Scalar on Function Regression

Jonathan Willnow, Jakob Juergens, Jonghun Baek

Presentation Day

Introduction

Jona
Introductory Example $\rightarrow$ Octane/NIR-spectrum

Theory

Jona
Motivation from multivariate regression (multivariate dgp).

Theory

Jonghun

- Random Functions (name square integrable functions)
- Motivate continuous stochastic processes (growth curves/electricity consumption/yield curves/stonks)
- Use curves to predict a scalar response (show typical dgp)

Theory

Jonghun

- Basis expansions (b-splines and fourier)
- Talk about purposes
- Plots and show bias variance tradeoff

Theory

Jakob

- Random function represented as linear combination of basis functions
- Just transform to multiple linear regression setting
- You already know that from the beginning

Theory - FPCA

Jakob

- Let's assume you know the theory of PCA (pc from varcov matrix)
- Introduce mean and covariance functions of random functions
- There is another cool basis $\rightarrow$ Eigenbasis (Karhunen-Loeve Expansion)
- Sample Analog! (create a basis from observations and use for basis regression)
- Plot fpcs and approximation of function realization

Spectral Representation of Random Vectors

Let $X(\omega)$ be a random vector realizing in $\mathbb{R}^p$.

- Let $\mu_x = \mathbb{E}(X)$ and $\Sigma_X = Cov(X)$
- Let $\{\gamma_i \mid i = 1, \dots, p\}$ be the orthonormal **Eigenvectors** of Σ_X
- Let $\{\lambda_i \mid i = 1, \dots, p\}$ be the corresponding **Eigenvalues** of Σ_X

Then X can also be represented as

$$X(\omega) = \mu_x + \sum_{i=1}^{p} \xi_i(\omega)\gamma_i$$

where the $\xi_i(\omega)$ have the following properties

1. $\mathbb{E}[\xi_i(\omega)] = 0$
2. $Var(\xi_i(\omega)) = \lambda_i$
3. $Cov(\xi_i(\omega), \xi_j(\omega)) = 0$ for $i \neq j$

Principal Component Analysis

Σ_X unknown $\rightarrow$ **sample analogues**

- Let $\mathbf{X} \in \mathbb{R}^{n \times p}$ be the matrix containing the standardized regressors in the usual configuration.
- Let $\hat{\Sigma}_X = \frac{\mathbf{X}'\mathbf{X}}{n}$
- Let $\{\hat{\gamma}_i \,|\, i = 1, \dots, p\}$ be the orthonormal **Eigenvectors** of $\hat{\Sigma}_X$
- Let $\{\hat{\lambda}_i \,|\, i = 1, \dots, p\}$ be the corresponding **Eigenvalues** of $\hat{\Sigma}_X$

Karhunen-Loéve Expansion

Mean Function:

$$\mu(t) = \mathbb{E}\left[F(\omega)(t)\right]$$

Autocovariance Function:

$$c(t,s) = \mathbb{E}\Big[\,(F(\omega)(t) - \mu(t))\,(F(\omega)(s) - \mu(s))\,\Big]$$

The **Eigenvalues** and **Eigenfunctions**: $\{(\lambda_i, \nu_i) \mid i \in \mathcal{I}\}$ are solutions of the following equation:

$$\int_0^1 c(t,s)\nu(s)\mathrm{d}s = \lambda\nu(t)$$

Karhunen-Loéve Expansion

A random function F can be expressed in terms of its mean function and its Eigenfunctions:

$$F(\omega)(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j(\omega)\nu_j(t)$$

Where the ξ_j are scalar-valued random variables with the following properties.

1. $\mathbb{E}[\xi_i(\omega)] = 0$
2. $Var(\xi_i(\omega)) = \lambda_i$
3. $Cov(\xi_i(\omega), \xi_j(\omega)) = 0$ for $i \neq j$

This representation is called the **Karhunen-Loéve Expansion** of the random function F and the Eigenfunctions can serve as a basis to represent the function.

Simulation Setup & Application

Jona

- Compare b-spline / fourier regression chosen via criterion (cv/aic/...)
- Similar for fpca
- generate new curves from observed curves motivated by Karhunen-Loeve expansion
- Compare optimal variants with test and training sets
- Connect to Application

Summary

Jona
Just summarize what we have done...

further reading

Put footnotes here!