FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

BACHELOR THESIS

Viktor Dolník

Functional ANOVA

Department of Probability and Mathematical Statistics

Supervisor of the bachelor thesis:	RNDr. Jiří Dvořák, Ph.D.
Study programme:	Mathematics
Study branch:	General Mathematics

Prague 2018

I declare that I carried out this bachelor thesis independently, and only with the cited sources, literature and other professional sources.

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 .... signature of the author

Title: Functional ANOVA

Author: Viktor Dolník

Department: Department of Probability and Mathematical Statistics

Supervisor: RNDr. Jiří Dvořák, Ph.D., Department of Probability and Mathematical Statistics

Abstract: We introduce the concept of functional data and the problem of functional analysis of variance, which differs from the univariate case in the fact that random functions, not random variables, are the subject of comparison. We continue by deriving an asymptotic test for functional one-way ANOVA from the elementary univariate F-test. We describe the simulation envelope test, whose global version suffers from the multiple comparisons problem. Then, an ordering is defined, based on which we create the rank envelope test, a stronger alternative to the simulation envelope test. We also describe how the rank test can be interpreted graphically. Using the rank envelope test, we devise another test for functional one-way ANOVA, which is also graphically interpretable and thus does not need a post-hoc analysis to identify which groups caused rejection of the null hypothesis. We compare the one-way ANOVA tests on a real-case study and a simulation study.

Keywords: Analysis of variance, F-test, envelope tests, permutation tests

I would like to thank my supervisor RNDr. Jiří Dvořák, Ph.D. for his valued advice and insight.

Contents

Introduction

In recent years, mathematical statistics has marked a growing popularity in functional data statistics. Due to an ever-increasing digitalisation and rapid collection of data, we mark a shift from dealing with random variables to dealing with random functions on an interval, usually representing time. Thus, tests that deal with random functions are becoming essential for correct statistical inference.

In univariate statistics, the analysis of variance (ANOVA), which deals with testing an equality of means of multiple groups of samples, has proven to be exceptionally useful in statistical inference. The functional case, however, was until recently merely transformed into a univariate test statistic, where all information about specific parts of functions that lead to rejection vanish.

The main goal of this thesis is to describe a functional one-way ANOVA test which includes a graphical interpretation, thereby providing a means for a detailed analysis of results without any need for further testing after rejection of the null hypothesis.

The first chapter of this thesis introduces several basic definitions and results from univariate statistics, which will be relied on in the subsequent chapters. Then, the univariate analysis of variance is introduced, along with the F-test, which is an elementary test of the ANOVA problem.

The second chapter introduces the topic of functional data and the one-way ANOVA setting for functional data, which is followed by a straightforward extension of the F-test to the functional case.

The thesis continues by introducing an envelope test, from which the rank envelope test is rigorously derived. The rank envelope test is crucial in developing the new graphical functional one-way ANOVA test.

The last chapter describes how the new one-way ANOVA test is constructed from the rank envelope test. It is then applied to real data, as well as to simulations, where various versions of the test are compared with each other and with the F-test from Chapter 2.

1. Univariate Statistics

1.1 Preliminaries

In this section, we will review a few basic definitions and theorems of mathematical statistics on which we will rely throughout the rest of the thesis. We will be working with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ throughout the chapter.

Definition 1.1. Let $X_1, \dots, X_n,\ n \in \mathbb{N}$ be a sequence of independent, identically distributed real-valued random variables from distribution F_0. We call the sequence a *random sample* from distribution F_0 and we denote it as

$$X_1, \dots, X_n \sim F_0.$$

Definition 1.2. Let $X_1, \dots, X_n,\ n \in \mathbb{N}$, be a sequence of real-valued random variables. We call the random variables *exchangeable*, if $\forall B_1, \dots, B_n \in \mathcal{B}$, where $\mathcal{B}$ is a Borel σ-algebra on $\mathbb{R}$, and for every permutation $\pi : \{1, \dots, n\} \to \{1, \dots, n\}$, it holds that:

$$\mathbb{P}(X_1 \in B_1, \dots, X_n \in B_n) = \mathbb{P}(X_{\pi(1)} \in B_1, \dots, X_{\pi(n)} \in B_n).$$

Remark. *Notice that if the sequence $X_1, \dots, X_n$ is a random sample, it is exchangeable.*

Definition 1.3. Let $X_1, \dots, X_n$ be a random sample from distribution F_0. Reorder the sequence into a sequence $X_{(1)}, \dots, X_{(n)}$, such that

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}.$$

We call $X_{(k)}$ the k-th order statistic.

Let $R_i \in \{1, \dots, n\}$ be a random variable such that $X_i = X_{(R_i)}$. We call R_i the rank of X_i.

Notice that rank is not necessarily unique for some distributions. For example, in a random sample from a discrete distribution, a pair of random variables has a non-zero probability to be equal.

Theorem 1.1. *Let $X_1, \dots, X_n \sim F_0$ be a random sample, such that*

$$\forall i, j \in \{1, \dots, n\},\ i \neq j :\ \mathbb{P}(X_i = X_j) = 0$$

and let $R_1, \dots, R_n$ be the respective ranks of $X_1, \dots, X_n$. Then,

$$\mathbb{P}(R_i = k) = \frac{1}{n},\ \forall i, k \in \{1, \dots, n\}.$$

Proof. Directly from Casella and Berger [2002], Chapter 5, page 230. □

A generalised version of this theorem will be proved in Chapter 3.

Definition 1.4. Let $A \subset \mathbb{R}^n,\ n \in \mathbb{N}$. By $\mathbf{1}_A$ we will mean a function

$$\mathbf{1}_A : \mathbb{R}^n \to \{0, 1\},$$

such that

$$\begin{aligned} \mathbf{1}_A(x) &= 1 \text{ if } x \in A, \\ \mathbf{1}_A(x) &= 0 \text{ if } x \notin A. \end{aligned}$$

Definition 1.5. Let $X_n \sim F_n,\ n \in \mathbb{N},\ X \sim F$ be random variables. We say that random variables X_n converge to X for $n \to \infty$ *in distribution*, if

$$\lim_{n \to \infty} F_n(x) = F(x) \text{ for all } x \in \mathbb{R} \text{ where } F \text{ is continuous.}$$

We denote this convergence as $X_n \xrightarrow[n \to \infty]{d} X$.

1.2 A one-way ANOVA test

In this section, the one-way analysis of variance (or ANOVA) is introduced, along with the most basic test that belongs to this branch of statistics. As a reference, we use Bingham and Fry [2010], Section 2.6. For a fixed $k \in \mathbb{N}$, we will consider the model:

$$\mathcal{F} = \{F_i = N(\mu_i, \sigma^2);\ \mu_i \in \mathbb{R};\ i = 1, \ldots, k;\ \sigma^2 > 0\},$$

where $N(\mu, \sigma^2)$ denotes the normal distribution with a mean of μ and a variance of σ^2.

The name ANOVA is used to denote the collection of procedures that deal with the following problem:

Let $k, n_1, \ldots, n_k \in \mathbb{N}$ and let:

$$\begin{aligned} &X_{1,1}, \ldots, X_{1,n_1} \sim F_1, \\ &X_{2,1}, \ldots, X_{2,n_2} \sim F_2, \\ &\qquad \vdots \\ &X_{k,1}, \ldots, X_{k,n_k} \sim F_k \end{aligned}$$

be random variables of the model $\mathcal{F}$, such that:

- Every group $X_{i,1}, \ldots, X_{i,n_i}$ is a random sample from the distribution F_i.
- Groups are independent of each other.

We test the following null hypothesis:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k,$$

where $\mu_i := \mathbb{E}X_{i,j},\ \forall i \in \{1, \ldots, k\},\ \forall j \in \{1, \ldots, n_i\}$. The alternative hypothesis is:

$$H_1 : \exists i, j \in \{1, \ldots, k\} \text{ such that } \mu_i \neq \mu_j.$$

Example

The analysis of variance was first described by Ronald Fischer, whose motivation was the following problem in agriculture, as is described in [Bingham and Fry, 2010].

We want to compare the effects of fertilisers on crop yields. For $k \in \mathbb{N}$ fertilisers and $n_i \in \mathbb{N}$ trials for each fertiliser $i \in \{1, \ldots, k\}$, we measure the yield of a crop. We assume that the yields are independent of each other and that their distribution is normal and equal in variance.

For $k = 2$, we can simply use the two-sample t-test. For $k > 2$, we compare the variability between groups with the variability within groups, which should be similar if the effect of fertilisers is generally the same. We will formalise this method into an *F-test*.

Definition 1.6. For random variables $X_{i,j}$ defined as above and for $n = \sum_{i=1}^{k} n_i$, we call:

$$\overline{X}_{i,+} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j} \text{ the sample mean of group } i,$$

$$\overline{X}_{+,+} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} X_{i,j} \text{ the overall sample mean,}$$

$$SS_T = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{i,j} - \overline{X}_{+,+})^2 \text{ the total sum of squares,}$$

$$SS_A = \sum_{i=1}^{k} n_i(\overline{X}_{i,+} - \overline{X}_{+,+})^2 \text{ the sum of squares between groups,}$$

$$SS_e = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{i,j} - \overline{X}_{i,+})^2 \text{ the residual sum of squares.}$$

Theorem 1.2. *Let $X_{i,j},\ i \in \{1, \ldots, k\},\ j \in \{1, \ldots, n_i\}$ be random variables. Then,*

$$SS_T = SS_A + SS_e.$$

Proof. see Bingham and Fry [2010], page 43. □

Theorem 1.3. *Let random variables $X_{i,j}$ be defined as above and assume that the model $\mathcal{F}$ holds. Then,*

$$\frac{SS_e}{\sigma^2} \sim \chi^2_{n-k};\ \mathbb{E}\left(\frac{SS_e}{\sigma^2}\right) = \sigma^2, \tag{1.1}$$

where χ^2_{n-k} is the chi-squared distribution with $n - k$ degrees of freedom.

Furthermore, under the null hypothesis H_0, the following is true:

$$\frac{SS_T}{\sigma^2} \sim \chi^2_{n-1};\ \mathbb{E}\left(\frac{SS_T}{\sigma^2}\right) = \sigma^2, \tag{1.2}$$

$$\frac{SS_A}{\sigma^2} \sim \chi^2_{k-1};\ \mathbb{E}\left(\frac{SS_A}{\sigma^2}\right) = \sigma^2, \tag{1.3}$$

$$SS_A \text{ and } SS_e \text{ are independent.} \tag{1.4}$$

Proof. see Bingham and Fry [2010], page 43-44. □

Theorems 1.2. and 1.3. yield that the ratio between the 'external' (between groups) sum of squares and the 'internal' (within groups) sum of squares has a high probability of being close to zero if the null hypothesis is true. The test statistic will be:

$$F_A := \frac{SS_A \times (n-k)}{SS_e \times (k-1)}.$$

Theorem 1.4. *If the model F holds and the null hypothesis is true, then* $F_A \sim F_{k-1,n-k}$*, where* $F_{k-1,n-k}$ *is the Fischer-Snedecor distribution with (k-1, n-k) degrees of freedom.*

Proof. see Anděl [1998], page 115. □

The critical region is naturally defined by an F-quantile:

$$H_0 \text{ is rejected} \iff F_A \geq F_{k-1,n-k}(\alpha),$$

with F-quantile is defined for $F_A \sim F_{k-1,n-k}$ and $\alpha \in (0,1)$ by the equality

$$\mathbb{P}(F_A \geq F_{k-1,n-k}(\alpha)) = \alpha.$$

The one-way ANOVA F-test is an exact test – if the model $\mathcal{F}$ holds, the significance level is exactly α. It is also a one-tailed test since we consider a test statistic that is close to zero to be in support of the null hypothesis.

2. ANOVA Tests on Functional Data

This chapter introduces the concept of functional data and functional one-way ANOVA. It continues by describing a commonly used functional one-way ANOVA test, which will be subjected to comparison with the graphical ANOVA test later on.

2.1 Functional Data

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let I be a set of indices, such that $X(t) : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is a random variable $\forall t \in I$. We denote a stochastic (or random) process as

$$\{X(t),\ t \in I\} : (\Omega, \mathcal{F}) \to (\mathbb{R}^I, \mathcal{B}^I),$$

where $(\mathbb{R}^I, \mathcal{B}^I)$ is the product measurable space

$$(\bigtimes_{t \in I} \mathbb{R}, \bigotimes_{t \in I} \mathcal{B}).$$

The set of indices I is commonly interpreted as a set of points in time, thus the name 'process'. This mirrors the actual applications of the definition – in practice, it is common to measure a process for a period of time and infer on the whole measurement. The 'randomness' steps into the problem either as an error of measurement or as a natural irregularity of the observed phenomenon.

Notice that for a fixed $\omega \in \Omega$, our stochastic process $\{X(\omega)(t) : t \in I\}$ becomes a function $X(\omega) : I \to \mathbb{R}$. Therefore, the term 'random function' is sometimes used instead. We will be using this term as well.

In practice, the interval I is often discretised and instead of $\{X(t) : t \in I\}$, we only measure a random vector. Nonetheless, it makes sense to construct a theory on random functions and approximate them by a random vector at the latest moment.

A random function $\{X(t), t \in I\}$ has a mean function, denoted as

$$\mathbb{E}(X)(t),\ t \in I,$$

differing from the univariate case only in the fact that it is a real function, not a real number. We also consider an autocovariance function, denoted as

$$\mathrm{Cov}(X)(t, s),\ t, s \in I,$$

which is the covariance between random variables $X(t)$ and $X(s)$. By variance of a random function, we will mean

$$\mathrm{Var}(X)(t) := \mathrm{Cov}(X)(t, t),\ t \in I.$$

In the rest of this thesis, we will consider all random functions to have finite variance functions $\mathrm{Var}(X)(t) < \infty$.

Example

Consider the problem of measuring the temperature of water in one year. We can measure the temperature continuously, such that one random element of our observations is a single function, with an interval on the x-axis representing the time of one year and the y-axis representing temperature.

Since weather is a very complex phenomenon, it appears to be partially random in time. Thus, we can measure water temperature for several years and statistically examine the collection of functions.

As in univariate statistics, there are numerous statistical problems and tests that deal with them. In this thesis, we are interested in the problem of testing an equality of mean functions – in other words, we are dealing with the extension of ANOVA to functional data, sometimes named as fANOVA, to which we will now move on.

2.2 Functional ANOVA

Let $K,\ n_1, \dots, n_K \in \mathbb{N},\ K > 2,\ N := \sum_{i=1}^{K} n_i,\ I = (a, b) \subset \mathbb{R}$ and let:

$$\begin{gathered}
\{X_{1,1}(t), t \in I\}, \dots, \{X_{1,n_1}(t), t \in I\} \sim \mathbb{P}_{X_1}, \\
\{X_{2,1}(t), t \in I\}, \dots, \{X_{2,n_2}(t), t \in I\} \sim \mathbb{P}_{X_2}, \\
\vdots \\
\{X_{K,1}(t), t \in I\}, \dots, \{X_{K,n_K}(t), t \in I\} \sim \mathbb{P}_{X_K}
\end{gathered}$$

be random functions, such that $\forall i \neq j,\ i, j \in \{1, \dots, K\}$:

- Each group $\{X_{i,1}(t), t \in I\}, \dots, \{X_{i,n_i}(t), t \in I\}$ is a random sample from distribution $\mathbb{P}_{X_i}$.
- $\forall X \sim \mathbb{P}_{X_i},\ \forall Y \sim \mathbb{P}_{X_j},\ \forall t, s \in I$: random variables $X(t)$ and $Y(s)$ are independent.
- $\forall X \sim \mathbb{P}_{X_i},\ \forall Y \sim \mathbb{P}_{X_j},\ \forall t, s \in I : \mathrm{Cov}(X)(t, s) = \mathrm{Cov}(Y)(t, s)$.
  This assumption is called '*homoscedasticity*'.

We have already assumed that mean and autocovariance functions exist and are finite for each group. We test the following null hypothesis:

$$H_0 : \forall t \in I :\ \mu_1(t) = \mu_2(t) = \cdots = \mu_K(t).$$

where $\mu_i(t) := \mathbb{E}X_{i,j}(t),\ \forall j \in \{1, \dots, n_i\},\ i \in \{1, \dots, K\}$.

By $\mu(t), t \in I$, we will denote the overall mean, which exists under the null hypothesis. The variance function, which is the same for all $\{X_{i,j}(t), t \in I\}$, will be denoted as $\mathrm{Var}(X)(t),\ t \in I$.

The alternative hypothesis would be:

$$H_1 : \exists i, j \in \{1, \dots, K\},\ \exists t_0 \in I :\ \mu_i(t_0) \neq \mu_j(t_0).$$

It is important to mention that for:

$$t, s \in I,\ t \neq s,\ i \in \{1, \dots, K\},\ j \in \{1, \dots, n_i\},$$

random variables $X_{i,j}(t)$ and $X_{i,j}(s)$ are not necessarily independent. For example, the autocovariance function $\mathrm{Cov}(X)(t,s)$ does not need to be a zero function.

In the context of functional ANOVA, the interval I is fixed and the same for all random functions. Therefore, we will denote random functions as $X(t)$. The rest of the notation that was introduced in this section will persist in all tests that deal with functional ANOVA.

Example

We will use an example from [Mrkvička et al., 2018]. In years 1979 - 2014, the temperature of water at the water level of Římov reservoir in the Czech Republic was measured every day. Let us denote the temperatures as $Y_1(t), \ldots, Y_{36}(t)$ for t spanning one year. Temperatures are clearly real functions of time, which we have represented in a discretised form as a vector of 365 elements. We would like to know whether the mean of the water temperature has changed over the 36 years. We divide the temperatures into three groups of 12 measured years, which we consider as independent and identically distributed, as well as independent of the other groups.

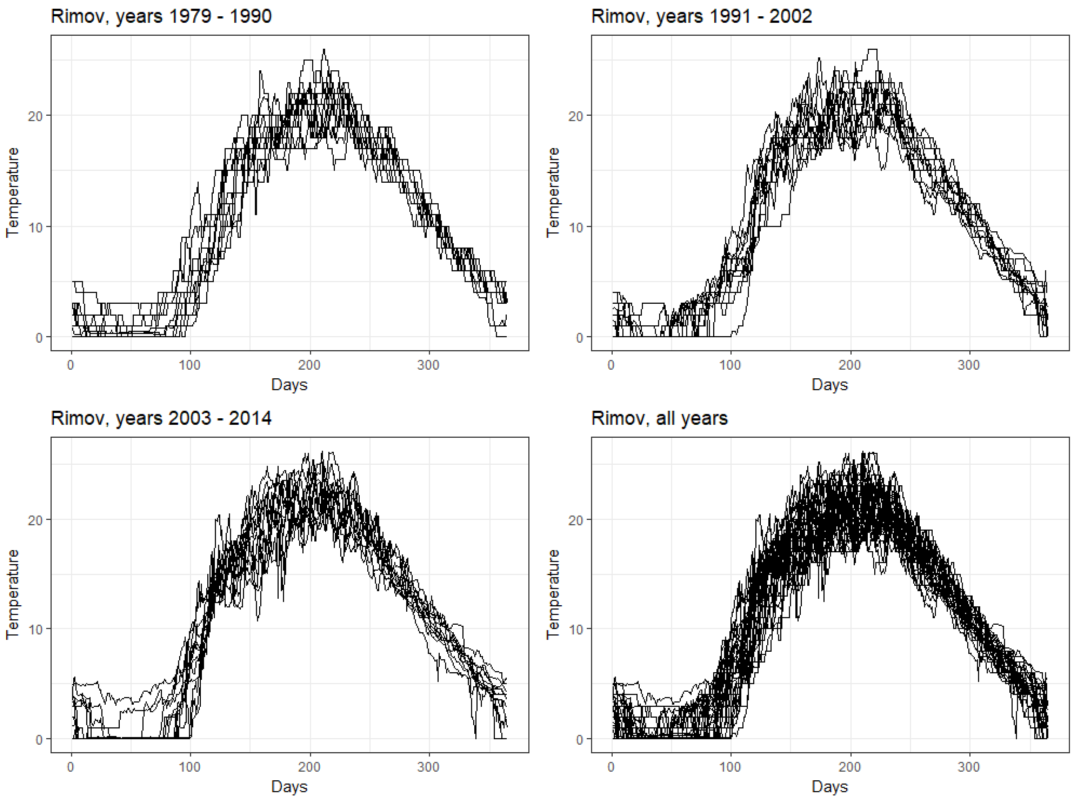

Figure 2.1: Water temperature in Římov water reservoir.

We now have three groups of random samples, which we denote as:

$$X_{1,1}(t), \ldots, X_{1,12}(t),$$
$$X_{2,1}(t), \ldots, X_{2,12}(t),$$
$$X_{3,1}(t), \ldots, X_{3,12}(t).$$

The null hypothesis is:

$$H_0 : \mu_1(t) = \mu_2(t) = \mu_3(t),\ \forall t \in I.$$

Thus, we have arrived at the problem of functional ANOVA.

2.2.1 Asymptotic F-test

We will now describe a simple functional one-way ANOVA test, which was introduced in [Cuevas, 2004]. It is a direct extension of the univariate ANOVA F-test.

When dealing with the problem of functional ANOVA, one might consider extending the test statistic of the univariate ANOVA F-test to the functional case. One way around the issue of multidimensionality would be by performing the F-test on every point of a discretised interval and applying Bonferroni's inequality. However, if the discretised interval contains too many points, the resulting test is very conservative – even if the null hypothesis holds, on average, it is still rejected in more than $(100 \cdot \alpha)\%$ of the cases. The other way could be by applying the $\mathcal{L}_2$-norm:

$$\|X\|_2 := \Big(\int_a^b X^2(t)dt \Big)^{\frac{1}{2}}$$

on whole functions instead of using the sums of squares in a pointwise fashion. This has been a subject of study in [Cuevas, 2004].

We get the following test statistic:

$$F_N = \frac{\frac{\sum_{i=1}^K n_i \|\overline{X}_{i,+} - \overline{X}_{+,+}\|^2}{K-1}}{\frac{\sum_{i=1}^K \sum_{j=1}^{n_i} \|X_{i,j} - \overline{X}_{i,+}\|^2}{N-K}}.$$

The difficulty with this approach is that F_N has a very complicated distribution. Therefore, the idea is to choose an asymptotic test, in which the distribution can be considerably simplified.

We will utilise one important idea behind ANOVA – the comparison of the 'external' sums of squares against the 'internal' sums of squares. Asymptotically, the denominator of F_N almost surely approaches the overall variance of the samples

$$\sigma^2 := \int_a^b [X_{i,j}(t) - \mu(t)]^2 dt.$$

Therefore, we can use the nominator as our test statistic, which shall reflect how much the group means differ from the overall mean and, therefore, from each other.

Due to some technical reasons, [Cuevas, 2004] use a slighly different test statistic, denoted as

$$V_N := \sum_{i<j}^{K} n_i \|\overline{X}_{i,+} - \overline{X}_{j,+}\|_2^2,$$

which converges to the same distribution as the original one. Notice that if H_0 is not true, V_N tends to infinity for $N \to \infty$. We will now define a Gaussian process, which appears in the theorem that deals with the asymptotic distribution of V_N.

Definition 2.2. Gaussian process is a random function $Z(t),\ t \in I$, such that for any finite subset $\{t_1, \dots, t_n\} \subset I,\ n \in \mathbb{N}$, the random vector $(Z(t_1), \dots, Z(t_n))$ has a multivariate normal distribution.

The theorem that follows gives us the asymptotic distribution of V_N. Notice that it does not assume homoscedasticity.

Theorem 2.1. *Let* $N, n_i \to \infty,\ \frac{n_i}{N} \to p_i \in \mathbb{R}\ \forall i \in \{1, \dots, K\}$. *Assume that* $X_{i,j},\ i \in \{1, \dots, K\},\ j \in \mathbb{N}$ *are random functions in* I *with finite variances and*

$$\begin{aligned} \mathbb{E}X_{i,j}(t) &= 0,\ \forall t \in I, \\ \mathrm{Cov}(X_{i,j})(s,t) &=: K_i(s,t),\ \forall t,s \in I,\ \forall j \in \{1, \dots, n_i\}. \end{aligned}$$

Then, under the null hypothesis of functional ANOVA:

$$V_N \xrightarrow[N\to\infty]{d} V := \sum_{i<j}^{K} \|Z_i(t) - C_{i,j} Z_j(t)\|_2^2,$$

where $C_{ij} = \sqrt{\frac{p_i}{p_j}}$ *and* $Z_1(t), \dots, Z_K(t)$ *are independent Gaussian processes with*

$$\begin{aligned} \mathbb{E}Y_i(t) &= 0,\ \forall t \in I, \\ \mathrm{Cov}(Y_i)(s,t) &= K_i(s,t),\ \forall t,s \in I. \end{aligned}$$

Proof. See [Cuevas, 2004], Theorem 1, page 114. □

Notice that we assume $\mathbb{E}X_{i,j}(t) = 0,\ \forall t \in I$. Under the null hypothesis, we can transform $X_{1,1}(t), \dots, X_{K,n_K}(t)$ easily by substracting the overall sample mean $\overline{X}_{+,+}(t)$ for all points on I.

The test itself is performed in the following way:

1. We discretise the interval I into a finite set of points R and calculate values of the random samples $X_{1,1}(t), \dots, X_{n_K,K}(t)$ for every $t \in R$. In practice, the random sample is rarely continuous.

2. We calculate $\hat{K}_i(s,t)$, which is a consistent estimate of $K_i(s,t)$:

$$\hat{K}_i(s,t) = \sum_{j=1}^{n_i} \frac{(X_{i,j}(s) - \overline{X}_{i,+}(s))(X_{i,j}(t) - \overline{X}_{i,+}(t))}{n_i - 1}.$$

   If we assume the covariances between groups to be equal, we can enhance the calculations by using an overall estimate.

3. We choose a number of simulations S and simulate each of the K functions $Z_i(t)$ S times on the discretised interval. Thus, we get

$$\{\hat{Z}(t)_{i,l}, t \in R\},\ i \in \{1, \dots, K\},\ l \in \{1, \dots, S\}.$$

4. For all i and r, we approximate $\|Z_i(t) - C_{i,j} Z_j(t)\|_2$ S times by

$$\|\hat{Z}_{i,l}(t) - C_{i,j} \hat{Z}_{j,l}(t)\|_e,\ l \in \{1, \dots, S\},$$

   where $\|\cdot\|_e$ is the Euclidean norm.

5. We calculate $\hat{V}_l \coloneqq \sum_{i<j}^{K} \|\hat{Z}_{i,l} - C_{i,j}\hat{Z}_{i,l}\|_e^2$.

6. For a chosen level of confidence α, we calculate an empirical quantile $\hat{V}_\alpha$ from the random sample $\hat{V}_1, \ldots, \hat{V}_S$, and then compare the value with V_N. If $V_N > \hat{V}_\alpha$, we reject the null hypothesis.

This test has a major disadvantage, which it shares with most tests for functional ANOVA – it lacks a graphical interpretation. The output of this test is binary and does not provide us with any information regarding the contribution of specific groups to the rejection.

In the next chapter, we describe a rank envelope test, based on which we construct a functional ANOVA test with a graphical interpretation in Chapter 4.

3. Envelope Tests

This chapter introduces the concept of envelopes, along with an envelope test that is used in practice and a novel envelope test from the article by [Myllymäki et al., 2016]. We will use them to test if the distribution of a random function is equal to the distribution of a random sample of functions. Even though this test appears to depart from the topic of this thesis, it will eventually be used in Chapter 4 to construct a functional ANOVA test with graphical interpretation.

3.1 Global envelopes

One of the most popular methods for testing equality of distributions on functional data are the envelope Monte Carlo tests.

Assume that we have a random function $X_1 \sim \mathbb{P}_1$. In this chapter, we will test a hypothesis $H_0 : X_1 \sim \mathbb{P}_0$, for some distribution $\mathbb{P}_0$ that is known either analytically (as a specific distribution with known parameters) or empirically (as a sample distribution), so that we can simulate s functions $X_2, \dots, X_{s+1} \sim \mathbb{P}_0$. We now define an envelope along with the concept of an *envelope test*.

Definition 3.1. Let $X_2, \dots, X_{s+1} \sim \mathbb{P}_0$ be a random sample of functions and let $k \in \{1, \dots, s\},\ t \in I$.

We call the random function $X_{low}^{(k)}(t)$ a *lower k-th envelope*, if

$$\forall t \in I : X_{low}^{(k)}(t) = X(t)_{(k)},$$

where $X(t)_{(k)}$ is the k-th order statisic of the random variables $X_2(t), \dots, X_{s+1}(t)$ for a fixed $t \in I$. Similarly, we call the random function $X_{upp}^{(k)}(t)$ an *upper k-th envelope*, if

$$\forall t \in I : X_{upp}^{(k)}(t) = X(t)_{(s-k)}.$$

We define a function $X_1(t),\ t \in I$ to be inside the k-th envelope, if:

$$\forall t \in I : X_1(t) \in [X_{low}^{(k)}(t), X_{upp}^{(k)}(t)].$$

Definition 3.2. Let $X_2, \dots, X_{s+1} \sim \mathbb{P}_0$ be a random sample of functions and let $k \in \{1, \dots, s\}$.

We call a test a *global envelope test*, if the null hypothesis is rejected if and only if the observed function X_1 is not *completely* covered by the k-th envelope for some fixed $k \in \{1, \dots, s\}$, i.e.:

$$\exists t_0 \in I : X_1(t_0) \notin [X_{low}^{(k)}(t_0), X_{upp}^{(k)}(t_0)].$$

We will now define an envelope test first introduced in [Ripley, 1977], from which we will proceed towards more sophisticated tests.

3.1.1 Ripley's envelope test

Let $X_1 \sim \mathbb{P}_1$ be a random function and let $X_2, \dots, X_{s+1} \sim \mathbb{P}_0$ be simulated random functions. Let H_0 be a null hypothesis defined as above. In Ripley's simulation envelope test, we choose

$$X_{low}^{(1)}(t) = \min_{i=2,\dots,s+1} X_i(t),$$
$$X_{upp}^{(1)}(t) = \max_{i=2,\dots,s+1} X_i(t).$$

Notice that under the null hypothesis, functions $X_1, X_2, \dots, X_{s+1}$ are independent and identically distributed. Therefore, under the assumption that

$$\forall t \in I,\ \forall i, j \in \{1, \dots, s+1\}:\ \mathbb{P}(X_i(t) = X_j(t)) = 0$$

and for a fixed $t \in I$, we can use Theorem 1.1. Thus, we get

$$\mathbb{P}(R_1(t) = k) = \frac{1}{s+1}, \forall k \in \{1, \dots, s+1\}.$$

A *pointwise* envelope test would then be based on the rule that the null hypothesis is rejected if and only if the observed random variable $X(t)$ falls outside of the interval $[X_{low}^{(1)}(t), X_{upp}^{(1)}(t)]$. For a fixed number of simulations s, we get an exact significance level. In other words, the following is true under the null hypothesis:

$$\mathbb{P}(X_{low}^{(1)}(t) \le X_1(t) \le X_{upp}^{(1)}(t)) = \mathbb{P}(X_{low}^{(1)}(t) < X_1(t) < X_{upp}^{(1)}(t)) =$$
$$= \mathbb{P}(1 < R_1(t) < s+1) = \sum_{k=2}^{s} \mathbb{P}(R_1(t) = k) = \frac{s-1}{s+1},\ \forall t \in I.$$

The significance level would then be the probability of the opposite event:

$$\mathbb{P}_{H_0}(X_1(t) \notin [X_{low}^{(1)}(t), X_{upp}^{(1)}(t)]) = 1 - \frac{s-1}{s+1} = \frac{2}{s+1}.$$

We can further enhance the test by choosing a number k and the envelopes as

$$X_{low}^{(k)}(t) \dots\ k\text{-th lowest } X_i(t),\ i = 2, \dots, s+1,$$
$$X_{upp}^{(k)}(t) \dots\ k\text{-th highest } X_i(t),\ i = 2, \dots, s+1.$$

Now, it becomes possible to choose a significance level for a fixed number of simulations. In other words, the significance level becomes a function $\alpha(s, k)$. The resulting significance level is $\alpha(s, k) = \frac{2 \cdot k}{s+1}$, which we can calculate in the same way as above. We also need to choose such a number of simulations that the number $k = \frac{\alpha(s+1)}{2}$ is an integer.

Notice that the pointwise envelope test only provides a significance level for a fixed $t \in I$. A global version of the test would need to have a single significance level on the whole interval I. Since we would be dealing with the problem of multiple comparison, a test using the pointwise envelope test on all points $t \in I$ would have a significance level much higher than the established α. We get an estimator of its type I error based on the results of the test, as is stated in [Loosmore and Ford, 2006].

Theorem 3.1. *Let s be a fixed number of simulations and r be the number of those simulations $X_i,\ i \in \{2, \dots, s+1\}$, for which the equation holds:*

$$X_{low}^{(1)}(t) \le X_i(t) \le X_{upp}^{(1)}(t)\ \forall t \in I.$$

The number $1 - \frac{r}{s}$ is an unbiased estimator of the type I error of the global simulation envelope test.

Proof. We can rewrite r as a function of $X_2, \dots, X_{s+1}$:

$$r(X_2, \dots, X_{s+1}) = \sum_{i=2}^{s+1} \mathbf{1}\{X_{low}^{(1)}(t) \le X_i(t) \le X_{upp}^{(1)}(t),\ \forall t \in I\}.$$

Then, we can see that

$$\mathbb{E}\Big(\frac{r(X_2, \dots, X_{s+1})}{s}\Big) = \frac{1}{s} \cdot \sum_{i=2}^{s+1} \mathbb{P}(X_{low}^{(1)}(t) \le X_i(t) \le X_{upp}^{(1)}(t),\ \forall t \in I) =$$
$$= \mathbb{P}(X_{low}^{(1)}(t) \le X_i(t) \le X_{upp}^{(1)}(t),\ \forall t \in I).$$

Let H_0 be true. Then:

$$\mathbb{P}(H_0 \text{ is rejected}) = \mathbb{P}(\forall t \in I : X_1(t) \notin [X_{low}^{(1)}(t), X_{upp}^{(1)}(t)]) =$$
$$= \mathbb{P}(\exists t \in I : X_1(t) > X_{upp}^{(1)}(t) \ or \ X_1(t) < X_{low}^{(1)}(t)) =$$
$$= 1 - \mathbb{P}(\ \forall t \in I :\ X_{low}^{(1)}(t) \le X_1(t) \le X_{upp}^{(1)}(t)),$$

whose unbiased estimator is $1 - \frac{r}{s}$. □

We could compute an upper bound of the type I error by using Bonferroni's inequality. A test with such a significance level would, however, have a very low power.

The global version of the simulation test was originally designed as a post-hoc test, whose purpose was to identify points that caused a rejection of the null hypothesis by another test, unrelated to the former one. Nevertheless, it has been used as a stand-alone test with a significance level misinterpreted as α. In spite of its weakness, this test is graphically interpretable – we reject the null hypothesis if the observed function is not *completely contained* inside the envelope. This attribute makes the test very desirable for applications. Therefore, the ambition is to transform it into a test with an exact significance level and sufficiently high power.

3.1.2 Extension of a classical Monte Carlo test

We will begin by extending Theorem 1.1 from Chapter 1. We will need a stronger theorem that deals with any ordering satisfying certain conditions. Thus, we will get the necessary tools to create a test with a controlled significance level from an ordering of whole functions. The theorem becomes a simple and a very general method of constructing quantiles for functional data. First, we will need to extend Definition 1.2 to the context of random functions.

Definition 3.3. Let $n \in \mathbb{N}, (X_1, \dots, X_n)$ be a vector of random functions on I. We call random functions $X_1, \dots, X_n$ exchangeable, if

$$\mathbb{P}_{X_1,\dots,X_n}(A) = \mathbb{P}_{X_{\pi(1)},\dots,X_{\pi(n)}}(A),$$

for any measurable set $A \in (\mathcal{B}^I)^n$ and for any permutation on $\{1, \dots, n\}$.

Remark. *If $X_1, \dots, X_n$ is a random sample of functions, functions $X_1, \dots, X_n$ are exchangeable.*

Definition 3.4. By an ordering '$\prec$' on random functions $X_1, \dots, X_n,\ n \in \mathbb{N}$, we will mean an ordering on a set of (non-random) functions

$$X = \{X_i(\omega),\ \omega \in \Omega,\ i \in \{1, \dots, n\}\}$$

that satisfies the property:

$$\{(f, g) \in X^2 : f \prec g\} \in (\mathcal{B}^I)^2.$$

Theorem 3.2. *Let $R_1, \dots, R_{s+1}$ be the ranks of exchangeable functions $X_1, \dots, X_{s+1}$, based on an ordering $\prec$, such that*

$$\forall i, j \in \{1, \dots, s+1\},\ i \neq j : \ \mathbb{P}(X_i \prec X_j \text{ or } X_j \prec X_i) = 1. \tag{3.1}$$

Then,

$$\mathbb{P}(R_i = k) = \frac{1}{s+1}\ \forall i, k \in \{1, \dots, s+1\}.$$

Proof. (3.1) implies that we can almost surely order $X_1, \dots, X_{s+1}$ by '$\prec$' without any ties. Let $M \coloneqq \{(f_1, \dots, f_{s+1}) \in (\mathbb{R}^I)^{s+1}, f_1 \prec f_2 \prec \cdots \prec f_{s+1}\}$.

M is a measurable set, e.g. from $(\mathcal{B}^I)^{s+1}$, as it is a countable intersection of measurable sets. Thus,

$$\mathbb{P}(X_1 \prec X_2 \prec \cdots \prec X_{s+1}) = \mathbb{P}((X_1, \dots, X_{s+1}) \in M) =$$
$$= \mathbb{P}((X_{\pi(1)}, \dots, X_{\pi(s+1)}) \in M) = \mathbb{P}(X_{\pi(1)} \prec X_{\pi(2)} \prec \cdots \prec X_{\pi(s+1)}),$$

for any permutation $\pi : \{1, \dots, s+1\}$, where the second equality holds due to the exchangibility assumption.

There are $(s+1)!$ permutations on $\{X_1, \dots, X_{s+1}\}$, implying that

$$\mathbb{P}(X_1 \prec \cdots \prec X_{s+1}) = \frac{1}{(s+1)!}.$$

Also, realise that $X_i \prec X_j$ is equivalent to $R_i < R_j$. Next, let

$$P_k \coloneqq \{\pi : \{1, \dots, k-1, k+1, \dots, s+1\} \to \{1, \dots, k-1, k+1, \dots, s+1\},$$
$$\pi \text{ is a permutation}\}.$$

For a fixed $i, k \in \{1, \dots, s+1\}$, we have

$$\mathbb{P}(R_i = k) = \sum_{\pi \in P_k} \mathbb{P}(X_{\pi(1)} \prec \cdots \prec X_{\pi(k-1)} \prec X_k \prec X_{\pi(k+1)} \prec \cdots \prec X_{\pi(s+1)})$$
$$= \sum_{\pi \in P_k} \frac{1}{(s+1)!} = \frac{s!}{(s+1)!} = \frac{1}{s+1},$$

which concludes the proof. □

With an appropriate ordering, this theorem gives us a tool to create exact tests on a random sample of functions without knowing much about its properties.

We will continue in the same way as Myllymäki et al. [2016] in Section 3. Now, we shall create an ordering of *functions* by their *extremity.* First, let us assume that we already have such an ordering '$\prec$' and that this ordering satisfies the condition $\mathbb{P}(X_i \prec X_j \text{ or } X_j \prec X_i) = 1,\ \forall i, j \in \{1, \dots, s+1\},\ i \neq j$. We will examine the case that the null hypothesis holds. Then, our observed function X_1 is independent and of the same distribution as the simulated functions $X_2, \dots, X_{s+1}$. We get the following theorem:

Theorem 3.3. *Let $X_1, X_2, \ldots, X_{s+1}$ be a random sample of functions and let '$\prec$' be an ordering satisfying condition* (3.1)*. Let s be a number such that $\alpha(s+1)$ is an integer. Then, a test that rejects the null hypothesis according to the rule*

$$1 + \sum_{i=2}^{s+1} \mathbf{1}(X_1 \prec X_i) \leq \alpha(s+1)$$

is an exact test on a significance level α.

Proof. See Myllymäki et al. [2016], Section 3, Lemma 1. □

The power of this test obviously depends on the ordering '$\prec$'. Notice that we can very easily modify any ordering to satisfy the condition (3.1) by *arbitrarily* sorting every equivalence class created by the ordering. Of course, the power of the test depends on the relevance of this sorting to the parameters that we consider to be "extreme". If we utilise this partially arbitrary ordering, we may get different p-values for multiple tests (based on this ordering) performed on the same data. These p-values would, however, belong to a fixed interval, which we denote as $[p_-, p_+]$.

3.2 Rank envelope test

Based on the previous approach, we now need to construct an ordering '$\prec$' . We will construct it in the same manner as was used in [Myllymäki et al., 2016], Section 4. In order to achieve high power of the test, we will look for a measure of *'extremeness'* of random functions. The measure will be called the *extreme rank* of a function and the ordering will be naturally devised from the measure.

Definition 3.5. Let s be a fixed number of simulations and let $X_2, \ldots, X_{s+1} \sim \mathbb{P}_0$ and X_1 be random functions. Then, for every $i \in \{1, \ldots, s+1\}$, we define the extreme rank of function X_i as:

$$R_i := \max\{k : X_{low}^{(k)}(t) \leq X_i(t) \leq X_{upp}^{(k)}(t),\ \forall t \in I\}.$$

The extreme rank can be interpreted in the following way:

If $R_i = k$ is large, the corresponding envelope $[X_{low}^{(k)}, X_{upp}^{(k)}]$ is constructed from high ranks of $X_j(t),\ j \in \{2, \ldots, s+1\}$. Thus, X_i is typical (unextreme).

We define the ordering '$\prec$' as follows:

$$\forall i, j \in \{1, \ldots, s+1\} : R_i < R_j \Rightarrow X_i \prec X_j.$$

Notice that the ordering '$\prec$' is weak – there is always a tie. For example, for a random sample of two and more functions, there are at least two functions with the extreme rank equal to 1. If this was not true, the only extreme function would have to be the upper and the lower envelope at the same time.

Therefore, we will need to resolve the ties with the observed function by deciding whether we want to consider the functions of the same rank as the observed function to be 'more extreme' or 'less extreme'. In the first case, we get a liberal test, whereas in the second case, we get a conservative test. In other words, the values

$$p_-(X_1) := \frac{1}{s+1} \sum_{i=1}^{s+1} \mathbf{1}(R_i < R_1),$$
$$p_+(X_1) := \frac{1}{s+1} \sum_{i=1}^{s+1} \mathbf{1}(R_i \leq R_1)$$

are the p-values of the liberal and the conservative rank envelope test, respectively. This result is formalised in the following theorem.

Theorem 3.4. *Let $s \in \mathbb{N}$ such that $\alpha(s+1) \in \mathbb{N}$. A test that rejects H_0 if and only if*

$$p_-(X_1) \leq \alpha \tag{3.2}$$

is a liberal test and a test that rejects H_0 if and only if

$$p_+(X_1) \leq \alpha \tag{3.3}$$

is a conservative test.

Proof. We will prove (3.2) in a similar fashion as in [Myllymäki et al., 2016]. We need to check if the probability of rejection is (under the null hypothesis) greater than α. The null hypothesis $H_0 : X_1 \sim \mathbb{P}_0$ holds, therefore $X_1, \dots, X_{s+1}$ is a random sample of functions, thus exchangeable.

We introduce new ranks $\tilde{R}_1, \dots, \tilde{R}_{s+1}$, so that $\mathbb{P}(\tilde{R}_i < \tilde{R}_j \textit{ or } \tilde{R}_j < \tilde{R}_i) = 1$. In other words, $\forall i, j \in \{1, \dots, s+1\}$:

$$R_i < R_j \Rightarrow \tilde{R}_i < \tilde{R}_j,$$
$$R_i = R_j \Rightarrow \text{ we order } \tilde{R}_i, \tilde{R}_j \text{ arbitrarily, so that } \tilde{R}_i \neq \tilde{R}_j,$$
$$\{\tilde{R}_1, \dots, \tilde{R}_{s+1}\} = \{1, \dots, s+1\}.$$

Thus, we get an ordering based on ranks $\tilde{R}_1, \dots, \tilde{R}_{s+1}$ that satisfies (3.1).

According to Theorem 3.2, the new rank of X_1 has a uniform distribution across $\{1, \dots, s+1\}$:

$$\mathbb{P}(\tilde{R}_1 = k) = \frac{1}{s+1}, \ \forall k \in \{1, \dots, s+1\}.$$

We know that $\{\tilde{R}_1, \dots, \tilde{R}_{s+1}\} = \{1, \dots, s+1\}$. Trivially, the following is true:

$$\forall k \in \{1, \dots, s+1\} : \sum_{i=1}^{s+1} \mathbf{1}(\tilde{R}_i < k) < k.$$

The probability of rejection is:

$$\mathbb{P}_{H_0}(p_-(X_1) \leq \alpha).$$

Then,

$$\begin{aligned}
\mathbb{P}_{H_0}(p_-(X_1) \le \alpha) &= \sum_{l=1}^{s+1} \mathbb{P}\Big(\frac{1}{s+1}\sum_{i=1}^{s+1} \mathbf{1}(\tilde{R}_i < \tilde{R}_1) \le \alpha \mid \tilde{R}_1 = l\Big) \cdot \mathbb{P}(\tilde{R}_1 = l) > \\
&> \sum_{l=1}^{s+1} \mathbb{P}\Big(\frac{\tilde{R}_1}{s+1} \le \alpha \mid \tilde{R}_1 = l\Big) \cdot \mathbb{P}(\tilde{R}_1 = l) = \\
&= \sum_{l=1}^{s+1} \mathbf{1}(l \le \alpha(s+1)) \cdot \mathbb{P}(\tilde{R}_1 = l) = \\
&= \sum_{l=1}^{\alpha(s+1)} \mathbb{P}(\tilde{R}_1 = l) = \alpha,
\end{aligned}$$

where the second equality holds due to the fact that

$$\begin{aligned}
\mathbb{P}(\tilde{R}_1 \le \alpha(s+1) \mid \tilde{R}_1 = l) &= 1 \text{ if } l \le \alpha(s+1), \\
&= 0 \text{ if } l > \alpha(s+1).
\end{aligned}$$

The case of (3.3) is proved analogically. □

3.2.1 Graphical interpretation

The main advantage of this test is its graphical interpretation, from which we can see how much each point $t \in I$ contributes to rejection. Based on the results of the rank envelope test, we will now create an envelope. Let $\alpha \in (0, 1)$ be the significance level of the test. Then, for

$$k_\alpha := \max_{k \in \mathbb{N}} \{\sum_{i=1}^{s+1} \mathbf{1}(R_i < k) \le \alpha(s+1)\},$$

we consider the k_α-th envelope, consisting of $X_{low}^{(k_\alpha)}$ and $X_{upp}^{(k_\alpha)}$. Notice that the envelope covers those functions whose extreme rank is greater than or equal to k_α and it may include the observed function X_1. At most $\alpha(s+1)$ functions stray out of the envelope, meaning that they are for at least one point $t \in I$ *strictly* above the upper envelope or *strictly* below the lower envelope.

Generally, the interval $[p_-, p_+]$ represents the result of the rank envelope test. Assume that there are no pointwise ties with probability 1. We need this assumption so that the pointwise rank is uniquely defined, from which we know that for all points $t \in I$, both the lower and the upper envelope coincides in t with just one function almost surely.

If $\alpha \notin [p_-, p_+)$, we have a clear result of the test:

- If $p_+ \le \alpha$, we clearly reject the null hypothesis. In graphical interpretation, the observed function strays out of the envelope, as there are $p_+(s+1) \le \alpha(s+1)$ functions from $X_1, \ldots, X_{s+1}$ whose rank is lower than or equal to the rank of X_1.

- If $p_- > \alpha$, there is no evidence for rejection of the null hypothesis. In graphical interpretation, the observed function is completely covered by the envelope and does not coincide with the envelope, as there are $p_-(s+1) > \alpha(s+1)$ functions whose rank is lower than the rank of X_1.

If $p_- \leq \alpha < p_+$, we do not have a clear result. In graphical interpretation, the observed function X_1 coincides with the envelope at some point $t \in I$. The observations are proved in [Myllymäki et al., 2016], Theorem 4.2.

There are several solutions to the lack of a single p-value, which is necessary for the case when $\alpha \in [p_-, p_+)$. Firstly, we can either choose a liberal or a conservative test. Secondly, we can choose the middle value

$$p_{mid} := \frac{p_- + p_+}{2}.$$

This corresponds to the case where one half of the functions of the same extreme rank as the observed function is considered as 'more extreme' and one half as 'less extreme'.

We can also introduce a stronger ordering – for example, for a random function X_i with rank $R_i = k$, we could count all the points where the function is part of the k-th envelope and order it with functions of the same rank based on this number. This notion is further discussed in [Myllymäki et al., 2016], Section 6.

4. A functional ANOVA using rank envelopes

In this chapter, we will describe a test for functional one-way ANOVA that was introduced in [Mrkvička et al., 2018]. We will employ the rank envelope test, which has been developed in Chapter 3. The test will thus be graphically interpretable, compared to the other functional ANOVA tests that have been covered in Chapter 2.

4.1 Rank envelope one-way ANOVA test

The rank envelope test is a test of equality of distributions on functional data. Our goal is to apply this test to the problem of functional ANOVA, which tests equality of means (non-random functions) in a group of random samples.

We will be using the same notation as in Chapter 2. For a fixed $i \in \{1, \ldots, K\}$, we have a random sample of n_i random functions $X_{i,j}(t),\ j \in \{1, \ldots, n_i\},\ t \in I$, where I is an interval.

The ambition is to perform one rank envelope test that would test the equality of all mean functions at the same time. This would result in a single overall level of significance, which is necessary for a statistical test. As the observed function, we will use a test statistic created as a function of group sample means. For the application of the rank envelope test, we also need to 'simulate' the test statistic. The lack of simulations of the test statistic will be resolved by assuming the null hypothesis to be true and applying the function of group sample means on permutations of $X_{i,j}$.

4.1.1 Test vectors

We shall use two variants of test statistics. Recall the null hypothesis:

$$H_0 : \mu_i(t) = \mu_j(t) = \mu(t),\ \forall i, j \in \{1, \ldots, K\},\ \forall t \in I.$$

We will use the fact that the mean of a random function is a (non-random) function on I to rewrite the null hypothesis H_0. Let $X_{i,j}$ be represented as:

$$X_{i,j}(t) = \mu(t) + \mu_i^*(t) + e_{i,j}(t),\ \forall t \in I,$$

where both μ and μ_i^* are (non-random) functions and $e_{i,j}$ is a random function, under the condition that

$$\sum_{i=1}^{K} \mu_i^*(t) = 0,\ \forall t \in I,$$

so that the overall mean function μ is uniquely identifiable.

In practice, functions $e_{i,j}$, which are called the *error functions*, are a combination of the error of measurement and the innate random nature of the phenomenon observed. We can now rewrite the null hypothesis in the following way:

$$H_0^{(1)} : \mu_i^*(t) = 0\ \forall i \in \{1, \ldots, K\},\ \forall t \in I.$$

We will now construct the first test statistic, which consists of group sample means:

$$\mathbf{X^{(1)}} = (\overline{X}_1, \dots, \overline{X}_K). \tag{4.1}$$

The test statistic is a single random function on the interval I repeated K times.

Notice that the null hypothesis $H_0^{(1)}$ is equivalent to the following one:

$$H_0^{(2)} : \mu_i^*(t) - \mu_j^*(t) = 0, \ \forall i, j \in \{1, \dots, K\}, \ \forall t \in I.$$

Based on this observation, we will construct the second test statistic, which consists of the differences between group sample means:

$$\mathbf{X^{(2)}} = (\overline{X}_1 - \overline{X}_2, \overline{X}_1 - \overline{X}_3, \dots, \overline{X}_{K-1} - \overline{X}_K). \tag{4.2}$$

From now on, we will only consider a discretised version of the interval I, which we denote as a sequence $t_1 < \cdots < t_L$ of length L. The discretisation is performed in the same fashion for every $X_{i,j}$. Thus, the first test vector $\mathbf{X^{(1)}}$ is of length $K \times L$, whereas the length of the second test vector $\mathbf{X^{(2)}}$ is $\frac{K(K-1)}{2} \times L$.

For $K > 3$, the second vector is longer than the first one. Therefore, $\mathbf{X^{(1)}}$ is useful if the number of groups is very large, whereas $\mathbf{X^{(2)}}$ is generally preferable, considering that it graphically illustrates the areas where any two group means differ significantly. In other functional ANOVA tests, this contrasting of group means is sometimes used as a post-hoc test – see Zhang [2013], Chapter 5, page 145.

4.1.2 Permutations

Let us now assume that we have a test vector $\mathbf{T}$, which is a function of the group sample means. If the null hypothesis is true, we know that all $X_{i,j}$ have the same mean, and so it is immaterial which functions $X_{i,j}$ we use to calculate a (group) sample mean. Under $H_0^{(1)}$, the sample mean $\overline{X}_{i,+}(t) = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{i,j}(t)$ is a consistent estimate of $\mu(t)$ for every i in $\{1, \dots, K\}$, $t \in \{t_1, \dots, t_L\}$, due to the strong law of large numbers.

Also, for a permutation (a bijection) $\pi(i, j) : A \to A$, where

$$A := \{[i, j] \in \mathbb{N} \times \mathbb{N} : 1 \leq i \leq K, 1 \leq j \leq n_i\},$$

the sample mean $\frac{1}{n_i}\sum_{j=1}^{n_i} X_{\pi(i,j)}$, where i is fixed, is a consistent estimate of μ as well. We construct a simulation T for a test vector $\mathbf{T}$ in the following way:

1. A permutation π is randomly generated.

2. Random sample $X_{i,j}$ is transformed by π into $X_{\pi(i,j)}$. We obtain a new random sample

$$\begin{gathered} X_{\pi(1,1)}(t), \dots, X_{\pi(1,n_1)}(t), \\ X_{\pi(2,1)}(t), \dots, X_{\pi(2,n_2)}(t), \\ \vdots \\ X_{\pi(K,1)}(t), \dots, X_{\pi(K,n_K)}(t). \end{gathered}$$

3. T is calculated as in (4.1) or (4.2), based on the type of test vector that was chosen.

This is repeated S times, where S is the number of simulations, which is a parameter that needs to be set before the execution of the test.

4.1.3 Correction for unequal variances

The method above is based on the assumption that the variances of groups are equal. If this was not so, the groups with higher variance would reach the extreme points of the rank envelope more frequently.

If it is not possible to assume homoscedasticity, we need to correct our observations. The first idea would be to standardise a function $X_{i,j}$ by transforming it into

$$\frac{X_{i,j}(t) - \mu(t)}{\sqrt{\mathrm{Var}(X_i)(t)}}.$$

We would, however, like to preserve its overall mean and variance, which is achievable by the following transformation:

$$S_{i,j}(t) := \frac{X_{i,j}(t) - \overline{X}_{+,+}(t)}{\sqrt{\hat{V}_i(t)}} \cdot \sqrt{\hat{V}(t)} + \overline{X}_{+,+}(t),$$

where $\overline{X}_{+,+}(t)$ denotes the overall sample mean and $\hat{V}_i, \hat{V}$ denote the sample variances of groups $i \in \{1, \ldots, K\}$ and the overall sample variance, respectively.

Thus, we obtain a scaled version $S_{i,j}$ of the original observation $X_{i,j}$.

Also, in order to smooth the resulting functions, we can use a moving average on the variances, as is further described in [Mrkvička et al., 2018].

4.2 Real case study

We have performed the functional ANOVA test on water temperature data from Římov, according to the example in Chapter 2. We assumed homoscedasticity and ran the functional ANOVA test on 10000 permutations.

Firstly, we ran the test using a test vector of the first type, $\mathbf{X^{(1)}}$. The resulting p-interval was $(0, 0.085)$, by choosing the middle value of the interval as the result: 0.043, we have rejected the null hypothesis.

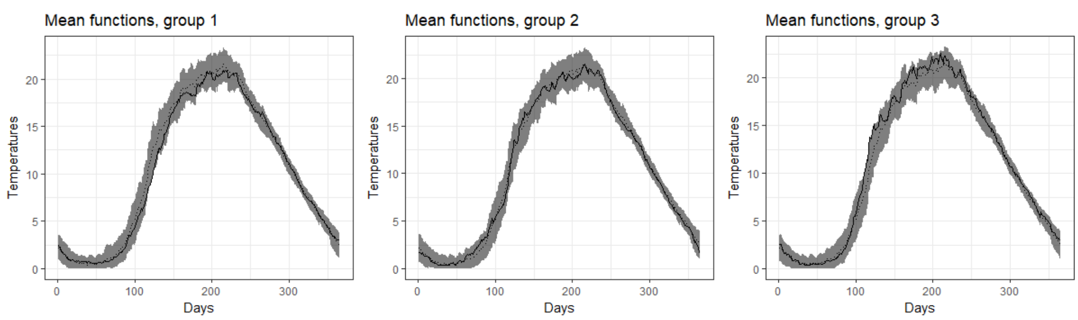

Figure 4.1: Mean functions of water temperature in the Římov reservoir dataset.

The barely visible coincidence with the envelope is around Day 120 (beginning of May) in the graph representing the mean function of the first group and the graph representing the third group. The black line depicts the observed function, whereas the grey area represents the global envelope.

Secondly, we ran the test using a test vector of the second type, $\mathbf{X^{(2)}}$. The resulting p-interval was $(0, 0.089)$ and, by choosing the middle value of the interval as the result: 0.045, we have rejected the null hypothesis.

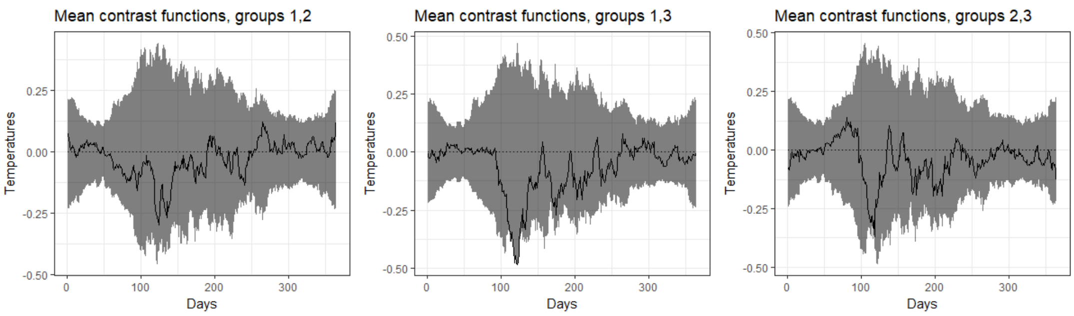

Figure 4.2: Comparison of mean functions of water temperature in the Římov reservoir dataset.

In the graphical output of the second test (4.2), we can see the locations that were responsible for rejection of the null hypothesis more clearly. The difference of means between the first and the third group, which is depicted as the observed black line, coincides with the envelope around Day 120. These results seem appropriate, as the largest difference in time is between the first and the third group of temperatures.

4.3 Simulation study

For mean functions, we have chosen the same four models M_1, M_2, M_3, M_4 as Cuevas [2004] or Mrkvička et al. [2018], on an interval $I = [0, 1]$ uniformly discretised into 100 points $0 = t_1 < \cdots < t_{100} = 1$. For the error functions, we have chosen:

- Gaussian process $E_1(t_i) \sim N(0, \sigma^2)$ that is independent for all points,
- Brownian process, defined for t_1 as $E_2(t_1) \sim N(0, \sigma^2)$ and recursively for $t_i,\ 1 < i \leq 100$ as a sum of $E_2(t_{i-1})$ with an independent random variable of the distribution $N(0, \sigma^2)$.

We have used standard deviations $\sigma \in \{0.05, 0.1, 0.5, 1, 2, 5\}$, which include deviations larger than those considered in [Mrkvička et al., 2018]. The same realisations of the error functions are used across all models of mean functions. For all models, we simulate $K = 3$ groups, $n_i = 10$ functions. The models for mean functions are:

- $M_1 : X_{i,j}(t) = t(1-t) + e_{i,j}(t),\ i \in \{1, \ldots, 10\},\ j \in \{1, 2, 3\}$,
- $M_2 : X_{i,j}(t) = t^i(1-t)^{6-i} + e_{i,j}(t),\ i \in \{1, \ldots, 10\},\ j \in \{1, 2, 3\}$,

- $M_3 : X_{i,j}(t) = t^{i/5}(1-t)^{6-\frac{i}{5}} + e_{i,j}(t),\ i \in \{1, \dots, 10\},\ j \in \{1, 2, 3\}$,
- $M_4 : X_{i,j}(t) = 1 + \frac{i}{50} + e_{i,j}(t),\ i \in \{1, \dots, 10\},\ j \in \{1, 2, 3\}$.

In the case of M_1, the null hypothesis is true. In the cases of M_2, M_3 and M_4, the alternative hypothesis is true. The means of M_2 and M_3 peak in different locations according to their group identity and M_4 has a constant mean which is slightly different for all groups. The mean functions are visualised in Figure 4.3.

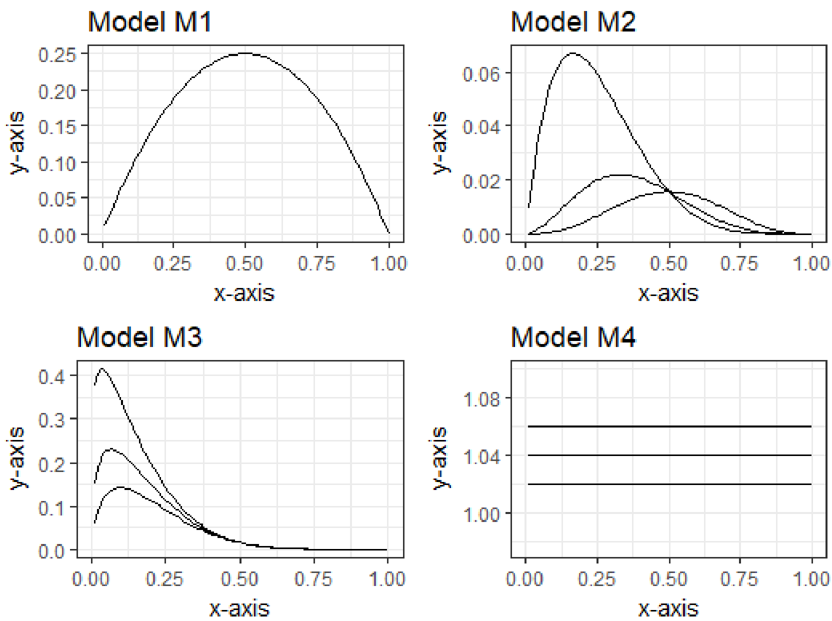

Figure 4.3: Mean function models (non-random).

We have compared three functional ANOVA tests:

- asymptotic F-test by [Cuevas, 2004], defined in Section 2, which we further denote as (AsF),
- the graphical functional ANOVA test by [Mrkvička et al., 2018] with the test vector $\mathbf{X}^{(1)}$, defined in Section 4, which we denote as (GFAM),
- the graphical functional ANOVA test with the test vector $\mathbf{X}^{(2)}$, which we denote as (GFAC).

For (GFAM) and (GFAC), [Mrkvička et al., 2018] use the extreme rank count method for calculating a unique p-value. We have used the liberal versions (using p_- as p-value), which we denote by suffix *-lib*, and the versions with p-value $p_{mid} = \frac{p_+ + p_-}{2}$, which we denote by suffix *-mid*. The p_{mid} version of the rank envelope test is somewhat weak in the context of functional ANOVA, as we usually have a long test vector with many ties. We have omitted the conservative version of the test, as it is even weaker than the p_{mid} version.

The number of simulations was 2000 for (GFAM) and (GFAC). For (AsF), we have chosen 300 simulations in order to achieve similar levels of time consumption.

For each model, error function and test, we have performed 100 repetitions of data generation and testing and calculated the number of times the null hypothesis was rejected. The results in the tables are the sample probabilities of rejection.

Table 4.1: Results for independent Gaussian process errors (sample prob. of rejection)

M_1	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	0.00	0.00	0.00	0.00	0.00	0.00
GFAM-lib	0.23	0.23	0.31	0.29	0.25	0.22
GFAM-mid	0.00	0.00	0.00	0.00	0.00	0.00
GFAC-lib	0.23	0.26	0.33	0.24	0.30	0.20
GFAC-mid	0.00	0.00	0.00	0.00	0.00	0.00
M_2	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	1.00	1.00	0.94	0.03	0.00	0.00
GFAM-lib	1.00	1.00	0.93	0.47	0.32	0.17
GFAM-mid	1.00	0.00	0.00	0.00	0.00	0.00
GFAC-lib	1.00	1.00	0.88	0.48	0.28	0.19
GFAC-mid	1.00	0.00	0.00	0.00	0.00	0.00
M_3	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	1.00	1.00	0.60	0.03	0.00	0.00
GFAM-lib	1.00	1.00	0.85	0.47	0.28	0.18
GFAM-mid	0.22	0.00	0.00	0.00	0.00	0.00
GFAC-lib	1.00	1.00	0.83	0.42	0.25	0.20
GFAC-mid	0.04	0.00	0.00	0.00	0.00	0.00
M_4	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	1.00	0.14	0.00	0.00	0.00	0.00
GFAM-lib	0.98	0.58	0.32	0.27	0.25	0.23
GFAM-mid	0.00	0.00	0.00	0.00	0.00	0.00
GFAC-lib	0.99	0.62	0.34	0.25	0.30	0.20
GFAC-mid	0.00	0.00	0.00	0.00	0.00	0.00

In the case of independent Gaussian errors, we can see in (4.3) that the asymptotic F-test fares very well. In model M_1, it does not reject the (true) null hypothesis. In models M_2, M_3, M_4, where the alternative is true, it rejects with a higher rate than p_{mid} versions of the graphical ANOVA tests. The liberal versions of the graphical ANOVA tests have a better rejection rate than (AsF) in cases with higher variances. The overall success of the (AsF) test is in part due to the assumption of normality, on which the original F-test relies. Notice that the p_{mid} versions are very conservative due to a high number of ties. This could be resolved by increasing the number of permutations. The difference in (GFAM) and (GFAC) is negligible, which is in part due to the fact that the length of both test vectors is identical for $K = 3$.

Table 4.2: Results for Brownian process errors (sample prob. of rejection)

M_1	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	0.05	0.06	0.04	0.06	0.05	0.11
GFAM-lib	0.07	0.05	0.05	0.05	0.09	0.05
GFAM-mid	0.07	0.04	0.04	0.04	0.06	0.05
GFAC-lib	0.08	0.04	0.07	0.07	0.06	0.06
GFAC-mid	0.05	0.02	0.05	0.02	0.06	0.04
M_2	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	0.42	0.13	0.04	0.06	0.05	0.11
GFAM-lib	0.45	0.14	0.05	0.04	0.09	0.05
GFAM-mid	0.42	0.11	0.05	0.04	0.06	0.05
GFAC-lib	0.46	0.13	0.08	0.07	0.06	0.06
GFAC-mid	0.39	0.09	0.06	0.03	0.05	0.04
M_3	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	0.31	0.08	0.04	0.06	0.05	0.11
GFAM-lib	1.00	1.00	0.13	0.06	0.08	0.05
GFAM-mid	1.00	1.00	0.12	0.03	0.05	0.05
GFAC-lib	1.00	1.00	0.12	0.07	0.07	0.05
GFAC-mid	1.00	1.00	0.09	0.04	0.06	0.04
M_4	$\sigma_1 = 0.05$	$\sigma_2 = 0.1$	$\sigma_3 = 0.5$	$\sigma_4 = 1$	$\sigma_5 = 2$	$\sigma_6 = 5$
AsF	0.04	0.05	0.04	0.06	0.05	0.11
GFAM-lib	0.14	0.06	0.06	0.05	0.08	0.05
GFAM-mid	0.12	0.05	0.05	0.04	0.05	0.05
GFAC-lib	0.15	0.05	0.05	0.07	0.06	0.06
GFAC-mid	0.12	0.02	0.04	0.02	0.06	0.04

In (4.3), we deal with a Brownian process error function. For the first model, where the null hypothesis holds, we can see that all tests reject the null hypothesis with a frequency being around the significance level α. The results of the second model appear to be balanced when compared with each other. In the third model, (AsF) test clearly struggles with rejection. The graphical tests, on the other hand, reject the null hypothesis for lower variances in all cases. Due to the incremental nature of the Brownian process, higher variances significantly influence the rejection rate. Thus, for higher variances, all models and tests have a rejection rate around the significance level α. Model M_4 suffers from this for deviations as low as $\sigma_2 = 0.1$ due to the fact that the mean functions differ by very little and Brownian error distorts this difference significantly.

Conclusion

In this Bachelor's Thesis, we have introduced the concept of functional data and the problem of functional one-way ANOVA, along with two functional one-way ANOVA tests. One of them represents the ANOVA tests that need a post-hoc evaluation to understand which groups caused rejection of the null hypothesis. The second test, whose development spans Chapters 3 and 4, is a test that does not suffer from this drawback. We have compared these tests in both a real case study and a simulation study.

This thesis brings a short introduction to functional data statistics. It rigorously describes how the novel functional one-way ANOVA test works and clarifies some deficiencies in the theory and the background behind this test. Namely, we prove the theorem concerning an estimator of the type I error for the global simulation envelope test, as well as the theorem for probability of ranks based on an ordering of random functions without ties, which is then used to prove the theorem that deals with the significance level of a test based on such an ordering. In the simulation study, models with previously unconsidered variances and p-values (p_{mid} and p_{lib}) were compared and analysed, using the R programming language of version 3.4.2, created by [R Core Team, 2017]. We have also used a package for the asymptotic F-test created by [Febrero-Bande and Oviedo de la Fuente, 2012] and a package from [Myllymäki et al., 2016].

The graphical functional one-way ANOVA test has not yet yielded many results, as it is a very recent addition to functional statistics. However, due to its applicable attributes inherited from the rank envelope test and its unique nature, we believe that it will find a prominent place in functional statistics and become an inspiration in the development of new ANOVA tests. The two-way variant of graphical functional ANOVA is also possible and under development by the authors of this test.

Bibliography

J. Anděl. *Statistické metody.* Druhé přepracované vydání. Matfyzpress, Praha, 1998. ISBN 80-85863-27-8.

N. H. Bingham and J. M. Fry. *Regression - Linear Models in Statistics*. Springer-Verlag London, London, 2010. ISBN 978-1-84882-968-8.

G. Casella and R. L. Berger. *Statistical Inference.* Second Edition. Duxbury-Thomson Learning, Duxbury, 2002. ISBN 0-534-24312-6.

A. Cuevas. An anova test for functional data. *Computational Statistics & Data Analysis*, 47:111 – 122, 2004.

Manuel Febrero-Bande and Manuel Oviedo de la Fuente. Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, 51(4):1–28, 2012. URL `http://www.jstatsoft.org/v51/i04/`.

N. B. Loosmore and E. D. Ford. Statistical inference using the g or k point pattern spatial statistics. *Ecology*, 87(8), 2006.

T. Mrkvička, U. Hahn, and M. Myllymäki. A one-way ANOVA test for functional data with graphical interpretation. *ArXiv e-prints*, February 2018.

M. Myllymäki, T. Mrkvička, P. Grabarnik, H. Seijo, and U. Hahn. Global envelope tests for spatial processes. *Journal of the Royal Statistical Society. Series B*, 79(2):381 – 404, 2016.

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

B. D. Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society. Series B*, 39(2):172 – 212, 1977.

J. Zhang. *Analysis of Variance for Functional Data.* CRC Press, New York, 2013. ISBN 978-1-4398-6273-5.