

**FACULTY
OF MATHEMATICS
AND PHYSICS**
Charles University

MASTER THESIS

Bc. Tomáš Jurica

Functional Data Analysis

Department of Probability, Mathematical Statistics and Econometrics

Supervisor of the master thesis: Doc. RNDr. Zdeněk Hlávka, Ph.D.

Study programme: Mathematics

Study branch: Probability, Mathematical Statistics
and Econometrics

Prague 2021

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In date

Author's signature

I want to thank my supervisor doc. RNDr. Zdeněk Hlávka, Ph.D. for his valuable advice and active help in writing the thesis. I would also like to thank my mother that, despite various obstacles, she still supported me during my studies in Prague and to my best friend Martin Majer for always accepting my lows and highs and nervousness during my whole three years of study at Matfyz. Finally, I want to thank my girlfriend Vesna Jarina, for the endless love, patience, support and positive energy she brought to my life. Without you, studying would be much more challenging than it actually was.

Title: Functional Data Analysis

Author: Bc. Tomáš Jurica

Department: Department of Probability, Mathematical Statistics and Econometrics

Supervisor: Doc. RNDr. Zdeněk Hlávka, Ph.D., Department of Probability, Mathematical Statistics and Econometrics

Abstract: The aim of the master thesis is to review of reconstruction techniques of functional data and existing one-way functional ANOVA (FANOVA) tests. Specifically, the work deals with $\mathcal{L}^2$ -norm based and F -type mean functions equality tests, $\mathcal{L}^2$ -norm based covariance functions equality tests and tests for distribution equality. Furthermore, for each type of the test, it is introduced test based on reconstructed functional data, using orthonormal basis functions of $\mathcal{L}^2$ space. Finally, simulation study was conducted for comparing properties of tests using orthonormal basis representation of functional data and tests applied on non-reconstructed data.

Keywords: functional data, one-way functional anova, functional data reconstruction

Contents

Introduction	3
1 Review of the theory for functional data	4
1.1 Examples of functional data	4
1.2 Definition of the random functions	5
1.3 Statistical model for functional data	7
1.4 Estimation of the mean function and covariance function	8
1.5 Asymptotic properties of the functional estimators	10
2 Reconstruction of functional data	14
2.1 Linear smoothing – orthonormal basis expansion	15
2.1.1 Example of orthonormal basis expansion	17
2.2 Kernel smoothing	18
2.2.1 Example of kernel smoothing	20
3 Functional Analysis of Variance (FANOVA)	21
3.1 FANOVA notation	23
3.2 Tests for equality of mean functions	25
3.2.1 Parametric bootstrap $\mathcal{L}^2$ - norm based test	26
3.2.2 Bias reduction method $\mathcal{L}^2$ -norm based test	28
3.2.3 $\mathcal{L}^2$ -norm based test using orthonormal basis expansion	32
3.2.4 F -type tests	33
3.2.5 F -type test using data reconstructions	39
3.3 Tests for equality of covariance functions	41
3.4 Tests for equality of distributions	43
3.4.1 Functional ANOVA based on the empirical characteristic functions	44
4 Simulation study	49
4.1 Simulations of equality of mean functions tests	49
4.1.1 Simulations under validity H_0	51
4.1.2 Empirical level	51
4.1.3 Simulation under H_1	55
4.1.4 Empirical power	55
4.2 Simulations for equality of covariance functions tests	60
4.2.1 Empirical size	60
4.2.2 Empirical power	60
4.3 Simulations for equality of distribution tests	64
4.3.1 Empirical levels	64
4.3.2 Empirical powers	64
Conclusion	68
Bibliography	71
List of Figures	73

List of Tables	74
List of Abbreviations	75
A Useful theorems and definitions	76
A.1 Hilbert spaces	76
A.2 Important results from Stochastic Processes theory	76
A.3 Additional proofs	81
A.4 Other simulation results	83

Introduction

In recent years, the term Functional Data Analysis (FDA) was rapidly increasing across the literature. The applications of functional data can be found in many scientific fields. For instance, Zhang et al. [2019] use functional data in the ischemic heart screening for detecting differences in normal and ischemia groups. Aguilera et al. [2021] used functional data for testing the homogeneity of three technical devices fabricated by different materials in Resistive Random Access Memories (RRAM's).

The functional data is a type of data which we can perceive either as a continuous stochastic process, with some mean function $\mu(t)$ and covariance function $\gamma(s, t)$, or as a smooth random function where its domain $\mathcal{T}$ is some finite interval (for example days, years, weeks, etc.). Both definitions are very similar. We usually make various assumptions about these functions such as smoothness or continuity or we require that they need to be specified on some space of function (e.g. $\mathcal{L}^p$ space). One of the common issues in the field of FDA is the form in which we observe functional data. In real world applications, it is almost impossible to observe data in every point of $\mathcal{T}$, and we are only able to observe some finite number of functional data. Moreover, it may happen that the observed functional data are evaluated on the different time points, and have small sample size, which lead to very sparse data. Consequently, the previous assumptions are not valid. A solution can be proposed in form of an application of some well-known smoothing methods to reconstruct such observed data to smoother functions evaluated at any time point of $\mathcal{T}$. After that, the statistical inference, such as hypothesis testing about the functional data should be valid.

The aim of this thesis is mainly focused on hypothesis testing, including functional ANOVA (FANOVA) testing problem and usage of reconstruction methods in this problem. Furthermore, we will present two approaches to the FANOVA problem. One approach is to describe the FANOVA problem using observed functional data, the second approach is FANOVA based on the reconstructed functional data.

In Chapter 1, we will introduce two practical examples to demonstrate the problems we can encounter in FDA. Next, we will describe the statistical model for functional data and necessary theory for $\mathcal{L}^2(\mathcal{T})$ space.

Chapter 2 is focused on reconstruction methods for functional data, mainly on linear and kernel smoothing methods. We introduce the main linear smoothing method – smoothing with the help of orthonormal basis functions. For kernel smoothing methods, we review the most common estimators. We will end the chapter with a practical implementation of both techniques on functional data.

The FANOVA problem is described in Chapter 3. The chapter is divided into three sections describing: the equality mean functions tests (EMFT), the equality covariance functions tests (ECFT), and the equality distributions tests (EDT). For each class of every test, we will show both approaches of the FANOVA problem.

In Chapter 4, we will conduct various simulations for comparing performances of the tests from Chapter 3. The main aim of the simulation study was to investigate the empirical power and size under various scenarios, e.g the size of the correlation in data, the size of the error, the level of data sparsity (or density) or if it useful to use tests based on orthonormal basis functions representation or not.

1. Review of the theory for functional data

In this Chapter, we will introduce two examples to demonstrate the practical issues we can encounter in functional data analysis (FDA). In the remaining sections, we will review all necessary theory for statistical inference of functional data.

1.1 Examples of functional data

Example 1. (Ramsey [2005]): Figure 1.1 shows the heights of ten boys measured at a discrete non-equidistant set of 31 distinct ages. Every point represents the height measured at a specific age. For purposes of this example, each point has been connected with a line, thereby creating ten curves. Every „curve” can be described by the following model (not necessarily correct)

$$X_i(t) = \mu(t) + \epsilon_i(t), i = 1, \dots, 10, t \in [0, 18], \quad (1.1)$$

where t is the age of the i -th boy, and for fixed $t \in \mathcal{T}$, $\epsilon(t) = (\epsilon_1(t), \dots, \epsilon_n(t))^\top \sim \mathcal{N}_n(\mathbf{0}_n, \Sigma)$ is a vector of measurement errors (with each error generated from some stochastic process such as Gaussian or MA(q)) and finally $\mu(t)$ is the called **mean function**. The mean function represents the overall trend of the evolution of the heights in the population from the which boys are chosen. Hence, based on the form of the model (1.1), heights of ten boys should be treated as a sample consisting of ten random curves, where each curve has form described by (1.1). Every random value in point t can be considered as functional value of height evaluated at each age. At the same time, it is often useful to plot the mean function, or the mean trend of the functional sample (the green line in Figure 1.1).

The main problem we face is the inability to work with uncountable real interval described in (1.1). Comparing curves shown in Figure 1.1 with the model (1.1) the difference becomes obvious. There are only 31 discrete points for each boy. Therefore, when working with real data, the interval should be replaced by the set of discrete time points $\{t_1, \dots, t_k\}$, where k is the number of discrete time points. In our example $k = 31$. This topic will be discussed in more detail in Chapter 2.

Example 2. Longitudinal data can be also treated as functional data. Figure 1.2 shows marginal bone levels [mm] in a jaw, under an installed tooth implant, during the years after the installation of the implant. The data was used for analysing the dependence of bone level evolution on various factors. As in the first example, each repeated measurement can be described as some functional at each year of the follow-up. The typical method used to analyse these type of data are LME models or GEE (for considering the within subject dependency). Another potential model (again, not necessarily correct and considering the discrete time point set) can be formulated as follows.

$$Y_{ij}(t_{ij}) = \mu(t_{ij}) + b_{i0} + b_{i1}t_{ij} + \epsilon_i, i = 1 \dots, 50, j = 1 \dots, n_i \quad (1.2)$$

Where $\mu(t_{ij}) = \beta_0 + \beta_1 t_{ij}$ is the mean function, $(b_{i0}, b_{i1})^\top$, are random effects independent

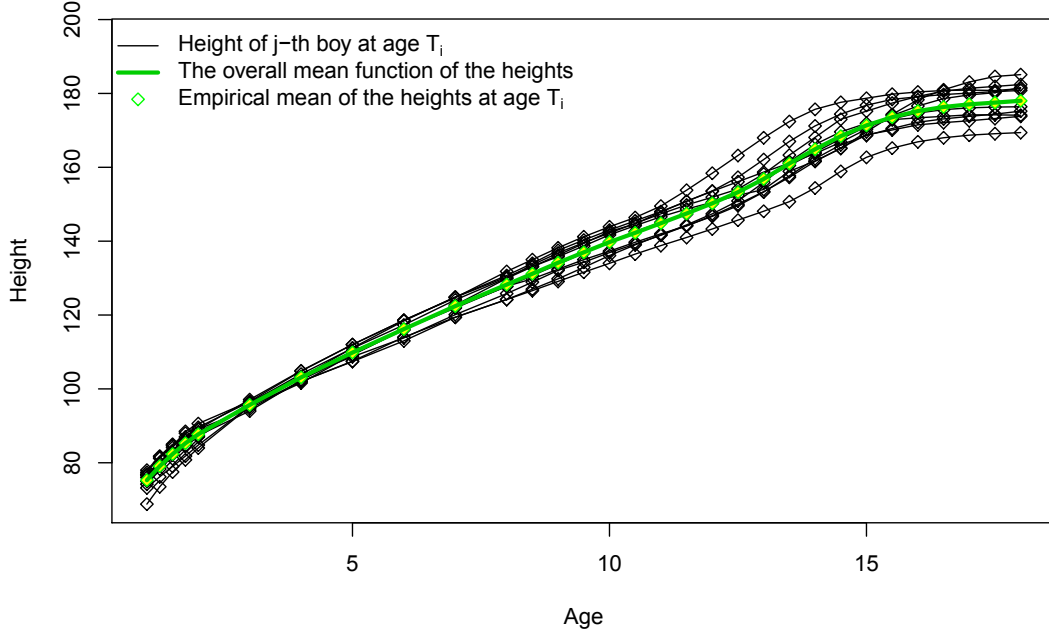


Figure 1.1: The heights of ten boys measured at 31 not equidistant ages (the black lines). Its mean function is represented by green line.

of random noise ϵ_i with $\text{cov}(b_{i0}, b_{i1}) = \sigma_{01}$, $E[b_{ij}] = 0, j = 0, 1$. This implies that $\text{cov}(Y_{ij}, Y_{ik}) = \sigma_{01}, j \neq k$. To sum up, we are considering a normal sample of random curves $\mathbf{Y}_1, \dots, \mathbf{Y}_{50}$, where $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^\top, i = 1, \dots, 50$, which are independent of each other, but they possess a within curve dependency, which is evaluated by σ_{01} . The problem in this example is, that the data are sparse for functional data analysis (or the mean trend is not a smooth function). This problem can be solved by smoothing methods allowing for evaluation of the mean function (or standalone curves) at any point. Therefore, a dense grid of discrete time points can be created for the purposes of smoothing the mean function (and the all curves). Specific smoothing methods will be described later in Chapter 2.

The previous two examples are perfect demonstrations of the type of data we will work with. In general, functional data will be treated as random curves or random processes with some specific mean function $\mu(t), t \in \mathcal{T}$, where $\mathcal{T}$ is a finite interval and a covariance function $\gamma(s, t), s, t \in \mathcal{T}$. Consequently, we need to specify the space, on which functional data are defined. This will be described in more detail in the following section.

1.2 Definition of the random functions

In univariate statistical analysis, we work with random variables defined as a measurable mapping $Z : (\Omega, \mathcal{S}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where Ω is the domain of this mapping or the set of all events (e.g. $\Omega = \mathbb{R}$), $\mathcal{S}$ is σ -algebra of the set Ω , $\mathcal{B}(\mathbb{R})$ is a Borel σ -algebra. If the range of the random variable Z is $\mathbb{R}^k$, we were talking about the random vector, where k is finite

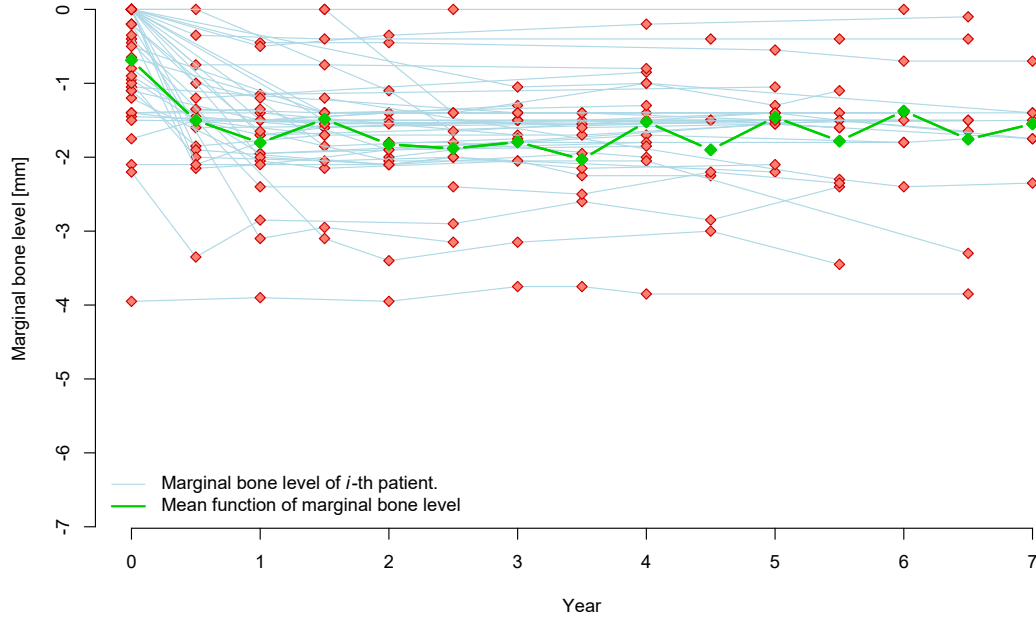


Figure 1.2: The marginal bone levels (blue lines) for 50 patients measured at equidistant time points (every half-year). Its mean function is represented by the green line.

number.

If we look back on the form of model (1.1), one can say that we can treat the functional data as a multivariate random variable, where the dimension is defined by the cardinality of the discretized time point set, or by the number of points in the interval $\mathcal{T}$ (this set is uncountable). In the case of functional data, we must abandon the concept of a multidimensional random variable and treat the data more like infinite dimensional random variables. The reason for such an approach are the limits when treating functional data as a multivariate random variable. Namely, we are limited to analyzing its trend or the way how such variable was constructed. Furthermore, there is no guarantee that the functional data are evaluated at the same time points, or at the same number of time points. Therefore, we should not talk about the random sample as it is usual. It is more appropriate to treat the functional datum as a function (smooth or continuous) and analyse the behavior of this function (derivatives, trend, shape, etc.) on some pre-specified time interval $\mathcal{T}$. One of the main challenges is to specify the statistic model for functional data, which is different across the literature. For example, Kokoszka and Horvath [2012] define such model as a random curve, with the following form

$$X = \{X(t); t \in \mathcal{T}\}, \quad (1.3)$$

where $\mathcal{T}$ is a finite interval and $X(t) : \mathcal{T} \mapsto \mathbb{R}$ is a random function, with values in some functional space and $X(t)$ is a random variable for some fixed point $t \in \mathcal{T}$. A functional random variable is defined by Zhang [2012] as a stochastic process $SP(\mu, \gamma)$ with mean function μ and covariance function γ , again with values in some functional

space. From a theoretical point of view, it is more suitable to treat the random curves as a random process and use the term random curve only in its practical meaning (as in Figure 1.1, where we described the height evolution as curves). Working with functional data as random process with values in some space gives us the ability to make inference about the functional data. Hence we will treat the functional datum X as some stochastic process defined on a probability space Ω , with values in the functional space $\mathcal{H}$, where $\mathcal{H}$ will be a separable Hilbert space, with properties defined in the A.1. Throughout this text we will use $\mathcal{H} = \mathcal{L}^2(\mathcal{T})$, which is the set of measurable real-valued functions g defined on the finite interval $\mathcal{T}$ and satisfying the square integrability condition

$$\int_{\mathcal{T}} g^2(t) dt < \infty. \quad (1.4)$$

As the space $\mathcal{L}^2$ is a separable Hilbert space, the inner product is well defined

$$\langle g, h \rangle = \int_{\mathcal{T}} g(t) h(t) dt, \quad (1.5)$$

and is equipped with the norm

$$\|g\| = \sqrt{\langle g, g \rangle} = \left\{ \int_{\mathcal{T}} g^2(t) dt \right\}^{1/2}. \quad (1.6)$$

Therefore, the distance between two functions f and g can be described as $d(f, g) = \|f - g\|$. Due to the square integrability condition, the functional data will be upper bounded. Moreover, if the condition of square integrability is violated, the norm will be equal to infinity (which also violates the definition of the norm defined on the Hilbert space) and we won't be able to use the properties of $\mathcal{L}^2$ space for analysis of functional data. In the next section, we define the notions above more formally.

1.3 Statistical model for functional data

In general, the starting point of a statistical analysis or statistical inference is random sample. For example, if we are interested in making inference about the weight of a population of men, we use the sequence $z_1, z_2, \dots, z_n$ (a realisation of the random sample $Z_1, \dots, Z_n$) representing the weight from the described population. We estimate the center (theoretical mean) and the variation (variance) parameters, construct confidence intervals for these parameters. In functional data analysis (FDA), we will work with samples of random curves. In Section 1.1, we examined a sample of random curves $X_1(t), \dots, X_{10}(t)$ representing the evolution of the heights of ten boys. Our ultimate goal is to say something about the trend of these curves, its variation, etc.

Throughout this text we will work with a functional random sample $X_1(t), \dots, X_n(t)$, where $t \in \mathcal{T}$ and $X_i(t) \sim SP(\mu(t), \gamma(s, t))$, where SP is notation for general stochastic process with mean function $\mu(t)$ and covariance function $\gamma(s, t)$, $s, t \in \mathcal{T}$. Therefore, for a fixed time point $t \in \mathcal{T}$ we assume that $X_i(t)$ is a random variable defined on a probability space Ω with a realisation $X(\omega)(t)$, $\omega \in \Omega$ in the space $\mathcal{L}^2$. We know, that $\mathcal{L}^2$ is a space of all square integrable functions, and as such the next assumption is

$$\|X(\omega)\|^2 = \int_{\mathcal{T}} \{X(\omega)(t)\}^2 dt < \infty, \quad \omega \in \Omega \quad (1.7)$$

In the field of FDA, it is more appropriate to change the term "square integrability" for the term "existence of the second moment". Consequently, if $\mathbf{E}\|X\|^2 < \infty$, we will say that random function has finite second moment and so exists. Finally, by saying that a random variable is a random element (or an element shortly) of $\mathcal{L}^2(\mathcal{T})$ space, we will mean that the second moment of the random variable exists, more specifically, $X(t) \sim SP(\mu(t), \gamma(s, t))$ with $\mu(t) \in \mathcal{L}^2(\mathcal{T})$. Furthermore, it is suitable to define the mean function and the covariance function as

$$\mu(t) = \mathbf{E}[X(t)], \quad (1.8)$$

$$\begin{aligned} \gamma(s, t) &= \text{cov}[X(t), X(s)] \\ &= \mathbf{E}[(X(t) - \mu(t))(X(s) - \mu(s))], \quad \forall s, t \in \mathcal{T}. \end{aligned} \quad (1.9)$$

The mean and covariance functions are defined on the interval $\mathcal{T}$. If we fix the time point t (or both t and s in the case of the covariance function), the mean function and covariance function is nothing else than the expected value and the covariance of a one-dimensional random variable. The crucial part of FDA is the inference about the aforementioned variables. This will be discussed in the following sections.

1.4 Estimation of the mean function and covariance function

In order to conduct further statistical inference, hypothesis testing and construction confidence intervals, we need to find suitable estimators of the mean and covariance functions, the main characteristics describing the stochastic process from which the functional data comes. The mean function $\mu(t)$ can be estimated by the sample mean function

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^n X_i(t) \quad \forall t \in \mathcal{T}, \quad (1.10)$$

and the covariance function $\gamma(s, t)$ can be estimated by pooled sample covariance

$$\widehat{\gamma(s, t)} = \frac{1}{n-1} \sum_{i=1}^n (X_i(t) - \hat{\mu}(t))(X_i(s) - \hat{\mu}(s)), \quad \forall s, t \in \mathcal{T}. \quad (1.11)$$

The question is, whether both introduced estimators are reasonable (consistent). Before investigating the properties of these estimators, we introduce the following necessary assumption.

Assumption 1. (A1) *The functional observations $X_1(t), \dots, X_n(t)$ are i.i.d. in $\mathcal{L}^2(\mathcal{T})$, square integrable random variables $\forall t \in \mathcal{T}$.*

The assumption above simply states, that we will assume that the random sequence $X_1(t), \dots, X_n(t)$ with $\mu(t) \in L^2(\mathcal{T})$, satisfies (1.7).

Theorem 1.4.1 (Properties of the estimators, Kokoszka and Horvath [2012]). *If assumption A1 is satisfied, then the followings holds:*

- i. $E\hat{\mu}(t) = \mu(t) \forall t \in \mathcal{T}$
- ii. $E\|\hat{\mu}(t) - \mu(t)\|^2 = O(n^{-1})$
- iii. $E\widehat{\gamma(s, t)} = \gamma(s, t)$

Proof.

i.

$$\begin{aligned} E\hat{\mu}(t) &= E\frac{1}{n} \sum_{i=1}^n X_i(t) = \frac{1}{n} \sum_{i=1}^n EX_i(t) = \frac{1}{n} \sum_{i=1}^n \mu(t) \\ &= \mu(t) \end{aligned}$$

- ii. The term $E\|\hat{\mu}(t) - \mu(t)\|^2$ is nothing else than mean squared error $\forall t \in \mathcal{T}$. Therefore, we can use $MSE(\hat{\theta}) = \text{var}(\hat{\theta}) + \text{bias}(\hat{\theta})$. So we have $\forall t \in \mathcal{T}$

$$\begin{aligned} MSE(\hat{\mu}(t)) &= \text{var}\mu(t) + \text{bias}\hat{\mu}(t) \quad (\hat{\mu} \text{ is unbiased estimator}) \\ &= \text{var}\hat{\mu}(t) \\ &= \text{var}\frac{1}{n} \sum_{i=1}^n X_i(t) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var}X_i(t) \quad (\text{var}X(t) = \gamma(t, t)) \\ &= \frac{\gamma(t, t)}{n}. \end{aligned}$$

Thus we obtain $E\|\hat{\mu}(t) - \mu(t)\|^2 \rightarrow 0$ as $n \rightarrow \infty$.

iii.

$$\begin{aligned} (n-1)E\hat{\gamma}(s, t) &= E \sum_{i=1}^n [X_i(t) - \mu(t) - (\hat{\mu}(t) - \mu(t))][X_i(s) - \mu(s) \\ &\quad - (\hat{\mu}(s) - \mu(s))] \quad (Z_i(t) = X_i(t) - \mu(t)) \\ &= E \sum_{i=1}^n [Z_i(t) - \bar{Z}(t)][Z_i(s) - \bar{Z}(s)] \\ &= \sum_{i=1}^n EZ_i(s)Z_i(t) - nE\bar{Z}(t)\bar{Z}(s) \\ &= n\gamma(s, t) - \frac{n}{n^2} \sum_{i,j} E(X_i(t) - \mu(t))(X_j(s) - \mu(s)) \\ &\stackrel{i.i.d}{=} n\gamma(s, t) - \frac{n}{n^2} \sum_i E(X_i(t) - \mu(t))(X_i(s) - \mu(s)) \\ &= n\gamma(s, t) - \gamma(s, t) \end{aligned}$$

Unbiasedness is obtained by dividing the last row of equation by $(n-1)^{-1}$

QED

The previous theorem states, that both introduced estimators of mean and covariance functions are unbiased estimators $\forall t \in \mathcal{T}$. In addition, the mean squared error of the mean function estimator is asymptotically negligible.

1.5 Asymptotic properties of the functional estimators

In statistical analysis, we rarely encounter random variables that will have exact distributions. However, we can approach those distribution asymptotically with the help of the Central Limit Theorems. FDA is no exception. For statistical test, we need to derive the asymptotic distributions of the mean and covariance functions. First, let us introduce the most important theorem required to derive asymptotic distributions of these functions, the central limit theorem for stochastic processes (denoted by FCLT).

Theorem 1.5.1 (CLT for Stochastic Processes, Kokoszka and Horvath [2012]). *Suppose that $Y_1, \dots, Y_n$ is a sequence of i.i.d. zero-mean random variables with values in separable Hilbert space such that the assumption A1 is satisfied. Then as $n \rightarrow \infty$*

$$\frac{1}{\sqrt{n}} \sum_i^n Y_i \xrightarrow{\mathcal{D}} Z, \quad (1.12)$$

where Z is zero-mean Gaussian process (see A.2.1).

Proof. See Andersen and Dobric [1987]

QED

FCLT serves a crucial role in making inference. The assumptions made in Section 1.2, namely the functional data being elements of $\mathcal{L}^2$ (a separable Hilbert space), guarantee that the assumptions of the above theorem will be satisfied throughout the entire text. Up to this point, basic properties of the sample mean function (its asymptotic distribution, asymptotic negligible error and unbiasedness) were shown. The last important property of the sample mean function is introduced by the following theorem.

Theorem 1.5.2. (Bosq [2000]) *Suppose that sequence $Y_1, \dots, Y_n$ is i.i.d. of random elements of separable Hilbert space such that $E\|Y_i\|^2 < \infty$ with $EY_i = \mu$. Then as $n \rightarrow \infty$*

$$\frac{1}{n} \sum_{i=1}^n Y_i \xrightarrow{a.s.} \mu \quad (1.13)$$

Proof. See Bosq [2000]

QED

In the theorem above, consistency is introduced for every random variable with values in a separable Hilbert space. Similarly to Theorem 1.5.1 (and under validity of A1) we can say, that the functional sample $X_1(t), \dots, X_n(t)$ satisfies the assumption above. Moreover, we can say that the sample mean function $\widehat{\mu}(t)$ is a consistent estimator of $\mu(t)$ for every point $t \in \mathcal{T}$. Using the following theorem, the asymptotic distribution of the mean function estimator will be proved.

Theorem 1.5.3 (Functional CLT, Zhang [2012]). *Let $X_1(t), \dots, X_n(t), t \in \mathcal{T}$ are random variables with values in $\mathcal{L}^2(\mathcal{T})$ such that random variables satisfy the assumption A1. Then as $n \rightarrow \infty$*

$$\sqrt{n}[\hat{\mu}(t) - \mu(t)] \rightarrow GP(0, \gamma(s, t)) \quad (1.14)$$

where $GP(0, \gamma(s, t))$ is a zero-mean Gaussian process with covariance function $\gamma(s, t)$.

Proof. Let us rewrite the term on the left side of (1.14) as

$$\begin{aligned}\sqrt{n}[\hat{\mu}(t) - \mu(t)] &= \sqrt{n} \left\{ \frac{1}{n} \sum_{i=1}^n X_i(t) - \mu(t) \right\} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i(t) - \mu(t)) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(t),\end{aligned}$$

where $Z_i(t) = X_i(t) - \mu(t)$. The sequence of random variables $Z_1(t), \dots, Z_n(t)$ is a sequence of i.i.d. zero mean random variables $\forall t \in \mathcal{T}$ with values in $\mathcal{L}^2(\mathcal{T})$. Applying FCLT to $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(t)$ we obtain the limiting distribution.

QED

Example 3. Let $X_1(t), \dots, X_n(t)$ be i.i.d. random curves, which takes the form

$$X_i(t) = A_i \cos(\phi + \lambda t), i = 1, \dots, n,$$

where A_i is random variable, such $EA_i = 0, EA_i^2 < \infty, t \in \mathbb{R}$. ϕ and A are independent and $\phi \sim U(0, 2\pi)$. It is obvious that $E\|X(t)\|^2 < \infty$. Furthermore, we have $EX_i(t) = EA_i \cos(\phi + \lambda t) = 0$. Therefore, A1 and the assumptions of Theorem 1.5.3 are satisfied and we have

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n A_i \cos(\phi + \lambda t) \xrightarrow{\mathcal{D}} GP(0, \gamma(s, t)),$$

where

$$\begin{aligned}\gamma(s, t) &= EX_i(t)X_i(s) \\ &= EA_i^2 E \cos(\phi + \lambda t) \cos(\phi + \lambda s) \\ &= EA^2 \cos \lambda(t - s)\end{aligned}$$

The covariance function describes the shape and the amount of noise of the random curves. In Section 1.4 we derived its unbiasedness and in the following theorem, its asymptotic distribution is introduced.

Assumption 2. (A2) The random sample $X_1(t), \dots, X_n(t)$ satisfy $EX_i(t) = 0$ and $E\|X_i(t)\|^4 < \infty$.

Theorem 1.5.4. (Asymptotic normality for covariance estimator, Zhang [2012]) Let the assumptions A1 and A2 be satisfied. Then as $n \rightarrow \infty$

$$\sqrt{n}[\hat{\gamma}(s, t) - \gamma(s, t)] \xrightarrow{\mathcal{D}} GP(0, \Gamma) \quad (1.15)$$

where $\Gamma = E[X(t)X(s)X(s')X(t')] - \gamma(s, t)\gamma(s', t')$, for $s, t, s', t' \in \mathcal{T}$.

Proof. Let us remind the representation of the estimator $\widehat{\gamma(s, t)}$ from the proof of Theorem 1.4.1

$$\widehat{\gamma(s, t)} = \frac{1}{n-1} \sum_{i=1}^n Z_i(t)Z_i(s) - \frac{n}{n-1} \bar{Z}(t)\bar{Z}(s).$$

where $Z_i(t) = X_i(t) - \mu(t)$ and $\hat{Z}(t) = \frac{1}{n} \sum_{i=1}^n Z_i(t)$. With little manipulation of previous terms

$$\sqrt{n}[\widehat{\gamma}(s, t) - \gamma(s, t)] = \frac{n}{n-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n [Z_i(t)Z_i(s) - \gamma(s, t)] \right\} - \frac{n\sqrt{n}}{n-1} \bar{Z}(t)\bar{Z}(s) - \frac{\sqrt{n}}{n-1} \gamma(s, t).$$

The second term $\frac{n\sqrt{n}}{n-1} \bar{Z}(t)\bar{Z}(s)$ can be written as

$$\frac{n\sqrt{n}}{n-1} \bar{Z}(t)\bar{Z}(s) = \frac{n}{n-1} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n [Z_i(t) - \mathbb{E}Z_i(t)] \right\} \bar{Z}(s) \quad (\mathbb{E}Z_i(t) = 0).$$

We have $[\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(t) - \mathbb{E}Z_i(t)] \xrightarrow{\mathcal{D}} GP(0, \gamma(s, t))$ from functional CLT, $\bar{Z}(s) \xrightarrow{a.s.} 0 \equiv \mathbb{E}Z_i(s)$ from Theorem 1.13 and term $\frac{n}{n-1} \rightarrow 1$ as $n \rightarrow \infty$. Therefore, from Slutsky theorem we have $\frac{n}{n-1} [\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(t) - \mathbb{E}Z_i(t)] \bar{Z}(s) \xrightarrow{\mathcal{D}} 0$. Furthermore $\frac{\sqrt{n}}{n-1} \gamma(s, t)$ is deterministic and converges to zero as $n \rightarrow \infty$. The final step is to prove the convergence of the first term on the right side of the above equation. The random variables $Z_1(t)Z_1(s) - \gamma(s, t), \dots, Z_n(t)Z_n(s) - \gamma(s, t)$ are i.i.d. zero-mean random variables. Furthermore, they satisfy A1 and A2 and we have

$$\begin{aligned} \mathbb{E}||Z(t)Z(s)||^2 &= \mathbb{E} \left[\int_{\mathcal{T}} \int_{\mathcal{T}} Z^2(t)Z^2(s)dsdt \right] \\ &= \mathbb{E} \left[\int_{\mathcal{T}} Z^2(t)dt \int_{\mathcal{T}} Z^2(s)ds \right] \\ &\leq \left[\mathbb{E} \left\{ \int_{\mathcal{T}} Z^2(t)dt \right\}^2 \right] \left[\mathbb{E} \left\{ \int_{\mathcal{T}} Z^2(s)ds \right\}^2 \right] \\ &= \mathbb{E}||Z(t)||^4 < \infty \end{aligned}$$

where in the third row, Cauchy-Schwarz inequality was used. Furthermore, we have from A.2.3

$$\text{cov}[V_1(s)V_1(t), V_1(s')V_1(t')] = \mathbb{E}[V_1(s)V_1(t), V_1(s')V_1(t')] - \gamma(s, t)\gamma(s', t')$$

QED

Example 4. (not every random curve is in $\mathcal{L}^2$) Suppose a sample of i.i.d. random curves $X_1(t), \dots, X_n(t)$, where $X_i(t)$ has the form

$$X(t) = A_i t^{-\frac{5}{4}}, \quad A_i \sim U(0, 1), t \in (0, 1)$$

First, let us verify the A1 is satisfied

$$\begin{aligned} E\|X\|^2 &= E \int_0^1 A^2 t^{-\frac{5}{2}} dt \\ &= EA^2 \infty = \infty. \end{aligned} \tag{1.16}$$

The A1 is not satisfied, therefore the sample of random curves are not elements of $\mathcal{L}^2(\mathcal{T})$ and all derived theory is not applicable and we are not able to estimate $EX(t)$ consistently. However, if the form of the random curves is defined in a more general way as

$$X_i(t) = A_i t^{-n}, \quad A_i \sim U(0, 1), t \in [0, 1], \tag{1.17}$$

for $n \in (-\infty, 1)$ the square integrability is satisfied and we can continue in the inference.

2. Reconstruction of functional data

In FDA we assume that the entire trajectory of the random curve $X(t)$ can be measured. Unfortunately, in real world applications, this is unrealistic. At best such random curves can be measured at a very dense discretized time grid. The ability to do so, at a sufficient level of granularity, is additionally dependent on the availability of computing resources. In practical applications, only pairs $(t_1, g_1), \dots, (t_n, g_n)$ are observed, where g_i is functional data evaluated at time point t_i and sequence $t_1, \dots, t_n$ is a discretized time grid, where $t_1 < \dots < t_n$. Furthermore, it may happen that the measurements are observed at different time point.

In Example 2 we observed pairs $(t_{ik}, g_{ik})^\top, k = 1, \dots, K, i = 1, \dots, n_k$, where g_{ik} represents marginal bone level for k -th patient evaluated at time point t_{ik} and the evaluation time points differ per patient. Moreover, only a small number of per-patient observations for the 7 years follow-up time is available, leading to very sparse data. In such cases, previous introduced theory is not entirely suitable. Furthermore, applying the theory introduced in the previous sections, and estimating only the mean and covariance functions, leads to a loss of information (e.g the trend of the mean function, shape of random curves etc). Different methods should be used to estimate the trend of the (often sparse) observed data and to reconstruct the "true" random process in order to make proper inference. For the purpose of this chapter, the following model for observed data is assumed

$$g_i(t) = f_i(t) + \epsilon_i(t), i = 1, \dots, n \quad (2.1)$$

where $f_i(t)$ is an unknown function, defining the shape of the functional datum and $\epsilon_i(t)$ is a measurement error. The measurement error is, most often, assumed to be some stochastic process $SP(0, \gamma(s, t))$. Furthermore, we assume that the function f is continuous but necessarily smooth on the observation space $\mathcal{S}$.

The model (2.1) introduced in Ramsey [2005], can additionally be extended. For instance, Zhang [2012] proposed that an inclusion of a subject specific error $\tau_i(t)$

$$g_i(t) = \mu(t) + \tau_i(t) + \epsilon_i(t) \quad (2.2)$$

where $f_i(t) = \mathbb{E}[g_i | \tau_i(t)] = \mu(t) + \tau_i(t)$, $\mu(t)$ is the population mean function, $\tau_i(t)$ is the subject specific deviation from that mean, and $\epsilon(t)$ is the population measurement error. Same as in the previous model, we would like to estimate the unknown function $\mu(t)$. In real world applications, assuming that the data is observed at every point of the follow-up time is unrealistic. In fact, we can only get data at a finite number of time points, thus it is more reasonable to assume the following model

$$g_i(t_{ik}) = f_i(t_{ik}) + \epsilon(t_{ik}), i = 1, \dots, n, k = 1, \dots, n_i, \quad (2.3)$$

The model above represents the most general case, where the observed data for different subjects are measured in different time points, but still in the same time interval $\mathcal{T}$. Our ultimate goal is the estimation of the function f and attempting to mimic the process (or its trend), which creates our observed response $g_1, \dots, g_n$. With the estimated function $\hat{f}$, we are able to reconstruct the whole set of functional data at desired time resolution. According to Ramsey [2005], there are two methods, how such functions can be estimated

- Using linear smoothing methods. These methods use so called *smoothing matrix* $\mathbb{C}$ to obtain the estimator of unknown function f of form $\hat{f} = \mathbb{C}g$. The various options, how to choose the smoothing matrix, can be found in Zhang [2012] or Ramsey [2005],
- Using non-parametric smoothing – e.g Kernel smoothing.

2.1 Linear smoothing – orthonormal basis expansion

The first method is based on parametric regression. In the previous section, we said, that the linear smoothing methods are based on the smoothing matrix $\mathbb{C}$. There exists infinitely many options how to choose such matrix. In this section, we focus on one class of smoothing matrices - matrices including orthonormal bases of $\mathcal{L}^2(\mathcal{T})$ space.

First, without loss of generality, we assume that each functional data is evaluated at identical time points. Let assume that the theoretical model for the functional data has the form

$$g_i = f_i(t) + \epsilon_i(t), \quad i = 1, \dots, n \quad (2.4)$$

where, for the fixed $t \in \mathcal{T}$, $\epsilon(t) = (\epsilon_1(t), \dots, \epsilon_n(t))^\top \sim \mathcal{N}_n(\mathbf{0}_n, \sigma^2 \mathbb{I}_n)$, where $\text{var}\epsilon_i = \sigma^2$ and $\mathbb{I}$ is $n \times n$ identity matrix and $f_i(t)$ is unknown function we would like to estimate. Unless stated otherwise, we assume that the errors arising from a general measurement process are homoscedastic. Linear smoothing procedure involves representing the function f_i as a linear combination of M known pre-specified basis functions $\phi_m, m = 1, \dots, M$ such that

$$\begin{aligned} f_i(t) &= \sum_{m=1}^M c_{im} \phi_m(t) \\ &= \mathbf{c}_i^\top \Phi(t), \quad i = 1, \dots, n, \end{aligned} \quad (2.5)$$

where $\mathbf{c}_i = (c_{i1}, \dots, c_{iM})^\top, i = 1, \dots, n$ is vector of unknown coefficients for each individual function, $\Phi(t) = (\phi_1(t), \dots, \phi_M(t))^\top$ is the vector of known orthonormal basis functions and M is its dimension. What follows is the idea behind the model (2.5).

We know that any space of some dimension M , can be generated by a linear span of some space-related basis

$$\text{span}(\phi_1, \dots, \phi_M) = \left\{ \sum_{m=1}^M c_m \phi_m, \mathbf{c} = (c_1, \dots, c_M) \in \mathbb{R}^M \right\} \quad (2.6)$$

In model (2.5) we assume that each functional observation comes from some functional space, where every function can be represented as a linear combination of some space-related basis, such as (2.6). Consequently, we would like to, with help of valid option for orthonormal basis functions, sufficiently high dimension M , to get closer to the true function f as possible. Notably, our goal is to reconstruct unknown function for each individual observation, with the same basis for each function.

The equation (2.4) allows the evaluation of the unknown function, or basis representation, for every $t \in \mathcal{T}$. In practice, this is not possible, and we must focus on the function reconstruction only at discrete time points. Furthermore, we assume the most general case, where the each functional data is evaluated at different time points, but still from the same time interval $\mathcal{T}$. Set $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})^\top$ as a vector of discrete points for each

observation i and set $\mathbb{Z}_i(\mathbf{t}_i) = (\Phi(\mathbf{t}_{i1}), \dots, \Phi(\mathbf{t}_{in_i}))^\top$, which is $n_i \times M$ matrix of basis representation for each discrete time point for the i -th observation.

$$\mathbb{Z}_i(\mathbf{t}_i) = \begin{pmatrix} \phi_1(t_{i,1}) & \dots & \phi_M(t_{i,1}) \\ \vdots & \ddots & \vdots \\ \phi_1(t_{i,n_i}) & \dots & \phi_M(t_{i,n_i}) \end{pmatrix} \quad (2.7)$$

With the help of the aforementioned notation, we may rewrite the equation (2.4) as

$$\mathbf{g}_i(\mathbf{t}_i) = \mathbb{Z}_i(\mathbf{t}_i)\mathbf{c}_i + \boldsymbol{\epsilon}, \quad i = 1, \dots, n, \quad (2.8)$$

where $\mathbf{g}_i = (g_{i1}, \dots, g_{in_i})^\top, i = 1, \dots, n$. The goal of this procedure is finding "the best" estimates of the coefficients $\mathbf{c}_i = (c_{i1}, \dots, c_{iM})^\top$ for each function. The equation (2.8) is similar to a linear regression model. The estimators of unknown coefficients $\mathbf{c}_i$ can be obtained by solving the constrained optimization problem

$$\min_{\mathbf{c}_i \in \mathbb{R}^M} (\mathbf{g}_i - \mathbb{Z}_i(\mathbf{t}_i)\mathbf{c}_i)^\top (\mathbf{g}_i - \mathbb{Z}_i(\mathbf{t}_i)\mathbf{c}_i) \quad (2.9)$$

$$\text{s.t. } \mathbf{A}\mathbf{c}_i = \mathbf{a}, \quad i = 1, \dots, n. \quad (2.10)$$

where the constraint is set for the estimability. The optimization problem above is nothing else than the well known least square optimization problem from linear regression theory. Therefore, the solution of this problem has a analytical solution (e.g see Komárek [2020]) of the form

$$\hat{\mathbf{c}}_i = \left(\mathbb{Z}_i^\top \mathbb{Z}_i \right)^{-1} \mathbb{Z}_i^\top \mathbf{g}_i. \quad (2.11)$$

solution (2.11) depends on discretized time points (where we do not assume that the unknown coefficients are time dependent). Therefore, the estimate of the trend of the i -th individual function is defined as

$$\hat{\mathbf{f}}_i(\mathbf{t}_i) = \Phi(\mathbf{t}_i) \left(\mathbb{Z}_i^\top \mathbb{Z}_i \right)^{-1} \mathbb{Z}_i^\top \mathbf{g}_i, \quad (2.12)$$

where $\hat{\mathbf{f}}_i(\mathbf{t}_i) = (f_i(t_{i1}), \dots, f_i(t_{in_i}))^\top$ is the estimated trend of the i -th observation. Furthermore, the form of the equation (2.12) allow us to evaluate the function at any discrete time point. Therefore, we are able to create a more smooth trend.

We can see that the estimated functions depends on the choice of basis, for which there are numerous options. Zhang [2012] and Ramsey [2005] introduced various choices for the basis function such as *Fourier series*, *B-spline smoothing*, *Wavelets transformations* or *P-splines*.

It is important to highlight the fact that we are trying to estimate the function $\hat{f}$ for every subject/individuals, at discretized time points. Choice of the basis function strongly depends on the type of the problem we face. For example, if the trend of the observed data is periodic, Fourier basis is relevant. On the other hand, if we have data without any cyclic or period trend, Fourier basis is useless and B-spline smoothing is more appropriate.

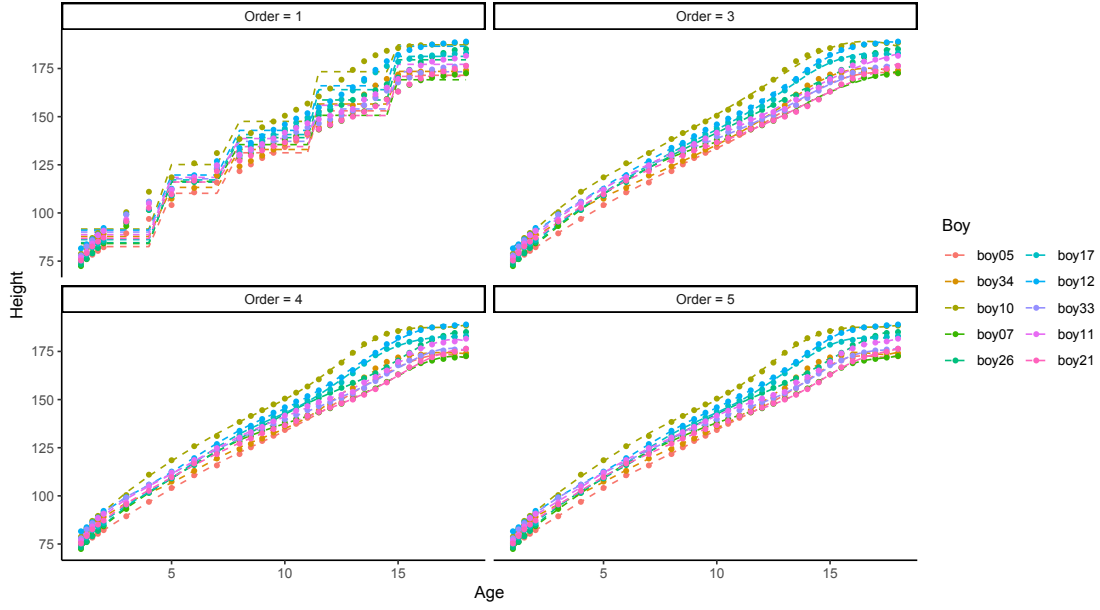


Figure 2.1: Observed heights of 10 boys from Example 1. Reconstruction of the data was done by B -splines.

2.1.1 Example of orthonormal basis expansion

In this section we introduce a practical application of the approximation using orthogonal basis expansions on real data. Let us revisit Example 1. We observe sample pairs $(t_1, H_{i,1})^\top, \dots, (t_{10}, H_{i,10})^\top, i = 1, \dots, 10$. Furthermore, the time grid is identical for all boys, or more precisely, height was measured at the same time for every boy. Our goal is to estimate the unknown function, or the trend of the evolution of boys' heights with the help of some basis function. Based on Figure 1.1, Fourier basis is irrelevant, as the data has no periodic trend. In general, the evolution of the height, mainly during early ages has an increasing trend (except for some outlying cases - illness, etc.). Therefore, a reasonable choice for the estimation of the unknown function for the height points are regression splines. Figure 2.1 shows the reconstruction of height for ten boys. We can see, that the quality of the reconstruction depends on the regression spline parameters - number of basis M (or dimension of basis vector) and the order of the polynomial. Increasing the polynomial order, improves the quality of the estimation.

Figure 2.2, describes the reconstruction of the same dataset with the help of the Fourier basis. There are two notable facts. First, the quality, compared to reconstruction using regression splines, is lower. The second being that the reconstruction has no practical sense. The trend of the heights is decreasing before birth age (we know that the height of unborn children does not have a decreasing trend), then there is a minimal cyclic trend, and finally after the follow-up the height begins to decrease (which is not true in general for young people). This example highlights the necessity to fully understand the nature of the analyzed data in order to correctly choose a reasonable orthonormal basis.

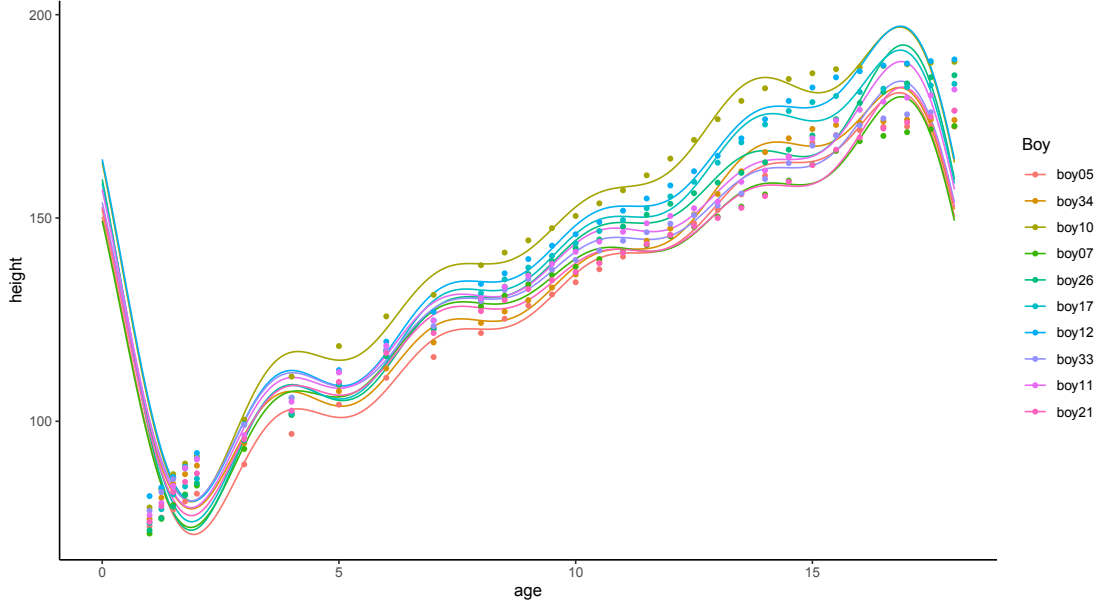


Figure 2.2: The reconstructed observed heights of 10 boys by Fourier basis from Example 1

2.2 Kernel smoothing

Similarly to the previous section, let us assume that we observe the pairs $(g_1, t_1)^\top, \dots, (g_n, t_n)^\top$, where we believe that g_i has the following functional form for a time point $t \in \mathcal{T}$

$$g_i(t) = f_i(t) + \epsilon(t), \quad i = 1, \dots, n \quad (2.13)$$

In this section, we describe an alternative way of smoothing functional data. We assume that the function f_i is p -times differentiable at the entire time interval $\mathcal{T}$. Then, we may approximate the unknown function $f(t)$ in any time point $t_0 \in \mathcal{T}$ using the Taylor polynomial

$$f_i(t_0) \approx f_i(t) + \sum_{k=1}^p (t - t_0)^k \frac{f_i^{(k)}(t_0)}{k!}, \quad t \in \mathcal{T}. \quad (2.14)$$

Denote $\beta_k(t_0) = \frac{f^{(k)}(t_0)}{k!}$. Then the *local polynomial estimator* of an unknown function f at time point $t_0 \in \mathcal{T}$ is defined as

$$\hat{\beta}(t_0) = (\hat{\beta}_0(t_0), \dots, \hat{\beta}_p(t_0))^\top \quad (2.15)$$

$$= \underset{b_0, \dots, b_p}{\operatorname{argmin}} \sum_{i=1}^n [g_i - b_0 - \sum_{k=1}^p b_k (t_i - t_0)^k]^2 K_i \left(\frac{t - t_0}{h} \right), \quad (2.16)$$

where $K_i(\cdot)$ is some Kernel function (for the various options, see Omelka [2020]). A Kernel function assigns the weight to each observed time point based on the distance between the current time point and the evaluation time point t_0 (higher weights for the time points close to t_0 , lower otherwise). To improve clarity, we denote the Kernel functions as $K_i(t, t_0, h) \equiv K_i \left(\frac{t - t_0}{h} \right)$

The kernel functions always depend on *bandwidth* h , which, in some sense, defines the range of values around point t_0 we take into account. Hand-choosing the bandwidth

by hand is difficult. Numerous techniques on how to choose this parameter are available. Most common techniques are based on *Cross-Validation* methods (e.g. *General Cross Validation criterion*), described in more detail in Omelka [2020]. For other techniques on bandwidth selection refer to the documentation of `fdasc` package Febrero-Bande and de la Fuente [2012].

To further clarify the kernel smoothing methods, let us rewrite equation (2.15) into matrix notation. This will allow us to find an explicit formula for $\widehat{\beta}_p(t_0)$. For the designed time points $t_1, \dots, t_n$ put

$$\mathbf{g} = \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} \quad \mathbb{A}(t_0) = \begin{pmatrix} 1 & (t_1 - t_0) & \dots & (t_1 - t_0)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & (t_n - t_0) & \dots & (t_n - t_0)^p \end{pmatrix} \quad (2.17)$$

and let $\mathbb{W}(t_0) = \text{diag}(K(\frac{t_1 - t_0}{h}), \dots, K(\frac{t_n - t_0}{h}))$. Then the equation can be rewritten as an optimization problem

$$\min_{\mathbf{b} \in \mathbb{R}^M} (\mathbf{g} - \mathbb{A}(t_0)\mathbf{b})^\top \mathbb{W}(t_0)(\mathbf{g} - \mathbb{A}(t_0)\mathbf{b}) \quad i = 1, \dots, n. \quad (2.18)$$

The optimization problem above is nothing else than the weighted linear square problem. As such its solution has an explicit form (see Omelka [2020] or Zhang [2012])

$$\hat{\beta}(t_0) = (\mathbb{A}(t_0)^\top \mathbb{W}(t_0) \mathbb{A}(t_0))^{-1} \mathbb{A}(t_0)^\top \mathbb{W}(t_0) \mathbf{g} = \mathbb{V}^{-1} \mathbb{S} \quad (2.19)$$

where

$$\mathbb{V} = \begin{pmatrix} \sum_{i=1}^n K_i(t, t_0, h) & \dots & \sum_{i=1}^n K_i(t_i - t_0)^p \\ \sum_{i=1}^n K_i(t, t_0, h)(t_i - t_0) & \dots & \sum_{i=1}^n K_i(t, t_0, h)(t_i - t_0)^{p+1} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n K_i(t, t_0, h)(t_i - t_0)^p & \dots & \sum_{i=1}^n K_i(t, t_0, h)(t_i - t_0)^{2p} \end{pmatrix} \quad (2.20)$$

$$\mathbb{S} = \begin{pmatrix} \sum_{i=1}^n K_i(t, t_0, h) g_i \\ \sum_{i=1}^n K_i(t, t_0, h)(t_i - t_0) \\ \vdots \\ \sum_{i=1}^n K_i(t, t_0, h)(t_i - t_0)^p \end{pmatrix} \quad (2.21)$$

At this point, it is necessary to know how many times does the function f need to be differentiated to incorporate the information about local behaviour. In general, as stated in Omelka [2020], our interests lie in approximation of f at $t_0 \in \mathcal{T}$, estimated by $\hat{\beta}_0(t_0) = f(t_0)$. If we set $p = 0$, we are essentially taking the observed data weighted by the chosen kernel. Such estimator is called *Nadaya-Watson estimator*, defined as

$$\hat{f}_i^{NW}(t_0) = \hat{\beta}_0(t_0) = \sum_{i=1}^n \frac{K_i(t_i, t_0, h) g_i}{\sum_{i=1}^n K_i(t_i, t_0, h)}, \quad i = 1, \dots, n \quad (2.22)$$

If our goal is to enrich the estimate of function f by local behaviour, then we set $p = 1$ and obtain a *local polynomial estimator*, defined as

$$\hat{f}^{LPK} = \frac{\sum_{i=1}^n \tau_0 - \tau_1(t_i - t_0)] K_i(t_i, t_0, h) g_i}{\tau_2 \tau_0 - \tau_1^2} \quad (2.23)$$

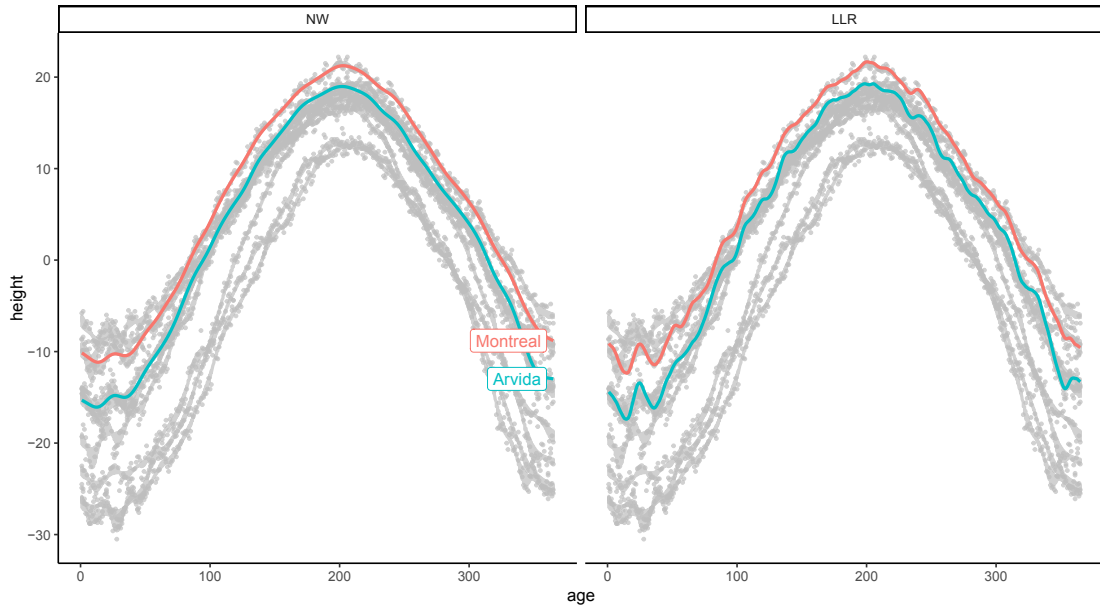


Figure 2.3: Reconstruction of the trend of the temperature for two canadian stations.

where $\tau_r = \sum_{i=1}^n K_i(t, t_0, h)(t - t_0)^r$.

One of the crucial disadvantage of both estimators is, that they are not robust. This can be solved, for example, choosing the LOWESS smoothing method. In the next section we show that it is quite easy to violate the trend with only one outlying observation. In real practical tasks, we encounter linear smoothing rather than Kernel smoothing. We will address this in Chapter 3.

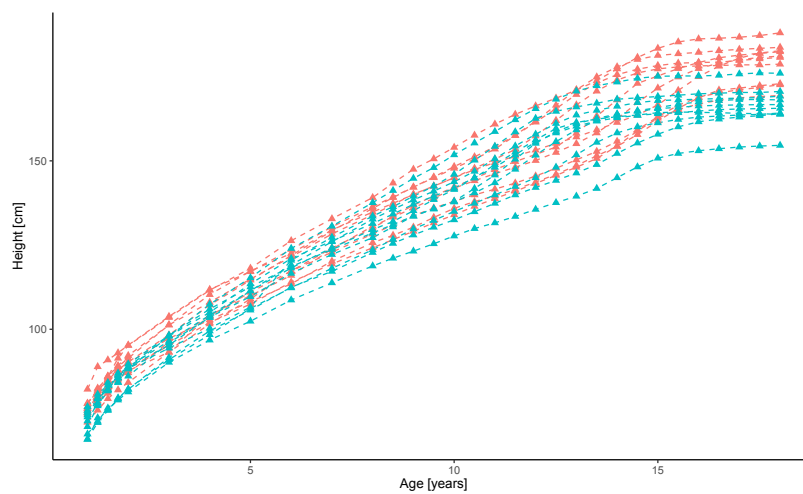
2.2.1 Example of kernel smoothing

In this section, we show a practical application of Kernel smoothing. The example will use the daily temperature data at 35 different locations in Canada¹. For simplicity, we take ten random locations. We observe the following pairs $(t_{k1}, C_{k1})^\top, \dots, (t_{kn_k}, C_{kn_k})^\top$, $k = 1, \dots, 10$, $i = 1, \dots, 365$, where t_{kj} is the day, where the temperature C_{kj} was measured. Same as in the boys example, we observe the temperatures at identical time points (days). Furthermore, the same model as (2.13) is assumed. In Figure 2.3, the trend reconstruction for two Canadian stations is shown. For trend reconstruction, the Nadarya-Watson estimator (left plot) and Local Linear estimator (right plot) were used. The optimal bandwidth h was chosen based on the *GCV* method. Noticeably, the LLE provides a better fit than the NW estimator, however the difference is sizeable and both smoothing options are relevant for this case.

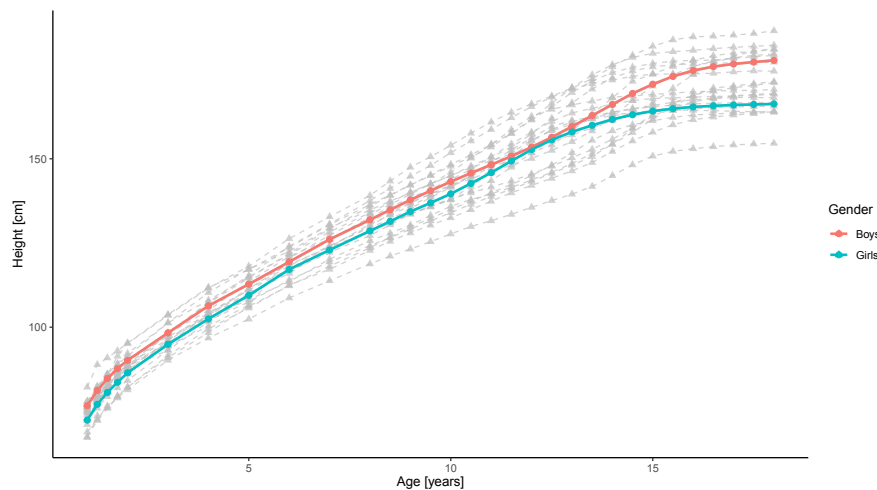
¹The name of the dataset is `CanadianWeather` and can be found in `fda` package

3. Functional Analysis of Variance (FANOVA)

In the introduction part, we were interested in the evolution of the height of the boys. In this chapter, we extend this example by evolution of the height of ten girls. The height curves for ten boys and for ten girls are shown in Figure 3.1. We should examine whether the girls and boys grow from childhood until their adolescent years with the same rate. More precisely, we would like to test if the heights are different between the boys and girls for the whole follow-up time. The inference will be mainly based on the mean functions of boys and girls which are shown in Figure 3.1 (b). We can see that the mean functions are different for almost the whole follow-up except the age of fourteen.



(a) Evolution of heights of ten boys and ten girls



(b) Mean functions of heights of girls and boys.

Figure 3.1: The evolution of the heights of ten boys and ten girls

Therefore, we are interested in testing the hypothesis

$$H_0 : \mu_{boys}(t) = \mu_{girls}(t) \forall t \in \mathcal{T} \text{ vs. } H_1 : \exists t \in \mathcal{T} \mu_{boys}(t) \neq \mu_{girls}(t). \quad (3.1)$$

where $\mu_{boys}(t)$ is the mean function of heights for the boys and $\mu_{girls}(t)$ is the mean function of heights for ten girls.

This type of testing is very similar to the two sample problem. In univariate statistical analysis, we should use F -test for comparison of two groups or two sample t -test, where the rejection of null hypothesis can lead to significant difference between two groups, and analysis is basically done at this point. In case of FDA, using t -test or F -test will be a little bit problematic, because we need to calculate the F -statistic or t -statistics for every time point t , which should be computationally expensive. Moreover, we can not ensure that the significance level α will be equal to the pre-set value. Basically, using this point-wise tests should lead to multiple testing problems. The another issue is that this type of test can not tell us in which point the groups are different, which is typically the object of the interest. In FDA we can use methods like change-point analysis for finding the precise time point, where the random curves are different.

The problem above can be generalized into K -sample problem or one-way ANOVA for functional data (Gorecki and Smaga [2015]). Let us describe the task. We have $K \geq 2$ independent functional groups

$$\begin{aligned} X_{11}(t), \dots, X_{1n_1}(t) &\stackrel{iid}{\sim} SP(\mu_1, \gamma), \\ X_{21}(t), \dots, X_{2n_2}(t) &\stackrel{iid}{\sim} SP(\mu_2, \gamma), \\ &\vdots \\ X_{K1}(t), \dots, X_{Kn_K}(t) &\stackrel{iid}{\sim} SP(\mu_K, \gamma), \end{aligned}$$

where $X_{ij}(t)$ is a random function defined on $t \in [a, b]$, $n_i, i = 1, \dots, K$ is the number of functional samples in i -th group, μ_i is a mean function of i -th group and γ is a common covariance function. Furthermore, we assume that the curve has the form

$$X_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t), \quad i = 1, \dots, K, \quad j = 1, \dots, n_i, \quad t \in \mathcal{T}, \quad (3.2)$$

where $\epsilon_{i1}(t), \dots, \epsilon_{in_i}(t) \stackrel{i.i.d.}{\sim} SP(0, \gamma(s, t))$.

Remark. In this thesis we will mainly work with a homoscedastic case, or more precisely, we will assume that all functional samples have the common covariance function. Although, there is also the FANOVA test, which can test the equality of mean functions also under a heteroscedastic case. For example, it is well-explained in Zhang [2012], Chapter 9.

We are interested in one-way ANOVA testing problem

$$\begin{aligned} H_0 : \mu_1(t) &= \mu_2(t) = \dots = \mu_K(t) \quad \forall t \in [a, b] \\ \text{vs.} \\ H_1 : \exists t \in [a, b] \quad \exists j \neq k \quad \mu_j(t) &\neq \mu_k(t), \end{aligned} \quad (3.3)$$

or in other words, we are interested in testing if functional samples are the same on the whole time grid $[a, b]$, against the alternative that there exists at least one time

point $t \in [a, b]$ (if we realise that the time points are discretized, the statement is correct), in which at least two functional samples are different. Unfortunately, this type of a test will not reveal in which time point (or time points) the functional samples are different and a further analysis is needed (multicomparison problem or change point analysis). In the following sections we will introduce tests which can be used to test (3.3). In general, we will not assume specific distribution for SP (for example the Gaussian process) as it is in usual ANOVA. Before we will introduce the FANOVA tests, we need to first define the statistical model and assumptions we will require for data to satisfy.

Model (Fanova). *Let the random variables $X_{1j}(t), \dots, X_{n_j j}(t), j = 1, \dots, K$ are i.i.d. random variables with values in $L^2(\mathcal{T})$, which are distributed as generating random variable $X_j(t) \sim SP(\mu_j, \gamma(s, t))$, defined on the interval $\mathcal{T} = [a, b]$, K is number of the groups, μ_j is a mean function related to the j -th group, γ is a covariance function.*

3.1 FANOVA notation

In this section we will introduce some of the well-known statistical tests for functional data, its properties and how the computation procedures are changed if we use reconstruction data techniques described in Chapter 2. Firstly, let us introduce the notation which will be useful for the whole section. Let

$$SSR(t) = \sum_{j=1}^K n_j [\bar{X}_j(t) - \bar{X}(t)]^2 \quad (3.4)$$

and

$$SSE(t) = \sum_{j=1}^K \sum_{i=1}^{n_j} [\bar{X}_{ij}(t) - \bar{X}_j(t)]^2 \quad (3.5)$$

be point-wise between subject and within subject variations, where $\bar{X}_j(t) = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}(t)$ is a subject specific empirical mean, which consistently estimates the subject specific mean function (or the mean trend) and $\bar{X}(t) = \frac{1}{n} \sum_{j=1}^K \sum_{i=1}^{n_j} X_{ij}(t)$ is the point-wise grand mean. Furthermore, with the help of FCLT, we have $\frac{1}{\sqrt{n}} \bar{X}_j(t) \xrightarrow{\mathcal{D}} GP(\mu_j, \gamma)$ as $n \rightarrow \infty$. Next, let us introduce the global versions of within and between subject sum of squares as

$$\begin{aligned} SSR &= \sum_{j=1}^K n_j \|\bar{X}_j - \bar{X}\|^2 \\ &= \sum_{j=1}^K n_j \int_{\mathcal{T}} [\bar{X}_j(t) - \bar{X}(t)]^2 dt, \end{aligned} \quad (3.6)$$

$$\begin{aligned} SSE &= \sum_{j=1}^K \sum_{i=1}^{n_j} \int_{\mathcal{T}} [\bar{X}_{ij}(t) - \bar{X}_j(t)]^2 dt \\ &= \sum_{j=1}^K \sum_{i=1}^{n_j} \|\bar{X}_{ij}(t) - \bar{X}_j(t)\|^2, \end{aligned} \quad (3.7)$$

where $\|\cdot\|^2$ represents the $\mathcal{L}^2$ -norm. The statistics above are the global version of the sum of squares. In some sense they measure the variations between groups on the whole follow-up time. For the purpose of this section, it is necessary to derive its exact (or asymptotic) distributions under validity of H_0 . First, we need to introduce necessary assumptions for deriving the distribution of the global sum of squares.

Assumption 3. (A3)[Zhang [2011]] $X_{1j}(t), \dots, X_{n_jj}(t), j = 1, \dots, K$ are elements of $\mathcal{L}^2(\mathcal{T})$. Furthermore, all functional are coming from $GP(\mu(t), \gamma(s, t))$, where $\text{tr}(\gamma) = \int_{\mathcal{T}} \gamma(t, t) dt < \infty$.

In the assumption above, we require the functional data to have a finite variance at each time point. However, in a real world application it is very hard to detect if functional data are generated by a Gaussian process. Fortunately, it is also possible to derive the asymptotic distribution of sum of squares with a help of the following assumption.

Assumption 4. (A4) Let $X_{j,1}, \dots, X_{j,n_j}, j = 1, \dots, K$ be stochastic processes with $\mu_j(t) \in L^2(\mathcal{T})$ and its covariance function satisfy $\text{tr}(\gamma) < \infty$. Furthermore, let the proportion $\frac{n_j}{n} \rightarrow b_j$ as $n \rightarrow \infty$, where $b_1, \dots, b_K \in (0, 1)$ and $\sum_{j=1}^K b_j = 1$.

Theorem 3.1.1 (exact distribution of SSR and SSE , Zhang [2011]). Under validity of A3, both global version of sum of squares SSR and SSE has under H_0 same distribution as

$$\sum_{l=1}^m \lambda_l A_l + \sum_{l=m+1}^{\infty} \pi_l^2, \sum_{l=1}^m \lambda_l B_l \quad (3.8)$$

where A_l, B_l are independent of each other and $A_l \sim \chi_{K-1}^2(\lambda_l^{-1} \pi_l^2), B_l \sim \chi_{n-K}^2$ and $\pi_l^2 = \|\int_{\mathcal{T}} \mu(t) \phi_l(t) dt\|^2, l = 1, \dots, m$.

Proof. Let us start with SSE

$$\begin{aligned} SSE &= \int_{\mathcal{T}} \sum_{j=1}^K \sum_{i=1}^{n_j} [X_{ij}(t) - \bar{X}_j(t)]^2 dt \\ &= \int_{\mathcal{T}} \sum_{j=1}^K (n_j - 1) \hat{\gamma}(t, t) dt \\ &= \int_{\mathcal{T}} (n - K) \hat{\gamma}(t, t) dt \\ &= \text{tr}((n - K) \hat{\gamma}) \end{aligned}$$

With the help of A3, and Theorem A.2.2 we have $(n - K) \hat{\gamma}(s, t) \sim WP(n - K, \gamma)$. Therefore $\hat{\gamma}(s, t)$ satisfy assumptions of Theorem A.2.3 and using the second point of theorem we have that SSE has the same distribution as $\sum_{l=1}^m \lambda_l B_l$, where $B_l \stackrel{iid}{\sim} \chi_{n-K}^2$. Right now it is necessary to prove the theorem for the SSR . Let us rewrite SSR in a well-known matrix notation

$$SSR(t) = \mathbf{V}(t) \left(\mathbb{I}_K - \mathbf{n}_n \mathbf{n}_n^{\top} \frac{1}{n} \right) \mathbf{V}(t),$$

where $\mathbf{V}(t) = (\sqrt{n_1}\bar{X}_1(t), \dots, \sqrt{n_K}\bar{X}_K(t))^\top$, $\mathbf{n}_n = (\sqrt{n_1}, \dots, \sqrt{n_K})^\top$. Under the validity of H_0 and $A3$ we have $\mathbf{V}(t) \sim GP(\mu(t), \gamma(s, t))$. Furthermore, we have to show that the matrix $\mathbb{I}_K - \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n}$ is an idempotent matrix of rank $K - 1$. If we rewrite the matrix multiplication as

$$\left(\mathbb{I}_K - \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n}\right) \left(\mathbb{I}_K - \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n}\right) = \mathbb{I}_K - \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n} - \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n} + \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n} \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n}. \quad (3.9)$$

The last term can be written as

$$\mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n} \mathbf{n}_n \mathbf{n}_n^\top \frac{1}{n} = \frac{1}{n^2} \begin{pmatrix} n_1 n & \dots & \sqrt{n_1 \dots n_K} n \\ \vdots & \dots & \vdots \\ n_K n & \dots & n_K n \end{pmatrix} = \frac{1}{n} \mathbf{n}_n \mathbf{n}_n^\top. \quad (3.10)$$

If we plugin (3.10) into (3.9), the idempotent property is proven. Then with the help of Theorem A.2.6 we obtain the desired result.

QED

As was mentioned, in a real case scenario it is hard to decide if the functional data are generated from the Gaussian process or not. In univariate statistical analysis it is easy to verify, if data are from a Gaussian distribution with the help of QQ -plots, histograms or with the help of some goodness of fit tests. In FDA it is much more complicated to detect if the data are from Gaussian process. From the definition of a Gaussian process, we know that if we take a random curve evaluated at a sufficient number of time points, which lead to a multivariate random variable, and this variable has a multivariate normal distribution, the random curve is coming from a Gaussian process. But it is very hard to detect how much points we need and which are the most representative ones. There also exists some tests to test if the functional data comes from the Gaussian process or not, but this topic is out of scope of this thesis. However, we need to have this fact in our minds before FDA. The following theorem describes the asymptotic distribution of the functional sum of squares statistics.

Theorem 3.1.2 (Zhang [2012], Theorem 5.9). *Let's assume that $A4$ holds. Then under the null hypothesis, as $n \rightarrow \infty$, we have*

$$\int_{\mathcal{T}} SSR(t) \xrightarrow{\mathcal{D}} \sum_{l=1}^m \lambda_l A_l, A_l \stackrel{i.i.d.}{\sim} \chi_{K-1}^2. \quad (3.11)$$

Proof. See Zhang [2012], Chapter 5.

QED

The results above are necessary for deriving the asymptotic distribution of the test statistic which will be introduced in the following sections.

3.2 Tests for equality of mean functions

It is a good idea to start with the simple and the most similar test to ANOVA test for the univariate data, the functional F -test. The test statistic is defined as

$$F(t) = \frac{\frac{SSR(t)}{k-1}}{\frac{SSE(t)}{n-k}}. \quad (3.12)$$

The F -statistic above is also called *F-ratio statistic*. Its interpretation is the same as in a non-functional case. F -ratio in some sense represents the amount of variation within each subject (or functional sample) against the total variation in the data, at every time point $t \in \mathcal{T}$. Unfortunately, a test based on this statistics is facing some serious problems

- If the time grid is very dense (e.g include 10^4 discrete time points), the evaluation of the test is very time consuming,
- Because we are evaluating the test for the each time point, we are facing the multi-comparison problem, and the true level of the test is much lower. It may be even close to 0. Hence it implies that test will be strongly conservative and from some point the test will not be able to detect the changes between the functional samples,
- As was explained by Cuevas et al. [2004], it is very complicated to derive the asymptotic distribution for (3.12). Therefore, the evaluation of the test has to be done at a fixed time point. For this fixed point, the statistic, under the normally distributed i.i.d. data, has a F -distribution with d_1 and d_2 degrees of freedom. We reject the null hypothesis, whenever $F(t) \geq F_{d_1, d_2}(1 - \alpha), \forall t \in \mathcal{T}$. It is obvious that a test based on such an evaluation method can never achieve the significance level and will be strongly conservative.
- This test does not work with data reconstruction introduced in Chapter 2.

All tests that will test the hypothesis (3.3) in some sense will be working with the statistics given in (3.12). However, some changes have to be made to resolve the issues outlined above.

3.2.1 Parametric bootstrap $\mathcal{L}^2$ - norm based test

The first class of tests, which are based on statistic (3.6) is called *$\mathcal{L}^2$ -norm based*. As was described by Cuevas et al. [2004], the main issue of previously mentioned F -ratio test statistic is that we are not able to derive the distribution under validity of H_0 , even if the functional sample comes from the Gaussian process. Authors of the mentioned paper propose a solution, how to get rid of this problem and still keep the idea of F -test. Their goal was to keeping the work with an idea: " H_0 should be rejected whenever the "external" variability between groups, as measured by the differences between their "sample means" is large enough at a prescribed level". In other words, their goal was to find a statistic, which still capture the between group variation, and at the same time, there will be a possibility to find its asymptotic distribution. The whole test is based on the statistic

$$L_n = \sum_{1 \leq i < j \leq K} n_i d(\bar{X}_j, \bar{X}_i) = \sum_{1 \leq i \leq j \leq K} \int_{\mathcal{T}} [\bar{X}_i(t) - \bar{X}_j(t)]^2 dt. \quad (3.13)$$

We can see that the statistic measures, in some sense, how far the functional groups are from the each other for the **whole follow up time**. It is obvious, that if at least one pair of the group are "far away" (or they are different from the each other for the whole follow-up time) from the each other, we will reject the null hypothesis. Furthermore, it can be shown that the statistics L_n and SSE are equivalent. The following theorem can help us to derive the asymptotic distribution of the statistic above.

Theorem 3.2.1. (Cuevas et al. [2004],) Let the assumption A4 hold. Then asymptotic distribution of statistic L_n under H_0 coincide with that of statistic

$$L = \sum_{1 \leq i < j \leq K} \int_{\mathcal{T}} [Z_i(t) - C_{ij} Z_j(t)]^2 dt \quad (3.14)$$

where $C_{ij} = \sqrt{\frac{p_i}{p_j}}$ and $Z_1(t), \dots, Z_K(t)$ are zero-mean Gaussian processes with covariance function $\gamma(s, t)$.

Proof. Let the assumption $\frac{n_i}{n} \rightarrow p_i > 0$ hold and let the H_0 be valid. Furthermore, define a continuous function $g : \mathcal{L}_1^2(\mathcal{T}) \times \mathcal{L}_2^2(\mathcal{T}) \times \dots \times \mathcal{L}_K^2(\mathcal{T}) \mapsto \mathbb{R}$, which has the form

$$g(w_1, \dots, w_K) = \sum_{1 \leq i < j \leq K} \int_{\mathcal{T}} [w_i(t) - w_j(t)]^2 dt. \quad (3.15)$$

Based on the assumption of the functional random sample and under the validity of H_0 we can use FCLT, as $n \rightarrow \infty$ we get

$$\left(\sqrt{n_1}[\bar{X}_1(t) - \mu(t)], \dots, \sqrt{n_K}[\bar{X}_K(t) - \mu(t)] \right) \xrightarrow{\mathcal{D}} (Z_1(t), \dots, Z_K(t)).$$

Now plug-in sequence $\left(\sqrt{n_1}[\bar{X}_1(t) - \mu(t)], \dots, \sqrt{n_K}[\bar{X}_K(t) - \mu(t)] \right)$ into (3.15) we have

$$\begin{aligned} & g(\sqrt{n_1}[\bar{X}_1(t) - \mu(t)], \dots, \sqrt{n_K}[\bar{X}_K(t) - \mu(t)]) \\ &= \sum_{1 \leq i < j \leq K} \int_{\mathcal{T}} \left[\sqrt{n_i}[\bar{X}_i(t) - \mu(t)] - \sqrt{n_j}[\bar{X}_j(t) - \mu(t)] \right]^2 dt \\ &= \sum_{1 \leq i < j \leq K} n_i \int_{\mathcal{T}} \left[(\bar{X}_i(t) - \mu(t)) - \sqrt{\frac{n_j/n}{n_i/n}} (\bar{X}_j(t) - \mu(t)) \right]^2 dt \end{aligned}$$

Therefore, from the continuous mapping theorem and applying the Slutsky theorem on the term $\sqrt{\frac{n_j/n}{n_i/n}} \rightarrow \sqrt{\frac{p_j}{p_i}}$, we have

$$g(\sqrt{n_1}[\bar{X}_1(t) - \mu(t)], \dots, \sqrt{n_K}[\bar{X}_K(t) - \mu(t)]) \xrightarrow{\mathcal{D}} g(Z_1(t), \dots, Z_K(t)),$$

where $g(Z_1(t), \dots, Z_K(t)) \equiv \sum_{1 \leq i < j \leq K} n_i \int_{\mathcal{T}} \left[Z_i(t) - \sqrt{\frac{p_j}{p_i}} Z_j(t) \right]^2 dt$.

QED

Remark. For the balanced design, $C_{ij} = 1$, the statistic introduced in Theorem 3.2.1 is just the $\mathcal{L}_2$ -distance between two Gaussian processes.

Remark. Theorem 3.2.1 assumes the general heteroscedastic case, but Cuevas et al. [2004] also considers the homoscedastic case, where all groups have the same covariance function. The procedure of test evaluation will not be affected by this, the only change will be in the Algorithm 1

However, as it can be seen, the theorem above does not provide the well-known distribution of statistic L . Therefore, a Monte Carlo simulation is needed for the approximation of the distribution of L . The procedure of the approximation of the distribution of test statistic L can be found in Algorithm 1. For the rest of this thesis, we will denote the described test as CH.

Algorithm 1: Monte Carlo simulation for approximation of distribution for L
Cuevas et al. [2004]

Input: number of groups K , observed L_n , group sample sizes $n_j, j = 1, \dots, K$
Data: $t_n = \{t_1, \dots, t_n\}$, a matrix $|t_n| \times Kn_j$ containing functional data
 $L_1^*, \dots, L_n^*$

for $j \leftarrow 1$ **to** K **do**
 $\hat{C}_j \leftarrow \text{CovEstimate}(X_{1,j}, \dots, X_{n_j,j})$;
 Generate the random samples $Z_{1,j}, \dots, Z_{N,j} \sim \mathcal{N}_{|t_n|}(\mathbf{0}_{|t_n|}, \hat{C}_j)$, $j = 1, \dots, K$;
end
for $l \leftarrow 1$ **to** N **do**
 Calculate $L_l^* = \sum_{i < j} \sum_{t \in t_n} [Z_{li}(t) - Z_{lj}(t)]^2$;
end
Result: Reject H_0 whenever $L_n > L_{1-\alpha}^*$.

3.2.2 Bias reduction method $\mathcal{L}^2$ -norm based test

The previous test solves the problem of a point-wise test. More precisely, the test statistic does not depend on a time point t , while taking into consideration the variation between groups for the whole follow-up time. On the other hand, this test has some limitations. Bootstrap $\mathcal{L}^2$ -norm based test does not include any information about the shape of each functional sample, or more precisely, there is not a covariance function in its calculation. This is partially solved by so called *bias-reduction* and *naïve* methods (see Zhang [2012], Chapter 4). The first idea of the aforementioned methods was proposed by Faraway [1997], where the author called such methods as *Multivariate Analysis-Based Methods*. The final form of the both methods was given by Zhang [2005]. In this section, we will describe the both methods. Calculations are based on the following statistic

$$\begin{aligned} T_n &= \sum_{i=1}^K n_i \|\bar{X}_i - \bar{X}\|^2 \\ &= \sum_{i=1}^K \int_T [\bar{X}_i(t) - \bar{X}(t)]^2 dt. \end{aligned} \quad (3.16)$$

The statistic T_n is nothing else than SSR , but for the purposes of this section, we will use the notation used in the mentioned articles.

We can see that the statistic T_n above has slightly different form (it is basically between groups sum of squares) than the statistic L_n . Statistic L_n measures how each groups varies from the others on the whole follow-up $\mathcal{T}$, where the T_n measures the variation of the groups from the population mean, where this population mean is the same for all groups if the H_0 is valid. Furthermore, in the balanced case where $n_1 = n_2 = \dots = n_K$, both are equal up to multiplicative constant.

Theorem 3.2.2. *Let T_n be statistic defined as in (3.16), L_n as in (3.13) If $n_1 = n_2 = \dots = n_K$, then both statistics are same up to multiplicative constant.*

Proof. Let us assume that the number of functional samples are the same across

the group. Then we have

$$\begin{aligned}
T_n &= \sum_{j=1}^K n_j \int_K [\bar{X}_j(t) - \bar{X}(t)]^2 dt \\
&= \int_T \sum_{j=1}^K n_j [\bar{X}_j^2(t) - 2\bar{X}_j(t)\bar{X}(t) + \bar{X}^2(t)] dt \\
&= \int_T n_1 \sum_{j=1}^K \bar{X}_j^2(t) - 2\bar{X} \sum_{j=1}^K n_1 \bar{X}_j + K n_1 \bar{X}^2 \\
&= \int_T \left\{ n_1 \sum_{j=1}^K \bar{X}_j^2(t) - 2K n_1 \bar{X} + K n_1 \bar{X}^2 \right\} dt \\
&= \int_T \left\{ n_1 \sum_{j=1}^K \bar{X}_j^2(t) - \frac{n_1}{K} \sum_{j=1}^K \bar{X}_j^2(t) - \frac{1}{K} \sum_{j \neq i} \bar{X}_j(t) \bar{X}_i(t) \right\} dt.
\end{aligned}$$

Same for the statistic L_n

$$\begin{aligned}
L_n &= \sum_{1 \leq i \leq j \leq K} \int_{\mathcal{T}} [\bar{X}_i(t) - \bar{X}_j(t)]^2 dt \\
&= \int_{\mathcal{T}} \sum_{1 \leq i \leq j \leq K} n_i [\bar{X}_i(t) - \bar{X}_j(t)]^2 dt \\
&= \int_{\mathcal{T}} \left\{ 2n_1 \sum_{j=1}^K \bar{X}_j^2(t) - \frac{2}{K n_1} \sum_{j=1}^K \bar{X}_j^2(t) - \frac{1}{K n_1} \sum_{i \neq j} \bar{X}_j(t) \bar{X}_i(t) \right\} dt
\end{aligned}$$

QED

The importance of the theorem above will be discussed later.

At this moment, we will focus on the approximation methods of distribution of test statistic T_n under validity of H_0 . If we assume that each functional sample group is the Gaussian with a mean function $\mu_j, j = 1, \dots, K$ and the groups are homoscedastic (they have the same covariance function), with the help of Theorem 3.1.1, T_n can be defined as χ^2 -type mixture described in the result of the mentioned theorem and its distribution can be described by the distribution of the random variable

$$T = \sum_{l=1}^m \lambda_l A_l, \quad (3.17)$$

where $A_l \sim \chi_{K-1}^2$. There exist various methods for a null distribution approximation of test statistic T_n . One of them, and most common approximation method is **normal approximation** method. This method is mostly useful when the generative process of our data is much violated from the Gaussian process. This topic was discussed in Zhang [2005], where this method was criticized by the author. The reason is that the χ^2 -mixture type variables are still χ^2 variables and so skewed. Normal distribution is a symmetric distribution around its mean and using this method for approximation of a distribution is not the right idea (because of possible negative values). The biggest problem of this approximation should be mainly in $(1 - \alpha)$ -quantile, where χ^2 distributions has much heavier tails than the normal distribution. The normal approximation

method is based on the approximation test statistic T_n by Normal distribution, where we set the expectation and the variance as

$$\begin{aligned}\mu_T &= \mathbb{E}T, \\ \sigma_T^2 &= \text{var}T.\end{aligned}$$

Using the described methods, we are rejecting the H_0 whenever $T_n > \phi_{\text{normal}}^{-1}(1 - \alpha)$, where $\phi_{\text{normal}}^{-1}$ is α -quantile of $\mathcal{N}(\mu_T, \sigma_T^2)$. The next approximation method is called *l-th cummulant matched χ^2 -Approximation* and was described by Zhang [2012]. The name of the cummulant may be confusing at first, but this is only a book-related term, and *l*-cummulant is nothing else than the *l*-th central moment of the some random variable. Describing the procedure of this method is out of the scope of this thesis and can be found in the mentioned book. In short, the method is based on the idea that we can approximate the distribution of one variable, with the help of a similar distribution, with the assumption that the both first *l* central moments are the same. In our case, as we have mentioned in the description of a normal approximation, we want to approximate the χ^2 -mixture type random variable by "normal" χ^2 variable. In the approximation of statistic T_n , we will use only the first three moments for using the cummulant matched methods

$$\mathcal{M}_1(X) = \mathbb{E}X, \quad \mathcal{M}_2(X) = \text{Var}X, \quad \mathcal{M}_3(X) = \mathbb{E}[X - \mathbb{E}X]^3.$$

Zhang [2012] introduced two methods for the null approximations with cummulant based methods: ***two-cummulant Matched χ^2 - Approximation*** and ***three-cummulant Matched χ^2 - Approximation***. The two-cummulant matched method is based on the following idea. As we mentioned in the description of the normal approximation method, χ^2 -mixture type variables are still χ^2 random variables, and so skewed. That it is natural to approximate distribution of T_n by the distribution of the similar random variable $U \sim \beta\chi_d^2$, where β is a scale parameter of χ^2 distribution and d are degrees of freedom. The both parameters β and d can be derived with the assumption of equality of two central moments $\mathcal{M}_1(T) = \mathcal{M}_1(U) = \beta d$, $\mathcal{M}_2(T) = \mathcal{M}_2(U) = 2\beta^2 d$. With a little manipulation, we have

$$\beta = \frac{\mathcal{M}_2(T)}{2\mathcal{M}_1(T)}, \quad d = \frac{2\mathcal{M}_1^2(T)}{\mathcal{M}_2(T)}. \quad (3.18)$$

Because of $\text{tr}(\gamma) < \infty$, using Theorem A.2.7 we have

$$\beta = \frac{\text{tr}(\gamma^{\otimes 2})}{\text{tr}(\gamma)}, \quad d = \frac{\text{tr}^2(\gamma)}{\text{tr}(\gamma^{\otimes 2})} \quad (3.19)$$

where $\text{tr}(\gamma^{\otimes r}) = \int_{\mathcal{T} \times \mathcal{T}} \gamma^r(s, t) dt ds$. For the successful approximation of the null distribution of T_n , we need to estimate both unknown parameters. According to the method of estimation, we distinguish two $\mathcal{L}^2$ -norm based tests using T_n : $\mathcal{L}^2$ -norm based test with **bias reduced** method of the estimation and $\mathcal{L}^2$ -norm based test with **naive** method of the estimation of unknown parameters. The naive method estimates the parameters in more intuitive way, where both unknown parameters are estimated by replacing the true covariance function γ for its unbiased estimator $\hat{\gamma}$. The naive estimators are defined as

$$\hat{\beta}^N = \frac{\text{tr}(\hat{\gamma}^{\otimes 2})}{\text{tr}(\hat{\gamma})}, \quad \hat{\kappa}^N = \frac{\text{tr}^2(\hat{\gamma})}{\text{tr}(\hat{\gamma}^{\otimes 2})}. \quad (3.20)$$

The bias-reduction method estimator is defined as

$$\hat{\beta}^B = \frac{\widehat{\text{tr}(\hat{\gamma}^{\otimes 2})}}{\widehat{\text{tr}(\gamma)}}, \quad \hat{\kappa}^B = \frac{\widehat{\text{tr}^2(\hat{\gamma})}}{\widehat{\text{tr}(\gamma^{\otimes 2})}} \quad (3.21)$$

and

$$\widehat{\text{tr}^2(\hat{\gamma})} = \frac{(n-K)(n-K+1)}{(n-K-1)(n-K+2)} \left[\text{tr}^2(\hat{\gamma}) - \frac{2\text{tr}(\hat{\gamma}^{\otimes 2})}{n-K+1} \right], \quad (3.22)$$

$$\widehat{\text{tr}(\gamma^{\otimes 2})} = \frac{(n-K)^2}{(n-K-1)(n-K+2)} \left[\text{tr}(\hat{\gamma}^{\otimes 2}) - \frac{\text{tr}^2(\hat{\gamma})}{n-K} \right]. \quad (3.23)$$

Remark. Both bias-reduced and naive methods for estimation the scale parameter β and degrees of freedom κ can be used only if the data are generated by Gaussian process. Otherwise, the above results are not valid.

Theorem 3.2.3. Let A3 hold. Then estimators (3.22) and (3.23) are unbiased estimators of $\text{tr}(\gamma)$ and $\text{tr}(\gamma^{\otimes 2})$.

Proof. Let us rewrite both equations (3.22) and (3.23) in the following way

$$\widehat{\text{tr}^2(\hat{\gamma})} = \frac{(n-K+1)\text{tr}^2(W)(n-K) + 2\text{tr}(W^{\otimes 2})}{(n-K)(n-K-1)(n-K+2)} \quad (3.24)$$

$$\widehat{\text{tr}(\gamma^{\otimes 2})} = \frac{\text{tr}(W^{\otimes 2}) + \text{tr}^2(W)}{(n-K-1)(n-K+2)} \quad (3.25)$$

where $W^{\otimes r} = (n-K)^r \hat{\gamma}^r(s, t)$. Furthermore, we know that under validity of A3, $(n-K)\hat{\gamma}(s, t) \sim WP(n-K, \gamma)$. Then assumptions of Theorem B.1.4 are satisfied and we have

$$\begin{aligned} \widehat{\text{Etr}^2(\hat{\gamma})} &= \frac{(n-K+1)\text{Etr}^2(W)(n-K) + 2\text{Etr}(W^{\otimes 2})}{(n-K)(n-K-1)(n-K+2)} \\ &\stackrel{\text{ThB.1.4c}}{=} \frac{(n-K)(n-K+1) - 2}{(n-K-1)(n-K+2)} \text{tr}^2(\gamma) = \text{tr}^2(\gamma) \end{aligned} \quad (3.26)$$

$$\begin{aligned} \widehat{\text{Etr}(\gamma^{\otimes 2})} &= \frac{\text{Etr}(W^{\otimes 2}) + \text{Etr}^2(W)}{(n-K-1)(n-K+2)} \\ &\stackrel{\text{ThB.1.4c}}{=} \frac{(n-K)(n-K+1) - 2}{(n-K-1)(n-K+2)} \text{tr}^2(\gamma^{\otimes 2}) = \text{tr}^2(\gamma) \end{aligned} \quad (3.27)$$

QED

The reason behind the name of biased-reduction method is that both $\hat{\beta}^B, \hat{\kappa}^B$ are unbiased estimators of both unknown parameters β and κ . We can see, that for the fixed number of groups, with increasing sample size of whole functional sample, both estimators are coincides. Furthermore, we should not care about the unbiasedness property, because both estimators are consistent, which is shown by the following theorem.

Theorem 3.2.4 (Zhang [2012], Chapter 4). Let A1, A2, A4 holds. Then as $n \rightarrow \infty$, we have $\text{tr}(\hat{\gamma}) \xrightarrow{P} \text{tr}(\gamma)$ and $\text{tr}(\hat{\gamma}^{\otimes 2}) \xrightarrow{P} \text{tr}(\gamma^{\otimes 2})$. Furthermore as $n \rightarrow \infty$, we have

$$\hat{\beta} \xrightarrow{P} \beta, \quad \hat{d} \xrightarrow{P} d \quad (3.28)$$

Proof. See Zhang [2012], Chapter 4.

QED

We reject the null hypothesis for a big distances from the population mean, or more precisely whenever $T_n > \hat{\beta} \chi_{(K-1)\hat{d}}^2(1-\alpha)$. Both $\mathcal{L}^2$ -norm based tests using bias-reduced or naive method of estimation will be denoted as L2B and L2N.

The bias reduction methods have added value of that they include the information about the shape of the functional data, or more precisely the covariance function in its estimate procedure. However, the both mentioned tests are not based on the reconstructed functional data, which was described in Chapter 2. This will be discussed in the following section.

3.2.3 $\mathcal{L}^2$ -norm based test using orthonormal basis expansion

In the previous sections, there was introduced tests which was based on „raw” functional data, or more precisely they do not take into consideration the reconstruction technique. It can be shown, that assuming that we are able to approximate the functional datum by linear combination of orthonormal basis function, we can calculate the test statistics which won't depend on functional data, but only on the coefficients $\mathbf{c}$. In this section we will assume, that the functional datum can be approximated by linear combination of M basis functions as in Chapter 2. Then, we can rewrite test statistics L_n and T_n as

$$\begin{aligned} L_n &= \sum_{1 \leq i < j \leq K} n_i \int_{\mathcal{T}} [\bar{X}_i(t) - \bar{X}_j(t)]^2 dt \\ &= \left[\sum_{1 \leq i < j \leq K} n_i (a_i - 2b + a_j) \right] \int_{\mathcal{T}} \phi^\top(t) \phi(t) dt \quad \left(\int_{\mathcal{T}} \phi^\top(t) \phi(t) dt = 1 \right) \\ &= \sum_{1 \leq i < j \leq K} n_i (a_i - 2b + a_j) \end{aligned} \quad (3.29)$$

$$\begin{aligned} T_n &= \sum_{i=1}^K n_i \int_{\mathcal{T}} [\bar{X}_i(t) - \bar{X}(t)]^2 dt \\ &= \left[\sum_{i=1}^K n_i a_i - 2c + d \right] \int_{\mathcal{T}} \phi^\top(t) \phi(t) dt \\ &= \sum_{i=1}^K n_i a_i - 2c + d, \end{aligned} \quad (3.30)$$

where

$$\begin{aligned} a_i &= \frac{1}{n_i^2} \sum_{j,k}^{n_i} \mathbf{c}_{ji}^\top \mathbf{c}_{ki}, \quad i = 1, \dots, K \\ b &= \frac{1}{n_i n_j} \sum_{l=1}^{n_i} \sum_{k=1}^{n_j} \mathbf{c}_{li}^\top \mathbf{c}_{kj}, \quad i = 1, \dots, K \\ c &= \frac{1}{n_i n} \sum_{i=1}^K \sum_{j,l=1}^{n_i} \mathbf{c}_{ji}^\top \mathbf{c}_{li}, \quad i = 1, \dots, K \\ d &= \frac{1}{n} \sum_{i,j=1}^K \sum_{l,r}^{n_i, n_j} \mathbf{c}_{il}^\top \mathbf{c}_{jr}. \end{aligned}$$

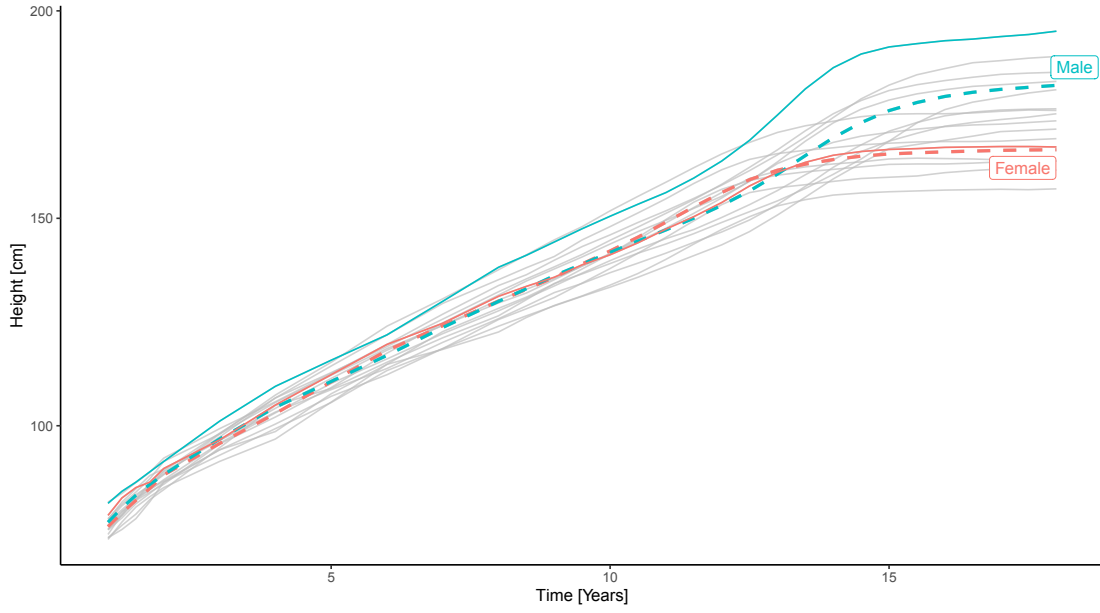


Figure 3.2: Growth data

As we can see on the derivations above, with help of the linear smoothing both, $\mathcal{L}^2$ -norm based test statistics are based only on unknown coefficients, which estimators are known. Therefore, no approximation distribution are needed and we can use for testing the (3.3) methods such as permutation test. In the practical examples in this thesis, we will use the test statistic (3.30). Finally, we denote the test based on (3.30) as **L2Recon**. The following example compares the aforementioned methods.

Example 5. *Let's take as an example the growth data from the introductory part. In this example, we are interested, if the mean evolution of heights of girls and boys is the same, or more precisely, whether girls and boys grow equally during adolescence. We take randomly 8 boys and 10 girls from the functional dataset. Furthermore, discretized time points $t_{31} = \{t_1, \dots, t_{31}\}$ represent the age of boy or girl, when the height was measured. Time points are not equidistant, but the domain of functional samples is the same (height is measured in the same times for boys and girls). The data can be found in Figure 3.2. We can see, that the empirical mean growth curve for boys and girls is the same until age 14, and after age 14, the curves are different. This makes sense, because in the adolescent years, the physical properties of boys are changing rapidly. We want to test*

$$H_0 : \mu_{\text{boys}}(t) = \mu_{\text{girls}} \quad \text{vs.} \quad H_1 : \mu_{\text{boys}}(t) \neq \mu_{\text{girls}}(t).$$

We applied for testing this hypothesis all L^2 -norm based test with help of **fdANOVA** package, where for Bootstrap parametric test we set $N = 10000$. The results of the test can be found in the table. We can see, that all of the test reject the H_0 on 5%-significance level and the evolution of heights is different across genders.

3.2.4 F -type tests

In the previous section, we introduced the alternatives to typical F -statistics. Class of $\mathcal{L}^2$ -norm based tests solved the so called globalization problem, where test statistics

Test	Test Statistic	Parameters
CS	$V_n = 12349.84$	$N = 10000$
CH	$V_n = 12349.84$	$N = 10000$
L2N	$T_n = 8233.23$	$\hat{\beta}^N = 784.254, \hat{\kappa}^N = 1.3699$
L2B	$T_n = 8233.23$	$\hat{\beta}^B = 721.9425, \hat{\kappa}^B = 1.416708$

Table 3.1: Result of the L^2 -norm based tests

is computed for every time point t at once, or more precisely, tests was not evaluated for the each $t \in \mathcal{T}$. In this section, we will return to tests, which will use the typical F statistic, but in the same time, they will solve all the mentioned problems at the beginning of Section 3.1. For the purposes of this section, let us introduce the test statistic we will work with

$$F_n = \frac{\frac{SSR}{K-1}}{\frac{SSE}{n-K}}, \quad (3.31)$$

where,

$$SSR = \int_{\mathcal{T}} SSR(t) dt \quad (3.32)$$

$$SSE = \int_{\mathcal{T}} SSE(t) dt. \quad (3.33)$$

The both statistics above measures, in some sense, the within subject (or one functional group) and between subject variation of functional samples on the whole follow-up interval, not just in one point.

Remark. We can rewrite denominator of (3.31) as

$$\frac{SSE}{n-K} = \int_{\mathcal{T}} (n-K)^{-1} \sum_{i=1}^K \sum_{j=1}^{n_i} [X_{ij}(t) - \bar{X}_i(t)]^2 dt = \int_{\mathcal{T}} \hat{\gamma}(t, t) dt = \text{tr}(\hat{\gamma}).$$

Therefore, F -type tests are including covariance function in its computations.

The class of tests based on the statistic (3.31) is called ***F-type tests***. Firstly, this type of the tests was introduced in functional linear models (see Faraway [1997] or Kokoszka and Horvath [2012]), but they can be used also in K -sample problems. Through this section, it will be shown, that the statistic above is nothing else than $\mathcal{L}^2$ -norm based taking into consideration the covariance function under some assumptions. The following theorem describe the distribution of statistic (3.31) under validity of some conditions, similar to these in $\mathcal{L}^2$ -norm based test.

Theorem 3.2.5 (Zhang [2011]). *Under A3, distribution of SSR and SSE is the same as the distributions of variables*

$$\sum_{l=1}^m \lambda_l A_l + \sum_{l=m+1}^{\infty} \pi_l^2, \quad \sum_{l=1}^m \lambda_l B_l, \quad (3.34)$$

where A_l, B_l are independent of each other and $A_l \sim \chi_{K-1}^2(\lambda_l^{-1} \pi_l^2), B_l \sim \chi_{n-K}^2$ and $\pi_l^2 =$

$\|\int_{\mathcal{T}} \mu(t)\phi_l(t)dt\|^2, l = 1, \dots, m$. Then, distribution of F_n is the same as the distribution of variable

$$\frac{\frac{\sum_{l=1}^m \lambda_l A_l + \sum_{l=m+1}^{\infty} \pi_l^2}{K-1}}{\frac{\sum_{l=1}^m \lambda_l B_l}{n-K}} \quad (3.35)$$

Proof. Distributions of SSR and SSE are discussed in Theorem 3.1.1. Therefore, it is sufficient to use the definition of F -type mixture and we get the desired result.

QED

Remark. The term $\sum_{l=m+1}^{\infty} \pi_l^2$ is more technical. The statistic above is mostly used to test the significance of main effect in one-way FANOVA, hence testing that the all mean functions all equal to 0 through the whole follow-up time. Then the F_n is nothing else than ratio of two χ^2 -mixture type variables. Theorem 3.2.5 above represent more general case.

As we said, the difference between $\mathcal{L}^2$ norm based test and F -type tests are that F -type tests are taking into consideration the covariance function of the sample. More precisely, F -type test statistic include in denominator $SSE = (n - K)\text{tr}(\hat{\gamma})$ (see proof of Theorem 3.1.1). Therefore, it is expected, that F -type test should outperform $\mathcal{L}^2$ -norm based tests (see Chapter 4). Furthermore, we can see that as in the case of $\mathcal{L}^2$ -norm based tests the distribution of test statistic F_n is hard to work with and some approximations are needed. The first method is typically bootstrap method. Bootstrap is mainly usable when we cannot say, that our data are generated from the Gaussian process. The algorithm for bootstrap based F -type test is shown in Algorithm 2.

Algorithm 2: Bootstrap-based F -type test

Input: B = Number of bootstrap simulations, observed F_n , number of observations per group n_j , number of groups K

Data: $t_n = \{t_1, \dots, t_n\}$, a $|t_n| \times Kn_j$ matrix $\mathbb{X}$ containing functional data

Output: Bootstrapped sample $F_1^*, \dots, F_B^*$

for $b \leftarrow 1$ **to** B **do**

$\mathbb{X}^* \leftarrow \text{ResampleWithReplacement} (X_{11}, X_{12}, \dots, X_{Kn_K};$

$\bar{X}_i^*(t) \leftarrow \frac{1}{n_i} \sum_{j=1}^{n_i} X_{i,j}^*, i = 1, \dots, K;$

$\bar{X}^*(t) \leftarrow \frac{1}{n} \sum_{i=1}^K \sum_{j=1}^{n_i} X_{i,j}^*(t);$

$SSR_b^* \leftarrow \sum_{t \in t_n} \sum_{i=1}^K n_i [\bar{X}_i^*(t) - \bar{X}^*(t)];$

$SSE_b^* \leftarrow \sum_{t \in t_n} \sum_{i=1}^K \sum_{j=1}^{n_i} [X_{ij}^*(t) - \bar{X}^*(t)] ;$

$F_b^* \leftarrow \frac{\frac{SSR_b^*}{K-1}}{\frac{SSE_b^*}{n-K}};$

end

Result: Reject H_0 whenever $F_n > F^*(1 - \alpha)$

The second method is normal approximation of test statistic F_n under the validity of H_0 . Because of statistic F_n include both χ^2 -mixture type variables, the usage of normal approximation for computing the critical values or p -values is not the best idea for the same reason as was mentioned for the normal approximation method

in $\mathcal{L}^2$ -norm based tests. In the case of F -type test, we set the mean and variance parameters of normal distributions as

(3.36)

$$\begin{aligned}\mu &= \frac{\mathcal{M}_1(A)}{\mathcal{M}_1(B)}, \\ &= \frac{n-1}{n-K} \\ \sigma^2 &= \frac{\mathcal{M}_1^2(A)}{\mathcal{M}_1^2(B)} \left[\frac{\mathcal{M}_2(A)}{\mathcal{M}_1^2(A)} + \frac{\mathcal{M}_2(B)}{\mathcal{M}_1^2(B)} \right] \\ &= \frac{(n-1)^2}{(n-K)^2} \left[\frac{\text{tr}(\gamma^{\otimes 2})}{n-1} + \frac{\text{tr}(\gamma^{\otimes 2})}{n-K} \right]\end{aligned}\tag{3.37}$$

where we can replace unknown $\text{tr}(\gamma^{\otimes 2})$ and $\text{tr}^2(\gamma)$ with naive or biased-reduced estimators.

The remaining method is the same as for the $\mathcal{L}^2$ -norm based tests - the cumulant matching method. Zhang [2011] or Zhang [2012] proposed the approximation of the null distribution of F_n by well known F -distribution with d_1 and d_2 degrees of freedom, where the goal is to estimate each parameter. With the help of Two-Cumulant Matched method, degrees of freedom has the form

$$d_1 = \frac{2\mathcal{M}_1^2(A)}{\mathcal{M}_2(A)}\tag{3.38}$$

$$= (K-1) \frac{\text{tr}(\gamma)}{\text{tr}(\gamma^{\otimes 2})}$$

$$\begin{aligned}d_2 &= \frac{2\mathcal{M}_1^2(B)}{\mathcal{M}_2(B)} \\ &= (n-K) \frac{\text{tr}(\gamma)}{\text{tr}(\gamma^{\otimes 2})}\end{aligned}\tag{3.39}$$

The estimators of both degrees of freedom is computed with help of both naive and bias reduction methods from section 2.3.3. The estimators are

$$\begin{aligned}\hat{d}_1^N &= (K-1)\hat{d}^N, \hat{d}_2^N = (n-K)\hat{d}^N \\ \hat{d}_1^B &= (K-1)\hat{d}^B, \hat{d}_2^B = (n-K)\hat{d}^B.\end{aligned}\tag{3.40}$$

where $\hat{d}^N, \hat{d}^B$, are estimated degrees of freedom from the naive and bias-reduced methods described in section 3.2.3. We can see that both estimates of degrees are not consistent estimators of d_1, d_2 . Therefore, we reject H_0 whenever $F_n > F_{\hat{d}_1^N, \hat{d}_2^N}$ (for naive method of estimation) or $F_n > F_{\hat{d}_1^B, \hat{d}_2^B}$ (for bias reduced method of an estimation). Unfortunately, every estimators introduced above works only under assumption A3, more specifically, if our functional samples are generated by a Gaussian process. If we do not have guarantee, that we are working with Gaussian data, class of F -type test are not applicable and the other type of test are needed - for example bootstrap-based $\mathcal{L}^2$ -norm and F -type tests. For the rest of this section we will abandon χ^2 -type and F -type mixtures and introduce the another test using the F -statistic

introduced at the beginning of this section. Until this point, we solve the problem of point-wise testing with test which measure the variation of each functional group over the whole follow-up time $\mathcal{T}$, or more precisely, we integrate each point-wise test statistic on the interval $\mathcal{T}$. Zhang et al. [2019] decided to treat the point-wise problem in different way, and take statistic F_n as the highest value of F -ratio statistic across whole $\mathcal{T}$, or more precisely, they introduced the new F -type test defined as

$$F_{\max} = \sup_{t \in \mathcal{T}} F_n(t). \quad (3.41)$$

Before we start discussion about the proposed test, it will be suitable to show the asymptotic distribution of $F_{\max}$. First, we need to introduce the following assumption.

Assumption 5. (A5)[Zhang et al. [2019]]

1. All functional samples are random elements from $\mathcal{L}^2(\mathcal{T})$,
2. sequence of differences $X_{ij}(t) - \mu_i(t), i = 1, \dots, K, j = 1, \dots, n_i$ are i.i.d $SP(0, \gamma)$,
3. As $n \rightarrow \infty$, the groups sample size satisfy $\frac{n_i}{n} \rightarrow c_i > 0, i = 1, \dots, K$ where $c_1, \dots, c_K \in (0, 1)$,
4. $E\|X_{ij}(t) - \mu_i(t)\|^4 < \infty$
5. covariance function satisfy $\text{tr}(\gamma) < \infty$ and variance at each point is positive,
6. $\text{cov}(X_{ij}(t) - \mu_i(t), X_{ij}(s) - \mu_i(s))$ is upper-bounded.

Theorem 3.2.6. Under validity of A5 and H_0 , $F_{\max} \xrightarrow{\mathcal{D}} R_0$, where R_0 has same distribution as

$$R = \sup_{t \in \mathcal{T}} \left\{ (K-1)^{-1} \sum_{i=1}^{K-1} w_i^2(t) \right\}, \quad (3.42)$$

where $w_1(t), \dots, w_{K-1}(t) \sim GP(0, \gamma_\rho)$.

Proof. See Zhang et al. [2019].

QED

What is interesting about this test is that, it includes the so-called *within-function correlation function* γ_ρ which has form

$$\gamma_\rho = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}}, \quad s, t \in \mathcal{T}, \quad (3.43)$$

and its estimator is defined as

$$\hat{\gamma}_\rho(s, t) = \frac{\widehat{\gamma(s, t)}}{\sqrt{\widehat{\gamma(s, s)}\widehat{\gamma(t, t)}}}, \quad s, t \in \mathcal{T}. \quad (3.44)$$

Authors of the proposed test find out, that this type of the test outperformed the F -type and $\mathcal{L}^2$ -norm based tests for the functional data, which has high within-function correlation. On the other hand, for the low within-function correlated functional data, $F_{\max}$ is slightly worse than aforementioned tests. For more details, see Zhang et al.

[2019], Section 4. This will be discussed more in Chapter 4, where we compare performance of test for the various amount of within-function correlations. For the evaluation of this test, the authors describe using parametric and non-parametric bootstrap for approximation of the null distribution of $F_{\max}$, where the procedure of those methods are described in Algorithm 4. Authors prefer more the non-parametric bootstrap, because parametric bootstrap require large sample size $n_1, \dots, n_K$. At the same time, it is well-known that the main drawbacks of both methods is the computation expensive. If we realize that the sequence $w_1(t), \dots, w_K(t)$ is generated from the zero-mean Gaussian process, so the statistic above is nothing else than the theoretical maximum of sample variance of $w_1(t), \dots, w_{K-1}(t)$. Therefore, the statistic R represent how different groups are in every time point $t \in \mathcal{T}$. If the covariance, or correlation of process is close to zero, or it is not significant in time point t , values of $w_i(t)$ will be jumping around zero, which implies low variation between groups, and the differences between groups are the result of randomness.

As in the preceding subsections, the approximation statistic R , or it distribution is hard to work with. Authors introduce parametric and non parametric bootstrap algorithms, which are demonstrated in 4

Algorithm 3: Bootstrapped-based $F_{\max}$ test

Input: B = Number of bootstrap simulations, number of groups K , observed $F_{n,\max}$, number of observations per group $n_j, j = 1, \dots, K$, bootstrap $\in \{\text{parametric, non-parametric}\}$

Data: $\mathbf{t}_n = \{t_1, \dots, t_n\}$, a $|\mathbf{t}_n| \times Kn_j$ matrix $\mathbb{X}$ containing functional data

Output: Bootstrapped sample $F_{1,\max}^*, \dots, F_{B,\max}^*$

if *bootstrap is parametric* **then**

for $b \leftarrow 1$ **to** B **do**

$\hat{\gamma}_\rho \leftarrow \text{CorrEstimate}(\mathbb{X})$ generate $w_i(t_k), i = 1, \dots, K-1, k = 1, \dots, m$,
 $w_i(t) \sim GP(0, \hat{\gamma}_\rho)$;
 $R_{b0} = \max_{t \in \{t_1, \dots, t_m\}} \left\{ (K-1)^{-1} \sum_{j=1}^{K-1} w_j^2(t) \right\}$

end

else

for $b \leftarrow 1$ B **do**

$V_{ij}(t) \leftarrow X_{ij}(t) - \bar{X}_j(t), i = 1, \dots, n_j, j = 1, \dots, K, t \in \mathbf{t}_n$;
 $V_{ij}^*(t) \leftarrow \text{BootstrapData}(V_{ij}(t))$;
 $SSR_b^*(t) \leftarrow \sum_{i=1}^K n_i \sum_{t \in \mathbf{t}_n} [\bar{V}_i^*(t) - \bar{V}^*(t)]^2, t \in \mathbf{t}_n$;
 $SSE_b^*(t) \leftarrow \sum_{i=1}^K n_i \sum_{t \in \mathbf{t}_n} [\bar{V}_i^*(t) - \bar{V}^*(t)]^2, t \in \mathbf{t}_n$;
 $F_{b,\max}^* \leftarrow \max_{t \in \{t_1, \dots, t_m\}} \left\{ \frac{SSR_b^*}{\frac{K-1}{n-K}} \right\}$

end

end

Result: Reject H_0 whenever $F_{n,\max} > F_{1-\alpha,\max}^*$

Algorithm above provide only one piece of approximation method of distribution of a random variable R . We need to make sure, that Algorithm 4 provide the reasonable estimate. The following theorem describe where the bootstrapped samples $F_{1,\max}^*, \dots, F_{B,\max}^*$ has desired distribution.

Theorem 3.2.7 (Zhang et al. [2019]). *Let A5 and let C^* be α -quantile of bootstrapped sample F_{max}^* and C_α be α -quantile of distribution of R_0 . Then as $n \rightarrow \infty$, we have $F_{max}^* \xrightarrow{\mathcal{D}} R_0$ and $C_\alpha^* \xrightarrow{\mathcal{D}} C_\alpha$.*

Proof. See Zhang et al. [2019].

QED

Theorem above is saying, that whole procedure is correct, when the sample size is sufficiently high. Therefore, we reject H_0 , whenever $F_{n,\max} > C_{1-\alpha}^*$, where $F_{n,\max}$ is observed test statistic and we denote this test as **Fmaxb**. **Fmaxb** test is one of the globalized F -type test.

Another interesting global version of F -type test is so called **Globalized point-wise F-test** (GPF), which uses test statistic $GPF_n = \int_{\mathcal{T}} F_n(t)$. Its asymptotic distribution, and approximation techniques for this distribution are in detail described in Zhang and Liang [2014].

3.2.5 F -type test using data reconstructions

In the previous two sections, we described tests, which calculations was based only on "raw" functional data, more precisely, they did not use the reconstruction techniques described in Chapter 2. For the first time, authors in Gorecki and Smaga [2015] introduce F -type test using techniques from the Chapter 2 and methods how to use such test for testing the hypothesis. Let us assume, that we can approximate functional data by a finite number of orthonormal basis functions

$$X_{ij}(t) = \sum_{m=0}^M c_{mij} \phi_m(t), t \in \mathcal{T}, \quad (3.45)$$

where $c_{mij}, m = 1, \dots, M$ are random variables with the finite variance. We can see, that we are assuming, that each functional group has different coefficients, but the orthonormal basis is the same across all groups. Then the reconstructed functional sample means are defined as

$$\bar{X}(t) = n^{-1} \sum_{j=1}^K \sum_{i=1}^{n_j} \mathbf{c}_{ij}^\top \phi(t), \quad \bar{X}_j(t) = n_j^{-1} \sum_{i=1}^{n_j} \mathbf{c}_{ij}^\top \phi(t), \quad (3.46)$$

where $\mathbf{c}_{ij} = (c_{ij1}, \dots, c_{ijM})^\top$ is random vector and $\phi(t) = (\phi_1(t), \dots, \phi_M(t))^\top$ are orthonormal basis functions. The test statistic is same as in the previous section

$$F_n = \frac{\frac{SSE}{K-1}}{\frac{SSR}{n-K}}. \quad (3.47)$$

Therefore, using the data reconstruction and the result from Section 3.2.4, the F_n has the form

$$F_n = \frac{\frac{\sum_{j=1}^K n_j^{-1} \sum_{m,l=1}^{n_j} \mathbf{c}_{jm}^\top \mathbf{c}_{jl} - \sum_{j,i=1}^K \sum_{m,l=1}^{n_j n_i} \mathbf{c}_{im}^\top \mathbf{c}_{jl}}{K-1}}{\frac{\sum_{j=1}^K \sum_{m=1}^{n_j} \mathbf{c}_{jm}^\top \mathbf{c}_{jl} - \sum_{j=1}^K n_j^{-1} \sum_{m,l=1}^{n_i} \mathbf{c}_{jm}^\top \mathbf{c}_{jl}}{n-K}} \quad (3.48)$$

We can see that with help of basis expansion method, the F -statistic is time independent and depend only on the number of groups, number of observations in each group. Fortunately, we derive the form of estimators of the unknown coefficients. Plugging the estimators of the unknown coefficients into (3.48), we have

$$F^* = \frac{\frac{\hat{a} - \hat{b}}{K-1}}{\frac{\hat{c} - \hat{a}}{n-K}}, \quad (3.49)$$

where

$$\hat{a} = \sum_{j=1}^K n_j^{-1} \sum_{k=1}^{n_j} \left[\left(Z_k^\top Z_k \right)^{-1} \right]^\top Z_k^\top \mathbf{X} \quad (3.50)$$

Evaluation of the test is different compared tests from the previous sections. The evaluation of the test is not based on some well-known distribution, but permutation method is used. The permutation procedure for this test is described in Algorithm 4.

Algorithm 4: Permutation based F -type test using basis expansion

Input: Number of permutations N , observed F_n , orthonormal basis function ϕ ,
Number of functional groups K , number of observations per sample
 $n_k, k = 1, \dots, K$

Data: $\mathbf{t}_n = \{t_1, \dots, t_n\}$, a $|t_n| \times Kn_j$ matrix $\mathbb{X}$ containing functional data

Output: Permutation sample $F_1^*, \dots, F_N^*$

$\hat{C} \leftarrow \text{FitCoefsBasisExpansion}(\mathbb{X});$

$\hat{b} \leftarrow \mathbf{b}(\hat{C});$

$\hat{c} \leftarrow \mathbf{c}(\hat{C});$

for $j \leftarrow 1$ **to** N **do**

$\mathbb{X}^* \leftarrow \text{PermuteAllData}(X_{11}, \dots, X_{K, n_K});$

for $k \leftarrow 1$ K **do**

$\hat{a}_k \leftarrow \mathbf{aPerGroup}(\mathbb{X}^*);$

end

$F_j^* \leftarrow \frac{\sum_k^K \hat{a}_k - \hat{b}}{\hat{c} - \sum_k^K \hat{a}_k};$

end

Result: Reject H_0 whenever $F_n > F_{1-\alpha}^*$.

The algorithm above represents the permutation-based F -type test using basis reconstruction. One of the input into this test is number of basis function. The choice of this parameter can be done by user, or can be chosen automatically with some well-known information criteria: BIC, AIC, eBIC, eAIC (for more details see Górecki and Smaga [2019]). For the rest of this thesis, described test as FP.

Until this point, we have introduced the various equality mean function tests, where we have shown that we have two options how to test the hypothesis (3.3): tests using the basis expansion or evaluation tests with the help of the distribution of $\mathcal{L}^2$ -norm tests and F -type tests. In Chapter 4 we will compare the performance of all the introduced tests, mainly comparison between aforementioned options for testing (3.3).

3.3 Tests for equality of covariance functions

Until this section, we were introduced tests, with assumptions, that functional samples comes from the stochastic process, with same covariance function. However, this assumption is often too strong, mostly for the functional data. It is problematic to assume that for the whole follow-up time, the functional samples have equal covariance functions. In this section we will investigate a more general case, where we assume that the covariance function is different across functional samples. The whole theory for this section will be mainly based on Theorem 1.5.4. First, let us remind some useful properties of the covariance function. If the functional data are coming from Gaussian process, we are able to derive the exact distribution of the covariance function $\gamma_i(s, t), i = 1, \dots, K$.

Theorem 3.3.1. *Let $X_1(t), \dots, X_n(t)$ be functional data generated by $GP(\mu(t), \gamma)$. Then*

$$(n-1)\hat{\gamma}_i(s, t) \sim WP(n-1, \gamma), i = 1, \dots, K \quad (3.51)$$

Proof. See Zhang [2012], Chapter 4.

QED

If the Gaussian assumption is violated, we know from Chapter 1 that as $n \leftarrow \infty$

$$\sqrt{n} [\hat{\gamma}_i(s, t) - \gamma_i(s, t)] \xrightarrow{\mathcal{D}} GP(0, \Gamma), i = 1, \dots, K. \quad (3.52)$$

where Γ is defined in Theorem 1.5.4. For the purpose of this section we will assume the following model.

Model. *Functional data $X_{i1}, \dots, X_{n_i, i}, i = 1, \dots, K$ are i.i.d. elements of $\mathcal{L}^2(\mathcal{T})$ and generated from $SP(\mu_i, \gamma_i(s, t))$.*

Remark. *It is important to note, that we are also assuming that the functional data has also same mean functions.*

In the testing of equality of covariance functions, we are interested in testing problem

$$H_0 : \gamma_1(s, t) = \dots = \gamma_K(s, t) \text{ vs. } H_1 : \neg H_0. \quad (3.53)$$

The important part of the inference is based on the common covariance estimator

$$\hat{\gamma}(s, t) = \sum_{i=1}^K (n_i - 1) \hat{\gamma}_i(s, t), \quad s, t \in \mathcal{T}. \quad (3.54)$$

The overall covariance function defined above has the similar functionality as the overall empirical mean function in the equality mean functions tests. Under validity of H_0 , the distance between group of covariance functions and the overall covariance functions should be close to zero. The following assumptions are necessary for deriving the distribution of the overall covariance function.

Assumption 6. *(A6) Let the model for equality covariance function test hold. Furthermore, let $tr(\gamma_i) < \infty, i = 1, \dots, K$ and as $n_{min} = \min_{i=1, \dots, K} n_i \rightarrow \infty$, group sample sizes satisfies $\frac{n_i}{n} \rightarrow c_i \in (0, 1), i = 1, \dots, K$.*

Theorem 3.3.2. (Zhang [2012]) Let A6 holds. Then as $n_{\min} \rightarrow \infty$

$$\left(\sqrt{n_1-1}\hat{\gamma}_1, \dots, \sqrt{n_K-1}\hat{\gamma}_K\right)^\top \xrightarrow{\mathcal{D}} GP_K(0, \text{diag}(\Gamma_1, \dots, \Gamma_K)) \quad (3.55)$$

,

Theorem 3.3.3. Let A6 holds. Then under validity of H_0 as $n_{\min} \rightarrow \infty$

$$\left(\sqrt{n_1-1}\hat{\gamma}_1, \dots, \sqrt{n_K-1}\hat{\gamma}_K\right)^\top \xrightarrow{\mathcal{D}} GP_K(0, \mathbb{I}). \quad (3.56)$$

Furthermore, under validity of null hypothesis

$$\hat{\gamma}(s, t) = \gamma(s, t) + o_p((n - K)^{-1/2}) \quad (3.57)$$

Remark. Theorem above tell us, that common covariance function estimator is $((n - K)^{-1/2})$ -consistent estimator of $\gamma(s, t)$.

In this section, we will describe only two tests for the testing the equality of the group covariance functions. Both of tests can be found in Zhang [2012], Chapter 10. First test will be based on the $\mathcal{L}^2$ -distance, similar to the one in Chapter 3 and we call it **$\mathcal{L}^2$ -norm based test of equality of covariance functions**. Test statistic is very similar to statistic (3.13) and is defined as

$$C_n = \sum_{i=1}^K (n_i - 1) \int_{\mathcal{T} \times \mathcal{T}} [\hat{\gamma}_i - \gamma(s, t)]^2 ds dt. \quad (3.58)$$

Same as in equality mean function tests, we need to find the distribution of (3.58). As in the case of mean equality tests, C_n can be approximated by χ^2 -type mixture distribution.

Theorem 3.3.4. Let A6 hold. Under validity of null hypothesis, as $n_{\min} \rightarrow \infty$

$$C_n \xrightarrow{\mathcal{D}} \sum_{r=1}^m \lambda_r A_r, \quad (3.59)$$

where $A_1, \dots, A_m$ is i.i.d sequence, $A_i \sim \chi_{K-1}^2$ and λ_r are eigenvalues of $\gamma(s, t)$.

Same as in the previous section, we want to approximate statistic C_n , for large sample size by statistic $T \sim \beta \chi_{(K-1)\kappa}^2$, where for estimating both β and κ we can again use bias reduction and naive estimation methods. Therefore, we are rejecting null hypothesis, or we are rejecting the equality of time variability of both groups, whenever $C_n > \hat{\beta} \chi_{(K-1)\kappa}^2$.

The second test we will use for testing the equality of covariance function is test based on the basis expansion. Let assume that we can approximate functional data by M basis orthonormal functions. Then we can, same is in the case F -type tests rewrite test statistic C_n in the means of unknown coefficients $\mathbf{c}$.

Theorem 3.3.5. *Let assume that functional data can be approximated by M basis functions. Then we can rewrite test statistic C_n as*

$$C_n = \sum_{i=1}^K (n_i - 1) \left[\frac{1}{(n_i - 1)^2} \left(\sum_{j,l}^{n_i} \Delta_{ijijilil} - \frac{2}{n} \sum_{j,l,m}^{n_i} \Delta_{ijijilim} + \frac{1}{n_i^2} \sum_{j,l,m,n}^{n_i} C_{ijilimin} \right) \right. \\ \left. + \frac{1}{(n_i - 1)(n - K)} \sum_{k=1}^K \left(\sum_{j,l}^{n_i, n_k} \Delta_{ijijklkl} - \frac{2}{n} \sum_{j,l,m}^{n_i, n_k} \Delta_{ijijklkm} + \frac{1}{n_i^2} \sum_{j,l,m,n}^{n_i, n_k, n_k} \Delta_{ijilkmkn} \right) \right. \\ \left. + \frac{1}{(n - K)^2} \sum_{k=1}^K \left(\sum_{j,l}^{n_k, n_k} \Delta_{kjkjklkl} - \frac{2}{n} \sum_{j,l,m}^{n_k} \Delta_{kjkjklkm} + \frac{1}{n_i^2} \sum_{j,l,m,n}^{n_k} \Delta_{kjkklkmkn} \right) \right],$$

where $\Delta_{abcdefgh} = \mathbf{c}_{ab}^\top \text{diag}(c_{cd1}c_{ef1}, \dots, c_{cdM}c_{efM}) \mathbf{c}_{gh}^\top$.

Proof. See the appendix. QED

For the evaluation of test using the statistic above we will use permutation test. Furthermore, we denote test above as **L2ecfRecon**. The procedure above is not the only way how we can include the basis expansion method in test statistic C_n . For example, we can use the functional principle components to express the covariance function of functional data. However, this method is out of the scope of this thesis, and details can be find Kokoszka and Horvath [2012], Chapter 3.

The final test proposed by Zhang [2012] is permutation test using the form of test statistic C_n . This test is mainly used, when we have small sample size per group. The testing procedure is described in Algorithm 5. Furthermore, we know that permutation test is feasible, when we have not sufficient high number of group sample sizes. In Chapter 4 we will investigate if it is necessary to use such as asymptotic approximation, or it will be more effective (in the means of computation cost vs. power of the test) use basis expansion test, or $\mathcal{L}^2$ -based test and permutation-based test.

Algorithm 5: Permutation test for test of equality of covariance functions

Output: $C_1^*, \dots, C_{\#Perm}^*$

Observe C_n ;

for $j \leftarrow 1$ **to** $\#Perm$ **do**

 Permute whole functional dataset;

$\hat{\gamma}_i^* \leftarrow \text{CovGroupEstimate}(X_{i1}^*, \dots, X_{in_i}^*), i = 1, \dots, K$;

$\hat{\gamma}^* \leftarrow \text{CovEstimate}(\hat{\gamma}_1^*, \dots, \hat{\gamma}_K^*)$;

$C_j^* \leftarrow \sum_{i=1}^K \sum_{t,s \in \mathcal{T}} [\hat{\gamma}_i^*(s, t) - \hat{\gamma}^*(s, t)]^2$

end

Result: Reject H_0 , whenever $C_n > C^*(1 - \alpha)$ or $\hat{p}^{\#Perm} \leq \alpha$

3.4 Tests for equality of distributions

In the previous sections, we introduced typical tests for testing the equality of main parameters describing the trend and shape of functional data - mean function and covariance function. However, we are able to construct dual hypothesis test to aforementioned tests. Such test is based on testing the equality of functional groups distributions.

Unfortunately, in this settings, this test is more general and we would not test the same thing. Therefore we need to make an additional assumptions.

Assumption 7. (A7) Functional data $X_{i1}, \dots, X_{in_i}, i = 1, 2$ are *i.i.d.* elements of $\mathcal{L}^2(\mathcal{T})$. Furthermore, they are coming from $SP(\mu_i, \gamma(s, t))$.

Our ultimate goal is to construct the dual hypothesis test to test of the equality of mean functions, which lead to the following hypothesis test

$$H_0 : SP_1 = SP_2 \text{ vs. } H_1 : SP_1 \neq SP_2. \quad (3.60)$$

If we reject the null hypothesis under these settings, we can say that in the average, the both functional samples are different. However, with using test to test the hypothesis above, which was created mainly to test the equality of both distributions, without such strong assumptions as we set in this section, we pay for less power test.

3.4.1 Functional ANOVA based on the empirical characteristic functions

In this section, we will introduce test for comparing two distributions which was introduced by Hlávka et al. [2021]. The authors proposed test in more general way, where they was interested into testing if the two stochastic process are equal, without any assumptions on the mean or the covariance functions.

We took this test as a dual test to the test of equality of mean. However, for the correct functionality of the mentioned test, the additionally assumptions needs to be specified. Therefore, for the whole section, we are assuming that A7 holds.

Test is based on the functional characteristic function of both functional samples, which is defined as

$$\psi_i(f) = \mathbb{E} \exp \left(i \int_{\mathcal{T}} f(t) X_p(t) dt \right), \quad i = 1, 2, \quad f \in L^2(\mathcal{T}). \quad (3.61)$$

Similar to the univariate random variables, random curves, or the functional data are determined by the values of $\psi(f), f \in \mathcal{L}^2(\mathcal{T})$. We are interested to test the hypothesis of equality of characteristic functions of both functional samples

$$H_0 : \psi_1 = \psi_2 \text{ vs. } H_1 : \psi_1 \neq \psi_2, \quad (3.62)$$

where $\psi_i, i = 1, 2$ represents the functional characteristic function for i -th group. Furthermore, we are assuming that both stochastic process differ only in its mean functions. If this assumption is violated we are no more talking about the FANOVA. Hlávka et al. [2021] proposed *Cramér-von Mises* type T -statistic for testing (3.62), which is defined as

$$T_{1,2}(Q) = \int_{\mathcal{T}} |\hat{\psi}_1(w) - \hat{\psi}_2(w)|^2 dQ(w), \quad (3.63)$$

where Q is Borel probability measure and $\hat{\psi}_k, k = 1, 2$ are empirical functional characteristic functions defined as

$$\hat{\psi}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} \exp \left(i \int_{\mathcal{T}} w(t) X_{k,j}(t) dt \right), \quad k = 1, 2, \quad (3.64)$$

It can be shown that the statistic above can be rewritten as

$$T_{1,2} = a_{1,1} + a_{2,2} - 2a_{1,2}, \quad (3.65)$$

where $a_{i,j} = \sum_{j=1}^{n_j} \sum_{i=1}^{n_i} \int_{L^2(\mathcal{T})} \cos\langle w, X_{k,i} - X_{k,j} \rangle dQ(w) / (n_i n_j)$ (see Hlávka et al. [2021], Section 3).

As we can see the proposed test statistic depends on the choice of measure Q . Right now let us focus on inner product or random projection $\langle w, X_{k,i} - X_{k,j} \rangle$. Let us assume that the Q is a centered Gaussian measure and let W be process with distribution Q . Then the functional characteristic function of projection $\langle W, X_{k,i} - X_{k,j} \rangle$ is defined as

$$\begin{aligned} \phi(w) &= \mathbb{E} \exp i\langle W, X_1 - X_2 \rangle \\ &= \mathbb{E} \cos\langle W, X_1 - X_2 \rangle + i\mathbb{E} \sin\langle W, X_1 - X_2 \rangle \end{aligned} \quad (3.66)$$

The term $i\mathbb{E} \sin\langle W, X_{k,i} - X_{k,j} \rangle = 0$ because

$$\begin{aligned} \mathbb{E} \sin\langle W, X_{k,i} - X_{k,j} \rangle &= \mathbb{E} \sin\langle W, X_{k,i} \rangle \cos\langle W, X_{k,i} \rangle - \mathbb{E} \sin\langle W, X_{k,j} \rangle \cos\langle W, X_{k,j} \rangle \\ &= \int \sin\langle w, X_{k,i} \rangle \cos\langle w, X_{k,i} \rangle dQ(w) \\ &\quad - \int \sin\langle w, X_{k,j} \rangle \cos\langle w, X_{k,j} \rangle dQ(w) \quad (\text{integrand is odd}) \\ &= 0. \end{aligned} \quad (3.67)$$

Therefore functional characteristic function of projection $\langle W, X_{k,i} - X_{k,j} \rangle$ is defined as

$$\phi(w) = \mathbb{E} \exp i\langle W, X_{k,i} - X_{k,j} \rangle = \mathbb{E} \cos\langle W, X_{k,i} - X_{k,j} \rangle. \quad (3.68)$$

Furthermore we are assuming that Q is Gaussian probability measure with covariance operator $V(w)(t) = \int_{\mathcal{T}} v(s, t) w(s) ds$, W has such distribution, therefore the above functional characteristic function has well known form

$$\begin{aligned} \phi_{X_{k,i} - X_{k,j}}(w) &= \exp\left(-\frac{1}{2}\langle W, X_{k,i} - X_{k,j} \rangle\right) \\ &= \exp\left(-\frac{1}{2} \int_{\mathcal{T}} \int_{\mathcal{T}} (X_{k,i}(t) - X_{k,j}(t))(X_{k,i}(s) - X_{k,j}(s))v(s, t) ds dt\right) \end{aligned} \quad (3.69)$$

Consequently, we obtain the more suitable form of test statistic $T_{1,2}$

$$\begin{aligned} T_{1,2} &= n_1^{-2} \sum_{j=1}^{n_1} \sum_{l=1}^{n_1} \exp\left(\frac{-1}{2} \int_{\mathcal{T}} \int_{\mathcal{T}} (X_{1,j}(t) - X_{1,l}(t))(X_{1,j}(s) - X_{1,l}(s))v(s, t) ds dt\right) \\ &\quad + n_2^{-2} \sum_{j=1}^{n_2} \sum_{l=1}^{n_2} \exp\left(\frac{-1}{2} \int_{\mathcal{T}} \int_{\mathcal{T}} (X_{2,j}(t) - X_{2,l}(t))(X_{2,j}(s) - X_{2,l}(s))v(s, t) ds dt\right) \\ &\quad - n_1^{-2} n_2^{-2} \sum_{j=1}^{n_1} \sum_{l=1}^{n_2} \exp\left(\frac{-1}{2} \int_{\mathcal{T}} \int_{\mathcal{T}} (X_{1,j}(t) - X_{2,l}(t))(X_{1,j}(s) - X_{2,l}(s))v(s, t) ds dt\right). \end{aligned} \quad (3.70)$$

Unfortunately the statistic above include still integrals, so they need to be estimated. The statistic above was constructed with help of projections or inner products, which will be used in estimating procedure. Furthermore, we need to take into consideration the fact that computations will be based on the discrete time grid, where discrete time are chosen from interval $[0, 1]$. Further, with the help of the assumption about distribution of Q we can take the random process $\mathbf{U}$ generated from Q -distribution (which is centered Gaussian), then the random vector $\mathbf{U} = (U(t_1), \dots, U(t_n))^{\top}$, where $t_1, \dots, t_n$

are discrete time points from the mentioned interval, has multivariate normal distribution with a mean vector equal to zero vector and variance matrix $\Sigma = \gamma(t_i, t_k), i, k = 1, \dots, n$ (see Definition A.2.1). We know, that each matrix component is continuous and bounded, therefore, we are able to estimate the inner product of space $\mathcal{L}^2(\mathcal{T})$ as

$$\widehat{\langle \mathbf{U}, \mathbf{z} \rangle} = \sum_{k=1}^s \frac{U(t_k)z(t_k)}{s} = \frac{1}{s} \langle \mathbf{U}, \mathbf{z} \rangle. \quad (3.71)$$

With the assumptions above, we have $\frac{1}{s} \langle \mathbf{U}, \mathbf{z} \rangle \sim \mathcal{N}(0, \frac{1}{s^2} \sum_k \sum_j z(t_k)z(t_j)\gamma(t_k, t_j))$. Because, we are assuming that W is a Gaussian process with a covariance operator V , the aforementioned estimation procedure holds, and we can estimate projection $\langle W, X_1 - X_2 \rangle$ by $\langle \mathbf{U}, X_1 - X_2 \rangle$. Finally, taking the all previous results, we get

$$\phi(w) = \mathbb{E} \exp i \langle \mathbf{U}, X_1 - X_2 \rangle \quad (3.72)$$

$$\begin{aligned} &= \mathbb{E} \cos \langle \mathbf{U}, X_1 - X_2 \rangle \\ &= \int_{\mathbb{R}^n} \cos \langle \mathbf{U}, X_{k,i} - X_{j,i} \rangle \quad (\text{using (3.68)}) \\ &= \exp \left(-\frac{1}{2} (X_{k,i} - X_{j,i})^\top \Sigma (X_{k,i} - X_{j,i}) \right) = R_{ki,kj}. \end{aligned} \quad (3.73)$$

If we plug-in the derivations above into (3.70), we get the final form of our test statistic for testing the hypothesis (3.62)

$$T_{1,2} = n_1^{-2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} R_{1i,1j} + n_2^{-2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} R_{2i,2j} - 2n_1^{-2}n_2^{-2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} R_{1i,2j}. \quad (3.74)$$

We can see that test statistic depend on the choosing of the matrix Σ . Hlávka et al. [2021] investigate performance of test under various options for Σ such as identity matrix, covariance matrix using functional principal components, or inverse of pooled sample covariance matrix and many more (for more details see mentioned article). For the evaluation of the proposed test, the authors choose the permutation test. They shown various options of using this test. One of the option is choose the matrix Σ and then permute both samples, or compute matrix Σ from the permuted samples. However, this step seems computation expensive. The whole testing procedure can be find in Algorithm 6 Evaluation of the test above is based on the estimated p -values, which has the form of $\hat{p}^* = \frac{\#T_j^* > T_{1,2}}{N}$. We reject H_0 , or more precisely equality of mean functions of both functional samples, whenever $\hat{p}^* \geq \alpha$, where α is pre-set significance level. For the rest of this thesis, we denote this test as CF.

As in the previous sections, we can also represent distribution equality test as test using orthonormal basis functions expansion. However, it will work only for the specific form of the orthonormal basis functions and the covariance operator V . Let us rewrite (3.69) as

$$\begin{aligned} \phi_{X_{k,i}-X_{k,j}}(w) &= \exp \left(-\frac{1}{2} \int_{\mathcal{T}} \int_{\mathcal{T}} (X_{k,i}(t) - X_{k,j}(t))(X_{k,i}(s) - X_{k,j}(s))v(s, t)dsdt \right) \\ &= \exp \left(-\frac{1}{2} \int_{\mathcal{T}} \int_{\mathcal{T}} (\mathbf{c}_{k,i}\phi(t) - \mathbf{c}_{k,j}\phi(t))(\phi^\top(s)\mathbf{c}_{k,i} - \phi^\top(s)\mathbf{c}_{k,j})v(s, t)dsdt \right) \\ &= \exp \left(-\frac{1}{2} (\mathbf{c}_{k,i}^\top - \mathbf{c}_{k,j}^\top) \left\{ \int_{\mathcal{T}} \int_{\mathcal{T}} \phi(t)\phi^\top(s)v(s, t)dsdt \right\} (\mathbf{c}_{k,i} - \mathbf{c}_{k,j}) \right) \\ &= R_{ki,kj}^* \end{aligned} \quad (3.75)$$

Algorithm 6: Permutation test for test of equality of functional characteristic functions

Input: Number of permutations N , observed $T_{1,2}$, choice of matrix Σ , number of observations per sample $n_k, k = 1, 2$

Data: $t_n = \{t_1, \dots, t_n\}$, a $|t_n| \times 2n$ matrix $\mathbb{X}$ containing functional data

Output: Permutation sample $T_{1,1,2}^*, \dots, T_{N,1,2}^*$

```

for  $j \leftarrow 1$  to  $N$  do
   $\mathbb{X}^* \leftarrow \text{PermuteAllData}(X_{11}, \dots, X_{2n_2});$ 
  if Covariance matrix is permutation dependent then
     $\hat{\Sigma} \leftarrow \text{CovEstimate}(\mathbb{X}^*);$ 
     $T_{j,1,2}^* \leftarrow T_{1,2}(\mathbb{X}^*, \hat{\Sigma});$ 
  else
     $\hat{\Sigma} \leftarrow \text{CovEstimate}(\mathbb{X});$ 
     $T_{j,1,2}^* \leftarrow T_{1,2}(\mathbb{X}^*, \hat{\Sigma});$ 
  end
end
Result: Reject  $H_0$  whenever  $T_{1,2} > T_{1,2}^*(1 - \alpha)$ .

```

The term $v(s, t)$ is nothing else than covariance function, and if $\text{tr}(v) < \infty$ holds, then by Zhang [2011]

$$v(s, t) = \sum_{r=1}^M \lambda_r \phi_r(t) \phi_r(s). \quad (3.76)$$

where $\phi_r(t), r = 1, \dots, M$ is an orthonormal basis of $\mathcal{L}^2(\mathcal{T})$ space. If we plugin (3.76) into (3.75), we have

$$\begin{aligned}
R_{ki,kj}^* &= \exp \left(-\frac{1}{2} (\mathbf{c}_{k,i}^\top - \mathbf{c}_{k,j}^\top) \left\{ \int_{\mathcal{T}} \int_{\mathcal{T}} \phi(t) \phi^\top(s) v(s, t) ds dt \right\} (\mathbf{c}_{k,i} - \mathbf{c}_{k,j}) \right) \\
&= \exp \left(-\frac{1}{2} (\mathbf{c}_{k,i}^\top - \mathbf{c}_{k,j}^\top) \left\{ \int_{\mathcal{T}} \int_{\mathcal{T}} \phi(t) \phi^\top(s) \sum_{r=1}^M \lambda_r \phi_r(t) \phi_r(s) ds dt \right\} (\mathbf{c}_{k,i} - \mathbf{c}_{k,j}) \right) \\
&= \exp \left(-\frac{1}{2} (\mathbf{c}_{k,i}^\top - \mathbf{c}_{k,j}^\top) \left\{ \sum_{r=1}^M \int_{\mathcal{T}} \int_{\mathcal{T}} \phi(t) \phi^\top(s) \lambda_r \phi_r(t) \phi_r(s) ds dt \right\} (\mathbf{c}_{k,i} - \mathbf{c}_{k,j}) \right) \\
&= \exp \left(-\frac{1}{2} (\mathbf{c}_{k,i}^\top - \mathbf{c}_{k,j}^\top) \text{diag}(\lambda_1, \dots, \lambda_M) (\mathbf{c}_{k,i} - \mathbf{c}_{k,j}) \right)
\end{aligned} \quad (3.77)$$

The unknown coefficient can be estimated by the methods from Chapter 2. Unfortunately, here comes the problem with estimating $\lambda_r, r = 1, \dots, M$. One can take the first M principal components or eigenvalues of covariance function estimator, but those variables will be related to other basis functions, and we will not get the correct $\text{diag}(\lambda_1, \dots, \lambda_M)$. However, if we realise that the term $\int_{\mathcal{T}} \int_{\mathcal{T}} \phi(t) \phi^\top(s) v(s, t) ds dt$ is nothing else than a matrix of constants, then we can choose $v(s, t) = \int_{\mathcal{T}} \int_{\mathcal{T}} \phi(t) \phi^\top(s) v(s, t) ds dt = \text{diag}(1, \dots, 1)$. Therefore $R_{ki,ki}^*$ can be written

as

$$\begin{aligned}
R_{ki,ki}^* &= \exp \left(-\frac{1}{2} (\mathbf{c}_{k,i}^\top - \mathbf{c}_{k,j}^\top) (\mathbf{c}_{k,i} - \mathbf{c}_{k,j}) \right) \\
&= \exp \left(-\frac{1}{2} (\mathbf{c}_{k,i}^\top \mathbf{c}_{k,i} - \mathbf{c}_{k,j}^\top \mathbf{c}_{k,j} - 2 \mathbf{c}_{k,j}^\top \mathbf{c}_{k,i}) \right)
\end{aligned} \tag{3.79}$$

Consequently, we can rewrite (3.74) as

$$B_{1,2} = n_1^{-2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} R_{1i,1j}^* + n_2^{-2} \sum_{i=1}^{n_2} \sum_{j=1}^{n_2} R_{2i,2j}^* - 2n_1^{-2} n_2^{-2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} R_{1i,2j}^*. \tag{3.80}$$

The evaluation of the test using the statistic above is equivalent to the procedure described in Algorithm (6). Various options for $v(s, t)$ can be found in an article by Hlávka et al. [2021], Section 4.2. One of the topics that will be examined in Chapter 4 will be a comparison of the performance of the mentioned tests under various time grid size, the choice of the orthonormal basis, its dimension and the matrix Σ . We denote this test as CFRecon.

4. Simulation study

In this chapter, we will conduct various simulations, where we compare tests introduced in Chapter 3 under various scenarios. In the whole chapter, we consider artificially generated data, where the data generation process is motivated by Gorecki and Smaga [2015], Cuevas et al. [2004]. We consider a theoretical data model for K -functional samples:

$$X_{ij}(t) = \mu_i(t) + \epsilon_{ij}(t), i = 1, \dots, K, j = 1, \dots, n_i. \quad (4.1)$$

where n_i is number of functional samples in i -th group and $\epsilon_{ij}(t) \sim SP(0, \gamma_i(s, t))$. However, in reality we do not have that much computational power to generate such functional data (we have not ability to generate infinitely many time points). Therefore, in our simulation study the artificially generated data (which we believe comes from the process (4.1)), has the form

$$X_{ij}(t_k) = \mu_i(t_k) + \epsilon_{ij}(t_k), t_k \in \mathcal{T}, \quad (4.2)$$

where $\mathcal{T}$ is interval including discretized time points.

For the whole simulation study, the follow up time $\mathcal{T} = [0, 1]$, where the time grid is generated as $t_k = \frac{k}{D+1}, k = 1, \dots, D$, and D is the size of the time grid. The simulations were performed with the help of the statistical software $\mathbb{R}$, where for the equality mean function tests `fdANOVA` (Górecki and Smaga [2019]) package was used and for the remaining tests (covariance and distribution equality tests) was implemented by a raw software code. The code can be found in the attachment of the thesis.

4.1 Simulations of equality of mean functions tests

In this section, we will conduct simulations on performance of proposed tests from Chapter 3, Section 3.2 - 3.3. For the clarity, we will use EMFT as abbreviation for the equality mean function tests. We will investigate:

- Power and size, under various scenarios - e.g the size of the time grid, the amount of error in data, the shape of the random curves (amount of within-function correlation),
- Power and size of tests using orthonormal basis expansion,
- How many basis functions are needed to reach the significance level and power.

The following models (Hypothesis models) for functional data are taken from Cuevas et al. [2004], Gorecki and Smaga [2015], Cuesta-Albertos and Febrero-Bande [2010] and Górecki and Smaga [2019]:

- Data generated under the validity of H_0
 - $\mathcal{F}_1$: $\mu_i^1(t) = t(t-1), \quad i = 1, \dots, K,$
 - $\mathcal{F}_2$: $\mu_i^2(t) = \sqrt{t}, \quad i = 1, \dots, K$
- Data generated in favor of the alternative
 - $\mathcal{F}_3$: $\mu_i^{3\delta}(t) = t(t-1) + \delta(i-1)t(t-1), \quad i = 3, \dots, K, \delta > 0,$

- $\mathcal{F}_4$: $\mu_i^{4\delta}(t) = \sqrt{t} + \delta(i-1)\sqrt{t}$, $i = 3, \dots, K$, $\delta > 0$,
- $\mathcal{F}_5$: $\mu_i^5(t) = t^{\frac{i}{5}}(1-t)^{6-\frac{i}{5}}$, $i = 1, \dots, K$,
- $\mathcal{F}_6$: $\mu_i^{6\alpha}(t) = 1 + \frac{i}{50} + (\alpha)(-1)^i[\sin(4\pi t)]$, $i = 1, \dots, K$, $\alpha \in \{0, 0.05, \dots, 0.45\}$,
- $\mathcal{F}_7$: $\mu_i^7(t) = 1 + \frac{i}{50}$, $i = 1, \dots, K$.

Model $\mathcal{F}_1$ should represent some process which has only one period. Model $\mathcal{F}_2$ should mimic the evolution of the height from Example 1. Models $\mathcal{F}_{5,6,7}$, with combination of error term generating from a Wiener process, represents some interesting examples of data. For example, model $\mathcal{F}_5$ generates functional data with different mean functions at the beginning of the follow-up time, but after some time point, their mean functions are almost same. We investigate if the proposed tests are able to detect those differences at beginning. Model $\mathcal{F}_6$ provides periodical functional data, which can mimic weather data `CanadianWeather` from package `fda`. The last model represents just constant functional with different absolute terms.

For each hypothesis model, we also need to include the variability and noise. We consider three types of the errors, same as was proposed in Cuevas et al. [2004] and Górecki and Smaga [2019]

- $\mathcal{E}_1$: $\epsilon_{ij}(t_k)$ are i.i.d. zero-mean random variables with variance equal to σ^2 ,
- $\mathcal{E}_2$: error terms $\epsilon_{ij}(t) \sim SP(0, \gamma)$ are i.i.d., where we set $\epsilon_{ij}(t) = \mathbf{a}^\top \boldsymbol{\phi}(t)$, where $\mathbf{a} = (a_1, \dots, a_{11})^\top$, $a_l = \sqrt{\lambda} z_l$, $\lambda = 1.5\rho^l$, $\rho = 0.1, \dots, 0.9$, $l = 1, \dots, 11$, z_l are i.i.d. zero-mean random variables with unit variance and $\boldsymbol{\phi}(t)$ is an orthonormal basis such that $\phi_1(t) = 1$, $\phi_{2k+1}(t) = \sqrt{2} \cos(2\pi kt)$, $\phi_{2k}(t) = \sqrt{2} \sin(2\pi kt)$. Therefore, the covariance function of functional data is equal to $\boldsymbol{\phi}^\top(t) \text{diag}(\lambda_1, \dots, \lambda_{11}) \boldsymbol{\phi}(s)$. The random variables z will be usually standard normal random variables or non-Gaussian random variables (to investigate behaviour of tests under violation of the assumptions about the distribution). Additionally, ρ represents the amount of within-function correlation. With larger values (e.g $\rho = 0.9$), data are low correlated. With lower values (e.g $\rho = 0.1$), data are highly correlated,
- $\mathcal{E}_3$: model for error terms is almost same as for the model $\mathcal{E}_2$, except one element of orthonormal basis. Here we set $\phi_2(t) = \phi_2(t) + (j-1)\omega$, $j = 1, \dots, K$, $\omega > 0$,
- $\mathcal{W}_1$: ϵ_{ij} are i.i.d. standard Wiener process with a dispersion parameter σ .

Errors generated by error model $\mathcal{E}_2$, are used for comparing $F_{\max}$ test with all mentioned tests. In this settings we are able to set the amount of the within-function correlation, where the behaviour of tests are point of interest. Zhang et al. [2019] described that the aforementioned test should over-perform the GPF and FP test, but they did not consider the remaining EMFT tests described in Chapter 3. Error models $\mathcal{E}_1$ and $\mathcal{W}_1$ control the amount of error in each functional model. Finally, error model $\mathcal{E}_3$ represents the model in the favour of alternative in the equality covariance tests.

Remark. With help of error model $\mathcal{E}_2$, Górecki and Smaga [2019] wanted to achieve the form of covariance function, which is define as $\gamma(s, t) = \sum_{w=1}^{11} \lambda_w \phi_w(s) \phi_w(t)$. However this is valid only if the condition $\text{tr}(\gamma) < \infty$ holds.

Finally, the functional data is generated as

$$X(t) = \text{Hypothesis model} + \text{Error model} \quad (4.3)$$

4.1.1 Simulations under validity H_0

We will consider two scenarios in this section

- **Scenario I:** Comparing all EMFT controlling only the amount of error in generated data. More precisely hypothesis models $\mathcal{F}_1$ and $\mathcal{F}_2$ with combination of error models $\mathcal{E}_1$ and $\mathcal{W}_2$,
- **Scenario II:** Comparing all EMFT under error model $\mathcal{E}_2$ and hypothesis models $\mathcal{F}_1$ and $\mathcal{F}_2$.

The simulation settings for both cases can be found in Table 4.1.

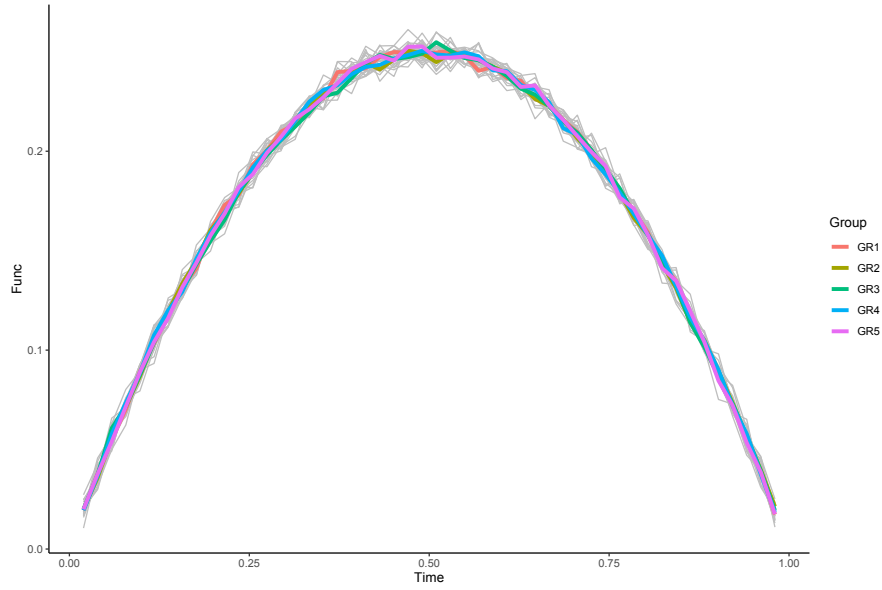
	Scenario I	Scenario II
Number of groups	$K = 5$	$K = 5$
Time grid size	$D \in \{25, 50, 70\}$	$D = 70$
The variance/correlation parameter	$\sigma \in \{0.2, 1, 4\}$	$\rho \in \{0.1, 0.3, \dots, 0.9\}$
Number of MC simulations	$B = 500$	$B = 500$
Number of samples per group	$n_j \in \{10, 15, 25\}$	$n_j = 10$

Table 4.1: Hyperparameters settings for simulation study under validity of H_0

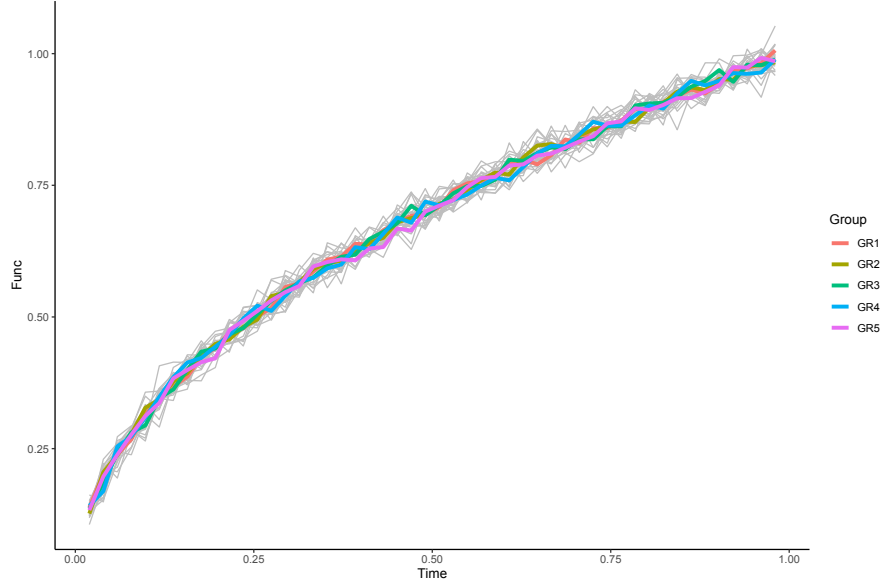
For both scenarios we artificially generate data under the model $\mathcal{F}_1$ and $\mathcal{F}_2$, where sample trajectories, with their mean functions are plotted on Figure 4.1. The model $\mathcal{F}_1$ mainly serves for comparing approximation EMFT (FB, L2B, FN, L2N), bootstrap based EMFT (Fb, L2b) with EMFT using data as basis representation (FP, L2recon) and globalized EMFT (GPF, $F_{\max}$). The second model $\mathcal{F}_2$ with error model $\mathcal{E}_2$ serves mainly for checking the behaviour of EMFT under controlling the amount of within-function correlation (the parameter ρ from $\mathcal{E}_2$). The behaviour of correlated functional data was mainly investigated only on $F_{\max}$ and GPF tests across the literature, but did not considering the other tests mentioned in Chapter 3 (see Zhang et al. [2019]). Therefore we consider two error models in this section – error model $\mathcal{E}_1$ with combination of model $\mathcal{F}_1$ (Figure 4.1) and error model $\mathcal{E}_2$ with combination $\mathcal{F}_1$ and $\mathcal{F}_2$ (see Figure 4.2). In the error model $\mathcal{E}_1$, error terms are normally distributed. With the both error models we are assuming that our data are generated from the Gaussian process and therefore all theory described in Chapter 3 is applicable (we can use naive and bias-reduced estimation methods). The goal of the simulation was to investigate the empirical power for each test under various settings (mainly number of observations, the variance parameter and the choice of the error model).

4.1.2 Empirical level

In Table 4.2 there can be found the results of the simulation under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, where the $\epsilon(t_k) \sim \mathcal{N}(0, \sigma^2)$. We can see that $\mathcal{L}^2$ -norm based test based on the approximation methods (L2B, L2N) with bootstrap based tests (CH, Fb) did not reach the desired significance level – bootstrap based test has empirical significance level around 3%, where the tests based on the χ^2 -type mixture approximation technique has empirical significance level around 8%. On the other hand, all test



(a) Artificial data generated under model $\mathcal{F}_1$ and the error model $\mathcal{E}_1$ with $\sigma = \frac{0.2}{51}$, $D = 50$, $n_j = 25$, $j = 1, \dots, 5$.

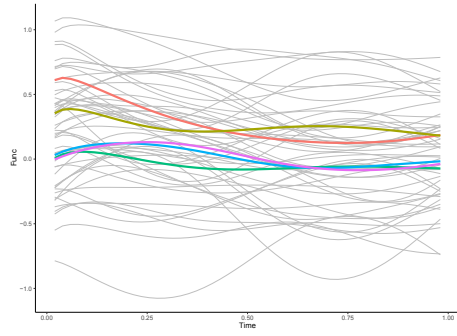


(b) Artificial data generated under model $\mathcal{F}_2$ and the error model $\mathcal{E}_1$ with $\sigma = \frac{1}{50}$, $D = 50$, $n_j = 25$, $j = 1, \dots, 5$.

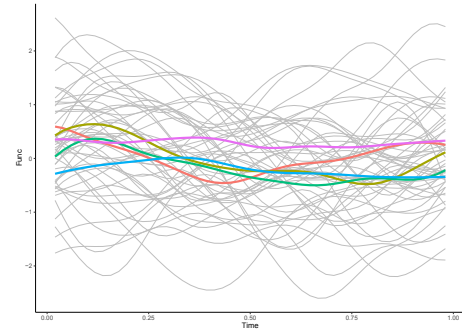
Figure 4.1: Artificial data generated under models $\mathcal{F}_1$ and $\mathcal{F}_2$ and error model $\mathcal{E}_1$.

using F -statistic (except the bootstrap based test) has approximately empirical levels around 4%-6%, where the permutation tests based on basis expansions (FP and L2recon)¹ and Fmax test was closest to the significance level. Finally, we can see that with normal errors, there is no dependency between the variance and the significance level. In the second simulation we compared the empirical sizes of EMFT under hypothesis models $\mathcal{F}_{2,3}$ and error model $\mathcal{E}_2$. This investigation was motivated by Górecki and

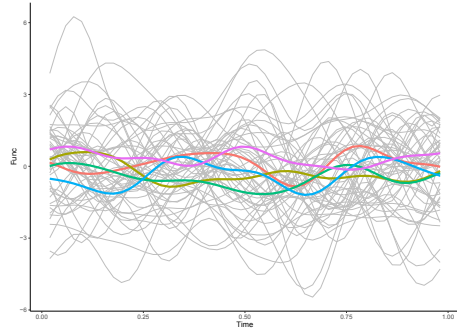
¹In these tests, we use orthonormal basis as Fourier series and the dimension of the basis was chosen automatically



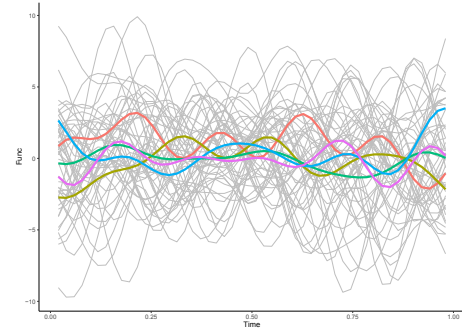
(a) Artificial data generated under model $\mathcal{F}_2$ and the error model $\mathcal{E}_1$ with $\rho = 0.1$, $D = 50$, $n_j = 10$, $j = 1, \dots, 5$.



(b) Artificial data generated under model $\mathcal{F}_2$ and the error model $\mathcal{E}_1$ with $\rho = 0.4$, $D = 50$, $n_j = 10$, $j = 1, \dots, 5$.



(c) Artificial data generated under model $\mathcal{F}_2$ and the error model $\mathcal{E}_1$ with $\rho = 0.7$, $D = 50$, $n_j = 10$, $j = 1, \dots, 5$.



(d) Artificial data generated under model $\mathcal{F}_2$ and the error model $\mathcal{E}_1$ with $\rho = 0.9$, $D = 50$, $n_j = 10$, $j = 1, \dots, 5$.

Figure 4.2: Artificial data generated under models $\mathcal{F}_2$ and error model $\mathcal{E}_2$.

Smaga [2019], where they compare only GPF and $F_{\max b}$ tests. We were considering also the $\mathcal{L}^2$ -norm based test and remaining F -type test. In this settings, for the fixed grid size and sample size, we strengthened the within-function correlation (and so controlling the shape of the functional data). In Table 4.5 there can be find empirical levels ($\delta = 0$) under various amount of within-correlation. As it can be seen, bootstrap-based tests **Fb** and **CH** did not reach the significance level even if the sample size is increased. $\mathcal{L}^2$ -norm based tests with naive method and bias-reduced method of estimation seems to be unstable. With decreasing amount of the correlation, they are close to the significance level, but with highly correlated data, they are far away from significance level. On the other hand, F -type test using naive and bias-reduction method of approximation, seems quite stable no matter the sample size or the amount of correlation. Interesting result provide also the tests using basis expansion (**L2recon** and **FP**) has not any problem to reach the pre-specified level.

Test	$n_j=10$			$n_j=15$			$n_j=25$		
	$\sigma=0.2$	$\sigma=1$	$\sigma=4$	$\sigma=0.2$	$\sigma=1$	$\sigma=4$	$\sigma=0.2$	$\sigma=1$	$\sigma=4$
FP	5.80	5.00	6.20	7.00	7.0	6.00	5.40	3.80	5.00
CH	4.20	2.00	4.00	5.40	5.00	3.60	4.40	3.00	4.40
L2N	8.00	6.60	8.20	8.60	7.60	7.20	5.80	4.40	5.00
L2B	8.20	7.40	8.60	9.20	8.00	8.20	6.80	5.00	5.00
L2b	7.20	5.80	8.00	8.20	7.00	6.80	5.60	3.80	4.00
FN	5.60	4.80	5.80	7.00	6.40	6.20	5.40	4.00	4.80
FB	5.80	5.00	6.00	7.00	6.60	6.60	5.40	4.00	4.80
Fb	3.60	2.20	3.40	5.20	4.40	4.00	4.20	2.80	4.20
GPF	7.40	6.20	8.00	8.40	7.40	7.00	6.00	4.40	5.00
Fmaxb	5.00	4.60	5.80	6.40	5.40	6.20	6.40	3.80	4.40
L2recon	4.80	5.40	5.80	4.40	5.20	4.90	5.80	4.40	5.00

Table 4.2: Empirical levels in % after 500 Monte Carlo simulations, error model $\mathcal{E}_1$, $D = 70$, $K = 5$.

4.1.3 Simulation under H_1

In this section, we will investigate the empirical power of mentioned tests. The simulations was mainly done under hypothesis models $\mathcal{F}_3$ and $\mathcal{F}_4$. The remaining simulations under models $\mathcal{F}_i, i = 3, \dots, 7$, where we mainly were working with error models $\mathcal{W}_1$ and $\mathcal{E}_2$ can be found in Appendix. The reason for considering $\mathcal{W}_1$ is that it provides (under model settings) interesting shapes of the functional samples mentioned in the Section 4.1.2. Furthermore we will investigate the performance of the test using orthonormal basis expansion L2Recon, FP with tests Fmaxb, GPF, L2B, L2N, FB, FN under various scenarios and hyperparameter settings, which can be found in Table 4.3. As in the previous section, we consider two scenarios

- **Scenario I:** Comparing all EMFT controlling only the amount of error in generated data. More precisely hypothesis models $\mathcal{F}_1$ and $\mathcal{F}_2$ with combination of error models $\mathcal{E}_1$ and $\mathcal{W}_1$,
- **Scenario II:** Comparing powers of test using basis expansion under various basis dimension and the size of the time grid.

	Scenario I	Scenario II
Number of groups	$K = 5$	$K = 5$
Time grid size	$D = 70$	$D \in \{5, 10, 15, 50, 200\}$
The variance/correlation parameter	$\sigma \in \{0.2, 1, 4\}$ $\rho \in \{0.1, 0.3, \dots, 0.9\}$	$\sigma \in \{0.1, 2, 5\}$
Number of MC simulations	$B = 500$	$B = 500$
Number of samples per group	$n_j \in \{10, 15, 25\}$	$n_j = 15$
Number of basis expansions	Number of basis is determined by the implemented information criterion	$M = 1, 5, 15, 50, 100, 200$

Table 4.3: Parameter settings for simulation study under validity of H_1

4.1.4 Empirical power

In Table 4.4 can be found the empirical power obtained under various variance settings for the error term. In total, with increasing variance in functional data, has problem detect the changes in the mean functions between the functional groups. The worst performance is shown mainly by bootstrap-based $\mathcal{L}^2$ -norm (CH) and F -type test (Fb), where with increasing amount of the error ($\sigma \in \{1, 5\}$), even under big sample sizes, they are not able to reach reasonable power. For example, for the parameter $\sigma = 0.2$ and $n_j = 20$, Fb test reaches only 56%, CH reaches 89%. For the other cases, they mostly reach 0% or maximum 10%. The almost same problem have tests using naive and biased-reduction approximation methods (L2N, L2B, FN, FB). With increasing amount of variance, they have ability to detect changes in the mean functions. Fortunately, it can be fixed with increasing sample size. For example, L2B test under $\sigma = 5, n_j = 5, \delta = 0.1$ has power 31.2%, but for $n_j = 20$, the power is 99%. Comparing the naive and bias-reduction

method of approximation, we can see that bias-reduction method has more power than the tests using the naive method, even under small sample sizes. In final, it seems that tests using basis expansion FP and L2recon have comparable power, where the latter have the highest power. More precisely, with high amount of error, L2recon has more empirical powers with small sample sizes (L2recon reaches 75 % with $n_j = 5$, and 99.2 % with $n_j = 10$). Surprisingly, the quite bad performance on the functional data has global test Fmaxb. With high amount of error, even with big sample sizes, it has power only 32%.

It can be seen, that the tests L2recon and FP has quite reasonable properties (sufficiently high power), even under the big amount of the error ($\sigma = 5$) or highly correlated data ($\rho = 0.1$). Unfortunately, Table 4.4 and Table 4.5 has shown only small piece of the information about the performance of FP and L2recon. In these tables, the number of basis functions are chosen under some automatically criterion and we have not controlled the size of time grid D .

In Table 4.6 can be found results of **Scenario II** simulations. In this scenario, we were trying to compare tests using basis expansion as the approximation method (FP, L2recon) with methods which has reasonable results in **Scenario I** (L2B, L2N, GPF, FB). The main investigation task in this was trade-off between power and computation powers under which M , test using reconstruction methods reach the desired power. It can be seen that the power of the tests using basis reconstruction, does not depend on on the basis dimension. More specifically, if we increase the number of basis functions or vice versa, the performance of the test is unchanged. For example, for a fixed grid size $D = 15$, shift parameter $\delta = 0.1$ and $\sigma = 5$, with the increasing number of basis functions, the power is around 17 – 18% for L2Recon and 12 – 15% for FP. Furthermore, if the time grid size is small, performance of the tests is not different than the default tests. At the bottom of the same table, there are shown computation times of each test. It can be seen that L2N, FB are the fastest (which is not surprising, because both test are not result of some iterative procedure). In final, using L2Recon and FP does not bring any advantage. Even with one basis function, tests have ability to reach the desired power. Furthermore, under Gaussian data, tests based on the orthonormal basis functions are comparable in the means of performance and computation time than L2B and FB. However, if the Gaussian assumptions is not valid, we can not use both L2B and FB and FP or L2recon can be more appropriate than the remaining tests.

		$\sigma = 0.2$				$\sigma = 1$				$\sigma = 5$			
		$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.05$	$\delta = 0.1$
FP	$n_j = 5$	45.20	100.00	100.00	100.00	4.60	7.60	100.00	100.00	4.40	6.80	8.00	25.60
	$n_j = 10$	91.20	100.00	100.00	100.00	6.80	15.60	100.00	100.00	4.80	6.00	16.20	70.80
	$n_j = 20$	100.00	100.00	100.00	100.00	8.20	32.60	100.00	100.00	6.20	7.40	31.80	99.60
CH	$n_j = 5$	0.00	12.00	100.00	100.00	0.00	0.00	9.80	100.00	0.00	0.00	0.00	0.00
	$n_j = 10$	25.80	100.00	100.00	100.00	0.00	0.00	100.00	100.00	0.00	0.00	0.20	5.00
	$n_j = 20$	99.60	100.00	100.00	100.00	0.00	6.00	100.00	100.00	1.60	0.40	8.80	89.60
L2N	$n_j = 5$	7.20	99.40	100.00	100.00	0.20	0.20	98.60	100.00	0.20	0.20	0.20	2.00
	$n_j = 10$	73.60	100.00	100.00	100.00	1.40	3.00	100.00	100.00	0.40	1.00	2.60	37.00
	$n_j = 20$	100.00	100.00	100.00	100.00	3.40	19.40	100.00	100.00	3.40	2.00	18.60	97.40
L2B	$n_j = 5$	56.00	100.00	100.00	100.00	6.40	10.20	100.00	100.00	6.60	9.40	11.40	31.20
	$n_j = 10$	93.00	100.00	100.00	100.00	7.80	17.60	100.00	100.00	4.80	6.80	18.60	72.40
	$n_j = 20$	100.00	100.00	100.00	100.00	9.40	33.60	100.00	100.00	7.20	7.40	33.40	99.20
L2b	$n_j = 5$	0.40	88.80	100.00	100.00	0.00	0.00	84.00	100.00	0.00	0.00	0.00	0.00
	$n_j = 10$	29.00	100.00	100.00	100.00	0.00	0.00	100.00	100.00	0.00	0.00	0.20	5.40
	$n_j = 20$	99.40	100.00	100.00	100.00	0.00	3.20	100.00	100.00	0.40	0.00	4.80	83.80
FN	$n_j = 5$	4.20	98.60	100.00	100.00	0.20	0.00	97.60	100.00	0.00	0.20	0.00	0.80
	$n_j = 10$	68.60	100.00	100.00	100.00	1.40	1.80	100.00	100.00	0.40	0.60	2.40	33.20
	$n_j = 20$	100.00	100.00	100.00	100.00	2.80	18.40	100.00	100.00	3.20	1.60	17.80	97.00
FB	$n_j = 5$	48.20	100.00	100.00	100.00	3.80	8.00	100.00	100.00	4.80	6.00	8.40	25.40
	$n_j = 10$	91.40	100.00	100.00	100.00	7.20	16.20	100.00	100.00	3.40	5.80	16.80	69.60
	$n_j = 20$	100.00	100.00	100.00	100.00	8.60	32.00	100.00	100.00	6.60	7.00	32.00	99.20
Fb	$n_j = 5$	0.00	0.00	100.00	100.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
	$n_j = 10$	0.60	100.00	100.00	100.00	0.00	0.00	100.000	100.00	0.00	0.00	0.00	0.00
	$n_j = 20$	98.60	100.00	100.00	100.00	0.00	0.80	100.00	100.00	0.00	0.00	0.40	56.00
GPF	$n_j = 5$	11.80	99.60	100.00	100.00	0.40	1.00	99.60	100.00	0.40	0.80	1.00	6.00
	$n_j = 10$	76.60	100.00	100.00	100.00	2.60	3.60	100.00	100.00	0.80	1.60	4.80	41.60
	$n_j = 20$	100.00	100.00	100.00	100.00	3.40	21.00	100.00	100.00	4.40	2.20	20.20	97.40
Fmaxb	$n_j = 5.00$	9.00	44.00	100.00	100.00	3.20	5.40	37.80	100.00	5.00	5.20	4.60	8.00
	$n_j = 10$	25.80	97.40	100.00	100.00	6.80	6.40	94.80	100.00	4.40	5.80	5.60	12.60
	$n_j = 20$	64.00	100.00	100.00	100.00	6.00	9.60	100.00	100.00	4.40	6.20	10.00	32.80
L2recon	$n_j = 5$	97.20	100.00	100.00	100.00	6.20	17.20	100.00	100.00	4.60	6.40	19.20	75.60
	$n_j = 10$	100.00	100.00	100.00	100.00	13.40	43.20	100.00	100.00	4.00	6.20	47.40	99.20
	$n_j = 20$	100.00	100.00	100.00	100.00	20.00	80.60	100.00	100.00	7.20	7.20	80.80	100.00

Table 4.4: Empirical power (in %) of each EMF-test under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with various variance parameter and shift parameter settings, $B = 500, K = 5$.

$\rho = 0.1$			$\rho = 0.5$			$\rho = 0.7$			$\rho = 0.9$		
$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 2$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 2$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 2$
$n_1 = 5$	5.00	4.40	31.80	100.00	4.20	7.60	80.20	4.20	4.80	6.60	56.40
$n_1 = 10$	5.20	7.20	65.80	100.00	5.00	13.20	99.80	5.20	4.00	8.40	92.20
$n_1 = 20$	4.20	9.00	97.80	100.00	3.60	27.60	100.00	4.60	7.40	18.20	100.00
$n_1 = 5$	0.60	1.00	11.00	100.00	0.40	1.00	44.00	0.20	0.20	0.80	15.60
$n_1 = 10$	3.40	3.80	54.00	100.00	2.80	1.20	98.60	1.60	1.40	3.00	85.60
$n_1 = 20$	3.00	7.40	96.80	100.00	2.60	4.20	100.00	1.60	3.00	14.40	100.00
$n_1 = 5$	9.40	7.60	41.60	100.00	5.80	11.00	87.40	4.80	5.20	8.40	60.80
$n_1 = 10$	6.60	9.40	71.00	100.00	5.60	15.20	99.80	5.60	4.20	9.40	93.00
$n_1 = 20$	5.20	10.00	98.00	100.00	4.20	6.60	100.00	5.40	7.20	19.40	100.00
$n_1 = 5$	11.80	9.20	46.60	100.00	8.40	13.40	90.20	7.80	8.20	10.80	68.40
$n_1 = 10$	8.40	10.80	74.00	100.00	7.00	16.60	100.00	6.80	5.80	10.80	94.20
$n_1 = 20$	5.60	10.40	98.20	100.00	4.80	31.20	100.00	5.60	8.00	20.60	100.00
$n_1 = 5$	10.00	8.20	41.60	100.00	5.40	9.40	84.60	3.60	3.80	5.80	55.00
$n_1 = 10$	5.80	8.40	68.80	100.00	4.60	13.60	99.60	3.60	2.80	6.20	91.40
$n_1 = 20$	4.80	9.80	98.00	100.00	3.40	27.00	100.00	2.80	5.00	16.20	100.00
$n_1 = 5$	5.20	4.80	32.00	100.00	4.40	8.00	81.20	3.00	3.00	5.00	54.80
$n_1 = 10$	5.40	7.40	66.00	100.00	5.40	14.00	99.60	4.40	3.80	7.60	92.60
$n_1 = 20$	4.20	9.00	97.80	100.00	3.80	27.80	100.00	4.00	6.60	18.60	100.00
$n_1 = 5$	5.40	5.40	32.80	100.00	4.80	9.60	84.40	4.60	5.00	8.20	60.40
$n_1 = 10$	5.40	7.40	66.00	100.00	5.40	14.60	99.80	5.60	4.20	9.00	93.00
$n_1 = 20$	4.20	9.20	97.80	100.00	4.00	28.20	100.00	5.20	7.20	19.20	100.00
$n_1 = 5$	0.80	1.80	11.40	100.00	0.40	1.20	43.20	0.20	0.00	0.00	11.20
$n_1 = 10$	3.60	4.00	53.40	100.00	2.60	5.80	98.00	1.40	0.80	2.20	82.80
$n_1 = 20$	2.60	7.40	96.80	100.00	2.60	21.60	100.00	1.20	2.40	11.80	100.00
$n_1 = 5$	9.20	7.20	40.60	100.00	8.20	12.40	87.40	9.20	7.80	11.60	64.60
$n_1 = 10$	6.60	9.00	69.80	100.00	7.40	16.00	100.00	6.80	5.60	10.20	93.60
$n_1 = 20$	5.00	10.00	98.00	100.00	4.60	31.00	100.00	5.00	8.00	21.00	100.00
$n_1 = 5$	4.40	5.00	42.80	100.00	5.20	3.20	75.40	4.00	4.40	5.40	35.80
$n_1 = 10$	5.60	7.00	88.20	100.00	6.20	4.40	99.60	5.40	4.00	7.00	80.40
$n_1 = 20$	4.60	10.80	100.00	100.00	4.80	27.00	100.00	5.20	6.20	14.20	99.80
$n_1 = 5$	5.20	4.60	31.00	100.00	4.40	8.00	79.60	4.40	4.80	6.60	55.00
$n_1 = 10$	5.20	7.60	65.00	100.00	5.40	13.40	99.40	5.20	4.00	8.20	91.60
$n_1 = 20$	4.20	9.00	97.00	100.00	3.60	27.40	100.00	5.00	6.80	18.40	100.00

Table 4.5: Empirical size ($\delta = 0$) and empirical power ($\delta > 0$) of EMFT under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with 500 MC simulations, $K = 5$ and $D = 70$.

		L2recon						FP			L2B	GPF	Fmaxb	FB
		M = 1	M = 5	M = 15	M = 50	M = 150	M = 1	M = 5	M = 15	M = 50	M = 150			
$\sigma = 0.2$	$D = 5$	$\delta = 0.01$	7.00	8.20	5.40	5.60	3.40	7.40	5.80	5.40	6.80	7.20	7.60	7.00
		$\delta = 0.1$	86.80	88.80	88.60	89.00	87.00	89.20	89.20	91.20	89.40	83.20	84.80	83.60
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 15$	$\delta = 0.01$	16.00	19.60	19.00	18.60	13.20	15.20	12.60	12.20	15.20	13.60	14.40	12.20
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 70$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\sigma = 5$	$D = 5$	$\delta = 0.01$	4.40	5.40	6.00	4.00	3.20	5.40	6.00	5.20	3.80	7.20	7.80	6.80
		$\delta = 0.1$	88.40	90.20	91.40	86.60	88.40	89.00	90.00	87.00	90.00	85.60	86.80	79.20
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 15$	$\delta = 0.01$	17.20	16.80	17.20	18.80	17.20	13.80	13.80	14.00	12.60	14.40	14.00	14.40
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 70$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\sigma = 100$	$D = 5$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 15$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 70$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\sigma = 100$	$D = 5$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 15$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
	$D = 70$	$\delta = 0.01$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.1$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		$\delta = 0.5$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 4.6: Empirical powers (in %) after 500 Monte Carlo simulations, hypothesis model $\mathcal{F}_3$ and error model $\mathcal{E}_1$ with error terms having standard normal distribution.

4.2 Simulations for equality of covariance functions tests

In this section we will investigate the behaviour of equality covariance function tests (ECFT) described in Chapter 3, section 3.3. The simulation settings is similar to EMFT, where the only change is in the error model. For simulations in this section we will use error model $\mathcal{E}_3$, where parameter ω represent the amount of change in the covariance function. Illustrative example of covariance function under H_0 and function under alternative can be found on Figure 4.4. Again, we will consider two scenarios

- **Scenario I:** We will assume, that the functional groups has equal mean functions, and we will shift only the covariance functions,
- **Scenario II:** In the second case, we will assume that the functional groups has different mean function across groups and therefore, we will investigate the behaviour of tests under the mean function shift.

The simulation settings can be found in Table 4.7.

	Scenario I	Scenario II
Number of groups	$K = 5$	$K = 5$
Time grid size	$D = 70$	$D = 70$
The variance/correlation parameter	$\rho \in \{0.1, 0.3, 0.6, 0.9\}$	$\rho \in \{0.1, 0.3, 0.6, 0.9\}$
Number of MC simulations	$B = 500$	$B = 500$
The number of permutations:	$N = 1000$	$N = 1000$
The size of the discrepancy between covariance functions	$\omega \in \{0, 0.01, 0.1, 1, 3, 8\}$	$\omega \in \{0, 0.01, 0.1, 1, 3, 8\}$
Number of samples per group	$n_j \in \{5, 15, 35\}$	$n_j \in \{5, 15, 35\}$
The mean function shift parameter	No shift parameter	$\delta \in \{0.2, 0.5\}$

Table 4.7: Hyperparameters settings for simulation study under validity of H_0 .

4.2.1 Empirical size

In Table 4.8 there are shown results from the first case simulations, where the empirical sizes (in %) are under $\omega = 0$. We can see that permutation test is quite resistant against the increasing within-function correlation and the empirical size is always approximately between 4.8 - 6 %. On the other hand, the remaining version of $\mathcal{L}^2$ -norm based test has problem with increasing correlation to reach the empirical level. However, it is fixed with increasing sample size. Furthermore, the unequal mean function has negligible effect on the reaching of empirical size for bias-reduced and permutation versions of $\mathcal{L}^2$ -norm based test. On the other hand, with increasing within-function correlation and considering unequal mean functions between functional samples, naive version of $\mathcal{L}^2$ -norm based test has problem to reach the significance level.

4.2.2 Empirical power

Let us start first with **Scenario I**, where empirical power can be found in Table 4.8 ($\omega > 0$). We can see that the power is increasing function of sample size and the distance of null hypothesis space (increasing ω). Furthermore, it seems that the tests

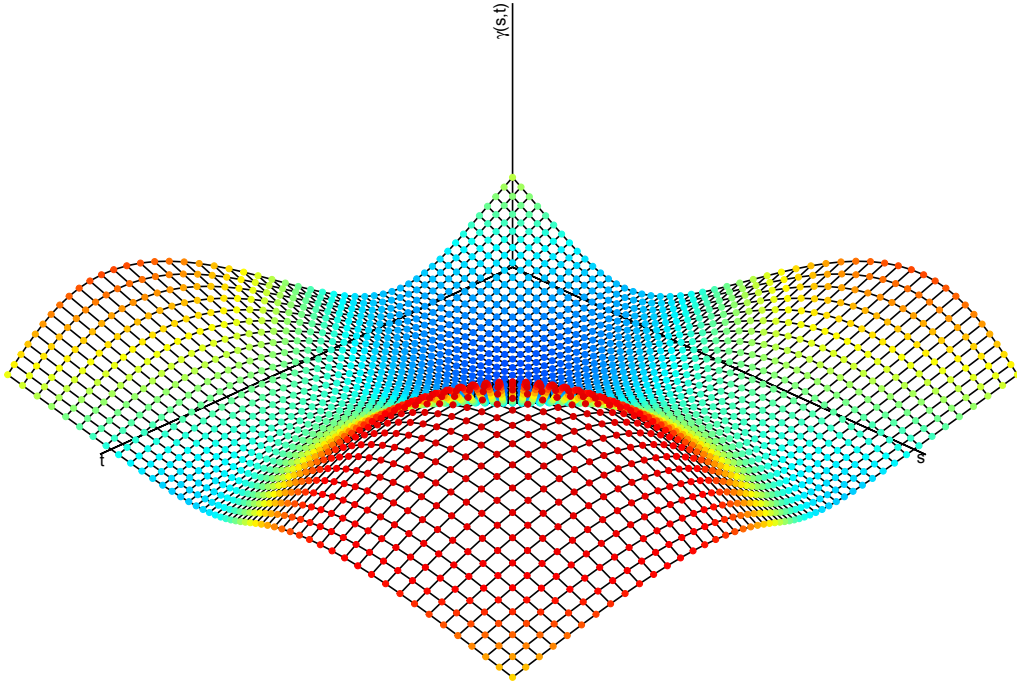


Figure 4.3: Covariance function under H_0 ($\omega = 0$) for the first group, $j = 1$.

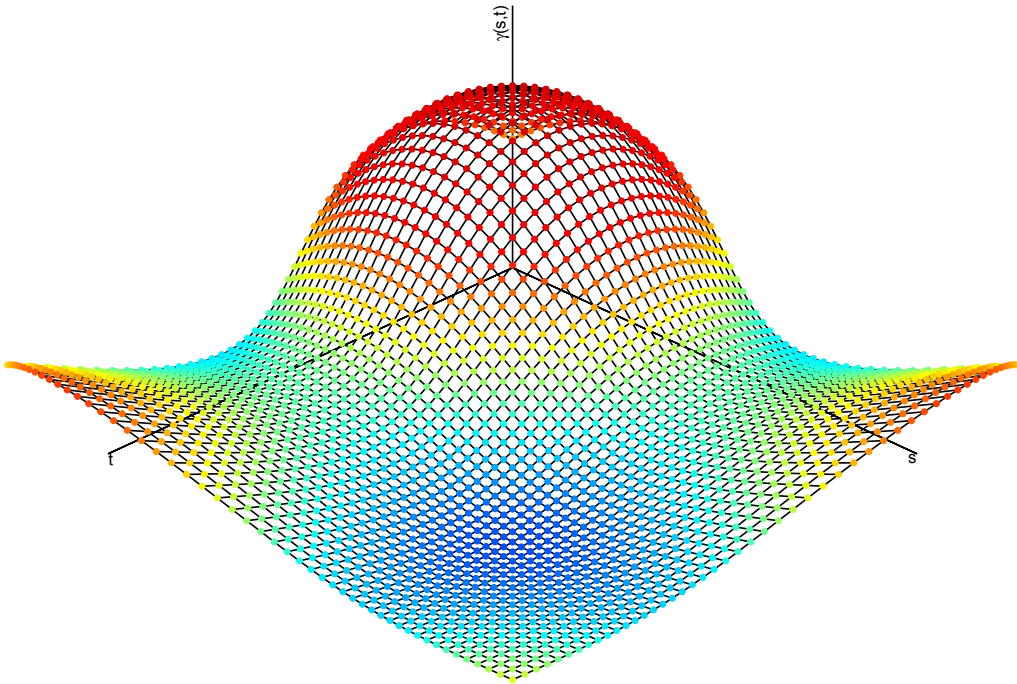


Figure 4.4: Covariance function under H_1 ($\omega = 3$) for the third group, $j = 3$.

using naive and biased reduction method of estimation outperforms the test based on permutation, even if the within-function correlation is high ($\rho = 0.1$) or moderate/low ($\rho = 0.6$ or $\rho = 0.9$). Furthermore, we can see, that if the sample sizes across groups are small, permutation test has poor performance for lower ρ ($n_j = 5, \rho = 0.1$). Compared to L2Pecf, L2Becf and L2Nect tests have the ability to reach quite high power even the sample sizes are small (for $n_j = 5, \omega = 8$, L2Becf has power 74% and L2Nect has power 70.8%), no matter what is the amount of within-function correlation. Therefore, the most appropriate method, with the assumption of equal mean function and lower sample sizes is $\mathcal{L}^2$ -norm based test, which can be used on highly or low correlated data. In total. On the other hand, on functional samples with bigger sample sizes, all test are feasible.

Let us focus on **Scenario II**, where results can be found in Table 4.9. We can see, that the performance remains almost unchanged. We can see, that with the assumption of different mean functions, the empirical powers are little bit higher, even if the sample sizes per group are small. The reason behind the power increase due to different mean functions is in the way how pooled sample covariance functions is computed. The empirical covariance function is the function of data and their empirical mean functions, and if they are different, also the covariance functions can be different. Same as in the previous scenario, both L2Becf and L2Necf are feasibly or low and high correlated functional data, with small sample sizes.

	ω	L2Becf			L2Necf			L2Pecf		
		$n_1 = 5$	$n_1 = 15$	$n_1 = 35$	$n_{1,2,3} = 5$	$n_1 = 15$	$n_1 = 35$	$n_1 = 5$	$n_1 = 15$	$n_1 = 35$
$\rho = 0.1$	0.00	4.00	4.60	4.00	2.60	4.00	3.60	3.60	5.00	5.00
	0.01	4.00	4.60	3.80	2.80	4.00	3.60	3.80	5.00	5.00
	0.10	4.20	5.00	4.00	2.80	4.60	3.60	4.40	5.60	4.60
	1.00	12.60	40.00	80.60	10.40	38.80	80.40	11.40	36.40	76.00
	3.00	47.40	98.60	100.00	43.60	98.40	100.00	31.40	91.40	100.00
	8.00	70.60	100.00	100.00	65.60	100.00	100.00	39.20	96.80	100.00
$\rho = 0.3$	0.00	4.00	5.00	3.60	1.80	4.00	3.40	3.40	5.60	4.40
	0.01	4.00	4.80	3.40	2.00	4.40	3.40	3.60	5.80	4.60
	0.10	4.20	5.60	5.80	3.60	4.80	5.80	4.60	6.60	6.40
	1.00	25.40	84.60	100.00	21.00	83.80	100.00	20.80	72.80	99.80
	3.00	62.00	99.80	100.00	56.80	99.80	100.00	36.80	95.80	100.00
	8.00	73.20	100.00	100.00	70.20	100.00	100.00	42.00	96.60	100.00
$\rho = 0.6$	0.00	4.40	6.20	3.40	1.60	4.40	3.40	4.40	5.40	4.80
	0.01	4.40	6.40	4.40	1.60	4.40	3.80	4.20	5.00	4.60
	0.10	4.60	8.60	9.20	2.20	6.60	8.40	6.00	8.60	10.40
	1.00	39.00	96.40	100.00	33.80	96.00	100.00	26.60	86.40	100.00
	3.00	69.80	100.00	100.00	63.20	100.00	100.00	38.40	96.60	100.00
	8.00	73.40	100.00	100.00	71.00	100.00	100.00	42.40	96.80	100.00
$\rho = 0.9$	0.00	4.20	5.40	4.00	0.60	2.80	2.80	5.60	4.80	3.40
	0.01	4.20	5.80	4.60	0.60	3.20	2.80	5.60	4.80	3.60
	0.10	4.40	8.20	8.60	0.80	4.20	6.60	5.40	7.80	8.40
	1.00	44.40	98.60	100.00	35.60	98.40	100.00	28.60	90.40	100.00
	3.00	70.00	100.00	100.00	66.40	100.00	100.00	39.80	96.60	100.00
	8.00	74.00	100.00	100.00	70.80	100.00	100.00	41.60	96.80	100.00

Table 4.8: Empirical size ($\omega = 0$) and empirical power ($\omega > 0$) for fixed time grid $D = 70$.

	ω	L2BEcf				L2NEcf				L2PEcf			
		$\delta = 0.2$		$\delta = 0.5$		$\delta = 0.2$		$\delta = 0.5$		$\delta = 0.2$		$\delta = 0.5$	
		$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$
$\rho = 0.1$	0.00	4.00	5.00	4.00	5.00	3.20	4.80	3.20	4.80	3.40	4.60	2.00	2.20
	0.01	4.20	5.00	4.20	5.00	3.20	4.60	3.20	4.60	3.20	4.40	2.00	2.20
	0.10	4.40	4.60	4.40	4.60	3.60	4.60	3.60	4.60	3.60	5.00	1.80	1.80
	1.00	23.60	61.60	23.60	61.60	21.60	61.00	21.60	61.00	17.80	55.00	11.20	44.00
	3.00	92.40	100.00	92.40	100.00	91.80	100.00	91.80	100.00	73.40	100.00	68.60	100.00
	8.00	99.60	100.00	99.60	100.00	99.60	100.00	99.60	100.00	85.20	100.00	84.80	100.00
$\rho = 0.5$	0	3.40	4.00	3.40	4.00	2.40	3.80	2.40	3.80	4.60	4.80	3.20	3.40
	0.01	3.20	4.40	3.20	4.40	2.40	3.60	2.40	3.60	4.20	4.80	3.00	3.40
	0.10	4.00	6.60	4.00	6.60	3.20	6.20	3.20	6.20	5.00	7.00	4.40	5.60
	1.00	82.80	100.00	82.80	100.00	81.00	100.00	81.00	100.00	64.20	99.80	62.60	99.60
	3.00	99.40	100.00	99.40	100.00	99.40	100.00	99.40	100.00	83.80	100.00	84.40	100.00
	8.00	99.60	100.00	99.60	100.00	99.60	100.00	99.60	100.00	85.80	100.00	86.20	100.00
$\rho = 0.9$	0.00	4.20	4.20	4.20	4.20	1.60	2.80	1.60	2.80	4.20	4.40	3.80	4.20
	0.01	4.20	4.00	4.20	4.00	1.60	2.80	1.60	2.80	4.20	4.40	4.00	4.00
	0.10	5.00	8.00	5.00	8.00	2.00	4.20	2.00	4.20	6.00	6.40	5.60	5.60
	1.00	91.40	100.00	91.40	100.00	89.60	100.00	89.60	100.00	72.60	100.00	71.00	100.00
	3.00	99.60	100.00	99.60	100.00	99.60	100.00	99.60	100.00	84.60	100.00	84.40	100.00
	8.00	99.60	100.00	99.60	100.00	99.60	100.00	99.60	100.00	86.40	100.00	86.40	100.00

Table 4.9: Empirical size ($\omega = 0$) and empirical power, with fixed grid size $D = 50$, after 500 simulations applied on artificial data generated under model $\mathcal{M}_2$ and error model $\mathcal{E}_3$

4.3 Simulations for equality of distribution tests

In this section, we will compare the performance of tests CF and CFRecon. Hlávka et al. [2021] have shown, that the performance of CF depends on the choice of the matrix Σ . It has turned out, that CF has best performance using $\Sigma = \{\gamma(s, t)\}_{s, t \in \mathcal{T}} = \frac{1}{n} \sum_{k=1}^K \sum_{j=1}^{n_k} (X_{kj}(t) - \bar{X}_k(t))(X_{kj}(s) - \bar{X}_k(s))$. Consequently, from the all possible choices of matrix Σ , we will consider only this option. However, CFRecon depends on the identity matrix, it will be reasonable also taking into consideration same option for CF. Through this section, we denote this options as **identity** and **PSCM** (inverse of pooled sample covariance matrix). We will conduct the similar simulations to the previous sections. The data will be generated under the hypothesis model $\mathcal{F}_2$ (validity of H_0), hypothesis model $\mathcal{F}_4$ (validity of H_1) and the error model $\mathcal{E}_2$, where error terms has standard normal distribution. The settings of the other parameters are the following

- Grid size: $D \in \{5, 10, 30, 70\}$,
- Number of the groups: $K = 2$
- Sample sizes per groups: $n_j \in \{5, 15, 25, j = 1, 2\}$,
- Number of the orthonormal basis functions: $M \in \{10, 30, 70, 150\}$,
- Amount of the within-function correlation: $\rho \in \{0.1, 0.9\}$.
- Shift parameter: $\delta \in \{0.2, 0.9\}$.

4.3.1 Empirical levels

In this section, we will investigate only the behaviour of CFRecon under validity of H_0 . The simulations for investigation of empirical sizes of CF, see Hlávka et al. [2021], Section 5.

In Table 4.10 can be found empirical levels for CFRecon. There are no obvious patterns in the results. It can be seen, that the empirical sizes are not a function of the grid size D or dimension of basis expansion M and it is sufficient, for reaching the pre-set significance level, choose the lower M .

4.3.2 Empirical powers

In Table 4.11 can be found empirical powers of both CF test and CFRecon test. As was expected, CF using PSCM has outperformed CF using the identity matrix. It can be seen that with the highly correlated data, CF is able to reach 100% power, even with small sample size and relative small time grid $D = 10$. If the data are low correlated $\rho = 0.9$, empirical power is increasing function of time grid D and sample sizes. However, interest results can be seen for sparse data ($D \in \{5, 10\}$) and low correlated data ($\rho = 0.9$). We can see, that if the data are sparse, with increasing sample sizes, the power of CF is decreasing. Fortunately, this behaviour has some possible explanation. Hlávka et al. [2021] in Chapter 4, Section 4.1, describe how to estimate inverse of the matrices, where they divide into two cases: $D > \sum_{j=1}^2 n_j$ and $D \leq \sum_{j=1}^2 n_j$. Unfortunately, we only followed the second option in our simulations. However, the latter should be one of the possible explanation, not the final one.

In the same table, there can be found empirical powers for CFRecon. There can be seen two patterns. It seems that in the case of high correlated data ($\rho = 0.1$), sparse

		$\rho = 0.1$			$\rho = 0.9$		
		$D = 10$	$D = 30$	$D = 70$	$D = 10$	$D = 30$	$D = 70$
$M = 10$	$n_j = 5$	5.80	5.40	6.60	5.60	6.20	6.60
	$n_j = 10$	7.20	6.40	5.20	6.60	5.20	4.20
	$n_j = 15$	4.60	4.80	6.20	6.80	5.80	6.60
$M = 30$	$n_j = 5$	6.00	5.00	6.40	5.80	6.20	6.80
	$n_j = 10$	7.20	6.40	5.20	6.40	5.20	4.40
	$n_j = 15$	4.60	4.80	6.20	6.80	5.80	6.60
$M = 70$	$n_j = 5$	6.00	5.20	6.60	5.40	6.00	7.00
	$n_j = 10$	7.20	6.40	5.20	6.40	5.20	4.40
	$n_j = 15$	4.60	4.80	6.20	6.80	5.80	6.60
$M = 150$	$n_j = 5$	5.80	4.80	6.60	5.20	6.20	6.40
	$n_j = 10$	7.20	6.40	5.20	6.40	5.20	4.40
	$n_j = 15$	4.60	4.80	6.20	6.80	5.80	6.60

Table 4.10: Empirical sizes (in %) of CFRecon after 500 Monte Carlo simulations, under hypothesis model $\mathcal{F}_2$ and error model $\mathcal{E}_2$, with error models having standard normal distribution.

data ($D = 5$) and using small number of orthonormal basis functions ($M = 10$), CF is more powerful than CFRecon. However, with increasing number of orthonormal basis functions, tests are comparable. For the low correlated data ($\rho = 0.1$), it seems that under small grid size ($D = 5$), CFRecon outperformed CF, even with smaller number of orthonormal basis functions ($M = 10$). With increasing time grid size, CF seems to have desired power. Although, we need to have in mind, that we are not sure about the reason of strange behaviour of CF test under small grid size. Consequently, in this case, we need to be more carefully about the final conclusions.

Comparing to the EMFT, empirical power of CFRecon is a function of grid size D and basis dimension M . For example, for $\rho = 0.1, \delta = 0.1$ Comparing to EMFT, empirical power of CF is the function of grid size D (or the sparsity level) and dimension of basis expansion M . We can see, that for $\rho = 0.1$, and increasing grid size D (or the data are more dense), for reaching the desired power, it is needed to use big number of orthonormal basis functions (more than 70). In addition, if the data are dense ($D = 70$), even 150 orthonormal basis functions were not enough to reach the desired empirical power (the reached power was only around 43 %). Even the significant difference in the mean functions $\delta = 0.9$ did not improve the performance. Further, the same pattern can be seen under $\rho = 0.9$. However, the more positive results are present. We can see that if data are dense ($D > 50$), although we had to use the large number of orthonormal basis functions ($M > 70$), but the desired 100% empirical power is achievable. From what was previously mentioned, it can be possible, that properties of CFRecon are affected by ρ . For the higher values of ρ , CFRecon has the ability to produce reasonable results, under sufficient large number of basis functions. Conversely, test is suitable if the data are sparse. However, these results are not convincing enough to rank this test first over commonly used tests.

We have shown, that there exists a possibility, that CF can reach the desired power, if

we choose large number of orthonormal basis functions. Unfortunately, this leads to very computational inefficient test. For example, if we want to apply this test, to test (3.62), for two samples, we need to estimate the unknown coefficients for $2n_j$ observations, where the vector of unknown coefficients is the length of M . Additionally, estimating each unknown coefficient is same as estimating the unknown coefficients in linear regression model. Consequently, the estimation procedure is equivalent to fitting $2n_j, j = 1, \dots, K$ regression models, which can lead, even with sparse data, to computationally inefficient test.

		$\rho = 0.1$										$\rho = 0.9$									
		$\delta=0.2$					$\delta=0.9$					$\delta=0.1$					$\delta=0.9$				
		$D = 5$	$D = 10$	$D = 30$	$D = 70$	$D = 100$	$D = 5$	$D = 10$	$D = 30$	$D = 70$	$D = 100$	$D = 5$	$D = 10$	$D = 30$	$D = 70$	$D = 100$	$D = 5$	$D = 10$	$D = 30$	$D = 70$	$D = 100$
CF	$n_j = 5$	5.00	4.95	4.75	4.80	4.85	4.85	4.75	6.05	4.80	5.40	5.40	4.80	6.20	6.80	5.80	4.20	6.20	6.80	6.80	
	$n_j = 10$	5.60	5.00	5.75	4.80	5.00	5.00	5.40	5.50	4.80	5.60	5.60	6.20	5.00	6.00	5.70	5.25	5.00	6.00	6.00	
	$n_j = 15$	4.50	5.90	5.00	7.00	5.65	5.65	5.80	6.05	7.00	6.50	6.50	5.20	5.00	6.20	5.55	4.95	5.00	6.20	6.20	
	$n_j = 5$	80.75	100.00	100.00	100.00	81.80	100.00	100.00	100.00	100.00	45.15	93.35	93.35	100.00	100.00	47.75	93.85	100.00	100.00	100.00	
	$n_j = 10$	85.90	100.00	100.00	100.00	85.40	100.00	100.00	100.00	100.00	19.10	63.35	63.35	100.00	100.00	18.60	64.40	100.00	100.00	100.00	
	$n_j = 15$	94.75	100.00	100.00	100.00	94.40	100.00	100.00	100.00	100.00	12.95	38.25	38.25	100.00	100.00	13.15	38.75	100.00	100.00	100.00	
CFRecon	$M = 10$	$n_j = 5$	47.60	7.20	0.00	0.00	44.40	11.00	0.00	0.00	72.60	10.80	10.80	0.00	0.00	76.40	11.80	0.00	0.00	0.00	
		$n_j = 10$	42.40	7.80	0.00	0.00	42.40	5.80	0.00	0.00	73.40	12.80	12.80	0.00	0.00	70.80	11.00	0.00	0.00	0.00	
		$n_j = 15$	40.00	8.00	0.00	0.00	41.40	7.20	0.00	0.00	74.00	9.00	9.00	0.00	0.00	70.20	8.40	0.00	0.00	0.00	
	$M = 30$	$n_j = 5$	90.00	66.20	8.60	0.00	90.40	66.80	7.80	0.00	100.00	99.80	99.80	8.40	0.00	100.00	99.60	8.40	0.00	0.00	
		$n_j = 10$	89.60	62.60	5.80	0.00	88.40	66.00	8.00	0.00	100.00	99.80	99.80	6.40	0.00	100.00	98.80	6.40	0.00	0.00	
		$n_j = 15$	90.80	66.00	5.40	0.00	86.20	63.80	7.20	0.00	100.00	99.60	99.60	8.40	0.00	100.00	99.60	8.40	0.00	0.00	
	$M = 70$	$n_j = 5$	99.40	91.40	55.00	5.40	98.80	93.20	55.60	5.40	100.00	100.00	100.00	96.40	5.40	100.00	100.00	96.40	5.40	5.40	
		$n_j = 10$	99.20	93.00	53.20	5.60	99.40	93.40	52.80	5.60	100.00	100.00	100.00	96.20	8.80	100.00	100.00	96.20	8.80	8.80	
		$n_j = 15$	99.40	93.00	48.60	4.60	99.80	94.20	53.80	4.60	100.00	100.00	100.00	96.40	5.60	100.00	100.00	96.40	5.60	5.60	
	$M = 150$	$n_j = 5$	100.00	99.80	86.00	48.40	100.00	99.60	90.00	48.40	100.00	100.00	100.00	100.00	95.60	100.00	100.00	100.00	95.60	95.60	
		$n_j = 10$	100.00	100.00	88.80	48.60	100.00	99.80	89.40	48.60	100.00	100.00	100.00	100.00	94.40	100.00	100.00	100.00	94.40	94.40	
		$n_j = 15$	100.00	99.80	90.60	44.60	100.00	99.80	91.00	44.60	100.00	100.00	100.00	100.00	93.80	100.00	100.00	100.00	93.80	93.80	

Table 4.11: Empirical power (in %) of CF and CFRecon tests after 500 Monte Carlo simulations, hypothesis model $\mathcal{F}_4$ and error model $\mathcal{E}_2$.

Conclusion

In the thesis, we dealt with functional ANOVA problem (FANOVA). Firstly, in Chapter 1, we introduced all necessary theory for the statistical inference of functional data, including the statistical model of the functional data, estimators of mean and covariance functions, and its properties.

In Chapter 2, we described the problem of reconstructing the observed functional data. Next, in Section 2.1, we described the linear smoothing method using a smoothing matrix of orthonormal basis functions for reconstructing the functional data. The last section serves as a review of the well-known Kernel smoothing for reconstruction of the observed functional data.

Chapter 3 was the core part of the thesis. We have reviewed all the well-known methods for FANOVA, including mean functions (Section 3.2) and covariance functions (Section 3.3) equality tests and tests for equality of distributions (Section 3.4). In Section 3.2.1, we introduced $\mathcal{L}^2$ -norm based tests, which was mainly discussed in Cuesta-Albertos and Febrero-Bande [2010] and Zhang [2012]. Firstly, we introduced $\mathcal{L}^2$ -norm based test based on parametric or non-parametric bootstrap, where we assume homoscedastic functional data. Section 3.2.2 was mainly devoted to describing the approximation of a distribution of the test statistic $T_n \equiv SSR$. The most common techniques are bias-reduction or naive methods of approximation of a distribution of the test statistic T_n . Furthermore, in Section 3.2.3 we tried to use functional data as a linear combination of M orthonormal basis function in a $\mathcal{L}^2$ -norm based test. This led to a new test, based only on the unknown coefficient, whose estimator was introduced in Chapter 2. Such test was evaluated with the help of the permutation test. The remaining Sections 3.2.4 - 3.2.5 dealt with F -type equality mean functions tests. This class of tests use the well-known ANOVA point-wise F -statistic. Unfortunately, as was shown across the literature, such point-wise test is not feasible in the field of FDA. In the mentioned sections, we introduced tests and methods, which can deal with the aforementioned problem. First, the statistic using both global SSR and SSE was introduced. We have shown, that under the Gaussian assumption, we can approximate such statistic by some scaled F -distribution with d_1 and d_2 , where the scale parameter and both degrees of freedom can be estimated by the naive and bias-reduced method of the estimation. Also, the additional global F -type tests were introduced. Namely *Globalized Point-wise F -type* test and *Maximal- F -type* test. The remaining section of mean function equality tests was devoted to F -type test based on a linear combination of M -orthonormal basis functions, which was proposed by Gorecki and Smaga [2015]. Similarly to the case of $\mathcal{L}^2$ -norm based test, the resulting test statistic depends only on the unknown coefficients. Furthermore, we have discussed that using a basis representation is not feasible for the remaining global tests.

Tests for equality of covariance functions were discussed in Section 3.3. From the all known tests, we introduced only one class of equality of covariance functions tests, the class of $\mathcal{L}^2$ -norm based tests, which are described in Zhang [2012]. Those tests are mainly based on the $\mathcal{L}^2$ -distance between group covariance function $\gamma_i(s, t)$ and the overall covariance function $\gamma(s, t)$, which represents, in some sense, the average population covariance function. Again, as in the previous sections, we have assumed that our functional data come from some a Gaussian process, where each group has a different mean function. At the end of this section, we introduced a test based on the orthonormal basis functions, which was again evaluated with the help of the permutation test.

The last section of Chapter 3 dealt with a test for equality of distribution of two stochastic processes proposed by Hlávka et al. [2021], which is based on the characteristic functional of a stochastic process. The aim of this test was to construct a dual testing problem to equality of mean functions test. However, if this dual problem should be valid, we had to made the assumption of equality of covariance functions for each group. In contrast to the previous tests, we only consider two-sample problem. We have shown in the detail the idea behind the test and all necessary derivations of the final test statistic. Similarly to the previous tests, we also represent this test as a test using orthonormal basis functions. However, the evaluation of such test was not straightforward. We had to make a naive assumption of choosing a covariance operator to get a diagonal matrix of one's in the final test statistic.

Chapter 4 includes the results of various simulations. In Sections 4.1.1 – 4.1.4 we were comparing the properties of tests for equality mean functions. It was shown, that the bootstrapped-based tests CH, Fb performed badly. More specifically, with the increasing amount of variance, they were not able to detect changes in the mean functions, even with big sample sizes per group (they reach 56% and 80% power under $n_j = 25$). At the same time, globalized test GPF had the worst performance with a high amount of variance ($\sigma = 5$), where this test reaches only 32% power under big sample sizes. Tests based on the orthonormal basis functions FP, L2recon have shown very good results. When controlling the amount of variance, both of tests were able to reach the desired power. The same is valid for the case, when we control the size of the within-function correlation, where they were comparable with the remaining tests. In addition, it seems that L2recon outperformed FP. In final, we investigated, if both L2recon and FP are feasible for usage, and how many basis function we need to get desired results. It was shown that in the case of EMFT tests, the empirical power (or size) does not depend on the number of basis functions. Furthermore, it was shown that under the Gaussian assumption, tests using naive and bias-reduced method of estimation (both $\mathcal{L}^2$ -norm based tests and F -type tests) outperformed tests using orthonormal basis functions, in the terms of empirical power and computation efficiency, even when the artificially generated data was sparse ($M = 5$). Consequently, it was shown, that there is no significant advantage of using the tests L2recon and FP instead of usual tests for FANOVA.

Section 4.2 was devoted to investigation of properties of tests for equality of covariance functions. We have shown, that the permutation based test is not appropriate.

Section 4.3 was devoted to comparing CF, proposed by Hlávka et al. [2021], and CFRecon, test for equality of distributions using orthonormal basis expansion. It was shown, that if we observed highly-correlated data ($\rho = 0.1$), evaluated in a few time points ($M = 5$), for a sufficiently high number of basis functions (150 basis functions), CFRecon has no problem to reach 100% and outperformed CF. On the other side, it seems that the empirical power of CFRecon depends on the grid size (or the level of sparsity) and the number of basis functions. More specifically, we need at least 70 basis functions included in the basis expansion. Additionally, one thing we need to keep in mind is that the approximation of functional data, with such large number of basis functions can be computationally expensive - we need to estimate the unknown coefficients for each functional observation, which can lead to $Kn_j, j = 1, \dots, K$ least square estimators. In final, same as in the EMFT tests, using tests based on basis expansion in testing the equality of distribution is not feasible due to high computing inefficiency and no added value.

Therefore, it turned out that tests using orthonormal basis functions as reconstruction method of functional data is not that bad idea, mostly for the highly correlated

data, or for sparse data. However, even if the data are sparse, using test based on the orthonormal basis functions are computational inefficient. In the end, however, if the data meet the Gaussian assumptions, one should use F -type tests and $\mathcal{L}^2$ -norm based tests, using bias-reduction method of estimation in FANOVA problem. These tests, even under small sample sizes, sparse data, high amount of error and various sizes of the within-correlation outperformed aforementioned tests based on basis expansion.

Bibliography

- Ana M. Aguilera, Christian Acal, M. Carmen Aguilera-Morillo, Francisco Jiménez-Molinos, and Juan B. Roldán. Homogeneity problem for basis expansion of functional data with applications to resistive memories. *Mathematics and Computers in Simulation*, 186:41–51, 2021. ISSN 0378-4754. doi: <https://doi.org/10.1016/j.matcom.2020.05.018>. URL <https://www.sciencedirect.com/science/article/pii/S0378475420301774>.
- N. T. Andersen and V. Dobric. The Central Limit Theorem for Stochastic Processes. *The Annals of Probability*, 15(1):164 – 177, 1987. doi: 10.1214/aop/1176992262. URL <https://doi.org/10.1214/aop/1176992262>.
- D. Bosq. *Linear Processes in Function Spaces*. Springer, Verlag New York, 2000. ISBN 978-1-4612-1154-9.
- Juan Cuesta-Albertos and Manuel Febrero-Bande. A simple multiway anova for functional data. *Test*, 19:537–557, 11 2010. doi: 10.1007/s11749-010-0185-3.
- Antonio Cuevas, Manuel Febrero-Bande, and Ricardo Fraiman. An anova test for functional data. *Computational Statistics & Data Analysis*, 47:111–122, 02 2004. doi: 10.1016/j.csda.2003.10.021.
- J. Faraway. Regression analysis for a functional response. *Technometrics*, 39:254–261, 1997.
- Manuel Febrero-Bande and Manuel Oviedo de la Fuente. Statistical computing in functional data analysis: The r package fda.usc. *Journal of Statistical Software, Articles*, 51(4):1–28, 2012. ISSN 1548-7660. doi: 10.18637/jss.v051.i04. URL <https://www.jstatsoft.org/v051/i04>.
- T. Gorecki and L. Smaga. A comparison of tests for the one-way anova problem for functional data. *Comput Stat*, 30:987–1010, 2015.
- Tomasz Górecki and Łukasz Smaga. fdanova: an r software package for analysis of variance for univariate and multivariate functional data. *Computational Statistics*, 34, 06 2019. doi: 10.1007/s00180-018-0842-7.
- Zděnek Hlávka, Daniel Hlubinka, and Kateřina Koňasová. Functional anova based on empirical characteristic functionals. *Article submitted for publication*, 2021.
- P. Kokoszka and L. Horvath. *Inference for Functional Data with Applications*. Springer, 2012. ISBN 978-1-4614-3654-6.
- A. Komárek. Linear regression, 2020. URL https://www2.karlin.mff.cuni.cz/~komarek/vyuka/2020_21/nmsa407/2020-NMSA407-notes.pdf.
- M. Omelka. Modern statistical methods, 2020. URL https://www2.karlin.mff.cuni.cz/~omelka/Soubory/nmst434/nmst434_course-notes.pdf.
- Silverman B.W Ramsey, J.O. *Functional Data Analysis*. Second Edition. Springer, New York, 2005. ISBN 978-0-387-22751-1.

- J. Zhang. *Analysis of Variance for Functional Data*. 2012.
- Jin-Ting Zhang. Approximate and asymptotic distributions of chi-squared-type mixtures with applications. *Journal of the American Statistical Association*, 100:273–285, 02 2005. doi: 10.1198/016214504000000575.
- Jin-Ting Zhang. Statistical inferences for linear models with functional responses. *Statistica Sinica*, 21(3):1431–1451, 2011. ISSN 10170405, 19968507. URL <http://www.jstor.org/stable/24309569>.
- Jin-Ting Zhang and Xuehua Liang. One-way anova for functional data via globalizing the pointwise f-test. *Scandinavian Journal of Statistics*, 41(1):51–71, 2014. doi: <https://doi.org/10.1111/sjos.12025>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1111/sjos.12025>.
- Jin-Ting Zhang, Ming-Yen Cheng, Hau-Tieng Wu, and Bu Zhou. A new test for functional one-way anova with applications to ischemic heart screening. *Computational Statistics & Data Analysis*, 132:3–17, 2019. ISSN 0167-9473. doi: <https://doi.org/10.1016/j.csda.2018.05.004>. URL <https://www.sciencedirect.com/science/article/pii/S0167947318301130>. Special Issue on Biostatistics.

List of Figures

1.1	The heights of ten boys measured at 31 not equidistant ages (the black lines). Its mean function is represented by green line.	5
1.2	The marginal bone levels (blue lines) for 50 patients measured at equidistant time points (every half-year). Its mean function is represented by the green line.	6
2.1	Observed heights of 10 boys from Example 1. Reconstruction of the data was done by B -splines.	17
2.2	The reconstructed observed heights of 10 boys by Fourier basis from Example 1	18
2.3	Reconstruction of the trend of the temperature for two canadian stations.	20
3.1	The evolution of the heights of ten boys and ten girls	21
3.2	Growth data	33
4.1	Artificial data generated under models $\mathcal{F}_1$ and $\mathcal{F}_2$ and error model $\mathcal{E}_1$	52
4.2	Artificial data generated under models $\mathcal{F}_2$ and error model $\mathcal{E}_2$	53
4.3	Covariance function under H_0 ($\omega = 0$) for the first group, $j = 1$	61
4.4	Covariance function under H_1 ($\omega = 3$) for the third group, $j = 3$	61

List of Tables

3.1	Result of the L^2 -norm based tests	34
4.1	Hyperparameters settings for simulation study under validity of H_0	51
4.2	Empirical levels in % after 500 Monte Carlo simulations, error model $\mathcal{E}_1$, $D = 70$, $K = 5$	54
4.3	Parameter settings for simulation study under validity of H_1	55
4.4	Empirical power (in %) of each EMF-test under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with various variance parameter and shift parameter settings, $B = 500$, $K = 5$	57
4.5	Empirical size ($\delta = 0$) and empirical power ($\delta > 0$) of EMFT under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with 500 MC simulations, $K = 5$ and $D = 70$	58
4.6	Empirical powers (in %) after 500 Monte Carlo simulations, hypothesis model $\mathcal{F}_3$ and error model $\mathcal{E}_1$ with error terms having standard normal distribution.	59
4.7	Hyperparameters settings for simulation study under validity of H_0	60
4.8	Empirical size ($\omega = 0$) and empirical power ($\omega > 0$) for fixed time grid $D = 70$	63
4.9	Empirical size ($\omega = 0$) and empirical power, with fixed grid size $D = 50$, after 500 simulations applied on artificial data generated under model $\mathcal{M}_2$ and error model $\mathcal{E}_3$	63
4.10	Empirical sizes (in %) of CFRecon after 500 Monte Carlo simulations, under hypothesis model $\mathcal{F}_2$ and error model $\mathcal{E}_2$, with error models having standard normal distribution.	65
4.11	Empirical power (in %) of CF and CFRecon tests after 500 Monte Carlo simulations, hypothesis model $\mathcal{F}_4$ and error model $\mathcal{E}_2$	67
A.1	Empirical sizes powers (in %) of each EMF-test under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with various variance parameter and shift param- eter settings, $B = 500$, $K = 5$	83
A.2	Empirical levels in % after 500 Monte Carlo simulations.	84
A.3	Empirical powers in % after 500 Monte Carlo simulations.	85
A.4	Empirical sizes ($\delta = 0$) and empirical powers ($\delta > 0$) of EMFT under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with 500 MC simulations, $K = 5$ and $M = 70$	86
A.5	Empirical sizes ($\omega = 0$) and empirical powers, with fixed grid size $M = 50$, after 500 simulations applied on artificial data generated under model $\mathcal{M}_2$ and error model $\mathcal{E}_3$	87
A.6	Empirical levels in % after 500 Monte Carlo simulations.	88

List of Abbreviations

A. Useful theorems and definitions

A.1 Hilbert spaces

For our work, we reduce the Hilbert Space only on $\mathbb{R}^n$.

Definition A.1.1. (*Inner product*): Inner product on vector space X over $\mathbb{R}$ is a mapping $\langle \cdot, \cdot \rangle : X \times X \mapsto \mathbb{R}$ with the following properties:

1. $\langle x, y \rangle = \langle y, x \rangle \quad \forall x, y \in X$,
2. $\langle x, x \rangle \geq 0 \quad \forall x \in X$,
3. $\forall x, y, z \in X \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
4. $\forall \alpha \in \mathbb{R} \langle \alpha x, y \rangle = \alpha \langle x, y \rangle \quad \forall x, y \in X$
5. $\langle x, x \rangle = 0$ if and only if $x = 0$

Pair $(X, \langle \cdot, \cdot \rangle)$ is called Inner product space.

Theorem A.1.1 (Cauchy Schwartz inequality). Let $\mathcal{X}$ be the inner product space. Then

1. $|\langle x, y \rangle| \leq \sqrt{\langle x, x \rangle} \sqrt{\langle y, y \rangle} \quad \forall x, y \in X$
2. Function $\|x\| : \mathcal{X} \mapsto \mathbb{R}_+ : \mathcal{X} \mapsto \sqrt{\langle x, x \rangle}$ for $x \in X$ is the norm on X

Definition A.1.2. A complete metric space, is a metric space, where every Cauchy sequence is convergent.

Definition A.1.3 (Hilbert space). A Hilbert space $\mathcal{H}$ is a real inner product space that is also a complete metric space with respect to the distance function induced by the inner product

A.2 Important results from Stochastic Processes theory

Theorem A.2.1. (Cramér-Slutsky theorem) Let $X_n \xrightarrow{\mathcal{D}} X, Y_n \xrightarrow{P} c$, then

- $X_n + Y_n \xrightarrow{\mathcal{D}} X + Y$
- $X_n Y_n \xrightarrow{\mathcal{D}} Xc$

Definition A.2.1. (Gaussian Processes) A process $Y(t), t \in \mathcal{T}$ is Gaussian with mean function $\mu(t), t \in \mathcal{T}$ and covariance function $\gamma(s, t), s, t \in \mathcal{T}$, if and only if for any p time points $t_j, j = 1, 2, \dots, p$, the random vector $(Y(t_1), \dots, Y(t_p))^T$ follows a multivariate normal distribution $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Gamma})$, where $\boldsymbol{\mu} = (\mu(t_1), \dots, \mu(t_p))^T$ and $\boldsymbol{\Gamma}$ is $p \times p$ covariance matrix with components $\gamma(t_j, t_i), i, j = 1, \dots, p$.

Definition A.2.2. (*Wishart Processes*) A general Wishart Process $W(s, t) \sim WP(n, \gamma)$ can be written as

$$W(s, t) = \sum_{i=1}^n W_i(s, t) = \sum_{i=1}^n v_i(s)v_i(t), \quad (\text{A.1})$$

where $W_i(s, t) = v_i(s)v_i(t)$, $i = 1, 2, \dots, n$ and $v_i(t) \stackrel{iid}{\sim} WP(1, \gamma)$ and $v_i(t)$, $i = 1, \dots, n \stackrel{iid}{\sim} GP(0, \gamma)$.

Theorem A.2.2. Assume that we have K functional samples $X_{1j}(t), \dots, X_{n_K j}(t)$, $j = 1, \dots, K$ where $\mu_j(t) \in L^2(\mathcal{T})$ and $\text{tr}(\gamma) < \infty$. Furthermore let each functional sample comes from Gaussian process. Then

$$(n - K)\hat{\gamma} \sim WP(n - K, \gamma), \quad (\text{A.2})$$

where $\hat{\gamma} = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (X_i(t) - \bar{X}_i(t))(X_i(s) - \bar{X}_i(s))$.

Theorem A.2.3 (Properties of Wishart Process, Zhang [2012]). Let $W(s, t) \sim WP(n, \gamma)$ with $\text{tr}(\gamma) < \infty$. Then we have

- $EW(s, t) = n\gamma(s, t)$,
- $\text{tr}(W)$ is χ^2 -mixture, more precisely $\text{tr}(W)$ has same distribution as $\sum_{m=1}^M \lambda_m A_m$, where $A_1, \dots, A_M$ is iid sequence and $A_m \sim \chi_n^2$,
- $E\text{tr}(W) = n\text{tr}(\gamma)$, $E\text{tr}^2(W) = 2n\text{tr}(\gamma^2) + n^2\text{tr}^2(\gamma)$, and $\text{Var}(\text{tr}(W)) = 2n\text{tr}(\gamma^2)$,
- $E\text{tr}(W^2) = n(n + 1)\text{tr}(\gamma^2) + n\text{tr}^2(\gamma)$

Definition A.2.3. Suppose that $Y_1(t), Y_2(t), \dots, Y_n(t) \stackrel{iid}{\sim} SP(0, \gamma(s, t))$ such that $E\|Y_1\|^4 < \infty$. Set $\bar{q}(s, t) = n^{-1} \sum_{i=1}^n q_i(s, t)$, where $q_i(s, t) = Y_i(s)Y_i(t)$, $s, t \in \mathcal{T}$, $i = 1, \dots, n$. Then as $n \rightarrow \infty$, we have

$$\sqrt{n}[\bar{q}(s, t) - \gamma(s, t)] \xrightarrow{\mathcal{D}} GP(0, \Gamma) \quad (\text{A.3})$$

where $\Gamma = \text{cov}(q_1(s, t), q_1(s', t')) = E[Y_1(s)Y_1(t)Y_1(s')Y_1(t')] - \gamma(s, t)\gamma(s', t')$.

Definition A.2.4 (χ^2 -type mixtures, Zhang [2005]). Random variable, which has form

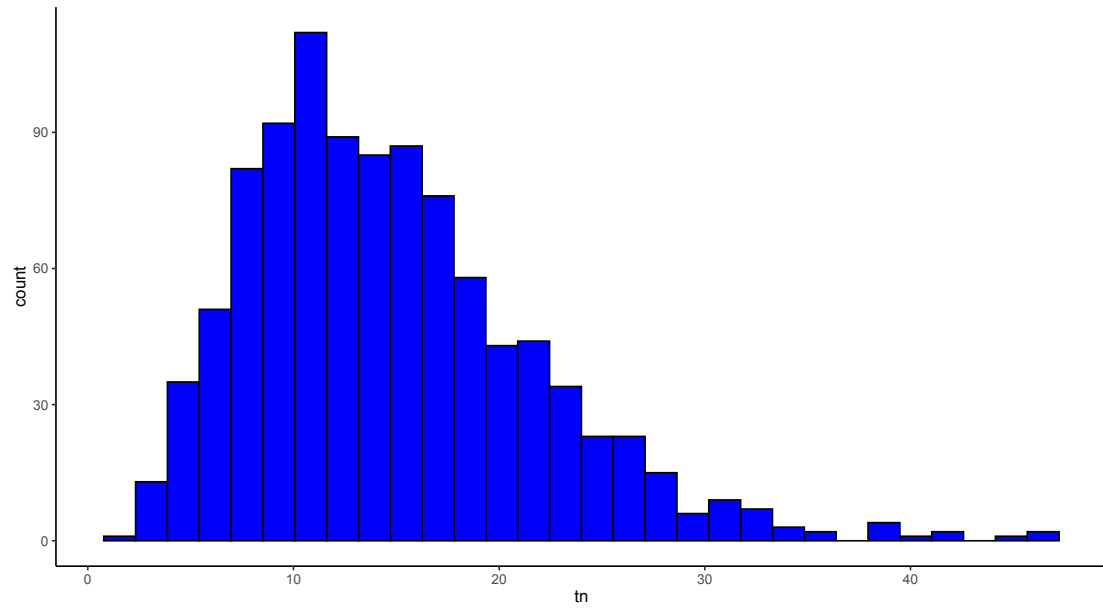
$$T = \sum_{l=1}^m c_l A_l, \quad (\text{A.4})$$

where $c_l, l = 1, \dots, m$ are real scalars and $A_1, A_2, \dots, A_m$ is a sequence of independent variables, where $A_l \sim \chi_1^2(u_l^2)$, where u_l^2 is non-central parameter of χ_1^2 distribution, is called χ^2 -type mixture.

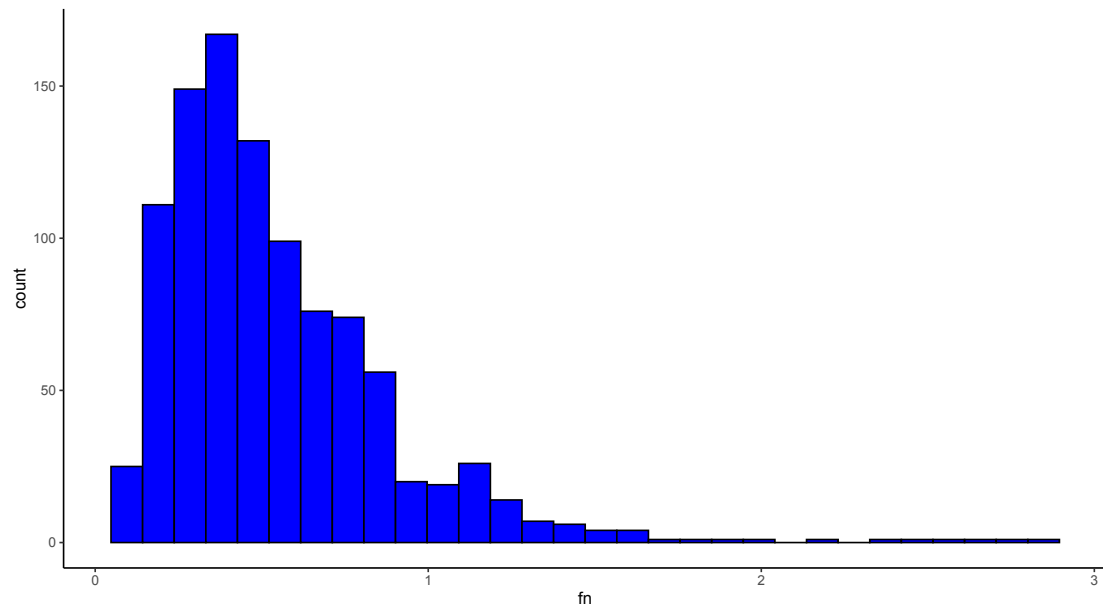
Definition A.2.5 (F -type mixtures, Zhang [2012]). Random variable, which has form

$$F = \frac{\sum_{l=1}^m c_l A_l}{\sum_{j=1}^k h_j B_j}, \quad (\text{A.5})$$

where A_l, B_j are independent of each other, $A_l \sim \chi_a^2(u_l^2)$, $B_j \sim \chi_c^2(v_j^2)$, u_l^2 and v_j^2 are non-central parameters, is called F -type mixture.



(a) Distribution of χ^2 -type mixture random variable.



(b) Distribution of F -type mixture random variable.

Theorem A.2.4 (Distribution of L^2 -norms Zhang [2012]). *Let $y(t) \sim GP(0, \gamma), t \in \mathcal{T}$ with $\mu(t) \in L^2(\mathcal{T})$ and $\text{tr}(\gamma) < \infty$, then the squared L^2 -norm of $y(t)$ has the same distribution as*

$$\|y\|^2 = \sum_{l=1}^m \lambda_l A_l + \sum_{l=m+1}^{\infty} \phi_l, \quad (\text{A.6})$$

where $\lambda_l, l = 1, \dots, \infty$ are the decreasing-ordered eigenvalues of $\gamma(s, t)$, $\phi_l(t), l = 1, \dots, \infty$ are the associated eigenfunctions, $\phi_l = \int_{\mathcal{T}} \mu(t) \phi_l(t) dt, l = 1, 2, \dots, \infty$ and m is the number of all the positive eigenvalues so that $\lambda_m > 0$ and $\lambda_l = 0, l > m$. In particular, when $\mu(t) \equiv 0$, we have $\pi_l = 0, l = 1, \dots, \infty$ so that L^2 norm of $y(t)$ has same distribution as

$$\|y\|^2 = \sum_{l=1}^m \lambda_l A_l, A_l \stackrel{i.i.d}{\sim} \chi_1^2, \quad (\text{A.7})$$

which is a central χ^2 -type mixture.

Theorem A.2.5. *If $\text{tr}(\gamma) < \infty$, then for any finite integer $k = 1, 2, \dots$, we have $\text{tr}(\gamma^{\otimes k}) = \sum_{r=1}^m \lambda_r^k$ where $\lambda_r, r = 1, 2, \dots, m$ are all the m positive eigenvalues of $\gamma(s, t)$.*

Theorem A.2.6 (Zhang [2012]). *Let $\mathbf{y}(t) \sim GP_n(\cdot, \gamma \mathbb{I}_n)$ with $(t) = [\mu_1(t), \dots, \mu_n(t)]^\top \in \mathcal{L}^2(\mathcal{T})$ and $\text{tr}(\gamma) < \infty$. Set $T = \text{tr}(q)$, where $q(s, t) = \mathbf{y}(s)^\top \mathbb{A} \mathbf{y}(t)$ is a quadratic form of $\mathbf{y}(t)$ and $\mathbb{A}$ is an idempotent matrix of rank K . Then T has same distribution as*

$$\sum_{r=1}^m \lambda_r A_r + \sum_{r=m+1}^{\infty} \delta_r^2, A_r \sim \chi_K^2(\lambda_r^{-1} \delta_r^2), \quad (\text{A.8})$$

where $\lambda_r, r = 1, \dots, \infty$ are the eigenvalues of $\gamma(s, t)$ with only first m eigenvalues being positive and $\delta_r^2 = \|\int_{\mathcal{T}} \mu(t) \phi_r(t) dt\|^2, r = 1, \dots, \infty$. In particular, when $(t) = 0$, we have $\delta_r^2 = 0, r = 1, \dots, \infty$ so that we have that T has same distribution as

$$\sum_{r=1}^m \lambda_r A_r, A_r \sim \chi_K^2 \quad (\text{A.9})$$

Theorem A.2.7 (Moments of χ^2 -mixture types). *Let assume that we are working with functional random sample $x_1(t), \dots, X_n(t) \sim SP(0, \gamma)$, and define statistic $T_n = g(X_1(t), \dots, X_n(t))$ which distribution can be approximated by χ^2 -mixture type random variable T . Further, let $\text{tr}(\gamma) < \infty$ hold. Then the first two moments of T are*

$$\begin{aligned} \mathcal{M}_1(T) &= \text{tr}(\gamma) \\ \mathcal{M}_2(T) &= \text{tr}(\gamma^{\otimes 2}) \end{aligned} \quad (\text{A.10})$$

Proof. Let us start with the first moment of central χ^2 -mixture type

$$\begin{aligned} \mathcal{M}_1(T) &= \mathbb{E} T \\ &= \mathbb{E} \sum_{l=1}^m \lambda_l A_l, A_l \sim \chi_d^2 \\ &= \sum_{l=1}^m \lambda_l d, \end{aligned}$$

we know that $\text{tr}(\gamma) < \infty$. Therefore using Theorem A.2.7 we have $\text{tr}(\gamma) = \sum_{l=1}^m$. Hence,

$$\mathcal{M}_1(T) = \sum_{l=1}^m \lambda_l d = d \text{tr}(\gamma). \quad (\text{A.11})$$

In the case of second moment, the approach is the same. Let us rewrite the second moment

$$\begin{aligned} \mathcal{M}_2(T) &= \text{Var} T \\ &= \sum_{l=1}^m \lambda_l^2 \text{Var} A_l \\ &= \sum_{l=1}^m \lambda_l^2 2d \end{aligned}$$

Again, with the assumption of finite trace of covariance function, we have $\text{tr}(\gamma^{\otimes 2}) = \sum_{l=1}^m \lambda_l^2$. Plug-in the previous into derivations above, we have $\mathcal{M}_2(T) = 2d \text{tr}(\gamma^{\otimes 2})$.

QED

A.3 Additional proofs

Proof. (Theorem 3.3.5) Let us rewrite (3.58) as following:

$$\begin{aligned} C_n &= \sum_{i=1}^K (n_i - 1) \int_{\mathcal{T} \times \mathcal{T}} [\hat{\gamma}_i - \gamma(s, t)]^2 ds dt \\ &= \sum_{i=1}^K (n_i - 1) \int_{\mathcal{T} \times \mathcal{T}} [\hat{\gamma}_i^2(s, t) - 2\hat{\gamma}_i(s, t)\hat{\gamma}(s, t) + \hat{\gamma}^2(s, t)] ds dt \end{aligned}$$

Furthermore, let us assume that $X_{ij}(t) = \sum_{l=1}^M c_{ijl} \phi_l(t) = \mathbf{c}_{ij}^\top \boldsymbol{\phi}(t)$, where $\boldsymbol{\phi}(t)$ is vector of orthonormal basis functions of $\mathcal{L}^2$ -space and $\mathbf{c}_{ij}^\top$ is vector of unknown coefficients. This proof will be done partially. It is sufficient to show the idea of the whole proof only for $\hat{\gamma}_i(s, t)$ and the rest is equivalent. With basis expansion representation of functional data, $\hat{\gamma}_i^2$, $i = 1, \dots, K$ can be written as

$$\begin{aligned} \hat{\gamma}_i^2(s, t) &= \frac{1}{(n_i - 1)^2} \left\{ \sum_{j=1}^{n_i} X_{ij}(t) X_{ij}(s) - n_i \bar{X}_i(t) \bar{X}_i(s) \right\} \left\{ \sum_{j=1}^{n_i} X_{ij}(t) X_{ij}(s) - n_i \bar{X}_i(t) \bar{X}_i(s) \right\} \\ &= \frac{1}{(n_i - 1)^2} \left\{ \sum_{j=1}^{n_i} \mathbf{c}_{ij}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ij} - n_i^{-1} \sum_{l,k} \mathbf{c}_{il}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ik} \right\} \\ &\quad \left\{ \sum_{j=1}^{n_i} \mathbf{c}_{ij}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ij} - n_i^{-1} \sum_{l,k} \mathbf{c}_{il}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ik} \right\} \\ &= \alpha_i + \beta_i + \theta_i, \end{aligned}$$

where

$$\begin{aligned} \alpha_i &= \frac{1}{n_i - 1} \sum_{j,l=1}^{n_i} \mathbf{c}_{ij}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ij} \mathbf{c}_{il}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{il}, \\ \beta_i &= -\frac{2}{n_i(n_i - 1)^2} \sum_{j,l,k=1}^{n_i} \mathbf{c}_{ij}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ij} \mathbf{c}_{il}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ik}, \\ \theta_i &= \frac{1}{n_i^2(n_i - 1)^2} \sum_{j,l,k,m=1}^{n_i} \mathbf{c}_{ij}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ij} \mathbf{c}_{ik}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{im}. \end{aligned}$$

Furthermore, it is easy to show, that term $\boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) \mathbf{c}_{ij} \mathbf{c}_{ik}^\top \boldsymbol{\phi}(t) \boldsymbol{\phi}^\top(s) = \Gamma$ can be written as

$$\Gamma = \begin{pmatrix} \sum_{k,r}^M c_{ijl} c_{ikr} \phi_1(t) \phi_l(t) \phi_1(s) \phi_l(s) & \dots & \sum_{k,r}^M c_{ijl} c_{ikr} \phi_1(t) \phi_l(t) \phi_M(s) \phi_l(s) \\ \vdots & \ddots & \vdots \\ \sum_{k,r}^M c_{ijl} c_{ikr} \phi_M(t) \phi_l(t) \phi_1(s) \phi_l(s) & \dots & \sum_{k,r}^M c_{ijl} c_{ikr} \phi_M(t) \phi_l(t) \phi_M(s) \phi_l(s) \end{pmatrix}.$$

We know that $\int_{\mathcal{T}} \phi_i(t) \phi_j(t) = \delta_{ij}$, where for $i, j = 1, \dots, M$, $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise. Therefore, it is obvious that $\int_{\mathcal{T}} \int_{\mathcal{T}} \Gamma ds dt = \text{diag}(c_{ij1} c_{ik1}, \dots, c_{ijM} c_{ikM})$. Therefore we have

$$\int_{\mathcal{T}} \int_{\mathcal{T}} \hat{\gamma}_i^2(s, t) ds dt = \int_{\mathcal{T}} \int_{\mathcal{T}} \alpha_i ds dt + \int_{\mathcal{T}} \int_{\mathcal{T}} \beta_i ds dt + \int_{\mathcal{T}} \int_{\mathcal{T}} \theta_i ds dt,$$

where

$$\begin{aligned} \int_{\mathcal{T}} \int_{\mathcal{T}} \alpha_i &= \frac{1}{n_i - 1} \sum_{j,l=1}^{n_i} \mathbf{c}_{ij}^{\top} \text{diag}(c_{ij1}c_{il1}, \dots, c_{ijM}c_{ilM}) \mathbf{c}_{il}^{\top}, \\ \int_{\mathcal{T}} \int_{\mathcal{T}} \beta_i &= - \frac{2}{n_i(n_i - 1)^2} \sum_{j,l,k=1}^{n_i} \mathbf{c}_{ij}^{\top} \text{diag}(c_{ij1}c_{ik1}, \dots, c_{ijM}c_{ikM}) \mathbf{c}_{il}^{\top}, \\ \int_{\mathcal{T}} \int_{\mathcal{T}} \theta_i &= \frac{1}{n_i^2(n_i - 1)^2} \sum_{j,l,k,m=1}^{n_i} \mathbf{c}_{ij}^{\top} \text{diag}(c_{il1}c_{ik1}, \dots, c_{ilM}c_{ikM}) \mathbf{c}_{im}^{\top}. \end{aligned}$$

If we apply the steps above on terms $\hat{\gamma}_i(s, t)\hat{\gamma}(s, t)$ and $\gamma^2(s, t)$, we obtain the desired result.

QED

A.4 Other simulation results

		$\sigma = 0.2$				$\sigma = 1$				$\sigma = 5$			
		$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.05$	$\delta = 0.1$	$\delta = 0.005$	$\delta = 0.01$	$\delta = 0.05$	$\delta = 0.1$
FP	$n_1 = 5$	45.2	100	100	100	4.6	7.6	100	100	4.4	6.8	8	25.6
	$n_1 = 10$	91.2	100	100	100	6.8	15.6	100	100	4.8	6	16.2	70.8
	$n_1 = 20$	100	100	100	100	8.2	32.6	100	100	6.2	7.4	31.8	99.6
CH	$n_1 = 5$	0	12	100	100	0	0	9.8	100	0	0	0	0
	$n_1 = 10$	25.8	100	100	100	0	0	100	100	0	0	0.2	5
	$n_1 = 20$	99.6	100	100	100	0	6	100	100	1.6	0.4	8.8	89.6
L2N	$n_1 = 5$	7.2	99.4	100	100	0.2	0.2	98.6	100	0.2	0.2	0.2	2
	$n_1 = 10$	73.6	100	100	100	1.4	3	100	100	0.4	1	2.6	37
	$n_1 = 20$	100	100	100	100	3.4	19.4	100	100	3.4	2	18.6	97.4
L2B	$n_1 = 5$	56	100	100	100	6.4	10.2	100	100	6.6	9.4	11.4	31.2
	$n_1 = 10$	93	100	100	100	7.8	17.6	100	100	4.8	6.8	18.6	72.4
	$n_1 = 20$	100	100	100	100	9.4	33.6	100	100	7.2	7.4	33.4	99.2
L2b	$n_1 = 5$	0.4	88.8	100	100	0	0	84	100	0	0	0	0
	$n_1 = 10$	29	100	100	100	0	0	100	100	0	0	0.2	5.4
	$n_1 = 20$	99.4	100	100	100	0	3.2	100	100	0.4	0	4.8	83.8
FN	$n_1 = 5$	4.2	98.6	100	100	0.2	0	97.6	100	0	0.2	0	0.8
	$n_1 = 10$	68.6	100	100	100	1.4	1.8	100	100	0.4	0.6	2.4	33.2
	$n_1 = 20$	100	100	100	100	2.8	18.4	100	100	3.2	1.6	17.8	97
FB	$n_1 = 5$	48.2	100	100	100	3.8	8	100	100	4.8	6	8.4	25.4
	$n_1 = 10$	91.4	100	100	100	7.2	16.2	100	100	3.4	5.8	16.8	69.6
	$n_1 = 20$	100	100	100	100	8.6	32	100	100	6.6	7	32	99.2
Fb	$n_1 = 5$	0	0	100	100	0	0	0	100	0	0	0	0
	$n_1 = 10$	0.6	100	100	100	0	0	100	100	0	0	0	0
	$n_1 = 20$	98.6	100	100	100	0	0.8	100	100	0	0	0.4	56
GPF	$n_1 = 5$	11.8	99.6	100	100	0.4	1	99.6	100	0.4	0.8	1	6
	$n_1 = 10$	76.6	100	100	100	2.6	3.6	100	100	0.8	1.6	4.8	41.6
	$n_1 = 20$	100	100	100	100	3.4	21	100	100	4.4	2.2	20.2	97.4
Fmaxb	$n_1 = 5$	9	44	100	100	3.2	5.4	37.8	100	5	5.2	4.6	8
	$n_1 = 10$	25.8	97.4	100	100	6.8	6.4	94.8	100	4.4	5.8	5.6	12.6
	$n_1 = 20$	64	100	100	100	6	9.6	100	100	4.4	6.2	10	32.8
L2recon	$n_1 = 5$	97.2	100	100	100	6.2	17.2	100	100	4.6	6.4	19.2	75.6
	$n_1 = 10$	100	100	100	100	13.4	43.2	100	100	4	6.2	47.4	99.2
	$n_1 = 20$	100	100	100	100	20	80.6	100	100	7.2	7.2	80.8	100

Table A.1: Empirical sizes powers (in %) of each EMF-test under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with various variance parameter and shift parameter settings, $B = 500, K = 5$.

Test	$M = 50$									$M = 70$								
	$n_j=10$			$n_j=15$			$n_j=25$			$n_j=10$			$n_j=15$			$n_j=25$		
	$\sigma=0.2$	$\sigma=1$	$\sigma=4$	$\sigma=0.2$	$\sigma=1$	$\sigma=4$	$\sigma=0.2$	$\sigma=1$	$\sigma=4$	$\sigma=0.2$	$\sigma=1$	$\sigma=4$	$\sigma=0.2$	$\sigma=1$	$\sigma=4$	$\sigma=0.2$	$\sigma=1$	$\sigma=4$
FP	5.8	5	6.2	7	7	6	5.4	3.8	5	4	5.8	4.6	4.8	5.6	4.4	4.4	6.4	6.2
CH	4.2	2	4	5.4	5	3.6	4.4	3	4.4	2.2	3.4	3.4	3.8	4.4	2.4	3.2	5.4	4.8
L2N	8	6.6	8.2	8.6	7.6	7.2	5.8	4.4	5	6.6	7.8	5.4	6.2	7	5.6	5	7	7.2
L2B	8.2	7.4	8.6	9.2	8	8.2	6.8	5	5	7.8	8.6	6	8	7.6	6	5.4	7	7.4
L2b	7.2	5.8	8	8.2	7	6.8	5.6	3.8	5	5.2	7.4	5.2	5.8	6.8	5	4.6	6.8	6.4
FN	5.6	4.8	5.8	7	6.4	6.2	5.4	4	4.8	4.2	6.4	4.8	5	6	4.8	4.4	6.4	6
FB	5.8	5	6	7	6.6	6.6	5.4	4	4.8	4.2	6.4	4.8	5	6	4.8	4.4	6.6	6
Fb	3.6	2.2	3.4	5.2	4.4	4	4.2	2.8	4.2	3	2.6	3	3.8	4.4	2.6	3.2	4.6	4.4
GPF	7.4	6.2	8	8.4	7.4	7	6	4.4	5	6.2	7.8	5.4	6.2	7	5.2	4.8	7	7.2
Fmaxb	5	4.6	5.8	6.4	5.4	6.2	6.4	3.8	4.4	4.8	5.8	4.4	4.6	6.6	3.6	3.8	5.2	6.4

Table A.2: Empirical levels in % after 500 Monte Carlo simulations.

$M = 50$												$M = 70$												
$n_j = 10$						$n_j = 15$						$n_j = 10$						$n_j = 15$						
$\sigma=0.01$	$\sigma=0.04$	$\sigma=0.1$	$\sigma=0.5$	$\sigma=1$		$\sigma=0.01$	$\sigma=0.04$	$\sigma=0.1$	$\sigma=0.5$	$\sigma=1$		$\sigma=0.01$	$\sigma=0.04$	$\sigma=0.1$	$\sigma=0.5$	$\sigma=1$		$\sigma=0.01$	$\sigma=0.04$	$\sigma=0.1$	$\sigma=0.5$	$\sigma=1$		
FP	0.932	0.916	0.926	0.932	0.91	0.982	0.984	0.988	0.982	0.982	0.814	0.934	0.928	0.92	0.912	0.984	0.976	0.984	0.984	0.974	0.978	0.984	0.984	0.974
CH	0.85	0.798	0.838	0.848	0.834	0.974	0.98	0.984	0.978	0.966	0.814	0.848	0.804	0.83	0.83	0.978	0.974	0.978	0.984	0.972	0.978	0.984	0.988	0.972
L2N	0.952	0.936	0.942	0.956	0.938	0.984	0.99	0.988	0.986	0.988	0.962	0.954	0.95	0.938	0.936	0.984	0.978	0.984	0.988	0.976	0.984	0.988	0.976	
L2B	0.96	0.944	0.944	0.964	0.95	0.986	0.99	0.99	0.99	0.99	0.97	0.96	0.956	0.948	0.94	0.986	0.98	0.984	0.988	0.982	0.984	0.988	0.982	
L2b	0.956	0.94	0.948	0.966	0.954	0.986	0.99	0.99	0.982	0.986	0.956	0.954	0.944	0.944	0.934	0.982	0.98	0.984	0.986	0.978	0.984	0.988	0.978	
FN	0.938	0.924	0.928	0.938	0.918	0.98	0.986	0.988	0.982	0.982	0.948	0.932	0.928	0.926	0.914	0.984	0.978	0.984	0.984	0.976	0.984	0.988	0.976	
FB	0.942	0.926	0.934	0.938	0.918	0.982	0.988	0.988	0.982	0.982	0.95	0.938	0.93	0.93	0.922	0.984	0.978	0.984	0.984	0.978	0.984	0.988	0.976	
Fb	0.738	0.706	0.76	0.76	0.768	0.966	0.962	0.966	0.962	0.952	0.73	0.772	0.736	0.758	0.742	0.952	0.962	0.952	0.976	0.956	0.984	0.988	0.976	
GPF	0.952	0.932	0.938	0.95	0.932	0.984	0.99	0.988	0.986	0.986	0.962	0.952	0.95	0.936	0.928	0.984	0.978	0.984	0.988	0.976	0.984	0.988	0.976	
Fmaxb	0.908	0.876	0.908	0.902	0.894	0.976	0.978	0.986	0.984	0.976	0.908	0.9	0.89	0.914	0.878	0.98	0.976	0.978	0.984	0.974	0.984	0.988	0.974	

Table A.3: Empirical powers in % after 500 Monte Carlo simulations.

		$\rho = 0.1$				$\rho = 0.5$				$\rho = 0.7$				$\rho = 0.9$			
		$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 2$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 2$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 2$	$\delta = 0$	$\delta = 0.1$	$\delta = 0.5$	$\delta = 2$
FP	$n_1 = 5$	5	4.4	31.8	100	4.2	4.2	7.6	80.2	4.2	4.8	6.6	56.4	4.6	5.4	6.8	31
	$n_1 = 10$	5.2	7.2	65.8	100	5	4.8	13.2	99.8	5.2	4	8.4	92.2	5.2	4.8	7.8	67.6
	$n_1 = 20$	4.2	9	97.8	100	3.6	6.6	27.6	100	4.6	7.4	18.2	100	4.2	5.8	12.4	98.2
CH	$n_1 = 5$	0.6	1	11	100	0.4	0.6	1	44	0.2	0.2	0.8	15.6	0	0	0	1
	$n_1 = 10$	3.4	3.8	54	100	2.8	1.2	6.6	98.6	1.6	1.4	3	85.6	1.6	1.4	1.8	42.2
	$n_1 = 20$	3	7.4	96.8	100	2.6	4.2	23.4	100	1.6	3	14.4	100	1.8	2.8	5.8	95.6
L2N	$n_1 = 5$	9.4	7.6	41.6	100	5.8	5.8	11	87.4	4.8	5.2	8.4	60.8	4	4	6.4	29.2
	$n_1 = 10$	6.6	9.4	71	100	5.6	5.6	15.2	99.8	5.6	4.2	9.4	93	5	4.6	6.4	66.8
	$n_1 = 20$	5.2	10	98	100	4.2	6.6	29.2	100	5.4	7.2	19.4	100	3.8	5.8	11	98.2
L2B	$n_1 = 5$	11.8	9.2	46.6	100	8.4	8.2	13.4	90.2	7.8	8.2	10.8	68.4	6.8	8.6	9.8	41.4
	$n_1 = 10$	8.4	10.8	74	100	7	6.4	16.6	100	6.8	5.8	10.8	94.2	6.4	6.6	9.4	71.6
	$n_1 = 20$	5.6	10.4	98.2	100	4.8	7	31.2	100	5.6	8	20.6	100	4.8	6.2	13.2	98.6
L2b	$n_1 = 5$	10	8.2	41.6	100	5.4	5.4	9.4	84.6	3.6	3.8	5.8	55	1.6	1.8	2.2	20.4
	$n_1 = 10$	5.8	8.4	68.8	100	4.6	5.4	13.6	99.6	3.6	2.8	6.2	91.4	3	2.2	3.8	56.2
	$n_1 = 20$	4.8	9.8	98	100	3.4	5.6	27	100	2.8	5	16.2	100	2.6	3.6	8.2	96.6
FN	$n_1 = 5$	5.2	4.8	32	100	4.4	4.2	8	81.2	3	3	5	54.8	2.2	2.2	2.6	22.4
	$n_1 = 10$	5.4	7.4	66	100	5.4	4.8	14	99.6	4.4	3.8	7.6	92.6	4.6	3.4	4.6	64
	$n_1 = 20$	4.2	9	97.8	100	3.8	6	27.8	100	4	6.6	18.6	100	3.2	4.6	10	98.2
FB	$n_1 = 5$	5.4	5.4	32.8	100	4.8	5	9.6	84.4	4.6	5	8.2	60.4	4.8	5.8	7.4	32.4
	$n_1 = 10$	5.4	7.4	66	100	5.4	5.6	14.6	99.8	5.6	4.2	9	93	5.4	4.8	7.8	68.6
	$n_1 = 20$	4.2	9.2	97.8	100	4	6.4	28.2	100	5.2	7.2	19.2	100	4	5.8	12.2	98.2
Fb	$n_1 = 5$	0.8	1.8	11.4	100	0.4	0.4	1.2	43.2	0.2	0	0	11.2	0	0	0	0.6
	$n_1 = 10$	3.6	4	53.4	100	2.6	0.8	5.8	98	1.4	0.8	2.2	82.8	0	0	0.2	30.6
	$n_1 = 20$	2.6	7.4	96.8	100	2.6	4	21.6	100	1.2	2.4	11.8	100	1.6	1.8	4.2	94.2
GPF	$n_1 = 5$	9.2	7.2	40.6	100	8.2	7	12.4	87.4	9.2	7.8	11.6	64.6	6.6	6.8	8.6	35.8
	$n_1 = 10$	6.6	9	69.8	100	7.4	6.2	16	100	6.8	5.6	10.2	93.6	5.4	5.8	7.6	67.2
	$n_1 = 20$	5	10	98	100	4.6	7	31	100	5	8	21	100	4.6	5.4	11.8	98.6
Fmaxb	$n_1 = 5$	4.4	5	42.8	100	5.2	3.2	6.6	75.4	4	4.4	5.4	35.8	6	5.2	5.8	16
	$n_1 = 10$	5.6	7	88.2	100	6.2	4.4	10.2	99.6	5.4	4	7	80.4	5.4	3.4	4.2	36.4
	$n_1 = 20$	4.6	10.8	100	100	4.8	7	27	100	5.2	6.2	14.2	99.8	5.8	5.8	8.2	80
L2recon	$n_1 = 5$	5.2	4.6	31	100	4.4	4.4	8	79.6	4.4	4.8	6.6	55	4.6	5	6.4	30.8
	$n_1 = 10$	5.2	7.6	65	100	5.4	4.6	13.4	99.4	5.2	4	8.2	91.6	5.2	5.2	8.6	68.4
	$n_1 = 20$	4.2	9	97	100	3.6	5.8	27.4	100	5	6.8	18.4	100	4.2	5.8	11.8	98.4

Table A.4: Empirical sizes ($\delta = 0$) and empirical powers ($\delta > 0$) of EMFT under hypothesis model $\mathcal{F}_1$ and error model $\mathcal{E}_1$, with 500 MC simulations, $K = 5$ and $M = 70$.

		L2BEcf				L2NEcf				L2PEcf			
		$\delta = 0.2$		$\delta = 0.5$		$\delta = 0.2$		$\delta = 0.5$		$\delta = 0.2$		$\delta = 0.5$	
	ω	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$	$n_1 = 10$	$n_2 = 25$
$\rho = 0.1$	0	4	5	4	5	3.2	4.8	3.2	4.8	3.4	4.6	2	2.2
	0.01	4.2	5	4.2	5	3.2	4.6	3.2	4.6	3.2	4.4	2	2.2
	0.1	4.4	4.6	4.4	4.6	3.6	4.6	3.6	4.6	3.6	5	1.8	1.8
	1	23.6	61.6	23.6	61.6	21.6	61	21.6	61	17.8	55	11.2	44
	3	92.4	100	92.4	100	91.8	100	91.8	100	73.4	100	68.6	100
	8	99.6	100	99.6	100	99.6	100	99.6	100	85.2	100	84.8	100
$\rho = 0.5$	0	3.4	4	3.4	4	2.4	3.8	2.4	3.8	4.6	4.8	3.2	3.4
	0.01	3.2	4.4	3.2	4.4	2.4	3.6	2.4	3.6	4.2	4.8	3	3.4
	0.1	4	6.6	4	6.6	3.2	6.2	3.2	6.2	5	7	4.4	5.6
	1	82.8	100	82.8	100	81	100	81	100	64.2	99.8	62.6	99.6
	3	99.4	100	99.4	100	99.4	100	99.4	100	83.8	100	84.4	100
	8	99.6	100	99.6	100	99.6	100	99.6	100	85.8	100	86.2	100
$\rho = 0.9$	0	4.2	4.2	4.2	4.2	1.6	2.8	1.6	2.8	4.2	4.4	3.8	4.2
	0.01	4.2	4	4.2	4	1.6	2.8	1.6	2.8	4.2	4.4	4	4
	0.1	5	8	5	8	2	4.2	2	4.2	6	6.4	5.6	5.6
	1	91.4	100	91.4	100	89.6	100	89.6	100	72.6	100	71	100
	3	99.6	100	99.6	100	99.6	100	99.6	100	84.6	100	84.4	100
	8	99.6	100	99.6	100	99.6	100	99.6	100	86.4	100	86.4	100

Table A.5: Empirical sizes ($\omega = 0$) and empirical powers, with fixed grid size $M = 50$, after 500 simulations applied on artificial data generated under model $\mathcal{M}_2$ and error model $\mathcal{E}_3$.

	$\mathcal{F}_1$						$\mathcal{F}_2$					
	$\rho = 0.1$	$\rho = 0.26$	$\rho = 0.42$	$\rho = 0.58$	$\rho = 0.74$	$\rho = 0.9$	$\rho = 0.1$	$\rho = 0.26$	$\rho = 0.42$	$\rho = 0.58$	$\rho = 0.74$	$\rho = 0.9$
FP	4	4.2	4.4	4.2	4	5	5.2	2.8	5.2	4.2	4.8	4.8
CH	2.6	2.8	3	2.6	2.6	2	3.4	1.4	3.6	3	1	1
L2N	5.8	4.8	5.2	4.4	4.2	5	6	3.6	6.6	5.2	5.2	3.4
L2B	6.6	6.2	5.4	5	5.2	6	6.6	3.6	7	5.4	6.2	5.8
<i>L2b</i>	5.4	4.2	4.8	3.8	3.8	2.6	5.2	3	5.4	4.2	3.2	1.2
FN	4.4	4	4.4	4	4	4.6	5	3	5.4	4.4	4.2	3
FB	4.6	4	4.6	4.2	4.4	5.6	5	3	5.6	4.8	5.4	4
Fb	2.6	2.8	2.6	1.6	1.2	0.6	3	1.4	3	1.6	1	0.6
GPF	5.4	4.8	5.2	5.6	6	5.2	5.6	3.4	6.6	5.8	5.2	4.6
Fmaxb	4.4	5	3.4	3.6	4.6	2.4	4.2	2.6	4.2	4.2	5.4	5.4

80 Table A.6: Empirical levels in % after 500 Monte Carlo simulations.