariance functions.

  - This is the method of PACE (a Matlab package for functional data), first proposed in Yao, Müller and W. (2005).
  (An R package, fdapace, has been released recently.)

  - It works for any type of sampling plan: regular, dense, or not.

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- So far, we have used "**scatter plot smoothers**" to estimate the mean and covariance functions.

  - This is the method of PACE (a Matlab package for functional data), first proposed in Yao, Müller and W. (2005).
  (An R package, fdapace, has been released recently.)

  - It works for any type of sampling plan: regular, dense, or not.

  - It assigns the same weight to each observation, so a subject with more measurements (large n_i) receives a larger total weight.

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- So far, we have used "**scatter plot smoothers**" to estimate the mean and covariance functions.

  - This is the method of PACE (a Matlab package for functional data), first proposed in Yao, Müller and W. (2005).
  (An R package, fdapace, has been released recently.)

  - It works for any type of sampling plan: regular, dense, or not.

  - It assigns the same weight to each observation, so a subject with more measurements (large n_i) receives a larger total weight.

- Another way to assign weights is to give equal weights, $\frac{1}{n}$, to all subjects as proposed in Li and Hsing (2010).

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- **Which weight assignment is better?**

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- **Which weight assignment is better?**
- In Zhang and W. (2016) we present asymptotic properties of these estimators and define what “dense” functional data means.

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- **Which weight assignment is better?**
- In Zhang and W. (2016) we present asymptotic properties of these estimators and define what "dense" functional data means.
- A referee requested that we develop a general theory using a general weight function, which covers both weighing schemes.

Theory: Mean and Covariance Estimation (Zhang and W., 2016)

- **Which weight assignment is better?**
- In Zhang and W. (2016) we present asymptotic properties of these estimators and define what “dense” functional data means.
- A referee requested that we develop a general theory using a general weight function, which covers both weighing schemes.
- As a result, an optimal weighing scheme based on convex combinations of the two weights was developed.

Estimation of Mean Function

- Equal weight per observation (Yao, Müller and W., 2005)

$$\hat{\mu}_{obs}(t) = \hat{\beta}_0 \text{ where}$$

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[Y_{ij} - \beta_0 - \beta_1 (T_{ij} - t) \right]^2 K_{h_\mu}(T_{ij} - t).$$

Estimation of Mean Function

- Equal weight per observation (Yao, Müller and W., 2005)

$$\hat{\mu}_{obs}(t) = \hat{\beta}_0 \text{ where}$$

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left[Y_{ij} - \beta_0 - \beta_1 (T_{ij} - t) \right]^2 K_{h_\mu}(T_{ij} - t).$$

- Equal weight per subject (Li and Hsing, 2010)

$$\hat{\mu}_{sub}(t) = \hat{\beta}_0 \text{ where}$$

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \mathbf{\frac{1}{n_i}} \sum_{j=1}^{n_i} \left[Y_{ij} - \beta_0 - \beta_1 (T_{ij} - t) \right]^2 K_{h_\mu}(T_{ij} - t).$$

Estimation of Mean Function

- A general weighing scheme (Zhang and W,, 2016),

  \- weight $\mathbf{w_i}$ for subject i, where $\sum_{i=1}^{n} n_i w_i = 1$.

  $\hat{\mu}_{obs}(t) = \hat{\beta}_0$ where

$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \mathbf{w_i} \sum_{j=1}^{n_i} \left[Y_{ij} - \beta_0 - \beta_1 (T_{ij} - t) \right]^2 K_{h_\mu}(T_{ij} - t).$$

Estimation of Mean Function

- A general weighing scheme (Zhang and W,, 2016),

  - weight $\mathbf{w_i}$ for subject i, where $\sum_{i=1}^{n} n_i w_i = 1$.

  $\hat{\mu}_{obs}(t) = \hat{\beta}_0$ where

  $$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}} \quad \sum_{i=1}^{n} \mathbf{w_i} \sum_{j=1}^{n_i} \Big[Y_{ij} - \beta_0 - \beta_1 (T_{ij} - t) \Big]^2 K_{h_\mu}(T_{ij} - t).$$

- The OBS scheme uses $w_i = \frac{1}{\sum_{i=1}^{n} n_i} \Rightarrow$ the same for all subjects.

  The SUBJ scheme uses $w_i = \frac{1}{n n_i}$.

Important notation

$n =$ Sample size, $n_i = \#$ observations of subject i.

Important notation

$n =$ Sample size, $n_i = \#$ observations of subject i.

- OBS: arithmetic mean of n_i and of n_i^2,

$$\bar{N} = \frac{1}{n}\sum_{i=1}^{n} n_i, \quad \bar{N}_{S2} = \frac{1}{n}\sum_{i=1}^{n} n_i^2.$$

Important notation

$n =$ Sample size, $n_i = \#$ observations of subject i.

- OBS: arithmetic mean of n_i and of n_i^2,

$$\bar{N} = \frac{1}{n}\sum_{i=1}^{n} n_i, \quad \bar{N}_{S2} = \frac{1}{n}\sum_{i=1}^{n} n_i^2.$$

- SUB: harmonic mean of n_i,

$$\bar{N}_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{n_i}}, \quad \frac{\bar{N}_{S2}}{(\bar{N})^2} \to 1.$$

Asymptotic Normality

- OBS:

$$[\Gamma_{obs}(t)]^{-1/2}\{\hat{\mu}_{obs}(t) - \mu(t) - \underbrace{\frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)}_{\text{asymptotic bias}}\} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{where}$$

$$\underbrace{\Gamma_{obs}(t)}_{\text{asymptotic variance}} = \|K\|^2\frac{\Sigma(t,t)+\sigma^2}{n\bar{N}h_\mu f(t)} + \frac{(\bar{N}_{S2}-\bar{N})}{n(\bar{N})^2}\Sigma(t,t).$$

Asymptotic Normality

- OBS:

$$[\Gamma_{obs}(t)]^{-1/2}\{\hat{\mu}_{obs}(t) - \mu(t) - \underbrace{\frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)}_{\text{asymptotic bias}}\} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{where}$$

$$\underbrace{\Gamma_{obs}(t)}_{\text{asymptotic variance}} = \|K\|^2\frac{\Sigma(t,t)+\sigma^2}{n\bar{N}h_\mu f(t)} + \frac{(\bar{N}_{S2}-\bar{N})}{n(\bar{N})^2}\Sigma(t,t).$$

- SUB

$$[\Gamma_{sub}(t)]^{-1/2}\{\hat{\mu}_{sub}(t) - \mu(t) - \underbrace{\frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)}_{\text{asymptotic bias}}\} \xrightarrow{d} \mathcal{N}(0,1), \quad \text{where}$$

$$\underbrace{\Gamma_{sub}(t)}_{\text{asymptotic variance}} = \|K\|^2\frac{\Sigma(t,t)+\sigma^2}{n\bar{N}_H h_\mu f(t)} + \frac{1}{n}\left(1-\frac{1}{\bar{N}_H}\right)\Sigma(t,t)$$

Rates of Convergence

1. Non-dense: Slower than $\sqrt{n}$-rate.

   Sparse data $=$ finite n_i, is a special case.

Rates of Convergence

1. Non-dense: Slower than $\sqrt{n}$-rate.
   Sparse data $=$ finite n_i, is a special case.

2. Dense:$\sqrt{n}$-rate, with asymptotic bias
   (between non- and parametric paradigm)

Rates of Convergence

1. Non-dense: Slower than $\sqrt{n}$-rate.
   Sparse data $=$ finite n_i, is a special case.

2. Dense:$\sqrt{n}$-rate, with asymptotic bias
   (between non- and parametric paradigm)

3. Ultra-dense: $\sqrt{n}$-rate, no asymptotic bias
   (Parametric paradigm)

Partition of Functional Data: OBS

Assume $\limsup_n (\bar{N}_{S2})/(\bar{N})^2 < \infty$.

1. **Non-Dense data**: When $\bar{N}/n^{1/4} \to 0$ and $h_\mu \asymp (n\bar{N})^{-1/5}$,

$$\sqrt{n\bar{N}h_\mu}\,\left[\hat{\mu}_{obs}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)\right] \xrightarrow{d} \mathcal{N}\left(0, \|K\|^2\frac{\Sigma(t,t)+\sigma^2}{f(t)}\right).$$

2. **Dense data**: When $\bar{N}/n^{1/4} \to C$ and $h_\mu/n^{-1/4} \to C_1$ where $0 < C, C_1 < \infty$,

$$\sqrt{n\frac{(\bar{N})^2}{\bar{N}_{S2}}}\,\left[\hat{\mu}_{obs}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)\right] \xrightarrow{d} \mathcal{N}\left(0, \|K\|^2\frac{\Sigma(t,t)+\sigma^2}{f(t)\cdot C_1} + \Sigma(t,t)\right).$$

3. **Ultra-Dense data**: When $\bar{N}/n^{1/4} \to \infty$, $h_\mu = o(n^{-1/4})$, and $h_\mu\bar{N} \to \infty$,

$$\sqrt{n\frac{(\bar{N})^2}{\bar{N}_{S2}}}\,\left[\hat{\mu}_{obs}(t) - \mu(t)\right] \xrightarrow{d} \mathcal{N}\left(0, \Sigma(t,t)\right).$$

Partition of Functional Data: SUB

1. **Non-Dense data**: When $\bar{N}_H/n^{1/4} \to 0$ and $h_\mu \asymp (n\bar{N}_H)^{-1/5}$,

$$\sqrt{n\bar{N}_H h_\mu}\,\left[\hat{\mu}_{sub}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)\right] \xrightarrow{d} \mathcal{N}\left(0, \|K\|^2\frac{\Sigma(t,t)+\sigma^2}{f(t)}\right).$$

2. **Dense data**: When $\bar{N}_H/n^{1/4} \to C$ and $h_\mu/n^{-1/4} \to C_1$ where $0 < C, C_1 < \infty$,

$$\sqrt{n}\,\left[\hat{\mu}_{sub}(t) - \mu(t) - \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)\right] \xrightarrow{d} N\left(0, \|K\|^2\frac{\Sigma(t,t)+\sigma^2}{f(t)C\cdot C_1} + \Sigma(t,t)\right).$$

3. **Ultra-Dense data**: When $\bar{N}_H/n^{1/4} \to \infty$, $h_\mu = o(n^{-1/4})$, and $h_\mu\bar{N}_H \to \infty$,

$$\sqrt{n}\,\left[\hat{\mu}_{sub}(t) - \mu(t)\right] \xrightarrow{d} \mathcal{N}\left(0, \Sigma(t,t)\right).$$

- This shows $\hat{\mu}_{sub}(t)$ is asymptotically equivalent to the sample mean when the true curves $X_i(t)$ can be observed without any measurement errors.

Comparison of Two Schemes

1. **Asymptotic bias**: both $= \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)$.

Comparison of Two Schemes

1. **Asymptotic bias**: both $= \frac{1}{2}h_\mu^2\sigma_K^2\mu^{(2)}(t)$.
2. **Asymptotic variance**:
   - **Non-dense data**: var$(\hat{\mu}_{obs}) \leq$ var$(\hat{\mu}_{sub})$, so OBS is more efficient.
   - **Ultra-dense data**: var$(\hat{\mu}_{obs}) \geq$ var$(\hat{\mu}_{sub})$, so SUB is more efficient.

L^2 Convergence

- Equal weight per observation:

$$\|\hat{\mu}_{obs} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\left(\frac{\bar{N}_{S2}}{(\bar{N})^2} + \frac{1}{\bar{N}h_\mu}\right)\frac{1}{n}}\right).$$

- Equal weight per subject:

$$\|\hat{\mu}_{sub} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\left(1 + \frac{1}{\bar{N}_H h_\mu}\right)\frac{1}{n}}\right)$$

Uniform Convergence

- Equal weight per observation:

$$\sup_{t\in[0,1]} \left|\hat{\mu}_{obs}(t) - \mu(t)\right| = O\left(h_\mu^2 + \sqrt{\left(\frac{\bar{N}_{S2}}{(\bar{N})^2} + \frac{1}{\bar{N}h_\mu}\right)\frac{\log(n)}{n}}\right) \quad a.s.$$

- Equal weight per subject:

$$\sup_{t\in[0,1]} \left|\hat{\mu}_{sub}(t) - \mu(t)\right| = O\left(h_\mu^2 + \sqrt{\left(1 + \frac{1}{\bar{N}_H h_\mu}\right)\frac{\log(n)}{n}}\right) \quad a.s.$$

Optimal Weighing Scheme

- Optimal weights for a convex combination of the two estimators exists.

$$w_i = \alpha \frac{1}{n\bar{N}} + (1-\alpha)\frac{1}{nN_i}$$

Optimal Weighing Scheme

- Optimal weights for a convex combination of the two estimators exists.

$$w_i = \alpha \frac{1}{n\bar{N}} + (1-\alpha)\frac{1}{nN_i}$$

- Optimal weights:

$$\alpha^* = \frac{c_{n2}}{c_{n1} + c_{n2}},$$

where c_{n1} and c_{n2} are respectively the asymptotic variance of $\hat{\mu}_{obj}$ and $\hat{\mu}_{sub}$.

Optimal Weighing Scheme

- Let $\hat{\mu}_{\alpha^*}$ be the estimate with the optimal weights, $w_i = \alpha^* \frac{1}{n\bar{N}} + (1-\alpha^*)\frac{1}{nN_i}$.

  $$\Rightarrow \|\hat{\mu}_{\alpha^*} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\frac{c_{n1}c_{n2}}{c_{n1}+c_{n2}}}\right).$$

Optimal Weighing Scheme

- Let $\hat{\mu}_{\alpha^*}$ be the estimate with the optimal weights, $w_i = \alpha^* \frac{1}{n\bar{N}} + (1-\alpha^*)\frac{1}{nN_i}$.

  $$\Rightarrow \|\hat{\mu}_{\alpha^*} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\frac{c_{n1}c_{n2}}{c_{n1}+c_{n2}}}\right).$$

- Since $\frac{c_{n1}c_{n2}}{c_{n1}+c_{n2}} \le \min\{c_{n1}, c_{n2}\}$, the rate for $\hat{\mu}_{\alpha^*}$ is always at least as good as those for $\hat{\mu}_{obj}$ and $\hat{\mu}_{sub}$.

Optimal Weighing Scheme

- Let $\hat{\mu}_{\alpha^*}$ be the estimate with the optimal weights, $w_i = \alpha^* \frac{1}{n\bar{N}} + (1-\alpha^*)\frac{1}{nN_i}$.

  $$\Rightarrow \|\hat{\mu}_{\alpha^*} - \mu\|_2 = O_p\left(h_\mu^2 + \sqrt{\frac{c_{n1}c_{n2}}{c_{n1}+c_{n2}}}\right).$$

- Since $\frac{c_{n1}c_{n2}}{c_{n1}+c_{n2}} \le \min\{c_{n1}, c_{n2}\}$, the rate for $\hat{\mu}_{\alpha^*}$ is always at least as good as those for $\hat{\mu}_{obj}$ and $\hat{\mu}_{sub}$.

- Simulation results in Zhang and W. (2016) show superior performance of $\hat{\mu}_{\alpha^*}$.

End of Theory for Mean Function

Covariance Estimate with General Weights

- Let $D_{ijl} = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{il} - \hat{\mu}(T_{il}))$ be the raw covariance based on a mean estimate $\hat{\mu}$.

- Let $\mathbf{v_i}$ be the weight attached to each observation for the ith subject with $\sum_{i=1}^{n} n_i(n_i - 1)\mathbf{v_i} = 1$.

Covariance Estimate with General Weights

- Let $D_{ijl} = (Y_{ij} - \hat{\mu}(T_{ij}))(Y_{il} - \hat{\mu}(T_{il}))$ be the raw covariance based on a mean estimate $\hat{\mu}$.

- Let $\mathbf{v_i}$ be the weight attached to each observation for the ith subject with $\sum_{i=1}^{n} n_i(n_i - 1)\mathbf{v_i} = 1$.

- A general covariance estimator based on the weights $\mathbf{v_i}$ is: $\hat{\Sigma}(s,t) = \hat{\beta}_0$, where

$$(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2) = \underset{\beta_0,\beta_1,\beta_2}{\operatorname{argmin}} \sum_{i=1}^{n} \mathbf{v_i} \sum_{1 \le j \ne l \le n_i} \Big[D_{ijl} - \beta_0 - \beta_1(T_{ij} - s) - \beta_2(T_{il} - t) \Big]^2 K_{h_\Sigma}(T_{ij} - s) K_{h_\Sigma}(T_{il} - t)$$

Covariance Estimators: Two Special Cases

1. $D_{ijl} = (Y_{ij} - \hat{\mu}_{obs}(T_{ij}))(Y_{il} - \hat{\mu}_{obs}(T_{il}))$ and $\mathbf{v_i} = \frac{1}{\sum_{i=1}^{n} n_i(n_i-1)}$ $\Longrightarrow$ OBS scheme.

2. $D_{ijl} = (Y_{ij} - \hat{\mu}_{sub}(T_{ij}))(Y_{il} - \hat{\mu}_{sub}(T_{il}))$ and $v_i = 1/(n_i(n_i-1))$ $\Longrightarrow$ SUB scheme.

Asymptotic Normality

Let

$$
\begin{aligned}
V_1(s,t) &= \mathrm{Var}[(Y_1-\mu(T_1))(Y_2-\mu(T_2)) \mid T_1=s, T_2=t]; \\
V_2(s,t) &= \mathrm{Cov}\left([Y_1-\mu(T_1)][Y_2-\mu(T_2)], [Y_1-\mu(T_1)][Y_3-\mu(T_3)] \mid T_1=s, T_2=t, T_3 \right. \\
V_3(s,t) &= \mathrm{Cov}\left([Y_1-\mu(T_1)][Y_2-\mu(T_2)], [Y_3-\mu(T_3)][Y_4-\mu(T_4)] \mid T_1=s, T_2=t, T_3 \right.
\end{aligned}
$$

$$
\Gamma_\gamma^{-1/2}\left[\hat{\gamma}(s,t)-\gamma(s,t)-\frac{1}{2}h_\gamma^2\sigma_K^2\left(\frac{\partial^2\gamma}{\partial s^2}(s,t)+\frac{\partial^2\gamma}{\partial t^2}(s,t)\right)+o_p(h_\gamma^2)\right] \xrightarrow{d} \mathcal{N}(0,1),
$$

where

$$
\begin{aligned}
\Gamma_\gamma = [1+\mathbf{I}(\mathbf{s}=\mathbf{t})]\Bigg[&\frac{\sum_{i=1}^n n_i(n_i-1)v_i^2}{h_\gamma^2}\|K\|^4\frac{V_1(s,t)}{f(s)f(t)}+\frac{\sum_{i=1}^n n_i(n_i-1)(n_i-2)v_i^2}{h_\gamma} \\
&\times\frac{f(s)V_2(t,s)+f(t)V_2(s,t)}{f(s)f(t)}\Bigg]+\left[\sum_{i=1}^n n_i(n_i-1)(n_i-2)(n_i-3)v_i^2\right]V_3(s,t),
\end{aligned}
$$

and $\mathbf{I}(\cdot)$ is the indicator function.

Summary for Covariance Estimation

Unified asymptotic normality for general weight functions:

1. Three partitions: non-dense, dense, and ultra-dense.

2. Two special weighing schemes:
   $\hat{\Sigma}_{obs}$ more efficient for non-dense data;
   $\hat{\Sigma}_{sub}$ more efficient for ultra-dense data.

Summary for Covariance Estimation

Unified asymptotic normality for general weight functions:

1. Three partitions: non-dense, dense, and ultra-dense.

2. Two special weighing schemes:
   $\hat{\Sigma}_{obs}$ more efficient for non-dense data;
   $\hat{\Sigma}_{sub}$ more efficient for ultra-dense data.

3. Discontinuity of the asymptotic variance of the covariance estimates:

   Asymptotic variance expressions are different between $s = t$ and $s \neq t$.

Discontinuity of the Asymptotic Variance

Technical Explanation:

$E[K_h(T-t)K_h(T-s)] = 0$, for $s \neq t$ when $h \to 0$ for K on an interval;

$E[K_h(T-t)K_h(T-s)] = \|K\|^2 f(t)/h + o(1/h)$, for $s = t$.

Discontinuity of the Asymptotic Variance

Technical Explanation:

$E[K_h(T-t)K_h(T-s)] = 0$, for $s \neq t$ when $h \to 0$ for K on an interval;

$E[K_h(T-t)K_h(T-s)] = \|K\|^2 f(t)/h + o(1/h)$, for $s = t$.

- **Intuitive explanation?**

End of Asymptotic Theory

References for Mean and Covariance Estimation

- Yao Mueller and W. (2005, JASA)
- Li and Hsing (2010, AoS)
- Zhang and W. (2016, Ann. Stat.)
- **Additional References:**

  Stainiswalis and Lee (1998, JASA),
  Yao (2005, J. Mult. Analy.),

End of Part I