Lecture 1: Fundamental concepts in Functional Data Analysis

Valeria Vitelli

Oslo Centre for Biostatistics and Epidemiology (OCBE)
Department of Biostatistics, University of Oslo, Norway
valeria.vitelli@medisin.uio.no

UiO : University of Oslo

PhD course
An Introduction to Functional Data Analysis: Theory and Practice

University of Palermo, Italy; March 25^{th} – 28^{th}, 2024

Outline

Main Reference: Chapters 1-5 in R&S[1]

[1] Ramsay & Silverman, 2005: Functional Data Analysis, 2^{nd} edt, *Springer*

What are Functional Data (FD)?

Some key characteristics of a Functional Data Analysis (FDA):

- **Objects of interest in the analysis:** curves, images, or functions on higher dimensional domains
- **Goals:** descriptive statistics, classification, regression, inference $\rightarrow$ often the same as for univariate / multivariate statistical analyses!
- Additional FDA-specific methodology: smoothing, registration

What are Functional Data (FD)?

To define a sample of FD, consider three aspects:

1. a sample of FD is the collection of the observations $x_1, \ldots, x_n$ on n subjects, where we can imagine that each x_i arises from the evaluation of a subject-specific random high-dimensional object, a curve in the univariate case, $\mathcal{X}_i(t), t \in [a, b]$
2. what one *actually observes* in practice are the discrete evaluations of the x_i's at a finite grid of points $\{t_{i1}, \ldots, t_{im_i}\}$, i.e., the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$, where $\boldsymbol{x}_i = (x_i(t_{i1}), \ldots, x_i(t_{im_i}))$
3. often such evaluations are contaminated with noise, so that we actually observe a not-so-smooth version of the process

$$\tilde{x}_{ij} = x_i(t_{ij}) + \epsilon_{ij}$$

Some key FD characteristics

- often high dimensional (m_i is *very large* for many i's)
- often unbalanced (the sets $\{t_{i1}, \ldots, t_{im_i}\}$ might not be the same for different i's)
- may have multiple measurements of the same process (several $\boldsymbol{x}_i$'s for the same $i \Rightarrow$ easy to exploit it to estimate ϵ_{ij})
- parametric assumptions on the underlying process are often not made (complexity / flexibility)

Spoiler: Smoothing

Reconstructing the *smooth version* $x_i(t)$ of the observed functional data $\boldsymbol{x}_i$ (Step 3 in previous slide) is a critical 1^{st} step of the analysis

Functional data vs longitudinal data

Longitudinal Data (LD)

Also called *panel data.* Include repeated observations over time on the same subjects, common in medicine and social sciences.
Main characteristics:

- few repeated measurements per observation unit
- potentially different (numbers of) time points for different observation units

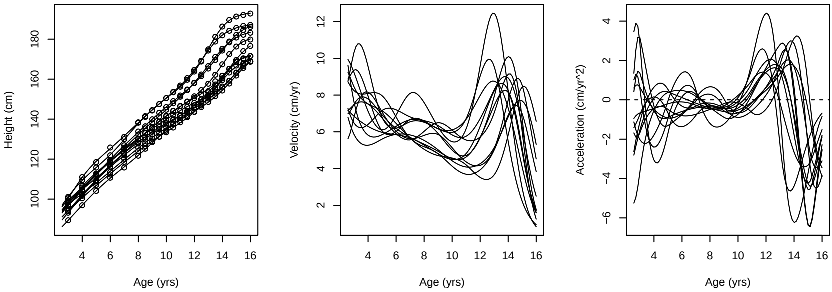
Methods to analyze LD traditionally differ in how data is **understood and treated:** repeated measurements accounted for by using **random effects in a parametric model.**

Functional data vs Time Series

Time Series

- Time Series analysis typically considers a **single observation of a stochastic process**
- Limited amount of data, so methods usually focus on **parametric estimation**
- **Aim** is good forecasting performance for unseen future time points
- **Goals and data structure** thus differ from LD and FDA
- If a **sample of different time series** is available (e.g., time series classification), setting closer to FDA

Examples (1): Berkeley Growth Data

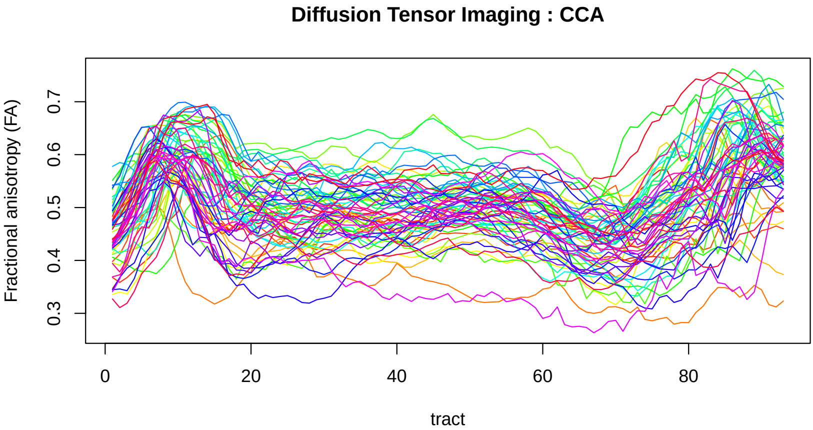


Examples (1): Berkeley Growth Data

Prototype for the type of data that we shall consider:

- heights of 54 girls and 39 boys measured at a set of 31 ages in the **Berkeley Growth Study** (Tuddenham and Snyder 1954) (here 12 girls and ages $3 < t < 16$ are selected)
- **ages are not equally spaced:** 4 measurements while the child is one year old, annual measurements from two to eight years, followed by heights measured biannually
- **uncertainty or noise in height values** that has a standard deviation of about three millimeters
- discrete values associated to the same observation **reflect the smooth variation in height** that could be assessed, in principle, as often as desired $\rightarrow$ *height function*
- data thus are a sample of **12 functional observations**

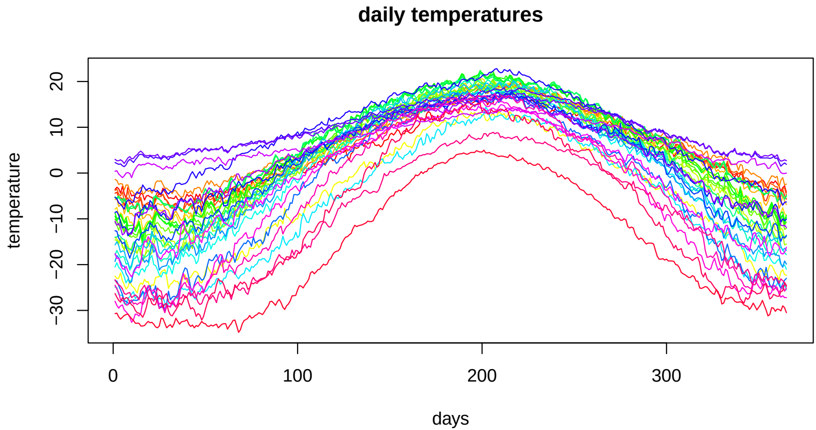
Examples (2): MRIs data from Multiple sclerosis patients



Examples (2): MRIs data from Multiple sclerosis patients

- **Tractography Application:** Multiple sclerosis (MS) is an demyalinating autoimmune disease associated with brain lesions
- DTI is an MRI technique which measures proxies of demyalination by quantifying the water diffusivity in the brain
- **Fractional anisotropy (FA)** is one measure of water diffusion in the brain
- **Data:** FA tract profiles for the corpus callosum (CCA) measured without missing values for 66 MS subjects at their first visit

Examples (3): Canadian Weather Data



Some functional data analyses

The **goals of functional data analysis** are essentially the same as those of any other branch of statistics, and include

- to represent the data in ways that aid further analysis
- to display the data so as to highlight various characteristics
- to study important sources of pattern and variation among the data
- to explain variation in an outcome or dependent variable by using input or independent variable information
- to compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same functions, or different functions for a common set of replicates (we focus on the former case!)

Useful FD notation and definitions I

Functional variable

A random variable $\mathcal{X}$ is called *functional variable* if it takes values in an infinite dimensional space (or functional space), i.e., it is a mapping $\mathcal{X} : \Omega \to \mathbb{H}$, where $\mathbb{H}$ is typically a Hilbert space (e.g., L^2). An observation x of $\mathcal{X}$ is called a *functional datum*.

Functional sample, or sample of FD

A functional dataset $x_1, \ldots, x_n$ is the observation of n functional variables $\mathcal{X}_1, \ldots, \mathcal{X}_n$ identically distributed as $\mathcal{X}$.

Useful FD notation and definitions II

"Simplified" notation

- for easiness of presentation, we implicitly identified $\mathcal{X}$ with the set $\{\mathcal{X}(t); t \in T\}$, because the "functional feature" or smoothness comes directly from the observations independently on the abscissa
- some specific situations (e.g., functional registration) require to explicitly refer to the abscissa (as it matters substantially for the analysis)
- the situation when the variable is a *curve* is associated with an unidimensional set $T \in \mathbb{R}$, but this needs not being the case in general (surfaces, manifolds, etc)

Useful FD notation and definitions III

Notation for infinite dimensional spaces & Operators

Unless stated differently, we assume $\mathbb{H}$ to be L^2, equipped with the usual *inner product*

$$\langle x, y \rangle = \int_T x(t)y(t)dt$$

and *norm*

$$||x|| = \langle x, x \rangle = \int_T x^2(t)dt$$

The *derivatives* follow the convention that:

- the derivative of order m of a function is $D^m x$ (as differentiation is an *operator*), with $D^0 x \equiv x$, and
- $D^{-1}x = \int x(t)dt$ ($\Rightarrow D^1 D^{-1} x = D^0 x = x$)

Functional means: functional expected value and sample (point-wise) mean

Formally, the mean of $\mathcal{X}$ is defined by

$$\mu := \mathbb{E}\left[\mathcal{X}\right] = \int_{\Omega} \mathcal{X}(\omega) dP(\omega) \tag{1}$$

where $(\Omega, \mathcal{A}, P)$ is the probability space. Usual estimator is its empirical version, i.e., the *functional sample mean*

$$\bar{\mathcal{X}}(t) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{X}_i,$$

which is in practice often estimated via the *point-wise mean,* i.e., the average of the FD in the sample point-wise across observations

$$\bar{x}(t) = \frac{1}{n} \sum_{i=1}^{n} x_i(t)$$

Variance functions I

The expectation in (1) **is well-defined if** $\mathcal{X}$ is **square integrable**. Then, $\mathbb{E}\left[||\mathcal{X}-\mu||^2\right]$ is the *variance function* associated with a random element $\mathcal{X}$ in a Hilbert space. The following result holds

Theorem

Assume that $\mathbb{E}\left[||\mathcal{X}||^2\right] < \infty$. Then

$$\mathbb{E}\left[||\mathcal{X}-\mu||^2\right] = \mathbb{E}\left[||\mathcal{X}||^2\right] - ||\mu||^2$$

where μ is the mean of $\mathcal{X}$. This gives an obvious recipe for deriving a functional estimator as done for the mean.

Variance functions II

Analogously to the case of the functional mean, a "universally" valid and well-known estimator of the variance function is the following *point-wise variance*

$$Var_{\mathcal{X}}(t) = \frac{1}{n-1} \sum_{i=1}^{n} [x_i(t) - \bar{x}(t)]^2 \,,$$

which can always be computed in practice.
The *standard deviation function* is the square root of the variance function

$$SD_{\mathcal{X}}(t) = \sqrt{Var_{\mathcal{X}}(t)}$$

Covariance and Correlation operators I

We now need to develop a concept of covariance for $\mathcal{X}$. Recall that, for $\boldsymbol{X} \in \mathbb{R}^p$, it holds

$$\mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})(\boldsymbol{X} - \mathbb{E}\boldsymbol{X})^T\right] = \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\boldsymbol{X}) \otimes (\boldsymbol{X} - \mathbb{E}\boldsymbol{X})\right],$$

which is a $p \times p$ matrix, so an element of the Borel[2] σ-field $\mathcal{B}(\mathbb{R}^p)$. The **covariance operator** builds on the same idea:

Definition: Covariance Operator

Assume $\mathbb{E}\left[||\mathcal{X}||^2\right] < \infty$. Then, the *covariance operator* for $\mathcal{X}$ is:

$$\mathcal{K} := \mathbb{E}\left[(\mathcal{X} - \mu) \otimes (\mathcal{X} - \mu)\right] = \int_\Omega (\mathcal{X}(\omega) - \mu) \otimes (\mathcal{X}(\omega) - \mu) dP(\omega)$$

[2] The Borel σ-field $\mathcal{B}(\mathbb{X})$ of a topological space $\mathbb{X}$ is the smallest σ-field containing all the open (relative to the norm-based metric) subsets of $\mathbb{X}$.

Covariance and Correlation operators II

Theorem

Since $\mathcal{X}(\omega) \in \mathbb{H}$ Hilbert, then $(\mathcal{X}(\omega) - \mu) \otimes (\mathcal{X}(\omega) - \mu)$ is a *Hilbert-Schmidt operator* with norm $||\mathcal{X}(\omega) - \mu||^2$.

The usual *point-wise estimators* for the covariance and correlation operators can be the easily derived as

$$Cov_{\mathcal{X}}(t_1, t_2) = \frac{1}{n-1} \sum_{i=1}^{n} \left[x_i(t_1) - \bar{x}(t_1)\right]\left[x_i(t_2) - \bar{x}(t_2)\right],$$

and

$$Corr_{\mathcal{X}}(t_1, t_2) = \frac{Cov_{\mathcal{X}}(t_1, t_2)}{\sqrt{Var_{\mathcal{X}}(t_1) Var_{\mathcal{X}}(t_2)}}$$

Functional Outliers

- the descriptive analysis of the data usually includes *visualizations* and *identification of outlying observations* that may be influential in the analysis
- this should be done for FDA as well → more complex task!
  - what does "functional outlier" mean?
- outlying curves can differ from other FD in the sample
  - by the **range** of their function values ("magnitude outliers")
  - by their **shape** ("shape outliers")
  - both

→ see Hyndman and Shang (2010)

Functional Outliers via Depth Functions

- Notion of *functional outlyingness or centrality* can be defined via the concept of **functional data depth**
- The first and most popular of such functional depths was the **(modified) functional band-depth** introduced by López-Pintado and Romo (2009):
  - first, consider all functional bands $B(x_{i_1}, x_{i_2}, \ldots, x_{i_j})$ determined any choice of j different curves $x_{i_1}, x_{i_2}, \ldots, x_{i_j}$ in the sample
    see next slide, left panel, for an example with 3 curves
  - then, for all x_i in the sample, calculate the proportion of such bands containing the whole graph of x_i
    see an example in the next slide, right panel
  - for $2 \leq J \leq n$ (fixed), the **functional band-depth** $BD^J(x_i)$ of x_i is the sum of such proportions for $j = 2, \ldots, J$
  - $J = 2$ or 3 are generally good choices (see López-Pintado and Romo (2009) for arguments on why)

Functional Band Depth: Illustration

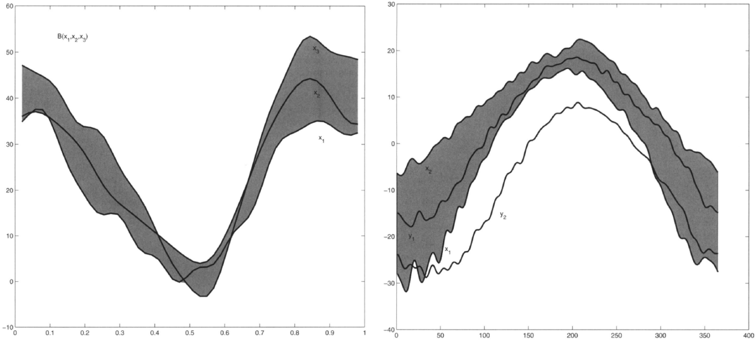


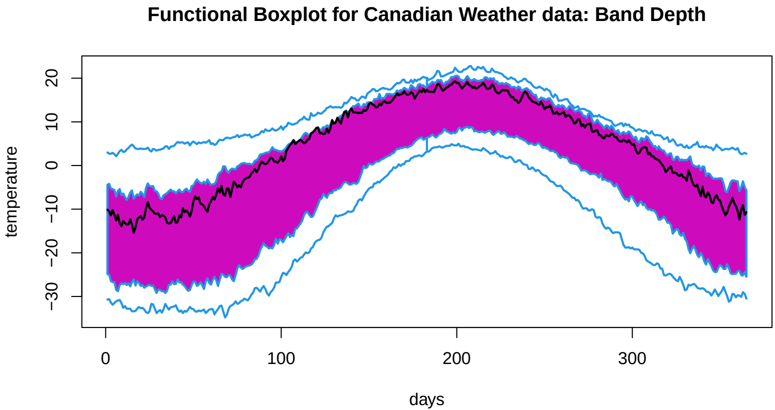
Figure from López-Pintado and Romo (2009):
Left panel, functional band determined by 3 curves; right panel, a practical example: y_1 is inside the band determined by x_1 and x_2, while y_2 is not.

Functional Boxplot

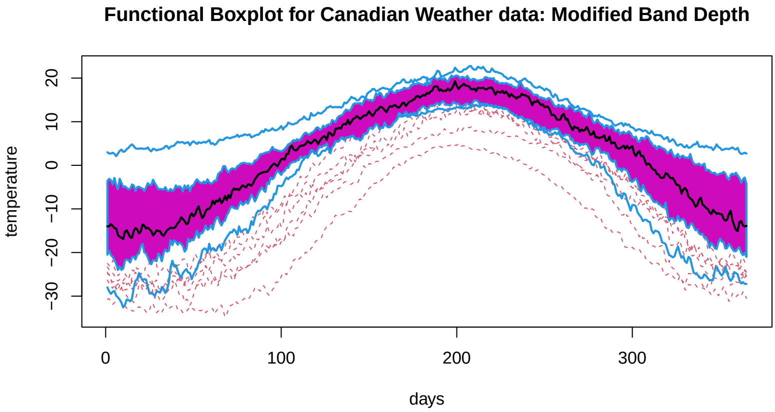
- The **sample median** is the curve with the *highest functional depth*[3]
- the 50% central region (corresponding to the **inter-quartile range** of a standard boxplot) is defined to envelop the 50% most central curves in the sample with the highest depth
- this region is inflated by 1.5 to generate the analog of whiskers in the boxplot
- any curve lying outside this region is flagged as an outlier

[3] ⇒ **the functional boxplot definition depends on the choice of functional depth!** See an illustration in the next slide

Functional Boxplot: Illustration



Functional Boxplot: Illustration



From discrete to functional data

The term *functional* refers to the intrinsic *smoothness* of the observed data, which means that we know / assume there exists an *underlying data generating process that is a function*

Definition: Smoothing

We refer to *smoothing* as the collection of methodologies aimed at **reconstructing the functional (smooth) form of the data from noisy discrete observations**

Why estimating a smooth representation at all?

- better summary statistics, description and interpretation
- allows evaluating / predicting the value of the process at any time point (be careful!)
- allows evaluation of derivatives (rates of change, if relevant)

Slightly different notation wrt to start

Observed data are the pairs:

$\{(y_{i1}, t_{i1}), \ldots, (y_{im_i}, t_{im_i})\}$, $i = 1, \ldots, n$

- y_{ij}'s are what we before called $\tilde{x}_{ij}$ (easier notation with y), they are "snapshots" of the smooth data $x_i(t_{ij})$ that *we would have observed if there was no noise*. Therefore:

$$y_{ij} = x_i(t_{ij}) + \epsilon_{ij}$$

  with ϵ_{ij} usually i.i.d. $\mathcal{N}(0, \sigma^2)$
- $x_i(t_{ij})$ is NOT $\mathcal{X}_i(t)$, the *latent functional random variable*
- $x_i(\cdot)$ is assumed *smooth* on T, and the realization of $\mathcal{X}_i$
- $t_{ij} \in T$ (may vary across i and j)

Definition: Smoothing, now precisely

We refer to *smoothing* as the collection of methodologies aimed at **reconstructing** $x_i(\cdot)$ $\forall$ $i = 1, \ldots, n$

From discrete to functional data: Basis expansion

Recall: Basis definition

A basis is a collection of functions that (i) are a linearly independent subset and (ii) can be chosen so that they "represent the mathematical space well". We will call the basis $\{\phi_k : k \geq 1\}$

Infinite dimensional spaces $\Rightarrow$ *infinite dimensional bases*

Recall: Orthonormal bases & basis expansion

If the basis is an orthonormal basis for the considered space $\mathbb{H}$, then $span\{\phi_k\} = \mathbb{H}$, and every $x \in \mathbb{H}$ can be written wrt the basis expansion:

$$x(t) = \sum_{k \geq 1} \langle x(t), \phi_k(t) \rangle \phi_k(t)$$

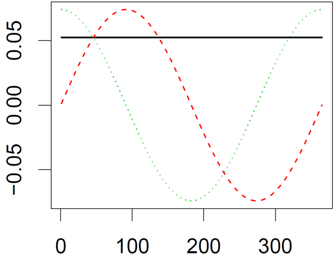
Truncating the expansion at some value $k = K$ is also a choice!

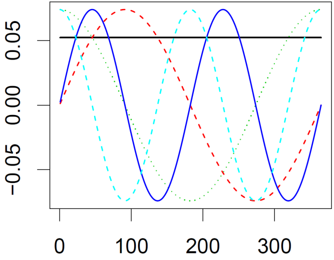
The Fourier Basis

Basis functions are sine and cosine of increasing frequency:

$$\{1, \{\sin(\omega t), \cos(\omega t)\}, \{\sin(2\omega t), \cos(2\omega t)\}, \ldots\}$$

and the constant ω has to be chosen according to the considered interval T ($|T| = 2\pi/\omega$). Canadian Wheather example, $|T| = 365$:





The Fourier Basis

Pros:

- very suitable choice for periodic / oscillating data
- computationally very efficient, especially for equally spaced t_{ij}'s

Cons:

- it is a very rigid basis, might not be good choice for non-periodic data
- still the default choice in many fields (due to FFT)

The B-spline Basis

Definition

Splines are polynomial segments that are joined end-to-end. The points at which the polynomial segments join are called *knots*

B-spline basis

A B-spline basis (or, shortly, *B-splines*) is completely defined by two elements

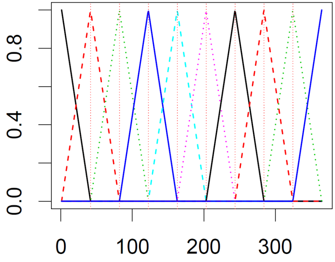
- the order (= degree + 1) of the polynomials used in the segments
- the location of knots (which define the segments)

Then, the number of B-splines functions in the basis is given by the order + number of internal knots

The B-spline Basis

Properties of B-splines of order m:

- derivatives up to $m-2$ are continuous ($\Rightarrow$ B-splines of order 4, cubic, popular as they have continuous 2^{nd} derivative)
- B-spline functions positive for *at most m* adjacent intervals $\Rightarrow$ fast computations!



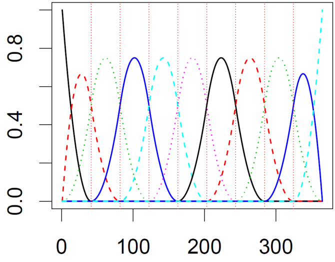


Figure: Example of B-spline functions. Left, B-splines of order 2, continuous; right, B-splines of order 3, continuous first derivative.

The B-spline Basis

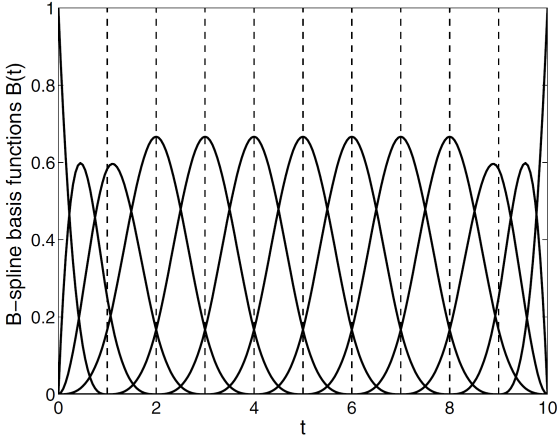


Figure 3.5 in R&S: The 13 basis functions defining an order 4 B-spline basis with 9 interior knots (shown as vertical dashed lines).

Smoothing FD using a basis via Least Squares (LS) fitting

Recall: our goal is to **fit the discrete observations $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$** using the model $y_{ij} = x_i(t_{ij}) + \epsilon_{ij}$

Model via Basis expansion (with matrix notation)

Fix the functional basis $\{\phi_k : k \geq 1\}$, and fix K. We are assuming that a good model for $x_i(\cdot)$ is:

$$x_i(t) = \sum_{k=1}^{K} c_{ik}\phi_k(t) = \boldsymbol{c}_i'\boldsymbol{\Phi}$$

where $\boldsymbol{c}_i \in \mathbb{R}^K$ includes basis coefficients for observation i.
Matrix notation: the basis $\{\phi_k : k \geq 1\}$ truncated at K elements is represented by the matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times K}$, where $\boldsymbol{\Phi}_{jk} = \phi_k(t_j)$.

Smoothing FD using a basis via Least Squares (LS) fitting

Simplest linear smoother: find the coefficients $\boldsymbol{c}_1, \dots, \boldsymbol{c}_n$ for all observations by minimizing

$$SMSSE(\boldsymbol{y}_i|\boldsymbol{c}_i) = \sum_{j=1}^{m} \left(y_{ij} - \sum_{k=1}^{K} c_{ik}\phi_k(t_j) \right)^2$$

which in matrix notation becomes

$$SMSSE(\boldsymbol{y}_i|\boldsymbol{c}_i) = (\boldsymbol{y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i)'(\boldsymbol{y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i) = ||\boldsymbol{y}_i - \boldsymbol{\Phi}\boldsymbol{c}_i||^2.$$

Obviously the LS solution is:

$$\hat{\boldsymbol{c}}_i = (\boldsymbol{\Phi}'\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}'\boldsymbol{y}_i \quad \text{and} \quad \boldsymbol{x}_i := \hat{\boldsymbol{y}}_i = \boldsymbol{\Phi}\hat{\boldsymbol{c}}_i \tag{2}$$

Smoothing FD using a basis via Least Squares (LS) fitting

Impact of K

K influences the *smoothness* of the fit.

- K small $\Rightarrow$ smooth fit: less flexibility but more generalization
- K large $\Rightarrow$ "jagged" fit: more flexibility but overfitting risk

How to choose the number of basis functions K?

- minimize mean-squared error (balance bias-variance trade-off)
- Cross-Validation (C-V) to tune K

Smoothing FD using a basis via Least Squares (LS) fitting

A simple Leave-One-Out (L-O-O) C-V scheme for $x_i(\cdot)$

1. Fix K. Repeat steps (i)-(iii) below for $j = 1, \ldots, m$:
   (i) exclude observation (y_{ij}, t_{ij}) (L-O-O)
   (ii) smooth the data $\boldsymbol{y}_{i,-j}$ by criterion (2), thus obtaining $\hat{\boldsymbol{c}}_{i,-j}$
   (iii) calculate the residual $e_j^K = y_{ij} - \sum_{k=1}^K \hat{c}_{ik,-j}\phi_k(t_j)$
2. Calculate the *cross-validation score:* the mean squared error of the m LS fits above on the removed observation

$$CV^K = \sum_{j=1}^{m} (e_j^K)^2$$

3. Repeat steps 1-2 above for reasonable choices of K, and select the optimal K^{opt} as $\text{argmin}_K\, CV^K$

Smoothing FD using a large K basis & Roughness Penalty

Recall: our goal is to **fit the discrete observations $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$** using the model $y_{ij} = x_i(t_{ij}) + \epsilon_{ij}$

- Fix the functional basis $\{\phi_k : k \geq 1\}$, with large K
- Assume $f(\cdot)$ is in the span of this basis: $f(t) = \sum_{k=1}^{K} c_k \phi_k(t)$
- Estimate $x_i(\cdot)$ by minimizing a goodness-of-fit criterion (Penalised SSE) over the choice of f :

$$PSSE_\lambda^i(f) = \sum_{j=1}^{m} (y_{ij} - f(t_j))^2 + \lambda P(f),$$

  where $P(f)$ measures the "roughness" of f, and λ controls the trade-off between goodness of fit and smoothness
- $\hat{x}_i(\cdot) = \text{argmin}_f \, PSSE_\lambda^i(f)$

Smoothing FD using a large K basis & Roughness Penalty

How to measure "roughness"?

- Square of the second derivative $\left[D^2 f(t)\right]^2$ of a function $f(t)$ at t is often called its *curvature* at t, since a straight line (= no curvature) has a zero second derivative
- Therefore, a natural measure of a function's *roughness* is the integrated squared second derivative

$$P_2(f) = \int \left[D^2 f(s)\right]^2 ds \tag{3}$$

- This suggests that we can generalize the roughness penalty (3) by allowing a derivative $D^m f(t)$ of arbitrary order

$$P_m(f) = \int [D^m f(s)]^2 ds$$

Smoothing FD using a large K basis & Roughness Penalty

The role of λ: Illustration

- $\lambda = 0 \;\Rightarrow$ *PSSE* reduces to *SMSSE*
- $\lambda \to \infty \;\Rightarrow$ more smoothness

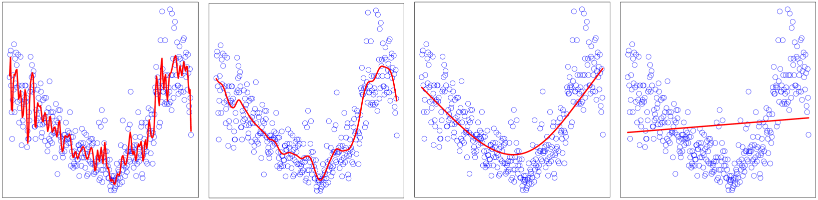


Figure: Vancouver mean temperature. Fits using 100 cubic splines (with equally spaced knots) and $\lambda = 0.1, 1000, 10^7, 10^{11}$ (left to right)

Recall: our goal is to **fit the discrete observations** $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ using the model $y_{ij} = x_i(t_{ij}) + \epsilon_{ij}$

Alternative method: Smoothing Splines

Smoothing technique that optimizes over the set of all possible f

- Estimate $x_i(\cdot)$ by minimizing the PSSE over the choice of f :

$$PSSE^i_\lambda(f) = \sum_{j=1}^{m} (y_{ij} - f(t_j))^2 + \lambda P(f),$$

  (same as before, but f is not chosen in $span\{\phi_k : k \geq 1\}$)

- $\hat{x}_i(\cdot) = \text{argmin}_f \, PSSE^i_\lambda(f)$
- **Solution:** *cubic splines* – piecewise cubic functions (splines of order 4), with knots at each t_j (Craven and Wahba 1978)

Smoothing techniques seen so far are called **presmoothing**, as resulting curves are treated as if they were truly functional observations. **Is there an alternative?**

Joint Smoothing and Modeling (JSM) approaches

They work with the observed data and **incorporate one of the smoothing methods into the analysis.**
Examples: sparse fPCA (Yao et al. 2005), functional regression (Scheipl et al. 2015; Greven and Scheipl 2017).

Presmoothing vs JSM

- **Presmoothing:** (+) can use theory and MANY methods developed for truly functional observations (-) do not propagate uncertainty (very serious when data are sparse)
- **JSM:** (+) uncertainty propagation, borrowing strength when data are sparse (-) limited amount of approaches available

Summary

- Basics of FDA:
  - Descriptive statistics (mean / variance functions, covariance operators)
  - Depth functions, Outliers and functional Boxplots
- Smoothing functional data (in a "classical" sense!):
  - Basis expansions (Fourier, B-splines)
  - LS-smoothing
  - Smoothing with roughness penalty
- Alternative methods:
  - Smoothing Splines
  - Join Smoothing and Modeling

References

Peter Craven and Grace Wahba. Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische mathematik*, 31(4):377–403, 1978.

Sonja Greven and Fabian Scheipl. A general framework for functional regression modelling. *Statistical Modelling*, 17(1-2):1–35, 2017.

Rob J Hyndman and Han Lin Shang. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45, 2010.

Sara López-Pintado and Juan Romo. On the concept of depth for functional data. *Journal of the American statistical Association*, 104(486):718–734, 2009.

Fabian Scheipl, Ana-Maria Staicu, and Sonja Greven. Functional additive mixed models. *Journal of Computational and Graphical Statistics*, 24(2):477–501, 2015.

R. D. Tuddenham and M. M. Snyder. Physical growth of california boys and girls from birth to age 18. Technical Report 1, University of California Publications in Child Development, 1954.

Fang Yao, Hans-Georg Müller, and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American statistical association*, 100(470): 577–590, 2005.