https://doi.org/10.15388/vu.thesis.463
https://orcid.org/0000-0001-5787-3855

VILNIUS UNIVERSITY

Tadas Danielius

Changepoint detection using p-variation of CUSUM process

DOCTORAL DISSERTATION

Natural Sciences,
Mathematics (N 001)

VILNIUS 2023

This dissertation was written between 2018 and 2022 at Vilnius University.

Academic supervisor – Prof. Habil. Dr. Alfredas Račkauskas (Vilnius University, Natural Sciences, Mathematics, N 001).
Academic consultant – Prof. Dr. Osvaldas Rukšėnas (Vilnius University, Natural Sciences, Biophysics, N 011).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Prof. Dr. Jonas Šiaulys (Vilnius University, Natural Sciences, Mathematics, N 001).

Members:
Assoc. Prof. Dr. Haeran Cho (Bristol University, Natural Sciences, Mathematics, N 001),
Assoc. Prof. Dr. Jurgita Markevičiūtė (Vilnius University, Natural Sciences, Mathematics, N 001),
Prof. Habil. Dr. Rimantas Norvaiša (Vilnius University, Natural Sciences, Mathematics, N 001),
Prof. Dr. Povilas Treigys (Vilnius University, Natural Sciences, Informatics, N 009).

The dissertation shall be defended at a public meeting of the Dissertation Defense Panel on June 6th, 2023, at 10:00 a.m. in Room 103 of the Faculty of Mathematics and Informatics at Vilnius University.
Address: Address: Naugarduko str. 24, LT03225, Vilnius, Lithuania
Tel.: +37052193050; e-mail: mif@mif.vu.lt.

The text of this dissertation can be accessed at the Vilnius University Library, as well as on the website of Vilnius University:
www.vu.lt/naujienos/ivykiu-kalendorius

https://doi.org/10.15388/vu.thesis.463
https://orcid.org/0000-0001-5787-3855

VILNIAUS UNIVERSITETAS

Tadas Danielius

Pasikeitimo taškų aptikimas naudojant dalinių sumų proceso p-variaciją

DAKTARO DISERTACIJA

Gamtos mokslai,
Matematika (N 001)

VILNIUS 2023

Disertacija rengta 2018–2022 metais Vilniaus universitete.

Mokslinis vadovas – prof. habil. dr. Alfredas Račkauskas (Vilniaus universitetas, gamtos mokslai, matematika, N 001).
Akademinis konsultantas – prof. dr. Osvaldas Rukšėnas (Vilnius University, gamtos mokslai, Biofizika, N 011).

Gynimo taryba:

Pirmininkas – prof. dr. Jonas Šiaulys (Vilniaus universitetas, gamtos mokslai, matematika, N 001).

Nariai:
prof. dr. Haeran Cho (Bristolio universitetas, gamtos mokslai, matematika, N 001),
doc. dr. Jurgita Markevičiūtė (Vilniaus universitetas, gamtos mokslai, matematika, N 001),
prof. habil. dr. Rimantas Norvaiša (Vilniaus universitetas, gamtos mokslai, matematika, N 001),
prof. dr. Povilas Treigys (Vilniaus universitetas, gamtos mokslai, informatika, N 009).

Disertacija ginama viešame Gynimo tarybos posėdyje 2023 m. birželio mėn. 6 d. 10:00 Vilniaus universiteto Matematikos ir informatikos fakultete, 103 auditorijoje.
Adresas: Naugarduko g. 24, LT03225, Matematikos ir informatikos fakultetas, Vilnius, Lietuva, tel.: +37052193050; el. paštas: mif@mif.vu.lt.

Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje ir VU interneto svetainėje adresu: www.vu.lt/naujienos/ivykiu-kalendorius

Contents

LIST OF ABBREVIATIONS

AMOC At Most One Change

CP Change point

CPA Change point Analysis

CPD Change point detection

CWT Continuous Wavelet Transform

ECG Electrocardiogram

EEG Electroencephalogram

FD Functional Data

FDA Funtional Data Analysis

fPCA Functional Principal Components Analysis

MDA Multivariate Data Analysis

MRI Magnetic resonance imaging

PCA Principal Components analysis

Introduction

With the rapid development of technology, great progress has been made in the digitization of data. Over the past decades, the amount and variety of data collected have grown exponentially. Data is collected in practically all areas: starting from physical phenomena observed in nature, physiological data in medicine, economic phenomena or smart watches that record the physiological data. Indeed, there is no field left untouched by the digitization revolution. Naturally, such rapid digitization raises many questions about data structure. One of the problems often encountered in data analysis is structural changes in time series. Changes may not be obvious or difficult to notice, therefore, solving such problems requires theoretically based mathematical instruments.

A change point, also known as a structural break or regime shift is a point in time at which the mean, variance, pattern, distribution, or other statistical property in time series changes abruptly or continuously. Change-points can occur in a variety of different contexts and can be caused by a wide range of factors, such as shifts in consumer behavior, changes in market conditions, or the introduction of new policies or regulations.

Change-points are often of interest in statistical analysis and data mining, as they can provide valuable insights into the underlying patterns and trends in the data. For example, detecting change-points in a time series data can help identify anomalies or irregularities in the data, and can be useful for predicting future trends or making informed decisions. Change point detection involves the analysis of two main questions:

1. Did the statistical characteristics of the data alter at any point in time?

2. If there was a change, when did it occur?

In certain situations, it is straightforward to identify structural changes in processes. For instance, the economy may experience a recession and then recover. The criteria that define a recession are well-defined and clearly stated. In other areas, it can be almost impossible to detect change points without using mathematical methods. Mathematical methods are particularly important when systems must automatically respond to a changed situation. For example, if the vibration level has changed, the engine must be shut off to prevent a crash. The analysis of change points has gained popularity with the emergence of smart devices, such as smart watches that can detect in real-time when a person starts running, climbing stairs, or falling asleep. Change point analysis is particularly relevant in the field of medicine, where medical measuring devices are used to continuously record data on a patient's physiological condition. This is important because detecting and reacting to changes as soon as

possible can help to prevent negative consequences. For example, if a medical device is able to detect that a patient's physiological condition has changed significantly, medical intervention can be provided in a timely manner to avoid serious health complications. One specific application of change point analysis is in the detection and real-time recording of epileptic seizures. As an example, Malladi et al. presented a method for real-time monitoring of epileptic seizures in their publication [46]. Another example where change point analysis is frequently employed in the field of medicine is when monitoring of cardiac activity through Electrocardiogram (ECG). To identify changes in the heart, various transformation methods, including wavelet transformations [53, 57], have been utilized. Hidden Markov models [41] and graph-constrained methods [25] are additional approaches that have been applied to ECG data. Fotoohinasab et al. provide a comprehensive overview of the different methods for change point analysis of ECG data in their publication [25].

Change point analysis is not limited to time series data. It is also frequently used to analyze images and audio signals. For example, magnetic resonance imaging produces 3D images that can be analyzed for change points [12, 47, 68]. In the case of audio recordings, change point analysis is often used to separate segments of speech from other sounds, which is important for tasks such as developing automatic speech recognition models, echo cancellation, and speech segmentation [35, 37, 73]. In many cases, the first step in solving these types of problems is to split the audio signal into distinct segments.

Overall, change point analysis is a useful technique for identifying changes in various types of data, including time series, images, and audio signals.

There is no universal definition of what constitutes a change. A change could be seen as a change in the data-generating model or model parameters, or it could be a change in distribution parameters such as mean (as shown in Fig. 1.4a in Chapter 1, section 1.3) or variance (as shown in Fig. 1.4b in Chapter 1, section 1.3). These types of changes are often considered in change point analysis. There are many methods that have been proposed for detecting changes in distributions (see [55]). However, other types of changes, such as changes in frequency and pattern, have received less attention. Changes in frequency (as shown in Fig. 1.5b in Chapter 1, section 1.3) are important for analyzing data with cyclical properties. These changes are often studied in the frequency domain using techniques such as the Fourier or wavelet transform (see [53, 57]). Detecting pattern changes (as shown in Fig. 1.5a in Chapter 1, section 1.3) is one of the most challenging tasks in change point analysis, and the coverage of methods for this problem in the scientific literature is not extensive. However, this is a relevant problem in areas such as brain wave analysis (see [72]).

The problem of change points has been extensively studied in the classical literature, and many detection algorithms have been proposed for one-

dimensional time series data. One of the first algorithms for change point detection is the Cumulative Sum (CUMSUM) algorithm, which was developed by Page in 1954 [50]. The CUMSUM algorithm was specifically designed to detect changes in the mean of a time series, and it has been widely used in quality control processes to ensure consistent production.

Change point detection methods can be classified into two main categories: parametric and nonparametric methods. Parametric methods assume a specific underlying probability distribution for the data, while nonparametric methods do not make such assumptions and instead use a range of heuristic and data-driven approaches to identify change points.

In change point detection, the concept of *online* and *offline* refers to the availability of the data being analyzed. *Offline* change point detection refers to the analysis of a complete, fixed dataset, while *online* change point detection, on the other hand, refers to the analysis of a dataset that is continuously growing.

In contrast, this thesis focuses on the *offline* problem of change point detection in a data sample that is invariant, meaning that the data set is fixed and not constantly growing.

Change point analysis is not limited to univariate time series data. In recent years, there has been a significant increase in research on change point detection in multidimensional time series data, which refers to data that has multiple variables measured over time. This type of analysis is useful in situations where there are multiple factors that may be influencing a system or process, and it can provide insights into how these factors are related and how they change over time.

Functional data analysis provides a natural framework for analyzing multivariate time series data. As a result, more and more classical methods are being adapted to work with functional data, which is a natural extension of multidimensional data from finite dimensions to infinite dimensions. In practice, functional data is often obtained by observing multiple subjects over time, space, or other continuous domains, and it can take the form of curves, surfaces, or other complex objects. Representing multidimensional data as functions greatly expands the range of analysis tools available, and functional principal components are often more informative than multidimensional principal components. As functional data analysis becomes more popular, the problem of detecting change points between curves (as shown in Fig. 1.6a and Fig. 1.6b, in Chapter 1, section 1.3) becomes increasingly relevant.

The first methods for analyzing functional data were mentioned as early as 1950, when Grenander (see [29]) attempted to apply statistical methods to stochastic processes. Later, in 1958, Rao (see [64]) used functional data analysis to compare the growth curves of organisms. The term "functional data analysis" itself was first coined by Ramsey in 1982 (see [61]). Extensive

discussions on the analysis and applications of functional data can be found in Ramsey and Silverman's book (see [63]), and a comprehensive review of the field is provided by Wang et al. (see [77]).

Aim and objectives

The primary objective of this dissertation is to investigate statistical techniques for the identification of change points in both univariate and functional samples. Specifically, the focus will be on proposing novel statistical tests based on the variation properties of the CUSUM process for change point detection.

To achieve the goal of the research, the following objectives have been formulated:

1. Define the objects and models under consideration.
2. Define and investigate a mean instability testing model based on the p-variation of the partial sum process.
3. Establish the limiting distribution of the test statistics under the null hypothesis and the contiguous alternative.
4. Define the $\mathcal{G}$-sum process and analyze its asymptotic behavior.
5. Construct statistical tests for change point detection in functional data based on the $\mathcal{G}$-sum process.
6. Analyze the proposed tests using simulation methods.
7. Apply the tests to real data.

Practical and scientific novelty

This thesis presents the development of novel statistical tests based on a p-variation of CUSUM process for the efficient detection of mean instabilities in both univariate and functional samples. The proposed tests include:

1. A test for At Most One Change (AMOC) in mean instabilities over time.
2. A test for detecting at most m change points.
3. A test for detecting an unknown number of change points.

The tests proposed in this thesis are effective for both large and small samples, and they can be practically applied in various fields such as medicine, image and sound analysis, climate change, and others. The tests are designed to detect changes in the mean, but the design of the tests can be generalized for other types of changes.

Layout of the dissertation

The dissertation consists of an summary, three chapters, general conclusions and a discussion of further research and a bibliography.

- **Chapter 1** introduces the reader to the main concept of function variation, as well as basic concepts of functional data analysis. It then focuses on change point problems, with an extensive review of available methods and their importance in different domains.
- In **Chapter 2**, a new method based on the p-variation properties of partial sums is proposed for detecting mean changes, and the theoretical aspects are examined. At the end of the chapter, the statistical power of the proposed test is evaluated through simulations and applied to real data.
- In **Chapter 3**, the proposed univariate test is generalized and adapted to functional data. The $\mathcal{G}$-sum and $\mathcal{G}$-cusum processes are defined and their asymptotic behavior is considered in the framework of the $\ell^\infty(\mathcal{G})$ space. This framework is used to derive the asymptotic distributions of the test statistics, and three tests are presented, following simulation studies. Finally, the tests are applied to real data.

Maintaining statements

- The limit distributions of the $\mathcal{G}$-sum and $\mathcal{G}$-cusum processes has been determined for random functions.
- A statistical test based on the p-variation of partial sums has been developed to detect changes in the mean.
- The asymptotic properties of $\mathcal{G}$-cusum process has been used for testing known and unknown number of change points in the mean of a functional sample.

Dissemination of results

Publications

[A1] T. Danielius, A. Račkauskas, p-Variation of CUSUM process and testing change in the mean, *Communications in Statistics-Simulation and Computation*, 1–13 (2020). https://doi.org/10.1080/03610918.2020.1844899

[A2] T. Danielius, A. Račkauskas, Multiple change point detection in functional sample via $\mathcal{G}$-sum process, *Mathematics* 10.13, (2022). https://doi.org/10.3390/math10132294

Conferences

[C1] T. Danielius. *Functional data analysis of neurophysiological data: case study.* NBBC19 : 7th Nordic-Baltic biometric conference, 3-5 June 2019, Vilnius, Lithuania.

[C2] T. Danielius, A. Račkauskas. *p-variation of cusum process and testing change in the mean*, 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2020) 19-21 December 2020, Virtual Conference.

[C3] T. Danielius, A. Račkauskas. *Multiple change point detection in functional sample via $\mathcal{G}$-sum process*, 63th Conference of the Lithuanian Mathematical Society, Kaunas University, 16-27 June 2022.

[C4] T. Danielius, A. Račkauskas. *Multiple change point detection in functional sample via $\mathcal{G}$-sum process*, 24th International Conference on Computational Statistics (CompStat 2022), 23-26 August 2022, Bologna, Italy.

[C5] T. Danielius. *Pasikeitimo taškų testai funkciniams duomenims paremti p-variacija*, Seminar Statistics and its applications, Vilnius University Institute of Applied Mathematics, 7 August 2022, Vilnius, Lithuania.

Main results

For a univariate sample $X_1, X_2, \ldots, X_n$ and the number $p > 2$, we define the test statistics

$$T_{p,n}(X_1,\ldots,X_n) = \max\left\{\sum_{j=1}^{m}\left|\sum_{i=k_{j-1}+1}^{k_j}(X_i-\overline{X}_n)\right|^p : 0=k_0<\cdots<k_m=n; 1\le m\le n\right\},$$

where $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$. It is important to review some necessary definitions as we continue to analyze these statistics. For a function $f : [0,1] \to \mathbb{R}$ and a number $0 < p < \infty$, the p-variation of f on the interval $[0,t]$ is

$$v_p(f;[0,t]) := \sup\left\{\sum_{j=1}^{m}|f(t_j)-f(t_{j-1})|^p\right\} \le +\infty,$$

where the supremum is taken over all partitions

$$0 = t_0 < t_1 < \cdots < t_m = t; \quad m = 1,2,\ldots,$$

of the interval $[0,t]$. We abbreviate $v_p(f) := v_p(f;[0,1])$. If $v_p(f) < \infty$ then we say that f has bounded p-variation, and $\mathcal{W}_p[0,1]$ is the set of all such functions.

The set $\mathcal{W}_p[0,1]$, $p \geq 1$, is a non separable Banach space with the norm

$$\|f\|_{[p]} := |f(0)| + v_p^{1/p}(f).$$

For each $n \geq 1$ and each $t \in [0,1]$, let

$$S_n(t) = 0 \text{ if } t \in [0, 1/n), \; S_n(t) := \sum_{i=1}^{\lfloor nt \rfloor} X_i, \text{ if } t \in [1/n, 1],$$

where for a real number $x \geq 0$, $\lfloor x \rfloor := \max\{k : k \in \mathbb{N}, k \leq x\}, \mathbb{N} = \{0, 1, \dots\}$. Then the cusum process $Z_n = (Z_n(t), t \in [0,1])$ is defined as

$$Z_n(t) = S_n(t) - \frac{\lfloor nt \rfloor}{n} S_n(1) = \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \overline{X}_n).$$

Classical examples of path spaces for the process Z_n include the Hilbert space $L_2[0,1]$ and the Skorohod space $D[0,1]$. Under various assumptions, a functional central limit theorem is established in these spaces. For example, the classical Donsker theorem for i.i.d. sequence (X_n) with finite second moment states that

$$n^{-1/2} Z_n \xrightarrow[n\to\infty]{\mathcal{D}} \sigma B \;\text{ in the space }\; D[0,1],$$

where $B = (B(s) := W(s) - sW(1), s \in [0,1])$ is a standard Brownian bridge, and $W = (W(s), s \in [0,1])$ is a standard Wiener process, and $\sigma^2 = \mathrm{var}(X_1)$.

The main result of the Chapter 2 is the following theorem

Theorem 1. Fix $p > 2$. Let $X_1, X_2, \dots$ be a sequence of independent identically distributed random variables and let $S_n = (S_n(t), t \in [0,1])$ be the partial sum process. If $\sigma^2 := EX_1^2 < \infty$, then the convergence

$$n^{-1/2} Z_n \xrightarrow[n\to\infty]{\mathcal{D}^*} \sigma B \;\text{ in }\; \mathcal{W}_p[0,1]$$

holds.

For independent random sample $X_1, \dots, X_n$ we consider the following model

$$X_i = \delta \mathbf{1}_{(k^*, n]}(i) + Y_i, \quad i = 1, \dots, n,$$

where $Y_1, \dots, Y_n$ are i.i.d. random variables with $E(Y_i) = 0$, $E(Y_i^2) = 1$ and $\delta \in R, k^* \in \{1, \dots, n\}$ are unknown parameters.

Under $H_0 : \delta = 0$, we prove that for any $p > 2$,

$$n^{-1/2} v_p^{1/p}(Z_n) \xrightarrow[n\to\infty]{\mathcal{D}} v_p^{1/p}(B).$$

Under contiguous alternative where $\delta = \delta_n \approx \sqrt{n}\delta^*$, and $k^* = \lfloor n\theta^* \rfloor$ with some

$\delta^* > 0$, and $\theta^* \in (0,1)$, we prove that

$$n^{-1/2} v_p^{1/p}(Z_n) \xrightarrow[n\to\infty]{\mathcal{D}} v_p^{1/p}(B - f),$$

where $f(t) = \begin{cases} \delta^* t(1-\theta^*) & \text{if } 0 \le t < \theta^* \\ \delta^* \theta^* (1-t) & \text{if } \theta^* \le t \le 1 \end{cases}$.

These two results are theoretical justification for statistics used to detect a change point in the mean. The power of the test was investigated by simulations.

Next, we consider a second-order stationary sequence of stochastic processes $Y_i = (Y_i(t), t \in [0,1]), i \in \mathbb{N}$, defined on a probability space $(\Omega, \mathcal{F}, P)$, having zero mean and covariance function $\gamma = \{\gamma(s,t), s,t \in [0,1]\}$. For a given functional sample $X_1(t), \ldots, X_n(t), t \in [0,1]$, we consider the model:

$$X_k(t) = g(k/n, t) + Y_k(t), \quad t \in [0,1], \quad k = 1, \ldots, n, \tag{1}$$

where the function $g : [0,1] \times [0,1] \to \mathbb{R}$ is deterministic, but unobserved. Our aim is to test the hypothesis:

$$H_0 : \quad g = 0 \quad \text{versus} \quad H_1 : g \neq 0$$

with emphasis on a case of change-point detection, which corresponds to a piecewise-constant function g with respect to the first argument.

This model covers a broad range of real-world problems such as climate change detection, image analysis, analysis of medical treatments, especially magnetic resonance images of brain activities, and speech recognition, to name a few. Besides, the change-point detection model (1) can be used for knot selection in spline smoothing as well as for trend changes in functional time series analysis.

The methodology we propose is based on some measures of variation of the process:

$$W_n(s) = \sum_{k=1}^{\lfloor ns \rfloor} (X_k - \overline{X}_n) + (ns - \lfloor ns \rfloor)(X_{\lfloor ns \rfloor + 1} - \overline{X}_n), \quad s \in [0,1],$$

where $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$.

Since this process is infinite-dimensional, we used the projections technique to reduce the dimension. To this aim, we assumed that Y_i is mean-squared continuous and jointly measurable and that γ has finite trace: $\operatorname{tr}(\gamma) = \int_0^1 \gamma(t,t)\,\mathrm{d}t < \infty$. In this case, Y_i is also an $L_2(0,1)$-valued random element, where $L_2 := L_2(0,1)$ is a Hilbert space of Lebesgue square integrable functions on $[0,1]$ endowed with the inner product $\langle f, g \rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t$ and the norm $\|f\| := \sqrt{\langle f, f \rangle}$. For two given sets $\mathcal{F}, \Psi \subset L_2$, we consider the $\mathcal{F} \times \Psi$-sum

process:

$$\nu_n = \Big(\sum_{k=1}^{n} \nu_{nk}(f,\psi), f \in \mathcal{F}, \psi \in \Psi \Big),$$

where $\nu_{nk}(f,\psi) = \langle X_k, \psi \rangle \lambda_{nk}(f)$, λ_{nk} is a uniform probability on the interval $[(k-1)/n, k/n]$ and $\lambda_{nk}(f) = \int_0^1 f(t)\,\mathrm{d}\lambda_{nk}(t)$. A natural framework for stochastic process ν_n is the space $\ell^\infty(\mathcal{G})$, where $\mathcal{G} = \mathcal{F} \times \Psi$. Recall for a class $\mathcal{G}$ that $\ell^\infty(\mathcal{G})$ is a Banach space of all uniformly bounded real-valued functions μ on $\mathcal{G}$ endowed with the uniform norm:

$$\|\mu\|_{\mathcal{G}} := \sup\{|\mu(g)| : g \in \mathcal{G}\}.$$

The limiting zero mean Gaussian process $\nu_\gamma = (\nu(f,\psi), f \in \mathcal{F}, \psi \in \Psi)$ is defined via covariance:

$$E\nu_\gamma(f,\psi)\nu_\gamma(f',\psi') = \mathcal{K}_\gamma((f,\psi),(f',\psi,)) := \langle \Gamma\psi, \psi' \rangle \langle f, f' \rangle, \quad \psi, \psi', f, f' \in L_2,$$

where $\Gamma : L_2 \to L_2$ is the covariance operator corresponding to the kernel γ.

Assumption 1. Random processes $Y, Y_1, Y_2, \ldots$ are i.i.d. mean square continuous, jointly measurable, with mean zero and covariance γ such that $\int_0^1 \gamma(t,t)\,\mathrm{d}t < \infty$.

For the model (1), we consider null hypothesis $H_0 : \ g = 0$ and two possible alternatives:

$$H_A : \ g = g_n = u_n q_n, \ \text{ where } \ u_n \to u \ \text{ in } \ \mathcal{W}_2[0,1], \ \ \sqrt{n} q_n \to q \ \text{ in } \ L_2,$$

and

$$H_A' : \ g = g_n = u_n q_n, \ \text{ where } \ u_n \to u \ \text{ in } \ \mathcal{W}_2[0,1], \ \ \sqrt{n} \sup_{\psi \in \Psi} |\langle q_n, \psi \rangle| \to \infty.$$

In both alternatives, the function u_n is responsible for the configuration of a drift within the sample, whereas the function q_n estimates a magnitude of the drift.

Theorem 2. Let the random processes (X_k) be defined by (1), where $Y, Y_1, Y_2, \ldots$ satisfy Assumption 1. Assume that, for some $1 \le q < 2$, the set $\mathcal{F} \subset \mathcal{W}_q[0,1]$ is bounded and the set $\Psi \subset L_2$ satisfies

$$\int_0^1 \sqrt{\log N(\varepsilon, \Psi, \rho)}\,\mathrm{d}\varepsilon < \infty. \tag{2}$$

Then, there exists a version of a Gaussian process ν_γ on $L_2 \times L_2$ such that its restriction on $\mathcal{F} \times \Psi$, $(\nu_\gamma(f,\psi), f \in \mathcal{F}, \psi \in \Psi)$ is continuous and the following hold:

(1a) Under H_0:

$$n^{-1/2}\nu_n \xrightarrow[n\to\infty]{\mathcal{D}} \nu_\gamma \quad \text{in} \quad \ell^\infty(\mathcal{F}\times\Psi). \tag{3}$$

(1b) Under H_A,

$$n^{-1/2}\nu_n \xrightarrow[n\to\infty]{\mathcal{D}} \nu_\gamma + \Delta, \quad \text{in} \quad \ell^\infty(\mathcal{F}\times\Psi), \tag{4}$$

where

$$\Delta(f,\psi) = \langle u, f\rangle\langle q,\psi\rangle.$$

If $u(s) = 1, s \in [0,1]$, then the alternative H_A corresponds to the presence of a signal in a noise. In this case, $\Delta(f,\psi) = \lambda(f)\langle q,\psi\rangle$. Therefore, the use of this theorem for testing a signal in a noise is meaningful provided $\langle q,\psi\rangle \neq 0$.

Assumption 2. The eigenvalues λ_r satisfy, for some $d > 0$,

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1} \geq 0.$$

One estimates Γ by

$$\widehat{\Gamma}_n := \frac{1}{n}\sum_{i=1}^{n}[(X_i - \overline{X}_n)\otimes(X_i - \overline{X}_n)],$$

where $\overline{X}_n(s) = n^{-1}(X_1(s) + \cdots + X_n(s))$. We denote the eigenvalues and eigenfunctions of $\widehat{\Gamma}$ by $\widehat{\lambda}_{nr}$ and $\widehat{\psi}_{nr}$, $r = 1,\ldots,n-1$, respectively. In order to ensure that $\widehat{\psi}_{nr}$ may be viewed as an estimator of ψ_r rather than of $-\psi_r$, we will in the following assume that the signs are such that $\langle\widehat{\psi}_{nr},\psi_r\rangle \geq 0$. Note that

$$\widehat{\Gamma}\widehat{\psi}_{nr} = \widehat{\lambda}_{nr}\widehat{\psi}_{nr}, \quad r = 1,\ldots,n-1, \tag{5}$$

and

$$\widehat{\lambda}_{nr} = \frac{1}{n-1}\sum_{i=1}^{n}\langle X_i - \overline{X}_n, \widehat{\psi}_{nr}\rangle^2, \quad r = 1,\ldots,n. \tag{6}$$

Define for $d > 0$,

$$\widehat{T}_{n,1}(d) := \max_{1\leq j\leq d}\frac{1}{\sqrt{\widehat{\lambda}_j}}\max_{1\leq k\leq n}\Big|\sum_{i=1}^{k}\langle X_i - \overline{X}_n, \widehat{\psi}_j\rangle\Big|. \tag{7}$$

This statistic is designed for at most one change-point alternative. Its limiting distribution is established in the following theorem.

Theorem 3. Let random functional sample (X_k) be defined by (1) where $Y, Y_1, Y_2, \ldots$ satisfies Assumptions 1 and 2. Then,

(a) Under H_0, it holds that

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)|,$$

where $B_1,\dots,B_d$ are independent standard Brownian bridge processes;

(b) Under H_A, it holds that

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)+\Delta(t)\langle q,\psi_k/\sqrt{\lambda_k}\rangle|,$$

where

$$\Delta(t)=\int_0^t u(s)\,\mathrm{d}s - t\int_0^1 u(s)\,\mathrm{d}s,\quad t\in[0,1]. \tag{8}$$

(c) Under H_A', it holds that

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Based on this result, we construct the testing procedure in a classical way. Choose for a given $\alpha\in(0,1)$, $C_\alpha>0$ such that

$$P(\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)|>C_\alpha)=\alpha.$$

According to Theorem 3, the test:

$$\widehat{T}_{n,1}(d)\ge\sqrt{n}C_\alpha \tag{9}$$

has asymptotic level α.

Let us note that, due to the independence of Brownian bridges $B_k, k=1,\dots,d$, we have

$$1-\alpha=P(\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)|\le C_\alpha)=P^d(\sup_{0\le t\le 1}|B_1(t)|\le C_\alpha).$$

This yields

$$P(\sup_{0\le t\le 1}|B_1(t)|\le C_\alpha)=(1-\alpha)^{1/d}.$$

Hence, C_α is the $(1-\alpha)^{1/d}$-quantile of the distribution of $\sup_{0\le t\le 1}|B_1(t)|$. This observation simplifies the calculations of critical values C_α.

In particular, if there is $s^*\in(0,1)$ such that $u(s)=\mathbf{1}_{[0,s^*]}(s), s\in[0,1]$, then we have one change-point model:

$$X_k(t)=\mathbf{1}_{[0,s^*]}(k/n)q_n(t)+Y_k(t),\quad t\in[0,1].$$

In this case, $\Delta(t)=\Delta^*(t):=\min\{t,s^*\}-ts^*,\quad t\in[0,1]$. Let us observe

that test statistic $\widehat{T}_{n,1}(d)$ tends to infinity when $d \to \infty$. On the other hand, with larger d, the approximation of X_j by series $\sum_{k=1}^{d} \langle X, \widehat{\psi}_j \rangle \widehat{\psi}_j$ is better and leads to better testing power. The following result establishes the asymptotic distribution of $\widehat{T}_{n,1}(d)$ as $d \to \infty$.

Theorem 4. Let random functional sample (X_k) be defined by (1) where $Y, Y_1, Y_2, \ldots$ satisfies Assumption 1. Then, under H_0,

$$\lim_{d\to\infty} \lim_{n\to\infty} P\Big(n^{-1/2}\widehat{T}_{n,1}(d) \le \frac{x}{a_d} + b_d\Big) = \exp\{-e^{-x}\}, \quad x \ge 0, \tag{10}$$

where

$$a_d = (8 \ln d)^{1/2}, \quad b_d = \frac{1}{4} a_d + \frac{\ln \ln d}{a_d}. \tag{11}$$

When d is large, the test (9) becomes

$$\widehat{T}_{n,1} \ge \sqrt{n}\Big[\frac{1}{a_d} \ln\Big(\frac{1}{\ln(1/\alpha)}\Big) + b_d\Big] \tag{12}$$

and has asymptotic level α as n and d tend to infinity.

For $m > 1$, let $\mathcal{N}_m$ be a set of all partitions $\kappa = (k_i, i = 0, 1, \ldots, m)$ of the set $\{0, 1, \ldots, n\}$ such that $0 = k_0 < k_1 < \cdots < k_{m-1} < k_m = n$. Next, consider for fixed integers d, $1 \le m < n$ and real $p > 2$,

$$\widehat{T}_{n,m}(d,p) := \max_{1\le j\le d} \frac{1}{\sqrt{\widehat{\lambda}_j}} \max_{\kappa\in\mathcal{N}_m} \Big\{ \sum_{i=1}^{m} \Big| \sum_{k=k_{i-1}+1}^{k_i} \langle X_k - \overline{X}_n, \widehat{\psi}_j \rangle \Big|^p \Big\}^{1/p}. \tag{13}$$

The statistic $\widehat{T}_{n,m}(d,p)$ is designed for testing at most m change-points in a sample.

Theorem 5. Let the random sample $(X_i, i = 1, \ldots, n)$ be as in Theorem 2. Then:

(a) Under H_0,

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j),$$

where $B_1, \ldots, B_d$ are independent standard Brownian bridges.

(b) Under H_A,

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j + \Delta\langle q, \widehat{\psi}_j/\sqrt{\widehat{\lambda}_j}\rangle).$$

(c) Under H_A',

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

According to this theorem, the test:

$$\widehat{T}_{n,m}(d,p) \geq \sqrt{n} C_\alpha(m,d,p) \tag{14}$$

has asymptotic level α, if $C_\alpha(m,d,p)$ is such that

$$P(v_{m,p}^{1/p}(B) \leq C_\alpha(m,d,p)) = (1-\alpha)^{1/d}.$$

Next, fixed integer d and real $p > 2$, we define

$$\widehat{T}_n(d,p) := \max_{1\leq j\leq d} \frac{1}{\sqrt{\widehat{\lambda}_j}} \max_{1\leq m\leq n} \max_{\kappa\in\mathcal{N}_m} \Big\{ \sum_{i=1}^{m} \Big| \sum_{k=k_{i-1}+1}^{k_i} \langle X_k - \overline{X}_n, \widehat{\psi}_j \rangle \Big|^p \Big\}^{1/p}. \tag{15}$$

The statistic $\widehat{T}_n(d,p)$ is designed for testing an unknown number of change-points in a sample.

Theorem 6. Let random sample $(X_i, i = 1,\ldots,n)$ be as in Theorem 2. Then:

(a) Under H_0,

$$n^{-1/2}\widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\leq j\leq d} v_p^{1/p}(B_j),$$

where $B_1,\ldots,B_d$ are independent standard Brownian bridges.

(b) Under H_A,

$$n^{-1/2}\widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\leq j\leq d} v_p^{1/p}(B_j + \Delta\langle q, \widehat{\psi}_j/\sqrt{\widehat{\lambda}_j}\rangle),$$

where $\Delta(t), t \in [0,1]$ is as defined in Theorem 2.

(c) Under H_A',

$$n^{-1/2}\widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

According to this result the test:

$$\widehat{T}_n(d,p) \geq \sqrt{n} C_\alpha(d,p) \tag{16}$$

has asymptotic level α, if $C_\alpha(d,p)$ is such that

$$P(v_p^{1/p}(B) \leq C_\alpha(d,p)) = (1-\alpha)^{1/d}.$$

Simulation results

In the univariate case, the statistical power was estimated using Monte-Carlo simulations. The sliding window approach was then used to identify the locations of change points and to detect multiple change points. The test was

applied to a dataset of annual Nile river flow measurements from 1871 to 1970, which contained a known change point around 1898 [18].

Consider a sequence $X_1, \ldots, X_n$, $n \geq 1$, of independent random variables. Suppose there exists a $\tau \in [0, 1]$ such that $X_1, \ldots, X_{n\tau}$ have the distribution $\mathcal{N}(\mu_1, \sigma^2)$ and $X_{n\tau+1}, \ldots, X_n$ have the distribution $\mathcal{N}(\mu_2, \sigma^2)$. Given α we construct consistent critical point a_α such that

$$P\left(v_p^{1/p}(Z_n) < a_\alpha\right) = \alpha$$

for $H_0 : \tau = 0$ versus $H_1 : \tau \in (0, 1]$.

The distribution of $v_p^{1/p}(Z_n)$ was estimated using Monte-Carlo simulations with $n = 1000$ and $k = 100000$ where k is number of replicated copies of the Z_n process. The resulting distribution for different values of p-variation ($p = 3, 4, 8$) is shown in Figure 2.1, with critical values at the asymptotic level of $\alpha = 0.95$ marked.

The power of detection was evaluated by studying Type II errors under the alternative hypothesis. The aim was to evaluate the statistical power in different scenarios. First, the magnitude of the change was varied by incrementally increasing the value of a, such that $\mu_2 = a$. The resulting statistical power with respect to different values of a and different numbers of observations ($n = 1000, 10000, 30000$) is shown in Figure 2.2a. The simulation results show that if the number of observations is large enough ($n \geq 30000$), the null hypothesis is correctly rejected more than 80% of the time when $\mu_2 > 0.035$.

In the second scenario, the statistical power was evaluated with respect to τ values with $n = 1000$. The value of τ represents the location of the change point. This simulation demonstrates the speed at which changes can be detected. The simulated results are shown in Figure 2.2b. It can be seen that if $|\mu_1 - \mu_2| \geq 0.3$, a τ value as small as 0.20 is sufficient to achieve a statistical power of 80%.

Simulation results demonstrate that the sliding window method is effective at identifying the location of the change points in data (see Figure 2.3). The red dashed line indicates the critical value with a significance level of $\alpha = 0.95$. It is typically observed that the statistic used to detect change points will approach this critical value around the point where a change is present.

We analyzed the annual flow of the River Nile at Aswan from 1871 to 1970, using data recorded in units of $10^8 m^3$. There is a known change point in the data near 1898 [18], and the measurements may provide insight into the patterns of rainfall in the region. The results of the proposed test show (see Figure 2.4) that the null hypothesis was rejected around the year 1898, which is very similar to what other researchers have indicated.

In our tests of the functional sample, we used three scenarios to evaluate statistical power. In the first scenario, the actual eigenvalues and eigenfunctions

were known, allowing us to avoid any data loss or measurement errors. In the second scenario, we reconstructed the generated functional sample from the first scenario by measuring the values at random times (10 and 200 data points) and using a different set of basis functions constructed functional sample. This allowed us to measure the impact of information loss due to measurements taken at discrete points and smoothing. The simulation results showed (see Figure 3.6) that the performance of the test is not significantly affected by the reconstruction of the random functional sample if we have enough data points. However, if the number of points is much lower (from 200 points to 10 points), we can see a degradation in performance.

In the third scenario the discrete observations $(i/M, y_{ij}), i = 0, 1, \ldots, M, j = 1, \ldots, n$, are generated by taking

$$y_{ij} = M^{-1/2} \sum_{k=1}^{i} \xi_{kj},$$

where (ξ_{jk}) are i.i.d. symmetrized Pareto random variables with index p (we used $p = 5$). The y_{ij} can be interpreted as the observation of a standard Wiener process at i/M. From $(y_{ij}, i = 1, \ldots, M)$, the function Y_j is obtained using the B-spline smoothing technique. During the simulation, we used $M = 1000$ and $D = 50$ B-spline functions, thus obtaining $n = 500$ functions $Y_1, \ldots, Y_n$.

Then, we define for $j = 1, \ldots, n$,

$$X_j = \begin{cases} Y_j, & \text{under null} \\ u_n(j/n)q_n + Y_j, & \text{under alternative} \end{cases}$$

and consider different configurations u_n of change-points and $q_n(t) = a_n\sqrt{Mt}, t \in [0, 1]$.

In the power studies, we tested two variants of the random functional samples. One had a single change point in the middle, and the other had two change points forming an epidemic change. In the first case, we modified 500 of the 1000 curves in the functional sample to violate the null hypothesis. In the second case, we modified 500 of the 1500 curves in the middle of the sample. For each repetition, we calculated two statistics in the single change-point simulation: $\widehat{T}_n(d, p)$ and $\widehat{T}_{n,1}(d)$. For the epidemic change simulation, we calculated $\widehat{T}_{n,m}(d, p)$ with $m = 2$.

Figure 3.8 presents the results of the statistical power simulation for a single change point (left) and an epidemic change (right). The results show that the epidemic change has weaker statistical power when using the statistic $\widehat{T}_{n,m}(d, p)$ compared to the unknown number of change point statistic $\widehat{T}_n(d, p)$. However, when restricting the partition count, the locations of the partitions often match or are very close to the actual locations of the change point.

For the real-world data, we used neurophysiological data and lickometer data from a long-term study on alcohol-consuming rats to demonstrate the performance of the proposed test for change-point detection. The rats were given two drinking bouts, one with alcohol and one with water, and were able to freely choose what to drink while their brain activity was monitored. In our analysis, we analyzed the first alcohol-drinking event, which lasted approximately 27 seconds. We included 10 seconds before and after the event, totaling 47 seconds. The time series was divided into processes of 100 ms, each containing 100 data points.

The results are visualized in Figure 3.9. We can see that the tests with statistics $\widehat{T}_n(d,p)$ and $\widehat{T}_{(n,m)}(d,p)$ strongly rejected the null hypothesis at around 2 seconds after the rat started to drink alcohol, indicating changes in brain activity during alcohol consumption in the CPu brain region. The statistic $\widehat{T}_{n,m}(d,p)$ had larger volatility in the Nacc brain region before the drinking event and lower volatility just after the event began.

Finally, the locations of the restricted ($m = 2$) p-variation partition points closely matched the beginning and end of the drinking period. In Figure 3.9, the gray vertical dashed lines indicate the actual beginning and end of the drinking period measured by the lickometer, and the black vertical lines indicate the location of the partitions calculated from the functional sample $\widehat{F}_{450}$. The first partition is located at 10.5 seconds, and the second partition point is at 38.4 seconds, which aligns well with the data collected from the lickometer. The test with a restricted partition count showed larger volatility but accurately determined the locations of the change points.

1. Background

This manuscript is devoted to problems related to Change point detection (CPD). Change point Analysis (CPA) is a powerful tool for evaluating structural changes in data and is widely used in many fields. The goal of this work is to extend the existing set of tools with a novel approach. First, we lay the foundation for the univariate process, based on measures of variation of the partial sum process. Then, we extend the method to a functional sample.

Before discussing change points in detail, we introduce the reader to concepts that are fundamental to the methods we propose in Chapters 2 and 3. Specifically, we focus on two concepts: the variation of functions (covered in section 1.1) and Funtional Data Analysis (FDA) concepts and tools (covered in Section 1.2).

Next, we introduce the reader to the field of change point analysis. Section 1.3 defines the terminology that is used throughout this thesis. Then, we review the existing methods and how our work relates to the existing literature on CPD. In Section 1.3, we present real-world situations where CPD is applied. Finally, we review the progress in CPD for functional data in Section 1.4.

1.1. p-Variation of the function

Camile Jordan in his paper [36], devoted to the convergence of Fourier series, introduced notation of the variation of the function. He used the new concept in order to prove the convergence theorem for the Fourier series of discontinuous period functions whose variation is bounded. Several generalizations and extensions of the concept of p-variation have been proposed, including the functions of bounded p-variation and the functions of bounded φ-variation. The latter notation was introduced by L.C. Young in 1937 [82] and has found widespread application. The p-variation term was first coined by Wiener in 1924 [78]. In his work, Wiener mainly focused on the case of the 2-variation ($p = 2$). He defined the p-variation of a function f as a collection of seminorms indexed by a real number $p \geq 1$, derived from the function defined on an ordered set of a metric space.

Later, major research with $p \neq 2$ was done by Young [81] and partly Love [45]. Young considered an inequality bearing a formal resemblance to that of Hölder and derived conditions for the existence of a Stieltjes integral. Young's integration theory allowed to define $\int ydx$ as soon as y had finite q-variation and x had a finite p-variation with $1/p + 1/q > 1$.

If $p = 1$, the variation is referred to as the total variation. Functions with a finite 1-variation are known as bounded variation functions. The concept of the total variation was introduced by Camile Jordan in 1881 [36]. The total varia-

tion has found numerous applications in various branches of mathematics and engineering, including the numerical analysis of differential equations (see [32]). Functions of bounded variation have received significant attention due to their use in the study of discontinuities and differentiability. These types of functions, which exhibit bounded oscillation or "roughness", have found numerous applications in the applied sciences, including physics, mechanics, chemistry, and more. Their ability to accommodate discontinuities has made them particularly useful in these fields. In image denoising, total variation was applied in order to reduce the noise of the image [15, 66].

For extensive use of *p-variation*, the reader may refer to the book by Dudley and Norvaiša [23].

Definition 1.1.1. A finite sequence $\kappa = \{t_i\}_{i=0}^n$ for a positive integer n is called a *partition* of $[a, b]$ if $a = t_0 < t_1 \cdots < t_n = b$.

Definition 1.1.2. Let f be any real-valued function on an interval $[a, b]$ with $-\infty < a \leq b < +\infty$ and let $0 < p < \infty$. If $a < b$, for a partition $\kappa = \{t_i\}_{i=0}^n$, the p-variation sum for f over κ is defined by

$$s_p(f; \kappa) := \left\{ \sum_{i=1}^{n} |f(t_i) - f(t_{i-1})|^p \right\}.$$

The p-variation of f on $[a, b]$ is defined as $v_p(f; [a, b]) := 0$ if $a = b$ and

$$v_p(f) := v_p(f; [a, b]) := \sup_{\kappa} s_p(f; \kappa)$$

if $a < b$, where the supremum is over all partitions κ of $[a, b]$. Then f is said to be of *bounded* p-variation on $[a, b]$, or $f \in \mathcal{W}_p[a, b]$, if and only if $v_p(f) < +\infty$ and $\mathcal{W}_p[0, 1]$ is the set of all such functions. The set $\mathcal{W}_p[0, 1]$, $p \geq 1$, is a non separable Banach space with the norm

$$\|f\|_{[p]} := |f(0)| + v_p^{1/p}(f).$$

The embedding $\mathcal{W}_p[0, 1] \hookrightarrow \mathcal{W}_q[0, 1]$ is continuous and

$$v_q^{1/q}(f) \leq v_p^{1/p}(f), \quad \text{for } 1 \leq p < q.$$

For a comparison with the α-Hölder, $\alpha \in (0, 1]$, property of f, if $p := 1/\alpha$ and $|f(t) - f(s)| \leq C|t - s|^\alpha, t, s \in [0, 1]$, then we have

$$\sum_{j=1}^{m} |f(t_j) - f(t_{j-1})|^p \leq C^p \sum_{j=1}^{m} (t_j - t_{j-1}) = C^p$$

and so $v_p(f) \leq C^p < \infty$.

It is possible for a f to have unbounded variation, but still $v_p(f) < \infty$ for some $p > 1$ and if $v_p(f) < \infty$ then $v_q(f) < \infty$ for $q > p$. For example, indicator functions have bounded p-variation for any $p > 0$.

Computational challenges

To calculate the p-variation of a function f, it is necessary to determine the supremum in the definition of $v_p(f)$ and find a partition κ of the interval $[0, 1]$ that achieves this maximum. If f is a piecewise monotone function with local extrema forming a partition $\kappa = (t_i)$ and $p = 1$, then the supremum in $v_p(f)$ is attainable using this partition. The algorithm for computing the total variation, given in reference 1, is straightforward and has a complexity of $O(n)$.

Algorithm 1: Computation of the total variation

```
Function 1-variation(X):
    /* n is the size of the sample X                    */
    n = |X|;
    v_1 = 0;
    for i = 0; i < n do
        v_1 = v_1 + |X[i] − X[i + 1]|
    end
    return v_1;
```

On the other hand, finding a maximizing partition is more complex when $p \neq 1$. Intuitively, it might seem that the maximizing partition points should be corner points. However, this is not necessarily the case if the function f is not piecewise monotone. In such a situation, it may be necessary to exclude certain points from the partition κ to achieve the maximum in the definition of $v_p(f)$. For example, intermediate points within an interval over which f is increasing should not be treated as corner points, as their inclusion would result in a smaller value of the variation. Instead, only the endpoints of the interval should be considered as corner points in this case. To show this we define the following lemma:

Lemma 1. Let f be a function and $\Pi = \{t_0, t_1, \ldots, t_n\}$ be a partition. Assume, that f is a monotone increasing function on interval $[t_{i-1}, t_{i+1}]$. Then if $\Pi' = \Pi_{t_i}$, $v_p(f, \Pi') \geq v_p(f, \Pi)$. If f is strictly increasing and $p > 1$, then inequality is strict.

In order to prove lemma 1 we use the following inequality $(a+b)^p \geq a^p + b^p$ for all $a, b \geq 0$ and $p \geq 1$.

Proof of lemma 1. From the inequality $(a + b)^p \geq a^p + b^p$ it follows

$$v_p(f,\Pi^{'})^p - v_p(f,\Pi)^p = (f(t_{i+1}) - f(t_{i-1}))^p - (f(t_{i+1}) - f(t_i))^p - (f(t_i) - f(t_{i-1}))^p$$
$$= \left(\Delta_{t_i} f + \Delta_{t_{i+1}} f\right)^p - (\Delta_{t_i} f)^p - \left(\Delta_{t_{i+1}} f\right)^p \geq 0$$

The inequality is strict if $\Delta_t, f > 0, \Delta_{t+1} f > 0$ and $p > 1$. □

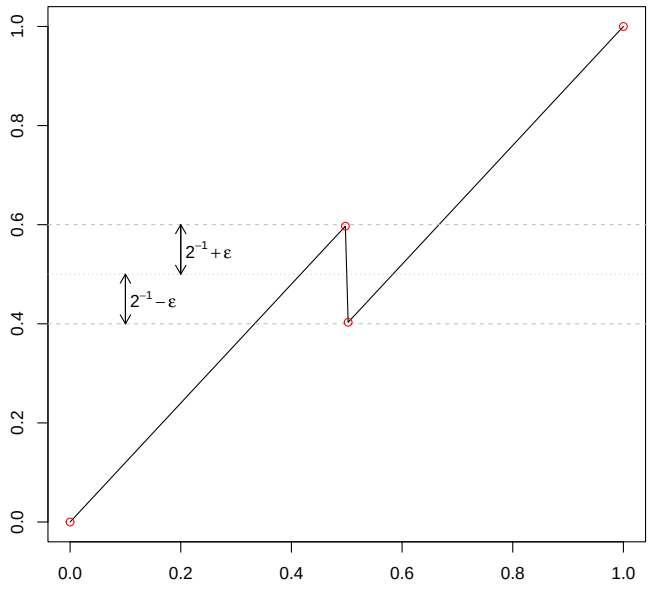


Figure 1.1: figure of the function

Example 1.1.1. Consider the function on an interval $[0, 1]$ illustrated in Figure 1.1 defined as

$$f(x) = \begin{cases} x(2\epsilon + 1), & \text{if } x < 2^{-1} \\ 2\epsilon(x-1) + x, & \text{otherwise} \end{cases}$$

With $0 < \epsilon < 0.16$, if we partition at the corner points, then

$$v_p(f,\Pi)^p = \left(2^{-1} + \epsilon\right)^p + (2\epsilon)^p + \left(2^{-1} + \epsilon\right)^p = (2\epsilon)^p + 2\left(2^{-1} + \epsilon\right)^p \approx 2^{1-p} < 1.$$

However, this approach does not necessarily result in the maximum p-variation. If we consider a trivial partition $\Pi' = 0, 1$, then $v_p(f,\Pi') = 1$, which is larger than $v_p(f,\Pi) < 1 \leq v_p(f,\Pi')$ for any $p > 1$. This demonstrates that the trivial

partition Π' may yield a larger p-variation than a partition Π that includes intermediate points.

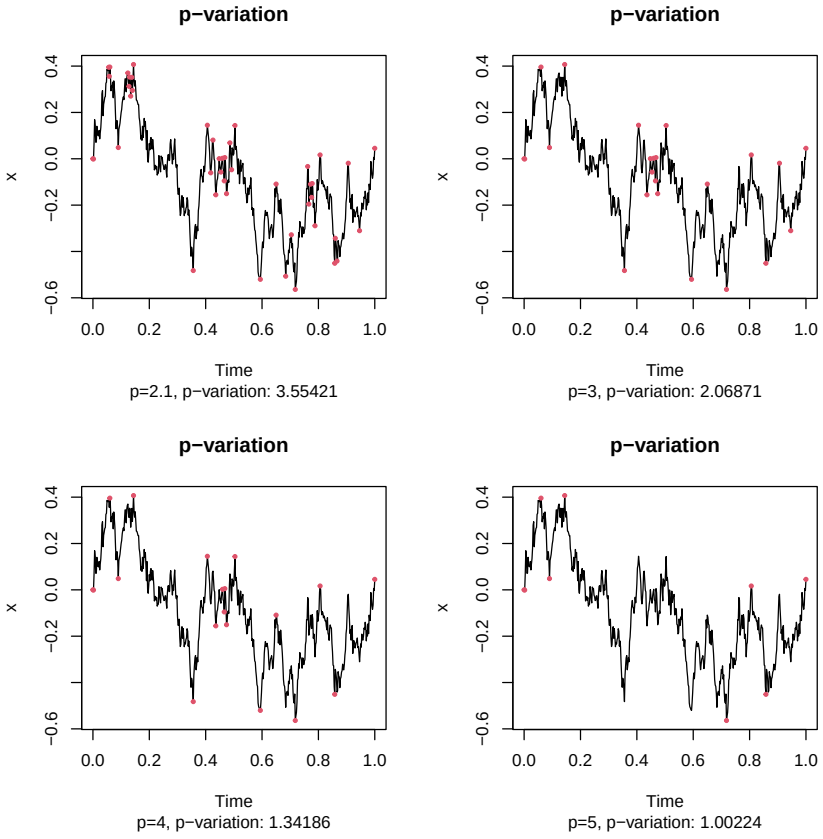


Figure 1.2: Illustration of the effect of varying p on the p-variation of the Brownian bridge.

The p-variation of the Brownian bridge can be illustrated through the example depicted in Figure 1.2. The figure shows how the locations of the maximal partition points, marked in red, vary as the value of p is changed. This demonstrates how the p-variation of the Brownian bridge is influenced by the value of p.

Note that, it is necessary to restrict the value of p to be greater than 2 when calculating the p-variation of the Brownian bridge, as shown by the well-known results of N. Wiener (1923) and P. Lévy (1940). Specifically, it has been established that

$$v_p(W) < \infty \quad \text{almost surely if and only if} \quad p > 2;$$

as well

$$v_p(B) < \infty \text{ almost surely if and only if } p > 2$$

and $v_2(W) = \infty$ and $v_2(B) = \infty$ almost surely.

Algorithm 2: Computation of the p-variation with complexity of $O(n^2)$

Input : sample X and scalar $p > 0$
Output: $v_p(X)$ - p-variation of the sample X

```
Function p-variation(X, p):
    /* n is the size of the sample X                    */
    n = |X|;
    /* P is an array of the cumulative p-variation      */
    P = array[n];
    for i = 1; i < n do
        for j = 0; j < i do
            d = P[j] + |X[i] − X[j]|^p;
            P[j] = max(P[i], d);
        end
    end
    v_p = P[n]^(1/p);
    return v_p;
```

Calculating v_p using a naive approach involves considering all combinations (see pseudo-code 2), which has a computational complexity of $O(n^2)$. This makes it impractical for larger values of n. An alternative method was proposed by Butkus and Norvaiša in their work, "Computing p-variation of functions," published in 2018 [14]. The algorithm, which has a lower computational complexity, is available on CRAN[1].

In their paper, Butkus and Norvaiša demonstrated that the p-variation of a piecewise monotonic function f is dependent only on a tuple $(f(t_0), f(t_1), \ldots, f(t_n))$, where $(t_i)_{i=0}^n$ is the minimal monotonicity partition of $[a, b]$. The general steps of the BN algorithm are as follows:

- *Find minimal monotonicity partition.* The goal is to find all points of strict local extrema of the sample function. The points which are not strict local extremum points are excluded.
- *Check small subsamples.* Each sample is checked for redundant points.
- *Merge smaller subsamples.* Let $\Pi_{0,i}$ and $\Pi_{i,n}$ be maximizing partitions for the p-variation over inerval $[0, i]$ and $[i, n]$, respectively. The union sample $\Pi_{0,i} \sqcup \Pi_{i,n}$ is used to find a maximizing partition $\Pi_{0,n}$. This operation is called merging. The merging is repeated for all small maximizing partitions until final maximization is found.

[1] https://cran.r-project.org/web/packages/pvar/index.html

- *Calculate the value of p-variation.* Having all maximizing partitions $(S_i)_{i=0}^n$ it is easy to compute the p-variation sum

$$v_p = \sum_{i_1}^{n} |S_{i-1} - S_i|^p.$$

The BN algorithm is efficient for calculating the p-variation of real-valued functions. However, it is not applicable to functions in $\mathbb{R}^2$. An alternative algorithm, developed by Alexey Korepanov, Terry Lyons, and Pavel Zorin-Kranich (KLZ), is available on GitHub[2]. According to the authors, this algorithm is fast enough for paths with a large number of points, and has a computational complexity that is close to $O(n \log(n))$. However, it should be noted that the worst-case complexity of the KLZ algorithm is $O(n^2)$. A comparison of the performance of these algorithms can be found in Table 1.1.

Table 1.1: A performance analysis of various algorithms for calculating p-variation was conducted on a machine with an Intel i7 CPU and 64GB of RAM running Linux. The evaluation was based on execution time in seconds for various combinations of p and n values.

	Algorithm	p=2.5	p=4	p=5.5	p=7	p=8.5
n=100	BN	0,0001	0.0001	0.0001	0.0001	0.0001
	KLZ	0.0037	0.0035	0.0033	0.0031	0.0029
	Simplistic $O(n^2)$	0.0166				
n=1000	BN	0.0002	0.0002	0.0002	0.0001	0.0001
	KLZ	0.0513	0.0456	0.045	0.0418	0.0425
	Simplistic $O(n^2)$	0.2921				
n=10000	BN	0.0013	0.0009	0.0009	0.0008	0.0008
	KLZ	0.6652	0.6111	0.5788	0.5606	0.567
	Simplistic $O(n^2)$	26.2577				
n=100000	BN	0.0128	0.0084	0.0077	0.0074	0.007
	KLZ	8.1725	7.4600	7.1236	7.0809	8.6697
	Simplistic $O(n^2)$	2621.3418 (approx. 43 minutes)				
n=1000000	BN	0.14	0.103	0.0954	0.4466	0.151
	KLZ	95.8404	89.2490	83.9256	95.8208	180.5138
	Simplistic $O(n^2)$	264822.6296 (approx. 3 days)				

1.2. Functional Data Analysis

When studying the environment, we often observe it not as a single point, but rather as a curve. For instance, a comprehensive view of the trajectory of a moving object over time is more informative than just a single coordinate at a particular moment. As such, for data samples of this form, it is more natural to treat each atomic observation as a curve rather than a discrete point. With the advancement of computational power and the proliferation of devices capable of

[2]https://github.com/khumarahn/p-var

continuously recording data, statistical inference and analysis of curve-shaped data has become increasingly challenging. This has led to the development of FDA in many research areas.

The history of Functional Data (FD) can be traced back to 1950 when Grenander attempted to apply statistical concepts and methods of inference to stochastic processes [29]. In 1958, Rao analyzed statistical methods for comparing growth curves in the study of growing organisms [64]. The term FDA was first coined by Ramsey in 1982 [61] and later by Ramsay and Dalzell in 1991 [62]. In their book [63], Ramsey and Silverman provide a thorough overview of FDA and offer numerous examples of its application in various domains.

In practice, FD are typically represented as discrete data points measured over time, space, or other continuum. These discrete data points for an individual function can be denoted as $X_i = (t_{i,j}, y_{i,j})$, where $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. The variables $t_{i,j}$ denote design time points, while $y_{i,j}$ are the responses at those time points. A random sample of FD typically consists of independent real-valued functions, $X_1(t), X_2(t), \ldots, X_n(t)$, on an interval $[0, T]$ on the real line. Often, it is convenient to view these functions as a one-dimensional stochastic process in a Hilbert space, such as $L^2([0, T])$. A process $X_i(t)$ is considered to be in L^2 if it satisfies the following condition:

$$\mathbb{E}\left(\int_{[0,T]} X^2(t)\,\mathrm{d}t\right) < \infty.$$

Zhang [83] described the FD as a generalization of multivariate data from a finite dimensional to an infinite dimensional. Wang et. al. [77] refer FD to the first generation and the next generation. The latter was said to be more complex objects, and possibly are multivariate, correlated, or involved images or shapes. Examples of the next-generation FD data include brain and neuroimaging data.

Before FDA gained popularity, data that were continuously recorded over a time interval at several discrete time points were historically analyzed using classical Multivariate Data Analysis (MDA) methods. However, interpreting a sample of curves in a multivariate fashion does not take into account the fact that the fundamental unit of the sample is a curve, therefore MDA has its limitations. Some problematic cases when analyzing FD within the MDA framework include:

- Observed data is sampled with non-equal spaces between data points.
- High-Frequency data can have more data points than a number of subjects.
- The total number of sampled data points may vary across the subjects.

Many MDA algorithms have an analogous version in FDA, such as Principal Components analysis (PCA), correlation analysis, and discriminant analysis. In addition, taking advantage of the unique characteristics of the functional framework, new tools are being developed. For example, the ability to compute derivatives allows the use of differential equations. For a comprehensive review of tools, readers may refer to the book by Ramsay and Silverman [63]. Nonparametric techniques for analyzing FD are also becoming well established (see [24, 34] for an extensive overview).

The high or infinite dimensionality of FD is a significant challenge for researchers. From a theoretical perspective, it can be difficult to develop statistical models and methods that are able to accurately capture the complex structure of FD. From a computational standpoint, the high dimensionality of FD can lead to poor scalability, making it difficult to process and analyze large datasets in a timely manner. This can result in insufficient statistical power, meaning that the results of statistical tests may not be reliable.

Despite these challenges, FD is a valuable source of information for researchers. The fact that FD is structured as an infinite dimensional space means that it has the potential to capture a wide range of complex patterns and trends. This opens up new opportunities for research and data analysis, and has made change point analysis for FD an important area of study.

It is important to note that in functional change point analysis, changes are measured between functions rather than at individual points. This is in contrast to traditional change point analysis, which typically focuses on changes at specific points in time. By focusing on changes between functions, functional change point analysis is able to capture more nuanced and subtle changes in data, making it a powerful tool for data analysis.

Construction of the functional objects

In practice, data collections involve observations at discrete points in time, resulting in non-continuous data. Even when the sampling rate is very high, the recorded data are still not continuous. The sampling rate, or the number of observations taken per unit of time, typically depends on the characteristics of the signal being measured. For example, a signal with high frequency components may require a higher sampling rate to capture the full range of variation. In any case, measurement error is almost always present in discrete observations, meaning that the recorded data may not perfectly reflect the true underlying signal.

Due to these considerations, the first step in FDA is often to construct continuous functional objects from discrete observations. One common approach to achieving this is through the use of a basis expansion. This involves expressing the data in terms of a set of basis functions, which are typically chosen to be smooth and well-behaved. By expressing the data in this way, it is possible

to capture the underlying structure of the data more accurately and to analyze it more effectively.

Basis expansion is a fundamental concept in FDA, and there are many different types of basis functions that can be used. Some common examples include B-splines, Fourier basis functions, and wavelets. The choice of basis functions will depend on the characteristics of the data and the goals of the analysis. In any case, the use of basis expansion allows for the construction of continuous functional objects from discrete observations, enabling more accurate and effective analysis of FD.

The basis expansion can be defined as follows:

$$x = \sum_{k=1}^{K} c_k \phi_k = c'\phi,$$

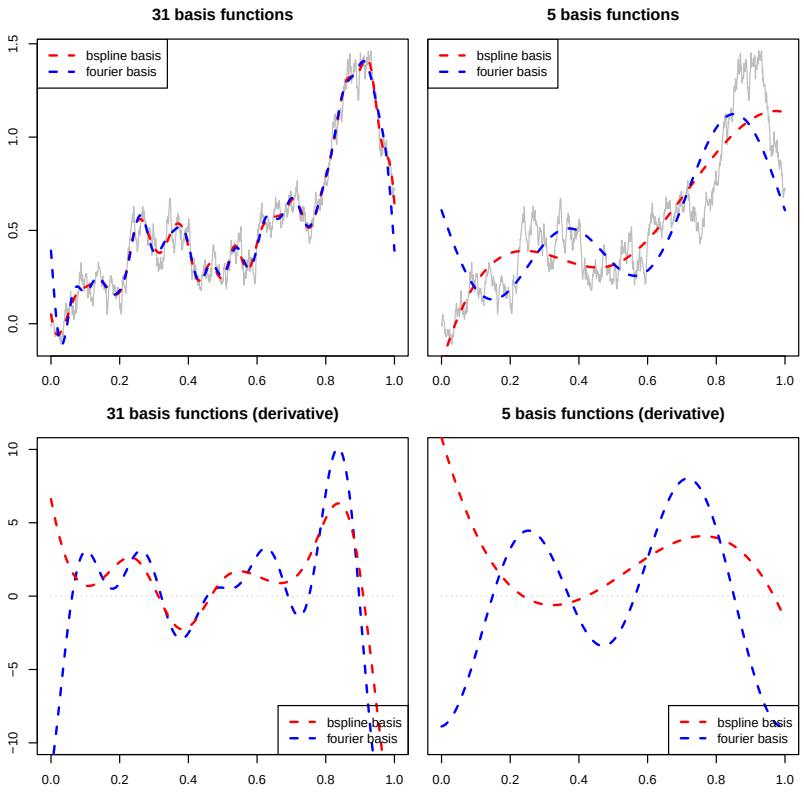


Figure 1.3: A smoothed Wiener process is demonstrated using B-spline and Fourier basis functions. The upper left figure displays 31 basis curves, and the upper right figure displays 5 basis functions. The derivative of the smoothed curves is depicted in the lower figures.

In basis expansion methods, a potentially infinite-dimensional space of functions is represented within the finite-dimensional framework of vectors like c, where K is the number of basis functions ϕ_k [63]. The smoothness of the data is controlled by the value of K, which can be chosen by the researcher. When selecting a basis system, it is important to consider two aspects. First, the basis should reflect the nature of the data as accurately as possible. For example, Fourier basis functions are well-suited for periodic curves, while B-spline functions are better at capturing nonperiodic and complex features [21]. Second, the derivatives of the basis should be considered. While the basis may represent the data well, it may not provide a good estimate of the derivative.

The choice of the parameter K depends on how closely the data should be represented in the functional object. When $K = n$, the data are represented exactly. However, using a smaller value of K allows for more degrees of freedom to test hypotheses and requires less computational power. The coefficients themselves can also be important descriptors of the data. The effect of different basis systems and values of K are illustrated in Figure 1.3. The upper left plot of the figure was smoothed using $K = 31$ basis functions, while the upper right plot was smoothed with $K = 5$.

One of the approaches of estimating c is by using the least square principle. This is equivalent to the minimization of the sum of squared errors:

$$SSE(y|c) = \sum_{j=1}^{J} \left[y_j - \sum_{k=1}^{K} c_k \phi_k(t_j) \right]^2 .$$

Let Φ be a $J \times K$ such that $\Phi_{kj} = \phi_k(t_j)$. Then setting the gradient of the loss to zero and solving for c we get:

$$\hat{c} = \left(\Phi'\Phi\right)^{-1} \Phi' y.$$

This method was further extended by Green and Silverman [28]. The authors introduced a roughness penalty to fit a more smooth curve. The measure of a function's roughness was measured by the integrated square second derivative

$$PEN_m(x) = \int \left[x^{(m)}(s) \right]^m \, \mathrm{d}s,$$

where $x^{(m)}(s)$ denotes the m order derivative of x evaluated at time s. Then, we can define the penalized residual sum of squares as

$$PENSSE_m(x|y) = [y - x]' [y - x] + \lambda PEN_m(x),$$

where λ is a smoothing parameter. Then, the authors obtained the expression of the estimated coefficient vector

$$\hat{c} = \left(\Phi'\Phi + \lambda R\right)^{-1} \Phi' y,$$

where

$$R_{ij} = \int \phi(s)_i^{(m)} \phi(s)_j^{(m)} \, \mathrm{d}s$$

is the penalty matrix.

Functional Principal Components Analysis

PCA was first proposed by Pearson [54] in 1901 and became the essential technique for data analysis. Mostly it was used for dimensionality reduction by projecting each data point onto a reduced number of components while preserving as much of the data variation as possible. PCA projects the data onto a subspace which minimizes the reconstruction error and maximizes the projected variance. The reduced space can reveal a latent structure in relations between variables. PCA became an essential part in the development of MDA. In general terms the goal is to find a transformation $\tau : \mathbb{R}^d \to \mathbb{R}^k$ that maps data points in $\mathbb{R}^d$ to data points in $\mathbb{R}^k$. In PCA case, the approach is to find a linear transformation that preserves the maximum *variance*. Before introducing the Functional Principal Components Analysis (fPCA), we briefly discuss the classical PCA. Consider m data points $x_1, \dots x_m \in \mathbb{R}^d$ and a positive number $k < d$. The k is the number of the target dimension. In particular, define a matrix X of an $m \times d$:

$$\mathbf{X} = \begin{bmatrix} -\mathbf{x}_1'- \\ \vdots \\ -\mathbf{x}_m'- \end{bmatrix} \in \mathbb{R}^{m \times d}.$$

Next, we need to subtract the mean from each row. Denote the sample mean by μ

$$\mu = m^{-1} \sum_{i=1}^{m} x_i$$

Then, denote

$$\hat{x}_i = x_i - \mu.$$

And then we denote $\hat{X}$ as a *mean-centered* data matrix:

$$\hat{X} = \begin{bmatrix} -\hat{\mathbf{x}}_1'- \\ \vdots \\ -\hat{\mathbf{x}}_m'- \end{bmatrix} \in \mathbb{R}^{m \times d}.$$

The next step is to calculate a covariance matrix S of $d \times d$ in terms of $\hat{X}$ as

$$\mathbf{S} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_i - \mu\right)\left(\mathbf{x}_i - \mu\right)' = \frac{1}{m-1}\hat{\mathbf{X}}'\hat{\mathbf{X}}.$$

fPCA technique is an analogous version of PCA for a functional data. It was one of the first MDA tools to be adopted to functional framework [20]. Basically, the adaptation was based on replacing vectors with functions matrices by compact linear operators, covariance matrices by covariance operators and scalar products in a vector space by scalar products in square-integrable functional space. Therefore, principal components became weight functions varying over the same interval $[a, b]$ as the data. An eigenbasis, representing data, is an *orthonormal* basis of the Hilbert space L^2, that consists of eigenfunctions of the autocovariance operator. Formally we can define the fPCA as follows. For a square-integrable stochastic process $X(t)$, let

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

and

$$G(s,t) = \mathrm{Cov}(X(s), X(t)) = \sum_{k=1}^{\infty}\lambda_k\varphi_k(s)\varphi_k(t),$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$ are the eigenvalues and $\varphi_1, \varphi_2, \ldots$ are the orthonormal eigenfunctions of the linear Hilbert–Schmidt operator

$$G : L^2(\mathcal{T}) \to L^2(\mathcal{T}),\ G(f) = \int_{\mathcal{T}} G(s,t)f(s)\,\mathrm{d}s.$$

By the Karhunen–Loève theorem, one can express the centered process in the eigenbasis

$$X(t) - \mu(t) = \sum_{k=1}^{\infty}\xi_k\varphi_k(t),$$

where $\xi_k = \int_{\mathcal{T}}(X(t) - \mu(t))\varphi_k(t)dt$ is the principal component associated with the k-th eigenfunction φ_k, with the following properties: $\mathbb{E}(\xi_k) = 0, \mathrm{Var}(\xi_k) = \lambda_k$ and $\mathbb{E}(\xi_k\xi_l) = 0$ for $k \neq l$. The centered process is then equivalent to $\xi_1, \xi_2, \ldots$ A common assumption is that X can be represented with good precision by only the first few eigenfunctions after subracting the mean function, i.e.

$$X(t) \approx X_m(t) = \mu(t) + \sum_{k=1}^{m}\xi_k\varphi_k(t),$$

where

$$\mathbb{E}\left(\int_{\mathcal{T}}\left(X(t) - X_m(t)\right)^2\,\mathrm{d}t\right) = \sum_{j>m}\lambda_j \to 0 \text{ as } m \to \infty.$$

A range of techniques have been developed for calculating fPCA, including parametric and nonparametric methods (e.g., Yao et al. 2007 [80], Sang et al.

2017 [67]). Wu et al. (2021) [79] offer a comprehensive review of these techniques, while Shang et al. (2014) [69] provide a review of the use of functional principal component analysis in explanatory analysis.

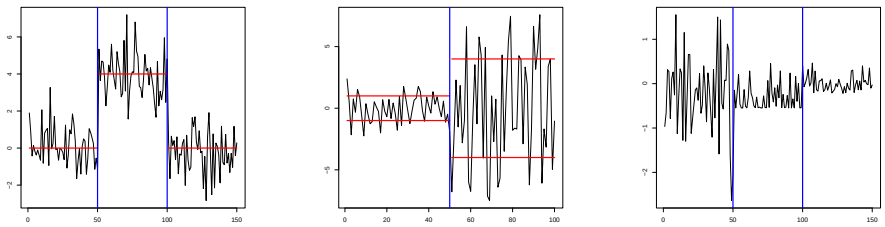
1.3. Change Point problem

Studies related to time series often assume that the process is stationary, meaning that its properties are independent of time. However, in practice, this assumption almost always fails in environments where there are shocks or other changes in the background. As a result, it is more reasonable to assume that the observed data is stationary only for certain segments of the whole sequence. This raises the question of how to identify and recover these segments.

Change point analysis involves partitioning the sequence into a number of segments, with points that divide the sequence into one or more segments that have different statistical properties known as change points. In general terms, there are two sets of problems related to CPD consists of the following two steps:

1. Did the statistical characteristics of the data alter at any point in time?
2. If there was a change, when did it occur?

Many researches consider this problem as classical hypothesis testing with null hypothesis (H_0) indicating of change, but Basseville [9] pointed out that change point estimation is different from classical hypothesis testing. When testing for change points, the multiple testing problem must be taken into account, meaning that almost every point is a *a priori* candidate for a Change point (CP). In order to identify the change points, an appropriately constructed test statistic is used. If the null hypothesis is rejected, the candidate point that provides the strongest evidence becomes the estimated CP.



(a) An example of the change in the mean

(b) An example of the change in the variance

(c) An example of the change in the distribution

Figure 1.4: Examples of change points.

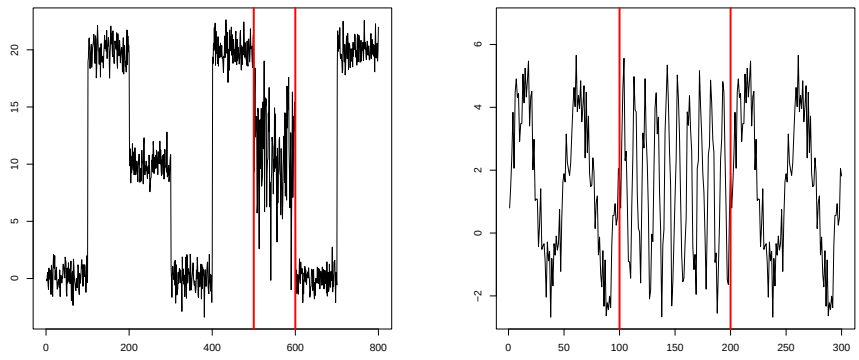
CP are not uniquely defined and the type of change must be taken into consideration. The change can happen in model parameters, in the model

itself, in the dependency structure, or even in a way that can't be specified mathematically. There exists a large set of cases that can be considered as changes in the sequence. The most widely studied problem includes the range of parameters of the distribution. The formulation can be rewritten as:

$$H_0 : X_i \sim \Phi_0 (x; \theta_0), \quad i = 1, \ldots, n,$$
$$H_1 : X_i \sim \begin{cases} \Phi_0 (x; \theta_0) & i = 1, 2, \ldots, \tau, \\ \Phi_1 (x; \theta_1) & i = \tau + 1, \tau + 2, \ldots, n, \end{cases}$$

where θ_i represents potentially unknown parameters of the probability distribution Φ. Many previous works have focused on the parametric characteristics such as the mean (see figure 1.4a) or the variance (see figure 1.4b) for retrospectively detecting changes, meaning that $\Phi_0 = \Phi_1$ but $\theta_0 \neq \theta_1$.

Another case of common CPD problem is when the probability distribution after the time τ changes together with the parameters of the corresponding distribution. The problem can also be extended to the multiple CP, when the sequence of the distributions exists (illustrative example in the figure 1.4c), meaning that $\Phi_0 \neq \Phi_1 \neq \ldots \neq \Phi_n$.



(a) Change in the pattern

(b) Change in the periodicity

Figure 1.5: Examples of the change in pattern and periodicity

There are cases that attracted less attention but are also relevant in many fields. Change in periodicity (see figure 1.5b) which is centered around time series with cyclic properties usually focuses on a frequency domain, for example by using Fourier transform or wavelet transform [17]. Finally, probably the most complex changes to detect are related to changes in the pattern (see figure 1.5a). Pattern changes are difficult to detect and there is not much coverage on detection methods. One approach was proposed by Shvetsov et. al. [72] in electrocardiography.

At Most One Change and multiple changes

The majority of CP techniques start with AMOC problem or a *single* CP as illustrated in the figure 1.4b. The *multiple* change-point detection is more relevant in an offline setting and it is much harder to solve. While the single change point detection can have at most $n-1$ outcomes, the latter can have 2^{n-1} outcomes. The sequence of size n can be segmented in k segments $\binom{n-1}{k-1}$, therefore the total number of possibilities can be $\sum_{k-1}^{n-1}\binom{n-1}{k-1} = 2^{n-1}$. There are other adaptations where the number of changes is known in advance, lowering the number of possible outcomes. Naturally, multiple change point detection algorithms can be used on datasets which have only one change, however, the opposite is not that trivial. Few possible procedures can be used together with the detection algorithm. One of the approaches is the sliding window approach, where the window is formed over some part of the sequence which satisfies the problem constraints, then, this window can slide over the whole sequence and capture different portions of that. Another procedure is called *binary segmentation* where one can search for a change in the entire dataset. If a change-point is found then the dataset sequence is divided into two parts and repeats the steps on both segments in a recursive fashion [76].

The focus of this thesis is on the detection of *single*, *multiple*, and *unknown number* of change points in an *offline* setting.

Online and Off-line environment

In CPD, the concept of "*online*" and "*offline*" refers to the availability of the data being analyzed.

"*Offline*" (see [74] selective survey of algorithms) change point detection refers to the analysis of a complete, fixed dataset. The goal of offline change point detection is to identify changes that have occurred in the past, based on the complete dataset. This type of change point detection is also known as "*a-posteriori*" change point detection, as the analysis is performed after the data has been collected. Another important aspect is that, with all the data, the number of changes can be more than one. In this case, the problem becomes more complex and an additional task is to determine the total number of changes. A particularly relevant problem is identifying epidemic changes when the change is temporal (as shown in Fig. 1.4a and 1.5b)

"*Online*" (see [65] for comparison of online CP algorithms) change point detection, on the other hand, refers to the analysis of a dataset that is continuously growing. The goal of online change point detection is to identify changes as soon as they occur, in real-time. This requires the use of algorithms that can continuously update their results as new data becomes available. Online change point detection is useful in situations where it is important to react to changes as soon as they happen, such as in monitoring the performance of a

machine or the health of a patient.

Online methods can be adapted for use in an offline setting by iteratively processing data points. However, adapting offline methods for use in an online setting presents more challenges. One potential approach is to utilize the entire recorded time series each time a new data point is received, though this may be impractical due to computational constraints, particularly for time series with a large number of data points recorded over an extended period. There may also be a lag in detection, as the algorithm may not have the necessary sensitivity to detect the change as promptly as online methods.

Offline methods have some advantages over online methods. They can potentially be more accurate as they can take into account the entire available data set. They also have the added flexibility of allowing for the analysis of multiple change points.

Parametric and non-parametric methods

Change point detection methods can be broadly classified into two main categories: *parametric* and *nonparametric* methods. Parametric methods assume a specific underlying probability distribution for the data and estimate the change points by maximizing the likelihood function or minimizing the distance between the data and the assumed distribution (see [16] for extensive review of parametric approaches). Nonparametric methods, on the other hand, do not assume any specific underlying distribution for the data, and instead use a variety of heuristic and data-driven approaches to identify the change points. The reader may refer to the book [13] by Brodsky and Darkhovsky for extensive review on nonparametric change point detection.

Both parametric and nonparametric methods have their own advantages and limitations, and the choice of method will depend on the specific characteristics of the data and the goals of the analysis. Some common parametric methods for change point detection include the CUSUM method, the Shewhart chart, and the Binary Segmentation method. Nonparametric methods include the Mann-Kendall test, the ChangeFinder method, and the Pelt method. There are many other methods and algorithms available, and researchers and practitioners may choose to use a combination of approaches to get the best results.

Approaches to change point detection

First, we provide a brief overview of some classical sequential procedures for CPD.

Control charts

One method for identifying change points in a process is to establish a threshold value and determine if the statistic being measured exceeds this threshold.

If it does, it can be interpreted as indicating a change in the process. The effectiveness of this approach depends heavily on the choice of the threshold value. The first effort to select a threshold in a more systematic manner was undertaken by Shewhart [71]. In this approach, it is assumed that the true mean, represented by μ', and the variance, represented by σ^2, of the process are known. The method is then applied to a batch of size n, with the goal of detecting any changes that may have occurred in the process. For a batch of size n

$$\overline{X}(t) = n^{-1} \sum_{i=n(k-1)+1}^{nk} X_i.$$

The change is detected if it satisfies the following inequality

$$|\overline{X}(t) - \mu'| > \beta \frac{\sigma}{\sqrt{n}},$$

where β is a constant.

CUSUM tests

Page [50] approach is to consider a CP not based on the statistic crossing threshold but if it moves far away from the historical minimum. For sequential detection, Page suggested the stopping rule

$$\min \left\{ m : S_m - \min_{0 \le i < m} S_i \ge h \right\},$$

where h is a positive constant, $S_m = \sum_{k=1}^{m} \psi(k)$, ψ is a score function of the k-th sample. It is common to use the *log-likelihood ratio*, assuming that X_i are drawn from the distribution $\Phi_i, i = 0, 1$ with a parameter $\theta \in \Theta$, equal θ_0 and θ_1 after the change, we set

$$\psi(k) = \sum_{i-n(k-1)+1}^{nk} \log \frac{\Phi_0(X_i; \theta_0)}{\Phi_1(X_i; \theta_1)}.$$

This test works only if we have a *positive* change in the mean of the score. One way to improve this, one can use two CUSUM algorithms together [50]. Barnard [7] suggested a special case, when X is normally distributed with mean 0 and σ^2, hence $\Phi_0(X_i; \theta_0) = \mathcal{N}(0, \sigma^2)$ and $\Phi_1(X_i; \theta_1)$ is normal with a mean σ. Then $\log[\Phi_0/\Phi_1] = \sigma x - \sigma^2/2$ is maximized with respect to σ at $\hat{\sigma} = x$. One should test for a change in a normal mean with an initial value of 0 by using the following expression:

$$\min \left\{ i : \max_{m \le i} \frac{|S_m - S_i|}{\sigma(i-m)^{1/2}} \ge h \right\}.$$

This equation tests for a change in mean by comparing the difference between the cumulative sum of the data and the cumulative sum of the mean with a threshold value h.

Filtered Derivatives

The filtered derivative method was proposed by Basseville [8]. The way of estimating the instants at which jumps occur consists in filtering a signal and comparing its derivative to a threshold. Define a moving average

$$g_k = \sum_{i=0}^{n-1} \gamma_i \log \frac{\Phi_0(X_{k-i}; \theta_0)}{\Phi_1(X_{k-i}; \theta_1)},$$

where the γ_i are positive weights. Then, consider a discrete derivative $\nabla_{g_k} := g_k - g_{k-1}$. If a sufficient number of discrete derivatives is above the threshold h

$$\sum_{i=0}^{n-1} \mathbf{1}_{\nabla_{g_k-1} > h} \geq \eta.$$

Note, that the parameter η normally should be small *i.e.* $\eta \approx 2$.

This concludes an overview of well-known sequential procedures for CPD. While most of the online algorithms are centered around stopping rule and threshold the idea for offline algorithms closely relates to the statistical hypothesis testing discussed in section 1.3. It can also be treated as an estimation problem, where CP is seen as parameters from the model one wish to estimate.

Likelihood ratio tests

In parametric CPD, a *likelihood ratio (LR)* is frequently found in the literature. Early works by Quandt [58, 59] about changes in the conditional mean of normally distributed data assumed that the likelihood theory was *Sup-F* or *Sup-Wald* test. Assume, the following form $\Phi_i = \hat{\Phi}(\theta_i)$, where $\theta_i = (\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,k})$ is a vector of the parameters which define the distribution Φ_i. Then, the likelihood estimation can be defined as

$$\mathcal{L}(\theta \mid \mathbf{x}) = \mathcal{L}(\theta_1, \ldots, \theta_k \mid x_1, \ldots, x_D) = \prod_{k=1}^{D} f(x_t \mid \theta_1, \ldots, \theta_k),$$

where $f(\cdot|\theta)$ denotes the probability density function of $\hat{\Phi}$. The likelihood function is maximized to estimate the parameter vector θ of x sequence. For the model containing a CP at some position τ to one without CP, we have to compare supremum over the parameter spaces. In case, the τ is not known, the double-supremum of the overall likelihood function over all candidate points τ and the corresponding parameter spaces are compared to the supremum of the likelihood over the parameter space without a change. This defines the

likelihood ratio for the estimation of a CP,

$$\mathcal{LR}(\mathbf{x}) = \frac{\sup_{\tau} \sup_{\theta_i \in \Theta_i, i=\{1,2\}} \mathcal{L}\left(\boldsymbol{\theta_1} \mid x_t, t = 1, \ldots, \tau_1\right) \mathcal{L}\left(\boldsymbol{\theta_2} \mid x_t, t = \tau_1 + 1, \ldots, D\right)}{\sup_{\theta \in \Theta} \mathcal{L}(\theta \mid x_t, t = 1, \ldots, D)}$$

where Θ and Θ_1, Θ_2 denotes parameter spaces with and without CP respectively. For any integer n, the *segmentation* integers $1 \leq D \leq n$ and $\tau_0 := 0 < \tau_1 < \cdots < \tau_{D-1} < \tau_D = n$. Then, *segments* are defined as $\{1, \ldots \tau_1\}, \ldots \{\tau_1 + 1, \ldots, n\}$. These segments represent the different parts of the data that have distinct characteristics.

CUSUM extensions

In an offline setting, the CUSUM approach is extended to choosing the H_0 that maximizes the likelihood function of the hypothesis. In this sense, τ is estimated by

$$\widehat{\tau_1} \in \underset{1 \leq t \leq n}{\arg\max} \left\{ \sum_{j=1}^{t} \log \Phi_{\theta_0}(x_j) + \sum_{j=t+1}^{n} \log \Phi_{\theta_1}(x_j) \right\}.$$

With a one-sided change in the mean and known σ^2 of Gaussian observations, Page [51] proposes to define the CUMSUM as

$$S_t := \sum_{i=1}^{t} \left(X_i - \theta_0 + \sigma\delta\right).$$

The H_0 is rejected if $S_n - \max_{t<n} S_t < -h$, $h > 0$.

Model selection

The concept of CPD is widely accepted within the academic community. However, selecting the most appropriate approach for a given situation is crucial, as it can greatly impact the accuracy of the results. No single model is capable of effectively detecting all types of changes, so it is important to carefully consider the specific characteristics of the process being analyzed and the types of changes that are expected to occur. Formally, CPD can be considered a model selection problem, in which the goal is to identify the optimal segmentation, represented by $\mathbf{T}$, based on a quantitative criterion, represented by $\mathbf{V}(\mathbf{T}, y)$, that should be minimized [74]. If the wrong model is chosen, it may lead to a high number of false positive results.

It is also important to consider the computational complexity of change point detection methods, especially when they are being used in online applications that require real-time processing. Sensors can generate a large volume of data per second, but may have limited computing resources. Some well-

known CPD methods can have a time complexity of $O(n^3)$ [1], which can be a significant limitation in terms of processing speed. To address this issue, researchers have developed search methods as a resolution procedure for discrete optimization problems in CPD. These methods aim to improve the efficiency and speed of the change point detection process, enabling it to be used in resource-constrained environments with high volumes of data.

Overall, CPD methods can generally be divided into two categories: optimal detection algorithms and approximate solution methods. Optimal detection algorithms are designed to find the exact solution to the CPD problem by considering all possible combinations and selecting the one that minimizes the objective function. However, this approach is not practical in most cases due to the high computational demands. In contrast, approximate solution methods provide approximate results that are faster to compute, but may be less accurate. These methods are more suitable for practical applications where computational efficiency is a concern.

Areas of application

The CPD problem was originally developed for statistical quality control, but it has since become an important tool in various fields where signal analysis is used. Extracting quantitative features from signals requires the use of solid mathematical models, and CPD is an active area of research in many different domains. In this section, we will provide a brief overview of some of the areas where CPD has been applied extensively.

Application on medical data

Nowadays there exist many sensors and diagnostic devices which help doctors monitor and diagnose health-related issues. No wonder that CPD is an important component of overall signal analysis.

German psychiatrist Hans Berger recorded the first human Electroencephalogram (EEG) in 1924 [30]. He showed that actions such as closing the eyes rise could be detected in the EEG. In contrast, during mental activity, the oscillations of the EEG has a higher frequency and was less regular. 70 years after the first EEG recordings it attracted the attention of many researchers who focused on mathematical models related to the analysis of the EEG signals. For example, during epileptic seizures the brain activity transitions between different states, which can be monitored with EEG. Malladi et. al. [46] propose an online bayesian CPD for an epileptic activity for real-time monitoring.

An ECG signal analysis is another area where CPD researches focus. An ECG is a test that is used to check the heart's rhythm in order to assess cardiovascular diseases. The key step for an ECG analysis is to segment the signal and locate its constitutive waves. Most of common approaches of ECG wave

detection are based on wavelet transformation [53, 57], Hidden Markov models [41] and simple mathematical operations such as differentiation, integration, and squaring. Other variants such as the graph-constrained changepoint detection approach proposed by Fotoohinasab et. al. [25]. We refer the reader to Beraza et. al [10] study for an in-depth review of a wide range of ECG segmentation algorithms.

Image and video analysis

Detecting changes in video or image data is a widely researched topic in the field of computer vision, with applications in areas such as surveillance, remote sensing, and medical image analysis. For a more detailed review of this topic, see Radke et al. [60]. Another closely related problem is shot detection, which involves identifying the longest continuous sequences within a video and organizing the data into more compact forms or extracting semantically meaningful information [19].

Magnetic resonance imaging (MRI) is a medical procedure that uses strong magnetic fields and radio waves to produce detailed images of the inside of the body. MRI scans are used to examine almost any part of the body, and the detection of small changes between scans can reveal how diseases progress over time. However, the change detection in 3D MRI data remains a challenging problem in computer vision. Some progress has been made in detecting state changes in MRI data [12, 47, 68].

Human activity analysis

There have been significant advancements in internet of things (IoT) wearable sensor technology, including devices like smartwatches and smartphones that continuously record human activity (e.g. [1, 42, 44]). One application of CPD in this field is the segmentation of signals and the mapping of human activity, such as sitting, standing, walking, or sleeping. By identifying changes in the signal data, it is possible to determine the specific activities that an individual is engaging in. This can be useful for a variety of purposes, such as monitoring physical activity levels, assessing health and wellness, and identifying patterns of behavior.

Speech signal analysis

There are many different applications for voice activity detection (VAD) in the field of speech analysis. VAD algorithms are used to identify periods of conversational speech within a signal, which can then be further analyzed or processed. This is useful for a variety of purposes, such as automatic speech recognition, speech communication over networks, speech coding, speech augmentation, and echo cancellation. In order to detect periods of speech within a

signal, VAD algorithms typically rely on CPD methods to identify changes in the signal data that correspond to transitions between speech and non-speech segments (see [35, 37, 73]).

In order to accurately identify speech segments, VAD algorithms must be able to distinguish between speech and non-speech sounds based on various characteristics of the signal data. These characteristics may include the spectral content of the signal, the presence of periodic patterns, the intensity of the signal, and other features that are unique to speech sounds. VAD algorithms may also take into account contextual information, such as the presence of other sounds or the history of the signal, in order to make more accurate predictions.

Overall, VAD algorithms are an important tool in the field of speech analysis, as they enable the automatic separation of speech segments from non-speech segments, which can be useful for a variety of purposes.

Application to climate changes

Climate change is a complex and multifaceted phenomenon that affects a wide range of environmental, social, and economic systems. As such, understanding and predicting the impacts of climate change requires the analysis of a wide range of data sources and variables. CPD methods have proven to be useful in this context, as they can help to identify trends and changes in climate data over time.

For example, CPD has been used to detect abrupt changes and variations in rainfall patterns [48, 52] and temperature trends [39]. By analyzing these variables, researchers can gain insight into the impacts of climate change on local and regional weather patterns, and how these patterns may evolve over time. This information can be used to inform the development of climate change mitigation and adaptation strategies, such as water resource management, land use planning, and the design of infrastructure and buildings to withstand extreme weather events.

In addition to analyzing data on weather and temperature, CPD methods have also been applied to other climate-related variables, such as sea level rise, atmospheric concentrations of greenhouse gases, and ocean temperatures and currents. By detecting changes in these variables, researchers can gain a more comprehensive understanding of the impacts of climate change and the complex interactions between different environmental systems.

In this section, we have highlighted a number of applications in which CPD plays a significant role. However, the list of applications is by no means exhaustive, as CPD methods are also widely used in other domains. Some examples of the types of data and applications that can also benefit from the use of CPD methods include:

- Financial data (e.g. [6,26]): CPD methods can be used to detect changes in financial markets, such as changes in stock prices, exchange rates, or interest rates. These methods can help analysts and investors to identify trends and make informed decisions about investments and risk management.

- Biological data (e.g. [56]): CPD methods can be used to analyze biological data, such as gene expression data, protein expression data, or microbiome data, in order to identify changes and trends that may be relevant to the understanding of various biological processes. This can be useful for researchers studying diseases, development, and evolution.

- Social network data (e.g. [38]): CPD methods can be used to analyze data from social networks, such as Twitter or Facebook, in order to identify changes and trends in social interactions and behavior. This can be useful for researchers studying social networks, marketing, and public opinion.

- Natural language data (e.g. [27]): CPD methods can be used to analyze data from natural language sources, such as text or speech data, in order to identify changes and trends in language use. This can be useful for researchers studying language evolution, language processing, and social interactions.

In fact, any analysis of a sequence of events can benefit from the use of CPD methods, as these methods are well-suited for identifying changes and trends in data over time.

1.4. Change point detection for functional data

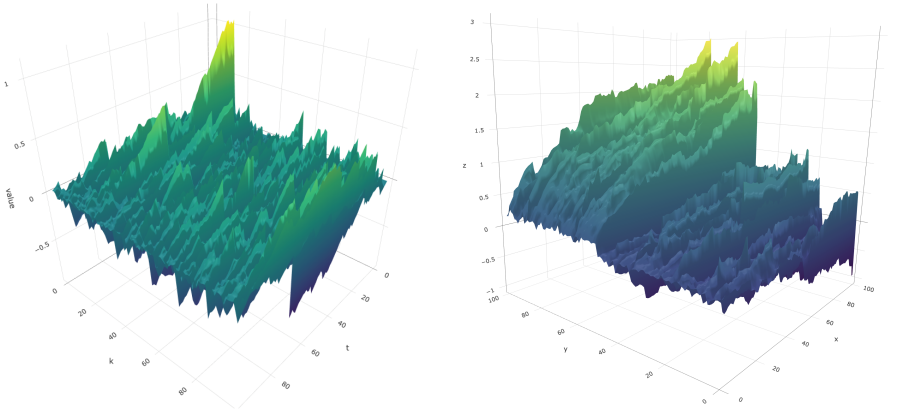
In this section, we will review some of the recent developments in the use of change point detection (CPD) methods for functional data (FD).

In section 1.2 we covered the basic principles of FDA. In this section we will review some of the recent developments of CPD for FD.

In a univariate case which we covered in previous sections, changes were measured between observed points. In FDA case, changes have to be measured between curves. This, makes the problem more challenging. The figure 1.6 show two functional samples: figure 1.6a without change, while figure 1.6b illustrate the functional sample with a change point in the middle of the dataset.

Despite the challenges, the adoption of a CPD statistical test to FDA is accelerating in recent years. A functional data CPD was successfully applied in many areas.

A cumulative sum (CUSUM) test was proposed by Berkes et al. [11] for independent functional data by using projections of the sample onto some principal



(a) An example of a simulated functional sample without change point.

(b) An example of a simulated functional sample with change.

Figure 1.6: Illustration of a Functional Data Sample with and without Change Points

components of a covariance γ. The test statistics proposed by Berkes et al. was defined as

$$S_{N,d} := \frac{1}{N^2} \sum_{l=1}^{d} \hat{\lambda}_\ell^{-1} \sum_{k=1}^{N} \left(\sum_{1 \le i \le k} \hat{\eta}_{i,\ell} - \frac{k}{N} \sum_{1 \le i \le N} \hat{\eta}_{i,\ell} \right)^2 ,$$

where $\hat{\lambda}_\ell$ are eigenvalues, $\hat{\eta}_{i,\ell}$ are scores corresponding to the largest d eigenvalues and $\hat{v}_l$ are eigenfunctions:

$$\hat{\eta}_{i,\ell} = \int \left\{ X_i(t) - \overline{X}_N(t) \right\} \hat{v}_l(t) \, \mathrm{d}t, \quad i = 1, \ldots, N, l = 1, \ldots, N.$$

Denoting by $c_d(\alpha)$ its $(1-\alpha)$ quantile, the test rejects H_0 hypothesis if $S_{N,d} > c_d(\alpha)$. The distribution of the random variable was derived by Kiefer [40].

Later, the problem was studied by Aue et al. [3], where its asymptotic properties were developed. The assumption of independence in many applications is too unrealistic. More often the observation is dependent to some degree on previous observations. Hörmann et al. [33] in their work weakly dependent functional data extended the test to accommodate the dependence.

The problem was further studied by Banerjee et al. [5]. Authors have proposed an alternative estimator of the covariance kernel, which is a consistent estimator of the true covariance kernel under the null hypothesis for both the

independent and the dependent data [5]. The test is defined as

$$H_{N,d} := \frac{1}{N^2} \sum_{l=1}^{d} \sum_{[Nu]=1}^{N} \frac{1}{\widehat{\lambda}_l^{(u)}} \left(\sum_{i=1}^{[Nu]} \hat{\eta}_{i,l}(u) - u \sum_{i=1}^{N} \hat{\eta}_{i,l}(u) \right)^2 .$$

It is proven [5], that $H_{N,d} \xrightarrow{d} \int_0^1 \sum_{l=1}^{d} B_l^2(u) \mathrm{d}u$ under the null hypothesis, where $B_1(\cdot), B_2(\cdot), \ldots, B_d(\cdot)$ denotes the standard independent Brownian bridges. The test rejects H_0 if value of $H_{N,d}$ is bigger than the tabulated $(1-\alpha)$th quantile $K_d(\alpha)$ in Berkes et al. [11].

Aston and Kirch [2] proposed an estimator of a CP in a model for an AMOC and an epidemic change as well as accounted for a wide class of dependency structures. Aue et al. [4] proposed a fully functional method for finding a change in the mean without losing information due to dimension reduction. Despite these advances, there are several outstanding issues with these approaches. Namely, computational scalability, an insufficient power to detect covariance and shape-based alternatives, and a lack of robustness. To address those problems T. Harris, Bo Li, and J. D. Tucker [31] proposed the multiple change-point isolation method for detecting multiple changes in the mean and covariance of a functional process. In contrast to all former approaches, Sharipov et al. [70] developed a test for structural changes in functional data which is based on Hilbert space theory. Recently, Li et. al. [43] proposed a method of finding multiple change points in a mean of functional data using a Bayesian approach, where the change was viewed as the result of an evolutionary process changing some wavelet coefficients at a time.

2. Testing change in the mean: univariate case

2.1. Introduction

In this chapter, we propose a new test statistic for the detection of a mean change in the sequence of observations which is built on the p-variation of a corresponding CUSUM process. More precisely, for a sample $X_1, X_2, \ldots, X_n$ and the number $p > 2$, we define the test statistics

$$T_{p,n}(X_1, \ldots, X_n) = \max\left\{\sum_{j=1}^{m}\left|\sum_{i=k_{j-1}+1}^{k_j}(X_i - \overline{X}_n)\right|^p : 0 = k_0 < \cdots < k_m = n; 1 \le m \le n\right\},$$

where $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$. As a theoretical support for changed segment tests based on this statistic, we establish its asymptotic distribution under the null hypothesis (see Theorem 9). We also evaluate finite sample performance through a simulation study.

The chapter is organized as follows. In section 2.2 we introduce a space of functions with bounded p-variation as a functional framework for the CUSUM process.

There we also present asymptotic results that are relevant to test statistics defined and discussed in section 2.3. Section 2.4 presents a Monte Carlo analysis of the finite sample properties of the tests and compares their size and power performance. We also compared the performance of other well-known change-point detection algorithms. Finally, in section 2.4.2 we demonstrate how the proposed method can be applied to real-world data.

2.2. Asymptotics of stepwise CUSUM process

Consider univariate time series, $X_k, k = 1, 2, \ldots$. For each $n \ge 1$ and each $t \in [0, 1]$, let

$$S_n(t) = 0 \text{ if } t \in [0, 1/n),\ S_n(t) := \sum_{i=1}^{\lfloor nt \rfloor} X_i,\ \text{if } t \in [1/n, 1],$$

where for a real number $x \ge 0$, $\lfloor x \rfloor := \max\{k : k \in \mathbb{N}, k \le x\}, \mathbb{N} = \{0, 1, \ldots\}$.

The p-variation for partial sums is defined as

$$v_p(f; [0, t]) := \sup\left\{\sum_{j=1}^{m} |S_j - S_{j-1}|^p\right\}.$$

The stepwise CUSUM process $Z_n = (Z_n(t), t \in [0,1])$ is defined as

$$Z_n(t) = S_n(t) - \frac{\lfloor nt \rfloor}{n} S_n(1) = \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \overline{X}_n).$$

Classical examples of path spaces for the process Z_n include the Hilbert space $L_2[0,1]$ and the Skorohod space $D[0,1]$. Under various assumptions, a functional central limit theorem is established in these spaces. For example, the classical Donsker theorem for i.i.d. sequence (X_n) with finite second moment states that

$$n^{-1/2} Z_n \xrightarrow[n\to\infty]{\mathcal{D}} \sigma B \ \text{ in the space } \ D[0,1],$$

where $B = (B(s) := W(s) - sW(1), s \in [0,1])$ is a standard Brownian bridge, and $W = (W(s), s \in [0,1])$ is a standard Wiener process, and $\sigma^2 = \mathrm{var}(X_1)$. The symbol $\xrightarrow[n\to\infty]{\mathcal{D}}$ means convergence in distribution.

In this section, we consider a step-wise CUSUM process Z_n in the path spaces of functions with bounded p-variation.

It is straightforward to find that for any $p \in (0, \infty)$,

$$v_p(Z_n) = \max \left\{ \sum_{j=1}^{m} \left| \sum_{k=k_{j-1}+1}^{k_j} (X_k - \overline{X}_n) \right|^p \right\},$$

where the maximum is taken over $0 = k_0 < \cdots < k_m = n$, and $1 \le m \le n$. Hence, $Z_n \in \mathcal{W}_p[0,1]$ for any $p > 0$, and $T_{p,n}(X_1, \ldots, X_n) = v_p(Z_n)$. To obtain limit distribution of the statistics $T_{p,n}(X_1, \ldots, X_n)$ we consider the limit behaviour of (Z_n) in the paths space $\mathcal{W}_p[0,1]$. Since finite-dimensional distributions of the process Z_n converge to those of a Brownian bridge we need to restrict to the index $p > 2$.

Since $\mathcal{W}_p[0,1]$ is a non-separable space we use convergence in law concept as defined by Hoffmann-Jørgensen denoting it by $\xrightarrow[n\to\infty]{\mathcal{D}^*}$.

Theorem 7. Fix $p > 2$. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables and let $S_n = (S_n(t), t \in [0,1])$ be the partial sum process. If $\sigma^2 := EX_1^2 < \infty$, then the convergence

$$n^{-1/2} Z_n \xrightarrow[n\to\infty]{\mathcal{D}^*} \sigma B \ \text{ in } \ \mathcal{W}_p[0,1]$$

holds.

Proof. It is proved by [49] that for $p > 2$,

$$n^{-1/2} \sigma^{-1} S_n \xrightarrow[n\to\infty]{\mathcal{D}^*} W \ \text{ in the space } \ \mathcal{W}_p[0,1].$$

Consider the mappings T_n and T defined for any function $f : [0,1] \to \mathbb{R}$ by

$$T_n f(t) = f(t) - \frac{\lfloor nt \rfloor}{n} f(1), \quad Tf(t) = f(t) - tf(1), \quad t \in [0,1].$$

Both functions T_n and T map $\mathcal{W}_p[0,1] \to \mathcal{W}_p[0,1]$ and both are linear and bounded, hence continuous. Since the continuous mapping theorem is valid for the space $\mathcal{W}_p[0,1]$ (see, e.g., [49]), we have

$$n^{-1/2}\sigma^{-1} T S_n \xrightarrow[n\to\infty]{\mathcal{D}^*} TW \text{ in the space } \mathcal{W}_p[0,1].$$

Next observe that

$$n^{-1/2}\sigma^{-1} Z_n = n^{-1/2}\sigma^{-1} T_n S_n = n^{-1/2}\sigma^{-1} T S_n + o_P(1). \tag{2.1}$$

Indeed

$$||T_n S_n - T S_n||_{[p]} \le ||I_n||_{[p]} |S_n(1)|,$$

where $I_n(t) = t - \lfloor nt \rfloor / n$. By the central limit theorem $n^{-1/2}\sigma^{-1}|S_n(1)| = O_P(1)$ whereas $||I_n||_{[p]} \le n^{-1+1/p}$. Hence, equation 2.1 is true and this relation completes the proof. □

The continuous mapping theorem yields

Corollary 8. Fix $p > 2$. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables and let Z_n be the step-wise CUSUM process. If $EX_1^2 = \sigma^2 \in (0, \infty)$ then

$$G(n^{-1/2}\sigma^{-1} Z_n) \xrightarrow[n\to\infty]{\mathcal{D}} G(B)$$

for any continuous functional $G : \mathcal{W}_p[0,1] \to \mathbb{R}$.

Among many interesting examples of continuous functional G we pick out the p-variation of a function $G_1(f) = v_p^{1/p}(f)$.

The following theorem is the theoretical justification for statistics used to detect a change point in the mean.

Theorem 9. Fix $p > 2$. Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables. If $EX_1^2 = \sigma^2 \in (0, \infty)$ then

$$n^{-1/2}\sigma^{-1} T_{p,n}^{1/p}(X_1, \ldots, X_n) \xrightarrow[n\to\infty]{\mathcal{D}} v_p^{1/p}(B).$$

Proof. The result follows directly from Corollary 8 □

Approximate values of the distribution of $v_p(B)$ can be obtained by using Theorem 9 with standard normal random variables. Assume that we have for

each $n \geq 1$ and each $M \geq 1$ a collection of independent standard normal random variables $\{Y_{kj}, k = 1, \ldots, n; j = 1, \ldots, M\}$. Let for each $j = 1, \ldots, M$,

$$v_p(n, j) = \max \left\{ \sum_{i=1}^{m} \left| \sum_{k=k_{i-1}+1}^{k_i} (Y_{kj} - \overline{Y_j}_n) \right|^p \right\},$$

where the maximum is taken over $0 = k_0 < \cdots < k_m = n$, and $1 \leq m \leq n$, and $\overline{Y_j}_n = n^{-1}(Y_{1j} + \cdots + Y_{nj})$. Consider

$$F_{nM}(x) := \frac{1}{M} \sum_{j=1}^{M} \mathbf{1}(n^{-1/2} v_p^{1/p}(n, j) \leq x), \ x \geq 0.$$

By the law of large numbers for any $n \geq 1$,

$$\lim_{M \to \infty} F_{nM}(x) = P(n^{-1/2} v_p^{1/p}(n, 1) \leq x), \ x \geq 0.$$

On the other hand, $n^{-1/2} v_p^{1/p}(n, 1) \xrightarrow[n \to \infty]{\mathcal{D}} v_p^{1/p}(B)$, by Corollary 8. Hence,

$$F_{n,M}(x) \approx P(v_p^{1/p}(B) \leq x).$$

for large n, M.

2.3. Application to change point problem

For independent random sample $X_1, \ldots, X_n$ consider the model

$$X_i = \delta \mathbf{1}_{(k^*, n]}(i) + Y_i, \quad i = 1, \ldots, n,$$

where $Y_1, \ldots, Y_n$ are i.i.d. random variables with $E(Y_i) = 0$, $E(Y_i^2) = 1$ and $\delta \in R, k^* \in \{1, \ldots, n\}$ are unknown parameters.

Under $H_0 : \delta = 0$, we have for any $p > 2$,

$$n^{-1/2} v_p^{1/p}(Z_n) \xrightarrow[n \to \infty]{\mathcal{D}} v_p^{1/p}(B).$$

Under contiguous alternative where $\delta = \delta_n \approx \sqrt{n}\delta^*$, and $k^* = \lfloor n\theta^* \rfloor$ with some $\delta^* > 0$, and $\theta^* \in (0, 1)$, it holds

$$n^{-1/2} v_p^{1/p}(Z_n) \xrightarrow[n \to \infty]{\mathcal{D}} v_p^{1/p}(B - f),$$

where $f(t) = \begin{cases} \delta^* t(1 - \theta^*) & \text{if } \ 0 \leq t < \theta^* \\ \delta^* \theta^* (1 - t) & \text{if } \ \theta^* \leq t \leq 1 \end{cases}$.

In the next section, we carry out some simulations to study the performance

of the proposed approach to the change-point detection problem.

2.4. Simulation experiments

The statistical power is estimated by Monte-Carlo simulations. Then, we investigate the "sliding window" approach for change-point localization and multiple change-point detections. Finally, the test is applied to a well-known dataset of the annual flow of the river Nile at Aswan (formerly Assuan), 1871–1970, in $10^8 m^3$ with apparent change-point near 1898 [18].

Throughout this section, the following parameters are used n - number of observations, k - number of simulations, τ - change point position, a_α - critical value.

Consider a sequence $X_1, \ldots, X_n$, $n \geq 1$, of independent random variables. Suppose there exists a $\tau \in [0, 1]$ such that $X_1, \ldots, X_{n\tau}$ have the distribution $\mathcal{N}(\mu_1, \sigma^2)$ and $X_{n\tau+1}, \ldots, X_n$ have the distribution $\mathcal{N}(\mu_2, \sigma^2)$. Given α we construct consistent critical point a_α such that

$$P\left(v_p^{1/p}(Z_n) < a_\alpha\right) = \alpha$$

for $H_0 : \tau = 0$ versus $H_1 : \tau \in (0, 1]$.

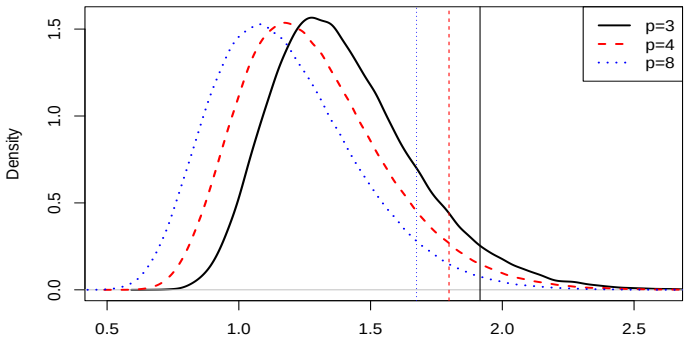
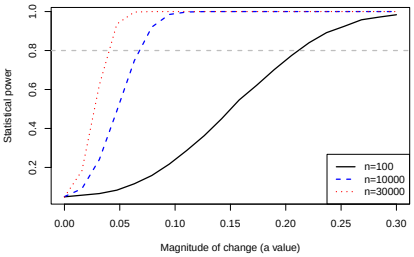


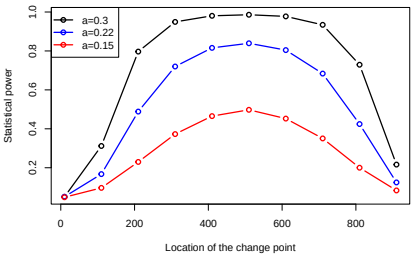
Figure 2.1: The probability distribution of $v_p^{1/p}(Z_n)$ with critical values at a significance level of $\alpha = 0.05$ highlighted

The distribution of the $v_p^{1/p}(Z_n)$ was estimated using Monte-Carlo technique, where parameters were set to $n = 1000$, $k = 100000$. The figure 2.1 show distribution with different p-variation $p = \{3, 4, 8\}$ values and marked critical values at asymptotic level $\alpha = 0.95$.

During experiments, every simulated data has a common structure. In a general setting, a sequence of observations $x_1, x_2, \ldots, x_n$ is drawn from the i.i.d.



(a) This figure illustrates the statistical power at different magnitudes of change (a value).



(b) This figure illustrates the statistical power at different change point location (τ value).

Figure 2.2: These figures illustrate the statistical power at different levels of magnitude change and change point location

random variables $X_1, X_2, \ldots, X_n$ and undergoes abrupt change in mean from μ_1 to μ_2 at a point τ. Recall that $X \sim \mathcal{N}(\mu, \sigma^2)$ means that X has Gaussian distribution with mean μ and variance σ^2. So, X_i is defined as

$$X_i \sim \begin{cases} \mathcal{N}(\mu_1, \sigma^2) & \text{if } i \leq n\tau \\ \mathcal{N}(\mu_2, \sigma^2) & \text{if } i > n\tau \end{cases}.$$

Consider the simulated copies of $Z_{nj}(t)$, $j = 0, \ldots, k$ of the process $Z_n(t)$ where k is subject to choice.

We estimate the power of detection by investigating Type II errors, given that a change-point exists. In this case, an alternative hypothesis is correct with level $\alpha = 0.95$.

Our aim is to evaluate the statistical power under different circumstances. First, we focus on the magnitude of the change by gradually increasing a value, $\mu_2 = a$. The figure 2.2a show statistical power with respect to different a values with different number of observations $n = \{1000, 10000, 30000\}$. The simulation results show that if the number of observations is large enough ($n \geq 30000$), then the null hypothesis is correctly rejected more than 80% times with $\mu_2 > 0.035$.

The second objective of this study is to evaluate the statistical power with respect to change point location (τ) using a sample size of $n = 1000$. The change point location (τ) parameter is used to indicate the location of the change point, which is a crucial aspect in determining the ability to detect changes quickly. The results of this simulation are presented in Figure 2.2b. It can be observed that with a difference in means ($|\mu_1 - \mu_2|$) of at least 0.3, a change point location (τ) as small as 0.20 is sufficient to achieve a statistical power of 80%

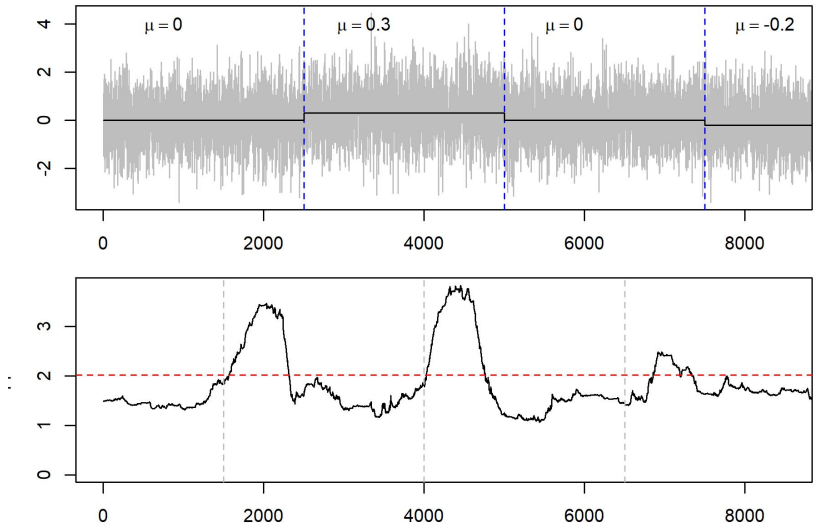


Figure 2.3: Sliding window approach. The top plot illustrates the whole process with changes marked in blue dashed lines. The figure at the bottom shows p-variation at each window position. The Red dashed line marks a critical value a_α, with $\alpha = 0.95$

2.4.1. Sliding window approach

In the previous section, we showed that the proposed algorithm is effective at identifying change points in data when all of the data are available for analysis, which is known as an "offline" approach. However, it is often the case that data arrive in batches or streams, rather than all at once. In this "semi-online" scenario, it can be useful to analyze only a subset of the data in order to find and localize multiple change points more efficiently. To address this, we have extended the method to use a sliding window approach, where a subset of the data is analyzed at a time. This allows for a more efficient and computationally less intensive analysis of the data, while still being able to detect change points.

The sliding window approach is a method used to analyze data that arrives in a stream or in batches, rather than all at once. It involves dividing the data into overlapping or non-overlapping windows of a fixed size, and then applying a statistical test to each window. By moving the window along the data stream and analyzing the data in each window, the sliding window approach allows for the analysis of data in a more efficient and computationally less intensive manner, compared to analyzing the entire dataset at once.

Sliding window approaches have been used in a variety of contexts, including signal processing, image processing, natural language processing, and data mining. They are particularly useful when dealing with large datasets that cannot be processed all at once, or when analyzing data in real-time, as the

window can be moved along the data stream as new data becomes available. However, the choice of window size and the overlap between windows can affect the performance of the sliding window approach and should be carefully considered.

To perform CPD using a sliding window approach, a specific window size h is chosen and applied to a large buffer of data points $x_1, \ldots, x_n, n \geq 1$. The window starts at the first point ($w = 1$) and is then shifted to the right by a defined number of elements λ at each iteration. The objective is to test hypotheses about the data in the present window, which consists of the points $x_{\lambda w}, \ldots, x_{\lambda w+h}$, as the window moves from $w = 1$ to $w = n/\lambda$.

Simulations have shown that the sliding window method is effective at localizing change points in data (see figure 2.3). The red dashed line indicates the critical value with a significance level of $\alpha = 0.95$. Typically, the statistic used to detect change points will approach this critical value line around the point where a change is present.

2.4.2. Real Data

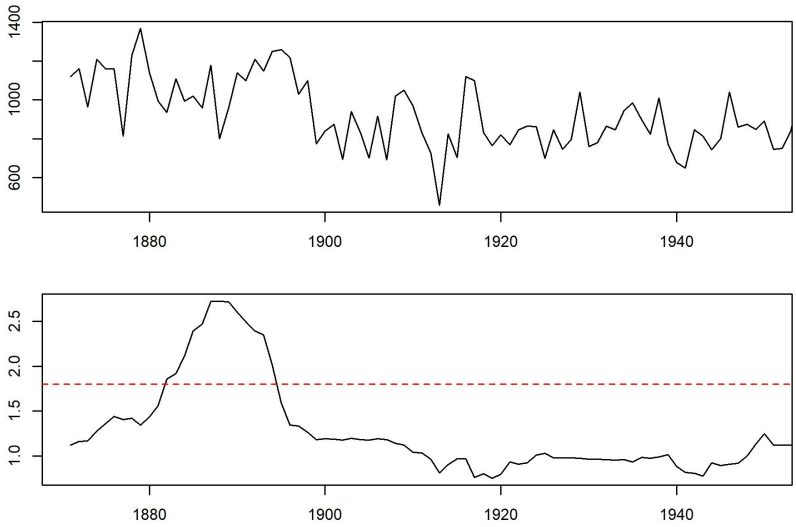


Figure 2.4: Data from the annual volume of discharge from the Nile River at Aswan. An apparent changepoint is visible near 1898.

As an example of a problem that can be addressed using change point detection, we consider the case of the River Nile. Specifically, we use measurements of the annual flow of the river at Aswan (formerly Assuan), which were recorded from 1871 to 1970 in units of $10^8 m^3$. According to [18], there appears to be a change point in the data near 1898. These measurements are meteorologically significant as they may provide evidence of a change in the pattern of rainfall in

the region. Figure 2.4 shows that there appears to be a decrease in the annual volume of the river after 1898.

2.5. Conclusions

In this chapter, we introduced a new test statistic for detecting changes in the mean of a sequence of observations. This statistic is based on the p-variation of the corresponding CUSUM process, and is designed to detect at most one change point in the process. We demonstrated through experiments that our approach is effective at detecting relatively small changes in the mean, and is computationally efficient for use with large data sets. While our approach works well in an offline setting with a single change point, it may not perform as well when multiple change points are present. To address this limitation, we demonstrated that using a "sliding window" approach can be an effective way to detect multiple change points.

3. Multiple change-point detection in a functional sample

The chapter is organized as follows. In Section 3.2, $\mathcal{G}$-sum and $\mathcal{G}$-CUSUM processes are defined and their asymptotic behavior is considered in a framework of the $\ell^\infty(\mathcal{G})$ space. The results presented in this section are used to derive the asymptotic distributions of the test statistics presented in Section 3.3. Section 3.4 is devoted to simulation studies of the proposed test algorithms. Section 3.5 contains a case study. Finally, Section 3.6 is devoted to the proofs of our main theoretical results.

3.1. Introduction

Consider a second-order stationary sequence of stochastic processes $Y_i = (Y_i(t), t \in [0,1]), i \in \mathbb{N}$, defined on a probability space $(\Omega, \mathcal{F}, P)$, having zero mean and covariance function $\gamma = \{\gamma(s,t), s,t \in [0,1]\}$. For a given functional sample $X_1(t), \ldots, X_n(t), t \in [0,1]$, consider the model:

$$X_k(t) = g(k/n, t) + Y_k(t), \quad t \in [0,1], \quad k = 1, \ldots, n, \tag{3.1}$$

where the function $g : [0,1] \times [0,1] \to \mathbb{R}$ is deterministic, but unobserved. Our main aim is to test for null and alternative hypothesis:

$$H_0 : \quad g = 0 \quad \text{versus} \quad H_1 : g \neq 0$$

with emphasis on a case of change-point detection, which corresponds to a piecewise-constant function g with respect to the first argument.

This model covers a broad range of real-world problems such as climate change detection, image analysis, analysis of medical treatments, especially magnetic resonance images of brain activities, and speech recognition, to name a few. Besides, the change-point detection model (3.1) can be used for knot selection in spline smoothing as well as for trend changes in functional time series analysis.

The methodology we propose is based on some measures of variation of the process:

$$W_n(s) = \sum_{k=1}^{\lfloor ns \rfloor} (X_k - \overline{X}_n) + (ns - \lfloor ns \rfloor)(X_{\lfloor ns \rfloor + 1} - \overline{X}_n), \quad s \in [0,1],$$

where $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$.

Since this process is infinite-dimensional, we used the projections technique to reduce the dimension. To this aim, we assumed that Y_i is mean-squared continuous and jointly measurable and that γ has finite trace: $\mathrm{tr}(\gamma) =$

$\int_0^1 \gamma(t,t)\,\mathrm{d}t < \infty$. In this case, Y_i is also an $L_2(0,1)$-valued random element, where $L_2 := L_2(0,1)$ is a Hilbert space of Lebesgue square integrable functions on $[0,1]$ endowed with the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t$ and the norm $\|f\| := \sqrt{\langle f, f\rangle}$.

In the case where the number of change points is known to be no bigger than m, our test statistics are constructed from (m,p)-variation (see the definition 1.1) of the processes $(\langle W_n(s), \psi\rangle, s \in [0,1])$, where $\psi \in \Psi \subset L_2(0,1)$ runs through a finite set Ψ of possibly random directions in $L_2(0,1)$. In particular, Ψ consists of estimated principal components. If the number of change-points is unknown, we consider the p-variation of the processes $(\langle W_n(s), \psi\rangle, s \in [0,1]), \psi \in \Psi$ and estimate the possible number of change-points.

3.2. $\mathcal{G}$-Sum Process and Its Asymptotic

Let $\mathcal{Q}$ be the set of all probability measures on $([0,1], \mathcal{B}_{[0,1]})$. For any $Q \in \mathcal{Q}$ and Q-integrable function f, $Qf := \int_0^1 f\,\mathrm{d}Q$. As usual, $\mathcal{L}_2([0,1], Q)$ is a set of measurable functions on $[0,1]$, which are square-integrable for the measure Q, and $L_2([0,1], Q)$ is an associated Hilbert space endowed with the inner product:

$$\langle f, g\rangle_Q = \int_0^1 f(t)g(t)Q(\,\mathrm{d}t), \quad f, g \in L_2([0,1], Q)$$

and corresponding distance $\rho_Q(f,g)$, $f, g \in L_2([0,1], Q)$. We abbreviate $L_2([0,1], \lambda)$ to L_2 and $\langle\cdot,\cdot\rangle_\lambda$ to $\langle\cdot,\cdot\rangle$ for Lebesgue measure λ. We use the norm $\|f\| := \sqrt{\langle f, f\rangle}$ and the distance $\rho(f,g) = \|f - g\|$ for the elements $f, g \in L_2$. On the set $L_2 \times L_2$, we use the inner product:

$$\langle (f,g), (f',g')\rangle_2 = \langle f, f'\rangle + \langle g, g'\rangle$$

and the corresponding distance:

$$\rho_2((f,g),(f',g')) = \Big(\|f - f'\|^2 + \|g - g'\|^2\Big)^{1/2}, \quad f, f', g, g' \in L_2.$$

For two given sets $\mathcal{F}, \Psi \subset L_2$, we consider the $\mathcal{F} \times \Psi$-sum process:

$$\nu_n = \Big(\sum_{k=1}^n \nu_{nk}(f,\psi), f \in \mathcal{F}, \psi \in \Psi\Big),$$

where $\nu_{nk}(f,\psi) = \langle X_k, \psi\rangle \lambda_{nk}(f)$, λ_{nk} is a uniform probability on the interval $[(k-1)/n, k/n]$ and $\lambda_{nk}(f) = \int_0^1 f(t)\,\mathrm{d}\lambda_{nk}(t)$. A natural framework for stochastic process ν_n is the space $\ell^\infty(\mathcal{G})$, where $\mathcal{G} = \mathcal{F} \times \Psi$. Recall for a class $\mathcal{G}$ that $\ell^\infty(\mathcal{G})$ is a Banach space of all uniformly bounded real-valued functions

μ on $\mathcal{G}$ endowed with the uniform norm:

$$\|\mu\|_{\mathcal{G}} := \sup\{|\mu(g)| : g \in \mathcal{G}\}.$$

Given a pseudometric d on $\mathcal{G}$, $UC(\mathcal{G}, d)$ is a set of all $\mu \in \ell^\infty(\mathcal{G})$, which are uniformly d-continuous. The set $UC(\mathcal{G}, d)$ is a separable subspace of $\ell^\infty(\mathcal{G})$ if and only if $(\mathcal{G}, d)$ is totally bounded. The pseudometric space $(\mathcal{G}, d)$ is totally bounded if $N(\varepsilon, \mathcal{G}, d)$ is finite for every $\varepsilon > 0$, where $N(\varepsilon, \mathcal{G}, d)$ is the minimal number of open balls of d-radius ε, which are necessary to cover $\mathcal{G}$.

It is worth noting that the process ν_n is continuous when $\mathcal{F} \times \Psi$ is endowed with the metric ρ_2. Indeed,

$$\begin{aligned}
|\nu_{nk}(f, \psi) - \nu_{nk}(f', \psi')| &\le |\langle Y_k, \psi\rangle \lambda_{nk}(f) - \langle Y_k, \psi'\rangle \lambda_{nk}(f')| \\
&\le |\lambda_{nk}(f)|\langle Y_k, \psi - \psi'\rangle + \langle Y_k, \psi'\rangle \lambda_{nk}(f - f')| \\
&\le \|Y_k\|[\sqrt{n}\|f\| \cdot \|\psi - \psi'\| + \|\psi'\| \cdot \|f - f'\|] \\
&\le \sqrt{2}\|Y_k\| \max\{\sqrt{n}\|f\|, \|\psi\|\}\rho_2((f, \psi), (f', \psi')),
\end{aligned}$$

since $|\lambda_{nk}(f)| \le \sqrt{n}\|f\|$ for every $f \in L_2$. If both sets $\mathcal{F}$ and Ψ are totally bounded, then the process ν_n is uniformly continuous so that ν_n takes values in the subspace $UC(\mathcal{G})$.

Next, we specify the set $\mathcal{F} \subset L_2$. To this aim, we recall some definitions. For a function $f : [0, 1] \to \mathbb{R}$, a positive number $0 < p < \infty$, and an integer $m \in \mathbb{N}$, the (m, p)-variation of f on the interval $[0, t]$ is

$$v_{m,p}(f; [0, t]) := \sup\Big\{ \sum_{j=1}^{m} |f(t_j) - f(t_{j-1})|^p \Big\},$$

where the supremum is taken over all partitions $0 = t_0 < t_1 < \cdots < t_m = t$, of the interval $[0, t]$. We abbreviate $v_{m,p}(f) := v_{m,p}(f; [0, 1])$. If $v_p(f) := \sup_{m \ge 1} v_{m,p}(f) < \infty$, then we say that f has finite p-variation and $\mathcal{W}_p[0, 1]$ is the set of all such functions. The set $\mathcal{W}_p[0, 1]$, $p \ge 1$, is a (non-separable) Banach space with the norm:

$$||f||_{[p]} := \sup_{0 \le t \le 1} |f(t)| + v_p^{1/p}(f).$$

The embedding $\mathcal{W}_p[0, 1] \hookrightarrow \mathcal{W}_q[0, 1]$ is continuous and

$$v_q^{1/q}(f) \le v_p^{1/p}(f), \quad \text{for } 1 \le p < q.$$

For more information on the space $\mathcal{W}_p[0, 1]$, we refer to [23].

The limiting zero mean Gaussian process $\nu_\gamma = (\nu(f, \psi), f \in \mathcal{F}, \psi \in \Psi)$ is

defined via covariance:

$$E\nu_\gamma(f,\psi)\nu_\gamma(f',\psi') = \mathcal{K}_\gamma((f,\psi),(f',\psi,)) := \langle \Gamma\psi,\psi'\rangle\langle f,f'\rangle, \quad \psi,\psi',f,f' \in L_2, \tag{3.2}$$

where $\Gamma : L_2 \to L_2$ is the covariance operator corresponding to the kernel γ. The function $\mathcal{K}_\gamma : \mathcal{G}\times\mathcal{G} \to \mathbb{R}$ is positive definite:

$$\sum_{k,j=1}^{m} c_j c_k \mathcal{K}_\gamma((f_j,\psi_j),(f_k,\psi_k)) \geq 0, \tag{3.3}$$

for all $c_1,\dots,c_m \in \mathbb{R}$, $(f_1,\psi_1),\dots,(f_m,\psi_m) \in \mathcal{G}$, and $m \geq 1$. Indeed, if we denote by $\mathcal{W} = (\mathcal{W}(f), f \in L_2)$ the isonormal Gaussian process on the Hilbert space L_2, we see that

$$\mathcal{K}_\gamma((f_j,\psi_j),(f_k,\psi_k)) = E\langle Y,\psi_j\rangle\langle Y,\psi_k\rangle E\mathcal{W}(f_j)\mathcal{W}(f_k);$$

hence,

$$\sum_{k,j=1}^{m} c_j c_k \mathcal{K}_\gamma((f_j,\psi_j),(f_k,\psi_k)) = E\Big(\sum_{k=1}^{m} c_k\langle Y,\psi_k\rangle\mathcal{W}(f_k)\Big)^2$$

and (3.3) follows. This justifies the existence of the process ν_γ.

Throughout, we shall exploit the following.

Assumption 3. Random processes $Y, Y_1, Y_2, \dots$ are i.i.d. mean square continuous, jointly measurable, with mean zero and covariance γ such that $\int_0^1 \gamma(t,t)\,\mathrm{d}t < \infty$. For the model (3.1), we consider null hypothesis $H_0 : \ g = 0$ and two possible alternatives:

$$H_A : \ g = g_n = u_n q_n, \ \text{ where } \ u_n \to u \ \text{ in } \ \mathcal{W}_2[0,1], \ \ \sqrt{n}q_n \to q \ \text{ in } \ L_2,$$

and

$$H'_A : \ g = g_n = u_n q_n, \ \text{ where } \ u_n \to u \ \text{ in } \ \mathcal{W}_2[0,1], \ \ \sqrt{n}\sup_{\psi\in\Psi}|\langle q_n,\psi\rangle| \to \infty.$$

In both alternatives, the function u_n is responsible for the configuration of a drift within the sample, whereas the function q_n estimates a magnitude of the drift.

Our main theoretical results are Theorems 10 and 12, which are proven in Section 3.6.

Theorem 10. Let the random processes (X_k) be defined by (3.1), where $Y, Y_1, Y_2, \dots$ satisfy Assumption 3. Assume that, for some $1 \leq q < 2$, the

set $\mathcal{F} \subset \mathcal{W}_q[0,1]$ is bounded and the set $\Psi \subset L_2$ satisfies

$$\int_0^1 \sqrt{\log N(\varepsilon, \Psi, \rho)}\, \mathrm{d}\varepsilon < \infty. \tag{3.4}$$

Then, there exists a version of a Gaussian process ν_γ on $L_2 \times L_2$ such that its restriction on $\mathcal{F} \times \Psi$, $(\nu_\gamma(f,\psi), f \in \mathcal{F}, \psi \in \Psi)$ is continuous and the following hold:

(1a) Under H_0:

$$n^{-1/2}\nu_n \xrightarrow[n\to\infty]{\mathcal{D}} \nu_\gamma \quad \text{in} \quad \ell^\infty(\mathcal{F} \times \Psi). \tag{3.5}$$

(1b) Under H_A,

$$n^{-1/2}\nu_n \xrightarrow[n\to\infty]{\mathcal{D}} \nu_\gamma + \Delta, \quad \text{in} \quad \ell^\infty(\mathcal{F} \times \Psi), \tag{3.6}$$

where

$$\Delta(f,\psi) = \langle u, f\rangle\langle q, \psi\rangle.$$

If $u(s) = 1, s \in [0,1]$, then the alternative H_A corresponds to the presence of a signal in a noise. In this case, $\Delta(f,\psi) = \lambda(f)\langle q,\psi\rangle$. Therefore, the use of this theorem for testing a signal in a noise is meaningful provided $\langle q,\psi\rangle \neq 0$.

As a corollary, Theorem 10 combined with the continuous mapping theorem gives the following result.

Theorem 11. Assume that conditions of Theorem 10 are satisfied. Then, the following hold:

(2a) Under H_0

$$\sup_{\psi\in\Psi, f\in\mathcal{F}} |n^{-1/2}\nu_n(f,\psi)| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{\psi\in\Psi, f\in\mathcal{F}} |\nu_\gamma(f,\psi)|.$$

(2b) Under H_A,

$$\sup_{\psi\in\Psi, f\in\mathcal{F}} |n^{-1/2}\nu_n(f,\psi)| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{\psi\in\Psi, f\in\mathcal{F}} |\nu_\gamma(f,\psi) + \langle u, f\rangle\langle q,\psi\rangle|.$$

(2c) Under H_A',

$$\sup_{\psi\in\Psi, f\in\mathcal{F}} |n^{-1/2}\nu_n(f,\psi)| \xrightarrow[n\to\infty]{\mathrm{P}} \infty. \tag{3.7}$$

Proof. Since both (2a) and (2b) are by-products of Theorem 10 and continuous

mappings, we need to prove only (2c). First, we observe that

$$\sup_{\psi\in\Psi,f\in\mathcal{F}}|\nu_n(f,\psi)| \geq \sup_{\psi\in\Psi,f\in\mathcal{F}}\Big|\sum_{k=1}^{n}\Big(\langle Y_k,\psi\rangle+\langle q_n,\psi\rangle u_n(k/n)\Big)\lambda_{nk}(f)\Big|$$
$$\geq \sup_{\psi\in\Psi,f\in\mathcal{F}}\Big|\sum_{k=1}^{n}u_n(k/n)\lambda_{nk}(f)\Big|\cdot|\langle q_n,\psi\rangle| - O_P(\sqrt{n}),$$

by (2a). Consider

$$I_n(f) := \Big|\sum_{k=1}^{n}u_n(k/n)\lambda_{nk}(f)\Big|.$$

We have

$$I_n(f) = n\Big|\sum_{k=1}^{n}u_n(k/n)\int_{(k-1)/n}^{k/n}f(t)\,\mathrm{d}t\Big|$$
$$\geq n\Big|\sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}u_n(t)f(t)\,\mathrm{d}t\Big| - n\Big|\sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}(u_n(t)-u_n(k/n))f(t)\,\mathrm{d}t\Big|$$
$$:= I_n'(f) - I_n''(f).$$

By the Hölder inequality,

$$I_n''(f) \leq n\sum_{k=1}^{n}\Big(\int_{(k-1)/n}^{k/n}(u_n(t)-u_n(k/n))^2\,\mathrm{d}t\Big)^{1/2}\Big(\int_{(k-1)/n}^{k/n}f^2(t)\,\mathrm{d}t\Big)^{1/2}$$
$$\leq n\Big(\sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}(u_n(t)-u_n(k/n))^2\,\mathrm{d}t\Big)^{1/2}\Big(\sum_{k=1}^{n}\int_{(k-1)/n}^{k/n}f^2(t)\,\mathrm{d}t\Big)^{1/2}$$
$$\leq n\Big(n^{-1}\sum_{k=1}^{n}v_2(u_n,[(k-1)/n,k/n])\Big)^{1/2}\|f\| \leq \sqrt{n}v_2^{1/2}(u_n)\|f\|.$$

Since $I_n'(f) = n|\langle u_n,f\rangle|$, we deduce

$$I_n(\psi,f) \geq n|\langle u_n,f\rangle| - \sqrt{n}v_2^{1/2}(u_n)\|f\|.$$

Hence,

$$n^{-1/2}\sup_{\psi\in\Psi,f\in\mathcal{F}}|\nu_n(f,\psi)| \geq \sqrt{n}\sup_{\psi\in\Psi,f\in\mathcal{F}}|\langle u_n,f\rangle|\cdot|\langle q_n,\psi\rangle| - O_P(1)$$

and this completes the proof of (2c). □

Next, we consider $\mathcal{G}$-sum process $\mu_n = (\mu_n(f,\psi), f\in\mathcal{F}, \psi\in\Psi)$ defined by

$$\mu_n(f,\psi) = \sum_{k=1}^{n}\langle X_k - \overline{X}_n,\psi\rangle\lambda_{nk}(f),$$

where $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$. Its limiting zero mean Gaussian process μ_γ is defined via covariance:

$$E\mu_\gamma(f,\psi)\mu_\gamma(f',\psi') = \langle \Gamma\psi, \psi'\rangle[\langle f, f'\rangle - \lambda(f)\lambda(f')], \quad \psi, \psi', f, f' \in L_2. \quad (3.8)$$

The existence of Gaussian process μ_γ can be justified as that of ν_γ above. Just notice that

$$\langle f, f'\rangle - \lambda(f)\lambda(f') - E(\mathcal{W}(f) - \lambda(f)\mathcal{W}(1))(\mathcal{W}(f') - \lambda(f')\mathcal{W}(1)),$$

where $1(t) = 1, t \in [0,1]$.

Theorem 12. Assume that the conditions of Theorem 10 are satisfied. Then, there exists a version of the Gaussian process μ_γ on $L_2(0,1) \times L_2(0,1)$ such that its restriction on $\mathcal{F} \times \Psi$, $(\nu(f,\psi), f \in \mathcal{F}, \psi \in \Psi)$ is continuous and the following hold:

(3a) Under H_0,

$$n^{-1/2}\mu_n \xrightarrow[n\to\infty]{\mathcal{D}} \mu_\gamma \ \text{ in } \ \ell^\infty(\mathcal{F} \times \Psi); \quad (3.9)$$

(3b) Under alternative H_A,

$$n^{-1/2}\mu_n \xrightarrow[n\to\infty]{\mathcal{D}} \mu_\gamma + \widetilde{\Delta} \text{ in } \ \ell^\infty(\mathcal{F} \times \Psi), \quad (3.10)$$

where

$$\widetilde{\Delta}(f,\psi) = [\langle u, f\rangle - \lambda(u)\lambda(f)]\langle q, \psi\rangle.$$

We see that the limit distribution of the $\mathcal{G}$-sum process separates the null and alternative hypothesis provided $[\langle u, f\rangle - \lambda(u)\lambda(f)]\langle q, \psi\rangle \neq 0$. As a corollary, Theorem 12 combined with the continuous mapping theorem gives the following results.

Theorem 13. Assume that the conditions of Theorem 10 are satisfied. Then, the following hold:

(4a) Under H_0,

$$\sup_{\psi\in\Psi, f\in\mathcal{F}} |n^{-1/2}\mu_n(f,\psi)| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{\psi\in\Psi, f\in\mathcal{F}} |\mu_\gamma(f,\psi)|. \quad (3.11)$$

(4b) Under H_A,

$$\sup_{\psi\in\Psi, f\in\mathcal{F}} |n^{-1/2}\mu_n(f,\psi)| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{\psi\in\Psi, f\in\mathcal{F}} |\mu_\gamma(f,\psi) + \widetilde{\Delta}(f,\psi)|. \quad (3.12)$$

(4c) Under H_A',

$$\sup_{\psi\in\Psi, f\in\mathcal{F}} |n^{-1/2}\mu_n(f,\psi)| \xrightarrow[n\to\infty]{\mathrm{P}} \infty. \quad (3.13)$$

Proof. Both (4a) and (4b) are by-products of Theorem 12 and continuous mappings, whereas the proof of (4c) follows the lines of the proof of Theorem 11 (2c). □

3.3. Test Statistics

Several useful test statistics can be obtained from the $\mathcal{G}$-sum process $\mu_n = (\mu_n(f,\psi), (f,\psi) \in \mathcal{G} = \mathcal{F} \times \Psi)$, by considering concrete examples of sets Ψ and $\mathcal{F}$.

Throughout this section, we assume that the sample $X_1, X_2, \ldots, X_n$ follows the model (3.1) and $Y, Y_1, Y_2, \ldots$ satisfies Assumption 3.

By Γ, we denote the covariance operator of Y: $\Gamma = E(Y \otimes Y)$. Recall

$$\Gamma x(t) = \int_0^1 \gamma(t,s)x(s)\,\mathrm{d}s, \quad t \in [0,1].$$

According to Mercer's theorem, the covariance γ has then the following singular-value decomposition:

$$\gamma(s,t) = \sum_{r=1}^{m} \lambda_r \psi_r(s)\psi_r(t), \quad t,s \in [0,1], \tag{3.14}$$

where $\lambda_1, \ldots, \lambda_m$ are all the decreasingly ordered positive eigenvalues of Γ and $\psi_1, \ldots, \psi_m$ are the associated eigenfunctions of Γ such that

$$\int_0^1 \psi_r^2(t)\,\mathrm{d}t = 1, \quad \int_0^1 \psi_r(t)\psi_\ell(t)\,\mathrm{d}t = 0, \quad r \neq \ell,$$

and m is the smallest integer such that, when $r > m$, $\lambda_r = 0$. If $m = \infty$, then all the eigenvalues are positive, and in this case, $\sum_r \lambda_r < \infty$. Note that $\lambda_r = E\langle Y, \psi_r\rangle^2$. Besides, we shall assume the following.

Assumption 4. The eigenvalues λ_r satisfy, for some $d > 0$,

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1}.$$

In statistical analysis, the eigenvalues and eigenfunctions of Γ are replaced by their estimated versions. Noting that, for each k,

$$E[(X_k - E(X_k)) \otimes (X_k - E(X_k))] = \Gamma,$$

one estimates Γ by

$$\widehat{\Gamma}_n := \frac{1}{n}\sum_{i=1}^{n}[(X_i - \overline{X}_n) \otimes (X_i - \overline{X}_n)],$$

where $\overline{X}_n(s) = n^{-1}(X_1(s) + \cdots + X_n(s)$. We denote the eigenvalues and eigenfunctions of $\widehat{\Gamma}$ by $\widehat{\lambda}_{nr}$ and $\widehat{\psi}_{nr}$, $r = 1, \ldots, n-1$, respectively. In order to ensure that $\widehat{\psi}_{nr}$ may be viewed as an estimator of ψ_r rather than of $-\psi_r$, we will in the following assume that the signs are such that $\langle \widehat{\psi}_{nr}, \psi_r \rangle \geq 0$. Note that

$$\widehat{\Gamma}\widehat{\psi}_{nr} = \widehat{\lambda}_{nr}\widehat{\psi}_{nr}, \quad r = 1, \ldots, n-1, \tag{3.15}$$

and

$$\widehat{\lambda}_{nr} = \frac{1}{n-1}\sum_{i=1}^{n}\langle X_i - \overline{X}_n, \widehat{\psi}_{nr}\rangle^2, \quad r = 1, \ldots, n. \tag{3.16}$$

The use of the estimated eigenfunctions and eigenvalues in the test statistics is justified by the following result. For a Hilbert–Schmidt operator T on L_2, we denote by $\|T\|_{HS}$ its Hilbert–Schmidt norm.

Lemma 2. Assume that Assumption 3 holds. Then, under H_A,

$$\|\widehat{\Gamma}_n - \Gamma\|_{HS} \to 0 \ \text{ as } \ n \to \infty.$$

Proof. First, we observe that

$$\widehat{\Gamma}_n = \widetilde{\Gamma}_n + T_{n1} + T_{n2} + T_{n3},$$

where

$$\begin{aligned}
\widetilde{\Gamma}_n &= \frac{1}{n}\sum_{k=1}^{n}(Y_k - \overline{Y}_n) \otimes (Y_k - \overline{Y}_n), \\
T_{n1} &= \frac{1}{n}\sum_{k=1}^{n}\Big[u_n(k/n) - \frac{1}{n}\sum_{j=1}^{n}u_n(j/n)\Big](Y_k - \overline{Y}_n) \otimes q_n, \\
T_{n2} &= \frac{1}{n}\sum_{k=1}^{n}\Big[u_n(k/n) - \frac{1}{n}\sum_{j=1}^{n}u_n(j/n)\Big]q_n \otimes (Y_k - \overline{Y}_n), \\
T_{n3} &= \frac{1}{n}\sum_{k=1}^{n}\Big[u_n(k/n) - \frac{1}{n}\sum_{j=1}^{n}u_n(j/n)\Big]^2 q_n \otimes q_n.
\end{aligned}$$

It is well known that $\|\widetilde{\Gamma}_n - \Gamma\|_{HS} \xrightarrow[n\to\infty]{\text{a.s.}} 0$ as $n \to \infty$. By the moment inequality for sums of independent random variables, we deduce

$$E\|T_{ni}\|_{HS}^2 \leq cn^{-2}\sum_{k=1}^{n}\Big[u_n(k/n) - \frac{1}{n}\sum_{j=1}^{n}u_n(j/n)\Big]^2 E\|Y\|^2\|q_n\|^2,$$

for both $i = 1, 2$. This yields $T_{ni} \xrightarrow[n\to\infty]{\mathrm{P}} 0$. Next, we have

$$\|T_{n3}\|_{HS} = \frac{1}{n}\sum_{k=1}^{n}\Big[u_n(k/n) - \frac{1}{n}\sum_{j=1}^{n} u_n(j/n)\Big]^2 \|q_n\|^2 \le \frac{1}{n}\sum_{k=1}^{n} u_n^2(k/n)\|q_n\|^2 \to 0$$

as $n \to \infty$ due to assumption H_A. This completes the proof. □

Lemma 3. Assume that Assumptions 3 and 4 for some finite d hold and $E(\|Y\|^4) < \infty$. Then, under H_0, as well as under H_A:

$$n^{1/2}|\widehat{\lambda}_{nj} - \lambda_j| = O_P(1), \quad \text{and} \quad n^{1/2}\|\widehat{c}_{nj}\widehat{\psi}_{nj} - \psi_j\| = O_P(1)$$

for each $1 \le j \le d$, where $\widehat{c}_{nj} = \langle\widehat{\psi}_{nj}, \psi_j\rangle$.

Proof. If the null hypothesis is satisfied, then $\widehat{\Gamma}_n = \widetilde{\Gamma}_n$ and the asymptotic results for the eigenvalues and eigenfunctions of $\widetilde{T}_n$ are well known (see, e.g., [34]). Under alternative H_A, the results follow from Lemma 2 and Lemma 2.2 and Lemma 2.3 in [34]. □

Next, we consider separately the test statistics for at most one, at most m, and for an unknown number of change-points.

3.3.1. Testing at Most One Change-Point

Define for $d > 0$,

$$T_{n,1}(d) := \max_{1\le j\le d}\frac{1}{\sqrt{\lambda_j}}\max_{1\le k\le n}\Big|\sum_{i=1}^{k}\langle X_i - \overline{X}_n, \psi_j\rangle\Big|. \tag{3.17}$$

This statistic is designed for at most one change-point alternative. Its limiting distribution is established in the following theorem.

Theorem 14. Let random functional sample (X_k) be defined by (3.1) where $Y, Y_1, Y_2, \ldots$ satisfies Assumptions 3 and 4. Then,

(a) Under H_0, it holds that

$$n^{-1/2}T_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)|,$$

where $B_1, \ldots, B_d$ are independent standard Brownian bridge processes;

(b) Under H_A, it holds that

$$n^{-1/2}T_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t) + \Delta(t)\langle q, \psi_k/\sqrt{\lambda_k}\rangle|,$$

where

$$\Delta(t) = \int_0^t u(s)\,\mathrm{d}s - t\int_0^1 u(s)\,\mathrm{d}s, \quad t \in [0,1]. \tag{3.18}$$

(c) Under H'_A, it holds that

$$n^{-1/2}T_{n,1}(d) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Proof. Consider the sets

$$\Psi_{d,\gamma} := \Big\{\frac{\psi_1}{\sqrt{\lambda_1}}, \ldots, \frac{\psi_d}{\sqrt{\lambda_d}}\Big\}, \text{ and } \mathcal{F}_1 = \{\mathbf{1}_{[0,t]}, t \in [0,1]\}. \tag{3.19}$$

Observing that

$$T_{n,1}(d) = \sup_{\psi\in\Psi_{d,\gamma}, f\in\mathcal{F}_1} |\mu_n(f,\psi)|$$

and $\mathcal{F}_1$ is a bounded set in $\mathcal{W}_q$, we complete the proof by applying Theorem 12. □

Based on this result, we construct the testing procedure in a classical way. Choose for a given $\alpha \in (0,1)$, $C_\alpha > 0$ such that

$$P(\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)| > C_\alpha) = \alpha.$$

According to Theorem 14, the test:

$$T_{n,1}(d) \ge \sqrt{n}C_\alpha \tag{3.20}$$

will have asymptotic level α. Under the alternative H_A, we have

$$\begin{aligned}\lim_{n\to\infty} P(n^{-1/2}T_{n,1}(d) \ge C_\alpha) &\ge P\Big(\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)| \le \max_{1\le k\le d}\max_{0\le t\le 1}|\Delta(t)\langle q, \psi_k/\lambda_k\rangle| - C_\alpha\Big)\\ &\ge 1-\alpha,\end{aligned}$$

when

$$\max_{1\le k\le d}\max_{0\le t\le 1}|\Delta(t)\langle q, \psi_k/\lambda_k\rangle| \ge 2C_\alpha. \tag{3.21}$$

Hence, if $g(s,t) = g_n(s,t) = u_n(s)q_n(t)$ and

$$\sqrt{n}\max_{1\le k\le d}\max_{0\le t\le 1}|\langle g_n(t,\cdot), \psi_k/\sqrt{\lambda_k}\rangle| \to \infty$$

as $n\to\infty$, then the test (3.20) is asymptotically consistent.

Let us note that, due to the independence of Brownian bridges $B_k, k = 1,\ldots,d$, we have

$$1-\alpha = P(\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)| \le C_\alpha) = P^d(\sup_{0\le t\le 1}|B_1(t)| \le C_\alpha).$$

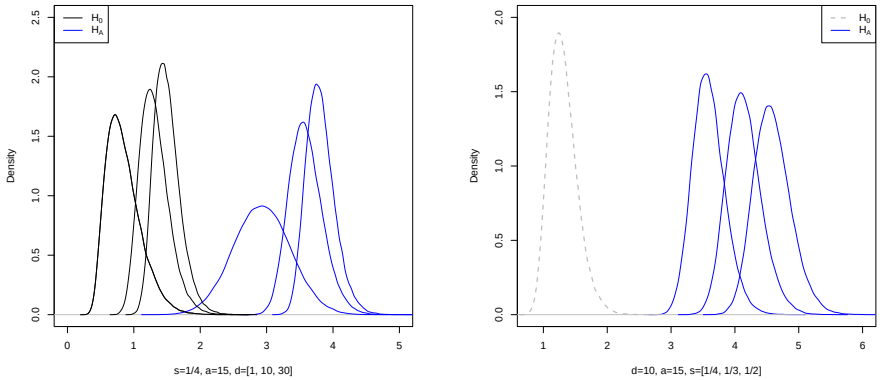


Figure 3.1: Density functions.

This yields

$$P(\sup_{0\le t\le 1}|B_1(t)|\le C_\alpha)=(1-\alpha)^{1/d}.$$

Hence, C_α is the $(1-\alpha)^{1/d}$-quantile of the distribution of $\sup_{0\le t\le 1}|B_1(t)|$. This observation simplifies the calculations of critical values C_α.

In particular, if there is $s^*\in(0,1)$ such that $u(s)=\mathbf{1}_{[0,s^*]}(s), s\in[0,1]$, then we have one change-point model:

$$X_k(t)=\mathbf{1}_{[0,s^*]}(k/n)q_n(t)+Y_k(t),\quad t\in[0,1].$$

In this case, $\Delta(t)=\Delta^*(t):=\min\{t,s^*\}-ts^*,\quad t\in[0,1]$.

Figure 3.1 below shows generated density functions of $\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)|$ and $\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)+\Delta^*(t)\langle q,\psi_k/\sqrt{\lambda_k}\rangle|$ for $d=[1,10,30]$, $s^*\in\{1/4,1/2,3/4\}$ where $q=a\psi_k\sqrt{\lambda_k}$ for a fixed k.

Let us observe that test statistic $T_{n,1}(d)$ tends to infinity when $d\to\infty$. On the other hand, with larger d, the approximation of X_j by series $\sum_{j=1}^d\langle X,\psi_j\rangle\psi_j$ is better and leads to better testing power. The following result establishes the asymptotic distribution of $T_{n,1}(d)$ as $d\to\infty$.

Theorem 15. Let random functional sample (X_k) be defined by (3.1) where $Y,Y_1,Y_2,\dots$ satisfies Assumption 3. Then, under H_0,

$$\lim_{d\to\infty}\lim_{n\to\infty}P\Big(n^{-1/2}T_{n,1}(d)\le\frac{x}{a_d}+b_d\Big)=\exp\{-e^{-x}\},\quad x\ge 0,\tag{3.22}$$

where

$$a_d=(8\ln d)^{1/2},\quad b_d=\frac{1}{4}a_d+\frac{\ln\ln d}{a_d}.\tag{3.23}$$

Proof. By Theorem 14, the proof reduces to

$$\lim_{d\to\infty} P(\sup_{1\le j\le d} \|B_j\|_\infty \le x/a_d + b_d) = \exp\{-e^{-x}\}, \quad x \ge 0. \tag{3.24}$$

It is known that

$$P(\|B_j\|_\infty > u) = 2e^{-2u^2}u^2(1+o(1)), \quad u \to \infty.$$

Since Brownian bridges $B_j, 1 \le j \le d$ are independent, we have

$$\begin{aligned} P(\sup_{1\le j\le d} \|B_j\|_\infty \le x/a_d + b_d) &= P^d(\|B_1\|_0 \le x/a_d + b_d) \\ &= \Big(1 - P(\|B_1\|_0 \ge x/a_d + b_d)\Big)^d \end{aligned}$$

and

$$\lim_{d\to\infty} dP(\|B_1\|_\infty \ge x/a_d + b_d) = e^{-x}.$$

This proves (3.24). □

When d is large, the test (3.20) becomes

$$T_{n,1} \ge \sqrt{n}\Big[\frac{1}{a_d}\ln\Big(\frac{1}{\ln(1/\alpha)}\Big) + b_d\Big] \tag{3.25}$$

and has asymptotic level α as n and d tend to infinity.

The dependence on d of critical values of the tests (3.20) and (3.25) is shown in Figure 3.2. A comparison was made for asymptotic level $\alpha = 0.05$. From Figure 3.2, we see that the critical values in (3.25) are smaller than those in (3.20). This means that the error of the first kind is more likely with the test (3.25), rather than with (3.36). This is confirmed by simulations.

If the eigenfunctions (ψ_k) are unknown, we use the statistics:

$$\widehat{T}_{n,1}(d) := \max_{1\le j\le d} \frac{1}{\sqrt{\widehat{\lambda}_j}} \max_{1\le k\le n} \Big|\sum_{i=1}^{k} \langle X_i - \overline{X}_n, \widehat{\psi}_j\rangle\Big|. \tag{3.26}$$

Theorem 16. Let random functional sample (X_k) be defined by (3.1), where $Y, Y_1, Y_2, \ldots$ satisfies Assumptions 3 and 4. Then:

(a) Under H_0,

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d} \sup_{0\le t\le 1} |B_k(t)|,$$

where $B_1, \ldots, B_d$ are independent standard Brownian bridge processes;

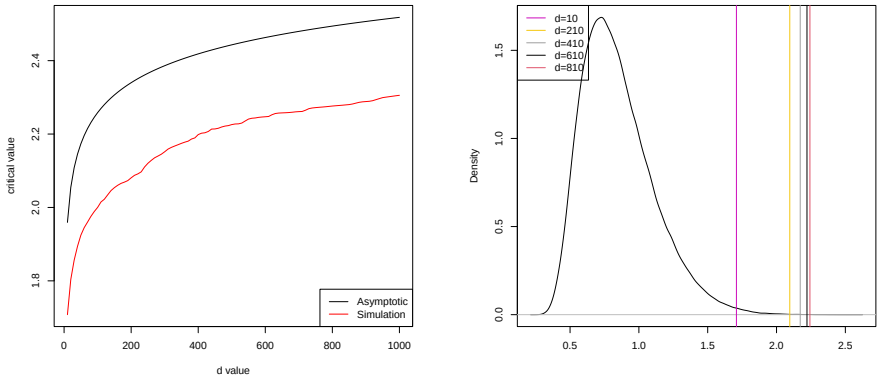


Figure 3.2: A comparison of the critical values in equations (3.20) and (3.25) is presented, with a significance level of $\alpha = 0.05$, along with the density function of $T_{n,1}(d)$.

(b) Under H_A, if $E\|Y\|^4 < \infty$, it holds that

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t) + \Delta(t)\langle q, \psi_k/\sqrt{\lambda_k}\rangle|,$$

where $\Delta(t) = \int_0^t u(s)\,\mathrm{d}s - t\int_0^1 u(s)\,\mathrm{d}s, \quad t \in [0,1]$.

(c) Under H_A', if $E\|Y\|^4 < \infty$, it holds that

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Proof. The result follows from Theorem 14 if we show that

$$D_n := n^{-1/2}|T_{n,1}(d) - \widehat{T}_{n,1}(d)| \xrightarrow[n\to\infty]{\mathrm{P}} 0. \tag{3.27}$$

On the set $\max_{1\le j\le d}|\lambda_j - \widehat{\lambda}_{nj}| + \max_{1\le j\le d}\|\psi_j - \widehat{c}_j\widehat{\psi}_{nj}\| \le An^{-1/2}$ and for $n \ge N_0$ such that $An^{-1/2} < \lambda_d/2$, simple algebra gives $D_n \le D_{n1} + D_{n2}$, where

$$\begin{aligned} D_{n1} &= \max_{1\le j\le d}\Big|\frac{1}{\lambda_j} - \frac{1}{\widehat{\lambda}_{nj}}\Big| \max_{1\le k\le 1}\Big|\sum_{i=1}^{k}\langle X_i - \overline{X}_n, \psi_j\rangle\Big| \\ &\le \frac{2}{\lambda_d}\max_{1\le j\le n}|\widehat{\lambda}_{nj} - \lambda_j| n^{-1/2}T_{n,1}(d) \to 0 \quad \text{as} \quad n\to\infty, \end{aligned}$$

and

$$D_{n2} \le n^{-1/2}\frac{2}{\varepsilon}\max_{1\le k\le n}\Big\|\sum_{i=1}^{k}[X_i-\overline{X}_n]\Big\|\max_{1\le j\le d}\|\widehat{\psi}_{nj}-\psi_j\|$$
$$\le \frac{2A}{\varepsilon}n^{-1}\max_{1\le k\le n}\Big\|\sum_{i=1}^{k}[X_i-\overline{X}_n]\Big\| \to 0 \ \text{ as } \ n\to\infty$$

by the law of large numbers. Lemma 3 concludes the proof. □

Test (3.20) now becomes

$$\widehat{T}_{n,1}(d) \ge \sqrt{n}C_\alpha \tag{3.28}$$

and has asymptotic level α by Theorem 16.

3.3.2. Testing at most m Change-Points

For $m > 1$, let $\mathcal{N}_m$ be a set of all partitions $\kappa = (k_i, i = 0, 1, \ldots, m)$ of the set $\{0, 1, \ldots, n\}$ such that $0 = k_0 < k_1 < \cdots < k_{m-1} < k_m = n$. Next, consider for fixed integers d, $1 \le m < n$ and real $p > 2$,

$$T_{n,m}(d,p) := \max_{1\le j\le d}\frac{1}{\sqrt{\lambda_j}}\max_{\kappa\in\mathcal{N}_m}\Big\{\sum_{i=1}^{m}\Big|\sum_{k=k_{i-1}+1}^{k_i}\langle X_k-\overline{X}_n,\psi_j\rangle\Big|^p\Big\}^{1/p}. \tag{3.29}$$

The statistics $T_{n,m}(d,p)$ are designed for testing at most m change-points in a sample.

Theorem 17. Let the random sample $(X_i, i = 1, \ldots, n)$ be as in Theorem 10. Then:

(a) Under H_0,

$$n^{-1/2}T_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j),$$

where $B_1, \ldots, B_d$ are independent standard Brownian bridges.

(b) Under H_A,

$$n^{-1/2}T_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j + \Delta\langle q, \psi_j/\sqrt{\lambda_j}\rangle),$$

where $\Delta(t), t \in [0,1]$ is as defined in Theorem 11.

(c) Under H'_A,

$$n^{-1/2}T_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

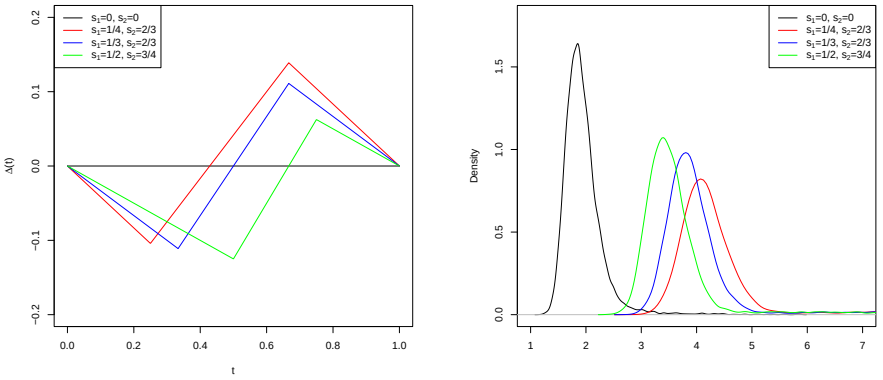


Figure 3.3: Functions Δ_2^* and density functions.

Proof. For $1 \le m \le n$ and $q = p/(p-1)$, set

$$\mathcal{F}_{m,q} := \Big\{ \sum_{j=1}^{m} b_j \mathbf{1}_{(t_{j-1},t_j]} : \quad \sum_{j=1}^{m} |b_j|^q \le 1, 0 = t_0 < t_1 < \cdots < t_m = 1 \Big\}. \quad (3.30)$$

It is easy to check that $\mathcal{F}_{m,q} \subset \mathcal{W}_q[0,1]$. Since

$$\sup \Big\{ \Big| \sum_{k=1}^{m} a_k b_k \Big| : \quad \sum_{k=1}^{m} |b_k|^q \le 1 \Big\} = \Big(\sum_{k=1}^{m} |a_k|^p \Big)^{1/p},$$

we have

$$T_{n,m}(d) = \max_{\psi \in \Psi_{d,\gamma}} \max_{f \in \mathcal{F}_{m,q}} |\mu_n(\psi, f)|,$$

and the results follow from Theorem 11. □

In particular, if there is $s_1^*, s_2^* \in (0,1)$ such that $u(s) = \mathbf{1}_{[s_1^*, s_2^*]}(s), s \in [0,1]$, then (3.1) corresponds to the so-called changed segment model. In this case, we have $\Delta(t) = \Delta_2^*(t) := \max\{0, \min\{t, s_2^*\} - s_1^*\} - t(s_2^* - s_1^*), \quad t \in [0,1]$. Figure 3.3) shows the generated density functions of $\max_{1 \le k \le d} v_{4,p}^{1/p}(B_k)$ and $\max_{1 \le k \le d} v_{4,p}^{1/p}(B_k + a_k \Delta_2^*)$ for different values of $d \ge 1$, $0 < s_1^* < s_2 < 1$, and $p > 2$. The numbers $a_1, \ldots, a_d$ were sampled from the uniform distribution on $[0, 15]$.

With the estimated eigenvalues and eigenfunctions, we define

$$\widehat{T}_{n,m}(d,p) := \max_{1 \le j \le d} \frac{1}{\sqrt{\widehat{\lambda}_{nj}}} \max_{\kappa \in \mathcal{N}_m} \Big\{ \sum_{i=1}^{m} \Big| \sum_{k=k_{i-1}+1}^{k_i} \langle X_k - \overline{X}_n, \widehat{\psi}_{nj} \rangle \Big|^p \Big\}^{1/p}. \quad (3.31)$$

Theorem 18. Let the functional sample $(X_k, k = 1, \ldots, n)$ be defined by (3.1)

where $Y, Y_1, Y_2, \ldots$ satisfies Assumptions 3 and 4. Then:

(a) Under H_0,

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j),$$

where $B_1, \ldots, B_d$ are independent standard Brownian bridges.

(b) Under H_A,

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j + \Delta\langle q, \psi_j/\sqrt{\lambda_j}\rangle),$$

where $\Delta(t), t \in [0,1]$ is as defined in Theorem 11.

(c) Under H_A',

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Proof. This goes along the lines of the proof of Theorem 16. □

According to Theorems 17 and 18, the tests:

$$T_{n,m}(d,p) \ge \sqrt{n}C_\alpha(m,d,p) \text{ and } \widehat{T}_{n,m}(d,p) \ge \sqrt{n}C_\alpha(m,d,p) \tag{3.32}$$

respectively, will have asymptotic level α, if $C_\alpha(m,d,p)$ is such that

$$P(v_{m,p}^{1/p}(B) \le C_\alpha(m,d,p)) = (1-\alpha)^{1/d}.$$

3.3.3. Testing Unknown Number of Change-Points

Next, consider for fixed integers d as above and real $p > 2$,

$$T_n(d,p) := \max_{1\le j\le d} \frac{1}{\sqrt{\lambda_j}} \max_{1\le m\le n} \max_{\kappa\in\mathcal{N}_m} \Big\{ \sum_{i=1}^{m} \Big| \sum_{k=k_{i-1}+1}^{k_i} \langle X_k - \overline{X}_n, \psi_j\rangle \Big|^p \Big\}^{1/p}. \tag{3.33}$$

The statistics $T_n(d,p)$ are designed for testing an unknown number of change-points in a sample.

Theorem 19. Let random sample $(X_i, i = 1, \ldots, n)$ be as in Theorem 10. Then:

(a) Under H_0,

$$n^{-1/2}T_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_p^{1/p}(B_j),$$

where $B_1, \ldots, B_d$ are independent standard Brownian bridges.

(b) Under H_A,

$$n^{-1/2}T_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_p^{1/p}(B_j + \Delta\langle q, \psi_j/\sqrt{\lambda_j}\rangle),$$

where $\Delta(t), t \in [0,1]$ is as defined in Theorem 10.

(c) Under H_A',

$$n^{-1/2} T_n(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Proof. For $q = p/(p-1)$, set

$$\mathcal{F}_q := \Big\{ \sum_{j=1}^{m} b_j \mathbf{1}_{(t_{j-1},t_j]} : \ \sum_{j=1}^{\infty} |b_j|^q \le 1, 0 = t_0 < t_1 < \cdots < t_m = 1, m \ge 1 \Big\}. \quad (3.34)$$

It is easy to check that $\mathcal{F}_q \subset \mathcal{W}_q[0,1]$. Since

$$\sup \Big\{ \Big| \sum_{k=1}^{\infty} a_k b_k \Big| : \ \sum_{k=1}^{\infty} |b_k|^q \le 1 \Big\} = \Big(\sum_{k=1}^{\infty} |a_k|^p \Big)^{1/p},$$

we have

$$T_n(d) = \max_{\psi \in \Psi_{d,\gamma}} \max_{f \in \mathcal{F}_{m,q}} |\mu_n(\psi, f)|,$$

and both statements (a) and (b) follow from Theorem 10. □

With the estimated eigenvalues and eigenfunctions, we define:

$$\widehat{T}_n(d,p) := \max_{1\le j\le d} \frac{1}{\sqrt{\widehat{\lambda}_{nj}}} \max_{1\le m\le n} \max_{\kappa \in \mathcal{N}_m} \Big\{ \sum_{i=1}^{m} \Big| \sum_{k=k_{i-1}+1}^{k_i} \langle X_k - \overline{X}_n, \widehat{\psi}_{nj} \rangle \Big|^p \Big\}^{1/p}. \quad (3.35)$$

Theorem 20. Let random sample (X_i) be as in Theorem 10. Then:

(a) Under H_0,

$$n^{-1/2} \widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_p^{1/p}(B_j),$$

where $B_1, \ldots, B_d$ are independent standard Brownian bridges.

(b) Under H_A,

$$n^{-1/2} \widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_p^{1/p}(B_j + \Delta \langle q, \psi_j / \sqrt{\lambda_j} \rangle),$$

where $\Delta(t), t \in [0,1]$ is as defined in Theorem 10.

(c) Under H_A',

$$n^{-1/2} \widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Proof. This goes along the lines of the proof of Theorem 16. □

According to Theorems 19 and 20, the tests:

$$T_n(d,p) \ge \sqrt{n} C_\alpha(d,p) \ \text{ and } \ \widehat{T}_n(d,p) \ge \sqrt{n} C_\alpha(d,p) \quad (3.36)$$

respectively, will have asymptotic level α, if $C_\alpha(d,p)$ is such that

$$P(v_p^{1/p}(B) \leq C_\alpha(d,p)) = (1-\alpha)^{1/d}.$$

The quantiles of distribution function of $v_p^{1/p}(B)$ were estimated in chapter 2.

3.4. Simulation Results

We examined the above-defined test statistics in a Monte Carlo simulation study. In the first subsection, we describe the simulated data under consideration. The statistical power analysis of the tests (3.36) and (3.32) is presented in Section 3.4.2.

3.4.1. Data

We used the following three scenarios:

(S1) Let (ξ_{jk}) be i.i.d. symmetrized Pareto random variables with index p (we used $p = 5$). Set

$$Y_j(t) = \sum_{k=1}^{d} \xi_{jk} \frac{\sqrt{2}\cos(k\pi t)}{k\sigma}, \quad t \in [0,1], \ j \geq 1, \tag{3.37}$$

where $\sigma^2 = E\xi_{11}^2$. Under the null hypothesis, we take $X_k = Y_k, k = 1, 2, \ldots, n$.

Under the alternative, we consider

$$X_j(t) = u_n(j/n) \sum_{k=1}^{d} a_{nk} \cos(k\pi t) + Y_j, \quad t \in [0,1], j = 1, \ldots, n,$$

where the function u_n defines the change-points' configuration and the coefficients (a_{nk}) are subject to choice.

(S2) We start with discrete observations $(x_{ij}, j = 0, 1, \ldots, M)$, $i = 1, \ldots, n$, by taking $x_{ij} = X_i(\tau_j)$, where the random sample $(X_j, j = 1, \ldots, n)$ is generated as in scenario (S1). Discrete observations are converted to the functional data $(X_j, j = 1, \ldots, n)$ by using B-spline bases.

(S3) Discrete observations $(i/M, y_{ij}), i = 0, 1, \ldots, M, j = 1, \ldots, n$, are generated by taking

$$y_{ij} = M^{-1/2} \sum_{k=1}^{i} \xi_{kj},$$

so that y_{ij} can be interpreted as the observation of a standard Wiener process at i/M. From $(y_{ij}, i = 1, \ldots, M)$, the function Y_j is obtained

using the B-spline smoothing technique. During the simulation, we used $M = 1000$ and $D = 50$ B-spline functions, thus obtaining $n = 500$ functions $Y_1, \ldots, Y_n$.

Then, we define for $j = 1, \ldots, n$,

$$X_j = \begin{cases} Y_j, & \text{under null} \\ u_n(j/n)q_n + Y_j, & \text{under alternative} \end{cases}$$

and consider different configurations u_n of change-points and $q_n(t) = a_n\sqrt{Mt}, t \in [0, 1]$.

We mainly concentrated on two possible change-point alternatives. The first is obtained with $u_n(t) = \mathbf{1}_{[0,\theta]}(t)$ and corresponds to one change-point alternative. Another is for the epidemic-type alternative, for which we take $u_n(t) = \mathbf{1}_{[\theta_1,\theta_2]}(t)$.

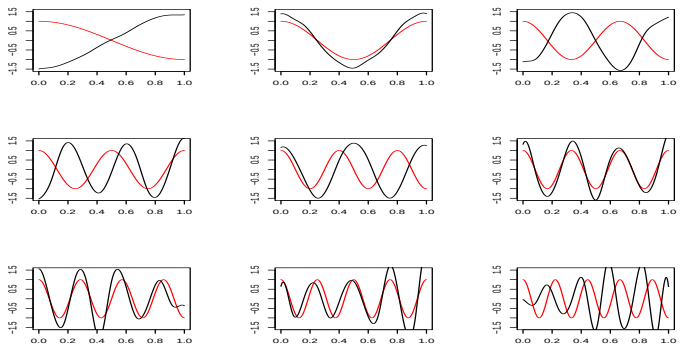


Figure 3.4: True basis functions (red) and "reconstructed" basis functions (black) using fPCA method.

Scenario (S1) is used as an optimal case situation where the actual eigenvalues and eigenfunctions are known. In this case, we are not required to approximate discrete functions, thus avoiding any data loss or measurement errors. The second scenario continues with the same random functional sample but goes through extra steps such as taking function values at discrete data points and reconstructing the random functional sample on a different set of basis functions. The aim of this exercise is to measure the impact when some information could be lost due to measurements taken at discrete points and smoothing. The simulation results show that, even after the reconstruction of the random functional sample, the performance of the test does not suffer too much.

Our simulation starts with the generation process of the random functional sample Y_j as described in the first scenario with $d = 30$.

First of all, we can compare the true eigenfunctions of covariance operator $\Gamma = E[Y_j \otimes Y_j]$ with the eigenfunctions of estimated operator $\widehat{\Gamma}_n$ (see Figure 3.4).

We see that the estimated harmonics have almost the same shape; only every second, the estimated eigenfunctions are phase shifted.

Next, for both scenarios (S1) and (S2), the density functions of the test statistic $T_{n,1}(d)$ (3.17) were estimated using Monte Carlo with 10,000 repetitions (see Figure 3.5). It shows four density plots: the red density functions of $T_{n,1}(d)$ are calculated using the true eigenfunctions and eigenvalues, while the black curves show the density of $\widehat{T}_{n,1}(d)$ (3.26) using the estimated eigenfunctions and eigenvalues. The left side density plots were estimated from the samples under the null hypothesis, while the plots on the right side show the density of $T_{n,1}(d)$ and $\widehat{T}_{n,1}(d)$ with the sample:

$$X_j(t) = \sum_{k=1}^{d} (\xi_{jk} + \mathbf{1}_{(\tau>j)} a) \frac{\sqrt{2}\cos(k\pi t)}{k\sigma}, t \in [0,1], \tau = 250, j = 1 \dots 500 \quad (3.38)$$

with added drift $a = 0.2$.

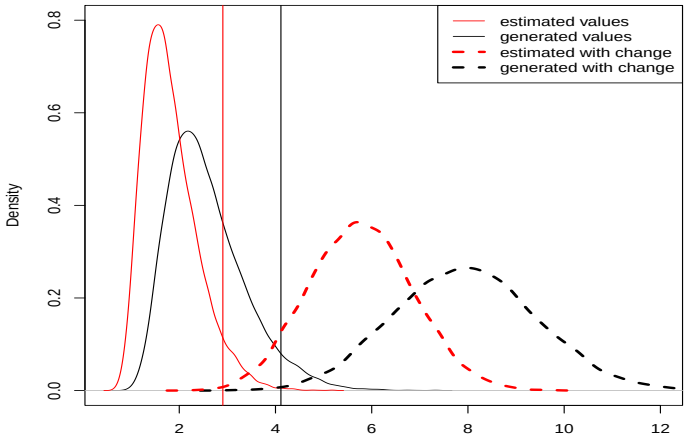


Figure 3.5: Density plots of the test statistic $T_{n,1}(d)$ and $\widehat{T}_{n,1}(d)$.

Since fPCA represents a functional data sample in the most parsimonious way, we can see that the density of the test statistics in scenario (S2) is more on the left side and more concise. Critical values $c_d(\alpha)$ with $\alpha = 0.05$ of the statistics $T_{n,1}(d)$ and $\widehat{T}_{n,1}(d)$ were also calculated and are shown in Figure 3.5.

3.4.2. Statistical Power Analysis

First, we compared the statistical power of the test (3.20) with statistic $T_{n,1}(d)$ of the scenario (S1) and scenario (S2) with statistic $\widehat{T}_{n,1}(d)$. To this aim, we used sample $(X_j, j = 1, \ldots, n)$ defined in (3.38), where $\tau = 250$, which is in the middle of the sample. We started with the no drift $a = 0$ (corresponding to the null hypothesis) and increasing the drift amount a by 0.03 up to the point when $a > 0.3$. At each a value, we repeated the simulation 1000 times. This gives a good indication of the statistical power with the amount of the added drift. The statistical power is illustrated in Figure 3.6. Based on the simulation results, we can see that, even if the random functional sample is approximated from the discrete data points, it still holds the same statistical power and the performance does not suffer from the information loss due to smoothing and fPCA. These are important results, because, normally, in observed real-world data, the true functions are unknown and have to be approximated, which almost always introduces measurement errors.

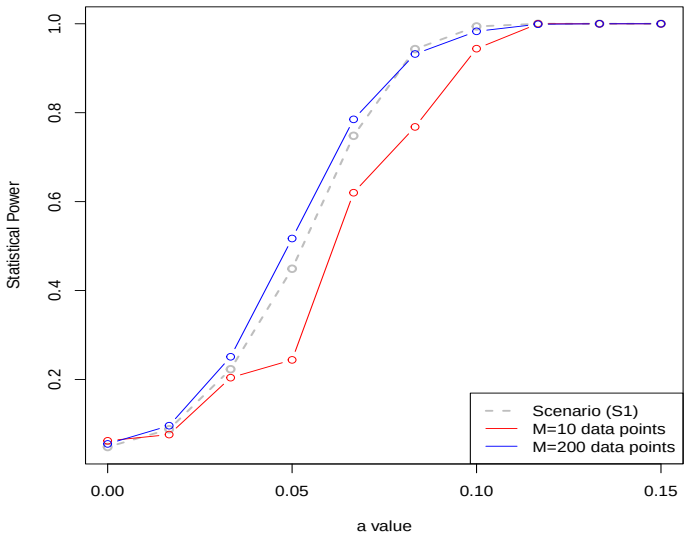
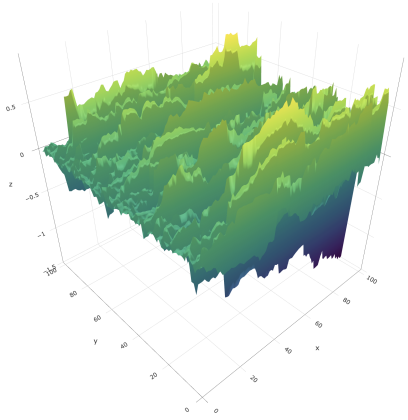


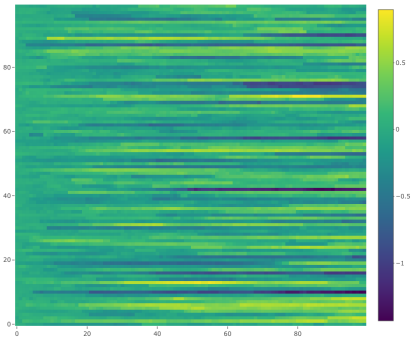
Figure 3.6: The comparison of the statistical power of scenario (S1) and (S2)

Next, we focus on the power tests (3.36) and (3.32) used directly on the functional data sets simulated in scenario (S3).In Figures 1.6a and 1.6b we presented the clear opposites of the functional data sets with respect to the change-point. The changes can be easily observed. However, especially working with functional data sets, the changes may not be that obvious. As an example, Figure 3.7 illustrates another functional data set with the change-point, where the presence of the change-point is not visible, but Monte Carlo experiments show that, with the same magnitude of change, for almost 80% of the cases,

H_0 was correctly rejected.



(a) The Three-dimensional plot illustrates the variations in mean across the dataset



(b) The heatmap presents a visual representation of the differences in mean across the dataset, with changes in mean represented by variations in color intensity

Figure 3.7: Sample with introduced drift of magnitude $a = 0.004$ after the change-point.

The density of the limiting distribution and asymptotic critical values were estimated using the Monte Carlo technique by simulating a Brownian bridge with 1000 points and running 100,000 replications.

In the power studies, we tested two variants of the random functional samples, one with a single change-point in the middle of the functional sample and the second with the two change-points forming epidemic change. In the first case, the functional sample $X_1, \ldots, X_n$ is constructed from $n = 1000$ random functions where 500 curves are changed in order to violate the null hypothesis. The model that violates the null hypothesis is defined as $X_k(t) = \Delta(t)\mathbf{1}\{i > n/2\} + Y_i(t)$, $\Delta(t) = a\sqrt{Mt}$, $t \in [0,1], M = 1000$, and the parameter a is used to control the magnitude of the drift after the change-point. In the second case, during each iteration, $n = 1500$ random functions are generated, where 500 curves in the middle were modified by taking $X_k(t) = \Delta(t)\mathbf{1}\{2n/3 > i > n/3\} + Y_i(t)$, $t \in [0,1]$. During each repetition, two statistics are calculated: $\widehat{T}_n(d,p)$ (3.35) and $\widehat{T}_{n,1}(d)$ (3.26) in the single change-point simulation. For the epidemic change simulation $\widehat{T}_{n,m}(d,p)$ (3.31), $m = 2$ statistic is calculated. We set the p-variation p parameter to 3. We also tested with different p-values, but this did not have any impact on the overall performance.

Figure 3.8 presents the results of the statistical power simulation of both cases: only one change point (left side figure) and epidemic change (right side plot). From the results, we can see that epidemic change has weaker statis-

tical power when using statistic $\widehat{T}_{n,m}(d,p)$ compared to an unknown number of change point statistic $\widehat{T}_n(d,p)$. On the other hand, when restricting the partition count, we observed one benefit, that the locations of the partitions in many cases match or are very close to the actual locations of the change-point.

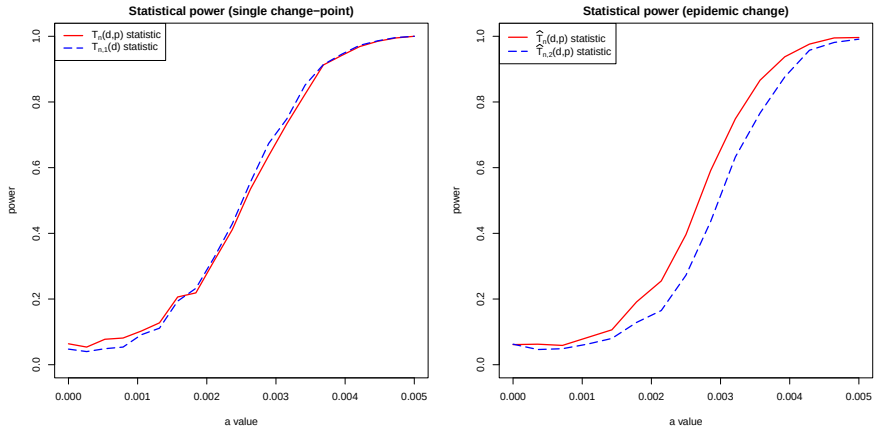


Figure 3.8: Power curves.

3.5. Application to Brain Activity Data

The findings of real data analysis to show the performance of the proposed test are demonstrated in this section. The data were collected during a long-term study on voluntary alcohol- consuming rats following chronic alcohol experience. The data consist of two sets: neurophysiological activity from the two brain centers (the dorsal and ventral striatum) and data from the lickometer device. The lickometer devices were used to monitor the drinking bouts. During the single trial, two locations of the brain were monitored for each rat. Rats were given two drinking bouts, one with alcohol and the other with water. Any time, they were able to freely choose what to drink. Electrodes were attached to the brains, and neurophysiological data were sampled at 1kHz intervals. It was not the goal of this study to confirm nor reject the findings, but to show the advantages of the functional approach for change-point detection. For this reason, the data are well-suited to illustrate the behavior of the test in real-world settings.

In our analysis, we took the first alcohol-drinking event, which lasted around 27s. We also included 10s before the drinking event and 10s after the event. The total time was 47 s long. The time series was broken down into processes

of 100ms. Each process had 100 data points.

$$A_{470,100} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,100} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,100} \\ \vdots & \vdots & \ddots & \vdots \\ a_{470,1} & a_{470,2} & \cdots & a_{470,100} \end{pmatrix}$$

All the processes were smoothed to the functions using 50 B-spline basis functions. The overall functional sample contained 470 functions $\widehat{F} = [f_1, f_2, \ldots, f_{470}]$. The functional sample was separated into sub-samples $\widehat{F}_i = [f_1, f_2, \ldots, f_{20+i}]$, $i = 0, 1, \ldots, 450$. For each sub-sample $\widehat{F}_i$, two statistics were calculated ($\widehat{T}_n(d,p)$ (3.35) and $\widehat{T}_{n,m}(d,p)$ (3.31), $m = 2$).

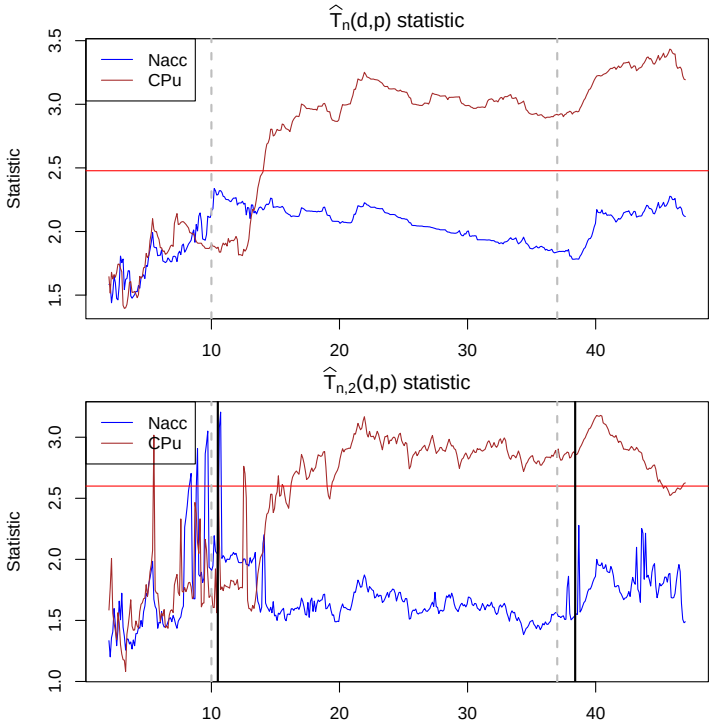


Figure 3.9: Statistics of the first alcohol drinking event, which lasted about 27 s. Ten seconds before and 10 s after were also included. The red horizontal line indicates the critical value with α level 0.95. Vertical gray dashed lines mark the beginning and the end of the drinking time. The black solid vertical lines mark the locations of the change points detected using the restricted p-variation method. Blue and light blue colors represent different brain regions.

The results are visualized in Figure 3.9. We can see that tests with statistics $\widehat{T}_n(d,p)$ and $\widehat{T}_{(n,m)}(d,p)$ strongly rejected the null hypothesis at around 2 s and onward after the rat started to consume the alcohol, which suggests that the

changes in the brain activity can be observed. However, the changes appear to happen only for the CPu brain region. Interestingly, the statistic $\widehat{T}_{n,m}(d,p)$ has much larger volatility compared to the unrestricted $\widehat{T}_n(d,p)$ in the Nacc brain region before the drinking event and lower volatility just after the drinking event started. However, it is not fully clear if this is the expected behavior or a Type I error.

Finally, the locations of the restricted ($m = 2$) p-variation partition points nearly matched the beginning and the end of the drinking period. In Figure 3.9, the gray vertical dashed lines indicate the actual beginning and the actual end of the drinking period measured by the lickometer and the black vertical lines indicate the location of the partitions calculated from the functional sample $\widehat{F}_{450}$. The first partition is located at 10.5 s and the second partition point at 38.4 s, which aligns well with the data collected from the lickometer.

The test with a restricted partition count showed weaker statistical power, but it did help determine the location of the change-points.

3.6. Proof of Theorems 10 and 12

The following theorem is a version of Theorem 2.11.1 in [75] adapted to the case of continuous processes.

Theorem 21. Assume that $\{Z_{ni}\colon 1 \le i \le m_n\}$ are independent continuous stochastic processes indexed by a totally bounded semi-metric space $(\mathcal{G}, d)$ such that

$$\lim_{n\to\infty} \sum_{i=1}^{m_n} E\|Z_{ni}\|_{\mathcal{F}}^2 1_{\{\|Z_{ni}\|_{\mathcal{F}}>\eta\}} = 0 \quad \text{for every } \eta > 0, \tag{3.39}$$

$$\lim_{n\to\infty} \sup_{d(f,g)<\delta_n} \sum_{i=1}^{m_n} E\big[Z_{ni}(f) - Z_{ni}(g)\big]^2 = 0 \quad \text{for every } \delta_n \downarrow 0, \tag{3.40}$$

$$\int_0^{\delta_n} \sqrt{\log N(\epsilon, \mathcal{G}, d_n)}\, \mathrm{d}\epsilon \xrightarrow[n\to\infty]{\mathrm{P}} 0 \quad \text{for every } \ \delta_n \downarrow 0, \tag{3.41}$$

where

$$d_n(f,g) = \Big(\sum_{k=1}^{m_n} [Z_{nk}(f) - Z_{nk}(g)]^2\Big)^{1/2}.$$

Then, the sequence $Z_n := \sum_{i=1}^{m_n}(Z_{ni} - EZ_{ni})$ is asymptotically d-equicontinuous, that is, for every $\varepsilon > 0$,

$$\lim_{\delta\downarrow 0} \limsup_{n\to\infty} P\Big(\sup_{d(f,g)<\delta} |Z_n(f) - Z_n(g)| > \varepsilon\Big) = 0.$$

Furthermore, (Z_n) converges in law in $\ell^\infty(\mathcal{G})$ provided that covariances converge pointwise on $\mathcal{G} \times \mathcal{G}$.

Proof of Theorem 10 (1a). Without loss of generality, we assumed that $\|\psi\| \le 1$ for all $\psi \in \Psi$ and $\|f\|_{\sup} \le 1$ for all $f \in \mathcal{F}$. To prove (1a), we applied Theorem 21 for $\mathcal{G} = \mathcal{F} \times \Psi$, $d = \rho_2$, and $Z_{nk} = n^{-1/2}\nu_{nk}$, $k = 1, \dots, n$, where, under H_0, $\nu_{nk}(f, \psi) = \langle Y_k, \psi\rangle \lambda_{nk}(f)$. Let us check first the conditions (3.39)–(3.41). We have

$$n^{-1/2}\|\nu_{nk}\|_{\mathcal{G}} = n^{-1/2} \sup_{\psi\in\Psi, f\in\mathcal{F}} |\langle Y_k, \psi\rangle \lambda_{nk}(f)| \le n^{-1/2}\|Y_k\| \sup_{\psi\in\Psi} \|\psi\| \sup_{f\in\mathcal{F}} \|f\|_{\sup}$$
$$\le n^{-1/2}\|Y_k\|.$$

Hence, (3.39) easily follows from $E\|Y\|^2 < \infty$. Since

$$\nu_{nk}(f, \psi) - \nu_{nk}(f', \psi') = \langle Y_k, \psi - \psi'\rangle \lambda_{nk}(f) + \langle Y_k, \psi'\rangle \lambda_{nk}(f - f')$$

and Y, Y_k are identically distributed, we have

$$E[\nu_{nk}(f, \psi) - \nu_{nk}(f', \psi')]^2 \le 2E\|Y\|^2[\lambda_{nk}^2(f - f')\|\psi'\|^2 + \lambda_{nk}^2(f)\|\psi - \psi'\|].$$

Summing this estimate and noting that for any $g \in L_2(0, 1)$,

$$n^{-1}\sum_{k=1}^n \lambda_{nk}^2(g) \le n^{-1}\sum_{k=1}^n \lambda_{nk}(g^2) = \|g\|^2$$

by the Hölder inequality, we find

$$n^{-1}\sum_{k=1}^n E[\nu_{nk}(f, \psi) - \nu_{nk}(f', \psi')]^2 \le 2E\|Y\|^2[\|\psi\|^2\|f - f'\|^2 + \|f'\|\|\psi - \psi'\|^2]$$
$$\le 2E\|Y\|^2[\|\psi\|^2 + \|f'\|^2]\delta_n$$
$$\le 4E\|Y\|^2\delta_n$$

if $\rho_2((f, \psi), (f', \psi')) < \delta_n$. This estimate yields (3.40). To check (3.41), we have

$$d_n((f, \psi), (f', \psi')) = \Big(n^{-1}\sum_{k=1}^n [\langle Y_k, \psi\rangle \lambda_{nk}(f) - \langle Y_k, \psi'\rangle \lambda_{nk}(f')]^2\Big)^{1/2}$$
$$= \Big(n^{-1}\sum_{k=1}^n [\langle Y_k, \psi - \psi'\rangle \lambda_{nk}(f) + \langle Y_k, \psi'\rangle \lambda_{nk}(f - f')]^2\Big)^{1/2}$$
$$\le \Big(n^{-1}\sum_{k=1}^n \langle Y_k, \psi - \psi'\rangle^2 \lambda_{nk}^2(f)\Big)^{1/2} + \Big(n^{-1}\sum_{k=1}^n \langle Y_k, \psi'\rangle \lambda_{nk}^2(f - f')\Big)^{1/2}$$
$$\le \Big(n^{-1}\sum_{k=1}^n \|Y_k\|^2 \lambda_{nk}(f^2)\Big)^{1/2} \rho_{2,\lambda}(\psi, \psi') + \Big(n^{-1}\sum_{k=1}^n \langle Y_k, \psi'\rangle^2 \lambda_{nk}(f - f')^2\Big)^{1/2}$$
$$\le A_n[\rho_{2,\lambda}(\psi, \psi') + \rho_{2,Q}(f, f')],$$

where

$$A_n = \Big(n^{-1}\sum_{k=1}^n \|Y_k\|^2\Big)^{1/2}, \quad \text{and} \quad Q = A_n^{-2} n^{-1} \sum_{k=1}^n \|Y_k\|^2 \lambda_{nk}$$

Hence,

$$N(\varepsilon, \mathcal{F}\times\Psi, d_n) \le N(A_n^{-1}\varepsilon, \mathcal{F}, \rho_{2,Q}) N(A_n^{-1}\varepsilon, \Psi, \rho_{2,\lambda}).$$

and the condition (3.41) is satisfied, provided that

$$I_1(\delta_n) := \int_0^{\delta_n} \sup_{Q\in\mathcal{Q}} \sqrt{\log N(A_n^{-1}\epsilon, \mathcal{F}, \rho_{2,Q})}\, \mathrm{d}\epsilon \xrightarrow[n\to\infty]{\mathrm{P}} 0 \quad \text{for every} \quad \delta_n \downarrow 0, \tag{3.42}$$

and

$$I_2(\delta_n) := \int_0^{\delta_n} \sqrt{\log N(B_n^{-1}\epsilon, \Psi, \rho_{2,\lambda})}\, \mathrm{d}\epsilon \xrightarrow[n\to\infty]{\mathrm{P}} 0 \quad \text{for every} \quad \delta_n \downarrow 0, \tag{3.43}$$

hold. Set

$$J_1(a) := \int_0^a \sup_{Q\in\mathcal{Q}} \sqrt{\log N(\epsilon, \mathcal{F}, \rho_{2,Q})}\, \mathrm{d}\epsilon, \quad J_2(a) := \int_0^a \sqrt{\log N(\epsilon, \Psi, \rho_{2,\lambda})}\, \mathrm{d}\epsilon.$$

It is known (see, e.g., [49]) that $J_1(1) < \infty$. Hence, $J_1(a) \to 0$ as $a \to 0$. By the condition (3.4), $J_2(a) \to 0$ as $a \to 0$. Changing the integration variables gives $I_1(\delta_n) = A_n J_1(A_n^{-1}\delta_n)$ and $J_2(\delta_n) = A_n J_2(A_n^{-1}\delta_n)$.

Set $\sigma^2 := E\|Y\|^2$. By the strong law of large numbers, $A_n^2 \xrightarrow[n\to\infty]{\mathrm{P}} \sigma^2$. Choosing $\eta < 3\sigma^2/4$, we have, for any $\delta > 0$,

$$\begin{aligned} P\Big(I_1(\delta_n) > \delta) &\le P(A_n J_1(A_n^{-1}\delta_n) > \delta, |A_n^2 - \sigma^2| < \eta\Big) + P(|A_n^2 - \sigma^2| > \eta) \\ &\le P(A_n J_1(A_n^{-1}\delta_n) > \delta, A_n^2 > \sigma^2/4\Big) + P(|A_n^2 - \sigma^2| > \eta) \\ &\le P\Big(A_n J_1(\eta\delta_n/2) > \delta) + P(|A_n^2 - \sigma^2| > \eta) \to 0 \end{aligned}$$

as $n \to \infty$. Similarly, we prove $I_2(\delta_n) \xrightarrow[n\to\infty]{\mathrm{P}} 0$.

Next, we have to check the pointwise convergence of the covariances of (Z_n). Since Y_k are independent, we have

$$\begin{aligned} E\Big(\sum_{k=1}^n \langle Y_k, \psi\rangle \lambda_{nk}(f) \sum_{k=1}^n \langle Y_k, \psi'\rangle \lambda_{nk}(f')\Big) &= \sum_{k=1}^n E\Big(\langle Y_k, \psi\rangle\langle Y_k, \psi'\rangle\Big)\lambda_{nk}(f)\lambda_{nk}(f') \\ &= (\Gamma\psi, \psi') \sum_{k=1}^n \lambda_{nk}(f)\lambda_{nk}(f'). \end{aligned}$$

We shall prove that

$$I_n := n^{-1}\sum_{k=1}^{n}\lambda_{nk}(f)\lambda_{nk}(f') \to \langle f, f'\rangle \ \text{ as } \ n\to\infty.$$

Set $\widetilde{I}_n := n^{-1}\sum_{k=1}^{n} f(k/n)f'(k/n)$. Evidently, $\lim_{n\to\infty}\widetilde{I}_n = \langle f, f'\rangle$, and we have to check

$$\lim_{n\to\infty}|I_n - \widehat{I}_n| \to 0.$$

We have

$$\Delta_n := |I_n - \widehat{I}_n| \le n^{-1}\sum_{k=1}^{n}[\lambda_{nk}(f)\lambda_{nk}(f') - f(k/n)f'(k/n)] \le |\Delta_n'| + |\Delta_n''|,$$

where

$$\Delta_n' := n^{-1}\sum_{k=1}^{n}[\lambda_{nk}(f)(\lambda_{nk}(f')-f'(k/n))],\ \ \Delta_n'' := n^{-1}\sum_{k=1}^{n}[f'(k/n)(\lambda_{nk}(f)-\lambda_{nk}(f'))].$$

Observing that

$$\begin{aligned}
|\lambda_{nk}(f) - f(k/n)| &= \Big|\int_0^1 (f(t)-f(k/n))\,\mathrm{d}\lambda_{nk}(t)\Big| \le n\Big|\int_{(k-1)/n}^{k/n}(f(t)-f(k/n))\,\mathrm{d}t\Big| \\
&\le \int_0^1 \sup\{|f(t)-f(k/n)| : t\in[(k-1)/n, k/n]\}\,\mathrm{d}\lambda_{nk} \\
&\le \sup\{|f(t)-f(k/n)| : t\in[(k-1)/n, k/n] \\
&\le v_2^{1/2}(f;[(k-1)/n, k/n)]),
\end{aligned}$$

we have

$$\begin{aligned}
|\Delta_n| &\le n^{-1}\lambda_{nk}(f)v_2^{1/2}(f,[(k-1]/n, k/n]) \\
&\le n^{-1}\Big(\sum_{k=1}^{n}\lambda_{nk}^2(f)\Big)^{1/2}\Big(\sum_{k=1}^{n}v_2(f,[(k-1)/n,k/n])\Big)^{1/2} \\
&\le n^{-1/2}\|f\|v_2^{1/2}(f).
\end{aligned}$$

This yields

$$\lim_{n\to\infty}E(Z_n(f,\psi)Z_n(f',\psi')) = \langle\Gamma\psi,\psi'\rangle\langle f, f'\rangle.$$

To complete the proof of (a), note that the existence of the continuous modification of Gaussian process $\nu = \nu(\psi, f), (\psi, f) \in \mathcal{G} = \Psi\times\mathcal{F}$) follows by Dudley [22], since the entropy condition $\int_0^1\sqrt{\log N(\epsilon,\mathcal{G},\rho_2)}\,\mathrm{d}\epsilon < \infty$ is satisfied. □

Lemma 4. It holds that

$$\lim_{n\to\infty} \sup_{||g||_{(2)}\le 1} \Big| n^{-1} \sum_{k=1}^{n} g(k/n) - \lambda(g) \Big| = 0.$$

Proof. We have

$$I_n := \frac{1}{n} \sum_{k=1}^{n} g(k/n) - \lambda(g) = \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} [g(k/n) - g(s)] \, \mathrm{d}s$$

For every $s \in [(k-1)/n, k/n]$,

$$|g(k/n) - g(s)| \le v_2^{1/2}(g, [(k-1)/n, k/n]).$$

Hence,

$$\begin{aligned} |I_n| &\le \sum_{k=1}^{n} \int_{(k-1)/n}^{k/n} v_2^{1/2}(g, [(k-1)/n, k/n]) \, \mathrm{d}s \\ &= \frac{1}{n} \sum_{k=1}^{n} v_2^{1/2}(g, [(k-1)/n, k/n]) \le \frac{1}{\sqrt{n}} \Big(\sum_{k=1}^{n} v_2(g, (k-1)/n, k/n]) \Big)^{1/2} \\ &\le \frac{1}{\sqrt{n}} \|g\|_{(2)} \end{aligned}$$

and this completes the proof. □

Proof of Theorem 10 (1*b*). Under H_A,

$$\begin{aligned} \langle X_k, \psi \rangle \lambda_{nk}(f) &= \langle Y_k, \psi \rangle \lambda_{nk}(f) + \langle g_n(k/n, \cdot), \psi \rangle \lambda_{nk}(f) \\ &= \langle Y_k, \psi \rangle \lambda_{nk}(f) + n^{-1/2} u(k/n) [\langle a, \psi \rangle + \langle a_n, \psi \rangle]. \end{aligned}$$

Hence,

$$\nu_n(f, \psi) = \widehat{\nu}_n(f, \psi) + \Delta_n(f, \psi) + r_n(f, \psi),$$

where

$$\widehat{\mu}_n(f, \psi) = \sum_{i=1}^{n} \langle Y_i, \psi \rangle \lambda_{ni}(f),$$

$$\Delta_n(f, \psi) = n^{-1/2} \sum_{k=1}^{n} u(i/n) \lambda_{ni}(f) \langle q, \psi \rangle$$

and

$$r_n(f, \psi) = n^{-1/2} \sum_{k=1} u(i/n)] \lambda_{ni}(f) \langle q_n, \psi \rangle.$$

We have by (1a)

$$n^{-1/2}\widehat{\nu}_n \xrightarrow[n\to\infty]{\mathcal{D}} \nu_\gamma \quad \text{in the space} \quad \ell^\infty(\mathcal{F}\times\Psi).$$

To complete the proof, we have to check

$$\lim_{n\to\infty}\sup_{f\in\mathcal{F},\psi\in\Psi}|n^{-1/2}\Delta_n(f,\psi)-\Delta(f,\psi)|=0. \tag{3.44}$$

and

$$\lim_{n\to\infty}\sup_{f\in\mathcal{F},\psi\in\Psi}|n^{-1/2}r_n(f,\psi)|=0. \tag{3.45}$$

To this aim, we involve lemma 4. We have

$$n^{-1/2}\Delta_n(f,\psi)-\Delta(f,\psi)=[I_{n1}(f)+I_{n2}(f)]\langle q,\psi\rangle,$$

where

$$I_{n1}(f)=\frac{1}{n}\sum_{k=1}^{n}u(k/n)[\lambda_{nk}(f)-f(k/n)],\ \ I_{n2}(f)=\frac{1}{n}\sum_{k=1}^{n}u(k/n)f(k/n)-\langle u,f\rangle.$$

By Lemma 4 applied to the function uf, we have $I_{n2}(f)\to 0$ uniformly over $f\in\mathcal{F}$. Consider I_{n1}. We have, as in the proof of Lemma 4,

$$|I_{n1}(f)|\le\sum_{k=1}^{n}|u(k/n)|\int_{(k-1)/n}^{k/n}[f(s)-f(k/n)|\,\mathrm{d}s\le n^{-1/2}\|u\|_\infty\|f\|_{(2)}.$$

Hence, $I_{n2}(f)\to 0$ uniformly over $f\in\mathcal{F}$. The convergence (3.45) follows by observing that

$$|n^{-1/2}r_n(f,\psi)|\le\|u\|_\infty\int_0^1|f(s)|\,\mathrm{d}s\|\psi\|\cdot\|q_n\|.$$

This proves (3.45) and completes the proof of (1b). □

Proof of Theorem 12 (3a). Consider the map $T:\ell^\infty(\mathcal{F})\to\ell^\infty(\mathcal{F}), T(x)(f)=x(f)-x(1)\lambda(f)$. The continuity of T is easy to check. Observing that $\widehat{\nu}_n=T(\nu_n)$, the convergence (3.9) is a corollary of Theorem 10 and a continuous mapping theorem.

To prove (3b), observe that, under H_A,

$$\begin{aligned}\langle X_k,\psi\rangle\lambda_{nk}(f)&=\langle Y_k,\psi\rangle\lambda_{nk}(f)+\langle g_n(k/n,\cdot),\psi\rangle\lambda_{nk}(f)\\&=\langle Y_k,\psi\rangle\lambda_{nk}(f)+n^{-1/2}u(k/n)[\langle v,\psi\rangle+\langle v_n,\psi\rangle]\end{aligned}$$

hence

$$\mu_n(f,\psi)=\widehat{\mu}_n(f,\psi)+\widetilde{\Delta}_n(f,\psi)+\widetilde{r}_n(f,\psi),$$

where

$$\widehat{\mu}_n(f,\psi) = \sum_{i=1}^{n} \langle Y_i - \overline{Y}_n, \psi\rangle \lambda_{ni}(f),$$

$$\widetilde{\Delta}_n(f,\psi) = n^{-1/2} \sum_{k=1}^{n} [u(i/n) - n^{-1} \sum_{j=1}^{n} u(j/n)] \lambda_{ni}(f) \langle q, \psi\rangle$$

and

$$\widetilde{r}_n(f,\psi) = n^{-1/2} \sum_{k=1} [u(i/n) - n^{-1} \sum_{j=1}^{n} u(j/n)] \lambda_{ni}(f) \langle q_n, \psi\rangle.$$

We have by (a)

$$n^{-1/2}\widehat{\mu}_n \xrightarrow[n\to\infty]{\mathcal{D}} \mu_\gamma \quad \text{in the space} \quad \ell^\infty(\mathcal{F}\times\Psi).$$

To complete the proof, we have to check

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F},\psi\in\Psi} |n^{-1/2}\widetilde{\Delta}_n(f,\psi) - \widetilde{\Delta}(f,\psi)| = 0. \tag{3.46}$$

and

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F},\psi\in\Psi} |n^{-1/2}\widetilde{r}_n(f,\psi)| = 0. \tag{3.47}$$

For this, we can use (3.44) and (3.45) and observe that $n^{-1}\sum_{k=1}^{n} u(k/n) \to \lambda(u)$ as $n\to\infty$. □

4. Concluding remarks and future works

In this dissertation, we aimed to develop a method for detecting change points in functional data samples. To this end, we proposed a mean instability testing model based on the p-variation of the process of partial sums and analyzed its statistical power through simulation methods. We also established the limiting distribution for the null and alternative hypotheses theoretically. Furthermore, we generalized the results of the univariate test and applied them to functional data.

In addition, we studied the asymptotic behavior of the $\mathcal{G}$-sums processes indexed by functions and established the limiting distribution theoretically. We proposed tests for detecting one change point, no more than m change points, and an unknown number of change points, and analyzed these tests using simulation methods on real data.

The results of this thesis contribute to the growing body of knowledge in the field of functional data analysis and provide a new method for detecting change points in functional data samples.

This study has identified several opportunities for further research and adaptation of the methods introduced here. At their core, all of these methods rely on the p-variation value, which is calculated by partitioning the sequence into partitions. The location of these partition points can provide additional interpretability and may help to localize the change points. In Chapter 3, we demonstrated that the location of the partition points in a real dataset closely matched the location of the true change points. This observation suggests that there may be potential for further exploitation of this relationship with a more robust justification.

Overall, there are many different tests and experiments that could be conducted in future work to further develop and refine these methods. This will be an important area of research as we continue to explore the capabilities of functional data analysis in detecting change points.

The Continuous Wavelet Transform (CWT) technique produces real-valued functions and represents nonstationary signals in a time-frequency domain. There is a possibility that by representing signal information in this way, the proposed functional change point detection tests could be adapted to detect changes in periodicity. Further research is needed to explore this idea and determine the feasibility and potential of using the CWT and functional change point detection techniques in combination.

Finally, this thesis has not covered many real-world applications where the methods introduced here could be applicable. Further research is needed to explore the potential for applying these methods in various contexts and to understand the full range of their capabilities and limitations.

Bibliography

[1] Samaneh Aminikhanghahi and Diane J. Cook. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, 5 2017.

[2] John A.D. Aston and Claudia Kirch. Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis*, 109:204–220, 8 2012.

[3] Alexander Aue, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis*, 100(10):2254–2269, 11 2009.

[4] Alexander Aue, Gregory Rice, and Ozan Sönmez. Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):509–529, 6 2018.

[5] Buddhananda Banerjee and Satyaki Mazumder. A more powerful test identifying the change in mean of functional data. *Annals of the Institute of Statistical Mathematics*, 70(3):691–715, 6 2018.

[6] Sayantan Banerjee and Kousik Guhathakurta. Change-point Analysis in Financial Networks. 11 2019.

[7] G A Barnard. Control Charts and Stochastic Processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 21(2):239–271, 1959.

[8] M. Basseville. Edge detection using sequential methods for change in level–Part II: Sequential detection of change in mean. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(1):32–50, 2 1981.

[9] Michèle Basseville. Detecting changes in signals and systems—A survey. *Automatica*, 24(3):309–326, 5 1988.

[10] Idoia Beraza and Iñaki Romero. Comparative study of algorithms for ECG segmentation. *Biomedical Signal Processing and Control*, 34:166–173, 4 2017.

[11] István Berkes, Robertas Gabrys, Lajos Horváth, and Piotr Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):927–946, 11 2009.

[12] Marcel Bosc, Fabrice Heitz, Jean-Paul Armspach, Izzie Namer, Daniel Gounot, and Lucien Rumbach. Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. *NeuroImage*, 20(2):643–656, 10 2003.

[13] B. E. Brodsky and B. S. Darkhovsky. *Nonparametric Methods in Change-Point Problems*. Springer Netherlands, Dordrecht, 1993.

[14] Vygantas Butkus and Rimas Norvaiša. Computation of p-variation. *Lithuanian Mathematical Journal*, 58(4):360–378, 10 2018.

[15] Vicent Caselles, Antonin Chambolle, and Matteo Novaga. The Discontinuity Set of Solutions of the TV Denoising Problem and Some Extensions. *Multiscale Modeling & Simulation*, 6(3):879–894, 1 2007.

[16] Jie Chen and Arjun K. Gupta. *Parametric Statistical Change Point Analysis*. Birkhäuser Boston, Boston, 2012.

[17] Hyunyoung Choi, Hernando Ombao, and Bonnie Ray. Sequential Change-Point Detection Methods for Nonstationary Time Series. *Technometrics*, 50(1):40–52, 2 2008.

[18] George W. Cobb. The Problem of the Nile: Conditional Solution to a Changepoint Problem. *Biometrika*, 65(2):243, 8 1978.

[19] C. Cotsaces, N. Nikolaidis, and I. Pitas. Video shot detection and condensed representation. a review. *IEEE Signal Processing Magazine*, 23(2):28–37, 3 2006.

[20] J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12(1):136–154, 3 1982.

[21] Carl de Boor. *A Practical Guide to Splines*, volume 27. Springer New York, New York, NY, revised edition edition, 2001.

[22] R. M. Dudley. Sample Functions of the Gaussian Process. *The Annals of Probability*, 1(1), 2 1973.

[23] R. M. Dudley and R. Norvaiša. *Concrete Functional Calculus*. Springer New York, New York, NY, 2011.

[24] Frédéric Ferraty and Philippe Vieu. *Nonparametric functional data analysis: theory and practice*, volume 76. Springer, 2006.

[25] Atiyeh Fotoohinasab, Toby Hocking, and Fatemeh Afghah. A Graph-constrained Changepoint Detection Approach for ECG Segmentation. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 332–336. IEEE, 7 2020.

[26] Vasile Georgescu. Online change-point detection in financial time series: challenges and experimental evidence with frequentist and bayesian setups. In *Methods for Decision Making in an Uncertain Environment*, pages 131–145. World Scientific, 7 2012.

[27] Cyril Goutte, Yunli Wang, Fangming Liao, Zachary Zanussi, Samuel Larkin, and Yuri Grinberg. Eurogames16: Evaluating change detection in online conversation. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.

[28] P.J. Green and Bernard. W. Silverman. *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall/CRC, 5 1993.

[29] Ulf Grenander. Stochastic processes and statistical inference. *Arkiv för Matematik*, 1(3):195–277, 10 1950.

[30] L F Haas. Hans Berger (1873-1941), Richard Caton (1842-1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(1):9–9, 1 2003.

[31] Trevor Harris, Bo Li, and James Derek Tucker. Scalable Multiple Changepoint Detection for Functional Data Sequences. 8 2020.

[32] Ami Harten. High resolution schemes for hyperbolic conservation laws. *Journal of Computational Physics*, 49(3):357–393, 3 1983.

[33] Siegfried Hörmann and Piotr Kokoszka. Weakly dependent functional data. 10 2010.

[34] Lajos Horváth and Piotr Kokoszka. *Inference for Functional Data with Applications*, volume 200. Springer New York, New York, NY, 2012.

[35] Jongseo Sohn, Nam Soo Kim, and Wonyong Sung. A statistical model-based voice activity detection. *IEEE Signal Processing Letters*, 6(1):1–3, 1 1999.

[36] C Jordan. Sur la série de Fourier.[On the Fourier Serie]. *Comptes Rendus de l'Académie des Sciences*, 92:228–230, 1881.

[37] J.-C. Junqua, B. Mak, and B. Reaves. A robust algorithm for word boundary detection in the presence of noise. *IEEE Transactions on Speech and Audio Processing*, 2(3):406–412, 7 1994.

[38] Lucy Kendrick, Katarzyna Musial, and Bogdan Gabrys. Change point detection in social networks—Critical review with experiments. *Computer Science Review*, 29:1–13, 8 2018.

[39] Elena A. Khapalova, Venkata K. Jandhyala, Stergios B. Fotopoulos, and James E. Overland. Assessing Change-Points in Surface Air Temperature Over Alaska. *Frontiers in Environmental Science*, 6, 10 2018.

[40] J. Kiefer. K-Sample Analogues of the Kolmogorov-Smirnov and Cramer-V. Mises Tests. *The Annals of Mathematical Statistics*, 30(2):420–447, 6 1959.

[41] Antti Koski. Modelling ECG signals with hidden Markov models. *Artificial Intelligence in Medicine*, 8(5):453–471, 10 1996.

[42] Shuang Li, Yao Xie, Hanjun Dai, and Le Song. M-Statistic for Kernel Change-Point Detection. In C Cortes, N Lawrence, D Lee, M Sugiyama, and R Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[43] Xiuqi Li and Subhashis Ghosal. Bayesian change point detection for functional data. *Journal of Statistical Planning and Inference*, 213:193–205, 7 2021.

[44] Song Liu, Makoto Yamada, Nigel Collier, and Masashi Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, 43:72–83, 7 2013.

[45] E Love and L Young. Sur une classe de fonctionnelles linéaires. *Fundamenta Mathematicae*, 28(1):243–257, 1937.

[46] Rakesh Malladi, Giridhar P Kalamangalam, and Behnaam Aazhang. Online Bayesian change point detection algorithms for segmentation of epileptic activity. In *2013 Asilomar Conference on Signals, Systems and Computers*, pages 1833–1837. IEEE, 11 2013.

[47] Ankita Mitra, Arunava De, and Anup Kumar Bhattacharjee. Detection of Progression of Lesions in MRI Using Change Detection. pages 467–473. 2014.

[48] Siti Nur Afiqah Mohd Arif, Mohamad Farhan Mohamad Mohsin, Azuraliza Abu Bakar, Abdul Razak Hamdan, and Sharifah Mastura Syed Abdullah. Change point analysis: A statistical approach to detect potential abrupt change. *Jurnal Teknologi*, 79(5), 6 2017.

[49] R. Norvaiša and A. Račkauskas. Convergence in law of partial sum processes in p-variation norm. *Lithuanian Mathematical Journal*, 48(2):212–227, 4 2008.

[50] E. S. Page. Continuous Inspection Schemes. *Biometrika*, 41(1/2):100, 6 1954.

[51] E. S. Page. On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44(1-2):248–252, 1957.

[52] Supriya Palaniswami and Krishnaveni Muthiah. Change Point Detection and Trend Analysis of Rainfall and Temperature Series over the Vellar River Basin. *Polish Journal of Environmental Studies*, 27(4):1673–1681, 3 2018.

[53] Jeong-Seon Park, Sang-Woong Lee, and Unsang Park. R Peak Detection Method Using Wavelet Transform and Modified Shannon Energy Envelope. *Journal of Healthcare Engineering*, 2017:1–14, 2017.

[54] Karl Pearson. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572, 11 1901.

[55] Moshe Pollak. Optimal Detection of a Change in Distribution. *The Annals of Statistics*, 13(1):206–227, 1985.

[56] Tatiana Polushina and Georgy Sofronov. Change-point detection in biological sequences via genetic algorithm. In *2011 IEEE Congress of Evolutionary Computation (CEC)*, pages 1966–1971. IEEE, 6 2011.

[57] Qin Qin, Jianqing Li, Yinggao Yue, and Chengyu Liu. An Adaptive and Time-Efficient ECG R-Peak Detection Algorithm. *Journal of Healthcare Engineering*, 2017:1–14, 2017.

[58] Richard E. Quandt. The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association*, 53(284):873, 12 1958.

[59] Richard E. Quandt. Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistical Association*, 55(290):324, 6 1960.

[60] R.J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. Image change detection algorithms: a systematic survey. *IEEE Transactions on Image Processing*, 14(3):294–307, 3 2005.

[61] J. O. Ramsay. When the data are functions. *Psychometrika*, 47(4):379–396, 12 1982.

[62] J O Ramsay and C J Dalzell. Some Tools for Functional Data Analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):539–572, 1991.

[63] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer New York, New York, NY, 2005.

[64] C. Radhakrishna Rao. Some Statistical Methods for Comparison of Growth Curves. *Biometrics*, 14(1):1, 3 1958.

[65] Anton Rauhameri, Katri Salminen, Jussi Rantala, Timo Salpavaara, Jarmo Verho, Veikko Surakka, Jukka Lekkala, Antti Vehkaoja, and Philipp Müller. A comparison of online methods for change point detection in ion-mobility spectrometry data. *Array*, 14:100151, 7 2022.

[66] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 11 1992.

[67] Peijun Sang, Liangliang Wang, and Jiguo Cao. Parametric functional principal component analysis. *Biometrics*, 73(3):802–810, 9 2017.

[68] Hae Jong Seo and Peyman Milanfar. A non-parametric approach to automatic change detection in MRI images of the brain. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 245–248. IEEE, 6 2009.

[69] Han Lin Shang. A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142, 4 2014.

[70] Olimjon Sharipov, Johannes Tewes, and Martin Wendler. Sequential block bootstrap in a Hilbert space with application to change point analysis. *Canadian Journal of Statistics*, 44(3):300–322, 9 2016.

[71] W A Shewhart. *Economic Control of Quality of Manufactured Product*. Bell Telephone Laboratories series. American Society for Quality Control, 1931.

[72] Nikolay Shvetsov, Nazar Buzun, and Dmitry V. Dylov. Unsupervised non-parametric change point detection in electrocardiography. In *32nd International Conference on Scientific and Statistical Database Management*, pages 1–4, New York, NY, USA, 7 2020. ACM.

[73] R. Tahmasbi and S. Rezaei. Change Point Detection in GARCH Models for Voice Activity Detection. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(5):1038–1046, 7 2008.

[74] Charles Truong, Laurent Oudre, and Nicolas Vayatis. Selective review of offline change point detection methods. *Signal Processing*, 167:107299, 2 2020.

[75] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer New York, New York, NY, 1996.

[76] Lyudmila Yur'evna Vostrikova. Detecting "disorder" in multidimensional random processes. In *Doklady akademii nauk*, volume 259, pages 270–274, 1981.

[77] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3(1):257–295, 6 2016.

[78] Norbert Wiener. The Quadratic Variation of a Function and its Fourier Coefficients. *Journal of Mathematics and Physics*, 3(2):72–94, 3 1924.

[79] Bohan Wu and Bradley Van Allen. A Comparison of Functional Principal Component Analysis Methods with Accelerometry Applications. 5 2021.

[80] Fang Yao. Functional principal component analysis for longitudinal and survival data. *Statistica Sinica*, 17(3):965–983, 2007.

[81] L. C. Young. An inequality of the Hölder type, connected with Stieltjes integration. *Acta Mathematica*, 67(0):251–282, 1936.

[82] L C Young. Sur une généralisation de la notion de variation de puissance p-ième bornée au sens de M. *Wiener, et sur la convergence des séries de Fourier, Comptes Rendus*, 204:470–472, 1937.

[83] Jin-Ting Zhang. *Analysis of Variance for Functional Data*. Chapman and Hall/CRC, 6 2013.

SANTRAUKA (summary in Lithuanian)

Įvadas

Sparčiai tobulėjant technologijoms, didelė pažanga buvo padaryta duomenų skaitmenizavime. Per pastaruosius dešimtmečius renkamų duomenų kiekis ir duomenų įvairovė augo eksponentiškai. Duomenys renkami praktiškai visose srityse: pradedant nuo gamtoje stebimų fizinių reiškinių, medicinoje fiziologinių duomenų, ekonominių reiškinių ar išmanių laikrodžių, kurie fiksuoja žmogaus fiziologinę būklę. Iš tiesų neliko nė vienos srities, kuri nebūtų paveikta skaitmenizavimo revoliucijos. Natūralu, kad toks spartus visų sričių skaitmenizavimas, kelia daug klausimų apie duomenų struktūrą. Viena iš problemų, su kuria dažnai susiduriama analizuojant duomenis yra struktūriniai pokyčiai laiko eilutėse. Pokyčiai gali būti neakivaizdūs ar sunkiai pastebimi, dėlto, sprendžiant šią problemą, reikalingi teoriškai pagrįsti matematiniai instrumentai.

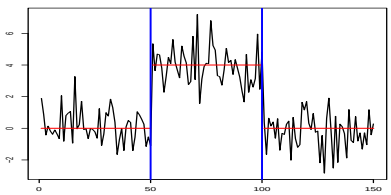
Pasikeitimo taškų analizę galima apibrėžti kaip sekos suskaidymą į segmentus, kurie pasižymi skirtingomis statistinėmis savybėmis. Taškai, kuriais laiko eilutė padalijama į vieną ar daugiau segmentų, yra vadinami pasikeitimo taškais (angl. *change-point*). Pasikeitimo taškų nustatymo uždaviniams yra keliami du pagrindiniai klausimai:

1. Ar duomenų statistinės savybės kuriuo nors laiko momentu pasikeitė?
2. Jeigu pasikeitė, kada tai įvyko?

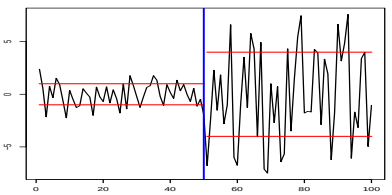
Kai kuriais atvejais procesų struktūriniai pokyčiai gali būti lengvai nustatomi, pavyzdžiui, ekonomika gali pereiti į recesiją, o vėliau sekti atsistatymas. Kriterijai, kurie apibrėžia recesijos sąvoką yra aiškiai apibrėžti. Kitose srityse pasikeitimo taškus pastebėti be matematinių metodų gali būti beveik neįmanoma. Matematiniai metodai ypač tampa aktualūs, kai sistemos automatiškai turi reaguoti į pasikeitusią situaciją, pavyzdžiui, išjungti variklį ir išvengti katastrofos, jei vibracijos lygis pasikeitė. Pasikeitimo taškų analizė sparčiai išpopuliarėjo atsiradus išmaniems įrenginiams, pavyzdžiui, išmanūs laikrodžiai realiu laiku nustato, kada žmogus pradėjo bėgti, lipti laiptais ar užmigo. Pokyčio taškų analizė yra ypač aktuali medicinoje. Naudojant medicininius matavimo prietaisus nuolatos registruojami duomenys apie ligonio fiziologinę būklę. Užfiksavus pasikeitimus svarbu kuo greičiau į tai reaguoti.

Laiku sureagavus galima išvengti neigiamų pasekmių, pavyzdžiui, epilepsijos priepuolio. Malladi et. al. [46] savo darbe pasiūlė metodą realiu laiku fiksuoti epilepsijos priepuolius. Kita medicinos sritis kurioje plačiai taikoma pokyčio taškų analizė yra širdies veiklos stebėjimas (EKG). Tam, kad nustatytume pokyčius širdies veikloje dažnai taikomi vilnelių (angl. *wavelet*) transformacijų metodai pokyčio taškams nustatyti [53, 57], Antti Koski [41] pritaikė paslėptuosius Markovo modelius (angl. *hidden Markov models*), Fotoohinasab et. al. [25] pritaikė grafų apribojimo (angl. *graph-constrained*) metodus. Plačiau apie įvairius metodus susijusius su pokyčio taškų analize EKG duomenims savo darbe pateikia Fotoohinasab et. al. [25].

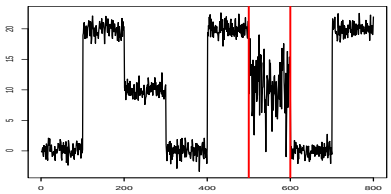
Pokyčio taškų analizė neapsiriboja vien tik laiko eilutėmis. Šis uždavinys dažnai sprendžiamas analizuojant vaizdus ir garso signalus. Atliekant magnetinį rezonansą yra kuriami 3D paveikslai (žr. [12, 47, 68]). Analizuojant garso įrašus dažnai bandoma atskirti segmentus tarp kalbos ir kitų garsu. Tai svarbu, kuriant automatinius kalbos atpažinimo modelius, aido panaikinimą (angl. *echo cancellation*), kalbos segmentavimą ir t.t. Daugelyje atvejų sprendžiant tokius uždavinius pirmi žingsniai yra suskaidyti audio signalą į skirtingus segmentus (žr. [35, 37, 73]).



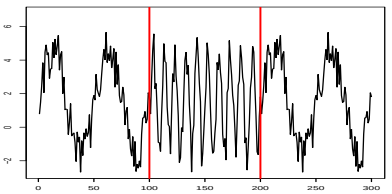
(a) Epideminis vidurkio pasikeitimas



(b) Dispersijos pasikeitimas



(c) Šablono pasikeitimas



(d) Epideminis dažnio pasikeitimas

Figure 4.1: Skirtingų pasikeitimų pavyzdžiai

Nėra vienareikšmiško apibrėžimo, kuris apibrėžtų, kas yra pokytis. Pokyčiu galime laikyti duomenis generuojančio modelio ar modelio parametrų

pasikeitimą. Dažniausiai yra nagrinėjami skirstinio parametrų tokių kaip vidurkis (4.1a pav.) ar dispersija (4.1b pav.), pasikeitimai. Daug metodų yra pasiūlyta, norint nustatyti ar pasikeitė skirstinys (žr. [55]). Verta paminėti ir kitas, mažiau dėmesio susilaukusias pasikeitimų kategorijas, kaip dažnio pasikeitimai ir šablono (angl. *pattern change*) pasikeitimai. Dažnio (4.1d pav.) pasikeitimai yra svarbūs analizuojant eilutes turinčias ciklišKumo savybių. Tokie pasikeitimai dažniausiai tiriami dažnių srityje, pavyzdžiui, naudojant Furjė ar vilnelės (angl. *wavelet*) transformaciją (žr. [53, 57]). Galiausiai, vienas iš sudėtingiausių uždavinių yra aptikti šabloninius pokyčius (4.1c pav.). Tokių aptikimo metodų aprėptis mokslinėje literatūroje nėra plati, nors, ši problema yra aktuali tokiose srityse kaip smegenų bangų analizė (žr. [72]).

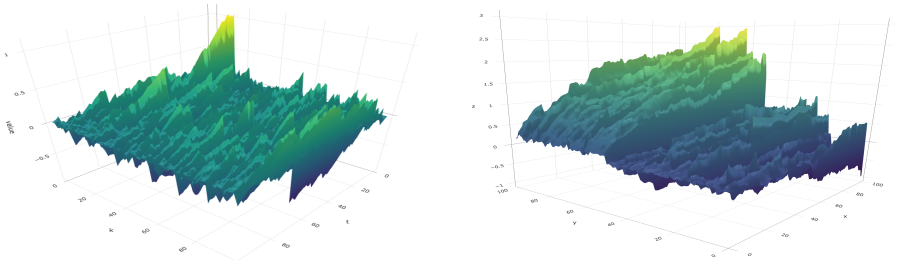
Pasikeitimo taškų problema yra skirstoma į du specifinius atvejus. Pirmuoju atveju (lot. *a-posteriori*) yra analizuojami pastovi ir nekintanti duomenų aibė (angl. *offline*). Turint visą duomenų imtį, tikslas yra atsakyti į aukščiau paminėtus klausimus. Tokių algoritmų efektyvumas vertinamas pagal tai kaip jautriai reaguojama į pasikeitimus ir pasikeitimo vietos nustatymo tikslumu. Kitas svarbus aspektas yra tai, kad turint visus duomenis, pasikeitimų skaičius gali būti daugiau nei vienas. Tada problema tampa sudėtingesnė ir sprendžiamas papildomas uždavinys siekiant nustatyti kiek iš viso egzistuoja pasikeitimų. Ypač aktuali problema yra nustatyti epideminius pasikeitimus, kai pasikeitimas yra laikinas (4.1a ir 4.1d pav.).

Antruoju atveju yra analizuojami realaus laiko duomenys (angl. *online*) kai duomenų aibė yra nuolatos auganti. Tikslas yra kuo anksčiau ir kuo tiksliau aptikti pasikeitimą jam tik įvykus. Šioje disertacijoje yra nagrinėjama problema, kai duomenų imtis yra nekintanti.

Pasikeitimo taškų problema nėra nauja ir plačiai išanalizuota klasikinėje literatūroje. Ypač daug aptikimo algoritmų yra pasiūlyta vienmatėms laiko eilutėms. Vienas iš pirmųjų algoritmų, skirtų pokyčiams nustatyti, yra kaupiamosios sumos (angl. *CUMSUM*) algoritmas (Page 1954 [50]), kuris buvo sukurtas vidurkio pokyčiui nustatyti. Algoritmas buvo taikomas kokybės gamybos kontrolei užtikrinti.

Pokyčio taškų paieškos analizė neapsiriboja tik vienmatėmis laiko eilutėmis. Pastaruoju metu didelis dėmesys yra skiriamas būtent daugiamatėms laiko eilutėms.

Funkcinių duomenų analizė sukuria natūralią aplinką daugiamačių laiko



(a) Funkcinių duomenų rinkinys be pasikeitimo

(b) Funkcinių duomenų rinkinys su pasikeitimu

Figure 4.2: Skirtingų pasikeitimų pavyzdžiai su funkciniais duomenimis

eilučių analizei. Todėl, vis daugiau klasikinių metodų yra adaptuojama darbui su tokiais duomenimis. Funkciniai duomenys yra natūralus daugiamačių duomenų apibendrinimas iš baigtinės dimensijos į begalinę. Praktikoje funkciniai duomenys gaunami stebint keletą tiriamųjų subjektų laike, erdvėje ar kitose tęstinėse srityse. Jie gali būti kreivės, paviršiai ar kiti sudėtingi objektai. Reprezentuojant daugiamačius duomenis kaip funkcijas stipriai išplečiamas analizės instrumentų spektras. Ypač funkcinės principinės komponentės yra daug informatyvesnės negu daugiamatės principinės komponentės. Populiarėjant tokių duomenų analizei vis aktualesnė tampa ir pasikeitimo taškų paieškos problema ne tarp fiksuotų taškų ar vektorių o tarp kreivių (4.2 pav.).

Pirmieji tokių duomenų analizės metodai buvo minimi jau 1950. Kai Grenanderis [29] savo publikacijoje bandė pritaikyti statistinius metodus stochastiniams procesams. Vėliau, Rao [64] 1958 metais savo analizėje lygino organizmo augimo kreives. Patį terminą funkcinė duomenų analizė pirmą kartą paminėjo Ramsey [61] 1982 metais. Šių duomenų analizė ir taikymai plačiai nagrinėjama Ramsey ir Silverman knygoje [63], taip pat išsamią apžvalgą galima rasti Wang et. al. [77] publikacijoje.

Tikslai ir uždaviniai

Pagrindinė disertacijos tema – pasikeitimo taškų aptikimas vienmatėse laiko eilutėse ir funkcinėse laiko eilutėse. Pagrindinis disertacijos tikslas pasiūlyti statistinius testus, paremtus dalinių sumų proceso variacinėmis savybėmis.

Tyrimo tikslui pasiekti keliami šie uždaviniai:

1. Apibrėžti nagrinėjamus objektus ir modelius.

2. Apibrėžti ir ištirti vidurkio nestabilumo testavimo modelį, paremtą dalinių sumų proceso p-variacija.

3. Nustatyti testo ribinį skirstinį prie nulinės ir alternatyvios hipotezių

4. Išnagrinėti procesų indeksuotų funkcijomis sumų ($\mathcal{G}$-sumų) asimptotiką.

5. Sudaryti pasikeitimo taškų statistinius testus funkciniams duomenims, remiantis $\mathcal{G}$-sumomis.

6. Sukonstruotus testus išanalizuoti imitaciniais metodais.

7. Pritaikyti testus realiems duomenims.

Mokslinis darbo naujumas

Disertacijoje pasiūlomas p-variacija paremtas statistinis testas, kuris leidžia efektyviai aptikti vidurkio nestabilumus vienmatėse laiko eilutėse. Šis testas apibendrinamas ir funkciniams duomenims. Pasiūlomi trys statistiniai testai:

- Ne daugiau kaip vieno pasikeitimo taško (angl. *at most one change point*) nustatymui.
- Ne daugiau kaip m pasikeitimo taškų nustatymui.
- Kai pasikeitimo taškų skaičius nežinomas.

Disertacijos praktinė vertė

Disertacijoje siūlomi testai, paremti p-variacija, išlieka efektyvus analizuojant tiek dideles tiek mažas duomenų aibes. Todėl gali būti plačiai taikomi įvairiose srityse, tokiose kaip medicina, vaizdų ir garso analizė, klimato kaita ir kitose. Siūlomi testai yra skirti vidurkio pokyčiui nustatyti, tačiau pačių testų konstrukcija gali būti apibendrinama ir kitiems pasikeitimų tipams nustatyti.

Darbo struktūra

Disertacija parašyta anglų kalba. Ją sudaro: įvadas, trys skyriai, bendrosios išvados bei diskusija apie tolimesnius tyrimus, priedai ir literatūros sąrašas. Pirmame disertacijos skyriuje pristatomos funkcijos variacijos savybės, kitų

autorių metodai, skirti pokyčio taškų nustatymui, bei pateikiami jų praktiniai taikymai ir svarba. Taip pat yra pateikiamos pagrindinės funkcinės duomenų analizės sąvokos ir pateikiami kitų autorių metodai, pritaikyti darbui su funkciniais duomenimis. Antrame disertacijos skyriuje yra pasiūlomas naujas metodas paremtas dalinių sumų proceso p-variacijos savybėmis vidurkio pasikeitimams nustatyti ir nagrinėjami teoriniai aspektai. Skyriaus pabaigoje imitacijomis yra įvertinama ir aprašoma pasiūlyto testo statistinė galia. Galiausiai statistinis testas yra pritaikomas realiems duomenimis. Trečiame disertacijos skyriuje yra pateikiamas pasiūlyto metodo apibendrinimas darbui su funkciniais duomenimis ir pasiūlomi trys statistiniai testai vidurkio nestabilumui patikrinti. Skyriaus pabaigoje teoriniai rezultatai patvirtinami imitacijomis ir testai patikrinami su realiais duomenimis.

Ginamieji teiginiai

- Teoriškai pagrįstas statistinis testas, paremtas dalinių sumų proceso p-variacijos savybėmis, vidurkio pasikeitimams aptikti.
- Nustatyti $\mathcal{G}$-sumų proceso ribiniai skirstiniai.
- Panaudojant $\mathcal{G}$-sumų asimptotines savybes sukonstruoti statistiniai testai tam, kad patikrinti funkcinės imties vidurkio pasikeitimo žinomam ir nežinomam taškų skaičiui

Disertacijos rezultatų aprobavimas

Disertacijos tema paskelbti du straipsniai, žurnaluose su citavimo indeksu *Clarivate Analytics Web of Knowledge* duomenų bazėje (WoS).

[A1] T. Danielius, A. Račkauskas, p-Variation of CUSUM process and testing change in the mean, *Communications in Statistics-Simulation and Computation*, 1–13 (2020). https://doi.org/10.1080/03610918.2020.1844899

[A2] T. Danielius, A. Račkauskas, Multiple change point detection in functional sample via $\mathcal{G}$-sum process, *Mathematics* 10.13, (2022). https://doi.org/10.3390/math10132294

Disertacijos keliami klausimai ir rezultatai pristatyti trijose tarptautinėse konferencijose ir vienoje konferencijoje, Lietuvoje:

[C1] T. Danielius. *Functional data analysis of neurophysiological data: case study.* NBBC19 : 7th Nordic-Baltic biometric conference, 3-5 June 2019, Vilnius, Lithuania.

[C2] T. Danielius, A. Račkauskas. *p-variation of cusum process and testing change in the mean*, 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2020) 19-21 December 2020, Virtual Conference.

[C3] T. Danielius, A. Račkauskas. *Multiple change point detection in functional sample via $\mathcal{G}$-sum process*, 63th Conference of the Lithuanian Mathematical Society, Kaunas University, 16-27 June 2022.

[C4] T. Danielius, A. Račkauskas. *Multiple change point detection in functional sample via $\mathcal{G}$-sum process*, 24th International Conference on Computational Statistics (CompStat 2022), 23-26 August 2022, Bologna, Italy.

[C5] T. Danielius. *Pasikeitimo taškų testai funkciniams duomenims paremti p-variacija*, Seminar Statistics and its applications, Vilnius University Institute of Applied Mathematics, 7 August 2022, Vilnius, Lithuania.

5. Vidurkio pasikeitimo testas: vienmatis atvejis

Antroje disertacijos dalyje yra pasiūlomas naujas statistinis testas, skirtas vidurkio pokyčiui stebėti vienmatėje laiko eilutėje. Testas remiasi p-variacijos pakitimais dalinių sumų (angl. *CUSUM*) procesuose. Turint imtį $X_1, X_2, \ldots, X_n$ ir skaičių $p > 2$, statistiką apibrėžiame kaip

$$T_{p,n}(X_1, \ldots, X_n) = \max\left\{\sum_{j=1}^{m}\left|\sum_{i=k_{j-1}+1}^{k_j}(X_i - \overline{X}_n)\right|^p : 0 = k_0 < \cdots < k_m = n; 1 \le m \le n\right\},$$

čia $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$.

Tęsiant šios statistikos analizę verta prisiminti kai kuriuos būtinus apibrėžimus. Funkcijai $f : [0,1] \to \mathbb{R}$ ir skaičiui $0 < p < \infty$, funkcijos f p-variacija intervale $[0,t]$ apibrėžiama kaip

$$v_p(f;[0,t]) := \sup\left\{\sum_{j=1}^{m}|f(t_j) - f(t_{j-1})|^p\right\} \le +\infty,$$

kai supremumas yra imamas iš visų skaidinių (angl. *partitions*)

$$0 = t_0 < t_1 < \cdots < t_m = t; \quad m = 1, 2, \ldots,$$

intervale $[0,t]$. Toliau, disertacijoje naudojamas trumpinys $v_p(f) := v_p(f;[0,1])$. Tuo atveju, kai $v_p(f) < \infty$, sakome, kad funkcijos f p-variacija yra aprėžta (angl. *bounded*) ir visų tokių funkcijų aibė yra neseparabili Banacho erdvė $\mathcal{W}_p[0,1]$ su norma

$$\|f\|_{[p]} := |f(0)| + v_p^{1/p}(f).$$

Įterpimai $\mathcal{W}_p[0,1] \hookrightarrow \mathcal{W}_q[0,1]$ yra tolydūs ir

$$v_q^{1/q}(f) \le v_p^{1/p}(f), \quad \text{kai} \;\; 1 \le p < q.$$

Nagrinėkime vienmates laiko eilutes $X_k, k = 1, 2, \ldots$. Tegul $\forall n \ge 1$ ir $\forall t \in$

$[0,1]$, tada

$$S_n(t) = 0 \text{ jei } t \in [0,1/n),\ S_n(t) := \sum_{i=1}^{\lfloor nt \rfloor} X_i,\ \text{čia } t \in [1/n, 1],$$

kai su kiekvienu realiu skaičiumi $x \geq 0$, $\lfloor x \rfloor := \max\{k : k \in \mathbb{N}, k \leq x\}, \mathbb{N} = \{0, 1, \dots\}$. Tada laipsniškas CUSUM procesas $Z_n = (Z_n(t), t \in [0,1])$ apibrėžiamas

$$Z_n(t) = S_n(t) - \frac{\lfloor nt \rfloor}{n} S_n(1) = \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \overline{X}_n).$$

Klasikiniai proceso Z_n trajektorijų erdvių pavyzdžiai yra Hilberto erdvė $L_2[0,1]$ ir Skorodo erdvė $D[0,1]$. Su įvairiomis prielaidomis šiose erdvėse galime įrodyti centrinę ribinę teoremą. Pavyzdžiui, klasikinė Donskerio teorema su nepriklausomais ir vienodai pasiskirsčiusiais atsitiktiniais dydžiais, turinčiais baigtinį antrąjį momentą, teigia, kad

$$n^{-1/2} Z_n \xrightarrow[n\to\infty]{\mathcal{D}} \sigma B \ \text{ erdvėje } \ D[0,1],$$

čia $B = (B(s) := W(s) - sW(1), s \in [0,1])$ yra standartinis Brauno tiltas.

5.1 teorema. (Su fiksuota $p > 2$ reikšme). Tarkime $X_1, X_2, \dots$ yra nepriklausomų ir vienodai pasiskirsčiusiais atsitiktinių dydžių seka ir dalinių sumų procesas $S_n = (S_n(t), t \in [0,1])$.

Jei, $\sigma^2 := EX_1^2 < \infty$, tada

$$n^{-1/2} Z_n \xrightarrow[n\to\infty]{\mathcal{D}^*} \sigma B \ \text{ erdvėje } \ \mathcal{W}_p[0,1].$$

5.2 teorema teoriškai pagrindžia, kad statistika gali būti naudojama vidurkio pasikeitimo taškams nustatyti.

5.2 teorema. Tegu $p > 2$ ir $X_1, X_2, \dots$ yra nepriklausomų ir vienodai pasiskirsčiusių atsitiktinių dydžių seka. Jei $EX_1^2 = \sigma^2 \in (0, \infty)$, tuomet:

$$n^{-1/2} \sigma^{-1} T_{p,n}^{1/p}(X_1, \dots, X_n) \xrightarrow[n\to\infty]{\mathcal{D}} v_p^{1/p}(B).$$

Nagrinėkime paprastosios atsitiktinės imties $X_1, \dots, X_n$ modelį

$$X_i = \delta \mathbf{1}_{(k^*,n]}(i) + Y_i, \quad i = 1, \dots, n,$$

kai $Y_1, \ldots, Y_n$ yra nepriklausomi ir vienodai pasiskirstę atsitiktiniai dydžiai su $E(Y_i) = 0$, $E(Y_i^2) = 1$ ir $\delta \in R, k^* \in \{1, \ldots, n\}$ yra nežinomieji parametrai.

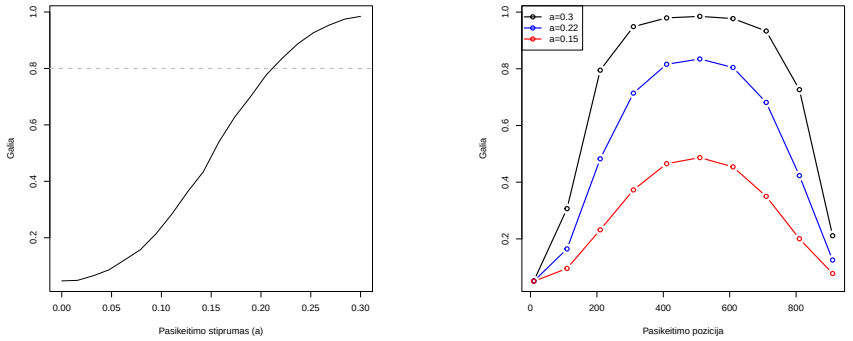
Nulinės hipotezės $H_0 : \delta = 0$, ir $\forall p > 2$, atveju turime,

$$n^{-1/2} v_p^{1/p}(Z_n) \xrightarrow[n\to\infty]{\mathcal{D}} v_p^{1/p}(B).$$

Alternatyvios hipotezės H_A atveju turime: $\delta = \delta_n \approx \sqrt{n}\delta^*$, ir $k^* = \lfloor n\theta^* \rfloor$ su kai kuriais $\delta^* > 0$ ir $\theta^* \in (0, 1)$, teisinga

$$n^{-1/2} v_p^{1/p}(Z_n) \xrightarrow[n\to\infty]{\mathcal{D}} v_p^{1/p}(B - f),$$

$$\text{kai } f(t) = \begin{cases} \delta^* t(1-\theta^*) & \text{kai } \ 0 \le t < \theta^* \\ \delta^* \theta^* (1-t) & \text{kai } \ \theta^* \le t \le 1 \end{cases}.$$



(a) Statistinė galia atlikus imitacijas, keičiant pasikeitimo stiprumą.

(b) Statistinė galia atlikus imitacijas, keičiant pasikeitimo poziciją.

Figure 5.1: Statistinės galios vertinimas

Statistinio testo efektyvumas buvo įvertintas imitaciniais metodais. Nagrinėkime nepriklausomą atsitiktinę imtį $X_1, \ldots, X_n$, kai X_i apibrėžiama kaip

$$X_i \sim \begin{cases} \mathcal{N}(\mu_1, \sigma^2) & \text{kai } i \le n\tau \\ \mathcal{N}(\mu_2, \sigma^2) & \text{kai } i > n\tau \end{cases}.$$

Atlikus 10000 imitacijų buvo suskaičiuota statistinė galia. Nulinė hipotezė

atmetama, kai

$$P\left(v_p^{1/p}(Z_n) > a_\alpha\right) = \alpha$$

čia $a_\alpha = 2.024352$ yra atsitiktinio dydžio $v_p^{1/p}(B)$ skirstinio $\alpha = 0.95$ kvantilis. Rezultatai (5.1a pav.) rodo, kad kai $a >= 0.22$, testo statistinė galia pasiekia 80% tikslumą.

Kitu atveju buvo nagrinėjama statistinė galia, keičiant pasikeitimo taško k poziciją, prie skirtingų $a = 0.15, 0.22, 0.3$ reikšmių. Rezultatai pavaizduoti 5.1b paveiksle.

6. Vidurkio pasikeitimo testai: funkcinis atvejis

Disertacijos trečiame skyriuje apibendrinami antro skyriaus rezultatai ir testas adaptuotas funkcinių duomenų analizei. Šiame skyriuje yra apibrėžiamas $\mathcal{G}$-CUSUM procesas ir nagrinėjami šio proceso teoriniai aspektai, įskaitant asimptotinį elgesį. Parenkant skirtingas aibes $\mathcal{G}$, pasiūlomi trys statistiniai testai pasikeitimo taškams aptikti:

1. Testas $T_{n,1}$ skirtas vieno pasikeitimo taškui nustatyti.
2. Testas $T_{n,m}$ skirtas ne daugiau kaip m pasikeitimo taškams nustatyti.
3. Testas T_n nežinomam pasikeitimo taškų skaičiui nustatyti.

Dauguma autorių didelį dėmesį skiria vieno pasikeitimo taško nustatymo problemai nagrinėti. Berkes et al. [11] pasiūlė CUSUM testą funkciniams duomenims, naudojant imties projekcijas į kai kurias pagrindines kovariacijos γ komponentes. Vėliau šią problemą ir asimptotines savybes nagrinėjo Aue [3] ir kiti. Aston ir Kirch [2] išplėtė šį testą silpnai priklausomiems funkciniams duomenims ir epideminiams pasikeitimams.

Nagrinėkime antros eilės stacionarią stochastinių procesų $Y_i = (Y_i(t), t \in [0,1]), i \in \mathbb{N}$ seką, apibrėžtą tikimybinėje erdvėje $(\Omega, \mathcal{F}, P)$, su nuliniu vidurkiu ir kovariacijos funkcija $\gamma = \{\gamma(s,t), s,t \in [0,1]\}$. Funkcinių duomenų imčiai $X_1(t), \ldots, X_n(t), t \in [0,1]$, nagrinėjamas modelis:

$$X_k(t) = g(k/n, t) + Y_k(t), \quad t \in [0,1], \quad k = 1, \ldots, n, \tag{6.1}$$

kai funkcija $g : [0,1] \times [0,1] \to \mathbb{R}$ yra deterministinė, bet nestebima. Pagrindinis tikslas yra, remiantis funkcijos variacijos savybėmis, sukurti testus tokiai nulinei hipotezei patikrinti prieš alternatyvą:

$$H_0 : \quad g = 0 \quad \text{prieš} \quad H_1 : g \neq 0$$

Disertacijoje didžiausias dėmesys yra skiriamas laiptinėms funkcijoms pirmojo argumento atžvilgiu, kas apima pasikeitimo taško aptikimo uždavinį.

Šis modelis neapsiriboja tik pasikeitimo uždavinio sprendimu, bet gali būti taikomas ir kitiems klausimams analizuoti, pvz.: vaizdų analizei, procesų seg-

mentavimui. Taip pat, analizuojant funkcines laiko eilutes, toks modelis gali būti taikomas nustatant tendencijas.

Šioje disertacijoje siūloma metodologija yra paremta proceso

$$W_n(s) = \sum_{k=1}^{\lfloor ns \rfloor} (X_k - \overline{X}_n) + (ns - \lfloor ns \rfloor)(X_{\lfloor ns \rfloor + 1} - \overline{X}_n), \quad s \in [0,1],$$

tam tikromis variacijos savybėmis. Čia $\overline{X}_n = n^{-1}(X_1 + \cdots + X_n)$.

Kiekvienas elementas X_k yra funkcija, todėl tokie procesai turi begalinę dimensiją, ir analizė tampa sudėtinga. Tam, kad pereitume iš begalinės dimensijos į baigtinę dimensiją yra naudojamos įvairios projektavimo technikos.

6.1 prielaida. $Y, Y_1, Y_2, \ldots$ yra nepriklausomi ir vienodai pasiskirstę atsitiktiniai procesai.

1. Kiekvienas procesas Y_i yra tolydus kvadratinio vidurkio prasme,
2. kiekvienas procesas Y_i yra matus abiejų argumentų atžvilgiu,
3. proceso γ kovariacinė funkcija turi baigtinį pėdsaką: $\mathrm{tr}(\gamma) = \int_0^1 \gamma(t,t)\,\mathrm{d}t < \infty$.

Kai 6.1 prielaida yra tenkinama, Y_i galime laikyti atsitiktiniu elementu su reikšmėmis erdvėje $L_2(0,1)$. $L_2 := L_2(0,1)$ yra Lebego prasme kvadratu integruojamų funkcijų, apibrėžtų intervale $[0,1]$ Hilberto erdvė su skaliarine daugyba $\langle f, g\rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t$ ir atstumo funkcija $\rho(f,g) = \|f - g\|$.

Nagrinėkime funkcijų klases $\mathcal{F}, \Phi \subset L_2$ ir $\mathcal{G}(= \mathcal{F} \times \Phi)$-sumų procesą

$$\nu_n = \Big(\sum_{k=1}^{n} \nu_{nk}(f,\psi), f \in \mathcal{F}, \psi \in \Psi \Big),$$

kai $\nu_{nk}(f,\psi) = \langle X_k, \psi\rangle \lambda_{nk}(f)$, λ_{nk} yra tolygioji tikimybė intervale $[(k-1)/n, k/n]$ ir $\lambda_{nk}(f) = \int_0^1 f(t)\,\mathrm{d}\lambda_{nk}(t)$. Šiuo žingsniu X_i yra projektuojama į pasirinktą kryptį $\psi \in \Psi$ Hilberto erdvėje. Natūrali aplinka, tokių stochastinių ν_n procesų analizei yra erdvė $\ell^\infty(\mathcal{G})$, kai $\mathcal{G} = \mathcal{F} \times \Phi$. $\ell^\infty(\mathcal{G})$ yra Banacho erdvė visų aprėžtų realiųjų funkcijų, apibrėžtų aibėje $\mathcal{G}$ su tolygia norma

$$\|\mu\|_{\mathcal{G}} := \sup\{|\mu(g)| : g \in \mathcal{G}\}, \mu \in \ell^\infty(\mathcal{G}).$$

Nagrinėjant $\mathcal{G}$-sumų procesus apibrėžiamas ribinis Gauso procesas $\nu_\gamma = (\nu(f,\psi), f \in \mathcal{F}, \psi \in \Psi)$ su kovariacija

$$E\nu_\gamma(f,\psi)\nu_\gamma(f',\psi') = \mathcal{K}_\gamma((f,\psi),(f',\psi,)) := \langle \Gamma\psi, \psi'\rangle\langle f, f'\rangle, \quad \psi, \psi', f, f' \in L_2,$$

kai $\Gamma : L_2 \to L_2$ yra kovariacijos operatorius γ branduoliui.

(6.1) modeliui, nagrinėjama nulinė hipotezė $H_0 : \quad g = 0$ ir dvi galimos alternatyvos

$$H_A : \quad g = g_n = u_n q_n, \quad \text{kai} \quad u_n \to u \quad \text{iš} \quad \mathcal{W}_2[0,1], \quad \sqrt{n} q_n \to q \quad \text{ervėje} \quad L_2,$$

ir

$$H'_A : \quad g = g_n = u_n q_n, \quad \text{kai} \quad u_n \to u \quad \text{iš} \quad \mathcal{W}_2[0,1], \quad \sqrt{n} \sup_{\psi\in\Psi} |\langle q_n, \psi\rangle| \to \infty.$$

Abiejose alternatyvose funkcija u_n aprašo pasikeitimo taškų konfigūraciją, o funkcija q_n įvertina dreifo dydį.

6.1 teorema. Tarkime, kad atsitiktiniai procesai $(X_k), k > 1$ yra apibrėžti (6.1) modeliu, kai $Y, Y_1, Y_2, \dots$ tenkina 6.1 prielaidą.

Tarkime, kad aibė $1 \le q < 2$, ir $\mathcal{F} \subset \mathcal{W}_q[0,1]$ yra aprėžta, o aibė $\Psi \subset L_2$ tenkina entropinę sąlygą

$$\int_0^1 \sqrt{\log N(\varepsilon, \Psi, \rho)}\, \mathrm{d}\varepsilon < \infty. \tag{6.2}$$

Tuomet egzistuoja tokia Gauso proceso ν_γ modifikacija erdvėje $L_2 \times L_2$, kad jo susiaurinimas $\mathcal{F} \times \Psi$, $(\nu_\gamma(f,\psi), f \in \mathcal{F}, \psi \in \Psi)$ yra tolydus procesas ir galioja tokie ribiniai teiginiai:

(1a) Prie H_0:

$$n^{-1/2}\nu_n \xrightarrow[n\to\infty]{\mathcal{D}} \nu_\gamma \quad \text{erdvėje} \quad \ell^\infty(\mathcal{F}\times\Psi). \tag{6.3}$$

(1b) Prie H_A,

$$n^{-1/2}\nu_n \xrightarrow[n\to\infty]{\mathcal{D}} \nu_\gamma + \Delta, \quad \text{erdvėje} \quad \ell^\infty(\mathcal{F}\times\Psi), \tag{6.4}$$

kai

$$\Delta(f,\psi) = \langle u, f\rangle\langle q, \psi\rangle.$$

Jei $u(s) = 1, s \in [0,1]$, tai alternatyvi H_A hipotezė rodo signalo egzistavimą triukšme. Šiuo atveju $\Delta(f,\psi) = \lambda(f)\langle q,\psi\rangle$. Todėl šios teoremos naudojimas signalui tirti triukšme yra prasmingas, jei $\langle q,\psi\rangle \neq 0$.

6.2 prielaida. Egzistuoja toks $d > 1$, kad tikrinės reikšmės λ_r tenkina sąlygą

$$\lambda_1 > \lambda_2 > \cdots > \lambda_d > \lambda_{d+1} \geq 0.$$

Statistinėje analizėje Γ tikrinės reikšmės ir tikrinės funkcijos yra pakeičiamos įvertintomis reikšmėmis. Atsižvelgiant į tai, kad kiekvienam k,

$$E[(X_k - E(X_k)) \otimes (X_k - E(X_k))] = \Gamma,$$

Γ galime įvertinti

$$\widehat{\Gamma}_n := \frac{1}{n}\sum_{i=1}^{n}[(X_i - \overline{X}_n) \otimes (X_i - \overline{X}_n)],$$

čia $\overline{X}_n(s) = n^{-1}(X_1(s) + \cdots + X_n(s))$.

Toliau $\widehat{\Gamma}$ tikrines reikšmes ir tikrines funkcijas atitinkamai žymėsime $\widehat{\lambda}_{nr}$ ir $\widehat{\psi}_{nr}$, $r = 1,\ldots,n-1$. Siekdami užtikrinti, kad $\widehat{\psi}_{nr}$ galėtų būti vertinamas kaip ψ_r, o ne $-\psi_r$ įvertis, toliau darysime prielaidą, kad ženklai yra tokie, kad $\langle\widehat{\psi}_{nr}, \psi_r\rangle \geq 0$. Pabrėždami, kad

$$\widehat{\Gamma}\widehat{\psi}_{nr} = \widehat{\lambda}_{nr}\widehat{\psi}_{nr}, \quad r = 1,\ldots,n-1, \tag{6.5}$$

ir

$$\widehat{\lambda}_{nr} = \frac{1}{n-1}\sum_{i=1}^{n}\langle X_i - \overline{X}_n, \widehat{\psi}_{nr}\rangle^2, \quad r = 1,\ldots,n. \tag{6.6}$$

Su $d > 0$, apibrėžiame

$$\widehat{T}_{n,1}(d) := \max_{1\leq j\leq d}\frac{1}{\sqrt{\widehat{\lambda}_j}}\max_{1\leq k\leq n}\Big|\sum_{i=1}^{k}\langle X_i - \overline{X}_n, \widehat{\psi}_j\rangle\Big|. \tag{6.7}$$

Ši kriterijaus $\widehat{T}_{n,1}$ statistika skirta daugiausia vieno pasikeitimo taško alternatyvai patikrinti. Jos ribinis pasiskirstymas nustatomas pagal šią teoremą.

6.2 teorema. Tegul atsitiktinė funkcinių duomenų (X_k) imtis apibrėžta 6.1 modeliu, kur $Y, Y_1, Y_2, \dots$ tenkina 6.1 ir 6.2 sąlygas. Tada:

(a) Jei H_0 teisinga, tai

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)|,$$

čia $B_1, \dots, B_d$ yra nepriklausomi standartiniai Brouno tilto procesai;

(b) Esant H_A, jei $E\|Y\|^4 < \infty$, tuomet

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t) + \Delta(t)\langle q, \widehat{\psi}_k/\sqrt{\widehat{\lambda}_k}\rangle|,$$

čia $\Delta(t) = \int_0^t u(s)\,\mathrm{d}s - t\int_0^1 u(s)\,\mathrm{d}s, \quad t \in [0,1]$.

(c) Kai H_A' teisinga ir jei $E\|Y\|^4 < \infty$, tai

$$n^{-1/2}\widehat{T}_{n,1}(d) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Remiantis šia teorema, galima klasikiniu būdu sukonstruoti testavimo procedūrą. Duotam $\alpha \in (0,1)$, $C_\alpha > 0$ tokiam, kad

$$P(\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)| > C_\alpha) = \alpha.$$

Pagal 6.2 teoremą, testas:

$$\widehat{T}_{n,1}(d) \ge \sqrt{n}C_\alpha \tag{6.8}$$

turi asimptotinį lygmenį α.

Pabrėžiame, kad dėl Brauno tiltų $B_k, k = 1, \dots, d$ nepriklausomumo,

$$1-\alpha = P(\sup_{1\le k\le d}\sup_{0\le t\le 1}|B_k(t)| \le C_\alpha) = P^d(\sup_{0\le t\le 1}|B_1(t)| \le C_\alpha).$$

Tuomet

$$P(\sup_{0\le t\le 1}|B_1(t)| \le C_\alpha) = (1-\alpha)^{1/d}.$$

Taigi, C_α yra $\sup_{0\le t\le 1}|B_1(t)|$ skirstinio $(1-\alpha)^{1/d}$ kvantilis. Tai supaprastina kritinių reikšmių C_α skaičiavimus.

Jei $s^* \in (0,1)$ yra toks, kad $u(s) = \mathbf{1}_{[0,s^*]}(s), s \in [0,1]$, tada turime modelį vienam pasikeitimui nustatyti:

$$X_k(t) = \mathbf{1}_{[0,s^*]}(k/n)q_n(t) + Y_k(t), \quad t \in [0,1].$$

Šiuo atveju, $\Delta(t) = \Delta^*(t) := \min\{t, s^*\} - ts^*, \quad t \in [0,1]$.

Akivaizdu, kad kriterijaus $\widehat{T}_{n,1}(d)$ statistika artėja į begalybę, kai $d \to \infty$. Kita vertus, su didesniu d, X_j aproksimacija pagal seką $\sum_{k=1}^{d}\langle X, \widehat{\psi}_j\rangle\widehat{\psi}_j$ yra tikslesnė ir užtikrina didesnę statistinę testo galią . Pagal žemiau suformuluotą teoremą nustatome asimptotinį skirstinį kriterijui $\widehat{T}_{n,1}(d)$, kai $d \to \infty$.

6.3 teorema. Tegul atsitiktinė funkcinė imtis (X_k) apibrėžta pagal (6.1) modelį kai $Y, Y_1, Y_2, \ldots$ tenkina 6.1 prielaidas. Tada, esant H_0,

$$\lim_{d\to\infty}\lim_{n\to\infty} P\Big(n^{-1/2}\widehat{T}_{n,1}(d) \le \frac{x}{a_d} + b_d\Big) = \exp\{-e^{-x}\}, \quad x \ge 0, \tag{6.9}$$

čia

$$a_d = (8\ln d)^{1/2}, \quad b_d = \frac{1}{4}a_d + \frac{\ln\ln d}{a_d}. \tag{6.10}$$

Kai d reikšmė yra pakankamai didelė, tada (6.8) tampa

$$\widehat{T}_{n,1} \ge \sqrt{n}\Big[\frac{1}{a_d}\ln\Big(\frac{1}{\ln(1/\alpha)}\Big) + b_d\Big] \tag{6.11}$$

su asimptotiniu lygmeniu α, kai n ir d artėja prie begalybės.

Tuo atveju , kai $m > 1$, gauname statistiką

$$\widehat{T}_{n,m}(d,p) := \max_{1\le j\le d}\frac{1}{\sqrt{\lambda_j}}\max_{\kappa\in\mathcal{N}_m}\Big\{\sum_{i=1}^{m}\Big|\sum_{k=k_{i-1}+1}^{k_i}\langle X_k - \overline{X}_n, \widehat{\psi}_j\rangle\Big|^p\Big\}^{1/p}. \tag{6.12}$$

Tokio kriterijaus $\widehat{T}_{n,m}(d,p)$ statistikos konstrukcija mums leidžia tikrinti imtį, kurioje gali būti ne daugiau kaip m pasikeitimo taškų.

6.4 teorema. Tegul funkcinių duomenų imtis $(X_k, k = 1,\ldots,n)$ apibrėžta pagal (6.1) modelį kai $Y, Y_1, Y_2, \ldots$ tenkina 6.1 ir 6.2 savybes. Tada:

(a) Jei hipotezė H_0 yra teisinga, tai

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j),$$

čia $B_1, \ldots, B_d$ yra nepriklausomi standartiniai Brauno tiltai.

(b) Jei alternatyvi hipotezė H_A yra teisinga, tai

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_{m,p}^{1/p}(B_j + \Delta\langle q, \widehat{\psi}_j/\sqrt{\widehat{\lambda}_j}\rangle),$$

(c) Jei alternatyvi hipotezė H_A' yra teisinga, tai

$$n^{-1/2}\widehat{T}_{n,m}(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Pagal 6.4 teoremą, testas:

$$\widehat{T}_{n,m}(d,p) \ge \sqrt{n} C_\alpha(m,d,p) \tag{6.13}$$

atitinkamai, turi asimptotinį lygmenį α, jeigu $C_\alpha(m,d,p)$ yra

$$P(v_{m,p}^{1/p}(B) \le C_\alpha(m,d,p)) = (1-\alpha)^{1/d}.$$

Toliau nagrinėjame atvejį, kai funkcinių duomenų imtyje pasikeitimų skaičius yra nežinomas.

Parenkant $d > 0$ ir $p > 2$, apibrėžiame

$$\widehat{T}_n(d,p) := \max_{1\le j\le d} \frac{1}{\sqrt{\widehat{\lambda}_{nj}}} \max_{1\le m\le n} \max_{\kappa\in\mathcal{N}_m} \Big\{ \sum_{i=1}^{m} \Big| \sum_{k=k_{i-1}+1}^{k_i} \langle X_k - \overline{X}_n, \widehat{\psi}_{nj}\rangle \Big|^p \Big\}^{1/p}.$$

6.5 teorema. Tegul atsitiktinė imtis (X_i) yra tokia, kaip apibrėžta 6.1 teoremoje. Tada:

(a) Jei H_0 teisinga, tai

$$n^{-1/2}\widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_p^{1/p}(B_j),$$

čia $B_1, \ldots, B_d$ yra nepriklausomi standartiniai Brauno tiltai.

(b) Jei H_A teisinga, tai

$$n^{-1/2}\widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathcal{D}} \max_{1\le j\le d} v_p^{1/p}(B_j + \Delta\langle q, \widehat{\psi}_j/\sqrt{\widehat{\lambda}_j}\rangle),$$

čia $\Delta(t), t \in [0,1]$ apibrėžta 6.1 teoremoje.

(c) Jei H'_A teisinga, tai

$$n^{-1/2}\widehat{T}_n(d,p) \xrightarrow[n\to\infty]{\mathrm{P}} \infty.$$

Pagal 6.5 teoremą, testas:

$$\widehat{T}_n(d,p) \geq \sqrt{n}C_\alpha(d,p) \tag{6.14}$$

turi asimptotinį lygmenį α, su $C_\alpha(d,p)$ jei

$$P(v_p^{1/p}(B) \leq C_\alpha(d,p)) = (1-\alpha)^{1/d}.$$

Tam, kad įvertintume statistinę testo galią, atlikome Monte Karlo imitacijas. Aktualūs buvo trys pagrindiniai scenarijai:

(S1) Tegu imtis (ξ_{jk}) yra sudaryta iš nepriklausomų ir vienodai pasiskirsčiusių simetrinių Pareto atsitiktinių dydžių su indeksu p (naudojame $p = 5$).

$$Y_j(t) = \sum_{k=1}^{d} \xi_{jk} \frac{\sqrt{2}\cos(k\pi t)}{k\sigma}, \quad t \in [0,1],\ j \geq 1, \tag{6.15}$$

čia $\sigma^2 = E\xi_{11}^2$. Kai nulinė hipotezė teisinga, naudojame $X_k = Y_k, k = 1, 2, \ldots, n$.

Kai alternatyvi hipotezė teisinga laikome, kad

$$X_j(t) = u_n(j/n)\sum_{k=1}^{d} a_{nk}\cos(k\pi t) + Y_j, \quad t \in [0,1], j = 1, \ldots, n,$$

čia funkcija u_n nusako pasikeitimo taškų konfigūraciją ir koeficientai (a_{nk}) yra laisvai pasirenkami.

(S2) Sakykime $(x_{ij}, j = 0, 1, \ldots, M)$, $i = 1, \ldots, n$ yra diskretūs stebėjimai, įvertinti taškuose $x_{ij} = X_i(\tau_j)$, kur atsitiktinė imtis $(X_j, j = 1, \ldots, n)$ generuojama taip pat, kaip ir (S1) scenarijuje. Diskretūs stebėjimai yra paverčiami į funkcinius duomenis $(X_j, j = 1, \ldots, n)$, naudojant B-spline bazes.

(S3) Imkime diskrečius stebėjimus $(i/M, y_{ij}), i = 0, 1, \ldots, M, j = 1, \ldots, n,$

kurie yra generuojami

$$y_{ij} = M^{-1/2} \sum_{k=1}^{i} \xi_{kj},$$

taip, kad y_{ij} galėtumėm interpretuoti, kaip standartinį Vynerio procesą ties i/M. Iš imties $(y_{ij}, i = 1, \ldots, M)$, funkcijos Y_j yra sukonstruojamos naudojant B-spline bazes. Imitacijos metu naudojame $M = 1000$ stebėjimų ir $D = 50$ bazinių funkcijų, taip sukonstruodami $Y_1, \ldots, Y_n$, funkcijų $(n = 500)$. Tada, apibrėžiame

$$X_j = \begin{cases} Y_j, & \text{prie nulinės hipotezės} \\ u_n(j/n)q_n + Y_j, & \text{prie alternatyvios hipotezės} \end{cases}$$

čia $j = 1, \ldots, n$; naudojame skirtingas pasikeitimo taškų konfigūracijas u_n su skirtingu dreifu, kurį nusako funkcija $q_n(t) = a_n\sqrt{Mt}$, $t \in [0, 1]$.

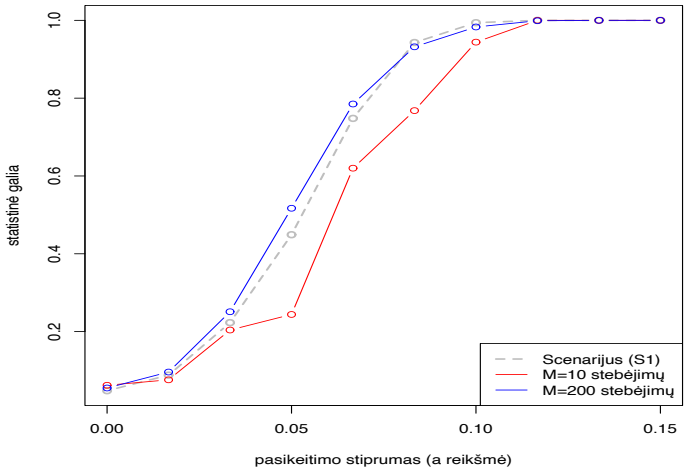


Figure 6.1: Statistinė galia atlikus imitacijas pagal (S1, S2) scenarijus

(S1) Scenarijus sukuria optimalią situaciją, kai tikrinės reikšmės ir tikrinės funkcijos yra žinomos. Šiuo atveju yra išvengiama duomenų praradimo ar matavimo klaidų. Antrasis (S2) scenarijus yra pirmo scenarijaus tęsinys, kur naudojama ta pati funkcinė duomenų imtis, tačiau atliekami papildomi žings-

niai: funkcijos įvertinamos atsitiktiniuose taškuose, diskrečioji imtis yra glodinama naudojant *b-spline* bazes, įvertinamos tikrinės reikšmės ir funkcijos. Galiausia, skaičiuojama statistika su įverčiais. Taip sukuriama imtis su matavimo paklaidos tikimybe.

Imitacijų rezultatai parodė, kad net ir po rekonstruotos atsitiktinės funkcinės imties, statistinio testo galia išlieka tokia pat tuo atveju, kai diskrečių stebėjimų skaičius pakankamai didelis ($n = 200$). Statistinės galios susilpnėjimas stebimas stipriai sumažinus diskrečių stebėjimų skaičių ($n = 10$). Rezultatai pavaizduoti 6.1 paveikslėlyje.

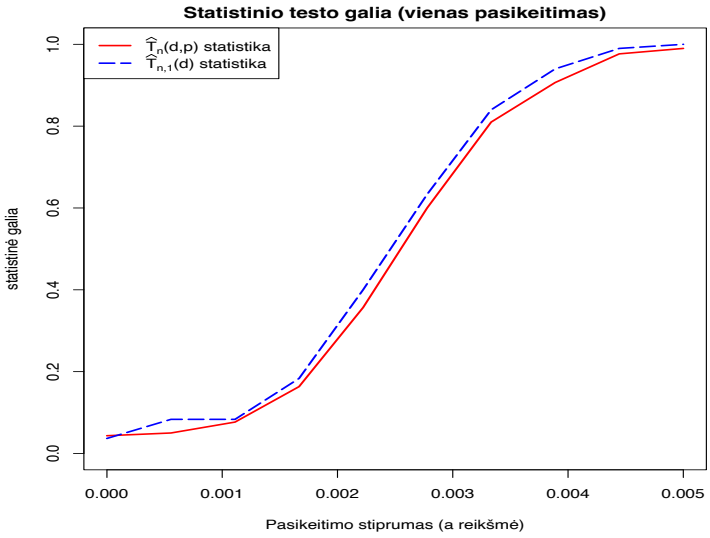


Figure 6.2: Statistinė galia atlikus imitacijas pagal (S3) scenarijų

($S3$) Scenarijaus metu buvo konstruojama imtis iš diskrečių stebėjimų, kai apie funkcinę duomenų imtį nieko nėra žinoma (išskyrus tai, kad egzistuoja epideminis pasikeitimas), kas atspindi realaus gyvenimo atvejį. Imitacijos buvo atliktos norint patikrinti (6.14) ir (6.13) statistinių testų galias. Rezultatai pavaizduoti 6.2 paveiksle.

Išvados

Šioje disertacijoje buvo apibrėžti nagrinėjami objektai ir modeliai vienmačiams ir funkciniams duomenims. Apibrėžtas vidurkio nestabilumo testavimo modelis, paremtas dalinių sumų proceso p-variacija. Testo statistinė galia buvo išanalizuota imitaciniais metodais ir teoriškai ištirtas testo suderinamumas. Taip pat buvo teoriškai nustatyti ribiniai skirstiniai prie nulinės ir alternatyvios hipotezių. Šie rezultatai buvo apibendrinami ir pritaikyti funkciniams duomenims. Buvo išnagrinėta procesų indeksuotų funkcijomis sumų ($\mathcal{G}$-sumų) asimptotika. Teoriškai buvo nustatyti ribiniai skirstiniai ir apibrėžti testai, tinkantys nustatyi vienam pasikeitimui, m pasikeitimams ir nežinomam skaičiui pasikeitimų. Testai, pritaikyti funkciniams duomenims, buvo išanalizuoti imitaciniais metodais.

Su pasikeitimo taškų problema susiduriama daugelyje sričių, kaip medicina, ekonomika, klimato kaita, vaizdų ir garsų analizė. Disertacijoje buvo pademonstruota, kaip pasiūlyti testai gali būti pritaikomi klimato kaitos duomenims ir fiziologiniams duomenims. Neapsiribojant tik šiomis sritimis, pasiūlyti testai gali būti plačiai taikomi ir kitur.

Šie testai gali būti plėtojami toliau, pavyzdžiui, adaptuojami dažnio pasikeitimams nustatyti, naudojant tolydžios vilnelės transformacijas (angl. *Continuous wavelet transform*). Tolydžios vilnelių transformacijos metodai leidžia atlikti nestacionarių signalų spektrinę analizę ir įvertinti spektro pokyčius laikui bėgant. Tokiu būdu sukonstruotiems funkciniams duomenims galime taikyti vieno ar daugiau pasikeitimų taškų nustatymo testus, pasiūlytus šioje disertacijoje.

Acknowledgements

I extend my sincerest gratitude to my supervisor, Professor Alfredas Račkauskas, for his unwavering support and guidance throughout my doctoral journey. His valuable insights, encouragement, and mentorship were instrumental in the successful completion of this research.

I am also deeply grateful to my wife, Milda, and children, Kostas and Fausta, as well as my friends, for their support and encouragement throughout this process. Their love and support has been invaluable and without it, this accomplishment would not have been possible.

Tadas Danielius

Vilnius

January 2023

Curriculum vitae

Name: Tadas

Surname: Danielius

Date of birth: 1982-03-27

Place of birth: Vilnius, Lithuania

E-mail: tadas.danielius@mif.vu.lt

Education:

2002–2006	Vilnius University, Faculty of Mathematics and informatics *Bachelor degree in Computer Science*
2016–2018	Vilnius University, Faculty of Mathematics and informatics *Master degree in Statistics*
2018–2022	Vilnius University, Faculty of Mathematics and informatics, *Ph.D. studies in Mathematics*

ERASMUS student exchange program:

2003–2004	Roskilde Computer Science college.

Work experience:

2021 - to date	Lead Data Scientist, chapter lead, IBM Lithuania.
2020 - 2021	Junior researcher at Vilnius University.
2020 - 2021	Senior Data Scientist, Euromonitor International
2020 - 2021	Senior System Architect, Lietuvos Bankas.
2020 - 2020	Principal Data Scientist, OAG Aviation.
2016 - 2020	Data Scientist, Tech. Lead, IBM Lithuania.
2008 - 2016	Software developer / System Architect, IBM UK.

Teaching experience:

2020 - 2022	Lecturer, Vilnius University

Publication by the author

On the dissertation topic

[A1] T. Danielius, A. Račkauskas, p-Variation of CUSUM process and testing change in the mean, *Communications in Statistics-Simulation and Computation*, 1–13 (2020). https://doi.org/10.1080/03610918.2020.1844899

[A2] T. Danielius, A. Račkauskas, Multiple change point detection in functional sample via $\mathcal{G}$-sum process, *Mathematics* 10.13, (2022). https://doi.org/10.3390/math10132294

Other publications

[A3] I. Poceviciute, R. Buisas, T. Danielius, R. Dulinskas, O. Rukšėnas, V. Vengelienė, The Anticonvulsant Lamotrigine Reduces Bout-Like Alcohol Drinking in Rats, *Alcohol and Alcoholism* 57.2, 242–245 (2021). https://doi.org/10.1093/alcalc/agab073

[A4] Ž. Chomanskis, V. Jonkus, T. Danielius, T. Paulauskas, M. Orvydaitė, K. Melaika, O. Rukšėnas, V. Hendrixson, S. Ročka, Remotely Programmable Deep Brain Stimulator Combined with an Invasive Blood Pressure Monitoring System for a Non-Tethered Rat Model in Hypertension Research, *Brain Sciences*, 2023; 13(3):504. https://doi.org/10.3390/brainsci13030504

NOTES

NOTES

Vilnius University Press
Saulėtekio al. 9, LT-10222 Vilnius
e-mail: info@leidykla.vu.lt,
www.leidykla.vu.lt
Print run copies 30