T-61.6030 Introductory Elements of Functional Data Analysis: Ramsay: Chapters 7,8,9

Ville Turunen

ville.t.turunen@tkk.fi

Outline

- Ch. 7: The registration and display of functional data
- Ch. 8: Principal components analysis for functional data
- Ch. 9: Regularized principal components analysis

Ch. 7: Introduction

- two types of variability: amplitude and phase
- physical time not always meaningful
  - e.g. human growth curve
- curves may have the same shape but different time scale $\Rightarrow$ direct comparison not possible
- time scale has to be transformed to register (align) the curves
- methods for registration
  - shift registration
  - feature or landmark registration
  - continuous registration

Shift registration

- simply move curves horizontally so that they are aligned
  - $x_i^*(t) = x_i(t + \delta_i)$
- *nuisance effects*: δ_i have no real interest
- *random effects*: δ_i are an important feature of each curve
- let $\mathcal{T}$ be the time interval $[T_1, T_2]$ where the curves are to be registered and $\hat{\mu}(t)$ the estimated mean
- least squares criterion to find δ_i
  - $\text{REGSSE} = \sum_{i=1}^{N} \int_{\mathcal{T}} [x_i(t + \delta_i) - \hat{\mu}(t)]^2 ds$ (7.1)
  - minimized with the Newton-Raphson algorithm
- *Procrustes method*: iteratively update in turns (1) mean $\hat{\mu}(t)$ and (2) shifts δ_i

Feature or landmark registration (1/2)

- *feature / landmark*: some characteristic that one can associate with specific argument value t
  - minima, maxima or zero crossings of the curve or its derivative etc.
- the goal is to construct a *time warping function* h_i for each curve so that the landmarks in the registered curves coincide
- registered curves: $x_i^*(t) = x_i[h_i(t)]$

Feature or landmark registration (2/2)

- the corresponding landmarks in each curve must be identified
- let t_{if} be the argument values for each curve x_i and each landmark $f = 1, \ldots, F$
- t_{0f} are the target timings i.e. landmarks in the mean function
- define $h_i(t)$ for each curve so that
  - $h(T_1) = T_1$ and $h(T_2) = T_2$
  - $h_i(t_{0f}) = t_{if}$ for all f
  - h_i is strictly monotonic
- the values between the landmarks are linearly interpolated

A more general warping function h

- linear interpolation for estimating h has limitations
  - no higher order derivatives
  - continuous registration not possible
- model time as a growth process (Section 6.3)
  - $h(t) = C_0 + C_1 \int_0^t exp[W(u)]du$ (7.2)
- $W(u) = 0 \Rightarrow h(t) = t$: physical time
- $W(u) < 0 \Rightarrow h(t) < t$: “running ahead”
- $W(u) > 0 \Rightarrow h(t) > t$: “running late”

Continuous registration (1/2)

- least squares criterion used for shift registration can not be used for general warping functions
  - tends to squeeze out regions where amplitude differs
- a criterion based on PCA can be defined
  - plotting the target curve $x_0(t)$ and correctly registered curve $x[h(t)]$ against each other should form a line
  - $\Rightarrow$ only one positive eigenvalue
- let $\mathbf{X}$ be a n by two matrix of sampled values $(x_0(t), x[h(t)])$
- functional analogue of $\mathbf{X}'\mathbf{X}$:
  - $$\mathbf{T}(h) = \left[\begin{array}{cc} \int \{x_0(t)\}^2 dt & \int x_0(t) x[h(t)] dt \\ \int x_0(t) x[h(t)] dt & \int \{x_0[h(t)]\}^2 dt \end{array} \right] \qquad (7.3)$$

Continuous registration (2/2)

- the fitting criterion:
  - $\texttt{MINEIG}(h) = \mu_2[\mathbf{T}(h)]$ (7.4)
  - i.e. the second eigenvalue of $\mathbf{T}(h)$
- when $\texttt{MINEIG}(h) = 0$, we have achieved registration and h is the warping function that does the job
- applying (7.2) and imposing smoothness regularization:
  - $\texttt{MINEIG}_\lambda(h) = \texttt{MINEIG}(h) + \lambda \int \{W^{(m)}(t)\}^2 dt$ (7.5)
- e.g. expand W in terms of B-splines

Some practical issues

- preprocessing: centering, rescaling
- zero crossings are good landmarks
  - register the derivate of the curve rather than the curve itself
- before continuous registration it is wise to register based on some clearly identifiable landmarks

Ch.8: PCA for functional data: Introduction

- user wants to find features characterizing the functions
- variance-covariance and correlation functions difficult to interpret
- $\Rightarrow$ principal components analysis (PCA)

PCA for multivariate data

- linear combination $f_i = \sum_{j=1}^{p} \beta_j x_{ij}, i = 1, \ldots, N$
- in vector form: $f_i = \beta' x_i, i = 1, \ldots, N$, where β' is the weight vector $(\beta_1, \ldots, \beta_p)'$ and x_1 is the vector $(x_{i1}, \ldots, x_{ip})'$
- PCA: find weight vectors that maximize variation in the f_i's
  1. find $\xi_1 = (\xi_{11}, \ldots, \xi_{1i})'$ so that
     - mean square $N^{-1} \sum_i f_{i1}^2$ is largest possible
     - here $f_{i1} = \xi_1' x_i$
     - constraint: $||\xi_1||^2 = 1$
  2. ... m. calculate second to mth weight vectors $\xi_2, \ldots, \xi_m$
     - additional constraints: $\xi_k' \xi_m = 0, k < m$
     - i.e. new weight vectors are orthogonal to the previous

PCA for functional data

- discrete index j in x_{ij} is replaced by continuous index s in $x_i(s)$
- instead of inner product $\beta' x_i = \sum_j \beta_j x_j$ use $\int \beta x = \int \beta(s) x(s) ds$
- weights become functions $\beta(s)$
- again, maximize $N^{-1} \sum_i f_{i1}^2 = N^{-1} \sum_i (\int \xi_{\mathbf{1}} x_i)^2$ with constraint $||\xi_{\mathbf{1}}||^2 = \int \xi_{\mathbf{1}}(s)^2 ds = 1$
- orthogonality constraint becomes: $\int \xi_k \xi_m = 0, k < m$

PCA and eigenvalues

- let $\mathbf{V}$ be the sample covariance matrix $\mathbf{V} = N^{-1}\mathbf{X}'\mathbf{X}$
- PCA for multivariate data can be performed by solving the eigenvector problem: $\mathbf{V}\xi = \rho\xi$
- for functional PCA, define the covariance function:
  - $v(s,t) = N^{-1}\sum_{i=1}^{N} x_i(s)x_i(t)$ (8.8)
- the PCA weight functions $\xi_j(s)$ satisfy:
  - $\int v(s,t)\xi(t)dt = \rho\xi(s)$ (8.9)
- define *covariance operator* V by:
  - $V\xi = \int v(\cdot,t)\xi(t)dt$ (8.10)
- functional PCA gets now the familiar form of:
  - $V\xi = \rho\xi$ (8.11)

Visualizing the results

- plotting the eigenfunctions $\xi(s)$
- plotting the mean and the functions obtained by adding and substracting a suitable multiple of the principal component function in question
- plotting principal component scores
- rotating principal components, e.g. VARIMAX rotation

Computational methods for functional PCA (1

- all methods involve converting the continuous functional eigenanalysis problem to an approximately equivalent matrix eigenanalysis task

1. discretize the functions

- take $N \times n$ data matrix $\mathbf{X}$ of finely sampled values of x_i
- solve the eigenvalue problem for $\mathbf{V} = N^{-1}\mathbf{X}'\mathbf{X}$
- to obtain an approximate eigenfunction $\xi(s)$ from the discrete values, we can use any convenient interpolation method

Computational methods for functional PCA (2

2. basis function expansion of the functions

- express each x_i as a linear combination of basis functions:
  - $x_i(t) = \sum_{k=1}^{K} c_{ik}\phi_k(t)$
  - in vector form: $\mathbf{x} = \mathbf{C}\phi$
- define $K \times K$ matrix $\mathbf{W} = \int \phi\phi'$
- suppose the eigenfunction ξ has expansion:
  - $\xi(s) = \sum_{k=1}^{K} b_k\phi_k(s) = \phi(s)'\mathbf{b}$
- now $\int v(s,t)\xi(t)dt = \int N^{-1}\phi(s)'\mathbf{C}'\mathbf{C}\phi(t)\phi(t)'\mathbf{b}dt = \phi(s)'N^{-1}\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b}$
- the eigenequation is purely matrix:
  - $N^{-1}\mathbf{W}^{1/2}\mathbf{C}'\mathbf{C}\mathbf{W}^{1/2}\mathbf{u} = \rho\mathbf{u}$
- eigenfunctions from $\mathbf{b} = \mathbf{W}^{-1/2}\mathbf{u}$

Bivariate and multivariate PCA

- simultaneous variation of more than one function
- both measured relative to the same function (e.g. time)
- both measure quantities in the same units (e.g. degrees or cm)

1. concatenate vectors into a single long bector Z_i
2. carry out PCA as in univariate case
3. separate the resulting components

- visualizing: e.g. plot one variable against the other

Ch.9: Regularized PCA: Introduction

- smoothing applied to PCA
- based on roughness penalty as in Chapter 5

Smoothing approach

- roughness penalty:
  - $\text{PEN}_2(\xi) = ||D^2\xi||^2$
- unsmoothed PCA maximizes sample variance $\texttt{var} \int \xi x_i$
- *penalized sample variance*:
  - $\text{PCAPSV}(\xi) = \frac{\texttt{var} \int \xi x_i}{||\xi||^2 + \lambda \text{PEN}_2(\xi)}$ (9.1)
- smoothing parameter $\lambda \geq 0$ chosen with cross-validation
- constraints:
  - $||\xi_j||^2 = 1$
  - $\int \xi_j(s)\xi_k(s)ds + \int D^2\xi_j(s)D^2\xi_k(s)ds = 0$ for $k = 1, \ldots, j-1$

Finding regularized PCA in practice (1/2)

- in practice, smoothed principal components are found by working in terms of a suitable basis

1. the periodic case

- Fourier basis $\{\phi_\nu\}$: $\{1, sin(\omega t), cos(\omega t), sin(2\omega t), cos(2\omega t)...\}$
- define $\omega_{2j-1} = \omega_{2j} = 2\pi j$
- the expansion: $x(s) = \sum_\nu c_\nu \phi_\nu(s) = \mathbf{c}'\phi(s)$
- roughness penalty becomes: $||D^2 x||^2 = \sum_\nu \omega_\nu^4 c_\nu^2$
- let $\mathbf{V}$ be the covariance matrix of the coefficient vectors $\mathbf{c_i}$
- let $\mathbf{S}$ be a diagonal matrix with entries $S_{\nu\nu} = (1 + \lambda\omega_\nu^4)^{-1/2}$
- $\mathbf{y}$ is the vector of Fourier coefficients for ξ

Finding regularized PCA in practice (2/2)

- penalized sample variance becomes:
  - $\text{PCAPSV}(\xi) = \frac{\mathbf{y}'\mathbf{V}\mathbf{y}}{\mathbf{y}'\mathbf{S}^{-2}\mathbf{y}}$ (9.6)
- eigenproblem is again purely matrix:
  - $\mathbf{V}\mathbf{y} = \rho\mathbf{S}^{-2}\mathbf{y}$ (9.7)
  - or: $(\mathbf{SVS})(\mathbf{S}^{-1}\mathbf{y}) = \rho(\mathbf{S}^{-1}\mathbf{y})$ (9.8)
- $\mathbf{SVS}$ is the covariance matrix of $\mathbf{S}\mathbf{c_i}$
- i.e. normal PCA on the smoothed coefficients $\mathbf{S}\mathbf{c_i}$

2. nonperiodic case

- e.g. B-splines are used instead of Fourier basis
- with similar (but bit more complicated) steps PCAPSV and the corresponding matrix eigenequation problem can be found

Alternative approaches

- smooth the data first, then carry out an unsmoothed PCA
- stepwise roughness penalty procedure: different λ_j for each principal component

Conclusions

- skipped in this presentation
  - lots of mathematical details
  - some nice examples and figures
  - *7.9 Computational details*
  - details of *8.3 Visualizing the results*
  - *8.4.3 More general numerical quadrature*
  - details of *8.5 Bivariate and multivariate PCA*
  - *9.3.3 Choosing the smoothing parameter by CV*