Functional Data Analysis

Lecture – 4 & 5

Mathematical foundation & Exploratory Data Analysis

May 8, 2018

Outline

Usual linear regression model

$$y = X\alpha + \epsilon$$

Corresponding assumptions:

Usual linear regression model

$$y = X\alpha + \epsilon$$

Corresponding assumptions: $\epsilon \sim \mathcal{N}(0, \Sigma)$ (major);

Usual linear regression model

$$y = X\alpha + \epsilon$$

Corresponding assumptions: $\epsilon \sim \mathcal{N}(0, \Sigma)$ (major); mostly, we assume that $\Sigma = \mathcal{I}$.

Usual linear regression model

$$y = X\alpha + \epsilon$$

Corresponding assumptions: $\epsilon \sim \mathcal{N}(0, \Sigma)$ (major); mostly, we assume that $\Sigma = \mathcal{I}$.

In fda, the data is usually some realisation of a stochastic process, as against a random variable.

Usual linear regression model

$$y = X\alpha + \epsilon$$

Corresponding assumptions: $\epsilon \sim \mathcal{N}(0, \Sigma)$ (major); mostly, we assume that $\Sigma = \mathcal{I}$.

In fda, the data is usually some realisation of a stochastic process, as against a random variable. Therefore, we're mostly interested in:

$$y(t) = f(t) + \epsilon(t),$$

where we wish to *estimate* $f(t)$, given the observation $y(t)$.

Comparing the finite and the infinite dimensional models

Usual regression

- Mostly linear, or functional form of the dependence is known.

fda

Comparing the finite and the infinite dimensional models

Usual regression

- Mostly linear, or functional form of the dependence is known.

fda

- Functional form of the dependence is not known, thus need some assumptions: the functional space and the basis

We know how to define mean and (co)variance of random vectors, but now we need to define the same for infinite dimensional random elements.

Comparing the finite and the infinite dimensional models

Usual regression

- Mostly linear, or functional form of the dependence is known.
- Estimation of the coefficients relies on the distributional assumptions on noise.

fda

- Functional form of the dependence is not known, thus need some assumptions: the functional space and the basis

We know how to define mean and (co)variance of random vectors, but now we need to define the same for infinite dimensional random elements.

Comparing the finite and the infinite dimensional models

Usual regression

- Mostly linear, or functional form of the dependence is known.
- Estimation of the coefficients relies on the distributional assumptions on noise.

fda

- Functional form of the dependence is not known, thus need some assumptions: the functional space and the basis
- What should be the distribution of noise?

We know how to define mean and (co)variance of random vectors, but now we need to define the same for infinite dimensional random elements.

Comparing the finite and the infinite dimensional models

Usual regression

- Mostly linear, or functional form of the dependence is known.
- Estimation of the coefficients relies on the distributional assumptions on noise.
-

fda

- Functional form of the dependence is not known, thus need some assumptions: the functional space and the basis
- What should be the distribution of noise?

We know how to define mean and (co)variance of random vectors, but now we need to define the same for infinite dimensional random elements.

Outline

A simplest model for the *data space* is **Hilbert space**– a collection of elements which has infinitely many (but countable) basis elements, and one can define an **inner product** between any pair of elements.

A simplest model for the *data space* is **Hilbert space**– a collection of elements which has infinitely many (but countable) basis elements, and one can define an **inner product** between any pair of elements.

For example:

$\ell^2 = \{(a_1, a_2, a_3, \ldots) : a_i \in \mathbb{R}, \text{ and } \sum_{i \geq 1} a_i^2 < \infty\}$.

A simplest model for the *data space* is **Hilbert space**– a collection of elements which has infinitely many (but countable) basis elements, and one can define an **inner product** between any pair of elements.

For example:

$\ell^2 = \{(a_1, a_2, a_3, \ldots) : a_i \in \mathbb{R}, \text{ and } \sum_{i\geq 1} a_i^2 < \infty\}$.

Looks similar to $\mathbb{R}^n$; and we know how to define standard normal on $\mathbb{R}^n$. So what about standard normal on ℓ^2?

A simplest model for the *data space* is **Hilbert space**– a collection of elements which has infinitely many (but countable) basis elements, and one can define an **inner product** between any pair of elements.

For example:

$\ell^2 = \{(a_1, a_2, a_3, \dots) : a_i \in \mathbb{R}, \text{ and } \sum_{i\geq 1} a_i^2 < \infty\}$.

Looks similar to $\mathbb{R}^n$; and we know how to define standard normal on $\mathbb{R}^n$. So what about standard normal on ℓ^2?

Let us recall how to characterise standard normal distribution on $\mathbb{R}^n$:

- density is given by $(2\pi)^{-n/2} \exp\left(-\frac{x^2}{2}\right)$

A simplest model for the *data space* is **Hilbert space**– a collection of elements which has infinitely many (but countable) basis elements, and one can define an **inner product** between any pair of elements.

For example:

$\ell^2 = \{(a_1, a_2, a_3, \ldots) : a_i \in \mathbb{R}, \text{ and } \sum_{i \geq 1} a_i^2 < \infty\}$.

Looks similar to $\mathbb{R}^n$; and we know how to define standard normal on $\mathbb{R}^n$. So what about standard normal on ℓ^2?

Let us recall how to characterise standard normal distribution on $\mathbb{R}^n$:

- density is given by $(2\pi)^{-n/2} \exp\left(-\frac{x^2}{2}\right)$
- or, all linear combinations are Normal with appropriate mean and variance.

Density?

Density is always defined with respect to the Lebesgue measure,

Density?

Density is always defined with respect to the Lebesgue measure, which does not exist when the dimension of the space goes to infinity.

Linear combinations?

Adaptable even to infinite dimensions.

[1] Discuss. Compare with matrices.

Linear combinations?

Adaptable even to infinite dimensions. Just need to know what kind of linear combinations should be admissible.

[1] Discuss. Compare with matrices.

Linear combinations?

Adaptable even to infinite dimensions. Just need to know what kind of linear combinations should be admissible.
One can define Gaussian distribution on ℓ^2 with mean m, and covariance $\mathcal{C}$ as long as

$$m \in \ell^2,$$

and $\mathcal{C}$ is a **linear operator**[1] on ℓ^2 such that

$$\sum_{i \geq 1} \langle \mathcal{C} e_i, e_i \rangle < \infty,$$

where $\{e_i\}$ is an orthonormal basis of ℓ^2.

---

[1] Discuss. Compare with matrices.

Linear combinations?

Adaptable even to infinite dimensions. Just need to know what kind of linear combinations should be admissible.
One can define Gaussian distribution on ℓ^2 with mean m, and covariance $\mathcal{C}$ as long as

$$m \in \ell^2,$$

and $\mathcal{C}$ is a **linear operator**[1] on ℓ^2 such that

$$\sum_{i \geq 1} \langle \mathcal{C} e_i, e_i \rangle < \infty,$$

where $\{e_i\}$ is an orthonormal basis of ℓ^2. But why?

[1] Discuss. Compare with matrices.

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian.

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i \geq 1}$ are i.i.d. standard normal.

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i \geq 1}$ are i.i.d. standard normal. Implying that $\mathcal{C} = \mathcal{I}$ on ℓ^2.

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i\geq 1}$ are i.i.d. standard normal. Implying that $\mathcal{C} = \mathcal{I}$ on ℓ^2.Clearly,

$$\sum_{i\geq 1} \langle \mathcal{I} e_i, e_i \rangle = \infty.$$

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i \geq 1}$ are i.i.d. standard normal. Implying that $\mathcal{C} = \mathcal{I}$ on ℓ^2.Clearly,

$$\sum_{i \geq 1} \langle \mathcal{I} e_i, e_i \rangle = \infty.$$

What is the magnitude/size of such a Gaussian element?

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i\geq 1}$ are i.i.d. standard normal. Implying that $\mathcal{C} = \mathcal{I}$ on ℓ^2.Clearly,

$$\sum_{i\geq 1} \langle \mathcal{I} e_i, e_i \rangle = \infty.$$

What is the magnitude/size of such a Gaussian element?

$$\mathbb{E}\left(\|X\|^2\right) = \mathbb{E}\left(\sum_{i\geq 1} X_i^2\right) = \infty$$

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i \geq 1}$ are i.i.d. standard normal. Implying that $\mathcal{C} = \mathcal{I}$ on ℓ^2.Clearly,

$$\sum_{i \geq 1} \langle \mathcal{I} e_i, e_i \rangle = \infty.$$

What is the magnitude/size of such a Gaussian element?

$$\mathbb{E}\left(\|X\|^2\right) = \mathbb{E}\left(\sum_{i \geq 1} X_i^2\right) = \infty$$

What about $\|X\|^2$ in general?

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i\geq 1}$ are i.i.d. standard normal. Implying that $\mathcal{C} = \mathcal{I}$ on ℓ^2.Clearly,

$$\sum_{i\geq 1} \langle \mathcal{I} e_i, e_i \rangle = \infty.$$

What is the magnitude/size of such a Gaussian element?

$$\mathbb{E}\left(\|X\|^2\right) = \mathbb{E}\left(\sum_{i\geq 1} X_i^2\right) = \infty$$

What about $\|X\|^2$ in general? $\|X\|^2 = \infty$ almost surely.

A quick explanation

Let us consider $X = (X_1, X_2, \ldots)$ an ℓ^2-valued random variable distributed as standard Gaussian. Meaning $\{X_i\}_{i\geq 1}$ are i.i.d. standard normal. Implying that $\mathcal{C} = \mathcal{I}$ on ℓ^2.Clearly,

$$\sum_{i\geq 1} \langle \mathcal{I} e_i, e_i \rangle = \infty.$$

What is the magnitude/size of such a Gaussian element?

$$\mathbb{E}\left(\|X\|^2\right) = \mathbb{E}\left(\sum_{i\geq 1} X_i^2\right) = \infty$$

What about $\|X\|^2$ in general? $\|X\|^2 = \infty$ almost surely. One can actually show that $\|X\|^2 < \infty$, whenever $\mathcal{C}$ is trace class.

Covariance between various linear combinations

In finite dimensions:

On $\mathbb{R}^n$, let $Y \sim \mathcal{N}(\mu, \Sigma)$, then $\langle a, Y \rangle$ and $\langle b, Y \rangle$ are both normally distributed, with covariance $\langle \Sigma a, b \rangle$.

Covariance between various linear combinations

In finite dimensions:
On $\mathbb{R}^n$, let $Y \sim \mathcal{N}(\mu, \Sigma)$, then $\langle a, Y \rangle$ and $\langle b, Y \rangle$ are both normally distributed, with covariance $\langle \Sigma a, b \rangle$.

In infinite dimensions:
On ℓ^2, let $X \sim \mathcal{N}(m, \mathcal{C})$, then $\langle a, X \rangle$ and $\langle b, X \rangle$ are both normally distributed, with covariance $\langle \mathcal{C} a, b \rangle$, i.e.,

$$\langle \mathcal{C} a, b \rangle = \mathbb{E}[\langle a, X \rangle \langle b, X \rangle]$$

Covariance between various linear combinations

In finite dimensions:
On $\mathbb{R}^n$, let $Y \sim \mathcal{N}(\mu, \Sigma)$, then $\langle a, Y\rangle$ and $\langle b, Y\rangle$ are both normally distributed, with covariance $\langle \Sigma a, b\rangle$.

In infinite dimensions:
On ℓ^2, let $X \sim \mathcal{N}(m, \mathcal{C})$, then $\langle a, X\rangle$ and $\langle b, X\rangle$ are both normally distributed, with covariance $\langle \mathcal{C}a, b\rangle$, i.e.,

$$\langle \mathcal{C}a, b\rangle = \mathbb{E}[\langle a, X\rangle\langle b, X\rangle]$$

After some analysis, one can conclude that

$$\mathcal{C} = \sum_{i\geq 1} \lambda_i \; \phi_i \otimes \phi_i \qquad \text{(Mercer's Theorem)}$$

for some ONB $\{\phi_i\}$. (compare with matrices, and discuss L^2 representation).

In fact, there exist a sequence $\{\xi_n\}$ of zero-mean, uncorrelated random variables such that $\mathbb{E}(\xi_i^2) = \lambda_i$, and

$$Y = \sum_{i \geq 1} \xi_i \, \phi_i \qquad \text{(Karhunen–Loéve expansion)}$$

In fact, there exist a sequence $\{\xi_n\}$ of zero-mean, uncorrelated random variables such that $\mathbb{E}(\xi_i^2) = \lambda_i$, and

$$Y = \sum_{i \geq 1} \xi_i \, \phi_i \quad \text{(Karhunen–Loéve expansion)}$$

Special case

In case the functional data is realisation of certain stochastic process, then the mean m is a **mean function** $m(t) = \mathbb{E}(X(t))$, and the covariance operator becomes a **covariance kernel**, $\mathcal{C}(s,t) = \mathrm{cov}(X(s)X(t))$.

Outline

- sample mean;
- sample covariance
- functional PCA (with Karhunen–Loéve –using probe functions)