

# Scalar on Function Regression with Applications to Near-Infrared Spectrography

Jonathan Willnow, Jakob Juergens, Jonghun Baek

Presentation Day

# Introduction

- **Near-Infrared (NIR) Spectroscopy** enables fast diagnostics by using the NIR region of the electromagnetic spectrum (from 780 nm to 2500 nm)
- Suited for field-monitoring / on-line analysis
- Spectroscopy results in high-dimensional dataset.
- This set of measurements serves as set of discretized approximations of smooth spectral curves
- Regression to determine relationship between octane rating and spectral curves

# Theory

A simple functional dataset is given by

$$\{x_i(t_{j,i}) \in \mathbb{R} \mid i = 1, 2, \dots, N, j = 1, 2, \dots, J_i, t_{j,i} \in [T_1, T_2]\}$$

- Continuous underlying process, where  $x_i(t)$  exists  $\forall t \in [T_1, T_2]$
- Only observed at  $x_i(t_{j,i})$
- Growth curves, financial data, human perception (pitch), ...
- To abstract information from the curves, they must be interpretable!

# Random Function

A **Random Variable** is a function  $X : \Omega \rightarrow \mathcal{S}$  which is defined on a common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  where  $\Omega$  is a probability space with a  $\sigma$ -algebra  $\mathcal{F}$  and a probability measure  $\mathbb{P}$ .

- If  $\mathcal{S} = \mathbb{R}$  then  $X$  is a random variable
- If  $\mathcal{S} = \mathbb{R}^n$  then  $X$  is a random vector
- If  $\mathcal{S}$  is a space of functions,  $X$  is called a **Random Function**

# Random Function

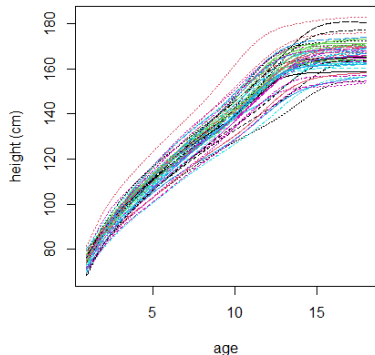
Let  $\mathbb{E}$  be the index set and this can be described as

$$\{X(t, \omega) : t \in \mathbb{E}, \omega \in \Omega\},$$

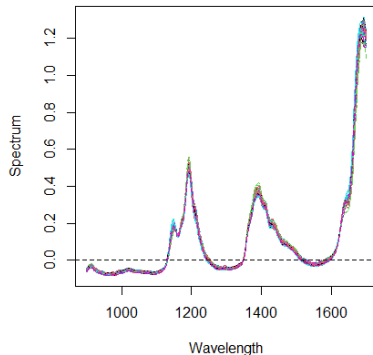
where  $X(t, \cdot)$  is  $\mathcal{F}$ -measurable function on the sample space  $\Omega$ .

- It can be shortened to  $X(t)$  by omitting  $\omega$
- The function is realized when the  $X(t)$  exists  $\forall t \in \mathbb{E}$

# Plots



Growth curves of  
54 girls age 1-18



NIR spectrum of 60  
gasoline samples

# Square Integrable Function

If a function  $f(t)$  satisfies

$$\int_0^1 (f(t))^2 dt < \infty$$

the function  $f(t)$  is called **Square Integrable Function** written  $f(t) \in \mathbb{L}^2[0, 1]$ .

- Without loss of generality, the interval is defined in  $[0, 1]$
- $\mathbb{L}^2$  is the set of all square integrable functions
- We focus on  $\mathbb{L}^2[0, 1]$  since the domain of our function is on the real line.

# Square Integrable Function

Let  $f, g \in \mathbb{L}^2[0, 1]$ , then

$$\langle f, g \rangle = \int_0^1 f(t)g(t)dt$$

Orthogonality of two different functions  $\langle f, g \rangle = 0$



# Basis Expansion

**Basis Expansion** is a linear combination of functions defining a function as described:

$$X_i(t) = \sum_{k=1}^{\infty} c_{ik} \phi_k(t) \approx \sum_{k=1}^K c_{ik} \phi_k(t), \quad i = 1, \dots, n, \quad \forall t \in \mathbb{E}$$

where  $\phi_k(t)$  is the  $k^{th}$  basis function of the expansion and  $c_{ik}$  is the corresponding coefficient. We truncate the basis at  $K$  to:

- make the function smoother
- replace the original curves  $X_i(t)$  by a smaller collection of  $c_{nm}$

# Two Typical Types of Basis Function

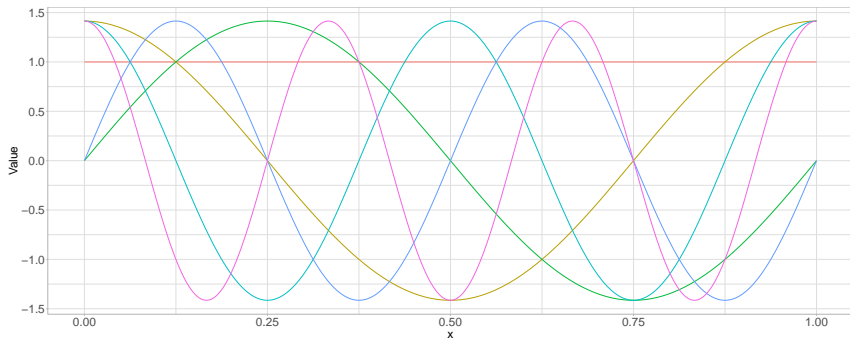
**Fourier Basis Function** is written as

$$f(x) = a_0 + \sum_{n=1}^{\infty} a_n \cos(2\pi nx) + b_n \sin(2\pi nx)$$

**B-spline Basis Function** is a flexible curve defined by degree and knots.

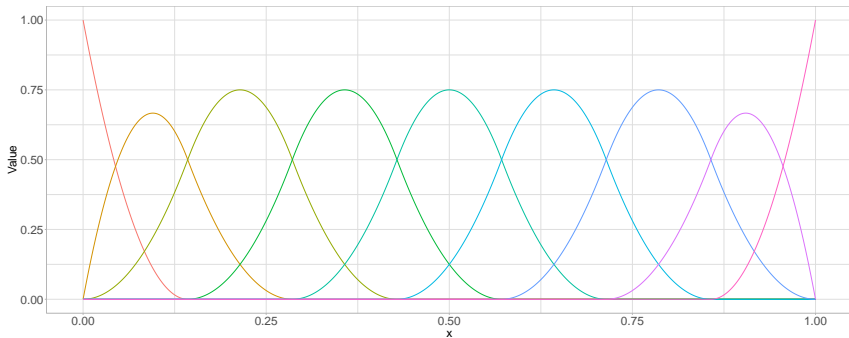
Each B-spline basis function,  $i$ -th B-spline basis function of degree  $p$ ,  $N_{i,p}(u)$  is defined on Cox-de Boor recursion formula.

# Plots of Basis Functions



Fourier basis functions with order 9

# Plots of Basis Functions



Bspline basis functions with order 9

# Trade Off between Bias and Variance

How do we choose the number  $K$  of basis functions?

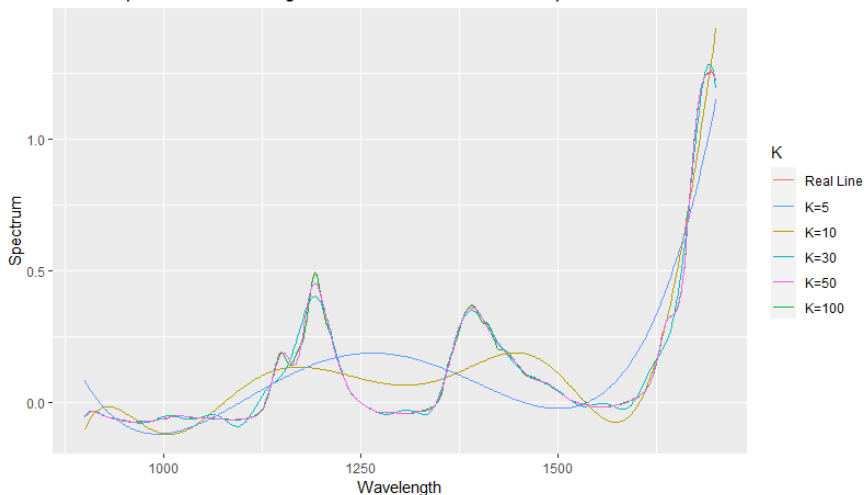
$$\mathbf{MSE}[\hat{x}(t)] = \mathbf{Bias}^2[\hat{x}(t)] + \mathbf{Var}[\hat{x}(t)]$$

## Integrated Mean Squared Error

- The larger  $K$ , the better fit to the data with also fitting noise
- If  $K$  is too small, it would miss some significant information that we want to estimate

# Trade Off between Bias and Variance

Basis expansions according to the different number of B-spline basis functions



# Estimation via Basis Representation

Assume the following **Data Generating Process**

$$Y(\omega) = \alpha + \int_0^1 \beta(s)X(\omega)(s)ds + \epsilon(\omega)$$

- $Y(\omega)$  and  $\epsilon(\omega)$  realize in  $\mathbb{R}$  and  $X(\omega)$  realizes in  $\mathbb{L}^2[0, 1]$

Let  $\{\phi_i(t) \mid i = 1, \dots, \infty\}$  be a basis of  $\mathbb{L}^2[0, 1]$  leading to the following representation of  $\beta(t)$

$$\beta(t) = \sum_{j=1}^{\infty} c_j \phi_j(t) \approx \sum_{j=1}^L c_j b_j(t)$$

# Estimation via Basis Representation

We can transform the data generating process into:

$$\begin{aligned} Y(\omega) &= \alpha + \int_0^1 \left[ \left( \sum_{j=1}^{\infty} c_j \phi_j(s) \right) X(\omega)(s) \right] ds + \epsilon(\omega) \\ &= \alpha + \sum_{j=1}^{\infty} \left[ c_j \int_0^1 X(\omega)(s) \phi_j(s) ds \right] + \epsilon(\omega) \\ &= \alpha + \sum_{j=1}^{\infty} c_j Z_j(\omega) + \epsilon(\omega) \end{aligned}$$

Where a  $Z_j(\omega)$  is a **scalar random variable**.



# Estimation via Basis Representation

Each combination of  $x_i(t)$  and  $\phi_j(t)$  gives us

$$Z_{i,j} = \int_0^1 x_i(s) b_j(s) ds$$

- This allows us to write each observation in the data set as  $(y_i, Z_{i,1}, Z_{i,2}, \dots)$
- Truncating the functional basis yields an approximation  $(y_i, Z_{i,1}, Z_{i,2}, \dots, Z_{i,L})$

Coefficients can then be estimated using theory from **multivariate regression** leading to an estimated coefficient vector  $\hat{c} \in \mathbb{R}^L$ .

# Estimation via Basis Representation

This can be translated into an estimated coefficient function  $\hat{\beta}(t)$ :

$$\hat{\beta}_L(t) = \sum_{j=1}^L \hat{c}_{Lj} \phi_j(t)$$

This is dependent on...

- The basis  $(\phi_j(t))_{j \in \mathcal{I}}$  for the estimation of  $\beta(t)$
- The truncation parameter  $L$
- The basis  $(\psi_j(t))_{j \in \mathcal{I}}$  used for the expansion of the observations
- The truncation parameter in the approximation of the observations  $K$

# Karhunen-Loève Expansion

**Mean Function:**

$$\mu(t) = \mathbb{E} [X(\omega)(t)]$$

**Autocovariance Function:**

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

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

$$\int_0^1 c(t, s) \nu(s) ds = \lambda \nu(t)$$

# Karhunen-Loève Expansion

A random function  $X(\omega)$  can be expressed in terms of its mean function and its Eigenfunctions:

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

Where the  $\xi_j$  are scalar-valued random variables with the following properties.

1  $\mathbb{E}[\xi_i(\omega)] = 0$

2  $\text{Var}(\xi_i(\omega)) = \lambda_i$

3  $\text{Cov}(\xi_i(\omega), \xi_j(\omega)) = 0$  for  $i \neq j$

This is called the **Karhunen-Loève Expansion** of  $X(\omega)$  and the Eigenfunctions can serve as a basis.

# Functional Principal Component Analysis

**Principal Component Analysis** can be extended to functional regressors in the form of **Functional Principal Component Analysis** (FPCA).

**Empirical Mean Function:**

$$\hat{\mu}(t) = \frac{1}{n} \sum_{j=1}^n f_j(t)$$

**Empirical Autocovariance Function:**

$$\hat{c}(t, s) = \frac{1}{n} \sum_{j=1}^n (f_j(t) - \hat{\mu}(t)) (f_j(s) - \hat{\mu}(s))$$

# Functional Principal Component Analysis

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

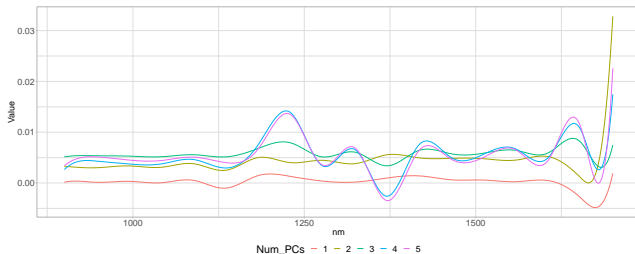
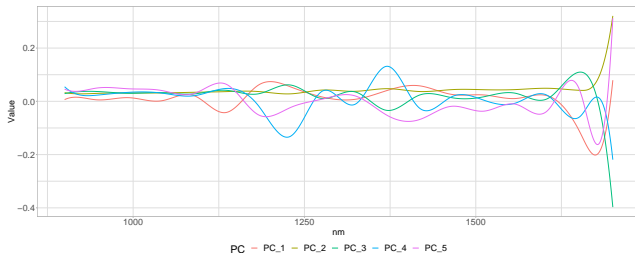
$$\int_0^1 \hat{c}(t, s) \hat{\nu}(s) ds = \hat{\lambda} \hat{\nu}(t)$$

The  $\{\hat{\nu}_i(s) \mid i \in \mathcal{I}\}$  are called **Functional Principal Components** and can serve as a basis for representing the original curves.

The corresponding scores  $\hat{\xi}_i$  can be derived as

$$\hat{\xi}_j(\omega) = \int_0^1 (F(\omega)(s) - \hat{\mu}(s)) \hat{\nu}_j(s) ds$$

# FPCA - Plots



The math is for intuition. In practice there are problems and the fpc's are derived differently.

# Simulation Setup & Application

- Use the **Gasoline Dataset** (NIR-spectroscopy,  $60 \times 401$ ) to predict octane ratings.
- Generate **similar curves** from gasoline dataset:

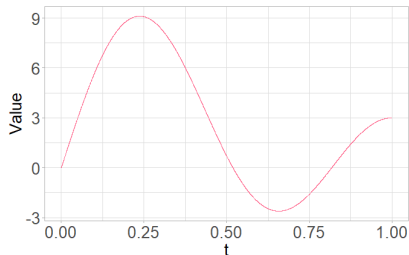
$$\tilde{F}(\omega)(t) = \hat{\mu}(t) + \sum_{j=1}^J \tilde{\xi}_j(\omega) \hat{\nu}_j(t)$$

- $\tilde{\xi}_j \sim \mathcal{N}(0, \hat{\lambda}_j)$  and  $\tilde{\xi}_j \perp \tilde{\xi}_k$  for  $j \neq k$
- Simplification: the  $\xi_j$  do not follow a normal
- $\tilde{F}(\omega)(t)$ ,  $\hat{\mu}(t)$  and  $\hat{\nu}_j(t)$  are approximated as vectors in  $\mathbb{R}^{401}$ .

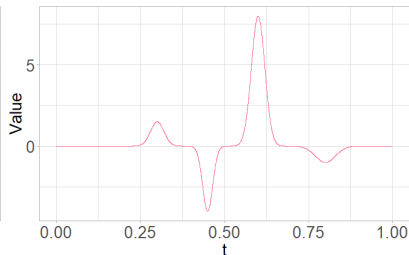


# Simulation Setup & Application cont.

Following **Reiss and Ogden (2007)**, let  $f_1(t)$  and  $f_2(t)$  be two coefficient functions:



$f_1(t)$ , smooth function



$f_2(t)$ , bumpy function

# Simulation Setup & Application cont.

Let

$$Y_{1,f} = \langle NIR, f \rangle + Z \frac{\text{var}(\langle NIR, f \rangle)}{0.9} - \text{var}(\langle NIR, f \rangle)$$

$$Y_{2,f} = \langle NIR, f \rangle + Z \frac{\text{var}(\langle NIR, f \rangle)}{0.6} - \text{var}(\langle NIR, f \rangle)$$

where  $Z \sim \mathcal{N}(0, 1)$  be two responses for  $f \in \{f_1(t), f_2(t)\}$ .

- Four combinations with different number of cubic basis-function  $n_{basis} \in (5, 6, \dots, 25)$  to perform regression using basis expansion and the FPCR approach.
- Compare results via criteria (CV, Mallows CP,...)
- **add results here!**

# Simulation Setup & Application cont.

- Use insights from the simulation study to uncover dependence.
- Similar setup, but using only bsplie basis expansion and initial 60 spectral curves.
- Validation set approach: Scores of testdata needs to be estimated by the trainingdata. **explain in detail?**
- Report results by MSE scaled by variance.

# Summary

Jona

Just summarize what we have done...

Thank you for your time!

# Further Reading

Put footnotes here!

# Spectral Representation of Random Vectors

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

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

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

# Principal Component Analysis

A related concept is **Principal Component Analysis** (PCA).

$\Sigma_X$  unknown  $\rightarrow$  **sample analogues**

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

Then  $Z_i(\omega) = \hat{\gamma}_i' X(\omega)$  is called the  $i$ 'th principal component and

- |   |  |
|---|--|
| 1 $\mathbb{E}[Z_i(\omega)] = 0$               | 3 $\text{Cov}(Z_i(\omega), Z_j(\omega)) = 0$ for |
| 2 $\text{Var}(Z_i(\omega)) = \hat{\lambda}_i$ | $i \neq j$                                       |