Deciding the Dimension of Effective Dimension Reduction Space for Functional Data

Yehua Li
Department of Statistics
University of Georgia

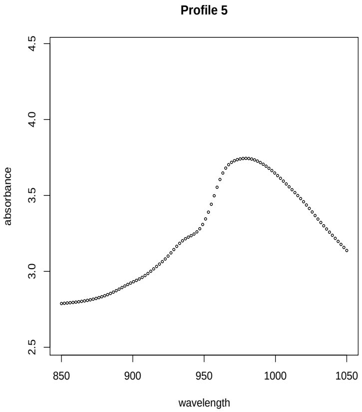
Tailen Hsing
Department of Statistics
University of Michigan

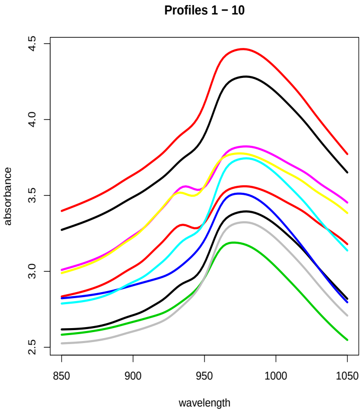
September 7, 2009

Near infrared spectroscopy data

- Meat samples were analyzed using near-infrared spectroscopy, which uses the near infrared region of the electromagnetic spectrum from 850 nm to 1050 nm. Each sample contains finely chopped pure meat with different moisture, fat and protein contents.
- For each meat sample the contents of water, fat and protein (in percent) were determined by analytic chemistry.
- See `http://lib.stat.cmu.edu/datasets/tecator` for details.

Near infrared spectroscopy data





Near infrared spectroscopy data

Data description

- number of meat samples: 215
- functional variable: absorbances
- multivariate data: moisture, fat and protein contents
- data size
  - absorbance: 100×215
  - moisture, fat and protein contents: 3×215

Near infrared spectroscopy data

Data description

- number of meat samples: 215
- functional variable: absorbances
- multivariate data: moisture, fat and protein contents
- data size
  - absorbance: 100×215
  - moisture, fat and protein contents: 3×215

Problem of interest

- estimating moisture, fat and protein contents based on absorbance

Basics of FDA

Let $\{X(t) : t \in T\}$ be a stochastic process with

- $\mu(t) = \mathbb{E}\{X(t)\}$ and $R(s,t) = \text{cov}\{X(s), X(t)\}$
- regularity conditions

Basics of FDA

Let $\{X(t) : t \in T\}$ be a stochastic process with

- $\mu(t) = \mathbb{E}\{X(t)\}$ and $R(s,t) = \text{cov}\{X(s), X(t)\}$
- regularity conditions

Data

- independent realizations $X_1, X_2, \ldots, X_n$ of X, (usually) densely observed, or
- independent realizations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ where Y is a covariate

Basics of FDA

Let $\{X(t) : t \in T\}$ be a stochastic process with

- $\mu(t) = \mathbb{E}\{X(t)\}$ and $R(s,t) = \text{cov}\{X(s), X(t)\}$
- regularity conditions

Data

- independent realizations $X_1, X_2, \ldots, X_n$ of X, (usually) densely observed, or
- independent realizations $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ where Y is a covariate

Inference

- distributional properties of X
- relationship between X and Y

Hilbert space and covariance operator

Let $\mathscr{H}$ be a Hilbert space of functions on T and $\langle \cdot, \cdot \rangle$ be the inner product of $\mathscr{H}$. Assume that $X \in \mathscr{H}$ a.s.

We will focus on $\mathscr{H} = L^2[a, b]$, the space of square-integrable functions on $[a, b]$, for some finite a, b, where

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt.$$

Hilbert space and covariance operator

Let $\mathscr{H}$ be a Hilbert space of functions on T and $\langle \cdot, \cdot \rangle$ be the inner product of $\mathscr{H}$. Assume that $X \in \mathscr{H}$ a.s.

We will focus on $\mathscr{H} = L^2[a,b]$, the space of square-integrable functions on $[a,b]$, for some finite a, b, where

$$\langle f, g \rangle = \int_a^b f(t)g(t)dt.$$

The *covariance operator* is

$$\Gamma_X : f \mapsto \int_a^b R(s,\cdot)f(s)ds, \;\; L^2[a,b] \mapsto L^2[a,b].$$

Eigen decomposition

Assume that $\mathbb{E}\|X\|^2 < \infty$. It follows that

- $\Gamma_X = \sum_{j=1}^{\infty} \omega_j \psi_j \otimes \psi_j$,
  - $\psi_1, \psi_2, \ldots$ are orthonormal
  - $\psi_j \otimes \psi_j$ is projection operator
  - $\omega_1 \geq \omega_2 \geq \cdots \geq 0$ and $\sum_i \omega_i < \infty$

Eigen decomposition

Assume that $\mathbb{E}\|X\|^2 < \infty$. It follows that

- $\Gamma_X = \sum_{j=1}^{\infty} \omega_j \psi_j \otimes \psi_j$,
  - $\psi_1, \psi_2, \ldots$ are orthonormal
  - $\psi_j \otimes \psi_j$ is projection operator
  - $\omega_1 \geq \omega_2 \geq \cdots \geq 0$ and $\sum_i \omega_i < \infty$
- [Mercer's Theorem]

  $R(s,t) = \sum_{j=1}^{\infty} \omega_j \psi_j(s) \psi_j(t)$

Eigen decomposition

Assume that $\mathbb{E}\|X\|^2 < \infty$. It follows that

- $\Gamma_X = \sum_{j=1}^{\infty} \omega_j \psi_j \otimes \psi_j$,
  - $\psi_1, \psi_2, \ldots$ are orthonormal
  - $\psi_j \otimes \psi_j$ is projection operator
  - $\omega_1 \geq \omega_2 \geq \cdots \geq 0$ and $\sum_i \omega_i < \infty$
- [Mercer's Theorem]

  $R(s,t) = \sum_{j=1}^{\infty} \omega_j \psi_j(s) \psi_j(t)$
- [Karhunen-Loève expansion]

  $X(t) = \mu(t) + \sum_{j=1}^{\infty} \omega_j^{1/2} \eta_j \psi_j(t)$
  - mean square expansion
  - $\mathbb{E}(\eta_j) = 0, \mathrm{cov}(\eta_j, \eta_k) = \delta_{j,k}$

A general functional regression model

Consider the model

$$Y = f(\langle \beta_1, X \rangle, \cdots, \langle \beta_K, X \rangle, \varepsilon),$$

where ε is unobserved and $f, K, \beta_1, \ldots, \beta_K$ are unknown, and the β_k's are linearly independent.

A general functional regression model

Consider the model

$$Y = f(\langle \beta_1, X \rangle, \cdots, \langle \beta_K, X \rangle, \varepsilon),$$

where ε is unobserved and $f, K, \beta_1, \ldots, \beta_K$ are unknown, and the β_k's are linearly independent.

Goals:

- estimate K
- estimate $\text{span}(\beta_1, \ldots, \beta_K)$, effective dimension reduction (EDR) space
- estimate (a version of) f

Special cases

- linear regression model: $Y = \beta_0 + \langle \beta_1, X \rangle + \varepsilon$

  Cardot, Ferraty and Sarda (2003), Cai and Hall (2006), Li and Hsing (2007)

- generalized linear model: $Y = f(\langle \beta, X \rangle, \varepsilon)$

  Müller and Stadtmüller (2005), Cardot and Sarda (2005)

- projection pursuit model: $Y = \sum_{k=1}^{K} f_k(\langle \beta_k, X \rangle) + \varepsilon$

  James and Silverman (2005)

Assumption $\mathscr{L}$: For any $\beta \in \mathscr{H}$,

$$\mathbb{E}(\langle \beta, X\rangle | \langle \beta_1, X\rangle, \cdots, \langle \beta_K, X\rangle) = c_0 + \sum_{k=1}^{K} c_k \langle \beta_k, X\rangle$$

for some constants $c_0, \cdots, c_K$.

Assumption $\mathscr{L}$ holds for processes with *elliptically contoured* distributions.

Inverse regression

Assumption $\mathscr{L}$: For any $\beta \in \mathscr{H}$,

$$\mathbb{E}(\langle \beta, X\rangle | \langle \beta_1, X\rangle, \cdots, \langle \beta_K, X\rangle) = c_0 + \sum_{k=1}^{K} c_k \langle \beta_k, X\rangle$$

for some constants $c_0, \cdots, c_K$.

Assumption $\mathscr{L}$ holds for processes with *elliptically contoured* distributions.

Theorem

[Li 1991, Ferré and Yao 2003] Under Assumption $\mathscr{L}$,

$$IR(t) := \mathbb{E}(X(t) - \mu(t) | Y) \in span\ (\Gamma_X \beta_1, \cdots, \Gamma_X \beta_K).$$

Elliptically contoured distribution

- A process X has an elliptically contoured distribution if

$$\mathbb{E}e^{i\langle f, X-\mu\rangle} = \phi(\langle f, Tf\rangle),\ f \in \mathscr{H},$$

  for some self-adjoint operator T and characteristic function ϕ.

- In the infinite-dimensional case, X has an elliptically contoured distribution if $X \stackrel{d}{=} \mu + RZ$ where Z is a zero-mean Gaussian process and R is a nonnegative random variable independent of Z. This can be proved based on a result by Schoenberg.

- Examples: Gaussian process, t-process

Inverse regression

By the Ferré and Yao Theorem,

$$\mathrm{Im}(\Gamma_{IR}) \subset \text{span } (\Gamma_X \beta_1, \cdots, \Gamma_X \beta_K).$$

If

$$\mathrm{Im}(\Gamma_{IR}) = \text{span } (\Gamma_X \beta_1, \cdots, \Gamma_X \beta_K),$$

then the β_k's can be estimated through the estimation of Γ_X and Γ_{IR}.

Inverse regression via principal components

- $X = \mu + \sum_{j=1}^{\infty} \omega_j^{1/2} \eta_j \psi_j$
- $\beta_k = \sum_{j=1}^{\infty} \omega_j^{-1/2} b_{kj} \psi_j$

Inverse regression via principal components

- $X = \mu + \sum_{j=1}^{\infty} \omega_j^{1/2} \eta_j \psi_j$
- $\beta_k = \sum_{j=1}^{\infty} \omega_j^{-1/2} b_{kj} \psi_j$
- $$\begin{aligned} \langle X - \mu, \beta_k \rangle &= \sum_{j=1}^{m} \eta_j b_{kj} + \sum_{j=m+1}^{\infty} \eta_j b_{kj} \\ &=: \boldsymbol{\eta}_{(m)}^T \mathbf{b}_{k,(m)} + \zeta_{k,(m)} \end{aligned}$$

Inverse regression based on principal components

The model $Y = f(\langle \beta_1, X\rangle, \cdots, \langle \beta_K, X\rangle, \varepsilon)$ can be written as

$$\begin{aligned} Y &= f_1(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{1,(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{K,(m)}, \varepsilon) \\ &= f_2(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)}, \varepsilon_{(m)}) \end{aligned}$$

Inverse regression based on principal components

The model $Y = f(\langle \beta_1, X\rangle, \cdots, \langle \beta_K, X\rangle, \varepsilon)$ can be written as

$$\begin{aligned} Y &= f_1(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{1,(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{K,(m)}, \varepsilon) \\ &= f_2(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)}, \varepsilon_{(m)}) \end{aligned}$$

- The EDR space of this problem is $\mathrm{span}(\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)})$.

Inverse regression based on principal components

The model $Y = f(\langle \beta_1, X \rangle, \cdots, \langle \beta_K, X \rangle, \varepsilon)$ can be written as

$$\begin{aligned} Y &= f_1(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{1,(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{K,(m)}, \varepsilon) \\ &= f_2(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)}, \varepsilon_{(m)}) \end{aligned}$$

- The EDR space of this problem is $\mathrm{span}(\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)})$.
- If $\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)}$ are linearly independent then the dimension of EDR space is K.

Inverse regression based on principal components

The model $Y = f(\langle \beta_1, X\rangle, \cdots, \langle \beta_K, X\rangle, \varepsilon)$ can be written as

$$\begin{aligned} Y &= f_1(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{1,(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)} + \zeta_{K,(m)}, \varepsilon) \\ &= f_2(\mathbf{b}_{1,(m)}^T \boldsymbol{\eta}_{(m)}, \cdots, \mathbf{b}_{K,(m)}^T \boldsymbol{\eta}_{(m)}, \varepsilon_{(m)}) \end{aligned}$$

- The EDR space of this problem is $\mathrm{span}(\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)})$.
- If $\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)}$ are linearly independent then the dimension of EDR space is K.
- Denote by $\Gamma_{IR,(m)}$ the inverse regression covariance matrix of this problem; namely

$$\Gamma_{IR,(m)} = \mathrm{cov}(\mathbb{E}(\boldsymbol{\eta}_{(m)} | Y)).$$

- The "pseudo-error", $\varepsilon_{(m)}$, is equal to

$$(\varepsilon, \zeta_{1,(m)}, \ldots, \zeta_{K,(m)}),$$

and is in general not independent of the $\boldsymbol{\eta}_{(m)}$.

Inverse regression based on principal components

- The "pseudo-error", $\varepsilon_{(m)}$, is equal to

$$(\varepsilon, \zeta_{1,(m)}, \ldots, \zeta_{K,(m)}),$$

  and is in general not independent of the $\boldsymbol{\eta}_{(m)}$.
- We do not directly observe $\boldsymbol{\eta}_{(m)}$.

Inverse regression based on principal components

- The "pseudo-error", $\varepsilon_{(m)}$, is equal to

$$(\varepsilon, \zeta_{1,(m)}, \ldots, \zeta_{K,(m)}),$$

  and is in general not independent of the $\boldsymbol{\eta}_{(m)}$.
- We do not directly observe $\boldsymbol{\eta}_{(m)}$.
- What's the role of m?

An example

- $$X(t) = \sum_{k=1}^{\infty} \omega_{2k-1}^{1/2} \eta_{2k-1} \sqrt{2} \cos(2k\pi t) + \sum_{k=1}^{\infty} \omega_{2k}^{1/2} \eta_{2k} \sqrt{2} \sin(2k\pi t), \quad t \in [0,1],$$

  where $\omega_k = 20(k+1.5)^{-3}$ and η_k's $\sim N(0,1)$.

An example

- $$X(t) = \sum_{k=1}^{\infty} \omega_{2k-1}^{1/2} \eta_{2k-1} \sqrt{2} \cos(2k\pi t) + \sum_{k=1}^{\infty} \omega_{2k}^{1/2} \eta_{2k} \sqrt{2} \sin(2k\pi t), \;\; t \in [0,1],$$

  where $\omega_k = 20(k+1.5)^{-3}$ and η_k's $\sim N(0,1)$.

- The principal components are

  $$\sqrt{2}\cos(2\pi t), \sqrt{2}\sin(2\pi t), \sqrt{2}\cos(4\pi t), \sqrt{2}\sin(4\pi t), \ldots$$

- $$\begin{aligned}\beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t) \\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t), \\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t) \\ &\quad + 1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4} \frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t).\end{aligned}$$

- $$\begin{aligned}\beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t) \\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t), \\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t) \\ &\quad + 1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4} \frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t).\end{aligned}$$

- $$Y = \langle \beta_1, X\rangle \times (2\langle \beta_2, X\rangle + 1) + \varepsilon \quad \Longrightarrow \quad K = 2$$

- $$\begin{aligned}\beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t) \\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t), \\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t) \\ &\quad + 1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4} \frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t).\end{aligned}$$

- $$Y = \langle \beta_1, X\rangle \times (2\langle \beta_2, X\rangle + 1) + \varepsilon \quad \Longrightarrow \quad K = 2$$

- $$\begin{array}{lrrrrrrr} \mathbf{b}_{1,(m)} = & .9, & 0, & 1.2, & 0, & 0, & 0, & \ldots \\ \mathbf{b}_{2,(m)} = & .45, & 0, & .6, & 0, & 0, & \text{-}3, & \ldots \end{array}$$

Sliced inverse regression

- Data: $(X_1, Y_1), \cdots, (X_n, Y_n)$, fully observed.
- Estimate Γ_X by $\widehat{\Gamma}_X$, the sample covariance operator

Sliced inverse regression

- Data: $(X_1, Y_1), \cdots, (X_n, Y_n)$, fully observed.
- Estimate Γ_X by $\widehat{\Gamma}_X$, the sample covariance operator
- Compute $\widehat{\psi}_j$ and $\widehat{\omega}_j$ by the eigen decomposition of $\widehat{\Gamma}_X$
- Compute the PC scores $\widehat{\eta}_{ij} = \widehat{\omega}_j^{-1/2} \langle \widehat{\psi}_j, X_i - \bar{X} \rangle$

Sliced inverse regression

- Data: $(X_1, Y_1), \cdots, (X_n, Y_n)$, fully observed.
- Estimate Γ_X by $\widehat{\Gamma}_X$, the sample covariance operator
- Compute $\widehat{\psi}_j$ and $\widehat{\omega}_j$ by the eigen decomposition of $\widehat{\Gamma}_X$
- Compute the PC scores $\widehat{\eta}_{ij} = \widehat{\omega}_j^{-1/2} \langle \widehat{\psi}_j, X_i - \bar{X} \rangle$
- Let $S_1, \cdots, S_H$ be a partition of the the range of Y. Let

$$
\begin{aligned}
n_h &= \sum_{i=1}^{n} I(Y_i \in S_h), \ \widehat{p}_h = \frac{n_h}{n} \\
\bar{\boldsymbol{\eta}}_{h,(m)} &= \frac{1}{n_h} \sum_{i=1}^{n} \widehat{\boldsymbol{\eta}}_{i,(m)} I(Y_i \in S_h) \\
\widehat{\Gamma}_{IR,(m)} &= \sum_{h=1}^{H} \widehat{p}_h \bar{\boldsymbol{\eta}}_{h,(m)} \bar{\boldsymbol{\eta}}_{h,(m)}^T
\end{aligned}
$$

- $\widehat{\Gamma}_{IR,(m)}$ estimates $\Gamma_{IR,(m)}$.
- Let $\widehat{\mathbf{b}}_{1,(m)}, \cdots, \widehat{\mathbf{b}}_{K,(m)}$ be the leading eigenvectors of $\widehat{\Gamma}_{IR,(m)}$.

  $\mathrm{span}(\widehat{\mathbf{b}}_{1,(m)}, \cdots, \widehat{\mathbf{b}}_{K,(m)})$ estimates $\mathrm{span}(\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)})$.
- $\widehat{\beta}_k := \sum_{j=1}^{m} \widehat{\omega}_j^{-1/2} \widehat{b}_{kj} \widehat{\psi}_j$

  $\mathrm{span}(\widehat{\beta}_1, \cdots, \widehat{\beta}_K)$ estimates $\mathrm{span}(\beta_1, \cdots, \beta_K)$.

- $\widehat{\Gamma}_{IR,(m)}$ estimates $\Gamma_{IR,(m)}$.
- Let $\widehat{\mathbf{b}}_{1,(m)}, \cdots, \widehat{\mathbf{b}}_{K,(m)}$ be the leading eigenvectors of $\widehat{\Gamma}_{IR,(m)}$.

  $\mathrm{span}(\widehat{\mathbf{b}}_{1,(m)}, \cdots, \widehat{\mathbf{b}}_{K,(m)})$ estimates $\mathrm{span}(\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)})$.
- $\widehat{\beta}_k := \sum_{j=1}^{m} \widehat{\omega}_j^{-1/2} \widehat{b}_{kj} \widehat{\psi}_j$

  $\mathrm{span}(\widehat{\beta}_1, \cdots, \widehat{\beta}_K)$ estimates $\mathrm{span}(\beta_1, \cdots, \beta_K)$.
- If X is p-dimensional and $m = p$, this is exactly Li's SIR.
- If X is infinite-dimensional, then in general m must increase to ∞ with n in establishing consistency [Ferré and Yao 2003].

- $\widehat{\Gamma}_{IR,(m)}$ estimates $\Gamma_{IR,(m)}$.
- Let $\widehat{\mathbf{b}}_{1,(m)}, \cdots, \widehat{\mathbf{b}}_{K,(m)}$ be the leading eigenvectors of $\widehat{\Gamma}_{IR,(m)}$.

  $\mathrm{span}(\widehat{\mathbf{b}}_{1,(m)}, \cdots, \widehat{\mathbf{b}}_{K,(m)})$ estimates $\mathrm{span}(\mathbf{b}_{1,(m)}, \cdots, \mathbf{b}_{K,(m)})$.
- $\widehat{\beta}_k := \sum_{j=1}^m \widehat{\omega}_j^{-1/2} \widehat{b}_{kj} \widehat{\psi}_j$

  $\mathrm{span}(\widehat{\beta}_1, \cdots, \widehat{\beta}_K)$ estimates $\mathrm{span}(\beta_1, \cdots, \beta_K)$.
- If X is p-dimensional and $m = p$, this is exactly Li's SIR.
- If X is infinite-dimensional, then in general m must increase to ∞ with n in establishing consistency [Ferré and Yao 2003].
- Prerequisite: Need to know K.

Deciding the dimension of EDR space

- To determine K, we sequentially test $H_0 : K \leq K_0$ vs. $H_a : K > K_0$ for $K_0 = 0, 1, \ldots$, and conclude $K = K_0$ the first time we fail to reject H_0.

Deciding the dimension of EDR space

- To determine K, we sequentially test $H_0 : K \leq K_0$ vs. $H_a : K > K_0$ for $K_0 = 0, 1, \ldots$, and conclude $K = K_0$ the first time we fail to reject H_0.
- If $K \leq K_0$ then $\mathrm{rank}(\Gamma_{IR,(m)}) \leq K_0$ and so $\lambda_j(\Gamma_{IR,(m)}) = 0$ for $j = K_0 + 1, \ldots, m$.

Deciding the dimension of EDR space

- To determine K, we sequentially test $H_0 : K \leq K_0$ vs. $H_a : K > K_0$ for $K_0 = 0, 1, \ldots$, and conclude $K = K_0$ the first time we fail to reject H_0.
- If $K \leq K_0$ then $\mathrm{rank}(\Gamma_{IR,(m)}) \leq K_0$ and so $\lambda_j(\Gamma_{IR,(m)}) = 0$ for $j = K_0 + 1, \ldots, m$.
- Test statistic:

$$\mathscr{T}_{K_0,(m)} = n \sum_{j=K_0+1}^{m} \lambda_j(\widehat{\Gamma}_{IR,(m)})$$

Deciding the dimension of EDR space

- To determine K, we sequentially test $H_0 : K \le K_0$ vs. $H_a : K > K_0$ for $K_0 = 0, 1, \ldots$, and conclude $K = K_0$ the first time we fail to reject H_0.
- If $K \le K_0$ then $\mathrm{rank}(\Gamma_{IR,(m)}) \le K_0$ and so $\lambda_j(\Gamma_{IR,(m)}) = 0$ for $j = K_0 + 1, \ldots, m$.
- Test statistic:

$$\mathscr{T}_{K_0,(m)} = n \sum_{j=K_0+1}^{m} \lambda_j(\widehat{\Gamma}_{IR,(m)})$$

- Reject H_0 for large values of $\mathscr{T}_{K_0,(m)}$.

Define $K_{(m)} = \mathrm{rank}(\Gamma_{IR,(m)})$.

Define $K_{(m)} = \mathrm{rank}(\Gamma_{IR,(m)})$.

Theorem

Assume that X is a Gaussian process. Assume that $K \le K_0$, and let $H > K_0 + 1$ and $m \ge K_0 + 1$. Denote by $\mathscr{X}$ a random variable having a χ^2 distribution with $(m - K_0)(H - K_0 - 1)$ degrees of freedom. Recall that $K_{(m)} \le K \le K_0$.

Define $K_{(m)} = \mathrm{rank}(\Gamma_{IR,(m)})$.

Theorem

Assume that X is a Gaussian process. Assume that $K \leq K_0$, and let $H > K_0 + 1$ and $m \geq K_0 + 1$. Denote by $\mathscr{X}$ a random variable having a χ^2 distribution with $(m - K_0)(H - K_0 - 1)$ degrees of freedom. Recall that $K_{(m)} \leq K \leq K_0$.

- *If $K_{(m)} = K_0$, then*

$$\mathscr{T}_{K_0,(m)} \xrightarrow{d} \mathscr{X} \text{ as } n \to \infty.$$

Define $K_{(m)} = \mathrm{rank}(\Gamma_{IR,(m)})$.

Theorem

Assume that X is a Gaussian process. Assume that $K \leq K_0$, and let $H > K_0 + 1$ and $m \geq K_0 + 1$. Denote by $\mathscr{X}$ a random variable having a χ^2 distribution with $(m - K_0)(H - K_0 - 1)$ degrees of freedom. Recall that $K_{(m)} \leq K \leq K_0$.

- *If $K_{(m)} = K_0$, then*

$$\mathscr{T}_{K_0,(m)} \xrightarrow{d} \mathscr{X} \text{ as } n \to \infty.$$

- *If $K_{(m)} < K_0$, then $\mathscr{T}_{K_0,(m)}$ is asymptotically stochastically bounded by $\mathscr{X}$; namely,*

$$\limsup_{n\to\infty} \mathbb{P}(\mathscr{T}_{K_0,(m)} > x) \leq \mathbb{P}(\mathscr{X} > x) \text{ for all } x.$$

Example, continued

- $$X(t) = \sum_{k=1}^{\infty} \omega_{2k-1}^{1/2} \eta_{2k-1} \sqrt{2} \cos(2k\pi t) + \sum_{k=1}^{\infty} \omega_{2k}^{1/2} \eta_{2k} \sqrt{2} \sin(2k\pi t), \quad t \in [0,1],$$

  where $\omega_k = 20(k+1.5)^{-3}$ and η_k's $\sim N(0,1)$.

Example, continued

- $$X(t) = \sum_{k=1}^{\infty} \omega_{2k-1}^{1/2} \eta_{2k-1} \sqrt{2} \cos(2k\pi t) + \sum_{k=1}^{\infty} \omega_{2k}^{1/2} \eta_{2k} \sqrt{2} \sin(2k\pi t), \ \ t \in [0, 1],$$

  where $\omega_k = 20(k + 1.5)^{-3}$ and η_k's $\sim N(0, 1)$.

- The principal components are

  $$\sqrt{2}\cos(2\pi t), \sqrt{2}\sin(2\pi t), \sqrt{2}\cos(4\pi t), \sqrt{2}\sin(4\pi t), \ldots$$

- $$\begin{aligned}\beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t) \\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t), \\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t) \\ &\quad + 1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4} \frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t).\end{aligned}$$

- $$\begin{aligned} \beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t) \\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t), \\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t) \\ &\quad + 1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4} \frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t). \end{aligned}$$

- $$Y = \langle \beta_1, X\rangle \times (2\langle \beta_2, X\rangle + 1) + \varepsilon \quad \Longrightarrow \quad K = 2$$

- $$\begin{aligned}\beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t) \\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t), \\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t) \\ &\quad +1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4} \frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t).\end{aligned}$$

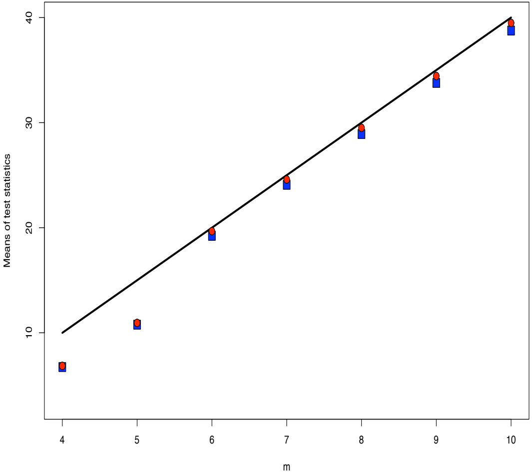
- $$Y = \langle \beta_1, X\rangle \times (2\langle \beta_2, X\rangle + 1) + \varepsilon \quad \Longrightarrow \quad K = 2$$

- $$\begin{aligned}\mathbf{b}_{1,(m)} &= .9, \quad 0, \quad 1.2, \quad 0, \quad 0, \quad 0, \quad \ldots \\ \mathbf{b}_{2,(m)} &= .45, \quad 0, \quad .6, \quad 0, \quad 0, \quad \text{-}3, \quad \ldots\end{aligned}$$

So $K_{(m)} = 1$ for $m \leq 5$ and $K_{(m)} = 2$ for $m \geq 6$.

So $K_{(m)} = 1$ for $m \leq 5$ and $K_{(m)} = 2$ for $m \geq 6$.

Let $K_0 = 2$ and $n = 500$. Compute $\mathbb{E}(\mathscr{T}_{K_0,(m)})$ and $\mathbb{E}(\mathscr{T}^*_{K_0,(m)})$.



- This results suggests a χ^2 test for testing $H_0 : K \leq K_0$ versus $H_a : K > K_0$, which is an extension of a test in Li (1991) to the functional data setting.

- This results suggests a χ^2 test for testing $H_0 : K \leq K_0$ versus $H_a : K > K_0$, which is an extension of a test in Li (1991) to the functional data setting.

- Ideally, case (i) holds and the χ^2 test has the correct size asymptotically.

- This results suggests a χ^2 test for testing $H_0 : K \leq K_0$ versus $H_a : K > K_0$, which is an extension of a test in Li (1991) to the functional data setting.

- Ideally, case (i) holds and the χ^2 test has the correct size asymptotically.

- For a number of reasons case (ii) may be true, for which the χ^2 test will be conservative.

Proposition

Let Z be a $p \times q$ random matrix and we write $Z = [\ Z_1 \mid Z_2\ ]$ where Z_1 and Z_2 have sizes $p \times r$ and $p \times (q - r)$, respectively, for some $0 < r < \min(p, q)$. Assume that Z_1 and Z_2 are independent, and Z_2 contains i.i.d. Normal $(0, 1)$ entries. Then $\sum_{j=r+1}^{p} \lambda_j(ZZ^T)$ is stochastically bounded by χ^2 with $(p - r)(q - r)$ degrees of freedom.

Proposition

Let Z be a $p \times q$ random matrix and we write $Z = [\ Z_1 \mid Z_2\ ]$ where Z_1 and Z_2 have sizes $p \times r$ and $p \times (q-r)$, respectively, for some $0 < r < \min(p,q)$. Assume that Z_1 and Z_2 are independent, and Z_2 contains i.i.d. Normal $(0,1)$ entries. Then $\sum_{j=r+1}^{p} \lambda_j(ZZ^T)$ is stochastically bounded by χ^2 with $(p-r)(q-r)$ degrees of freedom.

The case where Z is a matrix of i.i.d. Normal$(0,1)$ entries can be viewed as the special case, $r = 0$. For that the bound is exact since $\sum_{j=1}^{p} \lambda_j(ZZ^T)$ equals the sum of squares of all of the entries of Z and is therefore distributed as χ^2 with pq degrees of freedom.

Proof

$$\lambda_1(ZZ^T) = \max_{\|\mathbf{v}\|=1} (\mathbf{v}^T ZZ^T \mathbf{v})$$

Proof

$$\lambda_1(ZZ^T) = \max_{\|\mathbf{v}\|=1} (\mathbf{v}^T ZZ^T \mathbf{v})$$

$$\begin{aligned} &\sum_{j=r+1}^{p} \lambda_j(ZZ^T) \\ = \; &\min_{\Phi} \Big\{ \mathrm{tr}(\Phi^T ZZ^T \Phi), \; \Phi \text{ is a } p \times (p-r) \\ &\qquad\qquad \text{matrix with orthonormal columns} \Big\} \end{aligned}$$

Construct Φ by Gram-Schmidt procedure on the columns of Z.

Adjusted test

- For the non-Gaussian case, we need to eliminate nuisance parameters in the limit.
  - Let $\widehat{P}$ be the matrix whose columns are the eigenvectors that correspond to the $m - K_0$ smallest eigenvalues of $\widehat{V}$.
  - $\widehat{\tau}_h := \frac{1}{(m-K_0)n_h} \mathrm{tr}\{\widehat{P}\widehat{P}^T \sum_{i=1}^{n} (\widehat{\boldsymbol{\eta}}_{i,(m)} - \bar{\boldsymbol{\eta}}_{h,(m)}) \times (\widehat{\boldsymbol{\eta}}_{i,(m)} - \bar{\boldsymbol{\eta}}_{h,(m)})^T I(Y_i \in S_h)\}$
  - $\widehat{\Lambda} := \mathrm{diag}(\widehat{\tau}_1^{1/2}, \cdots, \widehat{\tau}_H^{1/2})$
  - $\widehat{\mathscr{J}} := I - (\widehat{p}_1^{1/2}, \cdots, \widehat{p}_H^{1/2})^T (\widehat{p}_1^{1/2}, \cdots, \widehat{p}_H^{1/2})$
  - $\widehat{G} = \mathrm{diag}\{\widehat{p}_1^{1/2}, \cdots, \widehat{p}_H^{1/2}\}$
  - $\widehat{M} = [\bar{\boldsymbol{\eta}}_{1,(m)}, \cdots, \bar{\boldsymbol{\eta}}_{H,(m)}]_{m \times H}$
- $\widehat{W}_{(m)} = \widehat{M}\widehat{G}\widehat{\mathscr{J}}\widehat{\Lambda}(\widehat{\Lambda}\widehat{\mathscr{J}}\widehat{\Lambda})^{-}(\widehat{M}\widehat{G}\widehat{\mathscr{J}}\widehat{\Lambda}(\widehat{\Lambda}\widehat{\mathscr{J}}\widehat{\Lambda})^{-})^T$

Consider the following test statistic

$$\mathscr{T}^*_{K_0,(m)} = n \sum_{j=K_0+1}^{m} \lambda_j(\widehat{W}_{(m)}).$$

Consider the following test statistic

$$\mathscr{T}^*_{K_0,(m)} = n \sum_{j=K_0+1}^{m} \lambda_j(\widehat{W}_{(m)}).$$

Theorem

Suppose X has an elliptically contoured distribution. Assume that the true dimension $K \le K_0$ and let $H > K_0 + 1$ and $m \ge K_0 + 1$. If $K_{(m)} = K_0$ then $\mathscr{T}^*_{K_0,(m)} \xrightarrow{d} \chi^2_{(m-K_0)(H-K_0-1)}$ as $n \to \infty$.

Discussions

- In our example, we need $m \geq 6$ to discover 2 directions.

Discussions

- In our example, we need $m \geq 6$ to discover 2 directions.
- A practical way of choosing m is making sure that the first m sample principal components explain most of the variation.

Discussions

- In our example, we need $m \geq 6$ to discover 2 directions.
- A practical way of choosing m is making sure that the first m sample principal components explain most of the variation.
-

What about choosing m liberally?

Discussions

- In our example, we need $m \geq 6$ to discover 2 directions.
- A practical way of choosing m is making sure that the first m sample principal components explain most of the variation.
-

What about choosing m liberally?

When the sample size is fixed, the power of the tests will decrease with m. Intuitively, when m is large the information contained in the individual components becomes diluted. This is also true for the "small n, large p" situation of multivariate data.

Discussions

- In our example, we need $m \geq 6$ to discover 2 directions.
- A practical way of choosing m is making sure that the first m sample principal components explain most of the variation.
- What about choosing m liberally?

  When the sample size is fixed, the power of the tests will decrease with m. Intuitively, when m is large the information contained in the individual components becomes diluted. This is also true for the "small n, large p" situation of multivariate data.
- If m is chosen appropriately and the level α is fixed for all of the tests, then the probability of correct identification of K tends to $1 - \alpha$ as $n \to \infty$. The choice of H is less important.

Adaptive Neyman test for the normal mean

Let $\mathbf{X} \sim N(\boldsymbol{\theta}, I_n)$ be an n-dimensional normal random vector. We wish to test $H_0 : \boldsymbol{\theta} = 0$ versus $H_a : \boldsymbol{\theta} \neq 0$.

- The maximum likelihood ratio test is: reject H_0 if $\|\mathbf{X}\|^2$ is large.

Adaptive Neyman test for the normal mean

Let $\mathbf{X} \sim N(\boldsymbol{\theta}, I_n)$ be an n-dimensional normal random vector. We wish to test $H_0 : \boldsymbol{\theta} = 0$ versus $H_a : \boldsymbol{\theta} \neq 0$.

- The maximum likelihood ratio test is: reject H_0 if $\|\mathbf{X}\|^2$ is large.
- The power of the test tends to the size of the test for any alternative $\boldsymbol{\theta}$ satisfying $\|\boldsymbol{\theta}\| = o(n^{1/4})$.

Adaptive Neyman test for the normal mean

Let $\mathbf{X} \sim N(\boldsymbol{\theta}, I_n)$ be an n-dimensional normal random vector. We wish to test $H_0 : \boldsymbol{\theta} = 0$ versus $H_a : \boldsymbol{\theta} \neq 0$.

- The maximum likelihood ratio test is: reject H_0 if $\|\mathbf{X}\|^2$ is large.
- The power of the test tends to the size of the test for any alternative $\boldsymbol{\theta}$ satisfying $\|\boldsymbol{\theta}\| = o(n^{1/4})$.
- Fan (1996) and Fan and Lin (1998) considered the test statistic

$$T_N = \max_{1 \leq m \leq N} \frac{1}{\sqrt{2m}} \sum_{j=1}^{m} (X_j^2 - 1).$$

"Adaptive Neyman" procedure

For $N > K_0$ define

$$M_N = \max_{K_0+1 \le m \le N} \frac{\mathscr{T}_{K_0,(m)} - (m - K_0)(H - K_0 - 1)}{\sqrt{2(m - K_0)(H - K_0 - 1)}}.$$

"Adaptive Neyman" procedure

For $N > K_0$ define

$$M_N = \max_{K_0+1 \le m \le N} \frac{\mathscr{T}_{K_0,(m)} - (m - K_0)(H - K_0 - 1)}{\sqrt{2(m - K_0)(H - K_0 - 1)}}.$$

Theorem

Suppose X is a Gaussian process. Assume that $K \le K_0$ and let $H > K_0 + 1$. Let χ_i^2, $i \ge 1$, be i.i.d. χ^2 random variables with $H - K_0 - 1$ degrees of freedom and define

$$\mathscr{X}_{(m)} = \sum_{i=1}^{m-K_0} \chi_i, \quad m \ge K_0 + 1.$$

Then, for all positive integers $N > K_0$, the collection of test statistics $\mathscr{T}_{K_0,(m)}, m = K_0 + 1, \cdots, N$, are jointly stochastically bounded by $\mathscr{X}_{(m)}, m = K_0 + 1, \cdots, N$, as n tends to ∞.

Simulations

- $$X(t) = \sum_{k=1}^{\infty} \omega_{2k-1}^{1/2} \eta_{2k-1} \sqrt{2} \cos(2k\pi t) + \sum_{k=1}^{\infty} \omega_{2k}^{1/2} \eta_{2k} \sqrt{2} \sin(2k\pi t),$$

  where $\omega_k = 20(k+1.5)^{-3}$ and η_k's $\sim N(0,1)$.

Simulations

- $$X(t) = \sum_{k=1}^{\infty} \omega_{2k-1}^{1/2} \eta_{2k-1} \sqrt{2} \cos(2k\pi t) + \sum_{k=1}^{\infty} \omega_{2k}^{1/2} \eta_{2k} \sqrt{2} \sin(2k\pi t),$$

  where $\omega_k = 20(k+1.5)^{-3}$ and η_k's $\sim N(0,1)$.

- The principal components are

  $$\sqrt{2}\cos(2\pi t), \sqrt{2}\sin(2\pi t), \sqrt{2}\cos(4\pi t), \sqrt{2}\sin(4\pi t), \ldots$$

- $$\begin{aligned}\beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t) \\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t), \\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t) \\ &\quad + 1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4} \frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t).\end{aligned}$$

- $$\begin{aligned}\beta_1(t) &= 0.9\sqrt{2}\cos(2\pi t) + 1.2\sqrt{2}\cos(4\pi t) - 0.5\sqrt{2}\cos(8\pi t)\\ &\quad + \sum_{k>4} \frac{\sqrt{2}}{(2k-1)^3}\cos(2k\pi t),\\ \beta_2(t) &= 0.45\sqrt{2}\cos(2\pi t) + 0.6\sqrt{2}\cos(4\pi t) - 3\sqrt{2}\sin(6\pi t)\\ &\quad +1.2\sqrt{2}\sin(8\pi t) + \sum_{k>4}\frac{(-1)^k\sqrt{2}}{(2k)^3}\sin(2k\pi t).\end{aligned}$$

- $$Y = \langle \beta_1, X\rangle \times (2\langle \beta_4, X\rangle + 1) + \varepsilon,$$

Simulations

- Assume that the X_i's are observed at 501 equally-spaced points on $[0, 1]$. FPCA is performed using Ramsay's `fda` package in `R`.

Simulations

- Assume that the X_i's are observed at 501 equally-spaced points on $[0,1]$. FPCA is performed using Ramsay's `fda` package in `R`.
- Decide the dimension of the EDR space using the χ^2 test and the adjusted χ^2 test. For comparison, we consider $m = 5$ and $m = 7$. (The first 5 PC explain about 90% of the variation in the curves, and the first 7 PC explain roughly 95%.)

Simulations

- Assume that the X_i's are observed at 501 equally-spaced points on $[0,1]$. FPCA is performed using Ramsay's `fda` package in `R`.
- Decide the dimension of the EDR space using the χ^2 test and the adjusted χ^2 test. For comparison, we consider $m = 5$ and $m = 7$. (The first 5 PC explain about 90% of the variation in the curves, and the first 7 PC explain roughly 95%.)
- For the adaptive Neyman test, we choose $N = K_0 + 30$ ($N \sim 30$ to 35).

Simulations

- Assume that the X_i's are observed at 501 equally-spaced points on $[0, 1]$. FPCA is performed using Ramsay's `fda` package in `R`.
- Decide the dimension of the EDR space using the χ^2 test and the adjusted χ^2 test. For comparison, we consider $m = 5$ and $m = 7$. (The first 5 PC explain about 90% of the variation in the curves, and the first 7 PC explain roughly 95%.)
- For the adaptive Neyman test, we choose $N = K_0 + 30$ ($N \sim 30$ to 35).
- We let $\alpha = .05$.

Simulations

	$n = 200$	$n = 500$
χ^2 test ($m = 5$)	0.040	0.047
Adj. χ^2 ($m = 5$)	0.068	0.068
χ^2 test ($m = 7$)	0.358	0.913
Adj. χ^2 ($m = 7$)	0.410	0.899
χ^2 test ($m = 30$)	0.085	0.566
Adj. χ^2 ($m = 30$)	0.170	0.616
Adaptive Neyman	0.229	0.885

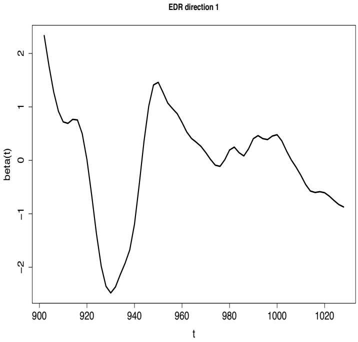
Table: Empirical frequencies of finding the correct model

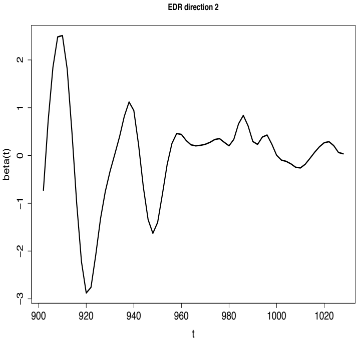
Tecator data

- Following the literature, we used the first 172 samples for training and the last 43 for testing, and we focused on the most informative part of the spectra, with wavelengths ranging from 902 to 1028 nm.

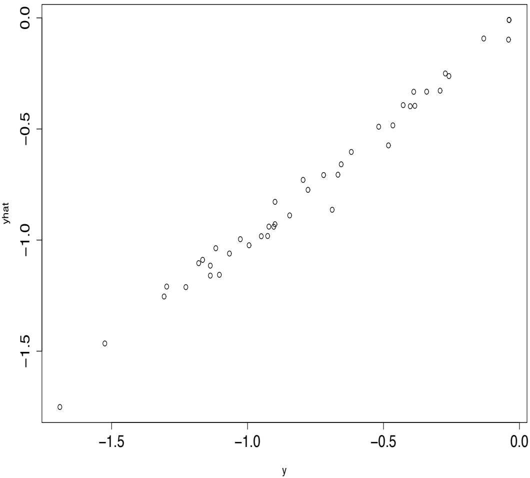
Tecator data

- Following the literature, we used the first 172 samples for training and the last 43 for testing, and we focused on the most informative part of the spectra, with wavelengths ranging from 902 to 1028 nm.
- All of our tests identified EDR dimension $K = 3$.





Tecator data



Concluding remarks

- In this talk we deal with the problem of deciding the dimension of the EDR space in a functional-data setting.

Concluding remarks

- In this talk we deal with the problem of deciding the dimension of the EDR space in a functional-data setting.
- We assume that the predictor is a random function residing in a Hilbert space and has an elliptically contoured distribution, and develop inference procedures based on rigorous statistical tests to determine the dimension of the EDR space.

Concluding remarks

- In this talk we deal with the problem of deciding the dimension of the EDR space in a functional-data setting.
- We assume that the predictor is a random function residing in a Hilbert space and has an elliptically contoured distribution, and develop inference procedures based on rigorous statistical tests to determine the dimension of the EDR space.
- While we focus on infinite-dimensional functional data, all of our results hold without modification for finite-dimensional data, including the "small n, large p" setting for which dimension reduction issues are especially important.

Concluding remarks

- Our procedures are defined by focusing on the information contained in sample principal component scores of the functional data. At the heart of our methodology is an asymptotic representation of the sum of small eigenvalues of the sliced-inverse-regression sample covariance matrix.

Concluding remarks

- Our procedures are defined by focusing on the information contained in sample principal component scores of the functional data. At the heart of our methodology is an asymptotic representation of the sum of small eigenvalues of the sliced-inverse-regression sample covariance matrix.

- In this work we focus on densely recorded functional data, for which standard nonparametric regression techniques can be applied to preprocess the data. A large proportion of commonly seen functional data fall in this category. Many authors have considered principal component analysis for sparsely observed functional data, e.g., data obtained from longitudinal studies. Extending those approaches to the context of this paper requires further research.