INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

UMI[®]
800-521-0600

Multiple Change-points with an Application to Financial Modeling

by

Markus R. Orasch, Dipl.-Ing.

A thesis submitted to
the Faculty of Graduate Studies
in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

School of
Mathematics and Statistics
Carleton University
Ottawa, Ontario, Canada
April 1999

© Copyright
1999 - Markus R. Orasch

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-42804-4

Canada

The undersigned hereby recommend to the Faculty of
Graduate Studies acceptance of the thesis

**Multiple Change-points with an
Application to Financial Modeling**

submitted by

Markus R. Orasch, Dipl.-Ing.

in partial fulfillment of the requirements for
the degree of *Doctor of Philosophy*

Kenneth S. Williams

Director, School of Mathematics and Statistics

Miklós Csörgő

Thesis Supervisor

Lajos Horváth

External Examiner

Carleton University

April 1999

“Cogito ergo sum.”

— *René Descartes*

Abstract

This dissertation deals with asymptotic methods in probability and statistics which may, for example, be used in business and finance. In particular, we study U-statistics based processes which may be used to detect (multiple or structural) changes in the distribution of independent observations.

In Chapter 1 we give an overview of well-known results which will be used throughout this thesis. Chapters 2 - 5 deal with change-point analysis.

Chapter 2 summarizes some basic properties of U-statistics and their use in the context of change-point analysis.

In Chapter 3 we review and study U-statistics based processes under the null hypothesis of no-change as well as under the alternative hypothesis of at most one change in the distribution.

Chapter 4 deals with U-statistics based processes when the alternative allows at most two changes in their distribution. As special cases, epidemic alternatives are investigated as well.

Using similar arguments as in Chapter 3 and Chapter 4, we extend our theory of at most one and two changes to multiple change-points in Chapter 5.

Besides the mentioned general theory, in Chapters 3 - 5 we study in particular changes in the mean and the variance, respectively. In addition to that we study multiple changes in the mean and/or variance and give an application in Chapter 6, which deals with changes of volatility in the financial stock market, using the Black-Scholes set-up.

The many repetitions in describing the hypotheses considered are done on purpose for the convenience of the reader. We prefer to give as much details as possible rather than to send the reader back to previous chapters.

In Chapter 1, Chapter 2 and Chapter 3 we summarize well-known results. In addition, we discuss in Chapter 3 some results which are believed to be new. Results in Chapters 4 and 5 as well as its application in Chapter 6 are also believed to be new. Theorems, corollaries and lemmas due to other authors will be accompanied by their names in brackets.

Acknowledgements

Before thanking all those people who guided and helped me over the past years and who were a very important part of my life I would like to quote a part of a story by Charles Krauthammer¹ about the extraordinary mathematician Pál Erdős:

A few years ago, Ron Graham (a friend and benefactor of Erdős) tells me, Erdős heard of a promising young mathematician who wanted to go to Harvard but was short the money needed. Erdős arranged to see him and lent him \$1,000. (The sum total of the money Erdős carried around at any one time was about \$30.) He told the young man he could pay it back when he was able to. Recently, the young man called Graham to say that he had gone through Harvard and was now teaching at Michigan and could finally pay the money back. What should he do?

Graham consulted Erdős. Erdős said, "Tell him to do with \$1,000 what I did."

The reason why I tell this story is that a similar thing happened to me. If it were not for all the help of my Professors Miklós Csörgő (Carleton University) and Pál Révész (Technical University of Vienna) it would not have been possible for me to study in Canada. I would like to thank them for the opportunity they have given me and I hope one day I will be able to do the same for someone else.

I would like to thank Professor Miklós Csörgő for all the discussions and meetings we had, for all his suggestions, his patience, perseverance and financial support. I have learned a lot from him. For that I am honored to have been his student. What makes Professor Csörgő truly exceptional is that he continues to teach his students, even though he is officially retired.

¹Washington Post Writers Group, *A life that added up to something*. This article may be found on the following web-page <http://cecm.sfu.ca/personal/jborwein/erdos.html>

As well I am thankful to the School of Mathematics and Statistics for all their financial support. In addition, my thanks also go to: the Pacific Institute for the Mathematical Sciences for inviting me to participate in the Industrial Problem Solving Workshop in Calgary last June; the Pál Erdős Center of Mathematics in Budapest for inviting me to the Workshop on Random Walks last July to give a talk on my research; Professor Barbara Szyszkowicz (Carleton University) for funding to attend the Meeting of the Canadian Mathematical Society in Kingston last December; and the Fields Institute for inviting me to attend the Workshop on Probability in Finance last January.

I would also like to thank Dr. Ričardas Zitikis of the Laboratory for Research in Statistics and Probability, Carleton University, for all the discussions and meetings we had, his encouragement and everything he taught me. As well many thanks to Professor Pál Révész who was my 'Diplomarbeit' (thesis) supervisor in Vienna, Austria, for teaching me all the basics I needed and for offering a helping hand the last couple of years.

My experience in Canada would not have been worthwhile if it were not for my friends who were there for me whenever I needed them and who made my stay in Ottawa so enjoyable and unforgettable. In particular, I would like to thank Kulwinder Saini, Paul Wear and Nick Xenos for their proof reading on several occasions. I would also like to express my gratitude to Mrs. Gillian Murray and Mrs. Adrienne Richter for the many questions they answered for me with so much patience.

To my best friend and fiancée, Silvia, her love, support and understanding sustained me over the past years. Finally, I would like to thank those without their help, love, understanding and support I would not have come this far: my family. Thanks for everything!

Gewidmet meinen Eltern, Melitta und Albert Orasch

Contents

Acceptance sheet	ii
Abstract	iii
Acknowledgements	v
Dedication	vii
Table of Contents	viii
List of Figures	xi
List of Symbols	xiii
1 Preliminaries	1
1.1 Basic Definitions	1
1.2 Stochastic Processes	3
2 U-Statistics in Change-point Analysis	8
2.1 Introduction	8
2.2 Motivation	11
2.3 Definition of U-Statistics	12
2.4 Generalized U-Statistics	14
2.5 Variance of U-Statistics	16
2.6 Some Convergence Results for U-Statistics	17

3	At Most One Change-point	21
3.1	Introduction	21
3.1.1	Notations under the Null Hypothesis H_0	22
3.1.2	Notations under the Alternative H_A	26
3.2	Antisymmetric Kernels	28
3.2.1	Asymptotic Results under H_0	28
3.2.2	Asymptotic Results under H_A	33
3.2.3	Estimating the Time of Change	33
3.3	Symmetric Kernels	35
3.3.1	Asymptotic Results under H_0	35
3.3.2	Asymptotic Results under H_A	40
3.3.3	Estimating the Time of Change	45
3.4	Change in the Mean	53
3.5	Change in the Variance	61
4	At Most Two Change-points	70
4.1	Introduction	70
4.1.1	Notations under the Null Hypothesis H_0	72
4.1.2	Notations under the Alternative $H_A^{(2)}$	74
4.2	Antisymmetric Kernels	76
4.2.1	Asymptotic Results under H_0	76
4.2.2	Asymptotic Results under $H_A^{(2)}$	85
4.3	Symmetric Kernels	92
4.3.1	Asymptotic Results under H_0	92
4.3.2	Asymptotic Results under $H_A^{(2)}$	100
4.4	Changes in the Mean	109
4.5	Changes in the Variance	115
4.6	Epidemic Alternatives	119

5	Multiple Change-points	131
5.1	Introduction	131
5.1.1	Notations under the Null Hypothesis H_0	133
5.1.2	Notations under the Alternative $H_A^{(s)}$	136
5.2	Antisymmetric Kernels	138
5.2.1	Asymptotic Results under H_0	138
5.2.2	Asymptotic Results under $H_A^{(s)}$	148
5.3	Symmetric Kernels	154
5.3.1	Asymptotic Results under H_0	154
5.3.2	Asymptotic Results under $H_A^{(s)}$	163
5.4	Multiple Changes in the Mean	172
5.5	Multiple Changes in the Variance	174
5.6	Multiple Changes in the Mean and/or Variance	175
5.7	Estimating the Number of Change-points	177
6	Applying Change-point Theory to the Financial Market	186
6.1	Introduction	186
6.2	Derivative Securities	187
6.2.1	Forward Contracts	187
6.2.2	Futures Contracts	188
6.2.3	Options	188
6.3	Modeling the Behavior of Stock Prices	190
6.4	The Black-Scholes Formula	194
6.5	Changes in the Volatility	198
	Bibliography	203

List of Figures

3.2.1 The limiting function $\bar{u}_{\frac{1}{3}}(t)$ with $\theta_{1,2} = 10$ takes its maximum value of 2.2 at $t = \lambda = \frac{1}{3}$. Note: The x -axis denotes t and the y -axis denotes $\bar{u}_{\lambda}(t)$	34
3.3.2 Summation Area	42
3.3.3 The limiting function $u_{\frac{1}{10}}(t)$ with $\theta_1 = 1$, $\theta_2 = 2$ and $\theta_{1,2} = 3$ takes its maximum value of 0.55 at $t = 0.475$	46
3.3.4 The limiting function $u_{\frac{1}{2}}(t)$ with $\theta_1 = 3$, $\theta_2 = 2$ and $\theta_{1,2} = 1$ takes its maximum value of $\frac{1}{3}$ at $t = \frac{1}{3}$	47
3.3.5 The limiting function $u_{\frac{7}{10}}(t)$ with $\theta_1 = 1$, $\theta_2 = 2$ and $\theta_{1,2} = 3$ takes its maximum value of 0.63 at $t = \lambda = \frac{7}{10}$	50
3.4.6 The data $X_1, \dots, X_{700}$ are i.i.d. $N(1, 1)$ -distributed and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(4, 1)$ -distributed.	54
3.4.7 A geometrical interpretation of $\mathbb{E}\{nS(k) - kS(n)\} = 0$ under H_0 . . .	56
3.5.8 The data $X_1, \dots, X_{700}$ are i.i.d. $N(0, 1)$ -distributed and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(0, 4)$ -distributed.	62
4.4.1 The data $X_1, \dots, X_{300}$ are i.i.d. $N(0, 1)$ -distributed, $X_{301}, \dots, X_{700}$ are i.i.d. $N(3, 1)$ -distributed and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(2, 1)$ -distributed.	110

4.4.2 A geometrical interpretation of $\mathbb{E}\{k_2 S(k_1) + (n - k_1) S(k_2) - k_2 S(n)\} = 0$ under H_0	111
4.5.3 The data $X_1, \dots, X_{100}$ are i.i.d. $N(0, 2)$ -distributed, $X_{101}, \dots, X_{300}$ are i.i.d. $N(0, 1)$ -distributed and $X_{301}, \dots, X_{1000}$ are i.i.d. $N(0, 3)$ -distributed.	116
4.6.4 The data $X_1, \dots, X_{300}$ and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(0, 1)$ -distributed and $X_{301}, \dots, X_{700}$ are i.i.d. $N(3, 1)$ -distributed.	120
4.6.5 Summation Area of $\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2)$	124
4.6.6 The limiting function $\bar{u}_{\frac{1}{3}, \frac{2}{3}}(t_1, t_2)$ with $\theta_{1,2} = 10$ takes its maximum value of 2.2 at the point $(\frac{1}{3}, \frac{2}{3})$	125
4.6.7 The limiting function $\bar{u}_{\frac{1}{10}, \frac{2}{10}}(t_1, t_2)$ with $\theta_{1,2} = 10$ takes its maximum value of 0.9 at the point $(\frac{1}{10}, \frac{2}{10})$	126
4.6.8 The limiting function $\bar{u}_{\frac{1}{10}, \frac{9}{10}}(t_1, t_2)$ with $\theta_{1,2} = 10$ takes its maximum value of 1.6 at the point $(\frac{1}{10}, \frac{9}{10})$	127
4.6.9 The limiting function $\bar{u}_{\frac{1}{3}, \frac{3}{4}}(t_1, t_2)$ with $\theta_{1,2} = 10$, $\theta_{1,3} = 20$ and $\theta_{2,3} = -12$ takes its maximum value of 3.05 at the point $(\frac{1}{120}, \frac{1}{3})$	130
6.3.1 Daily stock prices with different volatilities $\sigma_1 = 0.2$ and $\sigma_2 = 0.4$. . .	193
6.5.2 Price of an European call option depending on σ for $K = \$15$ (above) and $K = \$25$ (below).	199

List of Symbols

a. s.	statement holds almost surely
$\#A$	number of elements in the set A
$B = \{B(t), 0 \leq t \leq 1\}$	Brownian bridge
$\mathbb{E}X$	expectation value of the random variable X
H_0	null hypothesis
$H_A^{(s)}$	alternative hypothesis of s changes
$h(x, y)$	kernel function with $(x, y) \in \mathbf{R} \times \mathbf{R}$
$\mathbf{I}(X \leq x)$	indicator function: $\mathbf{I}(X \leq x) = 1$, if $X \leq x$, otherwise 0
i.i.d. r.v.'s	independent identically distributed random variables
$\log$	natural logarithm
$\mathbf{N}$	set of all positive integers: $\{1, 2, 3, \dots\}$
$(\Omega, \mathcal{F}, \mathbb{P})$	probability space
$\mathbb{P}\{X \leq x\}$	probability that the random variable X is less than or equal to x
$\mathbf{R}$	set of all real numbers: $\{x : -\infty < x < \infty\}$
$R(n) = \sum_{i=1}^n X_i^2$	sum of the squared random variables $X_1, \dots, X_n$
s	number of change-points
$S(t)$	price process of a stock, $0 \leq t \leq T$

$S(n) = \sum_{i=1}^n X_i$	sum of the random variables $X_1, \dots, X_n$
$W = \{W(t), 0 \leq t < \infty\}$	Wiener process
$X \stackrel{\mathcal{D}}{=} N(\mu, \sigma^2)$	X is a normal random variable with mean μ and variance σ^2
$[x]$	largest integer less than or equal to x
$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$	average of the random variables $X_1, \dots, X_n$
$X_n = O_P(a_n)$	$\lim_{c \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \mathbb{P}\left\{\left \frac{X_n}{a_n}\right > c\right\} = 0$
$X_n = O_P(1)$	X_n is bounded in probability
$X_n = o_P(a_n)$	$\frac{X_n}{a_n}$ converges to zero in probability
$X_n = o_P(1)$	X_n converges to zero in probability
$X_n \stackrel{a.s.}{=} o(1)$	X_n converges almost surely to zero
$Var X$	variance of the random variable X
$\stackrel{\mathcal{D}}{=}$	equality in distribution
$\longrightarrow$	convergence of real numbers
$\xrightarrow{a.s.}$	convergence almost surely
$\xrightarrow{P}$	convergence in probability
$\xrightarrow{\mathcal{D}}$	convergence in distribution
$\stackrel{=}{:}$ or $\stackrel{:=}{=}$	equal by definition, where the new expression is on the dotted side
$\square$	indicates the end of a proof

“A mathematician is a machine for
turning coffee into theorems.”

– Pál Erdős

Chapter 1

Preliminaries

In this chapter we give some important definitions used throughout this thesis which focuses on probability, statistics as well as on stochastics.

1.1 Basic Definitions

In this section some fundamental concepts of probability theory are given. The basic mathematical tool for this purpose is measure theory. We give a brief survey of some concepts that will be required in the following chapters. For further references, we refer, for example, to Billingsley (1986).

Definition 1.1.1 Ω , an *arbitrary space* or *set of points* ω , stands for the set of all possible outcomes ω of an experiment. A class $\mathcal{F}$ of subsets¹ of an arbitrary nonempty space Ω is called an *algebra* or *field* if the following are satisfied:

- $\Omega \in \mathcal{F}$,
- $A \in \mathcal{F} \implies A^C \in \mathcal{F}$, where A^C denotes the complement of A ,
- $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$.

A class $\mathcal{F}$ of subsets of Ω is called a σ -*field*, if in addition to being a field, the field also satisfies

- $A_1, A_2, \dots \in \mathcal{F} \implies A_1 \cup A_2 \cup \dots \in \mathcal{F}$.

¹Many authors also use the symbol $\mathcal{A}$ instead of $\mathcal{F}$.

Definition 1.1.2 A *set function* is a real-valued function defined on some class of subsets of Ω . A set function μ on a field $\mathcal{F}$ of subsets of Ω is a *countably additive measure* if the following is satisfied:

- $\mu(A) \in [0, \infty]$ for $A \in \mathcal{F}$,
- $\mu(\emptyset) = 0$,
- If $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ for a disjoint sequence of $\mathcal{F}$ -sets $A_1, A_2, \dots$ then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k),$$

the so-called *countable additivity*.

Moreover, the triple $(\Omega, \mathcal{F}, \mu)$ is a measure space if μ is a measure on a σ -field $\mathcal{F}$ of subsets of Ω . The pair $(\Omega, \mathcal{F})$ is a measurable space if $\mathcal{F}$ is a σ -field of subsets of Ω .

Definition 1.1.3 Consider two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ and the mapping $f : \Omega \rightarrow \Omega'$. Then f is *measurable* if $f^{-1}(A') = \{\omega \in \Omega : f(\omega) \in A', A' \in \mathcal{F}'\} \in \mathcal{F}$. In probability context, a real measurable function is called a *random variable*.

Definition 1.1.4 We call a set function $\mathbb{P}$ on a field $\mathcal{F}$ a *probability measure* if the following is satisfied

- $0 \leq \mathbb{P}(A) \leq 1$ for all $A \in \mathcal{F}$,
- $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$,
- If $\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}$ for a disjoint sequence of $\mathcal{F}$ -sets $A_1, A_2, \dots$ then

$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

We call the triple $(\Omega, \mathcal{F}, \mathbb{P})$ a *probability (measure) space* if $\mathcal{F}$ is a σ -field of subsets of Ω and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.

1.2 Stochastic Processes

In this section the general concept of a stochastic process and some of the most important properties of such processes will be introduced. We will also introduce the notion of Gaussian process which plays a fundamental role when talking about Wiener processes, Brownian bridges and Ornstein-Uhlenbeck processes. After introducing these stochastic processes we state Donsker's theorem which is of crucial importance when dealing with weak convergence of partial sums. For further references on concepts related to stochastic processes we refer, for example, to Cramér and Leadbetter (1967).

Definition 1.2.1 Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability (measure) space and T be a given parameter set. Then a finite and real-valued function $X = \{X(t) = X(t, \omega); t \in T\}$ which is a measurable function of $\omega \in \Omega$ for every fixed t is called a *stochastic process*.

The index set T is said to be *discrete* if it contains at most countably many points, otherwise it is said to be *continuous*.

A one-parameter stochastic process is a stochastic process with a one-dimensional index set T , while an s -parameter stochastic process has an s -dimensional index set T .

If we fix $t \in T$, then $X(t_{fix}, \omega)$, $\omega \in \Omega$, is a random variable and if we fix $\omega \in \Omega$, then $X(t, \omega_{fix})$, $t \in T$, is the so-called *sample path* or *trajectory function* of X .

Next we define a special stochastic process which plays a crucial role in the area of stochastics.

Definition 1.2.2 Let $X = \{X(t) : t \in T\}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. If all its finite-dimensional distributions are normal, or in other words, if every finite-dimensional vector $(X(t_1), \dots, X(t_k))$, $t_i \in T$, $i = 1, \dots, k$, $k \in \mathbb{N}$, of the process X has a multivariate normal distribution, then X is called a *Gaussian process*.

A Gaussian process is uniquely determined by its mean and its covariance function, since these two specify uniquely any multivariate normal distribution. One of

the most important Gaussian processes is the Wiener process. Throughout the rest of this section we use the notations from Csörgő and Révész (1981, Chapter 1).

Definition 1.2.3 A stochastic process $\{W(t; \omega) = W(t); 0 \leq t < \infty\}$, where $\omega \in \Omega$, and $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, is called a (*standard*) *Wiener process* or *Brownian motion*, if the following is satisfied:

- $W(t) - W(s) \stackrel{D}{=} N(0, t - s)$ for all $0 \leq s < t < \infty$ and $W(0) = 0$,
- $W(t)$ is an independent increment process, that is $W(t_2) - W(t_1), W(t_4) - W(t_3), \dots, W(t_{2i}) - W(t_{2i-1})$ are independent r.v.'s for all $0 \leq t_1 < t_2 \leq t_3 < t_4 \leq \dots \leq t_{2i-1} < t_{2i} < \infty$ ($i = 2, 3, \dots$),
- the sample path function $W(t; \omega)$ is continuous in t with probability 1.

The property that $W(0) = 0$ is merely a normalization and is a convenience rather than a basic requirement. If a process $\{\tilde{W}(t); 0 \leq t < \infty\}$ satisfies the other assumptions but not $\tilde{W}(0) = 0$, then the process $\{\tilde{W}(t) - \tilde{W}(0); 0 \leq t < \infty\}$ would be a standard Wiener process. The property that the sample path is continuous in t with probability 1 follows from the first two properties in the definition of a Wiener process in the sense that given a process $\tilde{W}$ with the first two properties, there always exists a version which satisfies the third property (cf. Doob (1953) and Resnick (1992)).

Clearly, the first two properties imply that the covariance function of a Wiener process is

$$\begin{aligned}
 \mathbb{E}W(s)W(t) &= \mathbb{E}W(s \wedge t)W(s \vee t) \\
 &= \mathbb{E}(W(s \wedge t)W(s \vee t) - W^2(s \wedge t) + W^2(s \wedge t)) \\
 &= \mathbb{E}W(s \wedge t)(W(s \vee t) - W(s \wedge t)) + \mathbb{E}W^2(s \wedge t) \\
 &= \mathbb{E}W(s \wedge t)\mathbb{E}(W(s \vee t) - W(s \wedge t)) + \mathbb{E}W^2(s \wedge t) \\
 &= s \wedge t.
 \end{aligned}$$

A constructive proof for the existence of this process is, for example, given by Csörgő and Révész (1981). For a historical summary and development of a Wiener process, as well as for further results, we refer to Csörgő (1979).

We now define another Gaussian process of crucial importance, the so-called *Brownian bridge*.

Definition 1.2.4 A stochastic process $\{B(t); 0 \leq t \leq 1\}$ is called a *Brownian bridge* or *tied down Brownian motion*, if the following is satisfied:

- the joint distribution of $B(t_1), \dots, B(t_n)$ ($0 \leq t_1 < t_2 < \dots < t_n \leq 1; n = 1, 2, \dots$) is Gaussian, with $\mathbb{E}B(t) = 0$,
- the covariance function of $B(t)$ is

$$\mathbb{E}B(s)B(t) = s \wedge t - st,$$

- the sample path function of $B(t; \omega)$ is continuous in t with probability 1.

From the second property it follows that

$$B(0) \stackrel{a.s.}{=} B(1) \stackrel{a.s.}{=} 0.$$

The existence of such a Gaussian process is an easy consequence of the following:

Lemma 1.2.1 Let $\{W(t); 0 \leq t < \infty\}$ be a Wiener process. Then

$$B(t) := W(t) - tW(1), \quad 0 \leq t \leq 1,$$

is a Brownian bridge.

When considering a Wiener process, we can see that the increments of the process over non-overlapping time intervals are stationary and independent, but the process itself is not stationary, that is to say, we do not have the property: Given a stochastic process $\{X(t); t \in T\}$, then for any positive integer k and any points $t_1, \dots, t_k \in T$ the joint distribution of $X(t_1), \dots, X(t_k)$ is the same as the joint distribution of $X(t_1 + h), \dots, X(t_k + h)$ for all $h \in T$, i.e. $(X(t_1), \dots, X(t_k)) \stackrel{\mathcal{D}}{=} (X(t_1 + h), \dots, X(t_k + h))$. This means that a stationary process is a stochastic process whose finite dimensional distributions remain unchanged through shifts in time. One can show that the covariance function $\mathbb{E}X(s)X(t)$ of such a process is

a function of $|s - t|$, the length of the interval. Hence, due to their respective covariance functions it is obvious that Wiener and Brownian bridge processes are not stationary.

But we can define a stochastic process by modifying a Wiener process, such that the new process is stationary. Consider the Gaussian process $\{U(t) = \frac{W(t)}{\sqrt{t}}; 0 \leq t < \infty\}$. By the definition of a Wiener process it follows that

$$\begin{aligned}\mathbb{E}U(t) &= \mathbb{E}\frac{W(t)}{\sqrt{t}} = 0, \\ \mathbb{E}U^2(t) &= \mathbb{E}\frac{W^2(t)}{t} = \frac{t}{t} = 1, \\ \mathbb{E}U(t)U(s) &= \mathbb{E}\frac{W(t)}{\sqrt{t}}\frac{W(s)}{\sqrt{s}} = \frac{s \wedge t}{\sqrt{st}} = \sqrt{\frac{s}{t}}, \quad \text{if } s < t.\end{aligned}$$

Definition 1.2.5 We define the stationary Gaussian process $V(t)$, the so-called *Ornstein-Uhlenbeck process*, via

$$V(t) := U(e^{2t}) = \frac{W(e^{2t})}{e^t}, \quad -\infty < t < \infty.$$

We have that $\mathbb{E}V(t) = 0$, $\mathbb{E}V^2(t) = 1$, and $\mathbb{E}V(s)V(t) = e^{-|t-s|}$.

Next, for further use, we quote the celebrated Donsker Theorem. Let $X_1, X_2, \dots$ be i.i.d.r.v.'s with $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$ and distribution function F . Further, we put $S_n = S(n) = \sum_{i=1}^n X_i$. We construct a sequence of stochastic processes $\{S_n(t); 0 \leq t \leq 1\}$ on $C(0, 1)$, the space of all continuous functions on the interval $(0, 1)$, from the partial sums $S_0, S_1, S_2, \dots, S_n$ as follows:

$$S_n(t) = n^{-1/2}\{S_{[nt]} + X_{[nt]+1}(nt - [nt])\}.$$

We quote Donsker's Theorem (1951) as follows:

Theorem 1.2.1 (Donsker, 1951) *We have, as $n \rightarrow \infty$,*

$$h(S_n(t)) \xrightarrow{\mathcal{D}} h(W(t))$$

for every continuous functional $h : C(0, 1) \rightarrow \mathbf{R}^1$ with respect to sup-norm topology.

Before applying Donsker's Theorem to some special functionals, we quote

Theorem 1.2.2 (Slutsky, 1925) *Let $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c$, where c is a finite constant, then, as $n \rightarrow \infty$,*

- $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$,
- $X_n Y_n \xrightarrow{\mathcal{D}} cX$,
- $X_n / Y_n \xrightarrow{\mathcal{D}} X/c$ if $c \neq 0$.

Donsker's theorem immediately implies that, as $n \rightarrow \infty$,

$$S_n(t) \xrightarrow{\mathcal{D}} W(t), \quad \text{for } t \text{ fixed in } (0, 1].$$

Combined with the fact that $\frac{X_{[nt]+1}}{\sqrt{n}}(nt - [nt]) \xrightarrow{P} 0$, Slutsky's theorem implies, as $n \rightarrow \infty$,

$$\frac{S([nt])}{\sqrt{n}} \xrightarrow{\mathcal{D}} W(t), \quad \text{for } t \text{ fixed in } (0, 1].$$

Moreover, we also have, as $n \rightarrow \infty$,

$$\sup_{0 \leq t \leq 1} \frac{S([nt])}{\sqrt{n}} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} W(t).$$

“If I feel unhappy, I do mathematics to become happy.

If I am happy, I do mathematics to keep happy.”

— *Alfréd Rényi*

Chapter 2

U-Statistics in Change-point Analysis

2.1 Introduction

The problem of abrupt parameter changes arises in many situations of daily life, as well as in a variety of experimental and mathematical sciences. For instance, in medicine one may be interested in testing whether treating HIV patients with a new drug stabilized the condition of the patients, and if not, in estimating the time(s) of change(s) in order to change the treatment. Another example is variance analysis in daily stock prices, where the use of change detection methods allows an investor to reduce his/her risk. In Chapter 6 a particular financial model is described and a special detection method is suggested. Detection of possible change-points is for instance also of interest in archaeology, in econometrics, in epidemiology, in nuclear physics and in quality control.

In practice, usually a (large) set of data is observed, for instance the daily stock prices over the past year. Then a statistical test should determine whether there was a change in the data or not. This set of data may be modeled by saying that we observe independent random variables over a special period of time, hence via a random process. Then we wish to detect whether a change could have occurred in the distribution that governs this random process as time goes by.

We wish to study such phenomena in terms of special stochastic processes based on U-statistics. It is needless to say that there are many other ways to study such phenomena. The construction of these processes is such that statistical tests can be

based on them for detecting possible changes in their distribution.

Change-point problems have originally arisen in the context of quality control, where one typically observes the output of a production line and would wish to signal deviations from an acceptable output level while observing the data. When one observes such a random process sequentially and stops observing at a random time of detecting change, then one speaks of a *sequential procedure* (e. g., stop a production line, if a specified percentage of the output is not good). Otherwise, one usually observes a change in a chronically ordered finite sequence for the sake of determining possible change(s) during the data collection (e. g., check whether a production line produced reasonable output or not). Most such *fixed sample size non-sequential procedures* are described in terms of asymptotic results ('infinite' sample size, i.e., $n \rightarrow \infty$).

Depending on whether the distribution of the data is assumed to be known we use either *parametric* or *nonparametric* models. For parametric models we refer to a survey of Csörgő and Horváth (1997) at the end of the first Chapter of their book. For nonparametric cases we refer, for example, to Brodsky and Darkhovsky (1993), Csörgő and Horváth (1988a, 1988b, 1997), Ferger and Stute (1992), Szyszkowicz (1992, 1995, 1998) as well as to their bibliographies.

We now state the problem of testing for a change in the data in a more mathematical way. Suppose we would like to test the null hypothesis

$$H_0 : X_1, \dots, X_n \text{ are independent identically distributed random variables}$$

against the alternative that there is at most one change-point in the sequence $X_1, \dots, X_n$, namely that we have

$$\begin{aligned} H_A : X_1, \dots, X_n \text{ are independent random variables and there is an integer } \tau, 1 \leq \tau < n, \text{ such that } & \mathbb{P}\{X_1 \leq t\} = \dots = \mathbb{P}\{X_\tau \leq t\}, \\ & \mathbb{P}\{X_{\tau+1} \leq t\} = \dots = \mathbb{P}\{X_n \leq t\} \text{ for all } t \text{ and } \mathbb{P}\{X_\tau \leq t_0\} \neq \\ & \mathbb{P}\{X_{\tau+1} \leq t_0\} \text{ for some } t_0. \end{aligned}$$

This means that we are testing, for having n independent random variables belonging to the same distribution, versus the first τ ones belonging to the same distribution

and the last $n - \tau$ ones to a different one. Therefore we will compare the first k observations to the last $n - k$ ones by using a bivariate function $h(x, y)$ that is often called the kernel function in the literature on U-statistics (see Section 2.3).

Instead of the alternative H_A of at most one change, we will also consider a more general one that allows at most s change-points¹ in the sequence $X_1, \dots, X_n$, namely that we have

$H_A^{(s)} : X_1, \dots, X_n$ are independent random variables and there exist s , $1 \leq s < n$, integers $\tau_1 = \tau_1(n)$, $\tau_2 = \tau_2(n)$, $\dots$, $\tau_s = \tau_s(n)$, $1 \leq \tau_1 \leq \tau_2 \leq \dots \leq \tau_s < n$, such that $\mathbb{P}\{X_1 \leq t\} = \dots = \mathbb{P}\{X_{\tau_1} \leq t\}$, $\mathbb{P}\{X_{\tau_1+1} \leq t\} = \dots = \mathbb{P}\{X_{\tau_2} \leq t\}$, $\dots$, $\mathbb{P}\{X_{\tau_s+1} \leq t\} = \dots = \mathbb{P}\{X_n \leq t\}$ for all t and $\mathbb{P}\{X_{\tau_i} \leq t_0\} \neq \mathbb{P}\{X_{\tau_i+1} \leq t_0\}$ for some t_0 and for all $1 \leq i \leq s$.

Similarly as before, we are now testing for at most s changes in the distribution. Hence, we will have to use a statistic that will somehow ‘feel’ the possibility of s changes. We will split the given sample of size n into $s + 1$, $1 \leq s < n$, blocks, compare each of them and combine the corresponding kernel functions $h(x, y)$ appropriately.

In view of Chapters 3 - 6, here we give an overview of the so-called *U-statistics* which, for example, will be used to detect changes in the mean or the variance. We give some basic results, definitions, and examples (cf. Serfling (1980), Chapter 5). This class of statistics was introduced in a fundamental paper by Hoeffding (1948). The members of this class have good consistency properties and we only assume that our n observations are independent and identically distributed. An appealing feature of a U-statistic (cf. Serfling (1980), Section 5.3) is its simple structure as a sum of identically, but not necessarily independent, distributed random variables. However, by the special device of ‘projection’, a U-statistic may be approximated by a sum of i.i.d. r.v.’s and then classical limit theory for sums does carry over to U-statistics. For proofs we again refer to Serfling (1980).

¹In this as well as in the following parts of this work s will always denote the number of changes. Hence, whenever we are talking about s changes, we think of s as being an integer, i. e., $s \in \mathbb{N}$.

2.2 Motivation

As mentioned earlier, large parts of this thesis focus on constructing stochastic processes based on U-statistics. The construction of these processes is such that statistical tests can be based on them for detecting possible changes in their distribution.

Tests for at most one change-point which are based on processes of U-statistics were first studied by Csörgő and Horváth (1986, 1988b, 1997). They investigate the asymptotic properties (as $n \rightarrow \infty$) of the U-statistics based process

$$Z_k = \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j), \quad 1 \leq k < n, \quad (2.2.1)$$

where the kernel $h(x, y)$ is either symmetric, i.e.,

$$h(x, y) = h(y, x), \quad \text{for all } x, y \in \mathbf{R}, \quad (2.2.2)$$

or antisymmetric, i.e.,

$$h(x, y) = -h(y, x), \quad \text{for all } x, y \in \mathbf{R}. \quad (2.2.3)$$

Typical choices for (2.2.2) are xy , $\frac{(x-y)^2}{2}$ (sample variance), $|x - y|$ (Gini's mean difference), or $\text{sign}(x + y)$ (Wilcoxon's one-sample statistic).

Typical choices for (2.2.3) are $(x - y)$ or $\text{sign}(x - y)$, since $\text{sign}(x - y) = -\text{sign}(y - x)$ and $\text{sign}(0) = 0$. Later on we will use $\frac{(x-y)^2}{2}$ to detect changes in the variance and $x - y$ to detect changes in the mean.

Csörgő and Horváth (1988b, 1997) give various asymptotic distributions of the U-statistics based process $\{Z_k, 1 \leq k < n, \}$ under the null hypothesis H_0 and the alternative H_A for symmetric and antisymmetric kernels. They also give tests that can be used to reject H_0 vs.² H_A .

²vs. = versus = against

2.3 Definition of U-Statistics

Throughout this section we refer to Serfling (1980) and Casella and Berger (1990).

Definition 2.3.1 A parameter θ is said to be *estimable of degree r* for a family of distributions $\mathcal{F}^*$, if r is the smallest sample size for which there exists a function $h^*(x_1, \dots, x_r)$ such that

$$\begin{aligned}\theta &= \theta(F) := \mathbb{E}_F(h^*(X_1, \dots, X_r)) \\ &= \int \dots \int h^*(x_1, \dots, x_r) dF(x_1) \dots dF(x_r),\end{aligned}\quad (2.3.1)$$

for all $F \in \mathcal{F}^*$, where $X_1, \dots, X_r$ are independent observations on a distribution F . $h^*(x_1, \dots, x_r)$ is called the kernel of θ and does not depend on F .

If $h^*(x_1, \dots, x_r) = h^*(x_{\alpha_1}, \dots, x_{\alpha_r})$ for all permutations of the integers $(1, \dots, r)$, then $h^*(x_1, \dots, x_r)$ is called a symmetric kernel.

Let

$$h(x_1, \dots, x_r) = \frac{1}{r!} \sum_p h^*(x_{\alpha_1}, \dots, x_{\alpha_r}), \quad (2.3.2)$$

where $\sum_p$ denotes the summation over all $r!$ permutations $(\alpha_1, \dots, \alpha_r)$ of $(1, \dots, r)$. Then h of (2.3.2) is always a symmetric kernel.

Definition 2.3.2 We define the *mean square error* of an estimator V_n of a parameter θ to be the function of θ defined by

$$\mathbb{E}_F(V_n - \theta)^2 = \text{Var}_F V_n + (\mathbb{E}_F V_n - \theta)^2.$$

Moreover, $(\mathbb{E}_F V_n - \theta)$ is called the *bias* of a point estimator V_n of a parameter θ . V_n depends only on the random sample $X_1, \dots, X_n$, i. e., $V_n = V(X_1, \dots, X_n)$. If the bias is identically (in θ) equal to 0, then the estimator V_n is called *unbiased* and satisfies

$$\mathbb{E}_F V_n = \theta \quad \text{for all } \theta.$$

Definition 2.3.3 A U-statistic of an estimable parameter of degree r is created with the symmetric kernel $h(\cdot)$ by forming

$$U_n = U(X_1, \dots, X_n) = \frac{1}{\binom{n}{r}} \sum_C h(X_{\alpha_1}, \dots, X_{\alpha_r}), \quad (2.3.3)$$

where $\sum_C$ denotes the summation over all $\binom{n}{r}$ combinations of r ($n \geq r$) distinct elements $(\alpha_1, \dots, \alpha_r)$ from $(1, \dots, n)$.

U_n is an unbiased estimator, since $\mathbb{E}_F U_n = \theta$.

Example 2.3.1 *The sample mean.* If $\theta(F) = \text{mean of } F = \mu(F) = \int x dF(x)$ and $h(x) = x$, then the corresponding U-statistic is

$$U(X_1, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i =: \bar{X}_n.$$

Example 2.3.2 *The sample variance.* If $\theta(F) = \text{variance of } F = \sigma^2(F) = \int (x - \mu)^2 dF(x)$ and $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$, then the corresponding U-statistic is

$$\begin{aligned} U(X_1, \dots, X_n) &= \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j) \\ &= \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n \bar{X}_n^2 \right) \\ &=: s_n^2. \end{aligned}$$

When dealing with one sample U-statistics in this work, we will always have $r = 2$. Therefore our kernel functions h will depend on two (real-valued) arguments only.

Next we will define special functions which will be used throughout this work for $r = 2$, e. g., change in the mean (see Section 3.4) or variance (see Section 3.5). We consider the symmetric kernel $h(x_1, \dots, x_r)$ where

$$\mathbb{E}_F \{h^2(X_1, \dots, X_r)\} < \infty.$$

We define

$$\begin{aligned} h_c(x_1, \dots, x_c) &:= \mathbb{E}_F\{h(x_1, \dots, x_c, X_{c+1}, \dots, X_r)\} \\ &= \mathbb{E}_F\{h(X_1, \dots, X_r) | X_1 = x_1, \dots, X_c = x_c\}, \quad 1 \leq c \leq r-1, \end{aligned}$$

and put $h_r := h$. We center at expectation, by defining

$$\begin{aligned} \theta(F) &:= \mathbb{E}_F\{h(X_1, \dots, X_r)\}, \\ \tilde{h}_c &:= h_c - \theta(F), \quad 1 \leq c \leq r, \end{aligned}$$

and put $\tilde{h} := \tilde{h}_r$. Since $\mathbb{E}_F h = \theta$,

$$\mathbb{E}_F \tilde{h}_c(X_1, \dots, X_c) = 0, \quad 1 \leq c \leq r.$$

Furthermore, we define $\zeta_0 := 0$ and, for $1 \leq c \leq r$,

$$\zeta_c := \text{Var}_F\{h_c(X_1, \dots, X_c)\} = \mathbb{E}_F\{\tilde{h}_c^2(X_1, \dots, X_c)\}. \quad (2.3.4)$$

The functions h_c and $\tilde{h}_c$ depend on F for $c \leq r-1$ and the role of these functions is technical only. E. g., they are used to calculate the variance of U-statistics (see Lemma 2.5.1). An application is given in Section 3.1.1, where we use the same methodology for $r = 2$.

2.4 Generalized U-Statistics

Throughout this section we refer to Serfling (1980).

In the case where we have one or several changes in the data, say we have s changes, $s \in \mathbb{N}$, we consider $s+1$ independent collections of independent observations $X_1^{(1)}, X_2^{(1)}, \dots, X_1^{(2)}, X_2^{(2)}, \dots, X_1^{(s+1)}, X_2^{(s+1)}, \dots$, taken from distribution functions $F^{(1)}, \dots, F^{(s+1)}$, respectively. Then for h being symmetric within each of

its $(s + 1)$ blocks of arguments we have

$$\begin{aligned}\theta_{1,\dots,s+1} &= \theta(F^{(1)}, \dots, F^{(s+1)}) \\ &:= \mathbb{E}_{F^{(1)}, \dots, F^{(s+1)}} h(X_1^{(1)}, \dots, X_{\alpha_1}^{(1)}; \dots; X_1^{(s+1)}, \dots, X_{\alpha_k}^{(s+1)}),\end{aligned}\quad (2.4.1)$$

where θ denotes a parametric function for which there is an unbiased estimator.

Definition 2.4.1 The U-statistic for estimating θ is defined as

$$U_n^{(s+1)} = \frac{1}{\prod_{j=1}^{s+1} \binom{n_j}{\alpha_j}} \sum_C h(X_{i_{11}}^{(1)}, \dots, X_{i_{1\alpha_1}}^{(1)}; \dots; X_{i_{(s+1)1}}^{(s+1)}, \dots, X_{i_{(s+1)\alpha_{s+1}}}^{(s+1)}), \quad (2.4.2)$$

where $\{i_{j1}, \dots, i_{j\alpha_j}\}$ denotes a set of α_j distinct elements of the set $\{1, 2, \dots, n_j\}$, $1 \leq j \leq (s + 1)$, and $\sum_C$ denotes summation over all such combinations.

In our specific setup right now we will deal with one single change-point ($s=1$) at time $\tau = \tau(n) := [n\lambda]$, $0 < \lambda < 1$, in the data set $X_1, X_2, \dots, X_n$ and therefore we will have the two independent collections $X_1, X_2, \dots, X_\tau$ and $X_{\tau+1}, X_{\tau+2}, \dots, X_n$ of independent observations taken from distributions $F^{(1)}$ and $F^{(2)}$, respectively. Hence, n_1 and n_2 in (2.4.2) are equal to τ and $n - \tau$, respectively. Using the previous notations with $\alpha_1 = \alpha_2 = 1$ and $s = 1$, we get

$$\begin{aligned}\theta_{1,2} &= \theta(F^{(1)}, F^{(2)}) \\ &:= \mathbb{E}_{F^{(1)}, F^{(2)}} h(X_\tau; X_{\tau+1}),\end{aligned}$$

and furthermore,

$$U_n^{(2)} = \frac{1}{\tau(n - \tau)} \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n h(X_i; X_j).$$

We mention again that τ depends on n .

Example 2.4.1 *The Wilcoxon 2-sample statistic.* Consider two random samples as above and, in addition, we also assume that $F^{(1)}$ and $F^{(2)}$ are continuous. Then the

two-sample U-statistic

$$U_n^{(2)} = \frac{1}{\tau(n-\tau)} \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n \mathbf{I}\{X_i \leq X_j\}$$

is an unbiased estimator for

$$\theta(F^{(1)}, F^{(2)}) = \int F^{(2)} dF^{(1)} = \mathbb{P}(X_{\tau} \leq X_{\tau+1}).$$

2.5 Variance of U-Statistics

In Section 2.3 we have seen that U-statistics are unbiased estimators. But in addition to that we will see that those classes of estimators are the best under all unbiased estimators (cf. Casella and Berger (1990) and Serfling (1980)).

Definition 2.5.1 An estimator V_n^* is called *best unbiased estimator* or *uniform minimum variance unbiased estimator* of θ , if $\mathbb{E}_F V_n^* = \theta$, for all θ , and for any other estimator V_n with $\mathbb{E}_F V_n = \theta$, we have $\text{Var}_F V_n^* \leq \text{Var}_F V_n$, for all θ .

We now state two lemmas due to Hoeffding (1948) which give an explicit formula for the calculation of the variance as well as upper and lower bounds and asymptotic properties.

Lemma 2.5.1 (Hoeffding, 1948) *The variance of U_n as in (2.3.3) is given by*

$$\text{Var}_F(U_n) = \frac{1}{\binom{n}{r}} \sum_{c=1}^r \binom{r}{c} \binom{n-r}{r-c} \zeta_c, \quad (2.5.1)$$

where ζ_c is defined in (2.3.4).

The calculation of the variance can be very difficult and therefore the next lemma is very useful. It gives us an upper and lower bound, the behavior, and the asymptotic behavior of the variance.

Lemma 2.5.2 (Hoeffding, 1948) *The variance of U_n as in (2.3.3) satisfies*

- $\frac{r^2}{n}\zeta_1 \leq \text{Var}_F(U_n) \leq \frac{r}{n}\zeta_r$,
- $(n+1)\text{Var}_F(U_{n+1}) \leq n\text{Var}_F(U_n)$,
- $\text{Var}_F(U_n) = \frac{r^2\zeta_1}{n} + O(n^{-2})$, $n \rightarrow \infty$.

That U-statistics are the best in the class of unbiased estimators of $\theta(F)$ is stated in the following theorem:

Theorem 2.5.1 (Serfling, 1980) *If $S_n = S(X_1, \dots, X_n)$ is an unbiased estimator of $\theta(F)$ based on the sample $X_1, \dots, X_n$ from the distribution F , then the corresponding U-statistic is also unbiased and $\text{Var}_F(U_n) \leq \text{Var}_F(S_n)$. Equality holds if and only if $S_n = U_n$.*

2.6 Some Convergence Results for U-Statistics

The following a.s. behavior result of U-statistics is due to Hoeffding (1961) and follows from the Strong Law of Large Numbers. It is a very important result and will for example be used to prove Theorem 3.3.2.

Theorem 2.6.1 (Hoeffding, 1961) *Let U_n be defined as in (2.3.3) and θ as in (2.3.1). If $\mathbb{E}_F|h| < \infty$, then, as $n \rightarrow \infty$,*

$$U_n \xrightarrow{\text{a.s.}} \theta.$$

For *generalized* $(s+1)$ -sample U-statistics, Sen (1977, Theorem 1) obtains strong convergence of $U_n^{(s+1)}$ under a stronger condition than that of Theorem 2.6.1.

Theorem 2.6.2 (Sen, 1977) *Let $U_n^{(s+1)}$ be defined as in (2.4.2), θ as in (2.4.1), and $\log^+ x = \log(x \vee 1)$, where $x \vee 1$ denotes the maximum of x and 1. If $\mathbb{E}_{F^{(1)}, \dots, F^{(s+1)}}\{|h| \log^+ |h|^s\} < \infty$, then, as $n_j \rightarrow \infty$, $1 \leq j \leq (s+1)$,*

$$U_n^{(s+1)} \xrightarrow{\text{a.s.}} \theta.$$

We mention that in Sen's theorem the condition

$$\mathbb{E}_{F^{(1)}, \dots, F^{(s+1)}}\{|h| \log^+ |h|^s\} < \infty, \quad (2.6.1)$$

could be replaced by the stronger condition

$$\mathbb{E}_{F^{(1)}, \dots, F^{(s+1)}} h^2 < \infty, \quad (2.6.2)$$

since (2.6.2) implies (2.6.1). We note that condition (2.6.1) is sufficient but not necessary (cf. Sen (1977, Section 3)).

In view of the next chapters, where we will establish convergence in probability results for a combination of U-statistics, we consider the two random samples $X_1, X_2, \dots, X_\tau$ and $X_{\tau+1}, X_{\tau+2}, \dots, X_n$ taken from distributions $F^{(1)}$ and $F^{(2)}$, respectively, where $\tau = \tau(n) := [n\lambda]$, $0 < \lambda < 1$. If we assume that $\mathbb{E}_{F^{(1)}}\{h(X_1, X_2)\} < \infty$ then, as $n \rightarrow \infty$, we have by Theorem 2.6.1 that

$$U(X_1, \dots, X_\tau) = \frac{2}{\tau(\tau-1)} \sum_{1 \leq i < j \leq \tau} h(X_i, X_j) \xrightarrow{a.s.} \theta_1, \quad (2.6.3)$$

where $\theta_1 := \mathbb{E}_{F^{(1)}}\{h(X_1, X_2)\}$.

Assuming that $\theta_2 := \mathbb{E}_{F^{(2)}}\{h(X_{\tau+1}, X_{\tau+2})\} < \infty$ and defining $Y_i = X_{i+\tau}$, $1 \leq i \leq n - \tau$, then

$$\begin{aligned} U(X_{\tau+1}, \dots, X_n) &= \frac{2}{(n-\tau)(n-\tau-1)} \sum_{\tau+1 \leq i < j \leq n} h(X_i, X_j) \\ &= \frac{2}{(n-\tau)(n-\tau-1)} \sum_{1 \leq i < j \leq n-\tau} h(X_{i+\tau}, X_{j+\tau}) \\ &= \frac{2}{(n-\tau)(n-\tau-1)} \sum_{1 \leq i < j \leq n-\tau} h(Y_i, Y_j). \end{aligned}$$

If $n \rightarrow \infty$, then we have almost sure convergence of the last expression to θ_2 . Hence we also have, as $n \rightarrow \infty$,

$$U(X_{\tau+1}, \dots, X_n) \xrightarrow{P} \theta_2. \quad (2.6.4)$$

When combining the two different random samples with $n_1 = \tau(n) := [n\lambda]$, $n_2 = n - \tau(n) := n - [n\lambda]$, $0 < \lambda < 1$ and $\alpha_1 = \alpha_2 = 1$, we have to assume that

$\mathbb{E}_{F^{(1)}, F^{(2)}} \{|h(X_i, X_j)| \log^+ |h(X_i, X_j)|\} < \infty$, $1 \leq i < j \leq n$. Then Theorem 2.6.2 applies and as $n \rightarrow \infty$ we have that

$$\begin{aligned} U^{(2)}(X_1, \dots, X_\tau; X_{\tau+1}, \dots, X_n) &= \frac{1}{\tau(n-\tau)} \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n h(X_i, X_j) \\ &= \frac{1}{\tau(n-\tau)} \sum_{i=1}^{\tau} \sum_{j=1}^{n-\tau} h(X_i, X_{j+\tau}) \\ &\xrightarrow{P} \theta_{1,2}, \end{aligned} \quad (2.6.5)$$

where $\theta_{1,2} := \mathbb{E}_{F^{(1)}, F^{(2)}} \{h(X_\tau, X_{\tau+1})\}$ and $\tau := [n\lambda]$, $0 < \lambda < 1$.

If $F = F^{(1)} = F^{(2)}$, then for $U^{(2)}(X_1, \dots, X_\tau; X_{\tau+1}, \dots, X_n)$ in (2.6.5) we have, as $n \rightarrow \infty$,

$$\begin{aligned} U(X_1, \dots, X_n) &= \frac{1}{\tau(n-\tau)} \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n h(X_i, X_j) \\ &= \frac{1}{\tau(n-\tau)} \left(\sum_{1 \leq i < j \leq n} - \sum_{1 \leq i < j \leq \tau} - \sum_{\tau+1 \leq i < j \leq n} \right) h(X_i, X_j) \\ &\xrightarrow{P} \frac{1}{\lambda(1-\lambda)} \left(\frac{1}{2} - \frac{\lambda^2}{2} - \frac{(1-\lambda)^2}{2} \right) \theta_1 \\ &= \theta_1 \end{aligned} \quad (2.6.6)$$

where $\theta_1 := \mathbb{E}_F \{h(X_1, X_2)\}$ and $\tau := [n\lambda]$, $0 < \lambda < 1$. Therefore, in this case, (2.6.5) holds true assuming $\mathbb{E}_F \{h(X_1, X_2)\} < \infty$ only, just like in (2.6.3). Moreover, (2.6.3), (2.6.4) and (2.6.5) converge in probability to the same finite value.

Example 2.6.1 Let us assume that $F^{(1)} = N(\mu_1, \sigma_1^2)$ and $F^{(2)} = N(\mu_2, \sigma_2^2)$ and define the kernel $h(x, y) = \frac{1}{2}(x - y)^2$. This kernel was used in Example 2.3.2 where we considered the sample variance. Using the results from above and letting again $\tau(n) := [n\lambda]$, $0 < \lambda < 1$, we have that, as $n \rightarrow \infty$,

$$U^{(2)}(X_1, \dots, X_\tau; X_{\tau+1}, \dots, X_n) \xrightarrow{P} \theta_{1,2} \quad (2.6.7)$$

where $\theta_{1,2} := \mathbb{E}_{F^{(1)}, F^{(2)}} \{h(X_\tau, X_{\tau+1})\} = \frac{\mu_1^2 + \mu_2^2}{2} + \frac{\sigma_1^2 + \sigma_2^2}{2} - \mu_1 \mu_2$.

If $F = F^{(1)} = F^{(2)}$ and $n \rightarrow \infty$, then

$$U(X_1, \dots, X_n) \xrightarrow{P} \theta_1.$$

where $\theta_1 := \mathbb{E}_F\{h(X_1, X_2)\} = \sigma^2$, an account of $\mu_1 = \mu_2$ and $\sigma_1^2 = \sigma_2^2 =: \sigma^2$.

“If the facts don’t fit the theory,
change the facts.”
– *Albert Einstein*

Chapter 3

At Most One Change-point

3.1 Introduction

We are to test the null hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables

against the alternative that there is at most one (single) change-point in the sequence $X_1, X_2, \dots, X_n$, namely that we have

$H_A : X_1, \dots, X_n$ are independent random variables and there is an integer τ , $1 \leq \tau < n$, such that $\mathbb{P}\{X_1 \leq t\} = \dots = \mathbb{P}\{X_\tau \leq t\}$, $\mathbb{P}\{X_{\tau+1} \leq t\} = \dots = \mathbb{P}\{X_n \leq t\}$ for all t and $\mathbb{P}\{X_\tau \leq t_0\} \neq \mathbb{P}\{X_{\tau+1} \leq t_0\}$ for some t_0 .

As mentioned in Section 2.2, tests for at most one change-point which are based on processes of U-statistics were studied by Csörgő and Horváth (1986, 1988b, 1997). Similarly we will investigate the asymptotic properties (as $n \rightarrow \infty$) of the U-statistics based process

$$Z_k = \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j), \quad 1 \leq k < n, \quad (3.1.1)$$

where the kernel $h(x, y)$ is either symmetric or antisymmetric. We will state their basic results and, moreover, we will impose conditions such that we will have a

good estimator for the time of change not only in the antisymmetric case (cf. Section 3.2.3), but also in the symmetric one (cf. Section 3.3.3), which they have not investigated. Then we will apply the theoretical results to detect at most one (single) change in the mean (h antisymmetric) and in the variance (h symmetric). In particular, testing for at most one change in the mean is illustrated by using a geometrical argument (cf. Section 3.4). We mention that though in part we build on them, most of the results of Csörgő and Horváth (1986, 1988b, 1997) in this section become immediate consequences of the results in Chapter 5 which deal with multiple changes (put $s = 1$).

3.1.1 Notations under the Null Hypothesis H_0

We define¹

$$\mathbb{E}h(X_i, X_j) =: \theta, \quad 1 \leq i < j \leq n, \quad (3.1.2)$$

and

$$\mathbb{E}h^2(X_i, X_j) =: \nu, \quad 1 \leq i < j \leq n. \quad (3.1.3)$$

We assume throughout the whole chapter that

$$\nu < \infty, \quad (3.1.4)$$

which of course implies

$$|\theta| < \infty. \quad (3.1.5)$$

¹Note that in Section 2.3 and Section 2.6 we used the equivalent notation $\theta := \mathbb{E}_F h(X_i, X_j)$, where F denotes the distribution function of the i.i.d. r.v.'s $X_1, \dots, X_n$, instead.

Furthermore, the expected value of Z_k , $1 \leq k < n$, defined by

$$Z_k = \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j), \quad 1 \leq k < n, \quad (3.1.6)$$

where the kernel $h(x, y)$ is either symmetric or antisymmetric, becomes

$$\begin{aligned} \mathbb{E}Z_k &= \mathbb{E} \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j) \\ &= \sum_{1 \leq i \leq k} \sum_{k < j \leq n} \mathbb{E}h(X_i, X_j) \\ &= k(n-k)\theta, \quad 1 \leq k < n. \end{aligned} \quad (3.1.7)$$

Let X and Y be independent identically distributed random variables and h be an *antisymmetric* kernel as defined in (2.2.3). Then we have that

$$\begin{aligned} \mathbb{E}\{h(X, Y)\} &= \mathbb{E}\{-h(Y, X)\} \\ &= -\mathbb{E}\{h(Y, X)\} \\ &= -\mathbb{E}\{h(X, Y)\}, \end{aligned}$$

which is possible only, if

$$\theta = \mathbb{E}\{h(X, Y)\} = 0. \quad (3.1.8)$$

Define

$$\tilde{h}(t) := \mathbb{E}\{h(t, X_2) - \theta\}. \quad (3.1.9)$$

Then condition (3.1.4) implies that

$$\mathbb{E}\tilde{h}^2(X_1) < \infty.$$

We also assume that

$$\bar{\sigma}^2 := \mathbb{E}\tilde{h}^2(X_1) > 0, \quad (3.1.10)$$

which is the so-called *non-degenerate* case when studying U-statistics. In this and the following chapters we will focus on this case, but we mention that the degenerate case in the context of change-point analysis was, for example, studied by Csörgő and Horváth (1997, Section 2.4).

We mention that we will use the symbol σ^2 in this chapter, as well as in the following ones, for the variance of the given data $X_1, \dots, X_n$, and $\bar{\sigma}^2$ for the expected value as in (3.1.10). The function in (3.1.9) induces the projection of U-statistics into sums of i.i.d. r.v.'s. We centralize Z_k by its mean, and put

$$U_k = Z_k - k(n-k)\theta, \quad 1 \leq k < n, \quad (3.1.11)$$

where the kernel function h in Z_k is *symmetric*. Actually, we define the same for an *antisymmetric* kernel function, but since (3.1.8) holds,

$$k(n-k)\theta = 0, \quad 1 \leq k < n,$$

and in this case we define

$$\bar{U}_k = Z_k, \quad 1 \leq k < n. \quad (3.1.12)$$

Moreover, we may write Z_k as the sum of three U-statistics and accordingly, $U_k, 1 \leq k < n$, equals to

$$U_k = U_n^{(3)} - (U_k^{(1)} + U_k^{(2)}), \quad 1 \leq k < n, \quad (3.1.13)$$

where

$$U_k^{(1)} = \sum_{1 \leq i < j \leq k} h(X_i, X_j) - \binom{k}{2} \theta,$$

$$\begin{aligned}
U_k^{(2)} &= \sum_{k+1 \leq i < j \leq n} h(X_i, X_j) - \binom{n-k}{2} \theta, \\
U_n^{(3)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j) - \binom{n}{2} \theta.
\end{aligned}$$

Similarly we define for an *antisymmetric* kernel

$$\bar{U}_k = Z_n^{(3)} - (Z_k^{(1)} + Z_k^{(2)}), \quad 1 \leq k < n, \quad (3.1.14)$$

where

$$\begin{aligned}
Z_k^{(1)} &= \sum_{1 \leq i < j \leq k} h(X_i, X_j), \\
Z_k^{(2)} &= \sum_{k+1 \leq i < j \leq n} h(X_i, X_j), \\
Z_n^{(3)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j).
\end{aligned}$$

For further use we also define

$$U_n(t) = \frac{1}{\tilde{\sigma} n^{3/2}} U_{[(n+1)t]}, \quad 0 < t < 1, \quad (3.1.15)$$

and

$$\bar{U}_n(t) = \frac{1}{\tilde{\sigma} n^{3/2}} \bar{U}_{[(n+1)t]}, \quad 0 < t < 1. \quad (3.1.16)$$

In this, and in the following chapters as well, U-statistic based processes that have a bar on top have an antisymmetric kernel and those without a bar a symmetric kernel. This will make it easier to distinguish between the antisymmetric and the symmetric cases.

3.1.2 Notations under the Alternative H_A

Let $F^{(1)}(t) = \mathbb{P}\{X_\tau \leq t\}$ and $F^{(2)}(t) = \mathbb{P}\{X_{\tau+1} \leq t\}$ be the respective distribution functions of the observations before and after the postulated change τ , and put²

$$\mathbb{E}h(X_i, X_j) =: \begin{cases} \theta_1, & 1 \leq i < j \leq \tau, \\ \theta_2, & \tau < i < j \leq n, \\ \theta_{1,2}, & 1 \leq i \leq \tau < j \leq n, \end{cases} \quad (3.1.17)$$

and

$$\mathbb{E}h^2(X_i, X_j) =: \begin{cases} \nu_1, & 1 \leq i < j \leq \tau, \\ \nu_2, & \tau < i < j \leq n, \\ \nu_{1,2}, & 1 \leq i \leq \tau < j \leq n. \end{cases} \quad (3.1.18)$$

Assume throughout the whole chapter that $\mathbb{E}h^2(X_i, X_j)$ is finite for all possible choices of i and j , namely

$$\nu_1, \nu_2, \nu_{1,2} < \infty, \quad (3.1.19)$$

which implies that

$$|\theta_1|, |\theta_2|, |\theta_{1,2}| < \infty. \quad (3.1.20)$$

Due to (3.1.17) and (3.1.20), we can calculate the expected value of Z_k , $1 \leq k < n$, which is defined by

$$Z_k = \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j), \quad 1 \leq k < n, \quad (3.1.21)$$

where the kernel $h(x, y)$ is either symmetric or antisymmetric. Namely, under H_A ,

²Note that in Section 2.4 and Section 2.6 we used the equivalent notations $\theta_1 := \mathbb{E}_{F^{(1)}} h(X_1, X_2)$, $\theta_2 := \mathbb{E}_{F^{(2)}} h(X_{\tau+1}, X_{\tau+2})$ and $\theta_{1,2} := \mathbb{E}_{F^{(1)}, F^{(2)}} h(X_\tau, X_{\tau+1})$ instead.

we have

$$\begin{aligned}
\mathbb{E}Z_k &= \mathbb{E} \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j) \\
&= \sum_{1 \leq i \leq k} \sum_{k < j \leq n} \mathbb{E}h(X_i, X_j) \\
&= \begin{cases} k(n - \tau)\theta_{1,2} + k(\tau - k)\theta_1, & 1 \leq k \leq \tau \\ \tau(n - k)\theta_{1,2} + (n - k)(k - \tau)\theta_2, & \tau \leq k < n. \end{cases} \quad (3.1.22)
\end{aligned}$$

Therefore, by using (3.1.8), we get that in the *antisymmetric* case

$$\theta_1 = \theta_2 = 0, \quad (3.1.23)$$

and (3.1.22) becomes

$$\mathbb{E}Z_k = \begin{cases} k(n - \tau)\theta_{1,2}, & 1 \leq k \leq \tau, \\ \tau(n - k)\theta_{1,2}, & \tau \leq k < n. \end{cases} \quad (3.1.24)$$

When searching for the possible time of change τ , one has to investigate the behavior of the process $Z_k, 1 \leq k < n$, under H_A , and when looking at the expected value of the process as a function in k , we see that it is increasing before the postulated change-point τ and decreasing after it, when assuming that h is antisymmetric and $\theta_{1,2}$ is positive. Moreover, it reaches its maximum at time τ , the change-point. Therefore, we have to find k , where the process $Z_k, 1 \leq k < n$, reaches its maximum. Consequently, for $\theta_{1,2}$ positive, we define $\hat{Z}_n$ to be the maximum (taken in k) over all Z_k 's, $1 \leq k < n$, i.e.,

$$\hat{Z}_n = \max_{1 \leq k < n} \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j), \quad (3.1.25)$$

and as an estimation for the change-point τ we define

$$\hat{\tau} = \min \left\{ k : \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j) = \hat{Z}_n \right\}. \quad (3.1.26)$$

Since (3.1.24) is easier to handle than (3.1.22), more is known when we have an

antisymmetric kernel than about a *symmetric* kernel. It is clear that we will need to impose special conditions on the expected value using a symmetric kernel, such that the maximum again will be reached at time τ .

We mention that the moments under the null hypothesis of no change may be derived from the formulae under the alternative. Since

$$\theta_1 = \theta_2 = \theta_{1,2},$$

and

$$\nu_1 = \nu_2 = \nu_{1,2},$$

and if we define

$$\theta := \theta_1,$$

$$\nu := \nu_1,$$

then the previous results reduce to the corresponding ones in Section 3.1.1.

3.2 Antisymmetric Kernels

We consider the processes Z_k , $1 \leq k < n$, $n \geq 2$, where the kernel h is antisymmetric as in (2.2.3). By using the notations and assumptions from Sections 3.1.1 and 3.1.2 here, we give some well known asymptotic results (cf. Csörgő and Horváth (1986, 1988b, 1997)) under H_0 (cf. Section 3.2.1) and H_A (cf. Section 3.2.2) as well as the time of change τ (cf. Section 3.2.3).

3.2.1 Asymptotic Results under H_0

We wish to study the limiting behavior of the stochastic process, as $n \rightarrow \infty$,

$$\bar{U}_k = Z_k = \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j), \quad 1 \leq k < n,$$

in the sup-norm under the null hypothesis of no-change.

The asymptotic behavior of $\bar{U}_k$ will be derived from the following *reduction principle* (Lemma 3.2.1). It is a consequence of Janson and Wichura (1983, Theorem 2.1) and given explicitly by Huse (1988, Lemma 2.1.6).

Lemma 3.2.1 (Huse, 1988) *Let h be an antisymmetric kernel and $\bar{U}_k$ as in (3.1.14). Then under H_0 the following statements hold true as $n \rightarrow \infty$:*

$$\begin{aligned} \max_{1 \leq k \leq n} \left| Z_k^{(1)} - \sum_{i=1}^k (k - 2i + 1) \tilde{h}(X_i) \right| &= O_P(n), \\ \max_{1 \leq k \leq n} \left| Z_k^{(2)} - \sum_{i=k+1}^n (n + k - 2i + 1) \tilde{h}(X_i) \right| &= O_P(n), \\ \left| Z_n^{(3)} - \sum_{i=1}^n (n - 2i + 1) \tilde{h}(X_i) \right| &= O_P(n). \end{aligned}$$

The basic idea for the proof is given by the above mentioned Janson and Wichura. Their results are stated in the context of stochastic area integrals. We will state their theorem (see Theorem 4.2.1) and will discuss how this lemma follows from their theorem in the context of at most two change-points in Chapter 4.

Janson and Wichura do not present a detailed proof. A detailed proof for this lemma is given by Huse (1988, Lemma 2.1.6). She proves that the sequence of random variables

$$Z_k^* = Z_k^{(1)} - \sum_{i=1}^k (k - 2i + 1) \tilde{h}(X_i), \quad k = 1, 2, \dots,$$

is a martingale. Using the notations of Billingsley (1986, Section 35), we quote:

Definition 3.2.1 Let $X_1, X_2, \dots$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{F}_1, \mathcal{F}_2, \dots$ be a sequence of σ -fields in $\mathcal{F}$. Then the sequence $\{(X_n, \mathcal{F}_n) : n = 1, 2, \dots\}$ is a *martingale* if the following four conditions hold:

1. $\mathcal{F}_n \subset \mathcal{F}_{n+1}$,
2. X_n is measurable $\mathcal{F}_n$,
3. $\mathbb{E}|X_n| < \infty$,

$$4. \mathbb{E}(X_{n+1}|\mathcal{F}_n) \stackrel{a.s.}{=} X_n.$$

If instead of 4. we have

$$5. \mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n \text{ a.s.},$$

then the sequence $\{(X_n, \mathcal{F}_n) : n = 1, 2, \dots\}$ is called a *submartingale*.

Having shown that Z_k^* is a martingale sequence, and using the *Hájek-Rényi inequality* (see Shorack and Wellner (1986, Appendix A)), Huse (1988) deduces

$$\lim_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\{\max_{1 \leq k \leq n} |Z_k^*| \geq cn\} = 0,$$

which shows that $\max_{1 \leq k \leq n} |Z_k^*| = O_P(n)$. The other two statements of Lemma 3.2.1 are proved in a similar way. As a consequence, we have (cf. Corollary 2.1.7 of Huse (1988))

Corollary 3.2.1 (Huse, 1988) *Under the same conditions as in Lemma 3.2.1 we have as $n \rightarrow \infty$*

$$\max_{1 \leq k \leq n} \left| \bar{U}_k - \left\{ n \sum_{i=1}^k \bar{h}(X_i) - k \sum_{i=1}^n \bar{h}(X_i) \right\} \right| = O_P(n).$$

Proof of Corollary 3.2.1 Since $\bar{U}_k = Z_n^{(3)} - (Z_k^{(1)} + Z_k^{(2)})$, $1 \leq k < n$, we have that

$$\begin{aligned} \kappa_n &:= \max_{1 \leq k \leq n} \left| \bar{U}_k - \left\{ n \sum_{i=1}^k \bar{h}(X_i) - k \sum_{i=1}^n \bar{h}(X_i) \right\} \right| \\ &= \max_{1 \leq k \leq n} \left| Z_n^{(3)} - (Z_k^{(1)} + Z_k^{(2)}) - \left\{ n \sum_{i=1}^k \bar{h}(X_i) - k \sum_{i=1}^n \bar{h}(X_i) \right\} \right|. \end{aligned}$$

We add and subtract special terms to apply the previous lemma. Accordingly,

$$\begin{aligned} \kappa_n &= \max_{1 \leq k \leq n} \left| \left(Z_n^{(3)} - \sum_{i=1}^n (n - 2i + 1) \bar{h}(X_i) \right) - \left(Z_k^{(1)} - \sum_{i=1}^k (k - 2i + 1) \bar{h}(X_i) \right) \right. \\ &\quad \left. - \left(Z_k^{(2)} - \sum_{i=k+1}^n (n + k - 2i + 1) \bar{h}(X_i) \right) - \left(n \sum_{i=1}^k \bar{h}(X_i) - k \sum_{i=1}^n \bar{h}(X_i) \right) \right| \end{aligned}$$

$$\begin{aligned}
& - \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i) + \sum_{i=1}^k (k-2i+1) \tilde{h}(X_i) \\
& + \sum_{i=k+1}^n (n+k-2i+1) \tilde{h}(X_i) \Big|.
\end{aligned}$$

Since,

$$\begin{aligned}
n \sum_{i=1}^k \tilde{h}(X_i) - k \sum_{i=1}^n \tilde{h}(X_i) - \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i) + \sum_{i=1}^k (k-2i+1) \tilde{h}(X_i) \\
+ \sum_{i=k+1}^n (n+k-2i+1) \tilde{h}(X_i) = 0,
\end{aligned}$$

we use the previous lemma and get the desired result, namely

$$\begin{aligned}
\kappa_n & \leq \left| Z_n^{(3)} - \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i) \right| + \max_{1 \leq k \leq n} \left| Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \tilde{h}(X_i) \right| \\
& + \max_{1 \leq k \leq n} \left| Z_k^{(2)} - \sum_{i=k+1}^n (n+k-2i+1) \tilde{h}(X_i) \right| + 0 \\
& = O_P(n).
\end{aligned}$$

□

We repeated here Huse's (1988) proof of Corollary 3.2.1 for the sake of demonstrating the usefulness of the reduction principle. Namely, $\bar{U}_k$ is approximated by sums of i.i.d. r.v.'s for which there are many limit theorems available.

Now we study the asymptotic behavior of $\bar{U}_k$, $1 \leq k < n$, in the sup-norm. As mentioned before the following results are due to Csörgő and Horváth (1988b, 1997).

Theorem 3.2.1 (Csörgő and Horváth, 1988b) *Assume that H_0 , (2.2.3), (3.1.4) and (3.1.10) hold. Then we can define a sequence of Brownian bridges $\{B_n(t), 0 \leq t \leq 1\}_{n \in \mathbb{N}}$ such that, as $n \rightarrow \infty$,*

$$\sup_{0 < t < 1} |\bar{U}_n(t) - B_n(t)| = o_P(1), \tag{3.2.1}$$

where

$$\{B_n(t), 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \{B(t), 0 \leq t \leq 1\},$$

for each n and with B being a standard Brownian bridge.

A proof is given by Csörgő and Horváth (1988b, Theorem 4.1). They show that $n \sum_{i=1}^k \tilde{h}(X_i)$ may be associated with $nW([nt])$ and $k \sum_{i=1}^n \tilde{h}(X_i)$ with $[nt]W(n)$, where $W(\cdot)$ is a standard Wiener process. This implies the theorem by using Corollary 3.2.1 via multiplying both sides of its statement by $\frac{1}{\sigma n^{3/2}}$.

Theorem 3.2.1 implies that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |\bar{U}_n(t)| \stackrel{\mathcal{D}}{\rightarrow} \sup_{0 < t < 1} |B(t)|. \quad (3.2.2)$$

This means that the sup-functionals of $|\bar{U}_n(t)|$ converge in distribution to the sup-functional of a Brownian bridge. Consequently, we can use tables for the supremum of a Brownian bridge to accept or reject H_0 . It is known that

$$\mathbb{P}\left\{\sup_{0 \leq t \leq 1} |B_n(t)| \leq x\right\} = 1 - \sum_{k \neq 0} (-1)^{k+1} \exp(-2k^2 x^2), \quad x > 0, \quad (3.2.3)$$

the well known limiting distribution of the two-sided Kolmogorov-Smirnov statistic (cf. Kolmogorov, 1933) which has been widely tabulated.

Tests based on $\sup_{0 < t < 1} |\bar{U}_n(t)|$ are not sensitive on the tails. Hence, we note, that a weighted version of Theorem 3.2.1 holds true as in Theorem 2.1 of Szyszkowicz (1991) if we replace $\frac{S(\lfloor (n+1)t \rfloor) - tS(n)}{\sigma}$ by $\bar{U}_n(t)$. In particular, by using the weight

$$q(t) = \left(t(1-t) \log \log \frac{1}{t(1-t)}\right)^{1/2},$$

we have, for example, that, as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} \frac{|\bar{U}_n(t)|}{q(t)} \stackrel{\mathcal{D}}{\rightarrow} \sup_{0 < t < 1} \frac{|B(t)|}{q(t)},$$

and we can use tables for $\sup_{0 < t < 1} \frac{|B(t)|}{q(t)}$ (cf. Eastwood and Eastwood (1998)) and

reject H_0 for large values of $\sup_{0 < t < 1} \frac{|\bar{U}_n(t)|}{q(t)}$.

3.2.2 Asymptotic Results under H_A

Csörgő and Horváth (1988b, Theorem 3.1) studied the asymptotic behavior of $\bar{U}_k$ under the alternative H_A and proved the following theorem.

Theorem 3.2.2 (Csörgő and Horváth, 1988b) *Assume that (2.2.3), H_A hold and*

$$\mathbb{E}\{|h(X_{[n\lambda]}, X_{[n\lambda]+1})| \log^+ (|h(X_{[n\lambda]}, X_{[n\lambda]+1})|)\} < \infty, \quad (3.2.4)$$

where $\tau = \tau(n) := [n\lambda]$, $0 < \lambda < 1$, and $\log^+ x = \log(x \vee 1)$. Then, as $n \rightarrow \infty$,

$$\frac{1}{n^2} \bar{U}_{[(n+1)t]} \xrightarrow{P} \bar{u}_\lambda(t) = \begin{cases} t(1-\lambda)\theta_{1,2}, & 0 < t \leq \lambda, \\ (1-t)\lambda\theta_{1,2}, & \lambda \leq t < 1. \end{cases}$$

We note that Theorem 3.2.2 also holds if we replace condition (3.2.4) by condition (3.1.19). This theorem follows from Theorem 3.3.2, where h is assumed to be symmetric, by using the fact that $\theta_1 = \theta_2 = 0$.

Theorem 3.2.2 implies the consistency of tests based on $\{\bar{U}_{[(n+1)t], 0 \leq t < 1}\}$, where $\bar{U}_k = Z_k$, $1 \leq k < n$. Assuming second finite moments as in (3.1.19) such that the results under H_0 hold then we can reject H_0 vs. H_A , when $\sup_{0 < t < 1} \frac{1}{n^{3/2}\bar{\sigma}} |\bar{U}_{[(n+1)t]}|$ is large and $\theta_{1,2} \neq 0$.

3.2.3 Estimating the Time of Change

We have seen in Section 3.1 that an antisymmetric kernel depends only on $\theta_{1,2}$ but not on θ_1 or θ_2 . Hence, if $\theta_{1,2}$ is positive we define

$$\hat{\tau} = \min \left\{ k : Z_k = \max_{1 \leq m < n} Z_m \right\},$$

since the point where Z_k , $1 \leq k < n$, reaches its unique maximum is at $k = \tau$, as it can be seen in Figure 3.2.1.

Figure 3.2.1: The limiting function $\bar{u}_{\frac{1}{3}}(t)$ with $\theta_{1,2} = 10$ takes its maximum value of 2.2 at $t = \lambda = \frac{1}{3}$. Note: The x -axis denotes t and the y -axis denotes $\bar{u}_{\lambda}(t)$.

Otherwise, if $\theta_{1,2}$ is negative, we define

$$\tilde{\tau} = \min \left\{ k : Z_k = \min_{1 \leq m < n} Z_m \right\}$$

since the point where Z_k , $1 \leq k < n$ reaches its unique minimum is at $k = \tau$.

In practice $\theta_{1,2}$ is unknown, but since $\frac{1}{n^2} Z_{[(n+1)t]} = \frac{1}{n^2} \bar{U}_{[(n+1)t]} \xrightarrow{P} \bar{u}_{\lambda}(t)$, $\frac{1}{n^2} Z_{[(n+1)t]}$ is likely to exhibit whether we have a $\smile$ -type or $\frown$ -type function. E.g., if we have a $\frown$ -type function then we take $\hat{\tau}$ to estimate τ . Ferger and Stute (1992) showed that $\hat{\tau}$ and $\tilde{\tau}$ respectively are strongly consistent estimators of τ and Gombay (1998) gave the asymptotic distribution of $\max_{1 \leq k < n} Z_k$ and that of $\tau - \hat{\tau}$ under the alternative H_A as we will discuss in Section 3.3.3.

Now we are interested in the null-distribution of $\hat{\tau}$ which may be described as the distribution of the argument of the maximum of Z_k , $1 \leq k < n$. From Theorem 3.2.1 we know that $U_n(t)$ converges weakly to a Brownian bridge. Hence, $\frac{\hat{\tau}}{n}$ converges in distribution to the time where a Brownian bridge reaches its maximum and it follows

from Birnbaum and Pyke (1958) (cf. Shorack and Wellner (1986, p. 385)) that the time where a Brownian bridge reaches its maximum is a Uniform(0,1)-distributed random variable. Hence, under H_0 , $\hat{\tau}$ takes every integer value from 1 to n with the same probability as stated in the following theorem by Csörgő and Horváth (1997, Theorem 2.4.14):

Theorem 3.2.3 (Csörgő and Horváth, 1997) *We assume that H_0 , (2.2.3), (3.1.4) and (3.1.10) hold, then*

$$\lim_{n \rightarrow \infty} P\left\{\frac{\hat{\tau}}{n} \leq x\right\} = x \quad \text{for all } 0 \leq x \leq 1.$$

Of course, using similar arguments as before we get a similar result for the minimizer $\tilde{\tau}$ which may be stated in a corollary as follows:

Corollary 3.2.2 *We assume that H_0 , (2.2.3), (3.1.4) and (3.1.10) hold, then*

$$\lim_{n \rightarrow \infty} P\left\{\frac{\tilde{\tau}}{n} \leq x\right\} = x \quad \text{for all } 0 \leq x \leq 1.$$

3.3 Symmetric Kernels

We consider the processes Z_k , $1 \leq k < \infty$, $n \geq 2$, where the kernel h is symmetric as in (2.2.2). By using notations and assumptions from Sections 3.1.1 and 3.1.2 we give some well known results (cf. Csörgő and Horváth (1997)) under H_0 (cf. Section 3.3.1) and H_A (cf. Section 3.3.2). Furthermore, we will investigate under what assumptions we may define an estimator for the time of change (cf. Section 3.3.3) as in the antisymmetric case in Section 3.2.3.

3.3.1 Asymptotic Results under H_0

We wish to study the limiting behavior of the stochastic process, as $n \rightarrow \infty$,

$$\begin{aligned} U_k &= Z_k - k(n-k)\theta \\ &= \sum_{1 \leq i \leq k} \sum_{k < j \leq n} h(X_i, X_j) - k(n-k)\theta, \quad 1 \leq k < n, \end{aligned}$$

under the null hypothesis of no-change in distribution. Similarly to the antisymmetric case, the asymptotic behavior of U_k will be derived from the following *reduction principle* (Lemma 3.3.1). It is a consequence of Theorem 1 of P. Hall (1979) and given explicitly by Huse (1988, Lemma 2.1.12).

Lemma 3.3.1 (Huse, 1988) *Let h be a symmetric kernel and U_k as in (3.1.13). Then under H_0 the following statements hold true as $n \rightarrow \infty$:*

$$\begin{aligned} \max_{1 \leq k \leq n} \left| U_k^{(1)} - k \sum_{i=1}^k \tilde{h}(X_i) \right| &= O_P(n), \\ \max_{1 \leq k \leq n} \left| U_k^{(2)} - (n-k) \sum_{i=k+1}^n \tilde{h}(X_i) \right| &= O_P(n), \\ \left| U_n^{(3)} - n \sum_{i=1}^n \tilde{h}(X_i) \right| &= O_P(n). \end{aligned}$$

The proof is similar to that of Lemma 3.2.1. Again we can combine the three statements above and get the following corollary as an immediate consequence.

Corollary 3.3.1 (Huse, 1988) *Under the same conditions as in Lemma 3.3.1 we have as $n \rightarrow \infty$*

$$\max_{1 \leq k \leq n} \left| U_k - \left\{ (n-k) \sum_{i=1}^k \tilde{h}(X_i) + k \left(\sum_{i=1}^n \tilde{h}(X_i) - \sum_{i=1}^k \tilde{h}(X_i) \right) \right\} \right| = O_P(n).$$

We shall see that the limit of $\{U_k, 1 \leq k < n\}$, as $n \rightarrow \infty$, is a Gaussian process, which is identified in the next theorem. Let $\{\Gamma(t), 0 \leq t \leq 1\}$ be a Gaussian process defined by

$$\Gamma(t) = (1-t)W(t) + t(W(1) - W(t)), \quad 0 \leq t \leq 1, \quad (3.3.1)$$

where W is a standard Wiener process. Since

$$\mathbb{E}W(t) = 0, \quad 0 \leq t \leq 1,$$

we have that

$$\begin{aligned}\mathbb{E}\Gamma(t) &= \mathbb{E}\big((1-t)W(t) + t(W(1) - W(t))\big) \\ &= (1-t)\mathbb{E}W(t) + t(\mathbb{E}W(1) - \mathbb{E}W(t)) \\ &= 0, \quad 0 \leq t \leq 1.\end{aligned}$$

By using the fact that

$$\text{Var}W(t) = t, \quad 0 \leq t \leq 1,$$

the variance of this Gaussian process is

$$\begin{aligned}\text{Var}\Gamma(t) &= \text{Var}\big((1-t)W(t) + t(W(1) - W(t))\big) \\ &= (1-t)^2\text{Var}W(t) + t^2(\text{Var}W(1) - \text{Var}W(t)) \\ &= (1-t)^2t + t^2(1-t) \\ &= (1-t)t, \quad 0 \leq t \leq 1.\end{aligned}$$

Let $\{B(t), 1 \leq t \leq 1\}$ be a Brownian bridge. Then we have that

$$\begin{aligned}\mathbb{E}\Gamma(t) &= \mathbb{E}B(t), \\ \text{Var}\Gamma(t) &= \text{Var}B(t),\end{aligned}$$

where $0 \leq t \leq 1$. Although these expected values and variances are the same, the two stochastic processes are not the same. This can be seen by checking the covariance function of these stochastic processes. Recall that a Gaussian process is uniquely determined by its covariance structure. By calculating the covariance function of the Brownian bridge $\{B(t), 1 \leq t \leq 1\}$ and using the facts that

$$B(t) = W(t) - tW(1), \quad 0 \leq t \leq 1,$$

and

$$\mathbb{E}W(s)W(t) = s \wedge t, \quad 0 \leq s, t \leq 1,$$

we get that

$$\begin{aligned}
 \text{Cov}[B(s), B(t)] &= \mathbb{E}B(s)B(t) \\
 &= \mathbb{E}(W(s) - sW(1))(W(t) - tW(1)) \\
 &= \mathbb{E}(W(s)W(t) - sW(1)W(t) - tW(s)W(1) + stW^2(1)) \\
 &= \mathbb{E}W(s)W(t) - s\mathbb{E}W(1)W(t) - t\mathbb{E}W(s)W(1) \\
 &\quad + st\mathbb{E}W^2(1) \\
 &= s \wedge t - s(1 \wedge t) - t(s \wedge 1) + st(1 \wedge 1) \\
 &= s \wedge t - st - ts + st \\
 &= s \wedge t - st, \quad 0 \leq s, t \leq 1.
 \end{aligned} \tag{3.3.2}$$

Similarly, we get the covariance function of our Gaussian process $\{\Gamma(t), 0 \leq t \leq 1\}$, namely

$$\begin{aligned}
 \text{Cov}[\Gamma(s), \Gamma(t)] &= \mathbb{E}\Gamma(s)\Gamma(t) \\
 &= \mathbb{E}\left((1-s)W(s) + s(W(1) - W(s))\right) \\
 &\quad \times \left((1-t)W(t) + t(W(1) - W(t))\right) \\
 &= \mathbb{E}\left((1-s)(1-t)W(s)W(t) + s(1-t)W(t)\right. \\
 &\quad \times (W(1) - W(s)) + (1-s)tW(s) \\
 &\quad \times (W(1) - W(t)) + st(W(1) - W(s))(W(1) - W(t))\left.)\right) \\
 &= (1-s)(1-t)\mathbb{E}W(s)W(t) + s(1-t)(\mathbb{E}W(t)W(1) \\
 &\quad - \mathbb{E}W(t)W(s)) + (1-s)t(\mathbb{E}W(s)W(1) \\
 &\quad - \mathbb{E}W(s)W(t)) + st(\mathbb{E}W^2(1) - \mathbb{E}W(s)W(1) \\
 &\quad - \mathbb{E}W(1)W(t) + \mathbb{E}W(s)W(t)) \\
 &= (1-s)(1-t)(s \wedge t) + s(1-t)((t \wedge 1) - (t \wedge s))
 \end{aligned}$$

$$\begin{aligned}
& + (1-s)t((s \wedge 1) - (s \wedge t)) \\
& + st((1 \wedge 1) - (s \wedge 1) - (1 \wedge t) + (s \wedge t)) \\
= & (1-s)(1-t)(s \wedge t) + s(1-t)(t - (t \wedge s)) \\
& + (1-s)t(s - (s \wedge t)) + st(1-s-t+(s \wedge t)) \\
= & (s \wedge t)((1-s)(1-t) - s(1-t) - (1-s)t + st) \\
& + s(1-t)t + (1-s)ts + st(1-s-t) \\
= & (1-2s)(1-2t)(s \wedge t) + (3-2s-2t)st, \\
& 0 \leq s, t \leq 1.
\end{aligned} \tag{3.3.3}$$

Since the covariance functions of the Gaussian processes $\{\Gamma(t), 0 \leq t \leq 1\}$ and $\{B(t), 0 \leq t \leq 1\}$ are not the same, the two processes are also not the same. Hence, we are now in the position to state the next theorem which is due to Csörgő and Horváth (1988b, 1997).

Theorem 3.3.1 (Csörgő and Horváth, 1988b) *Assume that H_0 , (2.2.2), (3.1.4) and (3.1.10) hold. Then we can define a sequence of Gaussian processes $\{\Gamma_n(t), 0 \leq t \leq 1\}_{n \in \mathbb{N}}$ such that, as $n \rightarrow \infty$,*

$$\sup_{0 < t < 1} |U_n(t) - \Gamma_n(t)| = o_P(1), \tag{3.3.4}$$

and, for each n , we have

$$\{\Gamma_n(t), 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \{\Gamma(t), 0 \leq t \leq 1\}. \tag{3.3.5}$$

A proof is given by Csörgő and Horváth (1997, Theorem 2.4.1). They use Corollary 3.3.1 and the fact that we can define a Wiener process $\{W(t), 0 \leq t < \infty\}$ such that

$$\frac{1}{n^{1/2}} \sup_{0 \leq t \leq 1} \left| \frac{1}{\bar{\sigma}} \sum_{i=1}^{[(n+1)t]} \tilde{h}(X_i) - W(nt) \right| = o_P(1).$$

Theorem 3.3.1 implies that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |U_n(t)| \xrightarrow{\mathcal{D}} \sup_{0 < t < 1} |\Gamma(t)|, \quad (3.3.6)$$

that is to say the supremum of $|U_n(t)|$ converges in distribution to the supremum of the absolute value of the Gaussian process $\Gamma(\cdot)$ as in (3.3.1). One can compute percentiles of its distribution function by producing tables for the latter random variable. Then we can use those tables to reject H_0 for large values of $\sup_{0 < t < 1} |U_n(t)|$.

Tests based on $\sup_{0 < t < 1} |U_n(t)|$ are not sensitive on the tails. Hence, we note, that a weighted version of Theorem 3.3.1 holds true as in Theorem 2.1 of Szyszkowicz (1991) if we replace $\frac{S(\lfloor (n+1)t \rfloor) - tS(n)}{\sigma}$ by $U_n(t)$ and the Brownian bridges by the Gaussian process $\Gamma_n(\cdot)$. In particular, using the weight function

$$q(t) = \left(t(1-t) \log \log \frac{1}{t(1-t)} \right)^{1/2},$$

we have, for example, that, as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} \frac{|U_n(t)|}{q(t)} \xrightarrow{\mathcal{D}} \sup_{0 < t < 1} \frac{|\Gamma(t)|}{q(t)}.$$

Producing tables for the distribution function of the random variable $\sup_{0 < t < 1} \frac{|\Gamma(t)|}{q(t)}$ is desirable for use in testing for H_0 .

3.3.2 Asymptotic Results under H_A

Csörgő and Horváth (1988b, Theorem 3.1) studied the asymptotic behavior of our U-statistics based process under the alternative H_A and showed the following theorem.

Theorem 3.3.2 (Csörgő and Horváth, 1988b) *Assume that (2.2.2), H_A hold and*

$$\mathbb{E}\{|h(X_{[n\lambda]}, X_{[n\lambda]+1})| \log^+ (|h(X_{[n\lambda]}, X_{[n\lambda]+1})|)\} < \infty, \quad (3.3.7)$$

where $\tau = \tau(n) := [n\lambda]$, $0 < \lambda < 1$, and $\log^+ x = \log(x \vee 1)$. Then, with Z_k as in (3.1.1), we have, as $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t]} \xrightarrow{P} u_\lambda(t) = \begin{cases} t(\lambda - t)\theta_1 + t(1 - \lambda)\theta_{1,2}, & 0 < t \leq \lambda, \\ (t - \lambda)(1 - t)\theta_2 + (1 - t)\lambda\theta_{1,2}, & \lambda \leq t < 1. \end{cases}$$

Again we mention that the theorem also holds, if condition (3.3.7) is replaced by condition (3.1.19). Since the way of proving this theorem will be of interest when having an alternative of more than one changes, we give a proof of the theorem. It is similar to the ones given by Csörgő and Horváth (1986, Theorem 3.1, and 1988b, Theorem 3.1) and Huse (1988, Theorem 2.3.7).

Proof of Theorem 3.3.2 We have $\tau = \tau(n) = [n\lambda]$, $1 \leq \tau < n$, the single change-point under the alternative H_A . We put $m = [(n+1)t]$ and first assume that $1 \leq m \leq \tau$.

In view of Theorem 2.6.1 by Hoeffding (1961) and Theorem 2.6.2 by Sen (1977), we have to change the summation of Z_m , $1 \leq m \leq \tau$, such that we can apply these theorems on U-statistics and generalized U-statistics. We have

$$\begin{aligned} Z_m &= \sum_{i=1}^m \sum_{j=m+1}^n h(X_i, X_j) \\ &= \sum_{i=1}^m \sum_{j=m+1}^{\tau} h(X_i, X_j) + \sum_{i=1}^m \sum_{j=\tau+1}^n h(X_i, X_j) \\ &= \sum_{i=1}^m \sum_{j=m+1}^{\tau} h(X_i, X_j) + \left(\sum_{i=1}^{\tau} \sum_{j=\tau+1}^n h(X_i, X_j) - \sum_{i=m+1}^{\tau} \sum_{j=\tau+1}^n h(X_i, X_j) \right). \end{aligned} \tag{3.3.8}$$

We have just split Z_m into three parts, where the last two are now of the forms of generalized U-statistics. We also have to rewrite the first part for the sake of applying any of the previously mentioned theorems. To do this, we look at the summation areas of the following sums

$$A_n = \sum_{i=1}^m \sum_{j=m+1}^{\tau},$$

Figure 3.3.2: Summation Area

$$\begin{aligned}
 A_n^{(1)} &= \sum_{1 \leq i < j \leq \tau}, \\
 A_n^{(2)} &= \sum_{1 \leq i < j \leq m}, \\
 A_n^{(3)} &= \sum_{m+1 \leq i < j \leq \tau},
 \end{aligned}$$

where A_n is the summation area we want to change. Figure 3.3.2 shows the areas $A_n, A_n^{(2)}$ and $A_n^{(3)}$ in comparison to that of $A_n^{(1)}$ and we can see that

$$A_n + A_n^{(2)} + A_n^{(3)} = A_n^{(1)}, \text{ and hence } A_n = A_n^{(1)} - (A_n^{(2)} + A_n^{(3)}).$$

Note that now $A_n^{(1)}, A_n^{(2)}$ and $A_n^{(3)}$ are U-statistics without their respective normalizing factors. In these U-statistics the underlying random variables are from the same distribution. Therefore, (3.3.8) becomes

$$\begin{aligned}
 Z_m &= \left(\sum_{1 \leq i < j \leq \tau} h(X_i, X_j) - \left(\sum_{1 \leq i < j \leq m} h(X_i, X_j) + \sum_{m+1 \leq i < j \leq \tau} h(X_i, X_j) \right) \right) \\
 &\quad + \left(\sum_{i=1}^{\tau} \sum_{j=\tau+1}^n h(X_i, X_j) - \sum_{i=m+1}^{\tau} \sum_{j=\tau+1}^n h(X_i, X_j) \right) \\
 &= \left(A_n^{(1)} - (A_n^{(2)} + A_n^{(3)}) \right) + \left(A_n^{(4)} - A_n^{(5)} \right), \tag{3.3.9}
 \end{aligned}$$

where $A_n^{(4)}$ and $A_n^{(5)}$ denote the last two summands in (3.3.8). We mention again that, except for their missing normalizing factors (see (2.3.3)), $A_n^{(1)}, A_n^{(2)}$ and $A_n^{(3)}$ are (non-degenerate) one-sample U-statistics and $A_n^{(4)}$ and $A_n^{(5)}$ are generalized two-sample

U-statistics. Hence, Hoeffding's Strong Law of Large Numbers (SLLN) applies and, as $n \rightarrow \infty$, it yields

$$\frac{1}{\binom{\tau}{2}} \sum_{1 \leq i < j \leq \tau} h(X_i, X_j) = \frac{2}{[n\lambda]([n\lambda] - 1)} \sum_{1 \leq i < j \leq [n\lambda]} h(X_i, X_j) \xrightarrow{a.s.} \theta_1. \quad (3.3.10)$$

This together with

$$\frac{[n\lambda]([n\lambda] - 1)}{n^2} \rightarrow \lambda^2, \quad n \rightarrow \infty,$$

implies

$$\frac{1}{n^2} A_n^{(1)} \xrightarrow{a.s.} \frac{\lambda^2 \theta_1}{2}, \quad n \rightarrow \infty. \quad (3.3.11)$$

Similarly, we obtain that

$$\frac{1}{n^2} A_n^{(2)} \xrightarrow{a.s.} \frac{t^2 \theta_1}{2}, \quad n \rightarrow \infty. \quad (3.3.12)$$

When looking at $A_n^{(3)}$, we see that we can not immediately apply Hoeffding's SLLN theorem, since the summation is taken over $m+1 \leq i < j \leq \tau$ and we would need a 1 to start with, instead of $m+1$. But since $X_1, \dots, X_\tau$ are i.i.d. r.v.'s, we have

$$\sum_{m+1 \leq i < j \leq \tau} h(X_i, X_j) \stackrel{D}{=} \sum_{1 \leq i < j \leq \tau-m} h(X_i, X_j), \quad (3.3.13)$$

where $\tau = \tau(n) := [n\lambda]$ and $m = m(n) := [(n+1)t]$. Using now Hoeffding's SLLN theorem and taking (3.3.13) into consideration we get the following convergence in probability result:

$$\frac{1}{n^2} A_n^{(3)} \xrightarrow{P} \frac{(\lambda - t)^2 \theta_1}{2}, \quad n \rightarrow \infty. \quad (3.3.14)$$

$A_n^{(4)}$ and $A_n^{(5)}$, are generalized two-sample U-statistics, except that the appropriate normalization factors are missing. Therefore Theorem 2.6.2 by Sen (1977) applies

and we get

$$\frac{1}{\tau(n-\tau)} \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n h(X_i, X_j) \xrightarrow{P} \theta_{1,2}, \quad n \rightarrow \infty. \quad (3.3.15)$$

Since $\frac{1}{\tau(n-\tau)} = \frac{1}{[n\lambda](n-[n\lambda])}$, we obtain

$$\frac{1}{n^2} A_n^{(4)} \xrightarrow{P} \lambda(1-\lambda)\theta_{1,2}, \quad n \rightarrow \infty, \quad (3.3.16)$$

and similarly, we get

$$\frac{1}{n^2} A_n^{(5)} \xrightarrow{P} (\lambda-t)(1-\lambda)\theta_{1,2}, \quad n \rightarrow \infty. \quad (3.3.17)$$

Finally, as $n \rightarrow \infty$, we arrive at

$$\begin{aligned} \frac{1}{n^2} Z_{[(n+1)t]} &= \frac{1}{n^2} \left((A_n^{(1)} - A_n^{(2)} - A_n^{(3)}) + (A_n^{(4)} - A_n^{(5)}) \right) \\ &\xrightarrow{P} \left(\left(\frac{\lambda^2\theta_1}{2} - \frac{t^2\theta_1}{2} - \frac{(\lambda-t)^2\theta_1}{2} \right) + (\lambda(1-\lambda)\theta_{1,2} - (\lambda-t)(1-\lambda)\theta_{1,2}) \right) \\ &= t(\lambda-t)\theta_1 + t(1-\lambda)\theta_{1,2}, \quad t \in [0, \lambda], \end{aligned} \quad (3.3.18)$$

and this proves the theorem for $1 \leq m \leq \tau$. The proof for $\tau < m \leq n$ is similar and hence omitted. $\square$

Theorem 3.3.2 can be used to study the consistency of tests based on $\{U_{[(n+1)t]}, 0 \leq t < 1\}$, where $U_k = Z_k - k(n-k)\theta$, $1 \leq k < n$. Assuming second finite moments as in (3.1.19) such that the results under H_0 hold, then we can consistently reject H_0 vs. H_A , when $\sup_{0 < t < 1} \frac{1}{n^{3/2\bar{\sigma}}} |U_{[(n+1)t]}|$ is large except in the case where $\theta_1 = \theta_2 = \theta_{1,2} = 0$.

3.3.3 Estimating the Time of Change

In Section 3.2.3 we considered an estimator for the time of change for an antisymmetric kernel. We either used the unique maximum or the unique minimum of Z_k , $1 \leq k < n$, to estimate the change-point $\tau = \tau(n) := [n\lambda]$, $0 < \lambda < 1$. In the symmetric case we will have similar results, but we will have to impose conditions³ on θ_1 , θ_2 , $\theta_{1,2}$ and λ , if we were to have that $u_\lambda(t)$ from Theorem 3.3.2 should have a unique maximum or minimum respectively at λ . To do this, we consider the function (cf. Theorem 3.3.2)

$$u_\lambda(t) = \begin{cases} t(1-\lambda)\theta_{1,2} + t(\lambda-t)\theta_1, & 0 < t \leq \lambda, \\ (1-t)\lambda\theta_{1,2} + (t-\lambda)(1-t)\theta_2, & \lambda \leq t < 1. \end{cases} \quad (3.3.19)$$

Differentiating with respect to t we get the derivative of $u_\lambda(t)$, namely

$$u'_\lambda(t) = \begin{cases} (1-\lambda)\theta_{1,2} + (\lambda-2t)\theta_1, & 0 < t \leq \lambda, \\ -\lambda\theta_{1,2} + (1+\lambda-2t)\theta_2, & \lambda \leq t < 1, \end{cases}$$

furthermore, we get the second derivative of $u_\lambda(t)$, i.e.,

$$u''_\lambda(t) = \begin{cases} -2\theta_1, & 0 < t \leq \lambda, \\ -2\theta_2, & \lambda \leq t < 1, \end{cases}$$

which tells us that $u_\lambda(t)$ is a concave function if θ_1 and θ_2 in (3.1.17) are positive. Recall that $\theta_1 \neq 0$ and $\theta_2 \neq 0$, since we are dealing with symmetric kernels. Using the fact that a concave function is a function of the form $\cap$, we have to search for a maximum. If both are negative, then $u_\lambda(t)$ is a convex function, which is a function of the form $\cup$, and we have to search for a minimum. So we define

$$\hat{\tau} = [n\hat{\lambda}] = \min \left\{ k : Z_k = \max_{1 \leq m < n} Z_m \right\}, \quad \text{if } \theta_1, \theta_2 > 0$$

³Csörgő and Horváth (1998, Section 2.4) and Ferger and Stute (1992) suggest to impose conditions, but do not give them explicitly.

and

$$\tilde{\tau} = [n\tilde{\lambda}] = \min \left\{ k : Z_k = \min_{1 \leq m < n} Z_m \right\}, \quad \text{if } \theta_1, \theta_2 < 0.$$

Let us first consider the case when θ_1 and θ_2 are positive. We know that

$$u_\lambda(0) = u_\lambda(1) = 0, \quad \text{and} \quad u_\lambda(\lambda) = \lambda(1 - \lambda)\theta_{1,2},$$

but we don't know, whether the maximum is taken at $t = \lambda$ or at $0 < t_{max_1} < \lambda$ and/or at $\lambda < t_{max_2} < 1$. It could happen that the maximum is taken at $t = t_{max_1}$ or $t = t_{max_2}$. Figure 3.3.3 shows the graph of the limiting function $u_\lambda(t)$ where the maximum is taken at $t = t_{max_2}$ and not at $t = \lambda$, and similarly, Figure 3.3.4 shows the the graph of $u_\lambda(t)$ when there are two local maxima at $t = t_{max_1}$ and $t = t_{max_2}$.

Figure 3.3.3: The limiting function $u_{\frac{1}{10}}(t)$ with $\theta_1 = 1$, $\theta_2 = 2$ and $\theta_{1,2} = 3$ takes its maximum value of 0.55 at $t = 0.475$.

Hence, we have to restrict ourselves to have a unique maximum at $t = \lambda$ only. Similarly, if θ_1 and θ_2 are negative, we have to restrict ourselves to have a unique minimum at $t = \lambda$ only.

Figure 3.3.4: The limiting function $u_{\frac{1}{2}}(t)$ with $\theta_1 = 3$, $\theta_2 = 2$ and $\theta_{1,2} = 1$ takes its maximum value of $\frac{1}{3}$ at $t = \frac{1}{3}$.

If $\theta_1 < 0 < \theta_2$ or $\theta_2 < 0 < \theta_1$ then we have a convex-concave or concave-convex function respectively. Hence, we are looking for a minimum or maximum respectively. Again, we have to restrict ourselves. We mention that in these cases there could be also a maximum or minimum respectively.

In view of finding a unique maximum we look at all possible extreme values of $u_\lambda(t)$. We get that

$$u'_\lambda(t) = 0, \text{ if } t = \begin{cases} t_{max_1} = \frac{(1-\lambda)\theta_{1,2}}{2\theta_1} + \frac{\lambda}{2}, & 0 < t \leq \lambda, \\ t_{max_2} = -\frac{\lambda\theta_{1,2}}{2\theta_2} + \frac{1+\lambda}{2}, & \lambda \leq t < 1. \end{cases} \quad (3.3.20)$$

Of course, we would like to have that $\lambda = t_{max_1} = t_{max_2}$. Hence, we have to investigate, if there is a maximum before and/or after the change-point λ , namely if

$$t_{max_1} < \lambda \text{ and/or } \lambda < t_{max_2}. \quad (3.3.21)$$

If now $\lambda < t_{max_2}$ is satisfied, then we have to show that the function value of t_{max_2} is

greater than the function value of λ . We calculate $u_\lambda(t_{max_2})$ by plugging the second part of (3.3.20) into (3.3.19). We consider

$$\begin{aligned}
u_\lambda(t_{max_2}) &= \lambda \left(1 - \left(-\frac{\lambda\theta_{1,2}}{2\theta_2} + \frac{1+\lambda}{2} \right) \right) \theta_{1,2} \\
&\quad + \left(-\frac{\lambda\theta_{1,2}}{2\theta_2} + \frac{1+\lambda}{2} - \lambda \right) \left(1 - \left(-\frac{\lambda\theta_{1,2}}{2\theta_2} + \frac{1+\lambda}{2} \right) \right) \theta_2 \\
&= \frac{\lambda(1-\lambda)}{2} \theta_{1,2} + \frac{\lambda^2\theta_{1,2}^2}{2\theta_2} + \frac{(1-\lambda)^2}{4} \theta_2 \\
&\quad - \frac{\lambda(1-\lambda)}{4} \theta_{1,2} + \frac{\lambda(1-\lambda)}{4} \theta_{1,2} - \frac{\lambda^2\theta_{1,2}^2}{4\theta_2} \\
&= \frac{\lambda^2\theta_{1,2}^2}{4\theta_2} + \frac{\lambda(1-\lambda)}{2} \theta_{1,2} + \frac{(1-\lambda)^2}{4} \theta_2 \\
&= \frac{1}{4\theta_2} (\lambda\theta_{1,2} + (1-\lambda)\theta_2)^2,
\end{aligned}$$

and show that the following inequality holds,

$$\begin{aligned}
u_\lambda(t_{max_2}) &\geq u_\lambda(\lambda) \\
\frac{1}{4\theta_2} (\lambda\theta_{1,2} + (1-\lambda)\theta_2)^2 &\geq \lambda(1-\lambda)\theta_{1,2} \\
\frac{1}{4\theta_2} (\lambda\theta_{1,2} + (1-\lambda)\theta_2)^2 &\geq \frac{1}{4\theta_2} (4\lambda(1-\lambda)\theta_2\theta_{1,2}) \\
\frac{1}{4\theta_2} (\lambda\theta_{1,2} - (1-\lambda)\theta_2)^2 &\geq 0,
\end{aligned} \tag{3.3.22}$$

where we assume that $\theta_2 > 0$. Otherwise the inequality is never satisfied, except when we are looking for the minimum. We have an '=' sign in (3.3.22), if

$$\lambda = \frac{\theta_2}{\theta_2 + \theta_{1,2}}.$$

Similarly, we get that

$$u_\lambda(t_{max_1}) = \frac{1}{4\theta_1} (\lambda\theta_1 + (1-\lambda)\theta_{1,2})^2$$

and

$$\begin{aligned} u_\lambda(t_{max_1}) &\geq u_\lambda(\lambda) \\ \frac{1}{4\theta_1}(\lambda\theta_1 + (1-\lambda)\theta_{1,2})^2 &\geq \lambda(1-\lambda)\theta_{1,2} \\ \frac{1}{4\theta_1}(\lambda\theta_1 - (1-\lambda)\theta_{1,2})^2 &\geq 0, \end{aligned} \quad (3.3.23)$$

where we assume that $\theta_1 > 0$. We have an '=' sign in (3.3.23), if

$$\lambda = \left(1 - \frac{\theta_1}{\theta_1 + \theta_{1,2}}\right).$$

By using the definition of t_{max_1} and t_{max_2} in (3.3.20), (3.3.21) leads to

$$1 - \frac{\theta_1}{\theta_1 + \theta_{1,2}} < \lambda \quad \text{and/or} \quad \lambda < \frac{\theta_2}{\theta_2 + \theta_{1,2}}. \quad (3.3.24)$$

If (3.3.24) is not satisfied, then $u_\lambda(\lambda)$ is the unique maximum. Therefore we have to restrict ourselves to the following choice of λ , namely

$$\lambda \in \left[\frac{\theta_2}{\theta_2 + \theta_{1,2}}, 1 - \frac{\theta_1}{\theta_1 + \theta_{1,2}} \right], \quad (3.3.25)$$

where $0 < \theta_1 < \infty$, $0 < \theta_2 < \infty$ and $|\theta_{1,2}| < \infty$, to have a unique maximum at λ .

If $\theta_1 < 0 < \theta_2$ or $\theta_2 < 0 < \theta_1$ then we may replace the right endpoint in (3.3.25) by 1 or the left endpoint by 0 respectively. Furthermore, (3.3.25) needs to hold true when $0 > \theta_1 > -\infty$, $0 > \theta_2 > -\infty$ and $|\theta_{1,2}| < \infty$, to have a unique minimum at λ .

Using these conditions on θ_1 , θ_2 , $\theta_{1,2}$ and λ from (3.3.25), we are now in the position to state a theorem by Ferger and Stute (1992) for symmetric kernel functions that are also assumed to be bounded.

Theorem 3.3.3 (Ferger and Stute, 1992) *Under suitable assumptions⁴ on the symmetric (or antisymmetric) bounded kernel function $h(x, y)$, such that λ is the unique*

⁴Note that these assumptions, which were not given by Ferger and Stute (1992) explicitly, are the ones in (3.3.25) when $0 < \theta_1 < \infty$, $0 < \theta_2 < \infty$ and $|\theta_{1,2}| < \infty$. As mentioned above, similar versions of (3.3.25) hold when θ_1, θ_2 and $\theta_{1,2}$ are chosen in a different way.

Figure 3.3.5: The limiting function $u_{\frac{7}{10}}(t)$ with $\theta_1 = 1$, $\theta_2 = 2$ and $\theta_{1,2} = 3$ takes its maximum value of 0.63 at $t = \lambda = \frac{7}{10}$.

maximizer (minimizer) of u_λ , and $u'_\lambda(t) \neq 0$, $t \neq \lambda$, we have that, as $n \rightarrow \infty$,

$$|\hat{\lambda} - \lambda| \stackrel{a.s.}{=} O\left(\frac{\log n}{n}\right),$$

which implies that $\hat{\lambda}$ is a strongly consistent estimator of λ . The same is true for $\tilde{\lambda}$.

Summarizing, we need to impose the special conditions from (3.3.25) on the bounded kernel h via θ_1 , θ_2 and $\theta_{1,2}$, and we will have a strongly consistent estimator for the time of change (cf. Theorem 3.3.3). In practice, we have to check whether the maximum or minimum respectively is taken in the latter interval. Moreover, since the indicated expected values of h are usually unknown, i. e., θ_1 , θ_2 and $\theta_{1,2}$ are unknown, we have to estimate them from the data by using appropriate estimators. If in (3.3.25) $\theta_{1,2} \rightarrow \infty$, i. e., when $\theta_{1,2}$ becomes very large (by our assumptions θ_1 , θ_2 and $\theta_{1,2}$ are all finite), then the interval in (3.3.25) converges to $[0, 1]$. Then the estimators $\hat{\tau}$ and $\tilde{\tau}$ work out fine. If on the other hand $\theta_1 \rightarrow \infty$ or $\theta_2 \rightarrow \infty$, then the interval in (3.3.25) converges to either $[0, 0]$ or $[1, 1]$, which means that

we can not use the estimators at all. Therefore, the optimal case is when we have $|\theta_1|, |\theta_2| \ll |\theta_{1,2}| < \infty$. If $\theta_{1,2}$ is between the others, then the interval will just cover a small area on the left or right half of the interval $[0, 1]$.

Gombay (1998) defines

$$\bar{\tau} = \bar{\tau}(n) = \min \{k : U_k = \max_{1 \leq m < n} U_m\} \quad (3.3.26)$$

as an estimator of τ and considered the distribution of $\bar{\tau}$ under H_A . She shows, for example, that for a symmetric non-degenerate kernel with finite second moment and under some technical assumptions, as $n \rightarrow \infty$, under H_A we have

$$|\tau(n) - \bar{\tau}(n)| = O_P(1),$$

and

$$\frac{\max_{1 \leq k < n} U_k - \tau(n - \tau)(\theta_{1,2} - \theta_1)}{n^{3/2} \left(\text{Var}(\mathbb{E}h(X_{\tau+1}, X_1))\lambda(1 - \lambda)^2 + \text{Var}(\mathbb{E}h(X_1, X_{\tau+1}))\lambda^2(1 - \lambda) \right)^{1/2}} \xrightarrow{\mathcal{D}} N(0, 1).$$

Moreover, she gives the distribution of $\bar{\tau} - \tau$ which depends on the underlying distribution function and on the change-point parameter λ . It behaves like the maximum of a two-sided random walk. In case of an antisymmetric kernel function h , Gombay (1998) shows that, under H_A , $\max_{1 \leq k < n} \bar{U}_k$ (cf. (3.1.12)) has the same limiting distribution as $\max_{1 \leq k < n} U_k$.

Similarly to the antisymmetric case, we would like to find the distribution of $\hat{\tau}$ under H_0 , the null hypothesis of no-change. Since U_k is defined by

$$U_k = Z_k - k(n - k)\theta, \quad 1 \leq k < n, \quad (3.3.27)$$

Z_k , $1 \leq k < n$, reaches its maximum when $U_k + k(n - k)\theta$ reaches its maximum. We define

$$Z_n(t) := \frac{1}{\bar{\sigma}n^{3/2}} Z_{[(n+1)t]}, \quad 0 < t < 1,$$

and get by using (3.3.27) that

$$\begin{aligned} Z_n(t) &= U_n(t) + \frac{[(n+1)t](n - [(n+1)t])\theta}{\bar{\sigma}n^{3/2}} \\ &\approx U_n(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta, \quad 0 < t < 1. \end{aligned}$$

Moreover, taking the sup on both sides,

$$\sup_{0 < t < 1} |Z_n(t)| \approx \sup_{0 < t < 1} \left| U_n(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta \right|. \quad (3.3.28)$$

But from Theorem 3.3.1 we know that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |U_n(t) - \Gamma_n(t)| = o_P(1), \quad (3.3.29)$$

which implies that $\sup_{0 < t < 1} |U_n(t)|$ behaves like $\sup_{0 < t < 1} |\Gamma(t)|$, where $\Gamma(t)$ is the Gaussian process from (3.3.1). Of course, we may add and subtract a term in (3.3.29) and get that, as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} \left| \left(U_n(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta \right) - \left(\Gamma_n(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta \right) \right| = o_P(1).$$

The latter statement implies that for n large

$$\left| \sup_{0 < t < 1} \left| U_n(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta \right| - \sup_{0 < t < 1} \left| \Gamma(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta \right| \right| = o_P(1).$$

Consequently, via (3.3.28) and the latter statement we conclude that $|Z_n(t)|$ in t reaches its maximum for large n where $|\Gamma(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta|$ does. But the latter expression goes to ∞ as $n \rightarrow \infty$, since $\frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta \rightarrow \infty$. Therefore, we do not know where the maximum of $Z_n(t)$ is taken. We may also think about using a weighted version of Theorem 3.3.1. Indeed, using an appropriate weight function q in (3.3.28), we get that

$$\sup_{0 < t < 1} \left| \frac{Z_n(t)}{q(t)} \right| \approx \sup_{0 < t < 1} \left| \frac{U_n(t) + \frac{n^{1/2}}{\bar{\sigma}}t(1-t)\theta}{q(t)} \right| \quad (3.3.30)$$

and as an estimation of τ we now define

$$\hat{\tau}_q = \min \left\{ k : \frac{Z_k}{q(\frac{k}{n})} = \max_{1 \leq m < n} \frac{Z_m}{q(\frac{m}{n})} \right\}.$$

For example, using the weight function

$$q(t) = \left(t(1-t) \log \log \frac{1}{t(1-t)} \right)^{1/2}$$

we still have that

$$\frac{\frac{n^{1/2}}{\bar{\sigma}} t(1-t)\theta}{\left(t(1-t) \log \log \frac{1}{t(1-t)} \right)^{1/2}} \rightarrow \infty, \quad \text{as } n \rightarrow \infty,$$

and therefore, as before, (3.3.30) goes to ∞ as $n \rightarrow \infty$. Hence, in order to estimate τ , it seems that we need to centralize Z_k , $1 \leq k < \infty$, by its mean, i. e., we have to work with U_k as in (3.3.27). This, however, will not give us the desired result anymore, since

$$\max_{1 \leq k < n} U_k \neq \max_{1 \leq k < n} Z_k. \quad (3.3.31)$$

In case of an antisymmetric kernel we had an '=' sign in (3.3.31) which made it possible to use Theorem 3.2.1 and get the distribution of $\hat{\tau}$ under H_0 . But in case of a symmetric kernel the fact that (3.3.31) holds does not allow us to use Theorem 3.3.1 to give the distribution of $\hat{\tau}$ under H_0 . Hence, the distribution of $\hat{\tau}$ under H_0 is still unknown.

In principle, we may, however, compute the distribution of $\tilde{\tau}$ in (3.3.26) under H_0 , since (3.3.29) implies that $|U_n(t)|$ reaches its maximum where $|\Gamma(t)|$ does. Hence $\frac{\tilde{\tau}}{n}$ converges to the distribution of the argument where $\sup_{0 < t < 1} |\Gamma(t)|$ is taken.

3.4 Change in the Mean

We are to test the *no-change in the mean* null-hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables
with $\mathbb{E}X_i = \mu$ and $0 < \sigma^2 = \text{Var}X_i < \infty$, $1 \leq i \leq n$,

against the *at most one change in the mean* alternative

$H_A : X_1, \dots, X_n$ are independent random variables and there is an integer τ , $1 \leq \tau < n$, such that $\mathbb{E}X_1 = \dots = \mathbb{E}X_\tau \neq \mathbb{E}X_{\tau+1} = \dots = \mathbb{E}X_n$ and $0 < \sigma^2 = \text{Var}X_1 = \dots = \text{Var}X_n < \infty$.

Taking simulated values of the therein indicated (k, X_k) , Figure 3.4.6 gives an example where a change in the mean occurs while the variance stays the same.

Figure 3.4.6: The data $X_1, \dots, X_{700}$ are i.i.d. $N(1, 1)$ -distributed and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(4, 1)$ -distributed.

If we define the *antisymmetric* kernel

$$h(x, y) = x - y, \quad (3.4.1)$$

then the stochastic process Z_k from (3.1.1) may be written as

$$\begin{aligned} \sum_{i=1}^k \sum_{j=k+1}^n (X_i - X_j) &= (n-k) \sum_{i=1}^k X_i - k \sum_{j=k+1}^n X_j \\ &= k(n-k) \left(\frac{1}{k} \sum_{i=1}^k X_i - \frac{1}{n-k} \sum_{j=k+1}^n X_j \right), \quad 1 \leq k < n, \end{aligned}$$

which may be interpreted as comparing the mean before the unknown time k , $1 \leq k < n$, of a possible change in the mean to the mean after change. Under H_0 this difference should be fluctuating near around zero, while under the alternative H_A of one change, the stochastic process Z_k , $1 \leq k < n$, will have a maximum or minimum at the time $k = \tau$.

When testing for more change-points we will use a different argument which in our present situation is similar to the following. Observe that under H_0

$$\begin{aligned} \mathbb{E}\{S(k)\} &= \mathbb{E} \sum_{i=1}^k X_i \\ &= \sum_{i=1}^k \mathbb{E} X_i \\ &= k\mu, \quad 1 \leq k \leq n, \end{aligned}$$

then we will see that testing for a change in the mean can be illustrated by using a geometrical argument. Consider the linear function $m(t) := t$, $t \in \mathbf{R}$, which joins under H_0 all the points $(k, \frac{1}{\mu} \mathbb{E}\{S(k)\})$, $k \in \mathbf{N}$, if $\mu \neq 0$, and it joins all the points $(k, \mathbb{E}\{S(k)\})$, $k \in \mathbf{N}$, if $\mu = 0$.

Without loss of generality let $\mu = 1$. Then in Figure 3.4.7 we join all the points $(k, \mathbb{E}\{S(k)\})$, $k \in \mathbf{N}$, via the straight line $m(t) = t$. We pick one $k \in \{1, \dots, n-1\}$,⁵ and draw a horizontal line starting from $B := (0, \mathbb{E}\{S(k)\})$, containing the point $(k, \mathbb{E}\{S(k)\})$, and with terminus, $C := (n, \mathbb{E}\{S(k)\})$. We draw a vertical line from the terminus and intersect the t -axis. We denote this intersection by $D := (n, \mathbb{E}\{S(0)\})$, where we define $S(0) := 0$. In this way we construct a rectangle,

⁵In Figure 3.4.7 we may think of k being defined as $k := \lceil nt^{(fix)} \rceil$, $0 < t^{(fix)} < 1$.

Figure 3.4.7: A geometrical interpretation of $\mathbb{E}\{nS(k) - kS(n)\} = 0$ under H_0 .

denoted by $ABCD$ (see Figure 3.4.7), where $A := (0, \mathbb{E}\{S(0)\})$, with length n and height $\mathbb{E}\{S(k)\}$.

Reflecting each point of the rectangle $ABCD$ around the 45 degree line $m = t$, we get a new rectangle $AEFG$, where $A := (0, \mathbb{E}\{S(0)\})$ is the reflection point of itself, $E := (0, \mathbb{E}\{S(n)\})$ is the reflection point of $D := (n, \mathbb{E}\{S(0)\})$, $F := (k, \mathbb{E}\{S(n)\})$ is the reflection point of $C := (n, \mathbb{E}\{S(k)\})$ and $G := (k, \mathbb{E}\{S(0)\})$ is the reflection point of $B := (0, \mathbb{E}\{S(k)\})$, with length k and height $\mathbb{E}\{S(n)\}$. Under H_0 , both rectangles have the same area. Consequently, we have that

$$\begin{aligned}
 \mathbb{E}\{S(k)\} \times (n - k) \times \mathbb{E}\{S(n)\} &= \mathbb{E}\{nS(k) - kS(n)\} \\
 &= \mathbb{E}\{(n - k)S(k) - k(S(n) - S(k))\} \\
 &= \mathbb{E} \sum_{i=1}^k \sum_{j=k+1}^n (X_i - X_j) \\
 &= 0, \quad 1 \leq k < n.
 \end{aligned} \tag{3.4.2}$$

Thus, in principle, for each given k , $1 \leq k < n$, we constructed an unbiased estimator of zero assuming that H_0 is true. We may also say that, viewed this way, testing for one change in the mean results in comparing areas of two different rectangles with each other.

Instead of defining $m(t) = t$ we may also define $m_\mu(t) := \mu t$ which joins all the points $(k, \mathbb{E}\{S(k)\})$, $k \in \mathbb{N}$. In a similar vein as before we get the same results. Moreover, under H_A the slope of the function m_μ will change exactly at k if $\mathbb{E}\{S(k)\} = \mathbb{E}\{S(\tau)\} = \mu\tau$. Hence the two areas will not be the same. Moreover, the maximal difference between those two areas will be when $k = \tau$.

Continuing with using the kernel $h(x, y) = x - y$ to test for at most one change in the mean, under H_0 and $1 \leq i, j \leq n$,

$$\mathbb{E}h(X_i, X_j) = \mathbb{E}(X_i - X_j) = \mathbb{E}X_i - \mathbb{E}X_j = 0.$$

Furthermore, assuming that

$$\mathbb{E}h^2(X_i, X_j) < \infty, \tag{3.4.3}$$

we have

$$\begin{aligned} \mathbb{E}h^2(X_i, X_j) &= \mathbb{E}(X_i - X_j)^2 \\ &= \mathbb{E}(X_i^2 - 2X_iX_j + X_j^2) \\ &= \mathbb{E}X_i^2 - 2\mathbb{E}X_iX_j + \mathbb{E}X_j^2 \\ &= 2\mathbb{E}X_i^2 - 2\mathbb{E}X_i\mathbb{E}X_i \\ &= 2\text{Var}X_i \\ &= 2\sigma^2, \end{aligned}$$

i. e., we assume that the variance σ^2 is finite. We define

$$\begin{aligned} \tilde{h}(t) &= \mathbb{E}\{h(t, X_2) - 0\} \\ &= \mathbb{E}\{t - X_2\} \end{aligned}$$

$$= t - \mu.$$

Since under H_0

$$\mathbb{E}\tilde{h}^2(X_1) = \mathbb{E}\{(X_1 - \mu)^2\} = \sigma^2,$$

it follows from (3.4.3) that

$$\tilde{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) = \sigma^2 < \infty,$$

and, in addition to (3.4.3), we also assume that

$$\tilde{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) = \sigma^2 > 0.$$

Since h is an *antisymmetric* kernel, we know from Theorem 3.2.1 that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 \leq t \leq 1} |\bar{U}_n(t) - B_n(t)| = o_P(1), \quad (3.4.4)$$

where for all $0 \leq t \leq 1$

$$\begin{aligned} \bar{U}_n(t) &= \frac{1}{\tilde{\sigma}n^{3/2}} \bar{U}_{[(n+1)t]} \\ &= \frac{1}{\sigma n^{3/2}} Z_{[(n+1)t]} \\ &= \frac{1}{\sigma n^{3/2}} \sum_{1 \leq i \leq [(n+1)t]} \sum_{[(n+1)t] < j \leq n} h(X_i, X_j) \\ &= \frac{1}{\sigma n^{3/2}} \sum_{1 \leq i \leq [(n+1)t]} \sum_{[(n+1)t] < j \leq n} (X_i - X_j). \end{aligned} \quad (3.4.5)$$

We recall also for each n

$$\{B_n(t), 0 \leq t \leq 1\} \stackrel{\mathcal{D}}{=} \{B(t), 0 \leq t \leq 1\},$$

with B being a standard Brownian bridge.

Assuming the same conditions as in Theorem 3.2.2, we obtain that under H_A , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |\bar{U}_n(t)| \xrightarrow{P} \infty.$$

Therefore we consistently reject H_0 , if $\sup_{0 < t < 1} |\bar{U}_n(t)|$ becomes too big and we have to think about ‘what is too big?’. On account of (3.4.4) it follows that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |\bar{U}_n(t)| \xrightarrow{D} \sup_{0 < t < 1} |B(t)|.$$

This in turn means that we can use tables for the supremum of the absolute value of a Brownian bridge to accept or reject H_0 .

Since the latter test statistic is not sensitive on the tails, we may also use a weighted version of (3.4.4) by using, for example, the weight function

$$q(t) = \left(t(1-t) \log \log \frac{1}{t(1-t)} \right)^{1/2}.$$

Consequently, with this q , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} \frac{|\bar{U}_n(t)|}{q(t)} \xrightarrow{D} \sup_{0 < t < 1} \frac{|B(t)|}{q(t)}$$

and now we may use tables (cf. Eastwood and Eastwood (1998)) for $\sup_{0 < t < 1} \frac{|B(t)|}{q(t)}$ to reject H_0 for large values of $\sup_{0 < t < 1} \frac{|\bar{U}_n(t)|}{q(t)}$.

The variance σ^2 in (3.4.5) is usually unknown. Consequently it has to be estimated on the basis of the same random sample. One possible way of estimating σ^2 is via the sample variance

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (3.4.6)$$

where $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$, the sample mean. According to Csörgő and Horváth (1997, Section 2.1), the use of the so called *pooled variances*

$$\hat{\sigma}_{[(n+1)t]}^2 = \begin{cases} \frac{1}{n} \left(\sum_{i=1}^{[(n+1)t]} (X_i - \hat{X}_{[(n+1)t]})^2 + \sum_{i=[(n+1)t]}^n (X_i - X_{[(n+1)t]}^*)^2 \right), & 0 \leq t < \frac{n}{n+1}, \\ \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_n)^2, & \frac{n}{n+1} \leq t \leq 1, \end{cases}$$

where

$$\hat{X}_{[(n+1)t]} = \frac{1}{[(n+1)t]} \sum_{i=1}^{[(n+1)t]} X_i \quad \text{and} \quad X_{[(n+1)t]}^* = \frac{1}{n - [(n+1)t]} \sum_{i=[(n+1)t]+1}^n X_i \quad (3.4.7)$$

is preferable to that of (3.4.6).

By using the pooled variances, we consider the possibility that there be a change-point in the data. Of course, this may also effect the variance. Therefore, we compute the variance before and after each time k , any of which could be a possible change, instead of the variance of all n data, since the latter does not take a possible change into consideration.

Csörgő and Horváth (1997, Section 2.1) show that by using the Law of Large Numbers, weak uniform consistency of the sequence of estimators $\hat{\sigma}_{[(n+1)t]}^2$ for estimating σ^2 can be established, namely, as $n \rightarrow \infty$,

$$\max_{1 \leq [(n+1)t] \leq n} |\hat{\sigma}_{[(n+1)t]}^2 - \sigma^2| = o_P(1).$$

Since $\hat{\sigma}_{[(n+1)t]}^2$ is a consistent estimator for σ^2 , $\bar{U}_n(t)$ in (3.4.5) may be estimated by

$$\hat{\bar{U}}_n(t) = \frac{1}{\hat{\sigma}_{[(n+1)t]}^2 n^{3/2}} \sum_{1 \leq i \leq [(n+1)t]} \sum_{[(n+1)t] < j \leq n} (X_i - X_j),$$

and our previous result from (3.4.4) carries over. In particular, as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |\hat{\bar{U}}_n(t) - B_n(t)| = o_P(1),$$

and, as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |\hat{\bar{U}}_n(t)| \xrightarrow{\mathcal{D}} \sup_{0 < t < 1} |B(t)|,$$

imply that we can still use tables for the distribution of $\sup_{0 < t < 1} |B(t)|$ and reject the null hypothesis of no change in the data, if our test statistic $\sup_{0 < t < 1} |\hat{\bar{U}}_n(t)|$ becomes too big.

Summarizing, the test statistic $\sup_{0 < t < 1} |\hat{\bar{U}}_n(t)|$ may be consistently used to test H_0 against H_A .

Moreover, if H_0 is rejected, then the time of change may be estimated by

$$\hat{\tau} = \min \left\{ k : Z_k = \max_{1 \leq m < n} Z_m \right\}, \quad \text{if } \theta_{1,2} > 0,$$

or by

$$\tilde{\tau} = \min \left\{ k : Z_k = \min_{1 \leq m < n} Z_m \right\}, \quad \text{if } \theta_{1,2} < 0.$$

3.5 Change in the Variance

We are to test the *no-change in the variance* hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables
with $\mathbb{E}X_i = \mu$ and $0 < \sigma^2 = \text{Var}X_i < \infty$, $1 \leq i \leq n$,

against the *at most one change in the variance* alternative

$H_A : X_1, \dots, X_n$ are independent random variables and there is an integer τ , $1 \leq \tau < n$, such that $\text{Var}X_1 = \dots = \text{Var}X_\tau \neq \text{Var}X_{\tau+1} = \dots = \text{Var}X_n$, $0 < \text{Var}X_\tau, \text{Var}X_{\tau+1} < \infty$, and $\mathbb{E}X_1 = \dots = \mathbb{E}X_n = \mu$.

Taking simulated values of the therein indicated (k, X_k) , Figure 3.5.8 gives an example where a change in the variance occurs while the mean stays the same.

Figure 3.5.8: The data $X_1, \dots, X_{700}$ are i.i.d. $N(0, 1)$ -distributed and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(0, 4)$ -distributed.

Let us assume throughout this section, that H_0 holds. If μ is known, then the problem seems to be very simple. Testing H_0 against H_A means that we are looking for the change in the mean of $(X_i - \mu)^2$, $1 \leq i \leq n$. So, similarly to Section 3.4, we look at the difference of

$$n \sum_{i=1}^k (X_i - \mu)^2 - k \sum_{i=1}^n (X_i - \mu)^2,$$

which under H_0 should be fluctuating near zero. Hence, we consider the stochastic process

$$\sum_{i=1}^k \sum_{j=k+1}^n \left((X_i - \mu)^2 - (X_j - \mu)^2 \right), \quad 1 \leq k < n.$$

If μ were known then this process is like Z_k of Section 3.4 and assuming appropriate moment conditions, a test statistic can be based on its sup-functional just like here. Indeed, assuming that $0 < \bar{\sigma}^2 = \text{Var}(X_i^2) < \infty$ the results of Section 3.4 apply in

this context as well. Moreover, Gombay, Horváth and Hušková (1996) also show that estimating μ if need be by its - under H_0 - consistent estimator $\bar{X}_n$, the just mentioned asymptotic results of Section 3.4 continue to hold true in this context as well, i. e., just like as if μ were known. When μ is unknown, as an alternative approach, we may also use the *symmetric* kernel (cf. Example 2.3.2)

$$h(x, y) = \frac{1}{2}(x - y)^2,$$

where under H_0 and $X_i \neq X_j$,

$$\begin{aligned} \mathbb{E}\left(\frac{1}{2}(X_i - X_j)^2\right) &= \frac{1}{2}\left(\mathbb{E}(X_i^2) - 2\mathbb{E}(X_i)\mathbb{E}(X_j) + \mathbb{E}(X_j^2)\right) \\ &= \frac{1}{2}\left(2\mathbb{E}(X_i^2) - 2(\mathbb{E}X_i)^2\right) \\ &= \text{Var}(X_i) \\ &= \sigma^2. \end{aligned}$$

Therefore, this h is an unbiased estimator for the variance σ^2 of the X_i 's. Furthermore we have to assume that

$$\mathbb{E}h^2(X_i, X_j) < \infty, \quad (3.5.1)$$

and, on account of $\mathbb{E}(X_i - \mu) = 0$ and $\sigma^2 = \mathbb{E}(X_i - \mu)^2$, we have that under H_0 and $X_i \neq X_j$,

$$\begin{aligned} \mathbb{E}h^2(X_i, X_j) &= \mathbb{E}\left\{\left(\frac{1}{2}(X_i - X_j)^2\right)^2\right\} \\ &= \frac{1}{4}\mathbb{E}\left\{\left((X_i - \mu) - (X_j - \mu)\right)^4\right\} \\ &= \frac{1}{4}\sum_{k=0}^4 \binom{4}{k} (-1)^{4-k} \mathbb{E}\{(X_i - \mu)^k\} \mathbb{E}\{(X_j - \mu)^{4-k}\} \\ &= \frac{1}{4}\left[\mathbb{E}(X_i - \mu)^4 - 4\mathbb{E}(X_i - \mu)^3\mathbb{E}(X_j - \mu) + 6\mathbb{E}(X_i - \mu)^2 \right. \\ &\quad \times \mathbb{E}(X_j - \mu)^2 - 4\mathbb{E}(X_i - \mu)\mathbb{E}(X_j - \mu)^3 + \mathbb{E}(X_j - \mu)^4] \\ &= \frac{1}{4}\left[\mathbb{E}(X_i - \mu)^4 - 0 + 6(\mathbb{E}(X_i - \mu)^2)^2 - 0 + \mathbb{E}(X_i - \mu)^4\right] \end{aligned}$$

$$= \frac{1}{4} [2\mathbb{E}(X_i - \mu)^4 + 6(\text{Var} X_i)^2].$$

Therefore our assumption in (3.5.1) becomes

$$\frac{1}{4} [2\mathbb{E}(X_i - \mu)^4 + 6(\sigma^2)^2] < \infty.$$

Define

$$\begin{aligned} \tilde{h}(t) &= \mathbb{E}\{h(t, X_2) - \sigma^2\} \\ &= \mathbb{E}\left\{\frac{1}{2}(t - X_2)^2 - \sigma^2\right\} \\ &= \frac{1}{2}(t^2 - 2t\mathbb{E}X_2 + \mathbb{E}X_2^2) - \sigma^2 \\ &= \frac{1}{2}(t^2 - 2t\mu + (\sigma^2 + \mu^2)) - \sigma^2 \\ &= \frac{1}{2}((t - \mu)^2 - \sigma^2). \end{aligned}$$

Since under H_0

$$\begin{aligned} \mathbb{E}\tilde{h}^2(X_1) &= \mathbb{E}\left\{\left[\frac{1}{2}((X_1 - \mu)^2 - \sigma^2)\right]^2\right\} \\ &= \frac{1}{4}\left\{\mathbb{E}(X_1 - \mu)^4 - 2\sigma^2\mathbb{E}(X_1 - \mu)^2 + (\sigma^2)^2\right\} \\ &= \frac{1}{4}\left\{\mathbb{E}(X_1 - \mu)^4 - (\sigma^2)^2\right\} \\ &= \frac{1}{4}\left\{\mathbb{E}(X_1 - \mu)^4 - (\mathbb{E}(X_1 - \mu)^2 - (\mathbb{E}(X_1 - \mu))^2)^2\right\} \\ &= \frac{1}{4}\left\{\mathbb{E}((X_1 - \mu)^2)^2 - (\mathbb{E}(X_1 - \mu)^2)^2\right\} \\ &= \frac{1}{4}\text{Var}((X_1 - \mu)^2), \end{aligned}$$

it follows from (3.5.1) that

$$\tilde{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) = \frac{1}{4}\text{Var}((X_1 - \mu)^2) < \infty,$$

and, in addition to (3.5.1), we also assume that

$$\bar{\sigma}^2 = \mathbb{E}\bar{h}^2(X_1) = \frac{1}{4}\text{Var}((X_1 - \mu)^2) > 0.$$

Since h is a *symmetric* kernel, we know from Theorem 3.3.1 that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |U_n(t) - \Gamma_n(t)| = o_P(1), \quad (3.5.2)$$

where we define $U_n(t)$ as in (3.1.15) for $0 < t < 1$ to be

$$\begin{aligned} U_n(t) &= \frac{1}{\bar{\sigma}n^{3/2}}U_{[(n+1)t]}, \\ &= \frac{1}{\bar{\sigma}n^{3/2}}\left(Z_{[(n+1)t]} - [(n+1)t](n - [(n+1)t])\sigma^2\right), \end{aligned} \quad (3.5.3)$$

with

$$\begin{aligned} Z_{[(n+1)t]} &= \sum_{i=1}^{[(n+1)t]} \sum_{j=[(n+1)t]+1}^n \frac{1}{2}(X_i - X_j)^2, \\ \bar{\sigma} &= \frac{1}{2}\left(\text{Var}((X_1 - \mu)^2)\right)^{1/2}, \\ \sigma^2 &= \text{Var}X_1, \end{aligned}$$

and

$$\Gamma(t) = (1-t)W(t) + t(W(1) - W(t)), \quad 0 \leq t \leq 1,$$

where

$$\{\Gamma_n(t), 0 \leq t \leq 1\} \stackrel{D}{=} \{\Gamma(t), 0 \leq t \leq 1\},$$

for each n .

Similarly to Section 3.4, where we investigated the change in the mean, we now have to estimate the usually unknown parameters σ^2 and $\bar{\sigma}^2$. Again we can

estimate σ^2 by the usual estimator of the variance $\hat{\sigma}_n^2$ from (3.4.6), or by the pooled variance $\hat{\sigma}_{[(n+1)t]}^2$ from (3.4.7). The second one incorporates the fact that there may be a change in the data better than the usual estimators for the variance, $\hat{\sigma}_n^2$ or $\frac{1}{n-1} \sum_{1 \leq i \leq n} (X_i - \bar{X}_n)^2$. Csörgő and Horváth (1997) suggest the use of the pooled variances.

We also have to estimate $\tilde{\sigma}^2$. Since

$$\begin{aligned} \tilde{\sigma} &= \frac{1}{2} \left(\text{Var}((X_1 - \mu)^2) \right)^{1/2} \\ &= \frac{1}{2} \left\{ \mathbb{E}((X_1 - \mu)^2)^2 - (\mathbb{E}(X_1 - \mu)^2)^2 \right\}^{1/2}, \end{aligned} \quad (3.5.4)$$

and $\mathbb{E}(X_1 - \mu)^2 = \text{Var} X_1$, we estimate the second part in (3.5.4) again by $\hat{\sigma}_n^2$ or by the pooled variances $\hat{\sigma}_{[(n+1)t]}^2$. To estimate $\mathbb{E}(X_1 - \mu)^4$, we either use the estimator for the 4-th moments

$$\hat{m}_n^4 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^4,$$

where $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$, or the pooled 4-th moments, since they similar to the pooled variances feel a possible change better. To do so, define

$$\hat{m}_{[(n+1)t]}^4 = \begin{cases} \frac{1}{n} \left(\sum_{i=1}^{[(n+1)t]} (X_i - \hat{X}_{[(n+1)t]})^4 + \sum_{i=[(n+1)t]}^n (X_i - X_{[(n+1)t]}^*)^4 \right), & 0 \leq t < \frac{n}{n+1}, \\ \frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_n)^4, & \frac{n}{n+1} \leq t \leq 1, \end{cases}$$

where $\hat{X}_{[(n+1)t]}$ and $X_{[(n+1)t]}^*$ are defined in (3.4.7). By using the Law of Large Numbers, weak uniform consistency of the sequence of estimators $\hat{m}_{[(n+1)t]}^4$ for estimating $\mathbb{E}(X_1 - \mu)^4$ can be established.

Hence, we estimate $U_n(t)$ in (3.5.3) by

$$\hat{U}_n(t) = \frac{\left(Z_{[(n+1)t]} - [(n+1)t](n - [(n+1)t])\hat{\sigma}_{[(n+1)t]}^2 \right)}{\frac{1}{2} \left(\hat{m}_{[(n+1)t]}^4 - (\hat{\sigma}_{[(n+1)t]}^2)^2 \right)^{1/2} n^{3/2}},$$

and the test statistic $\sup_{0 < t < 1} |\hat{U}_n(t)|$ may be used to test H_0 against H_A . This is due to the fact that, uniformly in $t \in (0, 1)$, $\hat{U}_n(t)$ is a consistent estimator of $U_n(t)$,

hence the results for $U_n(t)$ carry over. For instance, assuming the same conditions as in Theorem 3.3.2 we know that under H_A , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |\hat{U}_n(t)| \xrightarrow{P} \infty, \quad (3.5.5)$$

and from (3.5.2) it follows that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t < 1} |\hat{U}_n(t)| \xrightarrow{\mathcal{D}} \sup_{0 < t < 1} |\Gamma(t)|, \quad (3.5.6)$$

on account of $\sup_{0 < t < 1} |\hat{U}_n(t) - U_n(t)| = o_P(1)$. This means that the test statistic $\sup_{0 < t < 1} |\hat{U}_n(t)|$ converges in distribution to the supremum of the absolute value of the Gaussian process $\Gamma(t)$ and we can hope to use tables for the distribution of the supremum of this Gaussian process to accept or reject H_0 .

Moreover, if H_0 is rejected, then the time of change may be estimated by

$$\hat{\tau} = [n\hat{\lambda}] := \min \left\{ k : Z_k = \max_{1 \leq m < n} Z_m \right\}, \quad \text{if } Z_k, 1 \leq k < n, \text{ is } \cap\text{-shaped,}$$

or by

$$\tilde{\tau} = [n\tilde{\lambda}] := \min \left\{ k : Z_k = \min_{1 \leq m < n} Z_m \right\}, \quad \text{if } Z_k, 1 \leq k < n, \text{ is } \cup\text{-shaped.}$$

Let us assume that $\theta_1, \theta_2, \theta_{1,2} > 0$ (the other cases are similar). Then we have the unique maximum exactly at the change-point τ , if the change-point lies in the following interval (cf. (3.3.25)):

$$\lambda = \lim_{n \rightarrow \infty} \frac{\tau}{n} \in \left[\frac{\theta_2}{\theta_2 + \theta_{1,2}}, \left(1 - \frac{\theta_1}{\theta_1 + \theta_{1,2}}\right) \right], \quad (3.5.7)$$

where $\tau = \tau(n) = [n\lambda]$, $\theta_1 = \mathbb{E}\left(\frac{1}{2}(X_i - X_j)^2\right)$, $1 \leq i < j \leq \tau$, $\theta_2 = \mathbb{E}\left(\frac{1}{2}(X_i - X_j)^2\right)$, $\tau < i < j \leq n$, and $\theta_{1,2} = \mathbb{E}\left(\frac{1}{2}(X_i - X_j)^2\right)$, $1 \leq i \leq \tau < j \leq n$. These parameters depend on the change-point τ , hence we have to know them a priori to check whether (3.5.7) is satisfied or not.

Suppose we estimate θ_1 , θ_2 and $\theta_{1,2}$ by using $\hat{\tau}$, the point where the maximum is reached (or $\tilde{\tau}$ respectively). Hence, we estimate θ_1 and θ_2 by the usual estimator for the variance, the so-called second moment. To estimate the third parameter we use the fact that $\theta_{1,2} = (Var X_i + (\mathbb{E}X_i)^2) - 2\mathbb{E}X_i\mathbb{E}X_j + (Var X_j + (\mathbb{E}X_j)^2)$. Considering the $\sim$ -shaped case, the respective estimators are

$$\begin{aligned}\hat{\theta}_1 &= \left(\frac{1}{2} \sum_{i=1}^{\hat{\tau}} (X_i - \hat{X}_{[(n+1)\frac{\hat{\tau}}{n}]})^2 \right), \\ \hat{\theta}_2 &= \left(\frac{1}{2} \sum_{i=\hat{\tau}+1}^n (X_i - X_{[(n+1)\frac{\hat{\tau}}{n}]})^2 \right), \\ \hat{\theta}_{1,2} &= \left(\hat{\theta}_1 + (\hat{X}_{[(n+1)\frac{\hat{\tau}}{n}]})^2 \right) - 2\hat{X}_{[(n+1)\frac{\hat{\tau}}{n}]}X_{[(n+1)\frac{\hat{\tau}}{n}]}^* + \left(\hat{\theta}_2 + (X_{[(n+1)\frac{\hat{\tau}}{n}]}^*)^2 \right),\end{aligned}$$

where $\hat{X}_{[(n+1)\frac{\hat{\tau}}{n}]}$ and $X_{[(n+1)\frac{\hat{\tau}}{n}]}^*$ are defined in (3.4.7). But it is not true in general that

$$\lim_{n \rightarrow \infty} \frac{\hat{\tau}}{n} \in \left[\frac{\hat{\theta}_2}{\hat{\theta}_2 + \hat{\theta}_{1,2}}, \left(1 - \frac{\hat{\theta}_1}{\hat{\theta}_1 + \hat{\theta}_{1,2}}\right) \right], \quad (3.5.8)$$

implies

$$\lim_{n \rightarrow \infty} \frac{\hat{\tau}}{n} \in \left[\frac{\theta_2}{\theta_2 + \theta_{1,2}}, \left(1 - \frac{\theta_1}{\theta_1 + \theta_{1,2}}\right) \right]. \quad (3.5.9)$$

This is due to the fact that the estimators $\hat{\theta}_1$, $\hat{\theta}_2$ and $\hat{\theta}_{1,2}$ may be very bad, if $\hat{\tau}$ is far apart from τ . Therefore it is necessary to know the parameters θ_1 , θ_2 and $\theta_{1,2}$ a priori. Otherwise the procedure of estimating the unknown change-point τ by $\hat{\tau}$, the point where the stochastic process Z_k , $1 \leq k < n$ takes its maximum, will not be reliable. Moreover, Theorem 3.3.3 of Ferger and Stute does not hold, since we do not have a unique maximum at $k = \tau$. As mentioned before in Section 3.3.3, the above interval is very good, if $|\theta_1|, |\theta_2| \ll |\theta_{1,2}|$.

In conclusion, we should also say that the ‘change in the mean’ approach of Gombay, Horváth and Hušková (1996) to testing for changes in the variance via the

stochastic process (cf. also Section 2.8.7 in Csörgő and Horváth (1997))

$$\sum_{i=1}^k \sum_{j=k+1}^n \left((X_i - \bar{X}_n)^2 - (X_j - \bar{X}_n)^2 \right), \quad 1 \leq k < n,$$

appears to be preferable to that of (3.5.6) that we have just discussed.

“Courage is what it takes to stand up and speak.
Courage is also what it takes to sit down and listen.”

— Winston Churchill

Chapter 4

At Most Two Change-points

4.1 Introduction

We are to test the null hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables

against the alternative that there are at most two change-points in the sequence $X_1, \dots, X_n$, namely that we have

$H_A^{(2)} : X_1, \dots, X_n$ are independent random variables and there exist two integers $\tau_1 = \tau_1(n)$ and $\tau_2 = \tau_2(n)$, $1 \leq \tau_1 \leq \tau_2 < n$, such that $\mathbb{P}\{X_1 \leq t\} = \dots = \mathbb{P}\{X_{\tau_1} \leq t\}$, $\mathbb{P}\{X_{\tau_1+1} \leq t\} = \dots = \mathbb{P}\{X_{\tau_2} \leq t\}$, $\mathbb{P}\{X_{\tau_2+1} \leq t\} = \dots = \mathbb{P}\{X_n \leq t\}$ for all t , where $\mathbb{P}\{X_{\tau_1} \leq t_0\} \neq \mathbb{P}\{X_{\tau_1+1} \leq t_0\}$ and $\mathbb{P}\{X_{\tau_2} \leq t_0\} \neq \mathbb{P}\{X_{\tau_2+1} \leq t_0\}$ for some t_0 .

We mention that the alternative $H_A^{(2)}$ allows us to consider random variables $X_1, X_2, \dots, X_n$ with two changes in the distribution, but not necessarily involving three different distributions. For example, we could have a sample where the distribution before the first change is the same as the one after the second change, but in between them we have a different distribution. This is the so-called *epidemic alternative* as we shall see in Section 4.6

Suppose we were to test H_0 vs. $H_A^{(2)}$ by using a properly normalized sup-functional of the stochastic process $Z_{[(n+1)t]}$, $0 < t < 1$, from (3.1.1) as a test statistic. Borrowing the notations from (4.1.16) and (4.1.17) the in probability limiting function

of $\frac{1}{n^2} Z_{[(n+1)t]}$ under $H_A^{(2)}$ will be seen (cf. (5.7.2)) to be

$$u_{\lambda_1, \lambda_2}(t) = \begin{cases} t(\lambda_1 - t)\theta_1 + t(\lambda_2 - \lambda_1)\theta_{1,2} + t(1 - \lambda_2)\theta_{1,3}, & 0 < t \leq \lambda_1 < \lambda_2 < 1, \\ \lambda_1(\lambda_2 - t)\theta_{1,2} + \lambda_1(1 - \lambda_2)\theta_{1,3} + (t - \lambda_1) \\ \quad \times (\lambda_2 - t)\theta_2 + (t - \lambda_1)(1 - \lambda_2)\theta_{2,3}, & 0 < \lambda_1 < t \leq \lambda_2 < 1, \\ \lambda_1(1 - t)\theta_{1,3} + (\lambda_2 - \lambda_1)(1 - t)\theta_{2,3} \\ \quad + (t - \lambda_2)(1 - t)\theta_3, & 0 < \lambda_1 < \lambda_2 < t < 1. \end{cases} \quad (4.1.1)$$

Moreover, if we put $\theta_1 = \theta_2 = \theta_3 = 0$, $\theta_{1,2} = \frac{(1-\lambda_2)\theta_{2,3}}{\lambda_1}$ and $\theta_{1,3} = \frac{(\lambda_1-\lambda_2)\theta_{2,3}}{\lambda_1}$, then $u_{\lambda_1, \lambda_2}(t)$ is equal to zero for each t , $0 < t < 1$, when testing for two changes. Consequently, $\sup_{0 < t < 1} \left| \frac{Z_{[(n+1)t]}}{\sigma n^{3/2}} \right|$ is not consistent in general, when testing H_0 vs. $H_A^{(2)}$. This means that in some instances the alternative $H_A^{(2)}$ will be rejected although it may be true. Therefore, we have to define a different stochastic process which kind of ‘feels’ the possibility of two changes. This will then allow us to study the behavior of its sup-functional and suggest a consistent test statistic for testing H_0 vs. $H_A^{(2)}$.

Our experience gained from the previous chapter suggests to use a stochastic process that depends on two time variables. To construct such a two-time parameter stochastic process, we split the given sample $X_1, \dots, X_n$ into three blocks and compare each of the blocks with the other two. This corresponds to the idea of comparing the mean before (the first change-point), between (the two change-points) and after (the two change-points). Hence, we are using a kernel $h(x, y)$ of two variables x and y , and, since we will compare two blocks with each other at a time, we have three different possibilities to do so. Therefore, we define our new stochastic process in k_1 and k_2 as follows:

$$\begin{aligned} Z_{k_1, k_2} &:= \sum_{i=1}^{k_1} \sum_{j=k_1+1}^{k_2} h(X_i, X_j) + \sum_{i=1}^{k_1} \sum_{j=k_2+1}^n h(X_i, X_j) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n h(X_i, X_j) \\ &= \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n h(X_i, X_j) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n h(X_i, X_j), \\ &\quad 1 \leq k_1 < k_2 < n. \end{aligned} \quad (4.1.2)$$

In this way we compare the three blocks $(X_1, \dots, X_{k_1})$, $(X_{k_1+1}, \dots, X_{k_2})$ and $(X_{k_2+1}, \dots, X_n)$ with each other, where k_1 and k_2 vary from 1 to $n-2$ and 2 to $n-1$, respectively.

We are to study the asymptotic properties (as $n \rightarrow \infty$) of the process Z_{k_1, k_2} , $1 \leq k_1 < k_2 < n$, which may be expressed in terms of U-statistics. To do so we give some notations and definitions under the null hypothesis of no-change and under the alternative of at most two changes.

4.1.1 Notations under the Null Hypothesis H_0

We define

$$\mathbb{E}h(X_i, X_j) =: \theta, \quad 1 \leq i < j \leq n, \quad (4.1.3)$$

and

$$\mathbb{E}h^2(X_i, X_j) =: \nu, \quad 1 \leq i < j \leq n. \quad (4.1.4)$$

We assume throughout the whole chapter that

$$\nu < \infty, \quad (4.1.5)$$

which of course implies

$$|\theta| < \infty. \quad (4.1.6)$$

Computing now the expected value of Z_{k_1, k_2} , we obtain

$$\begin{aligned} \mathbb{E}Z_{k_1, k_2} &= \mathbb{E} \sum_{i=1}^{k_1} \sum_{j=k_1+1}^{k_2} h(X_i, X_j) + \mathbb{E} \sum_{i=1}^{k_1} \sum_{j=k_2+1}^n h(X_i, X_j) \\ &\quad + \mathbb{E} \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n h(X_i, X_j) \\ &= (k_2 - k_1)k_1\theta + (n - k_2)k_1\theta + (n - k_2)(k_2 - k_1)\theta \end{aligned}$$

$$= ((k_2 - k_1)k_1 + (n - k_2)k_2)\theta, \quad 1 \leq k_1 < k_2 < n. \quad (4.1.7)$$

We define, as in the previous chapter,

$$\tilde{h}(t) = \mathbb{E}\{h(t, X_2) - \theta\}, \quad (4.1.8)$$

and again we assume that

$$\tilde{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) > 0. \quad (4.1.9)$$

We centralize Z_{k_1, k_2} by its mean, and consider the process

$$\begin{aligned} U_{k_1, k_2} &= Z_{k_1, k_2} - ((k_2 - k_1)k_1 + (n - k_2)k_2)\theta, \\ &1 \leq k_1 < k_2 < n, \end{aligned} \quad (4.1.10)$$

where the kernel function h in Z_{k_1, k_2} is *symmetric*. For an *antisymmetric* kernel we define

$$\bar{U}_{k_1, k_2} = Z_{k_1, k_2}, \quad 1 \leq k_1 < k_2 < n. \quad (4.1.11)$$

Moreover, we can write Z_{k_1, k_2} as the sum of four U-statistics, and thus we get

$$\begin{aligned} U_{k_1, k_2} &= U_n^{(4)} - (U_{k_1}^{(1)} + U_{k_1, k_2}^{(2)} + U_{k_2, n}^{(3)}), \\ &1 \leq k_1 < k_2 < n, \end{aligned} \quad (4.1.12)$$

where

$$\begin{aligned} U_{k_1}^{(1)} &= \sum_{1 \leq i < j \leq k_1} h(X_i, X_j) - \binom{k_1}{2} \theta, \\ U_{k_1, k_2}^{(2)} &= \sum_{k_1 < i < j \leq k_2} h(X_i, X_j) - \binom{k_2 - k_1}{2} \theta, \\ U_{k_2, n}^{(3)} &= \sum_{k_2 < i < j \leq n} h(X_i, X_j) - \binom{n - k_2}{2} \theta, \end{aligned}$$

$$U_n^{(4)} = \sum_{1 \leq i < j \leq n} h(X_i, X_j) - \binom{n}{2} \theta.$$

Similarly, for an *antisymmetric* kernel we define

$$\begin{aligned} \bar{U}_{k_1, k_2} &= Z_n^{(4)} - (Z_{k_1}^{(1)} + Z_{k_1, k_2}^{(2)} + Z_{k_2, n}^{(3)}), \\ 1 &\leq k_1 < k_2 < n, \end{aligned} \quad (4.1.13)$$

where

$$\begin{aligned} Z_{k_1}^{(1)} &= \sum_{1 \leq i < j \leq k_1} h(X_i, X_j), \\ Z_{k_1, k_2}^{(2)} &= \sum_{k_1 < i < j \leq k_2} h(X_i, X_j), \\ Z_{k_2, n}^{(3)} &= \sum_{k_2 < i < j \leq n} h(X_i, X_j), \\ Z_n^{(4)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j). \end{aligned}$$

For further use, via (4.1.10) and (4.1.13) respectively, we also define

$$\begin{aligned} U_n(t_1, t_2) &= \frac{1}{\bar{\sigma} n^{3/2}} U_{[(n+1)t_1], [(n+1)t_2]}, \\ 0 &< t_1 < t_2 < 1, \end{aligned} \quad (4.1.14)$$

and

$$\begin{aligned} \bar{U}_n(t_1, t_2) &= \frac{1}{\bar{\sigma} n^{3/2}} \bar{U}_{[(n+1)t_1], [(n+1)t_2]}, \\ 0 &< t_1 < t_2 < 1. \end{aligned} \quad (4.1.15)$$

4.1.2 Notations under the Alternative $H_A^{(2)}$

Let $F^{(1)}(t) = \mathbb{P}\{X_{\tau_1} \leq t\}$, $F^{(2)}(t) = \mathbb{P}\{X_{\tau_1+1} \leq t\}$ and $F^{(3)}(t) = \mathbb{P}\{X_{\tau_2+1} \leq t\}$ be the respective distribution functions of the observations before the first, between

the first and second and after the second change, and put

$$\mathbb{E}h(X_i, X_j) =: \begin{cases} \theta_1, & 1 \leq i < j \leq \tau_1 \leq \tau_2 < n, \\ \theta_{1,2}, & 1 \leq i \leq \tau_1 < j \leq \tau_2 < n, \\ \theta_{1,3}, & 1 \leq i \leq \tau_1 \leq \tau_2 < j \leq n, \\ \theta_2, & 1 < \tau_1 < i < j \leq \tau_2 < n, \\ \theta_{2,3}, & 1 < \tau_1 < i \leq \tau_2 < j \leq n, \\ \theta_3, & 1 < \tau_1 \leq \tau_2 < i < j \leq n. \end{cases} \quad (4.1.16)$$

Furthermore, we define

$$\tau_1 := [n\lambda_1] \text{ and } \tau_2 := [n\lambda_2], \quad 0 < \lambda_1 \leq \lambda_2 < 1. \quad (4.1.17)$$

Similarly to (4.1.16), we also define the second moment of the kernel h by

$$\mathbb{E}h^2(X_i, X_j) =: \begin{cases} \nu_1, & 1 \leq i < j \leq \tau_1 \leq \tau_2 < n, \\ \nu_{1,2}, & 1 \leq i \leq \tau_1 < j \leq \tau_2 < n, \\ \nu_{1,3}, & 1 \leq i \leq \tau_1 \leq \tau_2 < j \leq n, \\ \nu_2, & 1 < \tau_1 < i < j \leq \tau_2 < n, \\ \nu_{2,3}, & 1 < \tau_1 < i \leq \tau_2 < j \leq n, \\ \nu_3, & 1 < \tau_1 \leq \tau_2 < i < j \leq n, \end{cases} \quad (4.1.18)$$

and assume throughout the whole chapter that $\mathbb{E}h^2(X_i, X_j)$ is finite for all possible choices of i and j , namely

$$\nu_1, \nu_{1,2}, \nu_{1,3}, \nu_2, \nu_{2,3}, \nu_3 < \infty. \quad (4.1.19)$$

Moreover, this implies that

$$|\theta_1|, |\theta_{1,2}|, |\theta_{1,3}|, |\theta_2|, |\theta_{2,3}|, |\theta_3| < \infty. \quad (4.1.20)$$

We will also use a weaker assumption than that of the existence of a finite second moment of h , namely that for random variables from two different distributions we

have

$$\begin{aligned}\mathbb{E}\{|h(X_{\tau_1}, X_{\tau_1+1})| \log^+ (|h(X_{\tau_1}, X_{\tau_1+1})|)\} &< \infty, \\ \mathbb{E}\{|h(X_{\tau_1}, X_{\tau_2+1})| \log^+ (|h(X_{\tau_1}, X_{\tau_2+1})|)\} &< \infty, \\ \mathbb{E}\{|h(X_{\tau_2}, X_{\tau_2+1})| \log^+ (|h(X_{\tau_2}, X_{\tau_2+1})|)\} &< \infty,\end{aligned}\tag{4.1.21}$$

where $\log^+ x = \log(x \vee 1)$.

4.2 Antisymmetric Kernels

We consider the processes Z_{k_1, k_2} , $1 \leq k_1 < k_2 < n$, $n \geq 3$, where the kernel h is antisymmetric as in (2.2.3). By using the notations and assumptions from Sections 4.1.1 and 4.1.2, we investigate the behavior of these processes under H_0 (cf. Section 4.2.1) and $H_A^{(2)}$ (cf. Section 4.2.2).

4.2.1 Asymptotic Results under H_0

We wish to study the limiting behavior of

$$\begin{aligned}\bar{U}_{k_1, k_2} &= Z_{k_1, k_2} \\ &= \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n h(X_i, X_j) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n h(X_i, X_j), \\ &\qquad\qquad\qquad 1 \leq k_1 < k_2 < n.\end{aligned}$$

in the sup-norm under the null hypothesis of no-change. For the indices k_1 and k_2 , we write $[(n+1)t_1]$ and $[(n+1)t_2]$ respectively.

The asymptotic behavior of $\bar{U}_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, will be derived from the fact that we may write it in terms of U-statistics as in (4.1.13) which in turn may be replaced by sums of i.i.d. r.v.'s. Furthermore, we may then approximate these sums of i.i.d. r.v.'s by Wiener processes.

We state Theorem 2.1 of Janson and Wichura (1983) in the degenerate case which, via the proof of Theorem 4.1 of Csörgő and Horváth (1988b) (cf. Lemma 3.2.1

and Corollary 3.2.1 in this thesis), will be the basis for our approach as well.

Theorem 4.2.1 (Janson and Wichura, 1983) *$X_1, X_2, \dots$ are i.i.d. r.v.'s with common distribution. If $\tilde{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) = 0$ (degenerate case), then*

$$\left(\frac{A_{[nt]}}{n}\right)_{0 \leq t \leq 1} \xrightarrow{\mathcal{D}} \left(\sum_r \lambda_r \mathbf{A}_r(t)\right)_{0 \leq t \leq 1}, \quad (4.2.1)$$

where $A_{[nt]} = \sum_{1 \leq i < j \leq [nt]} h(X_i, X_j)$, h is an antisymmetric kernel, the $\mathbf{A}_r$'s are independent stochastic area processes, and the sum on the right-hand side of (4.2.1) converges almost surely uniformly in t . Moreover, a stochastic area process is defined to be the Ito integral $\mathbf{A} = \int_0^t -V(s)dU(s) + U(s)dV(s)$, with a two-dimensional Wiener process $W(s) = \begin{pmatrix} U(s) \\ V(s) \end{pmatrix}$, $s \geq 0$.

We mention that the U-statistics in (4.1.13) are assumed to have non-degenerate kernels h . We may, however, construct kernels h^* that are degenerate in the same context. Recall that $\mathbb{E}\tilde{h}(X_1) = \mathbb{E}h(X_1, X_2) = 0$ and $\tilde{\sigma}^2 = \text{Var}\tilde{h}(X_1) = \mathbb{E}(\tilde{h}(X_1))^2 > 0$ and let

$$h(x, y) = h^*(x, y) + \tilde{h}(x) - \tilde{h}(y).$$

Since h is an antisymmetric kernel, it follows that

$$\begin{aligned} h^*(x, y) &= h(x, y) - \tilde{h}(x) + \tilde{h}(y) \\ &= -h(y, x) + \tilde{h}(y) - \tilde{h}(x) \\ &= -(h(y, x) - \tilde{h}(y) + \tilde{h}(x)) \\ &= -h^*(y, x), \end{aligned}$$

where for each x

$$\begin{aligned} \tilde{h}^*(x) &:= \mathbb{E}(h^*(x, X_2)) \\ &= \mathbb{E}h(x, X_2) - \mathbb{E}h(x, X_2) + \mathbb{E}h(X_2, X_2) \\ &= 0. \end{aligned}$$

Hence h^* is an antisymmetric kernel and it is degenerate ($\tilde{\sigma}^{*2} = \mathbb{E}\tilde{h}^{*2}(X_1) = 0$) as well. Moreover, for the U-statistic $Z_n^{(4)}$ defined in (4.1.13) we get

$$\begin{aligned}
Z_n^{(4)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j) \\
&= \sum_{1 \leq i < j \leq n} h^*(X_i, X_j) + \sum_{1 \leq i < j \leq n} \tilde{h}(X_i) - \sum_{1 \leq i < j \leq n} \tilde{h}(X_j) \\
&= Z_n^{*(4)} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \tilde{h}(X_i) - \sum_{j=2}^n \sum_{i=1}^{j-1} \tilde{h}(X_j) \\
&= Z_n^{*(4)} + \sum_{i=1}^{n-1} (n-i) \tilde{h}(X_i) - \sum_{j=2}^n (j-1) \tilde{h}(X_j) \\
&= Z_n^{*(4)} + \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i),
\end{aligned}$$

which implies that

$$Z_n^{*(4)} = Z_n^{(4)} - \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i), \quad (4.2.2)$$

where now $Z_n^{*(4)}$ is a U-statistic with an antisymmetric and degenerate kernel h^* , and Theorem 4.2.1 can now be applied. We mention that the $*$ in (4.2.2), as well as in the following formulas, indicates that we consider the difference of one of the U-statistics in (4.1.13) and a sum of i.i.d. r.v.'s.

For convenience we let

$$\begin{aligned}
\kappa_n^{(1)}(t_1, t_2) &:= [(n+1)t_2] \sum_{j=1}^{[(n+1)t_1]} \tilde{h}(X_j) + (n - [(n+1)t_1]) \sum_{j=1}^{[(n+1)t_2]} \tilde{h}(X_j) \\
&\quad - [(n+1)t_2] \sum_{j=1}^n \tilde{h}(X_j),
\end{aligned} \quad (4.2.3)$$

and show that

$$\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} |\bar{U}_{[(n+1)t_1], [(n+1)t_2]} - \kappa_n^{(1)}(t_1, t_2)| = O_P(n). \quad (4.2.4)$$

Consequently to (4.2.4) we will approximate $\kappa_n^{(1)}(t_1, t_2)$, which is a sum of i.i.d. r.v.'s, by appropriate Wiener processes and by doing this, we will get a Gaussian limiting

distribution for a properly normed version of $\bar{U}_{[(n+1)t_1],[(n+1)t_2]}$.

As to (4.2.4), we first let

$$H(t_i) := \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j), \quad 0 < t_i < 1,$$

and conclude

$$\begin{aligned} \kappa_n^{(1)}(t_1, t_2) &= ([(n+1)t_2] + [(n+1)t_1])H(t_1) + (n + [(n+1)t_2])H(t_2) \\ &\quad - [(n+1)t_1]H(t_1) - ([(n+1)t_2] + [(n+1)t_1])H(t_2) \\ &\quad - (n + [(n+1)t_2] - n)H\left(\frac{n}{n+1}\right) \\ &= -[(n+1)t_1]H(t_1) - ([(n+1)t_2] + [(n+1)t_1])(H(t_2) - H(t_1)) \\ &\quad - (n + [(n+1)t_2])(H\left(\frac{n}{n+1}\right) - H(t_2)) \\ &\quad + nH\left(\frac{n}{n+1}\right) + \sum_{j=1}^n (-2j+1)\tilde{h}(X_j) - \sum_{j=1}^{[(n+1)t_1]} (-2j+1)\tilde{h}(X_j) \\ &\quad - \sum_{j=[(n+1)t_1]+1}^{[(n+1)t_2]} (-2j+1)\tilde{h}(X_j) - \sum_{j=[(n+1)t_2]+1}^n (-2j+1)\tilde{h}(X_j) \\ &= \sum_{j=1}^n (n-2j+1)\tilde{h}(X_j) - \sum_{j=1}^{[(n+1)t_1]} ([(n+1)t_1] - 2j+1)\tilde{h}(X_j) \\ &\quad - \sum_{j=[(n+1)t_1]+1}^{[(n+1)t_2]} ([(n+1)t_2] + [(n+1)t_1] - 2j+1)\tilde{h}(X_j) \\ &\quad - \sum_{j=[(n+1)t_2]+1}^n (n + [(n+1)t_2] - 2j+1)\tilde{h}(X_j). \end{aligned}$$

Hence, we may now use formula (4.1.13), and get

$$\begin{aligned} \bar{U}_{[(n+1)t_1],[(n+1)t_2]} - \kappa_n^{(1)}(t_1, t_2) &= Z_n^{*(4)} - (Z_{[(n+1)t_1]}^{*(1)} + Z_{[(n+1)t_1],[(n+1)t_2]}^{*(2)} \\ &\quad + Z_{[(n+1)t_2],n}^{*(3)}), \quad 0 < t_1 < t_2 < 1. \end{aligned} \quad (4.2.5)$$

Based on (4.2.5), we arrive at the inequality

$$|\bar{U}_{[(n+1)t_1],[(n+1)t_2]} - \kappa_n^{(1)}(t_1, t_2)| \leq |Z_n^{*(4)}| + |Z_{[(n+1)t_1]}^{*(1)}| + |Z_{[(n+1)t_1],[(n+1)t_2]}^{*(2)}|$$

$$+ |Z_{[(n+1)t_2],n}^{*(3)}|, \quad 0 < t_1 < t_2 < 1, \quad (4.2.6)$$

and proceed to prove the following reduction principle.

Lemma 4.2.1 *Let h be an antisymmetric kernel and $Z_{[(n+1)t_1]}^{(1)}$, $Z_{[(n+1)t_1],[(n+1)t_2]}^{(2)}$, $Z_{[(n+1)t_2],n}^{(3)}$, $Z_n^{(4)}$ defined as in (4.1.13). Then under H_0 the following statements hold true as $n \rightarrow \infty$:*

$$\begin{aligned} & \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| Z_{[(n+1)t_1]}^{(1)} \right. \\ & \quad \left. - \sum_{i=1}^{[(n+1)t_1]} ((n+1)t_1 - 2i + 1) \tilde{h}(X_i) \right| = O_P(n), \\ & \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| Z_{[(n+1)t_1],[(n+1)t_2]}^{(2)} \right. \\ & \quad \left. - \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} ((n+1)t_2 + [(n+1)t_1] - 2i + 1) \tilde{h}(X_i) \right| = O_P(n), \\ & \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| Z_{[(n+1)t_2],n}^{(3)} \right. \\ & \quad \left. - \sum_{i=[(n+1)t_2]+1}^n (n + [(n+1)t_2] - 2i + 1) \tilde{h}(X_i) \right| = O_P(n), \\ & \left| Z_n^{(4)} - \sum_{i=1}^n (n - 2i + 1) \tilde{h}(X_i) \right| = O_P(n). \end{aligned}$$

Proof of Lemma 4.2.1 The first statement follows via (4.2.1) and the fact that, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |Z_{[(n+1)t_1]}^{*(1)}| \leq \sup_{0 < t_1 < 1} |Z_{[(n+1)t_1]}^{*(1)}| = O_P(n).$$

$Z_{[(n+1)t_1],[(n+1)t_2]}^{*(2)}$ is a U-statistic (with an antisymmetric and degenerate kernel) from $[(n+1)t_1] + 1$ to $[(n+1)t_2]$ minus a sum of i.i.d. r.v.'s. Hence, we have to shift the interval $[(n+1)t_1 + 1, [(n+1)t_2]]$ to the interval $[1, [(n+1)(t_2 - t_1)]]$ by using the fact that for i.i.d. r.v.'s $X_1, X_2, \dots, X_n$

$$\left\{ \sum_{[(n+1)t_1+1] \leq i < j \leq [(n+1)t_2]} h^*(X_i, X_j), 0 < t_1 < t_2 < 1 \right\}$$

$$\stackrel{\mathcal{D}}{=} \left\{ \sum_{1 \leq i < j \leq [(n+1)(t_2-t_1)]} h^*(X_i, X_j), 0 < t_1 < t_2 < 1 \right\}, \quad (4.2.7)$$

for each n . Moreover, by (4.2.1) and as $n \rightarrow \infty$,

$$\sup_{0 < t_2 - t_1 < 1} |Z_{[(n+1)(t_2-t_1)]}^{*(1)}| = O_P(n),$$

and we may write $\sup_{0 < t_2 - t_1 < 1} = \sup_{t_1 < t_2 < 1+t_1}$. Since (4.2.7) holds uniformly in t_1 and t_2 such that $0 < t_1 < t_2 < 1$, we have $\sup_{t_1 < t_2 < 1+t_1} = \sup_{0 < t_1 < t_2 < 1}$ and we arrive at, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |Z_{[(n+1)t_1], [(n+1)t_2]}^{*(2)}| = O_P(n),$$

which proves the second statement in this lemma.

Similarly, $Z_{[(n+1)t_2], n}^{*(3)}$ is a U-statistic (with an antisymmetric and degenerate kernel) from $[(n+1)t_2] + 1$ to n minus a sum of i.i.d. r.v.'s. Hence, we have to shift the interval $[(n+1)t_2 + 1, n]$ to the interval $[1, [n - (n+1)t_2]]$. Then

$$\begin{aligned} & \left\{ \sum_{[(n+1)t_2+1] \leq i < j \leq n} h^*(X_i, X_j), 0 < t_2 < 1 \right\} \\ & \stackrel{\mathcal{D}}{=} \left\{ \sum_{1 \leq i < j \leq [n-(n+1)t_2]} h^*(X_i, X_j), 0 < t_2 < 1 \right\}, \end{aligned} \quad (4.2.8)$$

for each n , and, as $n \rightarrow \infty$,

$$\sup_{0 < 1-t_2 < 1} |Z_{[n-(n+1)t_2]}^{*(1)}| = O_P(n).$$

Restricting t_2 to the interval $(t_1, 1) \subset (0, 1)$, where t_1 is between 0 and 1, then the supremum over this interval will also be $O_P(n)$, hence, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |Z_{[(n+1)t_2], n}^{*(3)}| = O_P(n),$$

which proves the third statement in this lemma.

By using (4.2.1) together with (4.2.2) we get, as $n \rightarrow \infty$,

$$|Z_n^{*(4)}| = O_P(n),$$

which proves the fourth and last statement in this lemma. $\square$

As a consequence of Lemma 4.2.1 and (4.2.6) we conclude also

Corollary 4.2.1 *Let h be an antisymmetric kernel and $\bar{U}_{[(n+1)t_1], [(n+1)t_2]}$ be defined as in (4.1.13). Under H_0 , as $n \rightarrow \infty$, we have*

$$\begin{aligned} \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| \bar{U}_{[(n+1)t_1], [(n+1)t_2]} - \left\{ [(n+1)t_2] \sum_{j=1}^{[(n+1)t_1]} \tilde{h}(X_j) \right. \right. \\ \left. \left. + (n - [(n+1)t_1]) \sum_{j=1}^{[(n+1)t_2]} \tilde{h}(X_j) - [(n+1)t_2] \sum_{j=1}^n \tilde{h}(X_j) \right\} \right| = O_P(n). \end{aligned}$$

We emphasize that $\bar{U}_{[(n+1)t_1], [(n+1)t_2]}$ is now approximated by sums of i.i.d. r.v.'s for which there are many limit theorems available. In particular, we now are in the position to study the asymptotic Gaussian behavior of $\bar{U}_{[(n+1)t_1], [(n+1)t_2]}$ in the sup-norm. We give the following theorem as well as a detailed proof.

Theorem 4.2.2 *Assume that H_0 , (2.2.3), (4.1.5) and (4.1.9) hold. Then we can define a sequence of Gaussian processes $\{\Gamma_n^a(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\}_{n \in \mathbb{N}}$ such that, as $n \rightarrow \infty$,*

$$\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2) - \Gamma_n^a(t_1, t_2)| = o_P(1), \quad (4.2.9)$$

and, for each n ,

$$\{\Gamma_n^a(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\} \stackrel{\mathcal{D}}{=} \{\Gamma^a(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\},$$

where the Gaussian process Γ^a is defined via a linear combination of a standard Wiener process W as follows:

$$\Gamma^a(t_1, t_2) = t_2 W(t_1) + (1 - t_1) W(t_2) - t_2 W(1), \quad 0 \leq t_1 \leq t_2 \leq 1.$$

$$(4.2.10)$$

Proof of Theorem 4.2.2 We define for each n the two-time parameter Gaussian process

$$\Gamma_n^a(t_1, t_2) := \frac{1}{n^{1/2}} (t_2 W(nt_1) + (1 - t_1)W(nt_2) - t_2 W(n)), \quad 0 \leq t_1 \leq t_2 \leq 1.$$

Indeed, we may write Γ_n^a in terms of a linear combination of independent increments of W , namely

$$\begin{aligned} \Gamma_n^a(t_1, t_2) = & \frac{1}{n^{1/2}} ((1 - t_1)W(nt_1) + (1 - t_1 - t_2)(W(nt_2) - W(nt_1)) \\ & - t_2(W(n) - W(nt_2))), \quad 0 \leq t_1 \leq t_2 \leq 1. \end{aligned}$$

The just mentioned independency follows from the definition of W .

Similarly, we conclude that Γ^a in (4.2.10) is a Gaussian process. Using the fact that Gaussian processes are uniquely determined by their covariance structure and that Γ_n^a (for each n) and Γ^a have the same covariance, it follows that for each n

$$\{\Gamma_n^a(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\} \stackrel{\mathcal{D}}{=} \{\Gamma^a(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\}.$$

Having shown that the above defined Γ_n^a is a sequence of Gaussian processes, we now prove (4.2.9) via using Corollary 4.2.1 and the fact that we can define a Wiener process $\{W(t), 0 \leq t < \infty\}$ such that (cf. Csörgő and Révész (1981, Theorem S.2.2.1 by Major (1979) combined with (S.2.2.2))), as $n \rightarrow \infty$,

$$\frac{1}{n^{1/2}} \sup_{0 \leq t \leq 1} \left| \frac{1}{\bar{\sigma}} \sum_{i=1}^{[(n+1)t]} \tilde{h}(X_i) - W(nt) \right| = o_P(1).$$

Hence, bounding above $\sup_{0 < t_1 < t_2 < 1}$ by $\sup_{0 < t_1 < 1}$ or $\sup_{0 < t_2 < 1}$ then we also have, as $n \rightarrow \infty$,

$$\frac{1}{n^{1/2}} \sup_{0 < t_1 < t_2 < 1} \left| \frac{1}{\bar{\sigma}} \sum_{i=1}^{[(n+1)t_j]} \tilde{h}(X_i) - W(nt_j) \right| = o_P(1), \quad \text{for } j = 1, 2. \quad (4.2.11)$$

Recall the definition of $\kappa_n^{(1)}(t_1, t_2)$ as in (4.2.3), and define

$$\rho_n := \sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2) - \Gamma_n^a(t_1, t_2)|.$$

Then, as $n \rightarrow \infty$,

$$\begin{aligned} \rho_n &= \sup_{0 < t_1 < t_2 < 1} \left| \bar{U}_n(t_1, t_2) - \frac{\kappa_n^{(1)}(t_1, t_2)}{\bar{\sigma} n^{3/2}} + \frac{\kappa_n^{(1)}(t_1, t_2)}{\bar{\sigma} n^{3/2}} - \Gamma_n^a(t_1, t_2) \right| \\ &\leq \sup_{0 < t_1 < t_2 < 1} \frac{1}{\bar{\sigma} n^{3/2}} \left| \bar{U}_{[(n+1)t_1], [(n+1)t_2]} - \kappa_n^{(1)}(t_1, t_2) \right| + \sup_{0 < t_1 < t_2 < 1} \left| \frac{\kappa_n^{(1)}(t_1, t_2)}{\bar{\sigma} n^{3/2}} - \Gamma_n^a(t_1, t_2) \right| \\ &= \frac{O_P(n)}{\bar{\sigma} n^{3/2}} + \sup_{0 < t_1 < t_2 < 1} \left| \frac{\kappa_n^{(1)}(t_1, t_2)}{\bar{\sigma} n^{3/2}} - \Gamma_n^a(t_1, t_2) \right| \\ &=: o_P(1) + \rho_n^{(1)}, \end{aligned}$$

where $\rho_n^{(1)}$ denotes the second expression on the right hand side. Next we observe that by using the definitions of $\kappa_n^{(1)}(t_1, t_2)$ and Γ_n^a we arrive, as $n \rightarrow \infty$, at

$$\begin{aligned} \rho_n^{(1)} &\leq \sup_{0 < t_1 < t_2 < 1} \frac{t_2}{n^{1/2}} \left| \frac{1}{\bar{\sigma}} \sum_{j=1}^{[(n+1)t_1]} \tilde{h}(X_j) - W(nt_1) \right| + \sup_{0 < t_1 < t_2 < 1} \frac{1-t_1}{n^{1/2}} \left| \frac{1}{\bar{\sigma}} \sum_{j=1}^{[(n+1)t_2]} \tilde{h}(X_j) \right. \\ &\quad \left. - W(nt_2) \right| + \sup_{0 < t_1 < t_2 < 1} \frac{t_2}{n^{1/2}} \left| \frac{1}{\bar{\sigma}} \sum_{j=1}^n \tilde{h}(X_j) - W(n) \right| \\ &= o_P(1), \end{aligned}$$

where the $o_P(1)$ statement follows from (4.2.11). Hence, it is proven that $\rho_n = o_P(1)$.

□

To avoid confusion with the notation, we denoted the two-time parameter Gaussian process by Γ^a , where the upper index a indicates that this Gaussian process corresponds to the case where we have an antisymmetric kernel. In case of a symmetric kernel, we will use the upper index sym instead. Furthermore we note that for $0 \leq t_1 \leq t_2 \leq 1$

$$\mathbb{E}\Gamma^a(t_1, t_2) = 0,$$

$$\text{Var}\Gamma^a(t_1, t_2) = t_2(1 - t_1)(1 + t_1 - t_2),$$

and

$$\mathbb{E}\Gamma^a(t_1, t_2)\Gamma^a(r_1, r_2) = \sum_{i=1}^2 \sum_{j=1}^2 (t_{i+1} - t_{i-1})(r_{j+1} - r_{j-1})(t_i \wedge r_j) - t_2 r_2,$$

where $t_0 = r_0 := 0$ and $t_3 = r_3 := 1$. Theorem 4.2.2 implies that under H_0

$$\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^a(t_1, t_2)|. \quad (4.2.12)$$

Moreover, this allows us to produce tables for the unknown distribution of $\sup_{0 < t_1 < t_2 < 1} |\Gamma^a(t_1, t_2)|$ and reject the null hypothesis of no-change, if $\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2)|$ becomes too large.

4.2.2 Asymptotic Results under $H_A^{(2)}$

We wish to study the limiting behavior of $Z_{[(n+1)t_1], [(n+1)t_2]} = \bar{U}_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, from (4.1.2). To do so, we will introduce the limiting function of $\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$. We will show that, as $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]} \xrightarrow{P} \bar{u}_{\lambda_1, \lambda_2}(t_1, t_2), \quad 0 < t_1 < t_2 < 1, \quad (4.2.13)$$

where we will assume that $h(x, y)$ is a nondegenerate antisymmetric kernel and (4.1.21) holds. Therefore (4.1.21) replaces the stronger assumption of a second finite moment as in (4.1.19) which was used to derive (cf. Section 4.2.1) the non-degenerate convergence in distribution, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^a(t_1, t_2)|,$$

where the stochastic process $\bar{U}_n(t_1, t_2)$ was defined as

$$\bar{U}_n(t_1, t_2) = \frac{1}{\bar{\sigma} n^{3/2}} Z_{[(n+1)t_1], [(n+1)t_2]}, \quad 0 < t_1 < t_2 < 1,$$

and the limiting Gaussian process $\Gamma^a(t_1, t_2)$ as

$$\Gamma^a(t_1, t_2) = t_2 W(t_1) + (1 - t_1) W(t_2) - t_2 W(1), \quad 0 \leq t_1 \leq t_2 \leq 1.$$

Although there will be a larger class of distributions $\{F^{(1)}, F^{(2)}, F^{(3)}\}$ that satisfies (4.1.21) than (4.1.19), naturally the non-degenerate results under H_0 require that the class of distributions where we are testing for 2 changes satisfies (4.1.19). Nevertheless, we will use the weaker assumption (4.1.21) to derive the degenerate asymptotic results under $H_A^{(2)}$.

The limiting function in (4.2.13) will depend on the location of the (at most) two change-points $[n\lambda_1]$ and $[n\lambda_2]$, $0 < \lambda_1 \leq \lambda_2 < 1$. Moreover, we will see that $\sup_{0 < t_1 < t_2 < 1} |\tilde{U}_n(t_1, t_2)|$ is consistent and goes to infinity in probability, as $n \rightarrow \infty$, under $H_A^{(2)}$.

The limiting function in (4.2.13) will be derived similarly to the results of Section 4.3.2 which deal with symmetric kernels. Since an antisymmetric kernel may be viewed as a symmetric kernel with $\theta_1 = \theta_2 = \theta_3 = 0$, we are jumping ahead and use the results from Section 4.3.2 here. There it is shown that $Z_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, may be split into many double sums, where the summation is taken over two blocks of r.v.'s that do not contain a change-point. Moreover, after proper normalization each of these double sums converges in probability to the appropriate $\theta_{i,j}$ (cf. (4.3.14) - (4.3.16)).

Since $Z_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, in (4.1.2) with $k_i = [(n+1)t_i]$, $1 \leq i \leq 2$, is the sum of the three double sums $\sum_{i=1}^{[(n+1)t_1]} \sum_{j=[(n+1)t_1]+1}^{[(n+1)t_2]} h(X_i, X_j)$, $\sum_{i=1}^{[(n+1)t_1]} \sum_{j=[(n+1)t_2]+1}^n h(X_i, X_j)$ and $\sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} \sum_{j=[(n+1)t_2]+1}^n h(X_i, X_j)$ we define for technical purposes

$$\begin{aligned} S(x_1, x_4, x_5, x_8, x_2, x_3, x_6, x_7) \\ &:= \sum_{i=[(n+1)x_1]+1}^{[(n+1)x_4]} \sum_{j=[(n+1)x_5]+1}^{[(n+1)x_8]} h(X_i, X_j) \\ &= \left(\sum_{i=[(n+1)x_1]+1}^{[(n+1)x_2]} + \sum_{i=[(n+1)x_2]+1}^{[(n+1)x_3]} + \sum_{i=[(n+1)x_3]+1}^{[(n+1)x_4]} \right) \end{aligned}$$

$$\times \left(\sum_{j=[(n+1)x_5]+1}^{[(n+1)x_6]} + \sum_{j=[(n+1)x_6]+1}^{[(n+1)x_7]} + \sum_{j=[(n+1)x_7]+1}^{[(n+1)x_8]} \right) h(X_i, X_j), \quad (4.2.14)$$

where $0 \leq x_1 < x_2 \leq x_3 \leq x_4 \leq x_5 < x_6 \leq x_7 \leq x_8 < 1$. Hence (4.2.14) is the sum of exactly nine double sums, where the summation is taken over two blocks of r.v.'s that do not contain a change-point. Here x_2, x_3, x_6 and x_7 play the role of dummy variables and are used to split blocks that contain change-points. At most two of them will actually be used, but since the location of the change-points is unknown, we need to consider two possible changes in each of the blocks $(X_{[(n+1)x_1]+1}, \dots, X_{[(n+1)x_4]})$ and $(X_{[(n+1)x_5]+1}, \dots, X_{[(n+1)x_8]})$. Hence, the convergence result from Section 4.3.2 may be applied to (4.2.14) and we get, as $n \rightarrow \infty$,

$$\begin{aligned} & \frac{1}{n^2} S(x_1, x_4, x_5, x_8, x_2, x_3, x_6, x_7) \\ & \xrightarrow{P} \bar{\eta}(x_1, x_4, x_5, x_8, x_2, x_3, x_6, x_7) \\ & := \sum_{q=5}^7 (x_{q+1} - x_q) \sum_{r=1}^3 (x_{r+1} - x_r) \theta_{l(x_{r+1}), l(x_{q+1})}, \end{aligned} \quad (4.2.15)$$

where $\theta_{1,1} = \theta_{2,2} = \theta_{3,3} = \theta = 0$ and¹

$$l(z) := \begin{cases} 1, & 0 < z \leq \lambda_1, \\ 2, & \lambda_1 < z \leq \lambda_2, \\ 3, & \lambda_2 < z \leq 1. \end{cases} \quad (4.2.16)$$

Moreover, if the distribution before the first change and after the second change are the same then we also have $\theta_{1,3} = \theta = 0$.

We are now in the position to go back to the definition of $Z_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, in (4.1.2) and write it in terms of double sums $S(\cdot)$ as in (4.2.14). Thus we arrive at

$$Z_{[(n+1)t_1], [(n+1)t_2]} = \sum_{j=1}^2 \sum_{i=j}^2 S(t_{j-1}, t_j, t_i, t_{i+1}, x_2, x_3, x_6, x_7),$$

¹We note that $\theta_{1,1} := \theta_1$, $\theta_{2,2} := \theta_2$ and $\theta_{3,3} := \theta_3$ which was defined in (4.1.16).

$$0 < t_1 < t_2 < 1, \quad (4.2.17)$$

where we define $t_0 := 0$ and $t_3 := \frac{n}{n+1}$. Since we do not know the location of the change-points $[n\lambda_1]$ and $[n\lambda_2]$, where $0 < \lambda_1 \leq \lambda_2 < 1$, we define the following functions $a_i = a_i(\lambda_1, \lambda_2, t_1, t_2)$, $1 \leq i \leq 6$, which will be used to derive a formula for (4.2.13) for all possible combinations of t_1, t_2, λ_1 and λ_2 :

$$\begin{aligned} a_1 &:= \begin{cases} \lambda_1, & \lambda_1 \leq t_1, \\ c_1, & \text{otherwise,} \end{cases} \\ a_2 &:= \begin{cases} \lambda_2, & \lambda_1 \leq \lambda_2 \leq t_1, \\ c_2, & \text{otherwise} \end{cases} \\ a_3 &:= \begin{cases} \lambda_1, & t_1 < \lambda_1 \leq t_2, \\ \lambda_2, & \lambda_1 \leq t_1 < \lambda_2 \leq t_2, \\ c_3, & \text{otherwise,} \end{cases} \\ a_4 &:= \begin{cases} \lambda_2, & t_1 < \lambda_1 \leq \lambda_2 \leq t_2, \\ c_4, & \text{otherwise,} \end{cases} \\ a_5 &:= \begin{cases} \lambda_1, & t_2 < \lambda_1 < 1, \\ \lambda_2, & \lambda_1 \leq t_2 < \lambda_2 < 1, \\ c_5, & \text{otherwise,} \end{cases} \\ a_6 &:= \begin{cases} \lambda_2, & t_2 < \lambda_1 \leq \lambda_2 < 1, \\ c_6, & \text{otherwise,} \end{cases} \end{aligned} \quad (4.2.18)$$

where $0 < a_1 \leq a_2 \leq \dots \leq a_6 < 1$ and $c_i \in [0, 1]$, $1 \leq i \leq 6$.

We need to define these a_i 's, $1 \leq i \leq 6$, since there are exactly six possibilities to place two change-points in three blocks. Moreover, exactly two of the latter a_i 's will change to one of the values λ_1 or λ_2 , while the other four a_i 's will get the value c_i and they will drop out of the limiting function $\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2)$. Hence, the actual values of the c_i 's are not important. The c_i 's are needed to give a general formula for the limiting function.

We are now in the position to state the following theorem which is an immediate consequence of the previous arguments in this section.

Theorem 4.2.3 *Assume that (2.2.3), (4.1.20), (4.1.21), and $H_A^{(2)}$ hold. Define $t_0 :=$*

0 and $t_3 := \frac{n}{n+1}$.² If $\tau_1 = \tau_1(n) := [n\lambda_1]$ and $\tau_2 = \tau_2(n) := [n\lambda_2]$, $0 < \lambda_1 \leq \lambda_2 < 1$, then, as $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]} \xrightarrow{P} \bar{u}_{\lambda_1, \lambda_2}(t_1, t_2),$$

where

$$\begin{aligned} \bar{u}_{\lambda_1, \lambda_2}(t_1, t_2) &= \sum_{j=1}^2 \sum_{i=j}^2 \bar{\eta}(t_{j-1}, t_j, t_i, t_{i+1}, a_{2(j-1)+1}, a_{2(j-1)+2}, a_{2i+1}, a_{2i+2}), \\ &0 < t_1 < t_2 < 1, \end{aligned} \quad (4.2.19)$$

and $\bar{\eta}$ and the a_i 's, $1 \leq i \leq 6$, are defined in (4.2.15) and (4.2.18), respectively.

We note that in Theorem 4.2.3 $\theta_{1,1} = \theta_{2,2} = \theta_{3,3} = \theta = 0$. Moreover, if the distributions before the first change and after the second change are the same, then we also have $\theta_{1,3} = \theta = 0$.

The limiting function in (4.2.19) is defined for every possible combination of t_1 , t_2 , λ_1 and λ_2 . If we write these cases individually, then (4.2.19) may also be written as³

²Note that $t_3 \rightarrow 1$ when $n \rightarrow \infty$. In (4.2.17) $t_3 := \frac{n}{n+1}$, since $n = \left[(n+1) \frac{n}{n+1} \right] = [(n+1)t_3]$ has to be satisfied, but in the limiting function (4.2.19) $t_3 = 1$.

³We note that in case of exactly one change-point, namely $\lambda_1 = \lambda_2$, the cases where $0 < t_1 \leq \lambda_1 < t_2 \leq \lambda_2 < 1$, $0 < \lambda_1 < t_1 < t_2 \leq \lambda_2 < 1$ and $0 < \lambda_1 < t_1 \leq \lambda_2 < t_2 < 1$ do not exist. We are left with the remaining three cases, when $\lambda_1 = \lambda_2$.

$$\left\{ \begin{array}{ll}
((\lambda_2 - \lambda_1)t_2\theta_{1,2} + (1 - \lambda_2)t_2\theta_{1,3}, & 0 < t_1 < t_2 \leq \lambda_1 \leq \lambda_2 < 1, \\
((t_2 - \lambda_1)t_1 + (\lambda_2 - t_2)\lambda_1)\theta_{1,2} \\
+ (1 - \lambda_2)\lambda_1\theta_{1,3} + (1 - \lambda_2)(t_2 - \lambda_1)\theta_{2,3}, & 0 < t_1 \leq \lambda_1 < t_2 \leq \lambda_2 < 1, \\
(\lambda_2 - \lambda_1)t_1\theta_{1,2} + ((1 - \lambda_2)t_1 \\
+ (\lambda_1 - t_1)(1 - t_2))\theta_{1,3} + (\lambda_2 - \lambda_1)(1 - t_2)\theta_{2,3}, & 0 < t_1 \leq \lambda_1 \leq \lambda_2 < t_2 < 1, \\
(\lambda_2 - t_1)\lambda_1\theta_{1,2} + (1 - \lambda_2)(t_2 - \lambda_1)\theta_{2,3} \\
+ (1 - \lambda_2)\lambda_1\theta_{1,3}, & 0 < \lambda_1 < t_1 < t_2 \leq \lambda_2 < 1, \\
(\lambda_2 - t_1)\lambda_1\theta_{1,2} + (1 - \lambda_2)\lambda_1\theta_{1,3} \\
+ ((1 - \lambda_2)(t_1 - \lambda_1) + (\lambda_2 - t_1)(1 - t_2))\theta_{2,3}, & 0 < \lambda_1 < t_1 \leq \lambda_2 < t_2 < 1, \\
(1 - t_1)\lambda_1\theta_{1,3} + (1 - t_1)(\lambda_2 - \lambda_1)\theta_{2,3}, & 0 < \lambda_1 \leq \lambda_2 < t_1 < t_2 < 1.
\end{array} \right. \quad (4.2.20)$$

Furthermore, if also $\theta_{1,3} = 0$, then the distribution before the first and after the second change are the same. Consequently then, $\theta_{1,2} = \theta_{2,3}$, and (4.2.19) may be written as

$$\left\{ \begin{array}{ll}
(\lambda_2 - \lambda_1)t_2\theta_{1,2}, & 0 < t_1 < t_2 \leq \lambda_1 \leq \lambda_2 < 1, \\
((t_2 - \lambda_1)(1 - \lambda_2 + t_1) + (\lambda_2 - t_2)\lambda_1)\theta_{1,2}, & 0 < t_1 \leq \lambda_1 < t_2 \leq \lambda_2 < 1, \\
(\lambda_2 - \lambda_1)(1 - t_2 + t_1)\theta_{1,2}, & 0 < t_1 \leq \lambda_1 \leq \lambda_2 < t_2 < 1, \\
((\lambda_2 - t_1)\lambda_1 + (1 - \lambda_2)(t_2 - \lambda_1))\theta_{1,2}, & 0 < \lambda_1 < t_1 < t_2 \leq \lambda_2 < 1, \\
((1 - \lambda_2)(t_1 - \lambda_1) + (\lambda_2 - t_1)(1 - t_2 + \lambda_1))\theta_{1,2}, & 0 < \lambda_1 < t_1 \leq \lambda_2 < t_2 < 1, \\
(1 - t_1)(\lambda_2 - \lambda_1)\theta_{1,2}, & 0 < \lambda_1 \leq \lambda_2 < t_1 < t_2 < 1.
\end{array} \right. \quad (4.2.21)$$

Assuming second finite moments $\mathbb{E}h^2$ instead of (4.1.21) Theorem 4.2.3 implies the consistency of tests based on sup-functionals of $\{\bar{U}_{[(n+1)t_1], [(n+1)t_2]}, 0 < t_1 < t_2 < 1\}$. This means that we can consistently reject H_0 vs. $H_A^{(2)}$, when

$$\sup_{0 < t_1 < t_2 < 1} \frac{1}{n^{3/2}\bar{\sigma}} |\bar{U}_{[(n+1)t_1], [(n+1)t_2]}| \text{ is too large,}$$

except in the case when $\theta_1 = \theta_2 = \theta_3 = \theta_{1,2} = \theta_{1,3} = \theta_{2,3} = 0$.

$\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2)$ is equal to 0 if and only if all $\theta_{i,j}$'s involved are equal to 0. To show this we observe that each of the six parts of the limiting function in (4.2.20) may be written as

$$u_{\lambda_1, \lambda_2}(t_1, t_2) = A_j t_1 t_2 + B_j t_1 + C_j t_2 + D_j, \quad 1 \leq j \leq 6,$$

where A_j , B_j , C_j and D_j depend on $\theta_{1,2}$, $\theta_{1,3}$ and $\theta_{2,3}$. Moreover, $u_{\lambda_1, \lambda_2}(t_1, t_2) = 0$, $0 < t_1 < t_2 < 1$, if and only if $A_j = 0$, $B_j = 0$, $C_j = 0$ and $D_j = 0$, $1 \leq j \leq 6$. For example, the linear system of the six independent equations $D_1 = 0$, $D_2 = 0$, $\dots$, $D_6 = 0$ involves only three unknown parameters which implies that there is only one possible solution, namely $\theta_{1,2} = \theta_{1,3} = \theta_{2,3} = 0$.

We observe that $\theta_1 = \theta_2 = \theta_3 = \theta_{1,2} = \theta_{1,3} = \theta_{2,3} = \theta$ implies

$$\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2) = 0.$$

Moreover, it follows then that under the null hypothesis H_0 of no-change

$$\begin{aligned} \frac{1}{n^2} \bar{U}_{[(n+1)t_1], [(n+1)t_2]} &= \frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]} \\ &\xrightarrow{P} \bar{u}_{\lambda_1, \lambda_2}(t_1, t_2) \\ &= 0 \end{aligned}$$

Thus, on assuming that $\theta_1 = \theta_2 = \theta_3 = \theta_{1,2} = \theta_{1,3} = \theta_{2,3} = \theta = 0$, the sequence

$$\bar{T}_n = \frac{1}{n^{3/2} \bar{\sigma}} \sup_{0 < t_1 < t_2 < 1} |\bar{U}_{[(n+1)t_1], [(n+1)t_2]}|, \quad n \in \mathbb{N}, \quad (4.2.22)$$

is not consistent against any class of alternatives. On the other hand, if at least one $\theta_{i,j}$ is not equal to 0 and we use $\bar{T}_n$, then

$$\mathbb{P}\{H_0 \text{ is rejected when using } \bar{T}_n | H_A^{(2)} \text{ is true}\} \xrightarrow{n \rightarrow \infty} 1.$$

This implies that the limits of the sequence $\{\bar{T}_n\}_{n \in \mathbb{N}}$ are different in probability under H_0 and $H_A^{(2)}$, namely we have consistency of $\{\bar{T}_n\}_{n \in \mathbb{N}}$.

4.3 Symmetric Kernels

We consider the processes U_{k_1, k_2} , $1 \leq k_1 < k_2 < n$, $n \geq 3$, where the kernel h is symmetric as in (2.2.2). By using the notations and assumptions from Sections 4.1.1 and 4.1.2 we investigate the behavior of these processes under H_0 (cf. Section 4.3.1) and $H_A^{(2)}$ (cf. Section 4.3.2) .

4.3.1 Asymptotic Results under H_0

We wish to study the limiting bahavior of

$$\begin{aligned} U_{k_1, k_2} &= Z_{k_1, k_2} - ((k_2 - k_1)k_1 + (n - k_2)k_2)\theta \\ &= \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n h(X_i, X_j) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n h(X_i, X_j) \\ &\quad - ((k_2 - k_1)k_1 + (n - k_2)k_2)\theta, \quad 1 \leq k_1 < k_2 < n, \end{aligned}$$

in the sup-norm under the null hypothesis of no-change. For the indices, we write $[(n+1)t_1]$ and $[(n+1)t_2]$ respectively.

The asymptotic behavior of $U_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, will be derived from the following *reduction principle* (Corollary 4.3.1). It is a consequence of Theorem 1 of Hull (1979) which may be reduced to the following corollary of Hall (1979):

Corollary 4.3.1 (Hall, 1979) *$X_1, X_2, \dots$ are i.i.d. r.v.'s with common distribution. If $\tilde{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) = 0$ (degenerate case) then $\frac{\sum_{1 \leq i < j \leq [nt]} h(X_i, X_j)}{2n}$ converges weakly to a stochastic process Y , where h is a symmetric kernel with mean 0 and finite second moment.*

We mention that the U-statistics in (4.1.12) are assumed to have a non-degenerate kernel h , but we may, in our context, construct kernels g_θ^* that are degenerate. Recall that $\mathbb{E}\tilde{h}(X_1) = \mathbb{E}(h(X_1, X_2) - \theta) = 0$ and $\tilde{\sigma}^2 = \text{Var}\tilde{h}(X_1) = \mathbb{E}(\tilde{h}(X_1))^2 > 0$ and let

$$h(x, y) = h^*(x, y) + \tilde{h}(x) + \tilde{h}(y).$$

Put

$$g_{\theta}^*(x, y) = h^*(x, y) - \theta,$$

then

$$h(x, y) - \theta = g_{\theta}^*(x, y) + \tilde{h}(x) + \tilde{h}(y).$$

Since h is a symmetric kernel, it follows that

$$\begin{aligned} g_{\theta}^*(x, y) &= h(x, y) - \theta - \tilde{h}(x) - \tilde{h}(y) \\ &= h(y, x) - \theta - \tilde{h}(y) - \tilde{h}(x) \\ &= g_{\theta}^*(y, x), \end{aligned}$$

where for each x

$$\begin{aligned} \tilde{g}_{\theta}^*(x) &:= \mathbb{E}(g_{\theta}^*(x, X_2)) \\ &= \mathbb{E}(h^*(x, X_2) - \theta) \\ &= \left((\mathbb{E}h(x, X_2) - \theta) - (\mathbb{E}h(x, X_2) - \theta) - (\mathbb{E}h(X_2, X_2) - \theta) \right) - \theta \\ &= 0. \end{aligned}$$

Hence g_{θ}^* is a symmetric kernel with mean 0, and it is degenerate ($\tilde{\sigma}^{*2} = \mathbb{E}\tilde{g}_{\theta}^{*2}(X_1) = 0$) as well. Finally, for the centralized U-statistic $U_n^{(4)}$ defined in (4.1.12) we get

$$\begin{aligned} U_n^{(4)} &= \sum_{1 \leq i < j \leq n} (h(X_i, X_j) - \theta) \\ &= \sum_{1 \leq i < j \leq n} g_{\theta}^*(X_i, X_j) + \sum_{1 \leq i < j \leq n} \tilde{h}(X_i) + \sum_{1 \leq i < j \leq n} \tilde{h}(X_j) \\ &= U_n^{*(4)} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \tilde{h}(X_i) + \sum_{j=2}^n \sum_{i=1}^{j-1} \tilde{h}(X_j) \\ &= U_n^{*(4)} + \sum_{i=1}^{n-1} (n-i) \tilde{h}(X_i) + \sum_{j=2}^n (j-1) \tilde{h}(X_j) \end{aligned}$$

$$= U_n^{*(4)} + \sum_{i=1}^n (n-1) \tilde{h}(X_i),$$

which implies that

$$\begin{aligned} U_n^{*(4)} &= \left(Z_n^{(4)} - \binom{n}{2} \theta \right) - \sum_{i=1}^n (n-1) \tilde{h}(X_i), \\ &= U_n^{(4)} - (n-1) \sum_{i=1}^n \tilde{h}(X_i), \end{aligned} \quad (4.3.1)$$

where $U_n^{*(4)}$ is a centralized U-statistic with a symmetric and degenerate kernel g_θ^* , and Corollary 4.3.1 can be applied. We note that $U_n^{*(4)}$ corresponds to the U-statistic $Z_n^{*(4)}$ in (4.2.2) in the antisymmetric case, but they are different. Also, we have to use different known results to get $O_P(n)$ statements, namely Corollary 4.3.1 and Theorem 4.2.1 respectively.

We mention that the $*$ in (4.3.1) as well as in the following formulas indicates that we consider the difference of one of the centralized U-statistics in (4.1.12) and its corresponding projection.

Similarly to the antisymmetric case (cf. (4.2.3)), let

$$\begin{aligned} \kappa_n^{(2)}(t_1, t_2) &:= \left([(n+1)t_2] - 2[(n+1)t_1] \right) \sum_{j=1}^{[(n+1)t_1]} \tilde{h}(X_j) + (n + [(n+1)t_1] \\ &\quad - 2[(n+1)t_2]) \sum_{j=1}^{[(n+1)t_2]} \tilde{h}(X_j) + [(n+1)t_2] \sum_{j=1}^n \tilde{h}(X_j), \end{aligned}$$

and we will show that (cf. Corollary 4.3.2)

$$\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} |U_{[(n+1)t_1], [(n+1)t_2]} - \kappa_n^{(2)}(t_1, t_2)| = O_P(n). \quad (4.3.2)$$

Consequently to (4.3.2), we approximate $\kappa_n^{(2)}(t_1, t_2)$, which is a sum of i.i.d. r.v.'s, by appropriate Wiener processes and by doing this, we will get a Gaussian limiting distribution for a properly normed version of $U_{[(n+1)t_1], [(n+1)t_2]}$.

We first prove

Lemma 4.3.1 *Let h be a symmetric kernel and $U_{[(n+1)t_1]}^{(1)}, U_{[(n+1)t_1], [(n+1)t_2]}^{(2)}, U_{[(n+1)t_2], n}^{(3)}$,*

$U_n^{(4)}$ defined as in (4.1.12). Then under H_0 the following statements hold true as $n \rightarrow \infty$:

$$\begin{aligned}
\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| U_{[(n+1)t_1]}^{(1)} - [(n+1)t_1] \sum_{i=1}^{[(n+1)t_1]} \tilde{h}(X_i) \right| &= O_P(n), \\
\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| U_{[(n+1)t_1], [(n+1)t_2]}^{(2)} - ([(n+1)t_2] \right. \\
&\quad \left. - [(n+1)t_1]) \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} \tilde{h}(X_i) \right| = O_P(n), \\
\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| U_{[(n+1)t_2], n}^{(3)} - (n - [(n+1)t_2]) \sum_{i=[(n+1)t_2]+1}^n \tilde{h}(X_i) \right| &= O_P(n), \\
\left| U_n^{(4)} - n \sum_{i=1}^n \tilde{h}(X_i) \right| &= O_P(n).
\end{aligned}$$

Proof of Lemma 4.3.1 For technical purposes we define

$$H(t_i) := \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j), \quad 0 < t_i < 1.$$

Thus we obtain

$$\begin{aligned}
\kappa_n^{(2)}(t_1, t_2) &= ([(n+1)t_2] - [(n+1)t_1])H(t_1) + (n - [(n+1)t_2])H(t_2) \\
&\quad - [(n+1)t_1]H(t_1) - ([(n+1)t_2] - [(n+1)t_1])H(t_2) \\
&\quad + (n + [(n+1)t_2] - n)H\left(\frac{n}{n+1}\right) \\
&= -[(n+1)t_1]H(t_1) - ([(n+1)t_2] - [(n+1)t_1])(H(t_2) - H(t_1)) \\
&\quad - (n - [(n+1)t_2])(H\left(\frac{n}{n+1}\right) - H(t_2)) + nH\left(\frac{n}{n+1}\right) \\
&= n \sum_{j=1}^n \tilde{h}(X_j) - [(n+1)t_1] \sum_{j=1}^{[(n+1)t_1]} \tilde{h}(X_j) - ([(n+1)t_2] - [(n+1)t_1]) \\
&\quad \times \sum_{j=[(n+1)t_1]+1}^{[(n+1)t_2]} \tilde{h}(X_j) - (n - [(n+1)t_2]) \sum_{j=[(n+1)t_2]+1}^n \tilde{h}(X_j), \\
&= (n-1) \sum_{j=1}^n \tilde{h}(X_j) - ([(n+1)t_1] - 1) \sum_{j=1}^{[(n+1)t_1]} \tilde{h}(X_j)
\end{aligned}$$

$$\begin{aligned}
& - \left([(n+1)t_2] - [(n+1)t_1] - 1 \right) \sum_{j=[(n+1)t_1]+1}^{[(n+1)t_2]} \tilde{h}(X_j) \\
& - (n - [(n+1)t_2] - 1) \sum_{j=[(n+1)t_2]+1}^n \tilde{h}(X_j),
\end{aligned}$$

and we now use formula (4.1.12) to arrive at

$$\begin{aligned}
U_{[(n+1)t_1], [(n+1)t_2]} - \kappa_n^{(2)}(t_1, t_2) &= U_n^{*(4)} - (U_{[(n+1)t_1]}^{*(1)} + U_{[(n+1)t_1], [(n+1)t_2]}^{*(2)} \\
&+ U_{[(n+1)t_2], n}^{*(3)}), \quad 0 < t_1 < t_2 < 1. \quad (4.3.3)
\end{aligned}$$

Thus we have the inequality

$$\begin{aligned}
|U_{[(n+1)t_1], [(n+1)t_2]} - \kappa_n^{(2)}(t_1, t_2)| &\leq |U_n^{*(4)}| + |U_{[(n+1)t_1]}^{*(1)}| + |U_{[(n+1)t_1], [(n+1)t_2]}^{*(2)}| \\
&+ |U_{[(n+1)t_2], n}^{*(3)}|, \quad 0 < t_1 < t_2 < 1. \quad (4.3.4)
\end{aligned}$$

By using Corollary 4.3.1 together with (4.3.1), as $n \rightarrow \infty$, we get

$$|U_n^{*(4)}| := |U_n^{(4)} - (n-1) \sum_{i=1}^n \tilde{h}(X_i)| = O_P(n). \quad (4.3.5)$$

Since $\tilde{h}(X_i)$, $1 \leq i \leq n$, are i.i.d. r.v.'s the central limit theorem with $\mathbb{E}\tilde{h}(X_i) = 0$ and $0 < \tilde{\sigma}^2 = \text{Var}\tilde{h}(X_i) < \infty$ implies

$$\begin{aligned}
\sum_{i=1}^n \tilde{h}(X_i) &= \sqrt{n}\tilde{\sigma} \cdot \frac{\sum_{i=1}^n \tilde{h}(X_i)}{\sqrt{n}\tilde{\sigma}} \\
&= O(\sqrt{n}) \cdot O_P(1) \\
&= O(\sqrt{n}). \quad (4.3.6)
\end{aligned}$$

Therefore (4.3.5), reduces to

$$|U_n^{(4)} - n \sum_{i=1}^n \tilde{h}(X_i)| = O_P(n),$$

which proves the fourth statement in this lemma.

The first statement in the lemma follows by combining (4.3.6) and the fact that,

as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_1]}^{*(1)}| \leq \sup_{0 < t_1 < 1} |U_{[(n+1)t_1]}^{*(1)}| = O_P(n).$$

$U_{[(n+1)t_1], [(n+1)t_2]}^{*(2)}$ is a U-statistic (with a symmetric and degenerate kernel with mean 0) from $[(n+1)t_1] + 1$ to $[(n+1)t_2]$ minus a sum of i.i.d. r.v.'s. Hence, we have to shift the interval $[(n+1)t_1 + 1, [(n+1)t_2]]$ to the interval $[1, [(n+1)(t_2 - t_1)]]$ by using the fact that for i.i.d. r.v.'s $X_1, X_2, \dots, X_n$

$$\left\{ \sum_{[(n+1)t_1+1] \leq i < j \leq [(n+1)t_2]} h^*(X_i, X_j), 0 < t_1 < t_2 < 1 \right\} \\ \stackrel{D}{=} \left\{ \sum_{1 \leq i < j \leq [(n+1)(t_2-t_1)]} h^*(X_i, X_j), 0 < t_1 < t_2 < 1 \right\}, \quad (4.3.7)$$

for each n . Note, that we also have equality in distribution when replacing h^* by $g_\theta^* := h^* - \theta$. Moreover, by Corollary 4.3.1,

$$\sup_{0 < t_2 - t_1 < 1} |U_{[(n+1)(t_2-t_1)]}^{*(1)}| = O_P(n),$$

and we may write $\sup_{0 < t_2 - t_1 < 1} = \sup_{t_1 < t_2 < 1+t_1}$. Since (4.3.7) holds uniformly in t_1 and t_2 such that $0 < t_1 < t_2 < 1$, we have $\sup_{t_1 < t_2 < 1+t_1} = \sup_{0 < t_1 < t_2 < 1}$ and we arrive at

$$\sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_1], [(n+1)t_2]}^{*(2)}| = O_P(n),$$

and similar arguments as in (4.3.6) yield

$$\sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_1], [(n+1)t_2]}^{(2)} - ([[(n+1)t_2] - [(n+1)t_1]]) \sum_{i=[(n+1)t_1+1]}^{[(n+1)t_2]} \bar{h}(X_i)| = O_P(n),$$

which proves the second statement in this lemma.

Similarly, $U_{[(n+1)t_2], n}^{*(3)}$ is a U-statistic (with a symmetric and degenerate kernel with mean 0) from $[(n+1)t_2] + 1$ to n minus a sum of i.i.d. r.v.'s. Hence, we have

to shift the interval $[(n+1)t_2+1, n]$ to the interval $[1, [n-(n+1)t_2]]$. Then

$$\left\{ \sum_{[(n+1)t_2+1] \leq i < j \leq n} h^*(X_i, X_j), 0 < t_2 < 1 \right\} \\ \stackrel{D}{=} \left\{ \sum_{1 \leq i < j \leq [n-(n+1)t_2]} h^*(X_i, X_j), 0 < t_2 < 1 \right\}, \quad (4.3.8)$$

for each n , and

$$\sup_{0 < 1-t_2 < 1} |U_{[n-(n+1)t_2]}^{*(1)}| = O_P(n),$$

and we may write $\sup_{0 < 1-t_2 < 1} = \sup_{0 > t_2-1 > -1} = \sup_{0 < t_2 < 1}$. If we restrict t_2 to the interval $(t_1, 1) \subset (0, 1)$, where t_1 is between 0 and 1, then the supremum over this interval will also be $O_P(n)$, hence

$$\sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_2], n}^{*(3)}| = O_P(n).$$

Moreover, we have

$$\sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_2], n}^{(3)} - (n - [(n+1)t_2]) \sum_{i=[(n+1)t_2+1]}^n \tilde{h}(X_i)| = O_P(n),$$

which proves the third statement in this lemma. $\square$

Moreover, using (4.3.4) we get

Corollary 4.3.2 *Let h be a symmetric kernel and $U_{[(n+1)t_1], [(n+1)t_2]}$ be defined as in (4.1.10). Under H_0 , as $n \rightarrow \infty$, we have*

$$\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq n} \left| U_{[(n+1)t_1], [(n+1)t_2]} - \left\{ ([n-(n+1)t_2] - 2[(n+1)t_1]) \sum_{j=1}^{[(n+1)t_1]} \tilde{h}(X_j) \right. \right. \\ \left. \left. + (n + [(n+1)t_1] - 2[(n+1)t_2]) \sum_{j=1}^{[(n+1)t_2]} \tilde{h}(X_j) + [(n+1)t_2] \sum_{j=1}^n \tilde{h}(X_j) \right\} \right| \\ = O_P(n).$$

The conclusion of Corollary 4.3.2 is that $U_{[(n+1)t_1], [(n+1)t_2]}$ is now approximated

by sums of i.i.d. r.v.'s. This enables us to study the Gaussian asymptotic behavior of $U_{[(n+1)t_1],[(n+1)t_2]}$ in the sup-norm.

Theorem 4.3.1 *Assume that H_0 , (2.2.2), (4.1.5) and (4.1.9) hold. Then we can define a sequence of Gaussian processes $\{\Gamma_n^{sym}(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\}_{n \in \mathbb{N}}$ such that, as $n \rightarrow \infty$,*

$$\sup_{0 < t_1 < t_2 < 1} |U_n(t_1, t_2) - \Gamma_n^{sym}(t_1, t_2)| = o_P(1), \quad (4.3.9)$$

and, for each n ,

$$\{\Gamma_n^{sym}(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\} \xrightarrow{\mathcal{D}} \{\Gamma^{sym}(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\},$$

where the Gaussian process Γ^{sym} is defined via a linear combination of a standard Wiener process W as follows:

$$\Gamma^{sym}(t_1, t_2) = (t_2 - 2t_1)W(t_1) + (1 + t_1 - 2t_2)W(t_2) + t_2W(1), \quad 0 \leq t_1 \leq t_2 \leq 1. \quad (4.3.10)$$

Proof of Theorem 4.3.1 The proof goes along the lines of the proof of Theorem 4.2.2. Namely we combine Corollary 4.3.2 with (4.2.11) and the theorem follows immediately. $\square$

We note that for $0 \leq t_1 \leq t_2 \leq 1$

$$\begin{aligned} \mathbb{E}\Gamma^{sym}(t_1, t_2) &= 0, \\ \text{Var}\Gamma^{sym}(t_1, t_2) &= (1 - t_2)t_2 + (t_2 - t_1)(4 - 3t_2)t_1, \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\Gamma^{sym}(t_1, t_2)\Gamma^{sym}(r_1, r_2) &= \sum_{i=1}^2 \sum_{j=1}^2 (t_{i+1} - 2t_i + t_{i-1})(r_{j+1} - 2r_j + r_{j-1})(t_i \wedge r_j) \\ &\quad + 2t_2(r_1r_2 - r_1^2 - r_2^2) + 2r_2(t_1t_2 - t_1^2 - t_2^2) + 3t_2r_2, \end{aligned}$$

where $t_0 = r_0 := 0$ and $t_3 = r_3 := 1$. Theorem 4.3.1 implies that under H_0

$$\sup_{0 < t_1 < t_2 < 1} |U_n(t_1, t_2)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^{sym}(t_1, t_2)|. \quad (4.3.11)$$

Moreover, this allows us to produce tables for the unknown distribution of $\sup_{0 < t_1 < t_2 < 1} |\Gamma^{sym}(t_1, t_2)|$ and reject the null hypothesis of no-change, if $\sup_{0 < t_1 < t_2 < 1} |U_n(t_1, t_2)|$ becomes too large.

4.3.2 Asymptotic Results under $H_A^{(2)}$

We wish to study the limiting behavior of $Z_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, from (4.1.2). We will show that (cf. Theorem 4.3.2), as $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]} \xrightarrow{P} u_{\lambda_1, \lambda_2}(t_1, t_2), \quad 0 < t_1 < t_2 < 1, \quad (4.3.12)$$

where we will assume that $h(x, y)$ is a nondegenerate symmetric kernel and (4.1.21) holds. Thus (4.1.21) replaces the stronger assumption of a second finite moment as in (4.1.19) which was used to derive the convergence in distribution result under H_0 (cf. Section 4.3.1), as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |U_n(t_1, t_2)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^{sym}(t_1, t_2)|,$$

where the stochastic process $U_n(t_1, t_2)$ was defined as

$$\begin{aligned} U_n(t_1, t_2) = & \frac{1}{\bar{\sigma} n^{3/2}} \left\{ Z_{[(n+1)t_1], [(n+1)t_2]} - \left(([(n+1)t_2] - [(n+1)t_1]) [(n+1)t_1] + \right. \right. \\ & \left. \left. + (n - [(n+1)t_2]) [(n+1)t_2] \right) \theta \right\}, \quad 0 < t_1 < t_2 < 1, \end{aligned}$$

and the limiting Gaussian process $\Gamma^{sym}(t_1, t_2)$ as

$$\Gamma^{sym}(t_1, t_2) = (t_2 - 2t_1)W(t_1) + (1 + t_1 - 2t_2)W(t_2) + t_2W(1), \quad 0 \leq t_1 \leq t_2 \leq 1.$$

The limiting function in (4.3.12) will depend on the location of the (at most) two change-points $[n\lambda_1]$ and $[n\lambda_2]$, $0 < \lambda_1 \leq \lambda_2 < 1$. Moreover, we will see that

$\sup_{0 < t_1 < t_2 < 1} |U_n(t_1, t_2)|$ is consistent and goes to infinity in probability, as $n \rightarrow \infty$, under $H_A^{(2)}$.

When looking at the definition of $Z_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, in (4.1.2) with $k_i = [(n+1)t_i]$, $1 \leq i \leq 2$, we see that it may be expressed in terms of the sum of the three double sums $\sum_{i=1}^{[(n+1)t_1]} \sum_{j=[(n+1)t_1]+1}^{[(n+1)t_2]} h(X_i, X_j)$, $\sum_{i=1}^{[(n+1)t_1]} \sum_{j=[(n+1)t_2]+1}^n h(X_i, X_j)$ and $\sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} \sum_{j=[(n+1)t_2]+1}^n h(X_i, X_j)$. Thus, each of these double sums is a sum of the form

$$\sum_{i=a+1}^b \sum_{j=c+1}^d h(X_i, X_j), \quad (4.3.13)$$

where $0 \leq a < b \leq c < d \leq n$ and $a, b, c, d \in \mathbb{N}$. These double sums may be associated with comparing the two blocks $(X_{a+1}, \dots, X_b)$ and $(X_{c+1}, \dots, X_d)$ with each other. In case of testing for (at most) one single change-point we compared the block $(X_1, \dots, X_k)$ with the block $(X_{k+1}, \dots, X_n)$. Now we have the three blocks $(X_1, \dots, X_{k_1})$, $(X_{k_1+1}, \dots, X_{k_2})$ and $(X_{k_2+1}, \dots, X_n)$ and therefore three possibilities to compare two different blocks with each other. Each block may contain at most two change-points, hence we split each of the sums in (4.3.13) into three sums. Since we do not know the location of the change-points we write the latter sum as follows,

$$\sum_{i=a+1}^b \sum_{j=c+1}^d h(X_i, X_j) = \left(\sum_{i=a+1}^{\gamma_1} + \sum_{i=\gamma_1+1}^{\gamma_2} + \sum_{i=\gamma_2+1}^b \right) \left(\sum_{j=c+1}^{\gamma_3} + \sum_{j=\gamma_3+1}^{\gamma_4} + \sum_{j=\gamma_4+1}^d \right) h(X_i, X_j), \quad (4.3.14)$$

where $0 \leq a < \gamma_1 \leq \gamma_2 \leq b \leq c < \gamma_3 \leq \gamma_4 \leq d \leq n$. Consequently, we are comparing blocks that do not have a change inside. Again we emphasize that there are at most two changes in total, which implies that some of the new small blocks may have no change inside or in between them. For example, if we were to have one change in the first block and one change in the second block, then (4.3.14) reduces to

$$\sum_{i=a+1}^b \sum_{j=c+1}^d h(X_i, X_j) = \left(\sum_{i=a+1}^{\gamma_1} + \sum_{i=\gamma_1+1}^b \right) \left(\sum_{j=c+1}^{\gamma_3} + \sum_{j=\gamma_3+1}^d \right) h(X_i, X_j),$$

where the blocks $(X_{a+1}, \dots, X_{\gamma_1})$, $(X_{\gamma_1+1}, \dots, X_b)$, $(X_{c+1}, \dots, X_{\gamma_3})$ and $(X_{\gamma_3+1}, \dots,$

X_d) do not contain a change-point, and, the two change-points now are postulated to be at the positions γ_1 and γ_3 , which themselves are also unknown. Since we do not know in advance the location of the change-points, we consider sums as in (4.3.14). For technical purposes we define the following function

$$l(z) := \begin{cases} 1, & 0 < z \leq \lambda_1, \\ 2, & \lambda_1 < z \leq \lambda_2, \\ 3, & \lambda_2 < z \leq 1, \end{cases} \quad (4.3.15)$$

which will be used to remind ourselves of the location (either before or between or after the change-points) of a block of r.v.'s which does not contain any change-point.

We split (4.3.14) into 3^2 different double sums which are of the form

$$\sum_{i=R_1+1}^{R_2} \sum_{j=R_3+1}^{R_4} h(X_i, X_j)$$

where $1 \leq R_1 = R_1(n) := [nr_1] < R_2 = R_2(n) := [nr_2] \leq R_3 = R_3(n) := [nr_3] < R_4 = R_4(n) := [nr_4] \leq n$ are chosen properly according to the double sums in (4.3.14). Furthermore we will see that for each of these sums we have that, as $n \rightarrow \infty$,

$$\frac{1}{(R_2 - R_1)(R_4 - R_3)} \sum_{i=R_1+1}^{R_2} \sum_{j=R_3+1}^{R_4} h(X_i, X_j) \xrightarrow{P} \theta_{l(r_2), l(r_4)}. \quad (4.3.16)$$

This is an immediate consequence of Theorem 2.6.2 for generalized two sample U-statistics by Sen (1977) (compare the proof of Theorem 3.3.2), and Hoeffding's SLLN (cf. Theorem 2.6.1) for two samples from the same distribution.

The double sum in (4.3.16) may be associated with comparing the two blocks $(X_{R_1+1}, \dots, X_{R_2})$ and $(X_{R_3+1}, \dots, X_{R_4})$ with each other, where both do not have any changes inside. If both belong to different distributions, then Theorem 2.6.2 applies, and (4.3.16) follows immediately. On the other hand, if both belong to the same distribution, then we may consider the two blocks as one block. We do this by deleting everything between those two blocks. We therefore consider the block

of i.i.d. r.v.'s $Y = (Y_{R_1+1} := X_{R_1+1}, Y_{R_1+2} := X_{R_1+2}, \dots, Y_{R_2} := X_{R_2}, Y_{R_2+1} := X_{[R_3-(R_3-R_2)]+1}, Y_{R_2+2} := X_{[R_3-(R_3-R_2)]+2}, \dots, Y_{[R_4-(R_3-R_2)]} := X_{[R_4-(R_3-R_2)]})$. We may then use Hoeffding's SLLN, but first we have to write the double sum in (4.3.16) in terms of U-statistics. Hence, we now write

$$\begin{aligned}
 \sum_{i=R_1+1}^{R_2} \sum_{j=R_3+1}^{R_4} h(X_i, X_j) &= \sum_{i=R_1+1}^{R_2} \sum_{j=R_2+1}^{R_4-(R_3-R_2)} h(Y_i, Y_j) \\
 &= \sum_{R_1 < i < j \leq R_4-(R_3-R_2)} h(Y_i, Y_j) - \sum_{R_1 < i < j \leq R_2} h(Y_i, Y_j) \\
 &\quad - \sum_{R_2 < i < j \leq R_4-(R_3-R_2)} h(Y_i, Y_j), \\
 &=: A_n^{(1)} - A_n^{(2)} - A_n^{(3)}
 \end{aligned}$$

and

$$\sum_{R_2 < i < j \leq R_4-(R_3-R_2)} h(Y_i, Y_j) \stackrel{D}{=} \sum_{R_1 < i < j \leq R_4-(R_3-R_2)-(R_2-R_1)} h(Y_i, Y_j),$$

for each fixed n . Using now Hoeffding's SLLN, we get the following convergence results for the U-statistics $A_n^{(1)}$, $A_n^{(2)}$ and $A_n^{(3)}$, as $n \rightarrow \infty$,

$$\begin{aligned}
 \frac{A_n^{(1)}}{n^2} &\xrightarrow{P} \frac{(\tau_4 - (\tau_3 - \tau_2) - \tau_1)^2}{2} \theta_{l(\tau_2), l(\tau_4-(\tau_3-\tau_2))}, \\
 \frac{A_n^{(2)}}{n^2} &\xrightarrow{P} \frac{(\tau_2 - \tau_1)^2}{2} \theta_{l(\tau_2), l(\tau_4-(\tau_3-\tau_2))}, \\
 \frac{A_n^{(3)}}{n^2} &\xrightarrow{P} \frac{(\tau_4 - (\tau_3 - \tau_2) - (\tau_2 - \tau_1) - \tau_1)^2}{2} \theta_{l(\tau_2), l(\tau_4-(\tau_3-\tau_2))}.
 \end{aligned} \tag{4.3.17}$$

Hence, as $n \rightarrow \infty$,

$$\frac{A_n^{(1)} - A_n^{(2)} - A_n^{(3)}}{n^2} \xrightarrow{P} (\tau_4 - \tau_3)(\tau_2 - \tau_1) \theta_{l(\tau_2), l(\tau_4-(\tau_3-\tau_2))}$$

and, therefore, (4.3.16) also holds, if two different blocks have the same distribution.

Note that $\theta_{l(\tau_2), l(\tau_4)} = \theta_{l(\tau_2), l(\tau_4-(\tau_3-\tau_2))}$ when using the block of i.i.d. r.v.'s Y .

Similarly to Section 4.2.2, we define

$$\begin{aligned}
S(x_1, x_4, x_5, x_8, x_2, x_3, x_6, x_7) \\
&:= \sum_{i=[(n+1)x_1]+1}^{[(n+1)x_4]} \sum_{j=[(n+1)x_5]+1}^{[(n+1)x_8]} h(X_i, X_j) \\
&= \left(\sum_{i=[(n+1)x_1]+1}^{[(n+1)x_2]} + \sum_{i=[(n+1)x_2]+1}^{[(n+1)x_3]} + \sum_{i=[(n+1)x_3]+1}^{[(n+1)x_4]} \right) \\
&\quad \times \left(\sum_{j=[(n+1)x_5]+1}^{[(n+1)x_6]} + \sum_{j=[(n+1)x_6]+1}^{[(n+1)x_7]} + \sum_{j=[(n+1)x_7]+1}^{[(n+1)x_8]} \right) h(X_i, X_j),
\end{aligned} \tag{4.3.18}$$

where $0 \leq x_1 < x_2 \leq x_3 \leq x_4 \leq x_5 < x_6 \leq x_7 \leq x_8 < 1$. Moreover, since (4.3.18) is a sum of nine double sums, (4.3.16) may be applied nine times, and the limiting function of (4.3.18) can be written as ($n \rightarrow \infty$)

$$\begin{aligned}
&\frac{1}{n^2} S(x_1, x_4, x_5, x_8, x_2, x_3, x_6, x_7) \\
&\xrightarrow{P} \eta(x_1, x_4, x_5, x_8, x_2, x_3, x_6, x_7) \\
&:= \sum_{q=5}^7 (x_{q+1} - x_q) \sum_{r=1}^3 (x_{r+1} - x_r) \theta_{l(x_{r+1}), l(x_{q+1})}.
\end{aligned} \tag{4.3.19}$$

We observe that for $\theta_1 = \theta_2 = \theta_3 = \theta_{1,2} = \theta_{1,3} = \theta_{2,3} = \theta$

$$\begin{aligned}
\eta(x_1, x_4, x_5, x_8, x_2, x_3, x_6, x_7) &= (x_8 - x_7) \left((x_2 - x_1)\theta + (x_3 - x_2)\theta \right. \\
&\quad \left. + (x_4 - x_3)\theta \right) + (x_7 - x_6) \left((x_2 - x_1)\theta \right. \\
&\quad \left. + (x_3 - x_2)\theta + (x_4 - x_3)\theta \right) + (x_6 - x_5) \\
&\quad \times \left((x_2 - x_1)\theta + (x_3 - x_2)\theta + (x_4 - x_3)\theta \right) \\
&= (x_8 - x_5)(x_4 - x_1)\theta.
\end{aligned}$$

We are now in the position to go back to the definition of $Z_{[(n+1)t_1], [(n+1)t_2]}$, $0 < t_1 < t_2 < 1$, in (4.1.2), and we write it in terms of double sums $S(\cdot)$ as

in (4.3.18). Consequently

$$Z_{[(n+1)t_1],[(n+1)t_2]} = \sum_{j=1}^2 \sum_{i=j}^2 S(t_{j-1}, t_j, t_i, t_{i+1}, x_2, x_3, x_6, x_7),$$

$$0 < t_1 < t_2 < 1, \quad (4.3.20)$$

where we define $t_0 := 0$ and $t_3 := \frac{n}{n+1}$. Since we do not know the location of the change-points $[n\lambda_1]$ and $[n\lambda_2]$, where $0 < \lambda_1 \leq \lambda_2 < 1$, we define the following functions $a_i = a_i(\lambda_1, \lambda_2, t_1, t_2)$, $1 \leq i \leq 6$:

$$\begin{aligned} a_1 &:= \begin{cases} \lambda_1, & \lambda_1 \leq t_1, \\ c_1, & \text{otherwise,} \end{cases} \\ a_2 &:= \begin{cases} \lambda_2, & \lambda_1 \leq \lambda_2 \leq t_1, \\ c_2, & \text{otherwise} \end{cases} \\ a_3 &:= \begin{cases} \lambda_1, & t_1 < \lambda_1 \leq t_2, \\ \lambda_2, & \lambda_1 \leq t_1 < \lambda_2 \leq t_2, \\ c_3, & \text{otherwise,} \end{cases} \\ a_4 &:= \begin{cases} \lambda_2, & t_1 < \lambda_1 \leq \lambda_2 \leq t_2, \\ c_4, & \text{otherwise,} \end{cases} \\ a_5 &:= \begin{cases} \lambda_1, & t_2 < \lambda_1 < 1, \\ \lambda_2, & \lambda_1 \leq t_2 < \lambda_2 < 1, \\ c_5, & \text{otherwise,} \end{cases} \\ a_6 &:= \begin{cases} \lambda_2, & t_2 < \lambda_1 \leq \lambda_2 < 1, \\ c_6, & \text{otherwise,} \end{cases} \end{aligned} \quad (4.3.21)$$

where $0 < a_1 \leq a_2 \leq \dots \leq a_6 < 1$ and $c_i \in [0, 1]$, $1 \leq i \leq 6$.

We need to define these a_i 's, $1 \leq i \leq 6$, since there are exactly six possibilities to place two change-points in three blocks. Moreover, exactly two of the latter a_i 's will change to one of the values λ_1 or λ_2 , while the other four a_i 's will get the value c_i and they will drop out of the limiting function $u_{\lambda_1, \lambda_2}(t_1, t_2)$. Hence, the actual values of the c_i 's are not important, but they are needed to give a general formula for the limiting function.

We are now in the position to state the following theorem which is an immediate

consequence of the previous arguments in this section.

Theorem 4.3.2 *Assume that (2.2.2), (4.1.20), (4.1.21), and $H_A^{(2)}$ hold. Define $t_0 := 0$ and $t_3 := \frac{n}{n+1}$.⁴ If $\tau_1 = \tau_1(n) := [n\lambda_1]$ and $\tau_2 = \tau_2(n) := [n\lambda_2]$, $0 < \lambda_1 \leq \lambda_2 < 1$, then, as $n \rightarrow \infty$,*

$$\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]} \xrightarrow{P} u_{\lambda_1, \lambda_2}(t_1, t_2),$$

where

$$u_{\lambda_1, \lambda_2}(t_1, t_2) = \sum_{j=1}^2 \sum_{i=j}^2 \eta(t_{j-1}, t_j, t_i, t_{i+1}, a_{2(j-1)+1}, a_{2(j-1)+2}, a_{2i+1}, a_{2i+2}),$$

$$0 < t_1 < t_2 < 1, \quad (4.3.22)$$

and η and the a_i 's, $1 \leq i \leq 6$, are defined in (4.3.19) and (4.3.21), respectively.

The limiting function in (4.3.22) is defined for every possible combination of t_1 , t_2 , λ_1 and λ_2 . If we spell out these cases individually, then (4.3.22) can also be written as

⁴Note that $t_3 \rightarrow 1$ when $n \rightarrow \infty$. In (4.3.20) $t_3 := \frac{n}{n+1}$, since $n = \left[(n+1) \frac{n}{n+1} \right] = [(n+1)t_3]$ has to be satisfied, but in the limiting function (4.3.22) $t_3 = 1$.

$$\left\{ \begin{array}{ll}
\begin{aligned}
& ((t_2 - t_1)t_1 + (\lambda_1 - t_2)t_2)\theta_1 + (\lambda_2 - \lambda_1)t_2\theta_{1,2} \\
& + (1 - \lambda_2)t_2\theta_{1,3},
\end{aligned} & 0 < t_1 < t_2 \leq \lambda_1 \leq \lambda_2 < 1, \\
\begin{aligned}
& (\lambda_1 - t_1)t_1\theta_1 + ((t_2 - \lambda_1)t_1 + (\lambda_2 - t_2)\lambda_1)\theta_{1,2} \\
& + (1 - \lambda_2)\lambda_1\theta_{1,3} + (\lambda_2 - t_2)(t_2 - \lambda_1)\theta_2 \\
& + (1 - \lambda_2)(t_2 - \lambda_1)\theta_{2,3},
\end{aligned} & 0 < t_1 \leq \lambda_1 < t_2 \leq \lambda_2 < 1, \\
\begin{aligned}
& (\lambda_1 - t_1)t_1\theta_1 + (\lambda_2 - \lambda_1)t_1\theta_{1,2} + ((1 - \lambda_2)t_1 \\
& + (\lambda_1 - t_1)(1 - t_2))\theta_{1,3} + (\lambda_2 - \lambda_1)(1 - t_2)\theta_{2,3} \\
& + (t_2 - \lambda_2)(1 - t_2)\theta_3,
\end{aligned} & 0 < t_1 \leq \lambda_1 \leq \lambda_2 < t_2 < 1, \\
\begin{aligned}
& (\lambda_2 - t_1)\lambda_1\theta_{1,2} + ((\lambda_2 - t_1)(t_1 - \lambda_1) \\
& + (t_2 - t_1)(\lambda_2 - t_2))\theta_2 + (1 - \lambda_2)(t_2 - \lambda_1)\theta_{2,3} \\
& + (1 - \lambda_2)\lambda_1\theta_{1,3},
\end{aligned} & 0 < \lambda_1 < t_1 < t_2 \leq \lambda_2 < 1, \\
\begin{aligned}
& (\lambda_2 - t_1)\lambda_1\theta_{1,2} + (1 - \lambda_2)\lambda_1\theta_{1,3} + ((1 - \lambda_2)(t_1 - \lambda_1) \\
& + (\lambda_2 - t_1)(1 - t_2))\theta_{2,3} + (t_1 - \lambda_1)(\lambda_2 - t_1)\theta_2 \\
& + (t_2 - \lambda_2)(1 - t_2)\theta_3,
\end{aligned} & 0 < \lambda_1 < t_1 \leq \lambda_2 < t_2 < 1, \\
\begin{aligned}
& (1 - t_1)\lambda_1\theta_{1,3} + (1 - t_1)(\lambda_2 - \lambda_1)\theta_{2,3} \\
& + ((1 - t_1)(t_1 - \lambda_2) + (1 - t_2)(t_2 - t_1))\theta_3,
\end{aligned} & 0 < \lambda_1 \leq \lambda_2 < t_1 < t_2 < 1.
\end{array} \right. \quad (4.3.23)$$

Assuming that $\mathbb{E}h^2$ is finite instead of (4.1.21) then Theorem 4.3.2 implies the consistency of tests based on sup-functionals of $\{U_{[(n+1)t_1], [(n+1)t_2]}, 0 < t_1 < t_2 < 1\}$. This means that we can consistently reject H_0 vs. $H_A^{(2)}$, when

$$\sup_{0 < t_1 < t_2 < 1} \frac{1}{n^{3/2}\bar{\sigma}} |U_{[(n+1)t_1], [(n+1)t_2]}| \text{ is too large,}$$

except in the case when $\theta_1 = \theta_2 = \theta_3 = \theta_{1,2} = \theta_{1,3} = \theta_{2,3} = 0$.

$u_{\lambda_1, \lambda_2}(t_1, t_2)$ is equal to 0 if and only if all $\theta_{i,j}$'s involved are equal to 0. To show this we observe that each of the six parts of the limiting function above may be written as

$$u_{\lambda_1, \lambda_2}(t_1, t_2) = \theta_{m_1} t_1^2 + \theta_{m_2} t_2^2 + A_j t_1 t_2 + B_j t_1 + C_j t_2 + D_j, \quad 1 \leq j \leq 6,$$

where A_j , B_j , C_j and D_j depend on $\theta_1, \theta_2, \theta_3, \theta_{1,2}, \theta_{1,3}$ and $\theta_{2,3}$ and $1 \leq m_1 \leq m_2 \leq 3$ have to be chosen properly. Moreover, $u_{\lambda_1, \lambda_2}(t_1, t_2) = 0$, $0 < t_1 < t_2 < 1$, if and only if $\theta_1 = 0$, $\theta_2 = 0$, $\theta_{2,3} = 0$, $A_j = 0$, $B_j = 0$, $C_j = 0$ and $D_j = 0$, $1 \leq j \leq 6$. But combining $\theta_{1,2} = \theta_{1,3} = \theta_{2,3} = 0$ with the arguments from the antisymmetric case we get consistency except in the trivial case.

We observe that if $\theta_1 = \theta_2 = \theta_3 = \theta_{1,2} = \theta_{1,3} = \theta_{2,3} = \theta$, then

$$\begin{aligned} u_{\lambda_1, \lambda_2}(t_1, t_2) &= \sum_{j=1}^2 \sum_{i=j}^2 (t_{i+1} - t_i)(t_j - t_{j-1})\theta \\ &= ((t_2 - t_1 + t_3 - t_2)(t_1 - t_0) + (t_3 - t_2)(t_2 - t_1))\theta \\ &= ((1 - t_1)t_1 + (1 - t_2)(t_2 - t_1))\theta \\ &= ((t_2 - t_1)t_1 + (1 - t_2)t_2)\theta, \end{aligned}$$

which is the limit ($n \rightarrow \infty$) of the expected value of $\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]}$ under H_0 (cf. (4.1.7)). Moreover, it follows then that under the null hypothesis H_0 of no-change

$$\begin{aligned} \frac{1}{n^2} U_{[(n+1)t_1], [(n+1)t_2]} &= \frac{1}{n^2} \left(Z_{[(n+1)t_1], [(n+1)t_2]} - \left(([(n+1)t_2] - [(n+1)t_1]) \right. \right. \\ &\quad \left. \left. \times [(n+1)t_1] + (n - [(n+1)t_2]) [(n+1)t_2] \right) \theta \right) \\ &= \frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2]} - \frac{([(n+1)t_2] - [(n+1)t_1])}{n^2} \\ &\quad \times [(n+1)t_1] + (n - [(n+1)t_2]) [(n+1)t_2] \theta \\ &\xrightarrow{P} u_{\lambda_1, \lambda_2}(t_1, t_2) - ((t_2 - t_1)t_1 + (1 - t_2)t_2)\theta \\ &= 0. \end{aligned}$$

Thus on assuming that $\theta_1 = \theta_2 = \theta_3 = \theta_{1,2} = \theta_{1,3} = \theta_{2,3} = \theta = 0$, the sequence

$$T_n = \frac{1}{n^{3/2} \bar{\sigma}} \sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_1], [(n+1)t_2]}|, \quad n \in \mathbb{N}, \quad (4.3.24)$$

is not consistent against any class of alternatives. On the other hand, if at least one

$\theta_{i,j}$ is not equal to 0 and we use T_n , then

$$\mathbb{P}\{H_0 \text{ is rejected} | H_A^{(2)} \text{ is true}\} \xrightarrow{n \rightarrow \infty} 1.$$

This implies that the limits of the sequence $\{T_n\}_{n \in \mathbb{N}}$ are different in probability under H_0 and $H_A^{(2)}$, namely we have consistency of $\{T_n\}_{n \in \mathbb{N}}$.

Example 4.3.1 We consider T_n and let c_α be such that

$$\mathbb{P}\left\{\frac{1}{n^{3/2}\tilde{\sigma}} \sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_1], [(n+1)t_2]}| \geq c_\alpha | H_0 \text{ is true}\right\} = \alpha$$

for some fixed $0 < \alpha < 1$ and $n \in \mathbb{N}$ large enough. We apply Theorem 4.3.2 and get that

$$\begin{aligned} \mathbb{P}\{H_0 \text{ is rejected when using } T_n | H_A^{(2)} \text{ is true}\} &= \\ &= \mathbb{P}\left\{\frac{1}{n^{3/2}\tilde{\sigma}} \sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_1], [(n+1)t_2]}| \geq c_\alpha | H_A^{(2)} \text{ is true}\right\} \\ &= \mathbb{P}\left\{\frac{1}{n^2} \sup_{0 < t_1 < t_2 < 1} |U_{[(n+1)t_1], [(n+1)t_2]}| \geq \frac{\tilde{\sigma}c_\alpha}{\sqrt{n}} | H_A^{(2)} \text{ is true}\right\} \\ &\geq \mathbb{P}\left\{\frac{1}{n^2} |U_{[(n+1)t_1], [(n+1)t_2]}| \geq \frac{\tilde{\sigma}c_\alpha}{\sqrt{n}} | H_A^{(2)} \text{ is true}\right\} \\ &\rightarrow 1 \quad \text{as } n \rightarrow \infty, \end{aligned}$$

where t_1 and t_2 are picked such that $\frac{1}{n^2} U_{[(n+1)t_1], [(n+1)t_2]} \not\rightarrow 0$. Hence, T_n in this example is, for $n \in \mathbb{N}$ large enough, unbiased at any level α and, as a sequence in n it is consistent against the class of all alternatives where at least one $\theta_{i,j} \neq 0$.

4.4 Changes in the Mean

We are to test the *no-change in the mean* null-hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables
with $\mathbb{E}X_i = \mu$ and $0 < \sigma^2 = \text{Var}X_i < \infty$, $1 \leq i \leq n$,

against the *at most two changes in the mean* alternative

$H_A^{(2)} : X_1, \dots, X_n$ are independent random variables and there are the
two integers τ_1 and τ_2 , $1 \leq \tau_1 \leq \tau_2 < n$, such that $\mathbb{E}X_1 = \dots =$
 $\mathbb{E}X_{\tau_1} \neq \mathbb{E}X_{\tau_1+1} = \dots = \mathbb{E}X_{\tau_2}$, $\mathbb{E}X_{\tau_1+1} = \dots = \mathbb{E}X_{\tau_2} \neq \mathbb{E}X_{\tau_2+1} =$
 $\dots = \mathbb{E}X_n$ and $0 < \sigma^2 = \text{Var}X_1 = \dots = \text{Var}X_n < \infty$.

Taking simulated values of the therein indicated (k, X_k) , Figure 4.4.1 gives an example where the mean changes two times while the variance stays the same.

Figure 4.4.1: The data $X_1, \dots, X_{300}$ are i.i.d. $N(0, 1)$ -distributed, $X_{301}, \dots, X_{700}$ are i.i.d. $N(3, 1)$ -distributed and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(2, 1)$ -distributed.

Similarly as in the case of testing for at most one change-point, testing for at most two change-points in the mean may be illustrated by using a geometrical argument. Consider the linear function $m(t) := t$, $t \in \mathbb{R}$, which joins under H_0 all the points $(k, \frac{1}{\mu} \mathbb{E}\{S(k)\})$, $k \in \mathbb{N}$, if $\mu \neq 0$, and it joins all the points $(k, \mathbb{E}\{S(k)\})$, $k \in \mathbb{N}$, if $\mu = 0$.

Figure 4.4.2: A geometrical interpretation of $\mathbb{E}\{k_2 S(k_1) + (n - k_1) S(k_2) - k_2 S(n)\} = 0$ under H_0 .

Without loss of generality let $\mu = 1$. Then in Figure 4.4.2 we join all the points $(k, \mathbb{E}\{S(k)\})$, $k \in \mathbb{N}$, via the straight line $m(t) = t$. We pick one $k_1 \in \{1, \dots, n - 2\}$ and then one $k_2 \in \{k_1 + 1, \dots, n - 1\}$.⁵ We draw a horizontal line starting from $B := (0, \mathbb{E}\{S(k_1)\})$, containing the point $(k_1, \mathbb{E}\{S(k_1)\})$, and with terminus, $C := (k_2, \mathbb{E}\{S(k_1)\})$. We draw a vertical line from the terminus and intersect the t -axis. We denote this intersection by $D := (k_2, \mathbb{E}\{S(0)\})$, where we define $S(0) := 0$. In this way we construct a rectangle, denoted by $ABCD$ (see Figure 4.4.2), where $A := (0, \mathbb{E}\{S(0)\})$, with length k_2 and height $\mathbb{E}\{S(k_1)\}$. We also draw a horizontal line starting from $F := (k_1, \mathbb{E}\{S(k_2)\})$, containing the point $(k_2, \mathbb{E}\{S(k_2)\})$, and with terminus, $G := (n, \mathbb{E}\{S(k_2)\})$. We then draw a vertical line from the starting point and another one from the terminus, both intersecting the t -axis at $E := (k_1, \mathbb{E}\{S(0)\})$ and $H := (n, \mathbb{E}\{S(0)\})$ respectively. Similarly, we construct another rectangle, denoted by $EFGH$ (see Figure 4.4.2), with length

⁵In Figure 4.4.2 we may think of k_1 and k_2 being defined as $k_1 := \lfloor nt_1^{(fix)} \rfloor$ and $k_2 := \lfloor nt_2^{(fix)} \rfloor$ respectively, $0 < t_1^{(fix)} < t_2^{(fix)} < 1$.

$(n - k_1)$ and height $\mathbb{E}\{S(k_2)\}$.

Reflecting each point of the rectangle $EFGH$ around the 45 degree line $m = t$, we get the new rectangle $BIJC$, where $B := (0, \mathbb{E}\{S(k_1)\})$ is the reflection point of $E := (k_1, \mathbb{E}\{S(0)\})$, $I := (0, \mathbb{E}\{S(n)\})$ is the reflection point of $H := (n, \mathbb{E}\{S(0)\})$, $J := (k_2, \mathbb{E}\{S(n)\})$ is the reflection point of $G := (n, \mathbb{E}\{S(k_2)\})$ and $C := (k_2, \mathbb{E}\{S(k_1)\})$ is the reflection point of $F := (k_1, \mathbb{E}\{S(k_2)\})$. This new rectangle has length k_2 and height $\mathbb{E}\{S(n) - S(k_1)\}$.

Combining the two rectangles $ABCD$ and $BIJC$ with each other, we have constructed the rectangle $AIJD$, which has length k_2 and height $\mathbb{E}\{S(n)\}$. Moreover, under H_0 , the new rectangle $AIJD$ has the same area as the sum of the two other ones, namely $ABCD$ and $EFGH$. Consequently, we have that

$$\begin{aligned}
 & k_2 \times \mathbb{E}\{S(k_1)\} + (n - k_1) \times \mathbb{E}\{S(k_2)\} - k_2 \times \mathbb{E}\{S(n)\} \\
 &= \mathbb{E}\{k_2 S(k_1) + (n - k_1) S(k_2) - k_2 S(n)\} \\
 &= \mathbb{E}\left\{\left((n - k_1) S(k_1) - k_1 (S(n) - S(k_1))\right) \right. \\
 &\quad \left. + \left((n - k_2) (S(k_2) - S(k_1)) - (k_2 - k_1) (S(n) - S(k_2))\right)\right\} \\
 &= \mathbb{E}\left(\sum_{i=1}^{k_1} \sum_{j=k_1+1}^n (X_i - X_j) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n (X_i - X_j)\right) \\
 &= 0, \quad 1 \leq k_1 < k_2 < n.
 \end{aligned}$$

Thus, in principle, for each given k_1 and k_2 , $1 \leq k_1 < k_2 < n$, we constructed an unbiased estimator of zero assuming that H_0 is true. We may also say that, viewed this way, testing for at most two changes in the mean results in comparing areas of three different rectangles with each other.

Continuing with using the kernel $h(x, y) = x - y$ to test for at most two changes in the mean and using the same arguments as in Section 3.4 we assume that under H_0

$$\mathbb{E}h^2(X_i, X_j) = 2\sigma^2 < \infty, \quad (4.4.1)$$

and

$$\mathbb{E}\tilde{h}^2(X_1) = \mathbb{E}\{(X_1 - \mu)^2\} = \sigma^2 > 0.$$

Since h is an *antisymmetric* kernel, we know from Theorem 4.2.2 that, as $n \rightarrow \infty$, under H_0

$$\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2) - \Gamma_n^a(t_1, t_2)| = o_P(1), \quad (4.4.2)$$

where for all $0 < t_1 < t_2 < 1$

$$\begin{aligned} \bar{U}_n(t_1, t_2) &= \frac{1}{\bar{\sigma}n^{3/2}} \bar{U}_{[(n+1)t_1], [(n+1)t_2]} \\ &= \frac{1}{\sigma n^{3/2}} Z_{[(n+1)t_1], [(n+1)t_2]} \\ &= \frac{1}{\sigma n^{3/2}} \left(\sum_{i=1}^{[(n+1)t_1]} \sum_{j=[(n+1)t_1]+1}^n (X_i - X_j) \right. \\ &\quad \left. + \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} \sum_{j=[(n+1)t_2]+1}^n (X_i - X_j) \right), \end{aligned} \quad (4.4.3)$$

and

$$\Gamma^a(t_1, t_2) = t_2 W(t_1) + (1 - t_1) W(t_2) - t_2 W(1), \quad 0 \leq t_1 \leq t_2 \leq 1. \quad (4.4.4)$$

We note that

$$\{\Gamma_n^a(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\} \stackrel{\mathcal{D}}{=} \{\Gamma^a(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\},$$

for each n .

Assuming the same conditions as in Theorem 4.2.3, we know that, as $n \rightarrow \infty$, under $H_A^{(2)}$

$$\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2)| \xrightarrow{P} \infty.$$

Therefore we consistently reject H_0 , if $\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2)|$ becomes too large. We note that (4.4.2) implies that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |\bar{U}_n(t_1, t_2)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^a(t_1, t_2)|, \quad (4.4.5)$$

and one would have to produce tables for the right hand side random variable in (4.4.5) for the sake of accepting or rejecting H_0 .

We note that by expressing (4.4.3) in terms of partial sums $S([(n+1) \cdot])$ we get that

$$\begin{aligned} \bar{U}_n(t_1, t_2) &= \frac{1}{\sigma n^{3/2}} \left([(n+1)t_2]S([(n+1)t_1]) + (n - [(n+1)t_1])S([(n+1)t_2]) \right. \\ &\quad \left. - [(n+1)t_2]S(n) \right) \\ &\approx \frac{1}{\sigma n^{1/2}} \left(t_2 S([(n+1)t_1]) + (1 - t_1)S([(n+1)t_2]) - t_2 S(n) \right), \end{aligned}$$

and now Donsker's theorem (see Theorem 1.2.1) yields (4.4.5) as well.

The variance σ^2 in (4.4.3) is usually unknown. Consequently it has to be estimated on the basis of the same random sample. One possible way of estimating σ^2 is via the sample variance, namely

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad (4.4.6)$$

where $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ is the sample mean. Since $\hat{\sigma}_n^2$ is a consistent estimator for σ^2 , $\bar{U}_n(t_1, t_2)$ in (4.4.3) may be estimated by

$$\hat{\bar{U}}_n(t_1, t_2) = \frac{1}{\hat{\sigma}_n n^{1/2}} \left(t_2 S([(n+1)t_1]) + (1 - t_1)S([(n+1)t_2]) - t_2 S(n) \right).$$

Now our previous result from (4.4.2) carries over. In particular, as $n \rightarrow \infty$, we obtain

$$\sup_{0 < t_1 < t_2 < 1} |\hat{\bar{U}}_n(t_1, t_2) - \Gamma_n^a(t_1, t_2)| = o_P(1),$$

and hence also

$$\sup_{0 < t_1 < t_2 < 1} |\hat{U}_n(t_1, t_2)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^a(t_1, t_2)|.$$

Consequently, just like in (4.4.5), tables for the distribution of $\sup_{0 < t_1 < t_2 < 1} |\Gamma^a(t_1, t_2)|$ could again be used to consistently reject the null hypothesis of no change in the data, if our test statistic $\sup_{0 < t_1 < t_2 < 1} |\hat{U}_n(t_1, t_2)|$ becomes too large.

4.5 Changes in the Variance

We are to test the *no-change in the variance* hypothesis

$$\begin{aligned} H_0 : X_1, \dots, X_n \text{ are independent identically distributed random variables} \\ \text{with } \mathbb{E}X_i = \mu \text{ and } 0 < \sigma^2 = \text{Var}X_i < \infty, 1 \leq i \leq n, \end{aligned}$$

against the *at most two changes in the variance* alternative

$$\begin{aligned} H_A^{(2)} : X_1, \dots, X_n \text{ are independent random variables and there exist the} \\ \text{two integers } \tau_1 \text{ and } \tau_2, 1 \leq \tau_1 \leq \tau_2 < n, \text{ such that } \text{Var}X_1 = \dots = \\ \text{Var}X_{\tau_1} \neq \text{Var}X_{\tau_1+1} = \dots = \text{Var}X_{\tau_2}, \text{Var}X_{\tau_1+1} = \dots = \text{Var}X_{\tau_2} \neq \\ \text{Var}X_{\tau_2+1} = \dots = \text{Var}X_n, 0 < \text{Var}X_1, \text{Var}X_{\tau_1+1}, \text{Var}X_{\tau_2+1} < \infty, \\ \text{and } \mathbb{E}X_1 = \dots = \mathbb{E}X_n = \mu. \end{aligned}$$

Taking simulated values of the therein indicated (k, X_k) , Figure 4.5.3 gives an example where the variance changes two times while the mean stays the same.

Similar arguments as in Section 3.5 suggest the use of the *symmetric* kernel

$$h(x, y) = \frac{1}{2}(x - y)^2.$$

Using the same arguments as in Section 3.5, we assume that under H_0

$$\mathbb{E}h^2(X_i, X_j) = \frac{1}{4} [2\mathbb{E}(X_i - \mu)^4 + 6(\sigma^2)^2] < \infty, \quad (4.5.1)$$

Figure 4.5.3: The data $X_1, \dots, X_{100}$ are i.i.d. $N(0, 2)$ -distributed, $X_{101}, \dots, X_{300}$ are i.i.d. $N(0, 1)$ -distributed and $X_{301}, \dots, X_{1000}$ are i.i.d. $N(0, 3)$ -distributed.

and

$$\bar{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) = \frac{1}{4}\text{Var}\left((X_1 - \mu)^2\right) > 0.$$

Since h is a *symmetric* kernel, we know from Theorem 4.3.1 that, as $n \rightarrow \infty$, under H_0

$$\sup_{0 < t_1 < t_2 < 1} |U_n(t_1, t_2) - \Gamma_n^{\text{sym}}(t_1, t_2)| = o_P(1), \quad (4.5.2)$$

where for all $0 < t_1 < t_2 < 1$

$$\begin{aligned} U_n(t_1, t_2) &= \frac{1}{\bar{\sigma}n^{3/2}} U_{[(n+1)t_1], [(n+1)t_2]} \\ &= \frac{1}{\bar{\sigma}n^{3/2}} \left\{ Z_{[(n+1)t_1], [(n+1)t_2]} - \left(([[(n+1)t_2] - [(n+1)t_1]] [(n+1)t_1] \right. \right. \\ &\quad \left. \left. + (n - [(n+1)t_2]) [(n+1)t_2] \right) \sigma^2 \right\}, \end{aligned} \quad (4.5.3)$$

with

$$\begin{aligned}
 Z_{[(n+1)t_1], [(n+1)t_2]} &= \sum_{i=1}^{[(n+1)t_1]} \sum_{j=[(n+1)t_1]+1}^n \frac{1}{2} (X_i - X_j)^2 \\
 &\quad + \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} \sum_{j=[(n+1)t_2]+1}^n \frac{1}{2} (X_i - X_j)^2, \quad 0 < t_1 < t_2 < 1, \\
 \bar{\sigma} &= \frac{1}{2} \left(\text{Var}((X_1 - \mu)^2) \right)^{1/2}, \\
 \sigma^2 &= \text{Var} X_1,
 \end{aligned}$$

and

$$\Gamma^{sym}(t_1, t_2) = (t_2 - 2t_1)W(t_1) + (1 + t_1 - 2t_2)W(t_2) + t_2W(1), \quad 0 \leq t_1 \leq t_2 \leq 1.$$

We note that

$$\{\Gamma_n^{sym}(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\} \stackrel{D}{=} \{\Gamma^{sym}(t_1, t_2), 0 \leq t_1 \leq t_2 \leq 1\},$$

for each n .

Let $S(k) = \sum_{i=1}^k X_i$, $R(k) = \sum_{i=1}^k X_i^2$, $k_1 = [(n+1)t_1]$ and $k_2 = [(n+1)t_2]$. Then $Z_{[(n+1)t_1], [(n+1)t_2]}$ from above can also be written as

$$\begin{aligned}
 Z_{k_1, k_2} &= \frac{1}{2} \left\{ (k_2 - 2k_1)R(k_1) + (n + k_1 - 2k_2)R(k_2) + k_2R(n) \right. \\
 &\quad \left. - 2(S(k_2) - S(k_1))S(k_1) - 2(S(n) - S(k_2))S(k_2) \right\}, \quad 1 \leq k_1 < k_2 < n.
 \end{aligned}$$

Assuming the same conditions as in Theorem 4.3.2 we know that under the alternative $H_A^{(2)}$, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |U_n(t_1, t_2)| \xrightarrow{P} \infty.$$

Similarly to previous sections, we have to estimate the usually unknown parameters σ^2 and $\bar{\sigma}^2$.

Again we estimate σ^2 by the usual estimator of the variance $\hat{\sigma}_n^2$ as in (3.4.6). We also have to estimate $\bar{\sigma}$ which is defined as

$$\bar{\sigma} = \frac{1}{2} \left\{ \mathbb{E}((X_1 - \mu)^2)^2 - (\mathbb{E}(X_1 - \mu)^2)^2 \right\}^{1/2}.$$

Hence, we estimate the second part of the latter equation by $\hat{\sigma}_n^2$ and the first part via the estimator for the 4-th centered sample moments

$$\hat{m}_n^4 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^4, \quad (4.5.4)$$

where $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$.

Hence, we estimate $U_n(t_1, t_2)$ in (4.5.3) by

$$\begin{aligned} \hat{U}_n(t_1, t_2) = & \frac{1}{\frac{1}{2}(\hat{m}_n^4 - (\hat{\sigma}_n^2)^2)^{1/2} n^{3/2}} \left\{ Z_{[(n+1)t_1], [(n+1)t_2]} - \left([(n+1)t_2] \right. \right. \\ & \left. \left. - [(n+1)t_1] \right) [(n+1)t_1] + (n - [(n+1)t_2]) [(n+1)t_2] \right\} \hat{\sigma}_n^2, \end{aligned}$$

Now our previous result from (4.5.2) carries over. In particular, as $n \rightarrow \infty$, we obtain

$$\sup_{0 < t_1 < t_2 < 1} |\hat{U}_n(t_1, t_2) - \Gamma_n^{sym}(t_1, t_2)| = o_P(1),$$

and hence also, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < 1} |\hat{U}_n(t_1, t_2)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^{sym}(t_1, t_2)|.$$

Thus, possible tables for the distribution of $\sup_{0 < t_1 < t_2 < 1} |\Gamma^{sym}(t_1, t_2)|$ could be used to reject the null hypothesis of no change in the data, if our test statistic $\sup_{0 < t_1 < t_2 < 1} |\hat{U}_n(t_1, t_2)|$ becomes too large.

4.6 Epidemic Alternatives

In the past two sections we studied alternatives with at most two changes in the mean and variance, respectively. Similarly, we will test for two changes and will also require that the distributions before the first and after the second change are the same. This kind of alternative, more or less as formulated by Levin and Kline (1985), has been called the epidemic alternative, on postulating that an epidemic state runs from time τ_1 through τ_2 , after which the normal state is restored. Applications of this model in econometric context are studied by Broemeling and Tsurumi (1987).

We start with the case, where we have an epidemic change in the mean as assumend in the alternative $H_A^{(2)}$ below. In particular, we test the *no-change in the mean* null-hypothesis

$$H_0 : X_1, \dots, X_n \text{ are independent identically distributed random variables} \\ \text{with } \mathbb{E}X_i = \mu \text{ and } 0 < \sigma^2 = \text{Var}X_i < \infty, 1 \leq i \leq n,$$

against the *epidemic change in the mean* alternative

$$H_A^{(2)} : X_1, \dots, X_n \text{ are independent random variables and there are two} \\ \text{integers } \tau_1 \text{ and } \tau_2, 1 < \tau_1 < \tau_2 < n, \text{ such that } \mathbb{E}X_1 = \dots = \mathbb{E}X_{\tau_1} = \\ \mathbb{E}X_{\tau_2+1} = \dots = \mathbb{E}X_n, \mathbb{E}X_{\tau_1+1} = \dots = \mathbb{E}X_{\tau_2}, \mathbb{E}X_{\tau_1} \neq \mathbb{E}X_{\tau_1+1} \text{ and} \\ 0 < \sigma^2 = \text{Var}X_1 = \dots = \text{Var}X_n < \infty.$$

This alternative postulates that the mean changed at an unknown time τ_1 and that, at another unknown time τ_2 , it changed back to its original level. Figure 4.6.4 gives an example where the mean changes at some point τ_1 , and at point τ_2 it changes back to the original mean before the first change τ_1 .

Nonparametric tests for epidemic alternatives were discussed in the literature in the past two or so decades (cf. Csörgő and Horváth (1997, Section 2.8.4) and Yao (1993)) and their related references. Levin and Kline (1985) suggested the test statistics

$$T_n^{(1)} = \max_{1 \leq k_1 < k_2 \leq n} \left(S(k_2) - S(k_1) - \frac{k_2 - k_1}{n} S(n) \right)$$

Figure 4.6.4: The data $X_1, \dots, X_{300}$ and $X_{701}, \dots, X_{1000}$ are i.i.d. $N(0, 1)$ -distributed and $X_{301}, \dots, X_{700}$ are i.i.d. $N(3, 1)$ -distributed.

and

$$T_n^{(2)} = \max_{1 \leq k_1 < k_2 \leq n} \left| S(k_2) - S(k_1) - \frac{k_2 - k_1}{n} S(n) \right|,$$

which may be interpreted as comparing the mean in the middle to the mean before and after. Using Donsker's theorem (cf. Theorem 1.2.1) we have that under H_0

$$\frac{1}{n^{1/2}\sigma} T_n^{(1)} \xrightarrow{\mathcal{D}} \sup_{0 \leq t_1 < t_2 \leq 1} (B(t_2) - B(t_1)), \quad \text{as } n \rightarrow \infty$$

and

$$\frac{1}{n^{1/2}\sigma} T_n^{(2)} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t), \quad \text{as } n \rightarrow \infty.$$

The latter convergence results uses the fact that

$$\sup_{0 \leq t_1 < t_2 \leq 1} |B(t_2) - B(t_1)| \stackrel{\mathcal{D}}{=} \sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t).$$

In a similar vein, Lombard (1987) proposed a quadratic version of the latter statistic, namely

$$T_n^{(3)} = \sum_{1 \leq k_1 < k_2 \leq n} \left(S(k_2) - S(k_1) - \frac{k_2 - k_1}{n} S(n) \right)^2.$$

Again we may use Donsker's theorem to compute its limiting distribution under H_0 , namely

$$\frac{1}{2n^3\sigma} T_n^{(3)} \xrightarrow{\mathcal{D}} \int \int_{0 \leq t_1 < t_2 \leq 1} (B(t_2) - B(t_1))^2 dt_2 dt_1, \quad \text{as } n \rightarrow \infty.$$

The likelihood ratio test statistic for testing H_0 vs. $H_A^{(2)}$, if the observations are normal with constant mean, is calculated to be

$$T_n^{(4)} = \max_{1 \leq k_1 < k_2 < n} \frac{|S(k_2) - S(k_1) - \frac{k_2 - k_1}{n} S(n)|}{\left\{ (k_2 - k_1) \left(1 - \frac{k_2 - k_1}{n} \right) \right\}^{1/2}},$$

where the distribution of $T_n^{(4)}$ is unknown. Instead of the latter expression Yao (1993) obtained large deviation approximations for

$$T_n^{(5)} = \max_{m_1 \leq k_2 - k_1 \leq m_2} \frac{S(k_2) - S(k_1) - \frac{k_2 - k_1}{n} S(n)}{\left\{ (k_2 - k_1) \left(1 - \frac{k_2 - k_1}{n} \right) \right\}^{1/2}},$$

assuming that $0 < \liminf_{n \rightarrow \infty} \frac{m_1}{n} < \limsup_{n \rightarrow \infty} \frac{m_2}{n} < 1$. This statistic is similar to $T_n^{(4)}$, but technical difficulties are avoided by trimming the endpoints. In practical situations one may face problems when choosing a proper m_1 and m_2 .

Based on the recursive residuals w_k defined by

$$w_k = \left(\frac{k-1}{k} \right)^{1/2} \left(X_k - \frac{S(k-1)}{k-1} \right), \quad 2 \leq k \leq n,$$

Yao (1993) introduced some new statistics. Let

$$\tilde{S}(k) = \sum_{i=2}^k w_i, \quad 2 \leq k \leq n,$$

and define

$$T_n^{(6)} = \max_{2 \leq k_1 < k_2 \leq n} \frac{\tilde{S}(k_2) - \tilde{S}(k_1)}{(k_2 - k_1)^{1/2}}$$

and

$$T_n^{(7)} = \max_{2 \leq k_1 < k_2 \leq n} \frac{|\tilde{S}(k_2) - \tilde{S}(k_1)|}{(k_2 - k_1)^{1/2}}.$$

The distributions of these two statistics are unknown due to technical difficulties arising by $(k_2 - k_1)^{1/2}$ in the denominator. To avoid these technical problems, Yao (1993) considered a trimmed version of $T_n^{(6)}$, namely

$$T_n^{(8)} = \max_{m_1 \leq k_2 - k_1 \leq m_2} \frac{\tilde{S}(k_2) - \tilde{S}(k_1)}{(k_2 - k_1)^{1/2}},$$

in case of normal observations and obtained large deviation approximations.

Csörgő and Horváth (1997) avoided technical difficulties by multiplying $T_n^{(6)}$ and $T_n^{(7)}$ with their denominator. Namely, similarly to the statistics $T_n^{(1)}$ and $T_n^{(2)}$ they considered the statistics

$$T_n^{(9)} = \max_{1 \leq k_1 < k_2 \leq n} (\tilde{S}(k_2) - \tilde{S}(k_1))$$

and

$$T_n^{(10)} = \max_{1 \leq k_1 < k_2 \leq n} |\tilde{S}(k_2) - \tilde{S}(k_1)|,$$

as well as their quadratic analogue

$$T_n^{(11)} = \sum_{1 \leq k_1 < k_2 \leq n} (\tilde{S}(k_2) - \tilde{S}(k_1))^2.$$

Using the fact that, as $n \rightarrow \infty$,

$$\frac{1}{n^{1/2}\sigma} \tilde{S}(nt) \xrightarrow{\mathcal{D}} B(t),$$

they obtained the following asymptotic distributions for the latter three statistics

$$\begin{aligned}\frac{1}{n^{1/2}\sigma}T_n^{(9)} &\xrightarrow{\mathcal{D}} \sup_{0 \leq t_1 < t_2 \leq 1} (B(t_2) - B(t_1)), \quad \text{as } n \rightarrow \infty, \\ \frac{1}{n^{1/2}\sigma}T_n^{(10)} &\xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq 1} B(t) - \inf_{0 \leq t \leq 1} B(t), \quad \text{as } n \rightarrow \infty,\end{aligned}$$

and

$$\frac{1}{2n^3\sigma}T_n^{(11)} \xrightarrow{\mathcal{D}} \int \int_{0 \leq t_1 < t_2 \leq 1} (B(t_2) - B(t_1))^2 dt_2 dt_1, \quad \text{as } n \rightarrow \infty,$$

which are the same as the ones for $\frac{1}{n^{1/2}\sigma}T_n^{(1)}$, $\frac{1}{n^{1/2}\sigma}T_n^{(2)}$ and $\frac{1}{2n^3\sigma}T_n^{(3)}$, respectively.

Using the results from Section 4.4 we suggest the use of the test statistic $\bar{T}_n = \bar{T}_n(X_1, X_2, \dots, X_n)$ which is defined via

$$\begin{aligned}\bar{T}_n &= \frac{1}{n} \max_{1 \leq k_1 < k_2 < n} |Z_{k_1, k_2}| \\ &= \frac{1}{n} \max_{1 \leq k_1 < k_2 < n} \left| \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n (X_i - X_j) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n (X_i - X_j) \right| \\ &= \max_{1 \leq k_1 < k_2 < n} \left| \frac{k_2}{n} S(k_1) + \frac{n - k_1}{n} S(k_2) - \frac{k_2}{n} S(n) \right|.\end{aligned}$$

By (4.4.2), or by using Donsker's theorem we have under H_0 that

$$\frac{1}{n^{1/2}\sigma} \bar{T}_n \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < 1} |\Gamma^a(t_1, t_2)|, \quad \text{as } n \rightarrow \infty,$$

where the limiting Gaussian process $\Gamma^a(t_1, t_2)$ is defined as

$$\Gamma^a(t_1, t_2) = t_2 W(t_1) + (1 - t_1) W(t_2) - t_2 W(1), \quad 0 \leq t_1 \leq t_2 \leq 1.$$

Moreover, under $H_A^{(2)}$, $\frac{1}{n^{1/2}\sigma} \bar{T}_n$ converges in probability to ∞ , since $\frac{1}{n\sigma} \bar{T}_n$ converges in probability to the supremum of the limiting function $\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2)$ defined as

$$\begin{cases} (\lambda_2 - \lambda_1)t_2\theta_{1,2}, & 0 < t_1 < t_2 \leq \lambda_1 < \lambda_2 < 1, \\ ((t_2 - \lambda_1)(1 - \lambda_2 + t_1) + (\lambda_2 - t_2)\lambda_1)\theta_{1,2}, & 0 < t_1 \leq \lambda_1 < t_2 \leq \lambda_2 < 1, \\ (\lambda_2 - \lambda_1)(1 - t_2 + t_1)\theta_{1,2}, & 0 < t_1 \leq \lambda_1 < \lambda_2 < t_2 < 1, \\ ((\lambda_2 - t_1)\lambda_1 + (1 - \lambda_2)(t_2 - \lambda_1))\theta_{1,2}, & 0 < \lambda_1 < t_1 < t_2 \leq \lambda_2 < 1, \\ ((1 - \lambda_2)(t_1 - \lambda_1) + (\lambda_2 - t_1)(1 - t_2 + \lambda_1))\theta_{1,2}, & 0 < \lambda_1 < t_1 \leq \lambda_2 < t_2 < 1, \\ (1 - t_1)(\lambda_2 - \lambda_1)\theta_{1,2}, & 0 < \lambda_1 < \lambda_2 < t_1 < t_2 < 1, \end{cases}$$

where $\tau_1 = [n\lambda_1]$, $\tau_2 = [n\lambda_2]$ and $\theta_{1,2} = \mathbb{E}(X_i - X_j) \neq 0$,⁶ $1 \leq i \leq \tau_1 < j \leq \tau_2 < n$ and $1 < \tau_1 < i \leq \tau_2 < j \leq n$.

The limiting function $\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2)$ depends on λ_1 , λ_2 , t_1 , t_2 and $\theta_{1,2}$. In Figure 4.6.5 we can see the six areas where t_1 and t_2 are defined. The areas A_1 , A_4 and A_6 are triangles, while A_2 , A_3 and A_5 are rectangles. Each of them depends on the location of at least one of the change-points λ_1 and λ_2 .

Figure 4.6.5: Summation Area of $\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2)$

⁶If $\theta_{1,2}$ were to be equal to 0 then $\frac{1}{n^{1/2}\sigma} \bar{T}_n$ is not consistent.

A special point in Figure 4.6.5 is (λ_1, λ_2) , which is the point where the areas A_2, A_3, A_4 and A_5 intersect. Moreover, assuming that $\theta_{1,2} > 0$, it turns out that the point where $\bar{u}_{\lambda_1, \lambda_2}(t_1, t_2)$ reaches its maximum is at (λ_1, λ_2) . We note that $\bar{u}_{\lambda_1, \lambda_2}(\lambda_1, \lambda_2) = (\lambda_2 - \lambda_1)(1 - \lambda_2 + \lambda_1)\theta_{1,2}$.

Figure 4.6.6: The limiting function $\bar{u}_{\frac{1}{3}, \frac{2}{3}}(t_1, t_2)$ with $\theta_{1,2} = 10$ takes its maximum value of 2.2 at the point $(\frac{1}{3}, \frac{2}{3})$.

In Figure 4.6.6 the graph of the limiting function $\bar{u}_{\frac{1}{3}, \frac{2}{3}}(t_1, t_2)$ is plotted. It can be seen that $\bar{u}_{\frac{1}{3}, \frac{2}{3}}(t_1, t_2)$ is defined in six different areas via six different functions. Three of these surfaces have the shape of a triangle, and three that of a rectangle. Moreover, the surfaces corresponding to the summation areas A_1, A_3, A_4 and A_6 are planes and the two others are not.

Figure 4.6.7 shows the limiting function $\bar{u}_{\frac{1}{10}, \frac{2}{10}}(t_1, t_2)$. Thus, in this case, we con-

sider a situation when two change-points are close to each other. This means that the epidemic change from one mean to another and back was very short. Nevertheless, the maximum is taken at point $(\lambda_1, \lambda_2) = (\frac{1}{10}, \frac{2}{10})$.

Figure 4.6.7: The limiting function $\bar{u}_{\frac{1}{10}, \frac{2}{10}}(t_1, t_2)$ with $\theta_{1,2} = 10$ takes its maximum value of 0.9 at the point $(\frac{1}{10}, \frac{2}{10})$.

Figure 4.6.8 shows the limiting function $\bar{u}_{\frac{1}{10}, \frac{9}{10}}(t_1, t_2)$, i. e., now the two change-points are far apart from each other. This means that there was a long period between the first and second change. Similar to the previous two cases the maximum is taken at point (λ_1, λ_2) .

Therefore, in addition to the test statistic $\bar{T}_n$ which tests for H_0 against $H_A^{(2)}$, we may also define an estimator for the times of change, say $\hat{\tau}$. Hence, we define as an

Figure 4.6.8: The limiting function $\bar{u}_{\frac{1}{10}, \frac{9}{10}}(t_1, t_2)$ with $\theta_{1,2} = 10$ takes its maximum value of 1.6 at the point $(\frac{1}{10}, \frac{9}{10})$.

estimator for the change-points τ_1 and τ_2

$$\hat{\tau} = \min \left\{ (k_1, k_2) : Z_{k_1, k_2} = \max_{1 \leq m_1 < m_2 < n} Z_{m_1, m_2} \right\}, \quad (4.6.1)$$

assuming that $\theta_{1,2} > 0$.⁷ If $\theta_{1,2} < 0$ then we define as an estimator for the change-points τ_1 and τ_2

$$\hat{\tau} = \min \left\{ (k_1, k_2) : Z_{k_1, k_2} = \min_{1 \leq m_1 < m_2 < n} Z_{m_1, m_2} \right\}. \quad (4.6.2)$$

⁷We use $\min\{(k_1, k_2) : \dots\}$ to denote the point (k_1, k_2) where the maximum (or minimum respectively) is taken and which has the smallest distance (euclidean norm) from the point $(0,0)$.

We note that $\hat{\tau}$ and $\tilde{\tau}$ are two-dimensional vectors.

Similarly to an epidemic change in the mean, we may assume an epidemic change in the variance. In particular, we now are to test the *no-change in the variance* null-hypothesis

$$H_0 : X_1, \dots, X_n \text{ are independent identically distributed random variables} \\ \text{with } \mathbb{E}X_i = \mu \text{ and } 0 < \sigma^2 = \text{Var}X_i < \infty, 1 \leq i \leq n,$$

against the *epidemic change in the variance* alternative

$$H_A^{(2)} : X_1, \dots, X_n \text{ are independent random variables and there are two} \\ \text{integers } \tau_1 \text{ and } \tau_2, 1 < \tau_1 < \tau_2 < n, \text{ such that } \text{Var}X_1 = \dots = \text{Var}X_{\tau_1} \\ = \text{Var}X_{\tau_2+1} = \dots = \text{Var}X_n, \text{Var}X_{\tau_1+1} = \dots = \text{Var}X_{\tau_2}, \text{Var}X_{\tau_1} \neq \\ \text{Var}X_{\tau_1+1}, 0 < \text{Var}X_{\tau_1}, \text{Var}X_{\tau_1+1} < \infty \text{ and } \mathbb{E}X_1 = \dots = \mathbb{E}X_n = \mu.$$

This alternative postulates that the variance changed at an unknown time τ_1 and that, at another unknown time τ_2 , it changed back to its original level.

Using the results from Section 4.5, we suggest the use of the test statistic $T_n = T_n(X_1, X_2, \dots, X_n)$ which is defined via

$$T_n = \frac{1}{\left(\hat{m}_n^4 - (\hat{\sigma}_n^2)^2\right)^{1/2} n^{3/2}} \max_{1 \leq k_1 < k_2 < n} \left| \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n (X_i - X_j)^2 + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n (X_i - X_j)^2 \right. \\ \left. - 2\left((k_2 - k_1)k_1 + (n - k_2)k_2\right) \hat{\sigma}_n^2 \right|,$$

where we replaced $\bar{\sigma}$ and σ^2 by the estimators suggested in Section 4.5 (cf. (4.4.6) and (4.5.4) respectively).

Assuming that the second moments of $(X_i - X_j)^2$, $1 \leq i < j \leq n$, and $\sigma^2 > 0$ are all finite then, under H_0 , T_n converges in distribution to $\sup_{0 < t_1 < t_2 < 1} |\Gamma^{sym}(t_1, t_2)|$, where the limiting Gaussian process $\Gamma^{sym}(t_1, t_2)$ is defined as

$$\Gamma^{sym}(t_1, t_2) = (t_2 - 2t_1)W(t_1) + (1 + t_1 - 2t_2)W(t_2) + t_2W(1), \\ 0 \leq t_1 \leq t_2 \leq 1.$$

Moreover, under H_A , T_n converges in probability to ∞ , since $\frac{T_n}{n^{1/2}}$ converges in

probability to the supremum of the limiting function $u_{\lambda_1, \lambda_2}(t_1, t_2)$ which is given by

$$\left\{ \begin{array}{ll} \left((t_2 - t_1)t_1 + (1 - \lambda_2 + \lambda_1 - t_2)t_2 \right) \theta_1 \\ \quad + (\lambda_2 - \lambda_1)t_2\theta_{1,2}, & 0 < t_1 < t_2 \leq \lambda_1 < \lambda_2 < 1, \\ \left((\lambda_1 - t_1)t_1 + (1 - \lambda_2)\lambda_1 \right) \theta_1 + \left((t_2 - \lambda_1) \right. \\ \quad \times (1 - \lambda_2 + t_1) + (\lambda_2 - t_2)\lambda_1 \left. \right) \theta_{1,2} \\ \quad + (\lambda_2 - t_2)(t_2 - \lambda_1)\theta_2, & 0 < t_1 \leq \lambda_1 < t_2 \leq \lambda_2 < 1, \\ \left((\lambda_1 - t_1)t_1 + (\lambda_1 - t_1 + t_2 - \lambda_2)(1 - t_2) \right) \theta_1 \\ \quad + (\lambda_2 - \lambda_1)(1 - t_2 + t_1)\theta_{1,2}, & 0 < t_1 \leq \lambda_1 < \lambda_2 < t_2 < 1, \\ (1 - \lambda_2)\lambda_1\theta_1 + \left((\lambda_2 - t_1)\lambda_1 + (1 - \lambda_2) \right. \\ \quad \times (t_2 - \lambda_1) \left. \right) \theta_{1,2} + \left((\lambda_2 - t_1)(t_1 - \lambda_1) \right. \\ \quad + (t_2 - t_1)(\lambda_2 - t_2) \left. \right) \theta_2, & 0 < \lambda_1 < t_1 < t_2 \leq \lambda_2 < 1, \\ \left((1 - \lambda_2)\lambda_1 + (t_2 - \lambda_2)(1 - t_2) \right) \theta_1 + \left((1 - \lambda_2) \right. \\ \quad \times (t_1 - \lambda_1) + (\lambda_2 - t_1)(1 - t_2 + \lambda_1) \left. \right) \theta_{1,2} \\ \quad + (t_1 - \lambda_1)(\lambda_2 - t_1)\theta_2, & 0 < \lambda_1 < t_1 \leq \lambda_2 < t_2 < 1, \\ \left((1 - t_1)(t_1 - \lambda_2 + \lambda_1) + (1 - t_2)(t_2 - t_1) \right) \theta_1 \\ \quad + (1 - t_1)(\lambda_2 - \lambda_1)\theta_{1,2}, & 0 < \lambda_1 < \lambda_2 < t_1 < t_2 < 1. \end{array} \right.$$

We note that this limiting function corresponds to the one in (4.3.23), when we put $\theta_1 = \theta_{1,3} = \theta_3$ and $\theta_{1,2} = \theta_{2,3}$. The limiting function depends on $\lambda_1, \lambda_2, t_1, t_2, \theta_1, \theta_{1,2}$ and θ_2 .

Similarly to the case of an epidemic change in the mean, the limiting function $u_{\lambda_1, \lambda_2}(t_1, t_2)$ is defined via six different functions over six different areas. Unfortunately, it is not true anymore that the maximum or minimum, respectively, of the limiting function is taken at (λ_1, λ_2) . This is due to the fact that there are two more parameters, namely θ_1 and θ_2 , involved than before. Figure 4.6.9 gives an example where neither the maximum nor the minimum is taken at (λ_1, λ_2) . Hence, an estimator for the times of change is only possible, if both changes satisfy special conditions. Such conditions can be found in a similar vein as in Section 3.3.3, where we deal with one change only.

Similar test statistics can also be found for different epidemic alternatives. The

Figure 4.6.9: The limiting function $\bar{u}_{\frac{1}{3}, \frac{3}{4}}(t_1, t_2)$ with $\theta_{1,2} = 10$, $\theta_{1,3} = 20$ and $\theta_{2,3} = -12$ takes its maximum value of 3.05 at the point $(\frac{1}{120}, \frac{1}{3})$.

main problem is to find an (anti-) symmetric kernel $h(x, y)$ which will be appropriate for the desired alternatives.

“Changing one thing for the better is worth more
than proving a thousand things are wrong.”

– *Anonymous*

Chapter 5

Multiple Change-points

5.1 Introduction

In the previous two chapters we investigated test statistics which allowed us to test for at most one or two change-points. In this chapter we generalize the main ideas obtained in these chapters, especially the one where we tested for at most two change-points, and study test statistics which allow us to test for at most s change-points, $1 \leq s < n$. Due to the increase of the number of possible change-points, the results become more complex and so do the notations. We note that proofs are done in a similar vein as in the previous chapters.

We are to test the null hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables

against the alternative that there are at most s change-points in the sequence $X_1, \dots, X_n$, namely that we have

$H_A^{(s)} : X_1, \dots, X_n$ are independent random variables and there exist s , $1 \leq s < n$, integers $\tau_1 = \tau_1(n)$, $\tau_2 = \tau_2(n)$, $\dots$, $\tau_s = \tau_s(n)$, $1 \leq \tau_1 \leq \tau_2 \leq \dots \leq \tau_s < n$, such that $\mathbb{P}\{X_1 \leq t\} = \dots = \mathbb{P}\{X_{\tau_1} \leq t\}$, $\mathbb{P}\{X_{\tau_1+1} \leq t\} = \dots = \mathbb{P}\{X_{\tau_2} \leq t\}$, $\dots$, $\mathbb{P}\{X_{\tau_s+1} \leq t\} = \dots = \mathbb{P}\{X_n \leq t\}$ for all t and $\mathbb{P}\{X_{\tau_i} \leq t_0\} \neq \mathbb{P}\{X_{\tau_i+1} \leq t_0\}$ for some t_0 and for all $1 \leq i \leq s$.

We note that, just like in the case of epidemic alternatives (cf. Section 4.6), the alternative $H_A^{(s)}$ allows us to consider random variables $X_1, X_2, \dots, X_n$ with s changes in the distribution, which do not necessarily result in $(s+1)$ different distributions.

Since we are testing for at most s , $1 \leq s < n$, changes we need to define a stochastic process which ‘feels’ the possibility of s changes. To do this we split the given sample $X_1, \dots, X_n$ into $s+1$ blocks and compare each of the blocks with the others. We continue using a kernel $h(x, y)$ of the two variables. Since, out of s , we always compare two blocks with each other, we have $\binom{s+1}{2}$ different possibilities to do so. Therefore, for the problem in hand, we define a sequence of s -time parameter stochastic processes as follows:

$$\begin{aligned}
 Z_{k_1, k_2, \dots, k_s} &:= \sum_{i=1}^{k_1} \left(\sum_{j=k_1+1}^{k_2} h(X_i, X_j) + \dots + \sum_{j=k_s+1}^n h(X_i, X_j) \right) \\
 &\quad + \sum_{i=k_1+1}^{k_2} \left(\sum_{j=k_2+1}^{k_3} h(X_i, X_j) + \dots + \sum_{j=k_s+1}^n h(X_i, X_j) \right) + \dots \\
 &\quad + \sum_{i=k_{s-1}+1}^{k_s} \sum_{j=k_s+1}^n h(X_i, X_j) \\
 &= \sum_{m=1}^s \sum_{i=k_{m-1}+1}^{k_m} \sum_{l=m}^s \sum_{j=k_l+1}^{k_{l+1}} h(X_i, X_j), \\
 &\qquad\qquad\qquad 1 \leq k_1 < k_2 < \dots < k_s < n, \quad (5.1.1)
 \end{aligned}$$

where $k_0 := 0$ and $k_{s+1} := n$. In this way we compare the $(s+1)$ blocks $(X_1, \dots, X_{k_1})$, $(X_{k_1+1}, \dots, X_{k_2})$, $\dots$, $(X_{k_s+1}, \dots, X_n)$ with each other, where the k_i ’s, $1 \leq i \leq s$, vary from i to $n+i-1-s$.

When testing for at most one change, this process reduces to

$$\begin{aligned}
 Z_{k_1} &= \sum_{m=1}^1 \sum_{i=k_{m-1}+1}^{k_m} \sum_{l=1}^1 \sum_{j=k_l+1}^{k_{l+1}} h(X_i, X_j), \\
 &= \sum_{i=k_0+1}^{k_1} \sum_{j=k_1+1}^{k_2} h(X_i, X_j), \\
 &= \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n h(X_i, X_j), \quad 1 \leq k_1 < n,
 \end{aligned}$$

the one we have already seen in Chapter 3. When testing for at most two changes the test statistic reduces to the one we studied in Chapter 4, namely to

$$\begin{aligned}
 Z_{k_1, k_2} &= \sum_{m=1}^2 \sum_{i=k_{m-1}+1}^{k_m} \sum_{l=m}^2 \sum_{j=k_l+1}^{k_{l+1}} h(X_i, X_j), \\
 &= \sum_{i=k_0+1}^{k_1} \left(\sum_{j=k_1+1}^{k_2} h(X_i, X_j) + \sum_{j=k_2+1}^{k_3} h(X_i, X_j) \right) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^{k_3} h(X_i, X_j) \\
 &= \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n h(X_i, X_j) + \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n h(X_i, X_j), \quad 1 \leq k_1 < k_2 < n.
 \end{aligned}$$

Here we study the asymptotic properties (as $n \rightarrow \infty$) of the sup-norm of the process $Z_{k_1, k_2, \dots, k_s}$, $1 \leq k_1 < k_2 < \dots < k_s < n$, which will be seen to be based on a combination of U-statistics. First we give some notations and definitions under the null hypothesis of no-change and the alternative of at most s changes.

5.1.1 Notations under the Null Hypothesis H_0

We define

$$\mathbb{E}h(X_i, X_j) =: \theta, \quad 1 \leq i < j \leq n, \quad (5.1.2)$$

and

$$\mathbb{E}h^2(X_i, X_j) =: \nu, \quad 1 \leq i < j \leq n. \quad (5.1.3)$$

We assume throughout the whole chapter that

$$\nu < \infty, \quad (5.1.4)$$

which of course implies

$$|\theta| < \infty. \quad (5.1.5)$$

Furthermore, we have that

$$\mathbb{E} \sum_{i=a+1}^b \sum_{j=c+1}^n h(X_i, X_j) = (n-c)(b-a)\theta,$$

where $0 \leq a < b \leq c < n$ and $a, b, c \in \mathbb{N}$. Hence,

$$\begin{aligned} \mathbb{E} Z_{k_1, k_2, \dots, k_s} &= \mathbb{E} \left\{ \sum_{i=1}^{k_1} \left(\sum_{j=k_1+1}^{k_2} h(X_i, X_j) + \dots + \sum_{j=k_s+1}^n h(X_i, X_j) \right) \right. \\ &\quad + \sum_{i=k_1+1}^{k_2} \left(\sum_{j=k_2+1}^{k_3} h(X_i, X_j) + \dots + \sum_{j=k_s+1}^n h(X_i, X_j) \right) + \dots \\ &\quad \left. + \sum_{i=k_{s-1}+1}^{k_s} \sum_{j=k_s+1}^n h(X_i, X_j) \right\} \\ &= \mathbb{E} \sum_{i=1}^{k_1} \sum_{j=k_1+1}^n h(X_i, X_j) + \mathbb{E} \sum_{i=k_1+1}^{k_2} \sum_{j=k_2+1}^n h(X_i, X_j) + \dots \\ &\quad + \mathbb{E} \sum_{i=k_{s-1}+1}^{k_s} \sum_{j=k_s+1}^n h(X_i, X_j) \\ &= (n-k_1)k_1\theta + (n-k_2)(k_2-k_1)\theta + (n-k_3)(k_3-k_2)\theta + \dots \\ &\quad + (n-k_s)(k_s-k_{s-1})\theta \\ &= (n-k_1)k_1\theta + (n-k_2)k_2\theta - (n-k_2)k_1\theta + (n-k_3)k_3\theta \\ &\quad - (n-k_3)k_2\theta + \dots + (n-k_s)k_s\theta - (n-k_s)k_{s-1}\theta \\ &= (k_2-k_1)k_1\theta + (k_3-k_2)k_2\theta + \dots + (n-k_s)k_s\theta \\ &= \sum_{i=1}^s (k_{i+1} - k_i)k_i\theta, \quad 1 \leq k_1 < k_2 < \dots < k_s < n, \quad (5.1.6) \end{aligned}$$

where $k_{s+1} := n$. We define, as in the previous chapters,

$$\tilde{h}(t) = \mathbb{E}\{h(t, X_2) - \theta\}, \quad (5.1.7)$$

and again we assume that

$$\tilde{\sigma}^2 = \mathbb{E}\tilde{h}^2(X_1) > 0. \quad (5.1.8)$$

We centralize $Z_{k_1, k_2, \dots, k_s}$ by its mean, and consider the process

$$U_{k_1, k_2, \dots, k_s} = Z_{k_1, k_2, \dots, k_s} - \sum_{i=1}^s (k_{i+1} - k_i) k_i \theta, \quad 1 \leq k_1 < k_2 < \dots < k_s < n, \quad (5.1.9)$$

where the kernel function h in $Z_{k_1, k_2, \dots, k_s}$ is *symmetric*. For an *antisymmetric* kernel we define

$$\bar{U}_{k_1, k_2, \dots, k_s} = Z_{k_1, k_2, \dots, k_s}, \quad 1 \leq k_1 < k_2 < \dots < k_s < n. \quad (5.1.10)$$

Moreover, we can write $Z_{k_1, k_2, \dots, k_s}$ as the sum of $(s+2)$ U-statistics, and thus we get that

$$U_{k_1, k_2, \dots, k_s} = U_n^{(s+2)} - (U_{k_1}^{(1)} + U_{k_1, k_2}^{(2)} + \dots + U_{k_s, n}^{(s+1)}), \quad 1 \leq k_1 < k_2 < \dots < k_s < n, \quad (5.1.11)$$

where

$$\begin{aligned} U_{k_1}^{(1)} &= \sum_{1 \leq i < j \leq k_1} h(X_i, X_j) - \binom{k_1}{2} \theta, \\ U_{k_1, k_2}^{(2)} &= \sum_{k_1 < i < j \leq k_2} h(X_i, X_j) - \binom{k_2 - k_1}{2} \theta, \\ &\vdots \\ U_{k_s, n}^{(s+1)} &= \sum_{k_s < i < j \leq n} h(X_i, X_j) - \binom{n - k_s}{2} \theta, \\ U_n^{(s+2)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j) - \binom{n}{2} \theta. \end{aligned}$$

Similarly, for an *antisymmetric* kernel we define

$$\bar{U}_{k_1, k_2, \dots, k_s} = Z_n^{(s+2)} - (Z_{k_1}^{(1)} + Z_{k_1, k_2}^{(2)} + \dots + Z_{k_s, n}^{(s+1)}), \quad 1 \leq k_1 < k_2 < \dots < k_s < n, \quad (5.1.12)$$

where

$$\begin{aligned}
Z_{k_1}^{(1)} &= \sum_{1 \leq i < j \leq k_1} h(X_i, X_j), \\
Z_{k_1, k_2}^{(2)} &= \sum_{k_1 < i < j \leq k_2} h(X_i, X_j), \\
&\vdots \\
Z_{k_s, n}^{(s+1)} &= \sum_{k_s < i < j \leq n} h(X_i, X_j), \\
Z_n^{(s+2)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j).
\end{aligned}$$

For further use we also define the sequence of s -time parameter stochastic processes

$$\begin{aligned}
U_n(t_1, t_2, \dots, t_s) &= \frac{1}{\bar{\sigma} n^{3/2}} U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}, \\
0 < t_1 < t_2 < \dots < t_s < 1, \quad (5.1.13)
\end{aligned}$$

and

$$\begin{aligned}
\bar{U}_n(t_1, t_2, \dots, t_s) &= \frac{1}{\bar{\sigma} n^{3/2}} \bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}, \\
0 < t_1 < t_2 < \dots < t_s < 1. \quad (5.1.14)
\end{aligned}$$

5.1.2 Notations under the Alternative $H_A^{(s)}$

Let $F^{(1)}(t) = \mathbb{P}\{X_{\tau_1} \leq t\}$, $F^{(2)}(t) = \mathbb{P}\{X_{\tau_1+1} \leq t\}$, $F^{(3)}(t) = \mathbb{P}\{X_{\tau_2+1} \leq t\}$, $\dots$, $F^{(s+1)}(t) = \mathbb{P}\{X_{\tau_s+1} \leq t\}$ be the respective distribution functions of the observations before the first, between the first and second, between the second and third, $\dots$, and after the s -th change respectively, and put

$$\mathbb{E}h(X_i, X_j) =: \theta_{q+1, r+1}, \quad (5.1.15)$$

where $\tau_q < i \leq \tau_{q+1}$ and $\tau_r < j \leq \tau_{r+1}$ for all $0 \leq q \leq r \leq s$, and where we define

$$\tau_0 := 1,$$

$$\begin{aligned}\tau_{s+1} &:= n, \\ \theta_i &:= \theta_{i,i}, \quad i = 1, \dots, s+1.\end{aligned}\tag{5.1.16}$$

Furthermore, we will put

$$\tau_i := [n\lambda_i], \quad 0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 1,\tag{5.1.17}$$

Similarly to (5.1.15), we also define the second moment of the kernel h by

$$\mathbb{E}h^2(X_i, X_j) =: \nu_{q+1, r+1}\tag{5.1.18}$$

where $\tau_q < i \leq \tau_{q+1}$ and $\tau_r < j \leq \tau_{r+1}$ for all $0 \leq q \leq r \leq s$. We assume throughout the whole chapter that $\mathbb{E}h^2(X_i, X_j)$ is finite for all possible choices of i and j , namely that

$$\nu_{i,j} < \infty, \quad 1 \leq i \leq j \leq s+1,\tag{5.1.19}$$

which implies that

$$|\theta_{i,j}| < \infty, \quad 1 \leq i \leq j \leq s+1.\tag{5.1.20}$$

For the sake of strong laws, we will also use a weaker assumption than a finite second moment of h , namely that for random variables from different distributions the following holds:

$$\mathbb{E}\{|h(X_{\tau_m}, X_{\tau_l+1})| \log^+ (|h(X_{\tau_m}, X_{\tau_l+1})|)\} < \infty, \quad 1 \leq m \leq l \leq s,\tag{5.1.21}$$

where $\log^+ x = \log(x \vee 1)$.

5.2 Antisymmetric Kernels

We consider the s -time parameter processes $Z_{k_1, k_2, \dots, k_s}$, $1 \leq k_1 < k_2 < \dots < k_s < n$, $n \geq s + 1$, where the kernel h is antisymmetric as in (2.2.3). By using the notations and assumptions from Sections 5.1.1 and 5.1.2, we are to investigate now the behavior of these processes under H_0 (cf. Section 5.2.1) and $H_A^{(s)}$ (cf. Section 5.2.2), as $n \rightarrow \infty$.

5.2.1 Asymptotic Results under H_0

We wish to study the limiting behavior of

$$\begin{aligned} \bar{U}_{k_1, k_2, \dots, k_s} &= Z_{k_1, k_2, \dots, k_s} \\ &= \sum_{m=1}^s \sum_{i=k_{m-1}+1}^{k_m} \sum_{l=m}^s \sum_{j=k_l+1}^{k_{l+1}} h(X_i, X_j), \\ &\quad 1 \leq k_1 < k_2 < \dots < k_s < n, \end{aligned}$$

in the sup-norm under the null hypothesis of no-change. For the indices $k_1, \dots, k_s$, we write $[(n+1)t_1], \dots, [(n+1)t_s]$, $0 < t_1 < t_2 < \dots < t_s < 1$, respectively.

The asymptotic behavior of $\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$, $0 < t_1 < t_2 < \dots < t_s < 1$, will be derived from the following *reduction principle* (Lemma 5.2.1), which follows from Theorem 4.2.1 by Janson and Wichura (1983).

Similarly as in case of testing for at most two change-points (cf. Section 4.2.1), we define the antisymmetric and degenerate ($\bar{\sigma}^{*2} = \mathbb{E}\bar{h}^{*2}(X_1) = 0$) kernel h^* with mean zero by

$$h^*(x, y) = h(x, y) - \bar{h}(x) + \bar{h}(y).$$

Hence, the U-statistic $Z_n^{(s+2)}$ defined in (5.1.12) can be written as

$$\begin{aligned} Z_n^{(s+2)} &= \sum_{1 \leq i < j \leq n} h(X_i, X_j) \\ &= \sum_{1 \leq i < j \leq n} h^*(X_i, X_j) + \sum_{1 \leq i < j \leq n} \bar{h}(X_i) - \sum_{1 \leq i < j \leq n} \bar{h}(X_j) \end{aligned}$$

$$\begin{aligned}
&= Z_n^{*(s+2)} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \tilde{h}(X_i) - \sum_{j=2}^n \sum_{i=1}^{j-1} \tilde{h}(X_j) \\
&= Z_n^{*(s+2)} + \sum_{i=1}^{n-1} (n-i) \tilde{h}(X_i) - \sum_{j=2}^n (j-1) \tilde{h}(X_j) \\
&= Z_n^{*(s+2)} + \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i),
\end{aligned}$$

and, moreover,

$$Z_n^{*(s+2)} = Z_n^{(s+2)} - \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i), \quad (5.2.1)$$

where $Z_n^{*(s+2)}$ is a U-statistic with an antisymmetric and degenerate kernel h^* , and Theorem 4.2.1 can be applied. We mention that the $*$ in (5.2.1), as well as in the following formulas, indicates that we consider the difference of one of the U-statistics in (5.1.12) and its corresponding projection.

We now proceed to prove the following reduction principle.

Lemma 5.2.1 *Let h be an antisymmetric kernel and $Z_{[(n+1)t_1]}^{(1)}, Z_{[(n+1)t_1], [(n+1)t_2]}^{(2)}, \dots, Z_n^{(s+2)}$ defined as in (5.1.12). Then under H_0 the following statements hold true as $n \rightarrow \infty$:*

$$\begin{aligned}
&\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| Z_{[(n+1)t_1]}^{(1)} \right. \\
&\quad \left. - \sum_{i=1}^{[(n+1)t_1]} ([(n+1)t_1] - 2i + 1) \tilde{h}(X_i) \right| = O_P(n), \\
&\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| Z_{[(n+1)t_1], [(n+1)t_2]}^{(2)} \right. \\
&\quad \left. - \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} ([(n+1)t_2] + [(n+1)t_1] - 2i + 1) \tilde{h}(X_i) \right| = O_P(n), \\
&\quad \vdots \\
&\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| Z_{[(n+1)t_s], n}^{(s+1)} \right. \\
&\quad \left. - \sum_{i=[(n+1)t_s]+1}^n (n + [(n+1)t_s] - 2i + 1) \tilde{h}(X_i) \right| = O_P(n),
\end{aligned}$$

$$\left| Z_n^{(s+2)} - \sum_{i=1}^n (n - 2i + 1) \bar{h}(X_i) \right| = O_P(n).$$

Proof of Lemma 5.2.1 The first statement follows from (4.2.1) and the fact that, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |Z_{[(n+1)t_1]}^{*(1)}| \leq \sup_{0 < t_1 < 1} |Z_{[(n+1)t_1]}^{*(1)}| = O_P(n).$$

$Z_{[(n+1)t_{i-1}], [(n+1)t_i]}^{*(i)}$, $2 \leq i \leq s$, is a U-statistic (with an antisymmetric and degenerate kernel) from $[(n+1)t_{i-1}] + 1$ to $[(n+1)t_i]$ minus a sum of i.i.d. r.v.'s. Hence, we have to shift the interval $[(n+1)t_{i-1}] + 1, [(n+1)t_i]$ to the interval $[1, (n+1)(t_i - t_{i-1})]$. Moreover, using the fact that for i.i.d. r.v.'s $X_1, X_2, \dots, X_n$

$$\begin{aligned} & \left\{ \sum_{[(n+1)t_{i-1}+1] \leq i < j \leq [(n+1)t_i]} h^*(X_i, X_j), 0 < t_{i-1} < t_i < 1 \right\} \\ & \stackrel{D}{=} \left\{ \sum_{1 \leq i < j \leq [(n+1)(t_i - t_{i-1})]} h^*(X_i, X_j), 0 < t_{i-1} < t_i < 1 \right\}, \quad 2 \leq i \leq s, \end{aligned} \quad (5.2.2)$$

for each n , we get, as $n \rightarrow \infty$,

$$\sup_{0 < t_i - t_{i-1} < 1} |Z_{[(n+1)(t_i - t_{i-1})]}^{*(1)}| = O_P(n).$$

Since (5.2.2) holds uniformly in t_{i-1} and t_i , $2 \leq i \leq s$, such that $0 < t_{i-1} < t_i < 1$, we have $\sup_{t_{i-1} < t_i < 1 + t_{i-1}} = \sup_{0 < t_{i-1} < t_i < 1}$ and we arrive, as $n \rightarrow \infty$, at

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |Z_{[(n+1)t_{i-1}], [(n+1)t_i]}^{*(i)}| \leq \sup_{0 < t_{i-1} < t_i < 1} |Z_{[(n+1)t_{i-1}], [(n+1)t_i]}^{*(i)}| = O_P(n),$$

$2 \leq i \leq s,$

which proves statements 2 to s in this lemma.

Similarly, $Z_{[(n+1)t_s], n}^{*(s+1)}$ is a U-statistic (with an antisymmetric and degenerate kernel) from $[(n+1)t_s] + 1$ to n minus a sum of i.i.d. r.v.'s. Hence, we have to shift the interval $[(n+1)t_s + 1, n]$ to the interval $[1, [n - (n+1)t_s]]$. Using now the fact

that

$$\left\{ \sum_{[(n+1)t_s+1] \leq i < j \leq n} h^*(X_i, X_j), 0 < t_s < 1 \right\} \\ \stackrel{p}{=} \left\{ \sum_{1 \leq i < j \leq [n-(n+1)t_s]} h^*(X_i, X_j), 0 < t_s < 1 \right\}, \quad (5.2.3)$$

for each n , we arrive, as $n \rightarrow \infty$, at

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |Z_{[(n+1)t_s], n}^{*(s+1)}| = O_P(n),$$

which proves statement $(s+1)$ in this lemma.

Combining (4.2.1) together with (5.2.1) we get

$$|Z_n^{*(s+2)}| = O_P(n),$$

which proves the last statement in this lemma. $\square$

Now we are in the position to show that the maximum of $\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$ minus a sum of projected i.i.d. r.v.'s is also of order n . We state the following corollary and give a detailed proof. The idea of the proof is similar to one in Corollary 3.2.1, however, here we give a more general version. This corollary clearly implies Corollary 3.2.1 and Corollary 4.2.1 when we put $s = 1$ and $s = 2$ respectively.

Corollary 5.2.1 *Let h be an antisymmetric kernel and $\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$ be defined as in (5.1.12). Under H_0 , as $n \rightarrow \infty$, we have*

$$\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| \bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} - \left\{ \left(\sum_{i=1}^s [(n+1)t_{i+1}] - [(n+1)t_{i-1}] \right) \sum_{j=1}^{[(n+1)t_i]} \bar{h}(X_j) \right\} - [(n+1)t_s] \sum_{i=1}^n \bar{h}(X_i) \right\} \right| \\ = O_P(n),$$

where $[(n+1)t_0] := 0$ and $[(n+1)t_{s+1}] := n$.

Proof of Corollary 5.2.1 Since $\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} = Z_n^{(s+2)} - (Z_{[(n+1)t_1]}^{(1)} + Z_{[(n+1)t_1], [(n+1)t_2]}^{(2)} + \dots + Z_{[(n+1)t_s], n}^{(s+1)})$, $0 \leq t_1 < t_2 < \dots < t_s < 1$, we have that

$$\begin{aligned} \kappa_n(t_1, t_2, \dots, t_s) &:= \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| \bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} \right. \\ &\quad - \left\{ \left(\sum_{i=1}^s ([(n+1)t_{i+1}] - [(n+1)t_{i-1}]) \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j) \right) \right. \\ &\quad \left. \left. - [(n+1)t_s] \sum_{i=1}^n \tilde{h}(X_i) \right\} \right| \\ &= \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| Z_n^{(s+2)} - (Z_{[(n+1)t_1]}^{(1)} \right. \\ &\quad + Z_{[(n+1)t_1], [(n+1)t_2]}^{(2)} + \dots + Z_{[(n+1)t_s], n}^{(s+1)} - \left\{ \left(\sum_{i=1}^s ([(n+1)t_{i+1}] \right. \right. \\ &\quad \left. \left. - [(n+1)t_{i-1}]) \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j) \right) - [(n+1)t_s] \sum_{i=1}^n \tilde{h}(X_i) \right\} \right| \end{aligned}$$

Define $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$ and let

$$\begin{aligned} \kappa_n^{(1)}(t_1, t_2, \dots, t_s) &:= \left(\sum_{i=1}^s ([(n+1)t_{i+1}] - [(n+1)t_{i-1}]) \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j) \right) \\ &\quad - [(n+1)t_s] \sum_{i=1}^n \tilde{h}(X_i) \end{aligned}$$

and

$$H(t_i) := \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j), \quad 0 \leq t_i \leq 1 \text{ with } H(t_0) := 0.$$

Since $t_{s+1} \rightarrow 1$ as $n \rightarrow \infty$, for convenience we write $H(1)$ instead of $H(\frac{n}{n+1})$. We rearrange the summation of $\kappa_n^{(1)}(t_1, t_2, \dots, t_s)$ and we get the following:

$$\begin{aligned} \kappa_n^{(1)}(t_1, t_2, \dots, t_s) &= \\ &= \sum_{i=1}^s ([(n+1)t_{i+1}] + [(n+1)t_i] - [(n+1)t_i] - [(n+1)t_{i-1}]) H(t_i) \\ &\quad - [(n+1)t_s] H(1) \\ &= \sum_{i=1}^s ([(n+1)t_{i+1}] + [(n+1)t_i]) H(t_i) - \sum_{i=1}^s ([(n+1)t_i] \end{aligned}$$

$$\begin{aligned}
& + [(n+1)t_{i-1}]H(t_i) - [(n+1)t_s]H(1) \\
= & \left(\sum_{i=0}^s ([(n+1)t_{i+1}] + [(n+1)t_i])H(t_i) - ([(n+1)t_1] + [(n+1)t_0])H(t_0) \right. \\
& - \left(\sum_{i=0}^s ([(n+1)t_{i+1}] + [(n+1)t_i])H(t_{i+1}) - ([(n+1)t_{s+1}] \right. \\
& \left. + [(n+1)t_s])H(t_{s+1}) \right) - [(n+1)t_s]H(1) \\
= & - \left(\sum_{i=0}^s ([(n+1)t_{i+1}] + [(n+1)t_i])(H(t_{i+1}) - H(t_i)) \right) + (n + [(n+1)t_s] \\
& - [(n+1)t_s])H(1) \\
= & n \sum_{i=1}^n \tilde{h}(X_i) - \left(\sum_{i=0}^s ([(n+1)t_{i+1}] + [(n+1)t_i]) \sum_{j=[(n+1)t_i]+1}^{[(n+1)t_{i+1}]} \tilde{h}(X_j) \right) \\
& + \sum_{j=1}^n (-2j+1)\tilde{h}(X_j) - \sum_{j=1}^{[(n+1)t_1]} (-2j+1)\tilde{h}(X_j) \\
& - \sum_{j=[(n+1)t_1]+1}^{[(n+1)t_2]} (-2j+1)\tilde{h}(X_j) - \dots - \sum_{j=[(n+1)t_s]+1}^n (-2j+1)\tilde{h}(X_j) \\
= & \sum_{j=1}^n (n-2j+1)\tilde{h}(X_j) - \left(\sum_{i=0}^s \sum_{j=[(n+1)t_i]+1}^{[(n+1)t_{i+1}]} ([(n+1)t_{i+1}] + [(n+1)t_i] \right. \\
& \left. - 2j+1) \tilde{h}(X_j) \right).
\end{aligned}$$

We are now in the position to split these sums in $\kappa_n^{(1)}(t_1, t_2, \dots, t_s)$ and apply the previous lemma. Since,

$$\begin{aligned}
\kappa_n(t_1, t_2, \dots, t_s) & = \\
= & \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| \left(Z_n^{(s+2)} - \sum_{i=1}^n (n-2i+1)\tilde{h}(X_i) \right) \right. \\
& - \left(Z_{[(n+1)t_1]}^{(1)} - \sum_{i=1}^{[(n+1)t_1]} ([(n+1)t_1] - 2i+1)\tilde{h}(X_i) \right) \\
& - \left(Z_{[(n+1)t_1], [(n+1)t_2]}^{(2)} - \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} ([(n+1)t_2] + [(n+1)t_1] - 2i+1)\tilde{h}(X_i) \right) \\
& - \dots - \left(Z_{[(n+1)t_s], [n]}^{(s+1)} - \sum_{i=[(n+1)t_s]+1}^n (n + [(n+1)t_s] - 2i+1)\tilde{h}(X_i) \right) \Big|,
\end{aligned}$$

we have that

$$\kappa_n(t_1, t_2, \dots, t_s) \leq$$

$$\begin{aligned}
&\leq \left| Z_n^{(s+2)} - \sum_{i=1}^n (n-2i+1) \tilde{h}(X_i) \right| + \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| Z_{[(n+1)t_1]}^{(1)} \right. \\
&\quad - \sum_{i=1}^{[(n+1)t_1]} ([(n+1)t_1] - 2i + 1) \tilde{h}(X_i) \Big| + \dots \\
&\quad + \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| Z_{[(n+1)t_s], [n]}^{(s+1)} \right. \\
&\quad - \sum_{i=[(n+1)t_s]+1}^n (n + [(n+1)t_s] - 2i + 1) \tilde{h}(X_i) \Big| \\
&= O_P(n).
\end{aligned}$$

□

We emphasize that $\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$ is now approximated by sums of i.i.d. r.v.'s. Hence, we now are in the position to study the asymptotic behavior of the weak convergence of $\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$.

Theorem 5.2.1 *Assume that H_0 , (2.2.3), (5.1.4) and (5.1.8) hold. Then we can define a sequence of Gaussian processes $\{\Gamma_n^a(t_1, t_2, \dots, t_s), 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1\}_{n \in \mathbb{N}}$ such that, as $n \rightarrow \infty$,*

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\bar{U}_n(t_1, t_2, \dots, t_s) - \Gamma_n^a(t_1, t_2, \dots, t_s)| = o_P(1), \quad (5.2.4)$$

and, for each n ,

$$\begin{aligned}
&\{\Gamma_n^a(t_1, t_2, \dots, t_s), 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1\} \\
&\quad \stackrel{\mathcal{D}}{=} \{\Gamma^a(t_1, t_2, \dots, t_s), 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1\},
\end{aligned}$$

where the Gaussian process Γ^a is defined via a linear combination of a standard Wiener process W as follows:

$$\begin{aligned}
\Gamma^a(t_1, t_2, \dots, t_s) &= \sum_{i=1}^s (t_{i+1} - t_{i-1}) W(t_i) - t_s W(1), \\
&\quad 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1,
\end{aligned} \quad (5.2.5)$$

where $t_0 := 0$ and $t_{s+1} := 1$.

Proof of Theorem 5.2.1 The proof is similar to that of Theorem 4.2.2. Again we use the fact that we can define a Wiener process $\{W(t), 0 \leq t < \infty\}$ such that (cf. Csörgő and Révész (1981, Theorem S.2.2.1 by Major (1979) combined with (S.2.2.2))), as $n \rightarrow \infty$,

$$\frac{1}{n^{1/2}} \sup_{0 \leq t \leq 1} \left| \frac{1}{\tilde{\sigma}} \sum_{i=1}^{[(n+1)t]} \tilde{h}(X_i) - W(nt) \right| = o_P(1).$$

Immediately, if we restrict $\sup_{0 < t_i < 1}$, $1 \leq i \leq s$, to $\sup_{0 < t_1 < t_2 < \dots < t_s < 1}$ then we also have, as $n \rightarrow \infty$,

$$\frac{1}{\tilde{\sigma} n^{1/2}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \left| \sum_{i=1}^{[(n+1)t_j]} \tilde{h}(X_i) - W(nt_j) \right| = o_P(1), \quad \text{for } j = 1, 2, \dots, s, s+1. \quad (5.2.6)$$

Combining this fact with Corollary 5.2.1 we get the desired result. $\square$

Similarly, as in Chapter 4, the upper index a symbolizes that this Gaussian process corresponds to the case where we have an antisymmetric kernel. Furthermore we note that for $0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1$

$$\begin{aligned} \mathbb{E}\Gamma^a(t_1, t_2, \dots, t_s) &= 0, \\ \text{Var}\Gamma^a(t_1, t_2, \dots, t_s) &= \sum_{i=0}^s (1 - t_{i+1} - t_i)^2 (t_{i+1} - t_i), \end{aligned}$$

and

$$\mathbb{E}\Gamma^a(t_1, t_2, \dots, t_s) \Gamma^a(r_1, r_2, \dots, r_s) = \sum_{i=1}^s \sum_{j=1}^s (t_{i+1} - t_{i-1})(r_{j+1} - r_{j-1})(t_i \wedge r_j) - t_s r_s,$$

where $t_0 = r_0 := 0$ and $t_{s+1} = r_{s+1} := 1$. The computation of the covariance is straight forward. We note that

$$\sum_{i=1}^s (r_{i+1} - r_{i-1}) r_i = r_s.$$

Theorem 5.2.1 implies that under H_0

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\bar{U}_n(t_1, t_2, \dots, t_s)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\Gamma^a(t_1, t_2, \dots, t_s)|. \quad (5.2.7)$$

Moreover, this allows us to produce tables for the unknown distribution of $\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\Gamma^a(t_1, t_2, \dots, t_s)|$ and reject the null hypothesis of no-change, if $\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\bar{U}_n(t_1, t_2, \dots, t_s)|$ becomes too large.

We mention that in case of at most one change, Theorem 5.2.1 reduces to Theorem 3.2.1 which is due to Csörgő and Horváth (1988b, 1997). This is easy to see, since with $s = 1$ we have

$$\begin{aligned} \Gamma^a(t_1) &= \sum_{i=1}^1 (t_{i+1} - t_{i-1})W(t_i) - t_1W(1) \\ &= (t_2 - t_0)W(t_1) - t_1W(1) \\ &= W(t_1) - t_1W(1) \\ &= B(t_1), \quad 0 \leq t_1 \leq 1. \end{aligned}$$

where $t_{s+1} = t_2 := 1$, $t_0 := 0$ and B is a Brownian bridge as in Theorem 3.2.1.

In case of at most two changes, namely $s = 2$, Γ^a reduces to

$$\begin{aligned} \Gamma^a(t_1, t_2) &= \sum_{i=1}^2 (t_{i+1} - t_{i-1})W(t_i) - t_2W(1) \\ &= (t_2 - t_0)W(t_1) + (t_3 - t_1)W(t_2) - t_2W(1) \\ &= t_2W(t_1) + (1 - t_1)W(t_2) - t_2W(1), \quad 0 \leq t_1 \leq t_2 \leq 1, \end{aligned}$$

where $t_{s+1} = t_3 := 1$, $t_0 := 0$ and $\Gamma^a(t_1, t_2)$ is the Gaussian process from (4.2.10) as in Theorem 4.2.2.

We are now in the position to combine test statistics for a different number of change-points. To do this, we define the i -dimensional vectors $(t_1, t_2, \dots, t_i) \in$

$(0, 1)^i \subset \mathbf{R}^i$ and $x_i \in \mathbf{R}$, $1 \leq i \leq s$, and the Sup-Euclidean norm

$$\|x_1, x_2, \dots, x_i\| = \sup_{0 < x_1 < x_2 < \dots < x_i < 1} \sqrt{x_1^2 + x_2^2 + \dots + x_i^2}. \quad (5.2.8)$$

Theorem 5.2.2 *Assume that H_0 , (2.2.3), (5.1.4) and (5.1.8) hold. Then we can define sequences of Gaussian processes $\{\Gamma_n^a(t_1), 0 \leq t_1 \leq 1\}$, $\{\Gamma_n^a(t_1, t_2), 0 \leq t_1 < t_2 \leq 1\}$, $\dots$, $\{\Gamma_n^a(t_1, t_2, \dots, t_s), 0 \leq t_1 < t_2 < \dots < t_s \leq 1, 1 \leq s < n\}$ such that with the Sup-Euclidean norm we have componentwise convergence in distribution, namely, as $n \rightarrow \infty$,*

$$(\bar{U}_n(t_1), \bar{U}_n(t_1, t_2), \dots, \bar{U}_n(t_1, t_2, \dots, t_s)) \xrightarrow{\mathcal{D}} (\Gamma^a(t_1), \Gamma^a(t_1, t_2), \dots, \Gamma^a(t_1, t_2, \dots, t_s)),$$

where for each n and $1 \leq i \leq s$

$$\begin{aligned} & \{\Gamma_n^a(t_1, t_2, \dots, t_i), 0 \leq t_1 \leq t_2 \leq \dots \leq t_i \leq 1\} \\ & \stackrel{\mathcal{D}}{=} \{\Gamma^a(t_1, t_2, \dots, t_i), 0 \leq t_1 \leq t_2 \leq \dots \leq t_i \leq 1\}. \end{aligned}$$

Proof of Theorem 5.2.3

$$\begin{aligned} & \|\bar{U}_n(t_1) - \Gamma_n^a(t_1), \bar{U}_n(t_1, t_2) - \Gamma_n^a(t_1, t_2), \dots, \bar{U}_n(t_1, t_2, \dots, t_s) - \Gamma_n^a(t_1, t_2, \dots, t_s)\| = \\ & = \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \sqrt{(\bar{U}_n(t_1) - \Gamma_n^a(t_1))^2 + \dots + (\bar{U}_n(t_1, t_2, \dots, t_s) - \Gamma_n^a(t_1, t_2, \dots, t_s))^2} \\ & \leq \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \left\{ \max_{1 \leq i \leq s} \sqrt{s} \cdot |\bar{U}_n(t_1, t_2, \dots, t_i) - \Gamma_n^a(t_1, t_2, \dots, t_i)| \right\} \\ & = \sqrt{s} \cdot \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \left\{ \max_{1 \leq i \leq s} |\bar{U}_n(t_1, t_2, \dots, t_i) - \Gamma_n^a(t_1, t_2, \dots, t_i)| \right\} \\ & = o_P(1) \end{aligned}$$

This implies that we have componentwise convergence in distribution of the $\bar{U}_n$'s to the appropriate Γ^a 's using the appropriate norm. $\square$

5.2.2 Asymptotic Results under $H_A^{(s)}$

We wish to study the limiting behavior of $Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$, $0 < t_1 < t_2 < \dots < t_s < 1$, from (5.1.1). To do so, we will introduce the limiting function of $\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$, $0 < t_1 < t_2 < \dots < t_s < 1$. We will show that (cf. Theorem 5.2.3), as $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} \xrightarrow{P} \bar{u}_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s),$$

$$0 < t_1 < t_2 < \dots < t_s < 1, \quad (5.2.9)$$

where we will assume that $h(x, y)$ is a nondegenerate antisymmetric kernel and (5.1.21) holds. Therefore, in this context, (5.1.21) replaces the stronger assumption of a second finite moment as in (5.1.19) which was used to derive (cf. Section 5.2.1) the convergence in distribution result, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\bar{U}_n(t_1, t_2, \dots, t_s)| \xrightarrow{D} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\Gamma^a(t_1, t_2, \dots, t_s)|,$$

where the stochastic process $\bar{U}_n(t_1, t_2, \dots, t_s)$ was defined as

$$\bar{U}_n(t_1, t_2, \dots, t_s) = \frac{1}{\sigma n^{3/2}} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$$

$$0 < t_1 < t_2 < \dots < t_s < 1,$$

and the limiting Gaussian process $\Gamma^a(t_1, t_2, \dots, t_s)$ as

$$\Gamma^a(t_1, t_2, \dots, t_s) = \sum_{i=1}^s (t_{i+1} - t_{i-1}) W(t_i) - t_s W(1),$$

$$0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1,$$

with $t_0 := 0$ and $t_{s+1} := 1$.

The limiting non-random function in (5.2.9) will depend on the location of the change-points $[n\lambda_1], [n\lambda_2], \dots, [n\lambda_s]$, $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 1$. Moreover, we will see that $\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\bar{U}_n(t_1, t_2, \dots, t_s)|$ is consistent and goes to infinity in probability, as $n \rightarrow \infty$, under $H_A^{(s)}$.

In a similar vein as in Section 5.3.2, we define

$$\begin{aligned}
& S(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \dots, x_{s+1}, x_{s+4}, x_{s+5}, \dots, x_{2s+3}) \\
&:= \sum_{i=[(n+1)x_1]+1}^{[(n+1)x_{s+2}]} \sum_{j=[(n+1)x_{s+3}]+1}^{[(n+1)x_{2s+4}]} h(X_i, X_j) \\
&= \left(\sum_{i=[(n+1)x_1]+1}^{[(n+1)x_2]} + \sum_{i=[(n+1)x_2]+1}^{[(n+1)x_3]} + \dots + \sum_{i=[(n+1)x_{s+1}]+1}^{[(n+1)x_{s+2}]} \right) \left(\sum_{j=[(n+1)x_{s+3}]+1}^{[(n+1)x_{s+4}]} \right. \\
&\quad \left. + \sum_{j=[(n+1)x_{s+4}]+1}^{[(n+1)x_{s+5}]} + \dots + \sum_{j=[(n+1)x_{2s+2}]+1}^{[(n+1)x_{2s+3}]} + \sum_{j=[(n+1)x_{2s+3}]+1}^{[(n+1)x_{2s+4}]} \right) h(X_i, X_j),
\end{aligned} \tag{5.2.10}$$

where $0 \leq x_1 < x_2 \leq x_3 \leq \dots \leq x_{s+2} \leq x_{s+3} < x_{s+4} \leq x_{s+5} \leq \dots \leq x_{2s+4} < 1$. Here $x_2, \dots, x_{s+1}, x_{s+4}, \dots, x_{2s+2}$ and x_{2s+3} play the role of dummy variables and are used to split blocks that may contain change-points. At most s of them will actually be used, but since the location of the change-points is unknown, we need to consider s possible changes in each of the blocks $(X_{[(n+1)x_1]+1}, \dots, X_{[(n+1)x_{s+4}]})$ and $(X_{[(n+1)x_{s+3}]+1}, \dots, X_{[(n+1)x_{2s+4}]})$. Hence, the convergence result from the following Section 5.3.2 can be applied to (5.2.10) and we get

$$\begin{aligned}
& \frac{1}{n^2} S(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \dots, x_{s+1}, x_{s+4}, x_{s+5}, \dots, x_{2s+3}) \\
& \xrightarrow{P} \bar{\eta}(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \dots, x_{s+1}, x_{s+4}, x_{s+5}, \dots, x_{2s+3}) \\
&:= \sum_{q=s+3}^{2s+3} (x_{q+1} - x_q) \sum_{r=1}^{s+1} (x_{r+1} - x_r) \theta_{l(x_{r+1}), l(x_{q+1})}.
\end{aligned} \tag{5.2.11}$$

where $\theta_{1,1} = \dots = \theta_{s+1,s+1} = \theta = 0$ and¹

$$l(z) := \begin{cases} 1, & 0 < z \leq \lambda_1, \\ 2, & \lambda_1 < z \leq \lambda_2, \\ \vdots & \\ s, & \lambda_{s-1} < z \leq \lambda_s, \\ s+1, & \lambda_s < z \leq 1. \end{cases} \tag{5.2.12}$$

¹We note that $\theta_{1,1} := \theta_1, \dots, \theta_{s+1,s+1} := \theta_{s+1}$ which was defined in (5.1.15).

Moreover, if the distribution between two different changes is the same then, by the property of antisymmetric kernels, the corresponding $\theta_{i,j}$ is also equal to 0. For example, if the distribution between the second and third and between the sixth and seventh change-point are the same, then $\theta_{3,7} = 0$.

We are now in the position to go back to the definition of $Z_{[(n+1)t_1],[(n+1)t_2],\dots,[(n+1)t_s]}$, $0 < t_1 < t_2 \dots < t_s < 1$ in (5.1.1), and write it in terms of double sums $S(\cdot)$ as in (5.2.10). Thus we arrive at

$$\begin{aligned} Z_{[(n+1)t_1],[(n+1)t_2],\dots,[(n+1)t_s]} &= \sum_{j=1}^s \sum_{i=j}^s S(t_{j-1}, t_j, t_i, t_{i+1}, x_2, x_3, \dots, x_{s+1}, \\ &\quad x_{s+4}, x_{s+5}, \dots, x_{2s+3}), \\ &\quad 0 < t_1 < \dots < t_s < 1, \end{aligned} \quad (5.2.13)$$

where we define $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$. Since we do not know the location of the change-points $[n\lambda_1], [n\lambda_2], \dots, [n\lambda_s]$, where $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 1$, we define the following functions $a_i = a_i(\lambda_1, \lambda_2, \dots, \lambda_s, t_1, t_2, \dots, t_s)$, $1 \leq i \leq (s+1)s$,² which will be used to derive a formula for (5.2.9) that will allow us to handle all possible combinations of $\lambda_1, \dots, \lambda_s$ and $t_1, \dots, t_s$:

$$\begin{aligned} a_1 &:= \begin{cases} \lambda_1, & \lambda_1 \leq t_1, \\ c_1, & t_1 < \lambda_1, \end{cases} \\ a_2 &:= \begin{cases} \lambda_2, & \lambda_1 \leq \lambda_2 \leq t_1, \\ c_2, & t_1 < \lambda_2, \end{cases} \\ &\vdots \\ a_s &:= \begin{cases} \lambda_s, & \lambda_1 \leq \dots \leq \lambda_s \leq t_1, \\ c_s, & t_1 < \lambda_s, \end{cases} \end{aligned}$$

²More precise, the functions a_i have to be written as follows: $a_1 = a_1(\lambda_1, t_1)$, $a_2 = a_2(\lambda_1, \lambda_2, t_1)$, $\dots$, $a_s = a_s(\lambda_1, \dots, \lambda_s, t_1)$ and $a_j = a_j(\lambda_1, \dots, \lambda_s, t_{(j-1) \bmod(s)}, t_{((j-1) \bmod(s))+1})$, $s+1 \leq j \leq s \cdot s$, and $a_{s \cdot s+1} = a_{s \cdot s+1}(\lambda_1, \dots, \lambda_s, t_s)$, $\dots$, $a_{s \cdot s+s} = a_{s \cdot s+s}(\lambda_1, \dots, \lambda_s, t_s)$ with mod beeing the modulo function.

$$\begin{aligned}
a_{s+1} &:= \begin{cases} \lambda_1, & t_1 < \lambda_1 \leq t_2, \\ \lambda_2, & \lambda_1 \leq t_1 < \lambda_2 \leq t_2, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-1} \leq t_1 < \lambda_s \leq t_2, \\ c_{s+1}, & \text{otherwise,} \end{cases} \\
a_{s+2} &:= \begin{cases} \lambda_2, & t_1 < \lambda_1 \leq \lambda_2 \leq t_2, \\ \lambda_3, & \lambda_1 \leq t_1 < \lambda_2 \leq \lambda_3 \leq t_2, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-2} \leq t_1 < \lambda_{s-1} \leq \lambda_s \leq t_2, \\ c_{s+2}, & \text{otherwise,} \end{cases} \\
&\vdots \\
a_{s+s} &:= \begin{cases} \lambda_s, & t_1 < \lambda_1 \leq \dots \leq \lambda_s \leq t_2, \\ c_{s+s}, & \text{otherwise,} \end{cases} \\
a_{2s+1} &:= \begin{cases} \lambda_1, & t_2 < \lambda_1 \leq t_3, \\ \lambda_2, & \lambda_1 \leq t_2 < \lambda_2 \leq t_3, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-1} \leq t_2 < \lambda_s \leq t_3, \\ c_{2s+1}, & \text{otherwise,} \end{cases} \\
&\vdots \\
a_{2s+s} &:= \begin{cases} \lambda_s, & t_2 < \lambda_1 \leq \dots \leq \lambda_s \leq t_3, \\ c_{2s+s}, & \text{otherwise,} \end{cases} \\
a_{s \cdot s+1} &:= \begin{cases} \lambda_1, & t_s < \lambda_1 < 1, \\ \lambda_2, & \lambda_1 \leq t_s < \lambda_2 < 1, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-1} \leq t_s < \lambda_s < 1, \\ c_{s \cdot s+1}, & \text{otherwise,} \end{cases} \\
&\vdots \\
a_{s \cdot s+s} &:= \begin{cases} \lambda_s, & t_s < \lambda_1 \leq \dots \leq \lambda_s < 1, \\ c_{s \cdot s+s}, & \text{otherwise,} \end{cases}
\end{aligned} \tag{5.2.14}$$

where $0 < a_1 \leq a_2 \leq \dots \leq a_{s \cdot s+s} < 1$ and $c_i \in [0, 1]$, $1 \leq i \leq (s+1)s$.

We need to define these a_i 's, $1 \leq i \leq (s+1)s$, since there are exactly $(s+1)s$ pos-

sibilities to place s change-points in $(s+1)$ blocks. Moreover, exactly s of the latter a_i 's will change to one of the values λ_j , $1 \leq j \leq s$, while the other s^2 a_i 's will get the value c_i and they will drop out of the limiting function $\bar{u}_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$.

We are now in the position to state the following theorem which is an immediate consequence of the above arguments in this section.

Theorem 5.2.3 *Assume that (2.2.3), (5.1.20), (5.1.21), and $H_A^{(s)}$ hold. Define $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$. If $\tau_i = \tau_i(n) := [n\lambda_i]$, $i = 1, 2, \dots, s$, $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 1$, then, as $n \rightarrow \infty$,*

$$\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} \xrightarrow{P} \bar{u}_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s),$$

where

$$\begin{aligned} \bar{u}_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s) &= \sum_{j=1}^s \sum_{i=j}^s \bar{\eta}(t_{j-1}, t_j, t_i, t_{i+1}, a_{s(j-1)+1}, a_{s(j-1)+2}, \dots, a_{s(j-1)+s}, \\ &\quad a_{si+1}, a_{si+2}, \dots, a_{si+s}), \\ &\quad 0 < t_1 < t_2 < \dots < t_s < 1. \end{aligned} \quad (5.2.15)$$

and $\bar{\eta}$ and the a_i 's, $1 \leq i \leq s(s+1)$, are defined in (5.2.11) and (5.2.14), respectively.

We note that in Theorem 5.2.3 $\theta_{1,1} = \theta_{2,2} = \dots = \theta_{s+1,s+1} = \theta = 0$. Moreover, if the distributions between the $(i-1)$ -th and i -th and between the $(j-1)$ -th and j -th change are the same, then we also have $\theta_{i,j} = \theta = 0$.

The limiting function in (5.2.15) is defined for every possible combination of $\lambda_1, \dots, \lambda_s$ and $t_1, \dots, t_s$. As we will discuss it in Section 5.2.2, the limiting function $\bar{u}_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$ involves $\binom{2s}{s}$ different cases. Furthermore, exactly $\binom{s+2}{2}$ different $\theta_{i,j}$'s, $1 \leq i \leq j \leq s+1$, appear. Removing the ones which are 0, namely $\theta_{1,1}, \theta_{2,2}, \dots, \theta_{s,s}$ and $\theta_{s+1,s+1}$, then exactly $\binom{s+1}{2}$ different $\theta_{i,j}$'s, $1 \leq i < j \leq s+1$, appear.

Assuming that $\mathbb{E}h^2$ is finite instead of (5.1.21), then Theorem 5.2.3 implies the consistency of tests based on sup-functionals of $\{\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}, 0 < t_1 <$

$t_2 < \dots < t_s < 1\}$. This means that we can consistently reject H_0 vs. $H_A^{(s)}$, when

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} \frac{1}{n^{3/2}\bar{\sigma}} |\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}| \text{ is too large,}$$

except in the case when $\theta_{i,j} = 0$, $1 \leq i \leq j \leq s+1$.

$u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$ is equal to 0 if and only if all $\theta_{i,j}$'s involved are equal to 0. Otherwise, by using the fact that $\lambda_1, \lambda_2, \dots, \lambda_{s-1}$ and λ_s are fixed a priori between 0 and 1 and do not depend on t_i , $1 \leq i \leq s$, one can show that there is at least one combination of $\lambda_1, \dots, \lambda_s$ and $t_1, \dots, t_s$ such that $u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$ is not equal to 0.

We observe that if $\theta_{i,j} = \theta$ for all possible choices of $1 \leq i \leq j \leq s+1$, then

$$\begin{aligned} \bar{u}_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s) &= \sum_{j=1}^s \sum_{i=j}^s (t_{i+1} - t_i)(t_j - t_{j-1})\theta \\ &= 0, \end{aligned}$$

since $\theta = 0$. Moreover, it follows that in this case under the null hypothesis H_0 of no-change

$$\begin{aligned} \frac{1}{n^2} \bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} &= \frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} \\ &\xrightarrow{P} \bar{u}_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s) \\ &= 0. \end{aligned}$$

Thus on assuming that $\theta_{i,j} = \theta = 0$ for all possible choices of $1 \leq i \leq j \leq n$, the sequence

$$\bar{T}_n = \frac{1}{n^{3/2}\bar{\sigma}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\bar{U}_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}|, \quad n \in \mathbb{N}, \quad (5.2.16)$$

is not consistent against any class of alternatives. On the other hand, if at least one $\theta_{i,j}$ is not equal to 0 and we use $\bar{T}_n$, then

$$\mathbb{P}\{H_0 \text{ is rejected when using } \bar{T}_n | H_A^{(s)} \text{ is true}\} \xrightarrow{n \rightarrow \infty} 1.$$

This implies that the limits of the sequence $\{\bar{T}_n\}_{n \in \mathbb{N}}$ in this case are different in probability under H_0 and $H_A^{(s)}$, and hence we have consistency of $\{\bar{T}_n\}_{n \in \mathbb{N}}$.

5.3 Symmetric Kernels

We consider the s -time parameter processes $U_{k_1, k_2, \dots, k_s}$, $1 \leq k_1 < k_2 < \dots < k_s < n$, $n \geq s + 1$, where the kernel h is symmetric as in (2.2.2). By using the notations and assumptions from Sections 5.1.1 and 5.1.2, we investigate the behavior of these processes under H_0 (cf. Section 5.3.1) and $H_A^{(s)}$ (cf. Section 5.3.2), as $n \rightarrow \infty$.

5.3.1 Asymptotic Results under H_0

We wish to study the limiting behavior of

$$\begin{aligned} U_{k_1, k_2, \dots, k_s} &= Z_{k_1, k_2, \dots, k_s} - \sum_{i=1}^s (k_{i+1} - k_i) k_i \theta, \\ &= \sum_{m=1}^s \sum_{i=k_{m-1}+1}^{k_m} \sum_{l=m}^s \sum_{j=k_l+1}^{k_{l+1}} h(X_i, X_j) - \sum_{i=1}^s (k_{i+1} - k_i) k_i \theta, \\ &\quad 1 \leq k_1 < k_2 < \dots < k_s < n, \end{aligned}$$

in the sup-norm under the null hypothesis of no-change. Again we will write $[(n+1)t_1], \dots, [(n+1)t_s]$, $0 < t_1 < t_2 < \dots < t_s < 1$, instead of the indices $k_1, \dots, k_s$.

The asymptotic behavior of $U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$, $0 < t_1 < t_2 < \dots < t_s < 1$, will be derived from the following *reduction principle* (Lemma 5.3.1). It is a consequence of Corollary 4.3.1 of Hull (1979).

Similarly, as in case of testing for at most two change-points (cf. Section 4.3.1) we define the symmetric degenerate ($\tilde{\sigma}^{*2} = \mathbb{E} \tilde{g}_\theta^{*2}(X_1) = 0$) kernel g_θ^* with mean 0 as

$$g_\theta^*(x, y) = h(x, y) - \theta - \bar{h}(x) - \bar{h}(y).$$

Hence the centralized U-statistic $U_n^{(s+2)}$ defined in (5.1.11) can be written as

$$\begin{aligned}
 U_n^{(s+2)} &= \sum_{1 \leq i < j \leq n} (h(X_i, X_j) - \theta) \\
 &= \sum_{1 \leq i < j \leq n} g_\theta^*(X_i, X_j) + \sum_{1 \leq i < j \leq n} \tilde{h}(X_i) + \sum_{1 \leq i < j \leq n} \tilde{h}(X_j) \\
 &= U_n^{*(s+2)} + \sum_{i=1}^{n-1} \sum_{j=i+1}^n \tilde{h}(X_i) + \sum_{j=2}^n \sum_{i=1}^{j-1} \tilde{h}(X_j) \\
 &= U_n^{*(s+2)} + \sum_{i=1}^{n-1} (n-i) \tilde{h}(X_i) + \sum_{j=2}^n (j-1) \tilde{h}(X_j) \\
 &= U_n^{*(s+2)} + \sum_{i=1}^n (n-1) \tilde{h}(X_i),
 \end{aligned}$$

which implies that

$$U_n^{*(s+2)} = U_n^{(s+2)} - (n-1) \sum_{i=1}^n \tilde{h}(X_i), \quad (5.3.1)$$

where $U_n^{*(s+2)}$ is a centralized U-statistic with a symmetric and degenerate kernel g_θ^* , and Corollary 4.3.1 can be applied.

We now proceed to prove the following reduction principle.

Lemma 5.3.1 *Let h be a symmetric kernel and $U_{[(n+1)t_1]}^{(1)}$, $U_{[(n+1)t_1], [(n+1)t_2]}^{(2)}$, $\dots$, $U_n^{(s+2)}$ defined as in (5.1.11). Then under H_0 the following statements hold true as $n \rightarrow \infty$:*

$$\begin{aligned}
 &\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| U_{[(n+1)t_1]}^{(1)} \right. \\
 &\quad \left. - [(n+1)t_1] \sum_{i=1}^{[(n+1)t_1]} \tilde{h}(X_i) \right| = O_P(n), \\
 &\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| U_{[(n+1)t_1], [(n+1)t_2]}^{(2)} \right. \\
 &\quad \left. - ([[(n+1)t_2] - [(n+1)t_1]]) \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} \tilde{h}(X_i) \right| = O_P(n), \\
 &\quad \vdots \\
 &\max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| U_{[(n+1)t_s], n}^{(s+1)} \right.
 \end{aligned}$$

$$\begin{aligned}
- (n - [(n+1)t_s]) \sum_{i=[(n+1)t_s]+1}^n \tilde{h}(X_i) &= O_P(n), \\
|U_n^{(s+2)} - n \sum_{i=1}^n \tilde{h}(X_i)| &= O_P(n).
\end{aligned}$$

Proof of Lemma 5.3.1 The first statement follows from Corollary 4.3.1 and the fact that, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_{[(n+1)t_1]}^{*(1)}| \leq \sup_{0 < t_1 < 1} |U_{[(n+1)t_1]}^{*(1)}| = O_P(n),$$

and combined with similar arguments as in (4.3.6) we have, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_{[(n+1)t_1]}^{(1)} - [(n+1)t_1] \sum_{i=1}^{[(n+1)t_1]} \tilde{h}(X_i)| = O_P(n).$$

$U_{[(n+1)t_{i-1}], [(n+1)t_i]}^{*(i)}$, $2 \leq i \leq s$, is a centralized U-statistic (with a symmetric and degenerate kernel) from $[(n+1)t_{i-1}] + 1$ to $[(n+1)t_i]$ minus a sum of i.i.d. r.v.'s. Hence, we have to shift the interval $[(n+1)t_{i-1}] + 1, [(n+1)t_i]$ to the interval $[1, [(n+1)(t_i - t_{i-1})]]$. Moreover, using the fact that for i.i.d. r.v.'s $X_1, X_2, \dots, X_n$

$$\begin{aligned}
&\left\{ \sum_{[(n+1)t_{i-1}]+1 \leq i < j \leq [(n+1)t_i]} h^*(X_i, X_j), 0 < t_{i-1} < t_i < 1 \right\} \\
&\stackrel{D}{=} \left\{ \sum_{1 \leq i < j \leq [(n+1)(t_i - t_{i-1})]} h^*(X_i, X_j), 0 < t_{i-1} < t_i < 1 \right\}, \quad 2 \leq i \leq s, \quad (5.3.2)
\end{aligned}$$

for each n , we get, as $n \rightarrow \infty$,

$$\sup_{0 < t_i - t_{i-1} < 1} |U_{[(n+1)(t_i - t_{i-1})]}^{*(1)}| = O_P(n).$$

Since (5.3.2) holds uniformly in t_{i-1} and t_i , $2 \leq i \leq s$, such that $0 < t_{i-1} < t_i < 1$, we have $\sup_{t_{i-1} < t_i < 1+t_{i-1}} = \sup_{0 < t_{i-1} < t_i < 1}$ and we arrive, as $n \rightarrow \infty$, at

$$\begin{aligned}
\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_{[(n+1)t_{i-1}], [(n+1)t_i]}^{*(i)}| &\leq \sup_{0 < t_{i-1} < t_i < 1} |U_{[(n+1)t_{i-1}], [(n+1)t_i]}^{*(i)}| = O_P(n), \\
&2 \leq i \leq s.
\end{aligned}$$

Similar arguments as in (4.3.6) yield, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_{[(n+1)t_{i-1}], [(n+1)t_i]}^{(i)} - ([(n+1)t_i] - [(n+1)t_{i-1}]) \sum_{i=[(n+1)t_{i-1}+1]}^{[(n+1)t_i]} \tilde{h}(X_i)| = O_P(n), \quad 2 \leq i \leq s,$$

which proves statements 2 to s in this lemma.

Similarly, $U_{[(n+1)t_s], n}^{*(s+1)}$ is a centralized U-statistic (with a symmetric and degenerate kernel) from $[(n+1)t_s] + 1$ to n minus a sum of i.i.d. r.v.'s. Hence, we have to shift the interval $[[(n+1)t_s + 1], n]$ to the interval $[1, [n - (n+1)t_s]]$. Using now the fact that

$$\left\{ \sum_{[(n+1)t_s+1] \leq i < j \leq n} h^*(X_i, X_j), 0 < t_s < 1 \right\} \stackrel{D}{=} \left\{ \sum_{1 \leq i < j \leq [n - (n+1)t_s]} h^*(X_i, X_j), 0 < t_s < 1 \right\}, \quad (5.3.3)$$

for each n , we arrive, as $n \rightarrow \infty$, at

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_{[(n+1)t_s], n}^{*(s+1)}| = O_P(n).$$

Moreover, we have

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_{[(n+1)t_s], n}^{(s+1)} - (n - [(n+1)t_s]) \sum_{i=[(n+1)t_s+1]}^n \tilde{h}(X_i)| = O_P(n),$$

which proves statement $(s+1)$ in this lemma.

Corollary 4.3.1 combined with (5.3.1) leads to

$$|U_n^{*(s+2)}| = O_P(n).$$

Using the definition of $U_n^{*(s+2)}$ we get

$$|U_n^{(s+2)} - (n-1) \sum_{i=1}^n \tilde{h}(X_i)| = O_P(n), \quad (5.3.4)$$

and by using (4.3.6) this reduces to

$$|U_n^{(s+2)} - n \sum_{i=1}^n \tilde{h}(X_i)| = O_P(n),$$

which proves the last statement in this lemma. $\square$

Now we are in the position to show that the maximum of $U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$ minus a sum of projected i.i.d. r.v.'s is also of order n . We state the following corollary and give a detailed proof. The idea of the proof is similar to one in Corollary 3.2.1, however, here we give a more general version. This corollary clearly implies Corollary 3.3.1 and Corollary 4.3.2 when we put $s = 1$ and $s = 2$ respectively.

Corollary 5.3.1 *Let h be a symmetric kernel and $U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$ be defined as in (5.1.11). Under H_0 , as $n \rightarrow \infty$, we have*

$$\begin{aligned} \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} & \left| U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} - \left\{ \left(\sum_{i=1}^s [(n+1)t_{i+1}] \right. \right. \right. \\ & \left. \left. + [(n+1)t_{i-1}] - 2[(n+1)t_i] \right) \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j) \right\} + [(n+1)t_s] \sum_{i=1}^n \tilde{h}(X_i) \right\} \\ & = O_P(n), \end{aligned}$$

where $[(n+1)t_0] := 0$ and $[(n+1)t_{s+1}] := n$.

Proof of Corollary 5.3.1 Since $U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} = U_n^{(s+2)} - (U_{[(n+1)t_1]}^{(1)} + U_{[(n+1)t_1], [(n+1)t_2]}^{(2)} + \dots + U_{[(n+1)t_s], n}^{(s+1)})$, $0 \leq t_1 < t_2 < \dots < t_s < 1$, we have that

$$\begin{aligned} \kappa_n(t_1, t_2, \dots, t_s) &:= \\ &:= \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} \right. \\ &\quad - \left\{ \left(\sum_{i=1}^s [(n+1)t_{i+1}] + [(n+1)t_{i-1}] - 2[(n+1)t_i] \right) \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j) \right\} \\ &\quad \left. + [(n+1)t_s] \sum_{i=1}^n \tilde{h}(X_i) \right\} \\ &= \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| U_n^{(s+2)} - (U_{[(n+1)t_1]}^{(1)} + U_{[(n+1)t_1], [(n+1)t_2]}^{(2)} \right. \end{aligned}$$

$$\begin{aligned}
& + \dots + U_{[(n+1)t_s], n}^{(s+1)} - \left\{ \left(\sum_{i=1}^s ([(n+1)t_{i+1}] + [(n+1)t_{i-1}]) \right. \right. \\
& \quad \left. \left. - 2[(n+1)t_i] \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j) \right) + [(n+1)t_s] \sum_{i=1}^n \tilde{h}(X_i) \right\} |
\end{aligned}$$

Recall that $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$ and let

$$\begin{aligned}
\kappa_n^{(2)}(t_1, t_2, \dots, t_s) &:= \left(\sum_{i=1}^s ([(n+1)t_{i+1}] + [(n+1)t_{i-1}] - 2[(n+1)t_i]) \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j) \right) \\
&\quad + [(n+1)t_s] \sum_{i=1}^n \tilde{h}(X_i)
\end{aligned}$$

and

$$H(t_i) := \sum_{j=1}^{[(n+1)t_i]} \tilde{h}(X_j), \quad 0 \leq t_i \leq 1 \text{ with } H(t_0) := 0,$$

where again we will write $H(1)$ instead of $H(\frac{n}{n+1})$ on account of being interested in n large. We rearrange the summation of $\kappa_n^{(2)}(t_1, t_2, \dots, t_s)$ and we get the following:

$$\begin{aligned}
\kappa_n^{(2)}(t_1, t_2, \dots, t_s) &= \\
&= \sum_{i=1}^s ([(n+1)t_{i+1}] - [(n+1)t_i]) H(t_i) - \sum_{i=1}^s ([(n+1)t_i] - [(n+1)t_{i-1}]) H(t_i) \\
&\quad + [(n+1)t_s] H(1) \\
&= \left(\sum_{i=0}^s ([(n+1)t_{i+1}] - [(n+1)t_i]) H(t_i) - ([(n+1)t_1] - [(n+1)t_0]) H(t_0) \right) \\
&\quad - \left(\sum_{i=0}^s ([(n+1)t_{i+1}] - [(n+1)t_i]) H(t_{i+1}) \right. \\
&\quad \left. - ([(n+1)t_{s+1}] - [(n+1)t_s]) H(t_{s+1}) \right) + [(n+1)t_s] H(1) \\
&= - \left(\sum_{i=0}^s ([(n+1)t_{i+1}] - [(n+1)t_i]) (H(t_{i+1}) - H(t_i)) \right) \\
&\quad + (n - [(n+1)t_s] + [(n+1)t_s]) H(1) \\
&= n \sum_{i=1}^n \tilde{h}(X_i) - \left(\sum_{i=0}^s ([(n+1)t_{i+1}] - [(n+1)t_i]) \sum_{j=[(n+1)t_i]+1}^{[(n+1)t_{i+1}]} \tilde{h}(X_j) \right)
\end{aligned}$$

We are now in the position to split these sums in $\kappa_n^{(2)}(t_1, t_2, \dots, t_s)$ and apply the

previous lemma. Since,

$$\begin{aligned}
\kappa_n(t_1, t_2, \dots, t_s) &= \\
&= \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| \left(U_n^{(s+2)} - n \sum_{i=1}^n \tilde{h}(X_i) \right) \right. \\
&\quad - \left(U_{[(n+1)t_1]}^{(1)} - [(n+1)t_1] \sum_{i=1}^{[(n+1)t_1]} \tilde{h}(X_i) \right) \\
&\quad - \left(U_{[(n+1)t_1], [(n+1)t_2]}^{(2)} - ([(n+1)t_2] - [(n+1)t_1]) \sum_{i=[(n+1)t_1]+1}^{[(n+1)t_2]} \tilde{h}(X_i) \right) \\
&\quad \left. - \dots - \left(U_{[(n+1)t_s], [n]}^{(s+1)} - (n - [(n+1)t_s]) \sum_{i=[(n+1)t_s]+1}^n \tilde{h}(X_i) \right) \right|,
\end{aligned}$$

we have that

$$\begin{aligned}
\kappa_n(t_1, t_2, \dots, t_s) &\leq \left| U_n^{(s+2)} - n \sum_{i=1}^n \tilde{h}(X_i) \right| + \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| U_{[(n+1)t_1]}^{(1)} \right. \\
&\quad \left. - [(n+1)t_1] \sum_{i=1}^{[(n+1)t_1]} \tilde{h}(X_i) \right| + \dots \\
&\quad + \max_{1 \leq [(n+1)t_1] \leq [(n+1)t_2] \leq \dots \leq [(n+1)t_s] \leq n} \left| U_{[(n+1)t_s], [n]}^{(s+1)} \right. \\
&\quad \left. - (n - [(n+1)t_s]) \sum_{i=[(n+1)t_s]+1}^n \tilde{h}(X_i) \right| \\
&= O_P(n).
\end{aligned}$$

□

We now have $U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$ approximated by sums of i.i.d. r.v.'s, and thus we can easily study the asymptotic behavior of the weak convergence of the s -time parameter stochastic process $U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$.

Theorem 5.3.1 *Assume that H_0 , (2.2.2), (5.1.4) and (5.1.8) hold. Then we can define a sequence of Gaussian processes $\{\Gamma_n^{sym}(t_1, t_2, \dots, t_s), 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1\}_{n \in \mathbb{N}}$ such that, as $n \rightarrow \infty$,*

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_n(t_1, t_2, \dots, t_s) - \Gamma_n^{sym}(t_1, t_2, \dots, t_s)| = o_P(1), \quad (5.3.5)$$

and, for each n ,

$$\begin{aligned} & \{\Gamma_n^{sym}(t_1, t_2, \dots, t_s), 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1\} \\ & \stackrel{\mathcal{D}}{=} \{\Gamma^{sym}(t_1, t_2, \dots, t_s), 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1\}, \end{aligned}$$

where the Gaussian process Γ^{sym} is defined via a linear combination of a standard Wiener processes W as follows:

$$\begin{aligned} \Gamma^{sym}(t_1, t_2, \dots, t_s) &= \sum_{i=1}^s (t_{i+1} + t_{i-1} - 2t_i)W(t_i) + t_s W(1), \\ & \quad 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1, \end{aligned} \quad (5.3.6)$$

where $t_0 := 0$ and $t_{s+1} := 1$.

Proof of Theorem 5.3.1 Combining Corollary 5.3.1 with (5.2.6) gives the desired result. $\square$

We note that for $0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1$

$$\begin{aligned} \mathbb{E}\Gamma^{sym}(t_1, t_2, \dots, t_s) &= 0, \\ \text{Var}\Gamma^{sym}(t_1, t_2, \dots, t_s) &= \sum_{i=0}^s (1 - t_{i+1} + t_i)^2 (t_{i+1} - t_i), \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}\Gamma^{sym}(t_1, t_2, \dots, t_s)\Gamma^{sym}(r_1, r_2, \dots, r_s) &= \sum_{i=1}^s \sum_{j=1}^s (t_{i+1} - 2t_i + t_{i-1})(r_{j+1} - 2r_j \\ & \quad + r_{j-1})(t_i \wedge r_j) + 2t_s \left(\sum_{i=1}^{s-1} r_{i+1}r_i - \sum_{i=1}^s r_i^2 \right) + 2r_s \left(\sum_{i=1}^{s-1} t_{i+1}t_i - \sum_{i=1}^s t_i^2 \right) + 3t_s r_s, \end{aligned}$$

where $t_0 = r_0 := 0$ and $t_{s+1} = r_{s+1} := 1$. To compute the covariance, we use the fact that

$$\sum_{i=1}^s (r_{i+1} + r_{i-1} - 2r_i)r_i = 2 \sum_{i=1}^{s-1} r_{i+1}r_i - 2 \sum_{i=1}^s r_i^2 + r_s.$$

Theorem 5.3.1 implies that under H_0

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_n(t_1, t_2, \dots, t_s)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\Gamma^{sym}(t_1, t_2, \dots, t_s)|. \quad (5.3.7)$$

Moreover, this allows us to produce tables for the unknown distribution of $\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\Gamma^{sym}(t_1, t_2, \dots, t_s)|$ and reject the null hypothesis of no-change, if $\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_n(t_1, t_2, \dots, t_s)|$ becomes too large.

We mention that in case of at most one change, Theorem 5.3.1 reduces to Theorem 3.3.1 which is due to Csörgő and Horváth (1988b, 1997). This is easy to see, since with $s = 1$ we have

$$\begin{aligned} \Gamma^{sym}(t_1) &= \sum_{i=1}^1 (t_{i+1} + t_{i-1} - 2t_i)W(t_i) + t_1 W(1) \\ &= (t_2 + t_0 - 2t_1)W(t_1) + t_1 W(1) \\ &= (1 - 2t_1)W(t_1) + t_1 W(1) \\ &= \Gamma(t_1), \quad 0 \leq t_1 \leq 1, \end{aligned}$$

where $t_{s+1} = t_2 := 1$, $t_0 := 0$ and Γ is the Gaussian process from (3.3.1) as in Theorem 3.3.1.

In case of at most two changes, namely $s = 2$, Γ^{sym} reduces to

$$\begin{aligned} \Gamma^{sym}(t_1, t_2) &= \sum_{i=1}^2 (t_{i+1} + t_{i-1} - 2t_i)W(t_i) + t_2 W(1) \\ &= (t_2 + t_0 - 2t_1)W(t_1) + (t_3 + t_1 - 2t_2)W(t_2) + t_2 W(1) \\ &= (t_2 - 2t_1)W(t_1) + (1 + t_1 - 2t_2)W(t_2) + t_2 W(1), \\ &\quad 0 \leq t_1 \leq t_2 \leq 1, \end{aligned} \quad (5.3.8)$$

where $t_{s+1} = t_3 := 1$, $t_0 := 0$ and $\Gamma^{sym}(t_1, t_2)$ is the Gaussian process from (4.3.10) as in Theorem 4.3.1.

Furthermore, we mention that the Gaussian processes Γ^a from (5.2.5) and Γ^{sym}

from (5.3.6) are different and their relationship is as follows:

$$\Gamma^a(t_1, t_2, \dots, t_s) = \Gamma^{sym}(t_1, t_2, \dots, t_s) + 2 \left(\sum_{i=1}^s (t_i - t_{i-1}) W(t_i) - t_s W(1) \right),$$

$$0 < t_1 < t_2 < \dots < t_s < 1.$$

We are now also in the position to combine test statistics for a different number of change-points and get the following theorem:

Theorem 5.3.2 *Assume that H_0 , (2.2.2), (5.1.4) and (5.1.8) hold. Then we can define sequences of Gaussian processes $\{\Gamma_n^{sym}(t_1), 0 \leq t_1 \leq 1\}$, $\{\Gamma_n^{sym}(t_1, t_2), 0 \leq t_1 < t_2 \leq 1\}$, ..., $\{\Gamma_n^{sym}(t_1, t_2, \dots, t_s), 0 \leq t_1 < t_2 < \dots < t_s \leq 1, 1 \leq s < n\}$ such that with the Sup-Euclidean norm as in (5.2.8) we have, as $n \rightarrow \infty$,*

$$\begin{aligned} & \left(U_n(t_1), U_n(t_1, t_2), \dots, U_n(t_1, t_2, \dots, t_s) \right) \\ & \xrightarrow{\mathcal{D}} \left(\Gamma^{sym}(t_1), \Gamma^{sym}(t_1, t_2), \dots, \Gamma^{sym}(t_1, t_2, \dots, t_s) \right), \end{aligned}$$

where for each n and $1 \leq i \leq s$

$$\begin{aligned} & \{ \Gamma_n^{sym}(t_1, t_2, \dots, t_i), 0 \leq t_1 \leq t_2 \leq \dots \leq t_i \leq 1 \} \\ & \xrightarrow{\mathcal{D}} \{ \Gamma^{sym}(t_1, t_2, \dots, t_i), 0 \leq t_1 \leq t_2 \leq \dots \leq t_i \leq 1 \}. \end{aligned}$$

The proof of this theorem goes along the lines of the proof of Theorem 5.2.2.

5.3.2 Asymptotic Results under $H_A^{(s)}$

We wish to study the limiting behavior of $Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}, 0 < t_1 < t_2 < \dots < t_s < 1$, from (5.1.1). To do so, we will introduce the limiting function of $\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}, 0 < t_1 < t_2 < \dots < t_s < 1$. We will show that (cf. Theorem 5.3.3), as $n \rightarrow \infty$,

$$\begin{aligned} \frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} & \xrightarrow{P} u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s), \\ & 0 < t_1 < t_2 < \dots < t_s < 1, \end{aligned} \quad (5.3.9)$$

where we will assume that $h(x, y)$ is a nondegenerate symmetric kernel and (5.1.21) holds. Therefore, in this context, (5.1.21) replaces the stronger assumption of a second finite moment as in (5.1.19) which was used to derive (cf. Section 5.3.1) the convergence in distribution statement, as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_n(t_1, t_2, \dots, t_s)| \xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} |\Gamma^{sym}(t_1, t_2, \dots, t_s)|,$$

where the stochastic process $U_n(t_1, t_2, \dots, t_s)$ was defined as

$$\begin{aligned} U_n(t_1, t_2, \dots, t_s) &= \frac{1}{\tilde{\sigma} n^{3/2}} \left(Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} - \sum_{i=1}^s \left([(n+1)t_{i+1}] - \right. \right. \\ &\quad \left. \left. - [(n+1)t_i] \right) \cdot [(n+1)t_i] \theta \right), \\ &\quad 0 < t_1 < t_2 < \dots < t_s < 1, \end{aligned}$$

with $t_{s+1} := \frac{n}{n+1}$ and the limiting Gaussian process $\Gamma^{sym}(t_1, t_2, \dots, t_s)$ as

$$\begin{aligned} \Gamma^{sym}(t_1, t_2, \dots, t_s) &= \sum_{i=1}^s (t_{i+1} + t_{i-1} - 2t_i) W(t_i) + t_s W(1), \\ &\quad 0 \leq t_1 \leq t_2 \leq \dots \leq t_s \leq 1, \end{aligned}$$

with $t_0 := 0$ and $t_{s+1} := 1$.

The limiting function in (5.3.9) will depend on the location of the change-points $[n\lambda_1], [n\lambda_2], \dots, [n\lambda_s]$, $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 1$. Moreover, we will see that $\sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_n(t_1, t_2, \dots, t_s)|$ is consistent and goes to infinity in probability, as $n \rightarrow \infty$, under $H_A^{(s)}$.

The general limiting function in (5.3.9) will involve many variables, since we have to handle all possible combinations of t_i 's and λ_i 's. This is due to the fact that we have to consider every possible combination of change-points.

When looking at the definition of $Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$, $0 < t_1 < t_2 < \dots < t_s < 1$ in (5.1.1), we see that we may split it into many double sums, where each of

these double sums is a sum of the form

$$\sum_{i=a+1}^b \sum_{j=c+1}^d h(X_i, X_j), \quad (5.3.10)$$

where $0 \leq a < b \leq c < d \leq n$ and $a, b, c, d \in \mathbb{N}$. These double sums may be associated with comparing the two blocks $(X_{a+1}, \dots, X_b)$ and $(X_{c+1}, \dots, X_d)$ with each other. In the case of testing for s changes, we have to compare each of the blocks $(X_1, \dots, X_{k_1})$, $(X_{k_1+1}, \dots, X_{k_2})$, $\dots$, $(X_{k_s+1}, \dots, X_n)$ with each other. Of course, each of these $s+1$ blocks may contain s change-points and therefore we split each of the sums in (5.3.10) into $s+1$ sums. Since we do not know the location of the s change-points in advance, we write the latter sum as follows,

$$\begin{aligned} \sum_{i=a+1}^b \sum_{j=c+1}^d h(X_i, X_j) &= \left(\sum_{i=a+1}^{\gamma_1} + \sum_{i=\gamma_1+1}^{\gamma_2} + \dots + \sum_{i=\gamma_{s-1}+1}^{\gamma_s} + \sum_{i=\gamma_s+1}^b \right) \left(\sum_{j=c+1}^{\gamma_{s+1}} \right. \\ &\quad \left. + \sum_{j=\gamma_{s+1}+1}^{\gamma_{s+2}} + \dots + \sum_{j=\gamma_{2s-1}+1}^{\gamma_{2s}} + \sum_{j=\gamma_{2s}+1}^d \right) h(X_i, X_j), \end{aligned} \quad (5.3.11)$$

where $0 \leq a < \gamma_1 \leq \gamma_2 \leq \dots \leq \gamma_s \leq b \leq c < \gamma_{s+1} \leq \gamma_{s+2} \leq \dots \leq \gamma_{2s} \leq d \leq n$.

Consequently, we are comparing blocks with each other that do not have a change inside. When using (5.3.11), the double sum in (5.3.10) is split into $(s+1)^2$ double sums. Again we emphasise that there are at most s changes in total which implies that some of the new small blocks may be viewed as bigger ones, since there is no change inside the blocks, nor in between them. Consider the case where we have $(s-2)$ changes in the block $(X_{a+1}, \dots, X_b)$ and 2 changes in the block $(X_{c+1}, \dots, X_d)$ then (5.3.11) reduces to

$$\begin{aligned} \sum_{i=a+1}^b \sum_{j=c+1}^d h(X_i, X_j) &= \left(\sum_{i=a+1}^{\gamma_1} + \sum_{i=\gamma_1+1}^{\gamma_2} + \dots + \sum_{i=\gamma_{s-2}+1}^{\gamma_{s-2}} + \sum_{i=\gamma_{s-2}+1}^b \right) \left(\sum_{j=c+1}^{\gamma_{s+1}} \right. \\ &\quad \left. + \sum_{j=\gamma_{s+1}+1}^{\gamma_{s+2}} + \sum_{j=\gamma_{s+2}+1}^d \right) h(X_i, X_j), \end{aligned}$$

where the blocks $(X_{a+1}, \dots, X_{\gamma_1})$, $(X_{\gamma_1+1}, \dots, X_{\gamma_2})$, $\dots$, $(X_{\gamma_{s-2}+1}, \dots, X_b)$ and $(X_{c+1},$

$\dots, X_{\gamma_{s+1}}), (X_{\gamma_{s+1}+1}, \dots, X_{\gamma_{s+2}}) (X_{\gamma_{s+2}+1}, \dots, X_d)$ do not contain a change-point.

Since we do not know in advance the location of the change-points, we consider sums as in (5.3.11). For technical purposes, we define the following function

$$l(z) := \begin{cases} 1, & 0 < z \leq \lambda_1, \\ 2, & \lambda_1 < z \leq \lambda_2, \\ \vdots & \\ s, & \lambda_{s-1} < z \leq \lambda_s, \\ s+1, & \lambda_s < z \leq 1, \end{cases} \quad (5.3.12)$$

which will be used to remind ourselves of the location (either before the first or between the first and second or between the second and third or $\dots$ or after the last change-point) of a block of r.v.'s which does not contain any change-point.

We split (5.3.11) into $(s+1)^2$ different double sums which are of the form

$$\sum_{i=R_1+1}^{R_2} \sum_{j=R_3+1}^{R_4} h(X_i, X_j),$$

where $1 \leq R_1 = R_1(n) := [nr_1] < R_2 = R_2(n) := [nr_2] \leq R_3 = R_3(n) := [nr_3] < R_4 = R_4(n) := [nr_4] \leq n$ are chosen properly according to the double sums in (5.3.11). Furthermore we know from Section 4.3.2 as an immediate consequence of Theorem 2.6.2 by Sen (1977) and Hoeffding's SLLN (cf. Theorem 2.6.1) that for each of these sums we have that, as $n \rightarrow \infty$,

$$\frac{1}{(R_2 - R_1)(R_4 - R_3)} \sum_{i=R_1+1}^{R_2} \sum_{j=R_3+1}^{R_4} h(X_i, X_j) \xrightarrow{P} \theta_{l(r_2), l(r_4)}. \quad (5.3.13)$$

The proof of (5.3.13) was discussed in the case of at most two change-points (cf. Section 4.3.2), and may also be applied here.

Similarly to (5.3.11), define now

$$S(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \dots, x_{s+1}, x_{s+4}, x_{s+5}, \dots, x_{2s+3})$$

$$\begin{aligned}
&:= \sum_{i=[(n+1)x_1]+1}^{[(n+1)x_{s+2}]} \sum_{j=[(n+1)x_{s+3}]+1}^{[(n+1)x_{2s+4}]} h(X_i, X_j) \\
&= \left(\sum_{i=[(n+1)x_1]+1}^{[(n+1)x_2]} + \sum_{i=[(n+1)x_2]+1}^{[(n+1)x_3]} + \dots + \sum_{i=[(n+1)x_{s+1}]+1}^{[(n+1)x_{s+2}]} \right) \left(\sum_{j=[(n+1)x_{s+3}]+1}^{[(n+1)x_{s+4}]} \right. \\
&\quad \left. + \sum_{j=[(n+1)x_{s+4}]+1}^{[(n+1)x_{s+5}]} + \dots + \sum_{j=[(n+1)x_{2s+2}]+1}^{[(n+1)x_{2s+3}]} + \sum_{j=[(n+1)x_{2s+3}]+1}^{[(n+1)x_{2s+4}]} \right) h(X_i, X_j),
\end{aligned} \tag{5.3.14}$$

where $0 \leq x_1 < x_2 \leq x_3 \leq \dots \leq x_{s+2} \leq x_{s+3} < x_{s+4} \leq x_{s+5} \leq \dots \leq x_{2s+4} < 1$.

Moreover, since (5.3.14) is a sum of double sums, (5.3.13) can be applied many times and the limiting function of (5.3.14) can be written as ($n \rightarrow \infty$)

$$\begin{aligned}
&\frac{1}{n^2} S(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \dots, x_{s+1}, x_{s+4}, x_{s+5}, \dots, x_{2s+3}) \\
&\xrightarrow{P} \eta(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \dots, x_{s+1}, x_{s+4}, x_{s+5}, \dots, x_{2s+3}) \\
&:= \sum_{q=s+3}^{2s+3} (x_{q+1} - x_q) \sum_{r=1}^{s+1} (x_{r+1} - x_r) \theta_{l(x_{r+1}), l(x_{q+1})}.
\end{aligned} \tag{5.3.15}$$

We note that if $\theta_{l(x_{r+1}), l(x_{q+1})} = \theta$ for all possible choices of r and q then

$$\begin{aligned}
&\eta(x_1, x_{s+2}, x_{s+3}, x_{2s+4}, x_2, x_3, \dots, x_{s+1}, x_{s+4}, x_{s+5}, \dots, x_{2s+3}) \\
&= (x_{2s+4} - x_{2s+3})((x_2 - x_1)\theta + (x_3 - x_2)\theta + \dots + (x_{s+2} - x_{s+1})\theta) \\
&\quad + (x_{2s+3} - x_{2s+2})((x_2 - x_1)\theta + (x_3 - x_2)\theta + \dots + (x_{s+2} - x_{s+1})\theta) \\
&\quad \vdots \\
&\quad + (x_{s+4} - x_{s+3})((x_2 - x_1)\theta + (x_3 - x_2)\theta + \dots + (x_{s+2} - x_{s+1})\theta) \\
&= (x_{2s+4} - x_{s+3})(x_{s+2} - x_1)\theta.
\end{aligned}$$

We are now in the position to go back to the definition of $Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$, $0 < t_1 < t_2 \dots < t_s < 1$ in (5.1.1), and write it in terms of double sums $S(\cdot)$ as in (5.3.14). Thus we obtain

$$Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} = \sum_{j=1}^s \sum_{i=j}^s S(t_{j-1}, t_j, t_i, t_{i+1}, x_2, x_3, \dots, x_{s+1},$$

$$x_{s+4}, x_{s+5}, \dots, x_{2s+3}),$$

$$0 < t_1 < \dots < t_s < 1, \quad (5.3.16)$$

where we define $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$. Since we do not know the location of the change-points $[n\lambda_1], [n\lambda_2], \dots, [n\lambda_s]$, where $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 1$, we define the following functions $a_i = a_i(\lambda_1, \lambda_2, \dots, \lambda_s, t_1, t_2, \dots, t_s)$, $1 \leq i \leq (s+1)s$, which will be used to derive a formula for (5.3.9) that will allow us to handle all possible combinations of $\lambda_1, \dots, \lambda_s$ and $t_1, \dots, t_s$:

$$\begin{aligned} a_1 &:= \begin{cases} \lambda_1, & \lambda_1 \leq t_1, \\ c_1, & t_1 < \lambda_1, \end{cases} \\ a_2 &:= \begin{cases} \lambda_2, & \lambda_1 \leq \lambda_2 \leq t_1, \\ c_2, & t_1 < \lambda_2, \end{cases} \\ &\vdots \\ a_s &:= \begin{cases} \lambda_s, & \lambda_1 \leq \dots \leq \lambda_s \leq t_1, \\ c_s, & t_1 < \lambda_s, \end{cases} \\ a_{s+1} &:= \begin{cases} \lambda_1, & t_1 < \lambda_1 \leq t_2, \\ \lambda_2, & \lambda_1 \leq t_1 < \lambda_2 \leq t_2, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-1} \leq t_1 < \lambda_s \leq t_2, \\ c_{s+1}, & \text{otherwise,} \end{cases} \\ a_{s+2} &:= \begin{cases} \lambda_2, & t_1 < \lambda_1 \leq \lambda_2 \leq t_2, \\ \lambda_3, & \lambda_1 \leq t_1 < \lambda_2 \leq \lambda_3 \leq t_2, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-2} \leq t_1 < \lambda_{s-1} \leq \lambda_s \leq t_2, \\ c_{s+2}, & \text{otherwise,} \end{cases} \\ &\vdots \\ a_{s+s} &:= \begin{cases} \lambda_s, & t_1 < \lambda_1 \leq \dots \leq \lambda_s \leq t_2, \\ c_{s+s}, & \text{otherwise,} \end{cases} \end{aligned}$$

$$\begin{aligned}
a_{2s+1} &:= \begin{cases} \lambda_1, & t_2 < \lambda_1 \leq t_3, \\ \lambda_2, & \lambda_1 \leq t_2 < \lambda_2 \leq t_3, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-1} \leq t_2 < \lambda_s \leq t_3, \\ c_{2s+1}, & \text{otherwise,} \end{cases} \\
&\vdots \\
a_{2s+s} &:= \begin{cases} \lambda_s, & t_2 < \lambda_1 \leq \dots \leq \lambda_s \leq t_3, \\ c_{2s+s}, & \text{otherwise,} \end{cases} \\
a_{s \cdot s+1} &:= \begin{cases} \lambda_1, & t_s < \lambda_1 < 1, \\ \lambda_2, & \lambda_1 \leq t_s < \lambda_2 < 1, \\ \vdots & \\ \lambda_s, & \lambda_1 \leq \dots \leq \lambda_{s-1} \leq t_s < \lambda_s < 1, \\ c_{s \cdot s+1}, & \text{otherwise,} \end{cases} \\
&\vdots \\
a_{s \cdot s+s} &:= \begin{cases} \lambda_s, & t_s < \lambda_1 \leq \dots \leq \lambda_s < 1, \\ c_{s \cdot s+s}, & \text{otherwise,} \end{cases}
\end{aligned} \tag{5.3.17}$$

where $0 < a_1 \leq a_2 \leq \dots \leq a_{s \cdot s+s} < 1$ and $c_i \in [0, 1]$, $1 \leq i \leq (s+1)s$.

We need to define these a_i 's, $1 \leq i \leq (s+1)s$, since there are exactly $(s+1)s$ possibilities to place s change-points in $(s+1)$ blocks. Moreover, exactly s of the latter a_i 's will change to one of the values λ_j , $1 \leq j \leq s$, the other s^2 a_i 's will get the value c_i and they will drop out of the limiting function $u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$.

The following theorem is an immediate consequence of the above arguments in this section.

Theorem 5.3.3 *Assume that (2.2.2), (5.1.20), (5.1.21), and $H_A^{(s)}$ hold. Define $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$.³ If $\tau_i = \tau_i(n) := [n\lambda_i]$, $i = 1, 2, \dots, s$, $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_s < 1$*

³Note that $t_{s+1} \rightarrow 1$ when $n \rightarrow \infty$. In (5.3.16) $t_{s+1} := \frac{n}{n+1}$, since $n = \left\lceil (n+1)\frac{n}{n+1} \right\rceil = [(n+1)t_{s+1}]$ has to be satisfied, but in the limiting function (5.3.18) $t_{s+1} = 1$.

1, then, as $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} \xrightarrow{P} u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s),$$

where

$$u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s) = \sum_{j=1}^s \sum_{i=j}^s \eta(t_{j-1}, t_j, t_i, t_{i+1}, a_{s(j-1)+1}, a_{s(j-1)+2}, \dots, a_{s(j-1)+s}, \\ a_{si+1}, a_{si+2}, \dots, a_{si+s}), \\ 0 < t_1 < t_2 < \dots < t_s < 1. \quad (5.3.18)$$

and η and the a_i 's, $1 \leq i \leq s(s+1)$, are defined in (5.3.15) and (5.3.17), respectively.

The limiting function in (5.3.18) is defined for every possible combination of $\lambda_1, \dots, \lambda_s$ and $t_1, \dots, t_s$. Moreover, there are $\binom{2s}{s}$ possibilities to choose $t_1, t_2, \dots, t_s$ given the postulated change-points $\lambda_1, \lambda_2, \dots, \lambda_s$, hence (5.3.18) has $\binom{2s}{s}$ different cases. This is, why the definition of the limiting function $u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$ involves so many variables. Otherwise we would have to state all different cases explicitly. For example, in case of 1 change-point we have 2 different cases, in case of 2 we have 6, in case of 3 we have 20, in case of 4 we have 70 and so on.

We note that in the limiting function in (5.3.18) exactly $\binom{s+2}{2}$ different $\theta_{i,j}$'s, $1 \leq i \leq j \leq s+1$, appear.⁴ This follows from the fact that there are exactly $\binom{s+2}{2}$ ways to compare 2 blocks (i.e. both before the first change, both between the first and the second change, ..., one before the first and the other between the first and the second change, ...), where none of these blocks contains any of the s changes. The problem is the same as choosing 2 balls (2 blocks) out of an urn with $(s+2)$ balls (2 blocks and s change-points).

Assuming that $\mathbb{E}h^2$ is finite instead of (5.1.21), then Theorem 5.3.3 implies the consistency of tests based on sup-functionals of $\{U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}, 0 < t_1 <$

⁴We note that some of these $\theta_{i,j}$'s can have the same value.

$t_2 < \dots < t_s < 1\}$. This means that we can consistently reject H_0 vs. $H_A^{(s)}$, when

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} \frac{1}{n^{3/2}\tilde{\sigma}} |U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}| \text{ is too large,}$$

except in the case when $\theta_{i,j} = 0$, $1 \leq i \leq j \leq s+1$.

The function $u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$ is equal to 0 if and only if all $\theta_{i,j}$'s involved are equal to 0. Otherwise, by using the fact that $\lambda_1, \lambda_2, \dots, \lambda_{s-1}$ and λ_s are fixed a priori for each n between 0 and 1 and do not depend on t_i , $1 \leq i \leq s$, one can show that there is at least one combination of $\lambda_1, \dots, \lambda_s$ and $t_1, \dots, t_s$ such that $u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s)$ is not equal to 0.

We observe that if $\theta_{i,j} = \theta$ for all possible choices of $1 \leq i \leq j \leq s+1$, then

$$\begin{aligned} u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t_1, t_2, \dots, t_s) &= \sum_{j=1}^s \sum_{i=j}^s (t_{i+1} - t_i)(t_j - t_{j-1})\theta \\ &= \sum_{j=1}^s (t_j - t_{j-1})(t_{j+1} - t_j + t_{j+2} - \\ &\quad - t_{j+1} + \dots + t_{s+1} - t_s)\theta \\ &= \sum_{j=1}^s (t_j - t_{j-1})(1 - t_j)\theta \\ &= \left(\sum_{j=1}^s (1 - t_j)t_j - \sum_{j=1}^s (1 - t_j)t_{j-1} \right)\theta \\ &= \left(\sum_{j=1}^s (1 - t_j)t_j - \sum_{j=1}^s (1 - t_{j+1})t_j \right)\theta \\ &= \sum_{j=1}^s (t_{j+1} - t_j)t_j\theta, \end{aligned}$$

which is the limit ($n \rightarrow \infty$) of the expected value of $\frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}$ (cf. (5.1.6)). Moreover, it follows that under the null hypothesis H_0 of no-change

$$\begin{aligned} \frac{1}{n^2} U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} &= \frac{1}{n^2} \left(Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} - \sum_{i=1}^s \left([(n+1)t_{i+1}] \right. \right. \\ &\quad \left. \left. - [(n+1)t_i] \right) [(n+1)t_i] \theta \right) \\ &= \frac{1}{n^2} Z_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]} - \frac{\sum_{i=1}^s \left([(n+1)t_{i+1}] \right)}{n^2} \end{aligned}$$

$$\begin{aligned}
& \frac{-[(n+1)t_i][(n+1)t_i]}{n^2} \theta \\
& \xrightarrow{P} \sum_{j=1}^s (t_{j+1} - t_j) t_j \theta - \sum_{j=1}^s (t_{j+1} - t_j) t_j \theta \\
& = 0.
\end{aligned}$$

Assuming that $\theta_{i,j} = \theta = 0$ for all possible choices of $1 \leq i \leq j \leq n$, the sequence

$$T_n = \frac{1}{n^{3/2} \bar{\sigma}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} |U_{[(n+1)t_1], [(n+1)t_2], \dots, [(n+1)t_s]}|, \quad n \in \mathbb{N}, \quad (5.3.19)$$

is not consistent against any class of alternatives. On the other hand, if at least one $\theta_{i,j}$ is not equal to 0 and we use T_n , then similarly to Example 4.3.1 we can show that

$$\mathbb{P}\{H_0 \text{ is rejected when using } T_n | H_A^{(s)} \text{ is true}\} \xrightarrow{n \rightarrow \infty} 1.$$

This implies that the limits of the sequence $\{T_n\}_{n \in \mathbb{N}}$ are different in probability under H_0 and $H_A^{(s)}$, and hence we have consistency of $\{T_n\}_{n \in \mathbb{N}}$.

5.4 Multiple Changes in the Mean

We are to test the *no-change in the mean* null-hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables
with $\mathbb{E}X_i = \mu$ and $0 < \sigma^2 = \text{Var}X_i < \infty$, $1 \leq i \leq n$,

against the *at most s changes in the mean* alternative

$H_A^{(s)} : X_1, \dots, X_n$ are independent random variables and there are s
integers $\tau_1, \tau_2, \dots, \tau_s$, $1 \leq \tau_1 \leq \tau_2 \leq \dots \leq \tau_s < n$, such that $\mathbb{E}X_1 =$
 $\dots = \mathbb{E}X_{\tau_1} \neq \mathbb{E}X_{\tau_1+1} = \dots = \mathbb{E}X_{\tau_2}$, $\mathbb{E}X_{\tau_2+1} = \dots = \mathbb{E}X_{\tau_3} \neq$
 $\mathbb{E}X_{\tau_3+1} = \dots = \mathbb{E}X_{\tau_s}$, $\dots$, $\mathbb{E}X_{\tau_{s-1}+1} = \dots = \mathbb{E}X_{\tau_s} \neq \mathbb{E}X_{\tau_s+1} =$
 $\dots = \mathbb{E}X_n$, and $0 < \sigma^2 = \text{Var}X_1 = \dots = \text{Var}X_n < \infty$.

Similarly, as in the previous sections, we note that testing for change-points in the mean can be illustrated via a geometrical argument, namely by comparing special areas with each other (cf. Section 3.4 and Section 4.4). Again we consider the linear function $m(t) := t$, $t \in \mathbf{R}$, which joins under H_0 all the points $(k, \frac{1}{\mu} \mathbb{E}\{S(k)\})$, $k \in \mathbf{N}$, if $\mu \neq 0$, and it joins all the points $(k, \mathbb{E}\{S(k)\})$, $k \in \mathbf{N}$, if $\mu = 0$.

Without loss of generality let $\mu = 1$. Then similarly to Section 4.4, we can draw the graph of the 45 degree line $m(t) = t$ which joins the points $(k, \mathbb{E}\{S(k)\})$, $k \in \mathbf{N}$. In a similar vein as in Figure 4.4.2, where we focused on at most two change-points, we can draw a more generalized graph, focusing on at most s change-points.

For each given $k_1, k_2, \dots, k_s$, $1 \leq k_1 < k_2 < \dots < k_s < n$, there correspond the s rectangles with endpoints $(k_{i-1}, \mathbb{E}\{S(0)\})$, $(k_{i-1}, \mathbb{E}\{S(k_i)\})$, $(k_{i+1}, \mathbb{E}\{S(k_i)\})$ and $(k_{i+1}, \mathbb{E}\{S(0)\})$, $1 \leq i \leq s$, respectively, where $k_0 := 0$ and $k_{s+1} := n$. Each of these rectangles with length $(k_{i+1} - k_{i-1})$ and height $\mathbb{E}\{S(k_i)\}$ has area $(k_{i+1} - k_{i-1}) \times \mathbb{E}\{S(k_i)\}$, $1 \leq i \leq s$, respectively, where $k_0 := 0$ and $k_{s+1} := n$.

By reflecting some parts of these rectangles⁵ around the 45 degree line $m = t$ we can construct a new rectangle with endpoints $(0, \mathbb{E}\{S(0)\})$, $(0, \mathbb{E}\{S(n)\})$, $(k_s, \mathbb{E}\{S(n)\})$ and $(k_s, \mathbb{E}\{S(0)\})$. This rectangle, with length k_s , height $\mathbb{E}\{S(n)\}$ and area $k_s \times \mathbb{E}\{S(n)\}$, has now, under H_0 , the same area as the sum of the previous areas. Consequently, for each given combination of $1 \leq k_1 < k_2 < \dots < k_s < n$,

$$T_{k_1, k_2, \dots, k_s}^{(n)} = \sum_{i=1}^s (k_{i+1} - k_{i-1}) S(k_i) - k_s S(n),$$

with $k_0 := 0$ and $k_{s+1} := n$, is an unbiased estimator of zero under H_0 . Moreover, $T_{k_1, k_2, \dots, k_s}^{(n)}$ is equivalent to $Z_{k_1, k_2, \dots, k_s}$ in (5.1.1) with $h(x, y) = x - y$. Since $T_{k_1, k_2, \dots, k_s}^{(n)}$ is a linear combination of partial sums and $t_i = \frac{k_i}{n}$, $1 \leq i \leq s$, Theorem 5.2.1 yields, as $n \rightarrow \infty$,

$$\begin{aligned} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \frac{1}{n^{3/2}\sigma} |T_{k_1, k_2, \dots, k_s}^{(n)}| &= \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \frac{1}{n^{1/2}\sigma} |T_{t_1, t_2, \dots, t_s}^{(n)}| \\ &\xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \left| \sum_{i=1}^s (t_{i+1} - t_{i-1}) W(t_i) - t_s W(1) \right|, \end{aligned}$$

⁵Note that these rectangles are overlapping.

with $t_0 := 0$ and $t_{s+1} := 1$, and where the limiting process is the same as $\Gamma^a(t_1, t_2, \dots, t_s)$ from (5.2.5).

5.5 Multiple Changes in the Variance

We are to test the *no-change in the variance* hypothesis

$H_0 : X_1, \dots, X_n$ are independent identically distributed random variables
with $\mathbb{E}X_i = \mu$ and $0 < \sigma^2 = \text{Var}X_i < \infty$, $1 \leq i \leq n$,

against the *at most s changes in the variance* alternative

$H_A^{(s)} : X_1, \dots, X_n$ are independent random variables and there exist s
integers $\tau_1, \tau_2, \dots, \tau_s$, $1 \leq \tau_1 \leq \tau_2 \leq \dots \leq \tau_s < n$, such that
 $\text{Var}X_1 = \dots = \text{Var}X_{\tau_1} \neq \text{Var}X_{\tau_1+1} = \dots = \text{Var}X_{\tau_2}$, $\text{Var}X_{\tau_1+1} =$
 $\dots = \text{Var}X_{\tau_2} \neq \text{Var}X_{\tau_2+1} = \dots = \text{Var}X_{\tau_3}$, $\dots$, $\text{Var}X_{\tau_{s-1}+1} =$
 $\dots = \text{Var}X_{\tau_s} \neq \text{Var}X_{\tau_s+1} = \dots = \text{Var}X_n$, $0 < \text{Var}X_1, \text{Var}X_{\tau_1+1},$
 $\text{Var}X_{\tau_2+1}, \dots, \text{Var}X_{\tau_s+1} < \infty$, and $\mathbb{E}X_1 = \dots = \mathbb{E}X_n = \mu$.

Again, as in Sections 3.5 and 4.5, we suggest the use of the *symmetric* kernel

$$h(x, y) = \frac{1}{2}(x - y)^2,$$

and, after some algebraic manipulations, $Z_{k_1, k_2, \dots, k_s}$ in (5.1.1) can be written as

$$\begin{aligned} Z_{k_1, k_2, \dots, k_s} &= \frac{1}{2} \left(\sum_{i=1}^s (k_{i+1} + k_{i-1} - 2k_i) R(k_i) + k_s R(n) \right. \\ &\quad \left. - 2 \sum_{i=1}^s (S(k_i) - S(k_{i-1}))(S(n) - S(k_i)) \right), \\ &\quad 1 \leq k_1 < k_2 < \dots < k_s < n, \end{aligned}$$

where $k_0 := 0$, $k_{s+1} := n$, $S(k) := \sum_{i=1}^k X_i$ and $R(k) := \sum_{i=1}^k X_i^2$. By Theorem 5.3.1, as $n \rightarrow \infty$, we have that under H_0

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} \frac{1}{n^{3/2} \tilde{\sigma}} \left| Z_{k_1, k_2, \dots, k_s} - \sigma^2 \sum_{i=1}^s (k_{i+1} - k_{i-1}) k_i \right|$$

$$\xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \left| \sum_{i=1}^s (t_{i+1} + t_{i-1} - 2t_i)W(t_i) + t_s W(1) \right|,$$

with $t_0 := 0$ and $t_{s+1} := 1$, and where the limiting process is the same as $\Gamma^{sym}(t_1, t_2, \dots, t_s)$ from (5.3.6).

5.6 Multiple Changes in the Mean and/or Variance

In the previous chapters, as well as in the previous sections of this chapter, we considered tests for changes in the mean or changes in the variance, separately only. Frequently, it is of interest to be able to test for changes in both the mean and the variance. It turns out that this is not an easy task in general. Based on U-statistics-type processes, here we propose a test statistic that will test for changes either in the mean or the variance or both. Unfortunately, in the following setup we are not able to distinguish between changes in both and changes in only one of them. Nevertheless, this test can be used when both depend on each other, namely the mean changes if and only if the variance changes (cf. Section 6.5). In case of independent normal variables Gombay and Horváth (1997) proposed an estimator for testing one single change in the mean and/or variance using the likelihood ratio test.

We are to test the *no-change in the mean and variance* hypothesis

$$H_0 : X_1, \dots, X_n \text{ are independent identically distributed random variables} \\ \text{with } \mathbb{E}X_i = \mu \text{ and } 0 < \sigma^2 = \text{Var}X_i < \infty, 1 \leq i \leq n,$$

against the *at most s changes in the mean and/or variance* alternative

$$H_A^{(s)} : X_1, \dots, X_n \text{ are independent random variables and there exist } s \\ \text{integers } \tau_1, \tau_2, \dots, \tau_s, 1 \leq \tau_1 \leq \tau_2 \leq \dots \leq \tau_s < n, \text{ such that} \\ \mathbb{E}X_1 = \dots = \mathbb{E}X_{\tau_1} \neq \mathbb{E}X_{\tau_1+1} = \dots = \mathbb{E}X_{\tau_2} \text{ and/or } \text{Var}X_1 = \\ \dots = \text{Var}X_{\tau_1} \neq \text{Var}X_{\tau_1+1} = \dots = \text{Var}X_{\tau_2}, \mathbb{E}X_{\tau_1+1} = \dots = \\ \mathbb{E}X_{\tau_2} \neq \mathbb{E}X_{\tau_2+1} = \dots = \mathbb{E}X_{\tau_3} \text{ and/or } \text{Var}X_{\tau_1+1} = \dots = \text{Var}X_{\tau_2} \neq$$

$$\begin{aligned} & VarX_{\tau_2+1} = \dots = VarX_{\tau_3}, \dots, \mathbb{E}X_{\tau_{s-1}+1} = \dots = \mathbb{E}X_{\tau_s} \neq \mathbb{E}X_{\tau_s+1} = \\ & \dots = \mathbb{E}X_{\tau_n} \text{ and/or } VarX_{\tau_{s-1}+1} = \dots = VarX_{\tau_s} \neq VarX_{\tau_s+1} = \\ & \dots = VarX_n, \text{ and } 0 < VarX_1, VarX_{\tau_1+1}, VarX_{\tau_2+1}, \dots, VarX_{\tau_s+1} < \\ & \infty. \end{aligned}$$

Since under H_0

$$\mathbb{E}X_i^2 = VarX_i + (\mathbb{E}X_i)^2, \quad 1 \leq i \leq n,$$

it is reasonable to consider symmetric kernels of the form

$$h(x, y) = \frac{x^2 + y^2}{2}.$$

Consequently, under H_0 , $h(X_i, X_j)$ is an unbiased estimator for $\theta = \mu^2 + \sigma^2$. It is obvious that changes in μ or σ^2 or in both will change θ . By using this kernel function, we can not distinguish which parameter changed. Nevertheless it may be used to detect, if there were any changes at all in any one, or in both of these parameters.

To apply our theory on U-statistic based processes, we assume that under H_0

$$\mathbb{E}h^2(X_i, X_j) = \frac{1}{2}Var(X_i^2) + (\mu^2 + \sigma^2)^2 < \infty, \quad (5.6.1)$$

and put

$$\begin{aligned} \tilde{h}(t) &:= \mathbb{E}\{h(t, X_2) - (\mu^2 + \sigma^2)\} \\ &= \frac{1}{2}(t^2 - (\mu^2 + \sigma^2)). \end{aligned}$$

Furthermore, we also assume that under H_0

$$\tilde{\sigma}^2 := \mathbb{E}\tilde{h}^2(X_1) = \frac{1}{4}Var(X_i^2)$$

is positive and finite. Hence, after some algebraic manipulations, $Z_{k_1, k_2, \dots, k_s}$ in (5.1.1)

with the kernel from above can be written as

$$Z_{k_1, k_2, \dots, k_s} = \sum_{i=1}^s (k_{i+1} + k_{i-1} - 2k_i) R(k_i) + k_s R(n),$$

$$1 \leq k_1 < k_2 < \dots < k_s < n,$$

where $k_0 := 0$, $k_{s+1} := n$ and $R(k) := \sum_{i=1}^k X_i^2$. By Theorem 5.3.1 we have that under H_0 , as $n \rightarrow \infty$,

$$\sup_{0 < t_1 < t_2 < \dots < t_s < 1} \frac{1}{n^{3/2} \tilde{\sigma}} \left| Z_{k_1, k_2, \dots, k_s} - (\mu^2 + \sigma^2) \sum_{i=1}^s (k_{i+1} - k_{i-1}) k_i \right|$$

$$\xrightarrow{\mathcal{D}} \sup_{0 < t_1 < t_2 < \dots < t_s < 1} \left| \sum_{i=1}^s (t_{i+1} + t_{i-1} - 2t_i) W(t_i) + t_s W(1) \right|,$$
(5.6.2)

with $t_0 := 0$ and $t_{s+1} := 1$, and where the limiting process is the same as $\Gamma^{sym}(t_1, t_2, \dots, t_s)$ from (5.3.6).

5.7 Estimating the Number of Change-points

When dealing with multiple change-points one of the questions that occurs is ‘how many (change-points) are there?’. This turns out to be not an easy task and there are not many papers written on this topic. Yao (1988) suggested an estimator for the number of changes in the mean when we have normal observations with common variance via a maximum likelihood argument and provided consistent way of estimating the true number of change-points. Serbinowska (1996) showed the consistency of the maximum likelihood estimator for the number of changes in case of binomial observations. Other related references about multiple change-points can be found in Csörgő and Horváth (1997) and Lee (1996).

Lee (1996) obtains a nonparametric estimator for the number of change-points and proved its consistency. Namely, we are to test the null hypothesis H_0 from Section 5.1 against the alternative

$$H_A^{(s)} : X_1, \dots, X_n \text{ are independent random variables and there exist } s_n$$

integers $\tau_1 = \tau_1(n), \tau_2 = \tau_2(n), \dots, \tau_s = \tau_{s_n}(n), 1 \leq \tau_1 < \tau_2 < \dots < \tau_{s_n} < n$, such that $\mathbb{P}\{X_1 \leq t\} = \dots = \mathbb{P}\{X_{\tau_1} \leq t\}, \mathbb{P}\{X_{\tau_1+1} \leq t\} = \dots = \mathbb{P}\{X_{\tau_2} \leq t\}, \dots, \mathbb{P}\{X_{\tau_{s_n}+1} \leq t\} = \dots = \mathbb{P}\{X_n \leq t\}$ for all t and $\mathbb{P}\{X_{\tau_i} \leq t_0\} \neq \mathbb{P}\{X_{\tau_i+1} \leq t_0\}$ for some t_0 and for all $1 \leq i \leq s_n$.

Using the idea of a window of observations (with length A_n), which is due to Lombard and Carroll (1992), Lee (1996) considered the difference D_j of two weighted empirical measures at each possible location j of change-point as follows:

$$D_j^{(n)}(x) = \frac{1}{\|c\|} \sum_{i=1}^{A_n} c_i \left(\mathbf{I}(X_{j-(i-1)} \leq x) - \mathbf{I}(X_{j-i} \leq x) \right), \quad A_n \leq j \leq n - A_n,$$

where $c_i, 1 \leq i \leq A_n$, is a sequence of positive numbers, $\|c\| = \sqrt{c_1^2 + \dots + c_{A_n}^2}$ and $x \in \mathbf{R}$. Assuming that the difference between two successive change-points is larger than $2A_n$, we can compute the expected value of $D_j^{(n)}(x)$ for each x . Note that for each j we can think of ‘comparing’ the block of random variables $(X_j, X_{j-1}, \dots, X_{j-A_n+2}, X_{j-A_n+1})$ to the block of random variables $(X_{j+1}, X_{j+2}, \dots, X_{j+A_n-1}, X_{j+A_n})$. Moreover, there is at most one change-point in the combined block $(X_{j-A_n+1}, \dots, X_{j+A_n})$, since by our assumption the difference between two successive change-points is larger than $2A_n$. Hence, if the compared random variables are both before or both after the change, then they have the same distribution. Therefore,

$$\begin{aligned} \Delta_j^{(n)}(x) &:= \mathbb{E} D_j^{(n)}(x) = \frac{1}{\|c\|} \sum_{i=1}^{A_n} c_i \left(\mathbb{E} \mathbf{I}(X_{j-(i-1)} \leq x) - \mathbb{E} \mathbf{I}(X_{j+i} \leq x) \right) \\ &= \frac{1}{\|c\|} \sum_{i=1}^{A_n} c_i \left(\mathbb{P}(X_{j-(i-1)} \leq x) - \mathbb{P}(X_{j+i} \leq x) \right) \\ &= \frac{1}{\|c\|} \left\{ \sum_{i=1}^{|\tau_r-j|} c_i \left(\mathbb{P}(X_{j-(i-1)} \leq x) - \mathbb{P}(X_{j+i} \leq x) \right) \right. \\ &\quad \left. + \sum_{i=|\tau_r-j|+1}^{A_n} c_i \left(\mathbb{P}(X_{j-(i-1)} \leq x) - \mathbb{P}(X_{j+i} \leq x) \right) \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\|c\|} \left\{ \sum_{i=1}^{|\tau_r-j|} c_i \left(\mathbb{P}(X_{\tau_r} \leq x) - \mathbb{P}(X_{\tau_r} \leq x) \right) \right. \\
&\quad \left. + \sum_{i=|\tau_r-j|+1}^{A_n} c_i \left(\mathbb{P}(X_{\tau_r} \leq x) - \mathbb{P}(X_{\tau_r+1} \leq x) \right) \right\} \\
&= \frac{1}{\|c\|} \sum_{i=|\tau_r-j|+1}^{A_n} c_i \left(\mathbb{P}(X_{\tau_r} \leq x) - \mathbb{P}(X_{\tau_r+1} \leq x) \right), \\
&\quad \text{for some } r \in 1, \dots, s_n, \text{ and } x \in \mathbf{R}.
\end{aligned}$$

Let

$$\delta_n(j) =: \begin{cases} \frac{1}{\|c\|} \sum_{i=|\tau_r-j|+1}^{A_n} c_i, & \text{if } |\tau_r - j| \leq A_n - 1 \text{ for some } \tau_r, \\ 0, & \text{otherwise.} \end{cases}$$

Then it is easy to see that $\delta_n(j)$ is increasing, if $j - A_n < j \leq \tau_r$, and decreasing, if $\tau_r \leq j \leq 2A_n$. Moreover, it takes its local maximum at the change-point τ_r . The same holds true for $\Delta_j^{(n)}$, if we assume, that $\sup_{x \in \mathbf{R}} (\mathbb{P}(X_{\tau_r} \leq x) - \mathbb{P}(X_{\tau_r+1} \leq x))$ is positive a.s., $1 \leq r \leq s_n$. Therefore, as an estimator for the true number of change-points, Lee (1996) suggests the estimator $\hat{s}_n$ with

$$\begin{aligned}
\hat{s}_n := \# \{ j : & \sup_{x \in \mathbf{R}} D_j^{(n)}(x) \geq C_n \text{ and } \sup_{x \in \mathbf{R}} D_j^{(n)}(x) > \sup_{x \in \mathbf{R}} D_m^{(n)}(x) \text{ for} \\
& j - A_n < m < j \text{ and } \sup_{x \in \mathbf{R}} D_j^{(n)}(x) \geq \sup_{x \in \mathbf{R}} D_m^{(n)}(x) \text{ for } j < m < j + A_n \},
\end{aligned} \tag{5.7.1}$$

where C_n satisfies $\frac{\sqrt{\log n}}{C_n} \xrightarrow{n \rightarrow \infty} 0$ and $\frac{C_n}{\sqrt{A_n}} \xrightarrow{n \rightarrow \infty} 0$. It is obvious that $\hat{s}_n = 0$ if $\sup_{x \in \mathbf{R}} D_j^{(n)}(x) < C_n$ for all possible j 's.

Under assumptions (A1) – (A3) of Lee (1996, page 297), Lee (1996) establishes that, as $n \rightarrow \infty$,

$$\hat{s}_n \stackrel{a.s.}{=} s_n.$$

We note that, for example, c_i , $1 \leq i \leq A_n$, may be chosen as $c_i = 1$ or $c_i = a + \frac{1}{i^b}$ for $a > 0$ and $b \geq \frac{1}{2}$.

Vostrikova (1981) suggested a binary segmentation method and also proved its

consistency. This binary segmentation procedure detects the number of change-points and their locations simultaneously. Namely, if we were to test for H_0 vs. $H_A^{(s)}$ (cf. Section 5.1), we first test H_0 , the null hypothesis of no-change, against H_A (cf. Section 3.1), the alternative of one single change in the distribution. If H_0 is not rejected, then we stop. On the other hand, if H_0 is rejected, then we know that at least one change-point is indicated and we test the two subsequences before and after the already located change-point for the possibility of having at most one further change-point in them. The procedure stops, when no further subsequences have change-points. The number of change-points found estimates the true number of change-points, s .

This procedure suggests that when looking for the true number of change-points, we should first test for one single change-point. On the other hand, this test statistic has to be consistent under the alternative of more than only one single change-point. But we saw in the beginning of Section 4.1 that, depending on the test statistics used, testing for more than one change can be inconsistent in special cases.

Recall notations from Section 5.1.2, especially

$$\begin{aligned}\lambda_0 &:= 0, \\ \lambda_{s+1} &:= 1, \\ \theta_i &:= \theta_{i,i}, \quad i = 1, \dots, s+1.\end{aligned}$$

Suppose we were to test H_0 vs. $H_A^{(s)}$ by using a properly normalized sup-functional of the stochastic process $Z_{[(n+1)t]}$, $0 < t < 1$, from (3.1.1) as a test statistics. After some algebraic manipulations it turns out that the limiting function of $\frac{1}{n^2} Z_{[(n+1)t]}$ under $H_A^{(s)}$ is of the form

$$u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t) = \theta_{j+1, j+1} \times t^2 + A_j \times t + B_j, \quad t \in [\lambda_j, \lambda_{j+1}) \text{ and } 0 \leq j \leq s,$$

where A_j and B_j are functions of λ_k , $1 \leq k \leq s+1$, and $\theta_{q,r}$, $1 \leq q \leq r \leq s+1$. We can see that $u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t) = 0$ for all $0 < t < 1$ if and only if $\theta_{j+1, j+1} = 0$, $A_j = 0$ and $B_j = 0$, where $0 \leq j \leq s$. Hence, any symmetric kernel will be consistent

when testing H_0 vs. $H_A^{(s)}$, but antisymmetric kernels may fail to be. This suggests to combine the procedure suggested by Vostrikova (1981) with the statistics studied in Section 3.3, namely $\sup_{0 < t < 1} |U_n(t)|$ as defined in in (3.1.15). Thus we aim at having a Vostrikova-type procedure in a non-parametric context. Define

$$\hat{i} := \min\{i : \lambda_i \leq t < \lambda_{i+1}, 0 < t < 1\},$$

and the following matrices that depend on the number of change-points s and on $\hat{i}$, which depends on t and the location of the change-points:

$$\mathbf{A}_{s,\hat{i}} := \left((1 - \lambda_s), (\lambda_s - \lambda_{s-1}), \dots, (\lambda_{\hat{i}+2} - \lambda_{\hat{i}+1}), (\lambda_{\hat{i}+1} - t) \right) \in \mathbf{R}^{1 \times (s+1-\hat{i})},$$

$$\mathbf{B}_{s,\hat{i}} := \begin{pmatrix} \theta_{\hat{i}+1,s+1} & \theta_{\hat{i},s+1} & \dots & \theta_{1,s+1} \\ \theta_{\hat{i}+1,s} & \theta_{\hat{i},s} & \dots & \theta_{1,s} \\ \vdots & \vdots & \dots & \vdots \\ \theta_{\hat{i}+1,\hat{i}+1} & \theta_{\hat{i},\hat{i}+1} & \dots & \theta_{1,\hat{i}+1} \end{pmatrix} \in \mathbf{R}^{(s+1-\hat{i}) \times (\hat{i}+1)},$$

$$\mathbf{C}_{\hat{i}} := \begin{pmatrix} t - \lambda_{\hat{i}} \\ \lambda_{\hat{i}} - \lambda_{\hat{i}-1} \\ \vdots \\ \lambda_2 - \lambda_1 \\ \lambda_1 - \lambda_0 \end{pmatrix} \in \mathbf{R}^{(\hat{i}+1) \times 1},$$

$$\mathbf{D}_s := \left((t - \lambda_s), (\lambda_s - \lambda_{s-1}), \dots, (\lambda_2 - \lambda_1), \lambda_1 \right) \in \mathbf{R}^{1 \times (s+1)}$$

$$\mathbf{E}_s := \begin{pmatrix} \theta_{s+1,s+1} \\ \theta_{s,s+1} \\ \vdots \\ \theta_{1,s+1} \end{pmatrix} \in \mathbf{R}^{(s+1) \times 1},$$

$$\mathbf{F} := \begin{pmatrix} 1 - t \end{pmatrix} \in \mathbf{R}^{1 \times 1}.$$

Obviously, the product of the first three matrices and the last three matrices is in $\mathbf{R}^{1 \times 1}$. The structure of the matrices is quite complicated. This is due to the fact that for fixed s we still have to consider where t is located. This, of course, will now change $\hat{i}$ which also changes the dimensions of the three matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. With similar arguments as in Section 5.2.2, namely using Theorem 2.6.2 and Hoeffding's SLLN (cf. Theorem 2.6.1) respectively, we arrive at the following theorem (cf. Theorem 5.3.3).

Theorem 5.7.1 *Assume that (2.2.2), (5.1.20), (5.1.21) and $H_A^{(s)}$ from Section 5.1 hold. Define $t_0 := 0$ and $t_{s+1} := \frac{n}{n+1}$. If $\tau_i = \tau_i(n) := [n\lambda_i]$, $i = 1, 2, \dots, s$, $0 < \lambda_1 < \lambda_2 < \dots < \lambda_s < 1$, then, as $n \rightarrow \infty$,*

$$\frac{1}{n^2} Z_{[(n+1)t]} \xrightarrow{P} u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t),$$

where

$$u_{\lambda_1, \lambda_2, \dots, \lambda_s}(t) := \begin{cases} \mathbf{A}_{s, \hat{i}} \times \mathbf{B}_{s, \hat{i}} \times \mathbf{C}_{\hat{i}}, & t \in (\lambda_{\hat{i}}, \lambda_{\hat{i}+1}], \text{ where } \hat{i} \in \{0, 1, \dots, s-1\}, \\ \mathbf{D}_s \times \mathbf{E}_s \times \mathbf{F}, & t \in (\lambda_s, 1), \end{cases}$$

and the matrices $\mathbf{A}_{s, \hat{i}}$, $\mathbf{B}_{s, \hat{i}}$, $\mathbf{C}_{\hat{i}}$, $\mathbf{D}_s$, $\mathbf{E}_s$ and $\mathbf{F}$ are defined above.

We observe that if $\theta_{i,j} = \theta$ for all possible choices of $1 \leq i \leq j \leq s+1$, then, as $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t]} \xrightarrow{P} \begin{cases} \left((1 - \lambda_s), \dots, (\lambda_{i+1} - t) \right) \begin{pmatrix} \theta & \dots & \theta \\ \vdots & & \vdots \\ \theta & \dots & \theta \end{pmatrix} \begin{pmatrix} t - \lambda_i \\ \lambda_i - \lambda_{i-1} \\ \vdots \\ \lambda_1 - \lambda_0 \end{pmatrix} \\ = t(1-t)\theta, \quad t \in (\lambda_i, \lambda_{i+1}], \\ \\ \left((1 - \lambda_s), (\lambda_s - \lambda_{s-1}), \dots, \lambda_1 \right) \begin{pmatrix} \theta \\ \vdots \\ \theta \end{pmatrix} (1 - t) \\ = t(1-t)\theta, \quad t \in (\lambda_s, 1), \end{cases}$$

which corresponds to the case where we don't have any changes as in (3.1.7) when we replace $\mathbb{E}Z_k$ by $\lim_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E}Z_{[(n+1)t]}$. Moreover, it follows that under the null hypothesis H_0 of no-change, as $n \rightarrow \infty$,

$$\begin{aligned} \frac{1}{n^2} U_{[(n+1)t]} &= \frac{1}{n^2} \left(Z_{[(n+1)t]} - [(n+1)t](n - [(n+1)t])\theta \right) \\ &= \frac{1}{n^2} Z_{[(n+1)t]} - \frac{[(n+1)t](n - [(n+1)t])}{n^2} \theta \\ &\xrightarrow{P} t(1-t)\theta - t(1-t)\theta \\ &= 0. \end{aligned}$$

Assuming that there are two changes in the distribution then the latter theorem implies the following. As $n \rightarrow \infty$,

$$\frac{1}{n^2} Z_{[(n+1)t]} \xrightarrow{P} \begin{cases} \begin{aligned} & \left((1 - \lambda_2), (\lambda_2 - \lambda_1), (\lambda_1 - t) \right) \begin{pmatrix} \theta_{1,3} \\ \theta_{1,2} \\ \theta_{1,1} \end{pmatrix} \begin{pmatrix} t - \lambda_0 \end{pmatrix} \\ & = t(1 - \lambda_2)\theta_{1,3} + t(\lambda_2 - \lambda_1)\theta_{1,2} + t(\lambda_1 - t)\theta_{1,1}, \quad t \in (0, \lambda_1], \end{aligned} \\ \\ \begin{aligned} & \left((1 - \lambda_2), (\lambda_2 - t) \right) \begin{pmatrix} \theta_{2,3} & \theta_{1,3} \\ \theta_{2,2} & \theta_{1,2} \end{pmatrix} \begin{pmatrix} t - \lambda_1 \\ \lambda_1 - \lambda_0 \end{pmatrix} \\ & = (1 - \lambda_2)((t - \lambda_1)\theta_{2,3} + \lambda_1\theta_{1,3}) + (\lambda_2 - t)((t - \lambda_1)\theta_{2,2} \\ & \quad + \lambda_1\theta_{1,2}), \quad t \in (\lambda_1, \lambda_2], \end{aligned} \\ \\ \begin{aligned} & \left((t - \lambda_2), (\lambda_2 - \lambda_1), (\lambda_1 - \lambda_0) \right) \begin{pmatrix} \theta_{3,3} \\ \theta_{2,3} \\ \theta_{1,3} \end{pmatrix} \begin{pmatrix} 1 - t \end{pmatrix} \\ & = (1 - t)((t - \lambda_2)\theta_{3,3} + (\lambda_2 - \lambda_1)\theta_{2,3} + \lambda_1\theta_{1,3}), \quad t \in (\lambda_2, 1). \end{aligned} \end{cases} \quad (5.7.2)$$

We note that this is the same limiting function as used in (4.1.1).

Theorem 3.3.2 implies that,

$$T_n = \frac{1}{n^{3/2}\bar{\sigma}} \sup_{0 < t < 1} |U_{[(n+1)t]}|, \quad n \in \mathbb{N}, \quad (5.7.3)$$

is consistent against any class of alternatives, if at least one $\theta_{i,j}$ is not equal to 0. Hence,

$$\mathbb{P}\{H_0 \text{ is rejected when using } T_n | H_A^{(s)} \text{ is true}\} \xrightarrow{n \rightarrow \infty} 1.$$

This implies that the limits of the sequence $\{T_n\}_{n \in \mathbb{N}}$ are different in probability under H_0 and $H_A^{(s)}$, and hence we have consistency of $\{T_n\}_{n \in \mathbb{N}}$.

Now we continue along the lines of Vostrikova (1981), and search for the argument, where the test statistic T_n takes its maximum. Unfortunately, the maximum is not always taken at one of the change-points, as we saw this in case of one single change-point in Section 3.3.3. Therefore we don't know, if the point where we split the intervals is a change-point or not. Hence it is possible that the number of possible change-points can be too big. Of course, if the change-points are in such a way that the maximum is always taken at one of them then the procedure works out fine.

Summarizing, we showed that T_n which was used in Section 3.3 in the context of testing for one single change-point is also consistent when testing for s , $1 \leq s < n$, change-points. Due to the fact that the local maximas are not necessarily taken at the change-points, one may, for example, prefer the use of the estimator $\hat{s}_n$ in (5.7.1) proposed by Lee (1996) to test for the true number of change-points.

“Imagination is more important
than knowledge.”

— *Albert Einstein*

Chapter 6

Applying Change-point Theory to the Financial Market

6.1 Introduction

In 1973, F. Black and M. Scholes derived a formula for option pricing, that has been called the Black-Scholes formula since. They worked closely together with R. Merton, and in 1997 Merton and Scholes were awarded the “Nobel Prize in Economical Sciences”. Black died in 1995 in his mid-fifties. Thousands of traders and investors now use this formula every day to value stock options in markets throughout the world. Black, Merton and Scholes thus contributed to the rapid growth of markets for derivatives in the last 10 years.

The derivation of the Black-Scholes formula involves many areas in Probability theory, for example, Martingale theory, Wiener processes, Itô processes, Stochastic integration and Stochastic differentiation. In the next sections we give a glimpse for some of the basic notions.

The Black-Scholes formula is used to compute the value of the so-called European options and other derivative securities. In their model the so-called volatility parameter σ is assumed to be a constant. It (the variance (volatility)) is frequently estimated via historical data. Using the results from the previous sections, we will propose a test procedure for testing for possible changes in the volatility in the Black-Scholes setup.

In practice, changes of the variance (volatility) are very important to know about, since these changes will affect the behavior of an investor. The higher the variance (volatility), the higher is the value of an option in the Black-Scholes model.

Using the methodologies from the previous chapters, we will aim at explaining how to detect changes in the variance (volatility). Also, we will see that changes in the mean of the stock price¹ do not affect the Black-Scholes formula, hence the value of an option.

For further reference and a more detailed description, we refer to J. Hull (1993), especially Chapters 1, 9 and 10. We will closely follow his presentation of the matter in hand. For a review of the Black-Scholes formula we also refer, for example, to Duffie (1996), and to Csörgő (1999), who details the original derivation of Black and Scholes (1973).

6.2 Derivative Securities

A derivative security is a variable depending on other more basic underlying variables. During the last few years, derivative securities have become more important in the financial market. Different possibilities of derivatives, for example, are forward and futures contracts or options. We will give an overview on these derivatives.

6.2.1 Forward Contracts

A forward contract is an agreement, for example between two financial institutions or a financial institution and one of its clients, to buy or sell an asset at a given time for a certain price, the so-called *delivery price*. It is not traded on an exchange and the parties usually know each other. One of the parties agrees to buy the underlying asset at a certain specified date and price, which is called the *long position*, and the other party agrees to sell it, and this is called the *short position*. The holder of the

¹We emphasize that we will denote throughout this chapter the stock price at time t by $S(t)$. This is the same notation as used widely in the finance literature. On the other hand, in mathematical literature $S(n)$ usually denotes the sum of n random variables. Although this was already used in previous chapters we will change our notation in this chapter.

short position delivers the asset to the holder of the long position in return for a cash amount, the delivery price. A forward contract is worth zero when it is first entered into. Depending on the movements in the price of the asset it can have a positive or negative value later on. The pay-off from a long position in a forward contract on one unit of an asset is

$$S(T) - K,$$

where K is the delivery price and $S(T)$ the spot price of the underlying asset at maturity time T of the contract. Of course, the pay-off from a short position is

$$K - S(T).$$

Both pay-offs can be positive or negative.

6.2.2 Futures Contracts

A futures contract is an agreement between two parties to buy or sell an asset at a given time for a certain price. It is traded on an exchange and the parties usually do not know each other. The underlying asset could be a commodity, e. g., sugar, lumber, gold or cows, or a financial asset, e. g., currencies, treasury bills or bonds. An exact delivery date is usually not specified, but there is a delivery month. The exchange specifies the period during the month when delivery must be made.

6.2.3 Options

There are two different types of options. The *call option* gives the holder the right, but not the obligation, to buy the underlying asset, e. g., stocks, foreign currencies, commodities, or futures contracts, at a given time $t = T$ for a certain price K , the so-called *exercise* or *strike price* or *maturity*. Similarly, a *put option* gives the holder the right to sell. There are different kinds of options.

Suppose, we sign at time $t = 0$ a contract which gives us the right (option) to

buy one share of a stock at a specified price K at a specified time T . Then we will exercise the option, this means realize the right to buy at the exercise price K , if the price of the stock $S(T)$ is greater than K , the price we agreed to pay for the stock. On the other hand, if $K \geq S(T)$, we do not exercise the option, since the price of the stock $S(T)$ is less than the specified price K . Of course, exactly the opposite is true, if we have the option to sell instead of the option to buy.

It is clear that an investor must pay a fixed amount of money to purchase an option contract. The Black-Scholes formula allows us to compute how much somebody should be willing to pay for an option contract. Moreover, there are two sides to every option contract, the investor who has taken the *long position* (i.e., who has bought the option) and the investor who has taken the *short position* (i.e., who has sold or written the option). The investor with the short position receives cash up front but may have liabilities later. The profit or loss of both investors is the reverse of each other.

For now, let us consider a call option in the long position for one single share of a stock. We recall that K is the delivery price. Then the pay-off when buying at a specified time T is

$$(S(T) - K)^+ = \max(S(T) - K, 0),$$

which is the so-called *European call option* (option exercised at maturity $t = T$). As mentioned before, there are different options and by allowing to buy (exercise the option) at any time from now ($t = 0$) to maturity ($t = T$) the pay-off is

$$\{(S(t) - K)^+, 0 \leq t \leq T\},$$

which is the so-called *American call option* (option exercised at any time from $t = 0$ to maturity $t = T$). If we allow to buy when the price of the stock is a maximum over a specified period then the pay-off is

$$\left\{ \left(\max_{t_0 \leq i \leq t} S(i) - K \right)^+, 0 \leq t_0 < t \leq T \right\},$$

which is the so-called *call on maximum option*, or *look back option*. Similarly, we may allow to buy when looking at the average of the stock price over a specified period of time. Then the pay-off is

$$\left\{ \left(\int_{t_0}^t S(x) dx - K \right)^+, 0 \leq t_0 < t \leq T \right\},$$

which is the so-called *Asian option*, or *call on average option*.

Some of these options are riskier than the others. Hence, each of them will have a different price. For example, an American option does not have the same risk as a European option, since we may exercise at any time from now to maturity. This will make the American option more expensive than the European. Of course, there are many other options and there is a lot of freedom about the characteristics of an option.

6.3 Modeling the Behavior of Stock Prices

Stock prices are usually assumed to follow a Markov process. This means that the probability distribution of the price of any particular future time depends only on the current stock price $S(t)$. Hence, the present price of a stock $S(t)$ impounds all the information contained in a record of past prices $S(t - t^*)$, $t^* < t$. If the Markov property does not hold, then one could make lots of profit by comparing charts of the past stock prices. Hence, it is a reasonable assumption.

While a Wiener process has mean or drift rate 0 and variance 1, we can define the so-called *generalized Wiener process* for a variable x which has drift μ and variance σ^2 , namely

$$dx = \mu dt + \sigma dW(t). \quad (6.3.1)$$

The μdt term means that x has an expected drift of μ per unit time, hence in a time interval of length T , x increases by an amount of μT . The σdW term is the so-called noise added to the path followed by x .

Since the parameters μ and σ may also depend on the variable x and the time t , we modify equation (6.3.1), and get a so-called *Itô process*

$$dx = \mu(x, t)dt + \sigma(x, t)dW(t). \quad (6.3.2)$$

Now we are in the position to define a process for the stock price $S(t)$. A first guess would be that $S(t)$ follows a generalized Wiener process with constant expected drift and constant variance. However, this process fails, since the expected percentage return required by investors from a stock is independent of $S(t)$. For example, an investor will require a 10% per annum expected return when $S(t)$ is \$10 as well as when it is \$20. Hence, we assume that the expected drift rate in $S(t)$ is $\mu S(t)$ for some constant parameter μ . Therefore,

$$S(t + \Delta t) - S(t) = \mu S(t)\Delta t$$

and, when compounded continuously

$$dS(t) = \mu S(t)dt,$$

so that

$$S(t) = S(0)e^{\mu t}, \quad 0 \leq t \leq T.$$

When considering a stock we also have to consider its volatility. We do this by assuming that the variance of the percentage return in a short period of time Δt is the same, independent of $S(t)$. Hence, we express $S(t)$ by an Itô process with instantaneous expected drift rate μ and instantaneous variance σ^2 , i. e., we assume

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad 0 \leq t \leq T. \quad (6.3.3)$$

We mention that $\mu \in \mathbf{R}$ is the constant *expected rate of return* and $\sigma > 0$ the so-called *stock price volatility*. μ depends on the level of interest rates in the economy.

The lower the level of interest rates, the lower the expected return required on any stock. Typical values for σ are in the range 0.2 to 0.4, i.e. 20% to 40%.

We even may simulate the movements in the stock price by using Monte Carlo Simulation. We give an example of one possible pattern of a stock price movement, since different random variables will produce another pattern. But by repeating this procedure many times, we may even estimate the distribution function $S(T)$ for a given T . Since

$$\frac{S(t + \Delta t) - S(t)}{S(t)} = \mu\Delta t + \sigma\epsilon\sqrt{\Delta t}$$

with $\epsilon \stackrel{\mathcal{D}}{=} N(0, 1)$ we have that

$$\frac{S(t + \Delta t) - S(t)}{S(t)} \stackrel{\mathcal{D}}{=} N(\mu\Delta t, \sigma^2\Delta t).$$

Hence, we may compute the price of the stock the following day, $S(t + \Delta t)$, as follows:

$$S(t + \Delta t) = S(t) + S(t)(\mu\Delta t + \sigma\epsilon\sqrt{\Delta t}),$$

where again $\epsilon \stackrel{\mathcal{D}}{=} N(0, 1)$.

We now consider $S(t)$ over a period of one year, $T = 1$, and we consider daily changes in $S(t)$, $\Delta t = \frac{1}{365}$. We assume to have an initial stock price $S(0)$ of \$20 and an expected return of 16% per annum, hence $\mu = 0.16$. Moreover, we draw a graph with a volatility of 20% per annum, $\sigma_1 = 0.2$, and 40% per annum, $\sigma_2 = 0.4$. For both patterns we will use the same generated random variables $\epsilon_1, \dots, \epsilon_{365}$ which are standard normal distributed. Again we mention that these are only two possible patterns, and they depend on the generated random variables $\epsilon_1, \dots, \epsilon_{365}$, but we can see the effect of different variances on the stock price.

In Figure 6.3.1 we show the patterns of the stock price for different volatilities and we see that the stock with a volatility of 40% per annum is much more fluctuating around the expected drift than the one with 20%. Hence, the higher the volatility

Figure 6.3.1: Daily stock prices with different volatilities $\sigma_1 = 0.2$ and $\sigma_2 = 0.4$.

the higher the risk for an investor. This is one of the motivations where change-point theory comes into the picture, for, as the volatility changes, the behavior of an investor might also change. This will also effect how to price an option which depends on the behavior of the stock $S(t)$ and hence also on its volatility.

For an owner of a stock, the movements in the stock tend to offset each other. But for an owner of an option this is different. For example, an owner of a call option (see Section 6.2.3), who has the right to buy an option at a specified time and price, benefits from price increase, but has limited downside risk if the premium decreases. Therefore the value of an option increases as the volatility increases.

6.4 The Black-Scholes Formula

Using the notions from Csörgő (1999, Section 4) in this section, we assume, as in (6.3.3), that the stock price process $\{S = S(t), t \geq 0\}$ is governed (driven) by a standard Wiener process $\{W = W(t), 0 \leq t < \infty\}$ on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ via the Itô process

$$dS(t) = \mu S(t)dt + \sigma S(t)dW(t), \quad 0 \leq t \leq T. \quad (6.4.1)$$

Assume further that the value or price C of a European call option at any time $t \in [0, T]$ depends only on the underlying stock price $S(t)$ and the time t , i.e., we have

$$\{C = C(t, S(t)), 0 \leq t \leq T\}, \quad (6.4.2)$$

and, in addition, that the real valued function $C = C(t, S(t))$ on $[0, T] \times (0, \infty)$ is continuously differentiable in t and twice continuously differentiable in S , denoted by $C \in C^{1,2}[[0, T] \times (0, \infty)]$. We note that the processes in (6.4.1) and (6.4.2) are driven by the same Wiener process. Itô showed that C is again an Itô process, and by Itô's chain rule formula

$$\partial C = \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S}dS + \frac{1}{2} \frac{\partial^2 C}{\partial S^2}(dS)^2, \quad (6.4.3)$$

where dS is given by (6.4.1). Hence we have

$$\begin{aligned} \partial C &= \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S}(\mu Sdt + \sigma SdW) + \frac{1}{2} \frac{\partial^2 C}{\partial S^2}(\mu Sdt + \sigma SdW)^2 \\ &= \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S}(\mu Sdt + \sigma SdW) + \frac{1}{2} \frac{\partial^2 C}{\partial S^2} \sigma^2 S^2 dt \\ &= \sigma S \frac{\partial C}{\partial S} dW + \left(\frac{\partial C}{\partial t} + \mu S \frac{\partial C}{\partial S} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \right) dt, \end{aligned} \quad (6.4.4)$$

where, for computing $(\mu Sdt + \sigma SdW)^2$, we used the following “multiplication table”

for differentials (cf. Karatzas and Shreve (1988, p. 154))

$$\begin{array}{c|cc} & dt & dW \\ \hline dt & 0 & 0 \\ dW & 0 & dt \end{array}.$$

The geometric Brownian motion process

$$S(t) = S(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)}, \quad 0 \leq t \leq T, \quad (6.4.5)$$

is a solution of (6.4.1), starting from $S(0)$ at time 0. Indeed, on letting $f(t, W) := S(t)$, then Itô's formula (cf. (6.4.3)) for the Itô process in (6.4.1) is

$$dS(t) = \frac{\partial f}{\partial t}dt + \frac{\partial f}{\partial W}dW + \frac{1}{2}\frac{\partial^2 f}{\partial W^2}(dW)^2 \quad (6.4.6)$$

By (6.4.5) we arrive at the differentials

$$\frac{\partial f}{\partial W} = \sigma S, \quad \frac{\partial^2 f}{\partial W^2} = \sigma \frac{\partial f}{\partial W} = \sigma^2 S, \quad \frac{\partial f}{\partial t} = \left(\mu - \frac{1}{2}\sigma^2\right)S,$$

which, in turn, via (6.4.6) yield (6.4.1) as desired, namely

$$\begin{aligned} dS(t) &= \sigma S(t)dW(t) + \left(\mu - \frac{1}{2}\sigma^2\right)S(t)dt + \frac{1}{2}\sigma^2 S(t) \underbrace{(dW(t))^2}_{dt} \\ &= \sigma S(t)dW(t) + \mu S(t)dt, \quad 0 \leq t \leq T. \end{aligned}$$

The uniqueness of the solution S in (6.4.5) for (6.4.1) follows from a general result of Itô, which states that a stochastic differential equation with Lipschitz continuous coefficients has a unique solution.

By (6.4.5), we have that for each $t \in [0, T]$

$$\begin{aligned} \log S(T) - \log S(t) &= \log \frac{S(T)}{S(t)} \\ &\stackrel{\mathcal{D}}{=} N\left(\left(\mu - \frac{\sigma^2}{2}\right)(T-t), \sigma^2(T-t)\right). \end{aligned} \quad (6.4.7)$$

Following Black and Scholes (1973), we now derive the Black-Scholes formula. Their assumptions are:

- The stock price $S(t)$ follows a geometric Brownian motion with constant μ and constant σ ,
- No transaction costs, no taxes, and no dividends involved,
- Short selling of securities is permitted,
- The trading is continuous,
- No arbitrage opportunities, which means no riskless profit by simultaneously entering into transaction in more than one market, e. g. to buy 1000 shares of a stock in New York for \$320 and sell it after considering the exchange rate for \$350 in Vienna,
- The risk-free interest rate r is constant and the same for all maturities.

Black and Scholes define a portfolio such that the holder of it is short one derivative security and long an amount of $\frac{\partial C}{\partial S}$ shares. Hence, the value of the portfolio, say P , is

$$P = -C + \frac{\partial C}{\partial S}S. \quad (6.4.8)$$

Using (6.4.1) and (6.4.4) we get that

$$\begin{aligned} dP &= -dC + \frac{\partial C}{\partial S}dS \\ &= -\left(\frac{\partial C}{\partial S}\mu S + \frac{\partial C}{\partial t} + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2\right)dt - \frac{\partial C}{\partial S}\sigma SdW + \\ &\quad + \frac{\partial C}{\partial S}\mu Sdt + \frac{\partial C}{\partial S}\sigma SdW \\ &= -\frac{\partial C}{\partial t}dt - \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2dt. \end{aligned}$$

Using the no-arbitrage condition, it follows that we have

$$dP = rPdt.$$

Consequently, we obtain

$$-\frac{\partial C}{\partial t}dt - \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2 dt = r\left(-C + \frac{\partial C}{\partial S}S\right)dt$$

which, in turn, implies the *Black-Scholes partial differential equation*

$$\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S}rS + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2 dt = rC, \quad (6.4.9)$$

where C is a function of S and t . It is very interesting to see that the expected return on the stock μ drops out of the equation and does not have any influence on pricing the option.

Note that this equation does not involve any variables that are affected by the risk preferences of investors. It would not be independent of risk preferences if it involved the expected return μ . In fact, the higher the level of risk aversion by investors, the higher μ will be for any given stock, hence μ depends on risk preferences (cf. Hull (1993, section 10.8)). Remarkably, under the Black-Scholes model, μ drops out of the equation.

Subject to $C(T, S) = (S - K)^+$, equation (6.4.9) can be solved, i. e., in case of a European call option as in Section 6.2.3 (cf. Black and Scholes (1973)), the price C of the option at any time $t = t_0 \in [0, T]$ is the *Black-Scholes formula* as follows,

$$\begin{aligned} C &= e^{-r(T-t_0)}\mathbb{E}\left(\max(S(T) - K, 0)\right) \\ &= S(t_0)\Phi(d) - Ke^{-r(T-t_0)}\Phi(d - \sigma\sqrt{T-t_0}), \end{aligned} \quad (6.4.10)$$

where

$$d = \frac{\ln\left(\frac{S(t_0)}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)(T - t_0)}{\sigma\sqrt{T - t_0}}.$$

The formula says that the option value C is higher the higher the price of the share today, $S(t_0)$, the higher the volatility of the share price, σ , the higher the risk-free interest rate, r , the longer the time to maturity, T , the higher the probability that the option will be exercised, and the lower the strike price, K . We note that Φ

denotes the standard normal distribution function.

6.5 Changes in the Volatility

In the previous section we stated the Black-Scholes formula that allows us to compute the value of a European call option assuming that $S(t)$ is a geometric Brownian motion with constants μ and σ^2 . Since most of the factors in this formula are fixed in advance, e. g., we know the price of the share today, $S(t)$, the time of maturity, T , the strike price, K , and more or less the risk free interest rate, r , we propose to test whether the volatility, σ , may have changed or not. Frequently, σ is estimated by historical data, for instance, via the stock prices of the last n days.

We will see that the problem of detecting changes in the volatility in stock prices is a very difficult task due to its complicated behavior. We will see that although under the assumptions of the Black-Scholes model we have a consistent estimator to test for changes in the volatility, but it will not be the case, if we change the model a bit. This will be clear when we exchange the assumption of geometric Brownian motion (cf. (6.4.5)) by geometric fractional Brownian motion (cf. (6.5.2)).

In Figure 6.5.2 we see the dependency of the price of a European call option on the volatility σ . As σ becomes big, the price of the option increases. Hence, when estimating the volatility using historical data, change in the volatility will effect its estimator. We fixed the spot price $S(0) = \$20$, the risk-free interest $r = 0.04$, $T = \frac{1}{2}$ which is equivalent to a period of half a year and then we can see graphs for the strike price $K = \$15$ and $K = \$25$, respectively. Of course, when the strike price is below the spot price then the option will be more expensive, since we buy at time $t = 0$ the right to buy in half a year a share for less money than the spot price $S(0)$. The price of the option goes down, if the spot price is below the strike price.

Let $S(t_1), \dots, S(t_{n+1})$ be the stock prices of the last $n + 1$ days, following a geometric Brownian motion under the Black-Scholes model (cf. (6.4.5), where μ is

Figure 6.5.2: Price of an European call option depending on σ for $K = \$15$ (above) and $K = \$25$ (below).

replaced by r), and let $\Delta t = t_i - t_{i-1}$, $1 \leq i \leq n+1$. Typically $t_i = \frac{i}{365}$. Thus,

$$\tilde{Z}_{i-1} := \log \frac{S(t_i)}{S(t_{i-1})} \stackrel{\mathcal{D}}{=} N\left((r - \frac{\sigma^2}{2})\Delta t, \sigma^2\Delta t\right), \quad 2 \leq i \leq n+1,$$

which is the continuously compounded return in the i -th interval and $S(t_i) = S(t_{i-1})e^{\tilde{Z}_{i-1}}$. We note that the $\tilde{Z}_{i-1}$'s are i.i.d. r.v.'s, normally distributed as indicated. Since we are interested in testing for changes in the volatility σ , we define

$$Z_i := \frac{\tilde{Z}_{i-1}}{\sqrt{\Delta t}} \stackrel{\mathcal{D}}{=} N\left((r - \frac{\sigma^2}{2})\sqrt{\Delta t}, \sigma^2\right), \quad 1 \leq i \leq n.$$

We note in passing that in this specific set-up the volatility σ in the Black-Scholes formula (cf. (6.4.10)) can be estimated via the maximum likelihood estimator (MLE).

Using the fact that, in this case, r and Δt are both known, we get

$$\sigma_{MLE}^2 = \frac{-1 + \sqrt{1 + \frac{\Delta t}{n} \sum_{i=1}^n (Z_i - r\sqrt{\Delta t})^2}}{\frac{\Delta t}{2}}, \quad (6.5.1)$$

which may replace the usually suggested sample variance for estimating σ in (6.4.10).

Suppose we are testing for two changes in the volatility (variance). From Section 4.5 we know that we may use a sup-functional of a properly normalized U-statistics based process with kernel function $h(x, y) = \frac{1}{2}(x - y)^2$ to test for changes in the variance.

Recall that for this test statistic we had to assume that under H_0 we have i.i.d. r.v.'s with finite non-zero variance σ^2 and under $H_A^{(2)}$ we have independent random variables with at most two changes in the variance, but no changes in the mean. Since r and Δt are constant, we observe that our random variables Z_i , $1 \leq i \leq n$, satisfy the assumptions made under H_0 , but violate the assumptions under $H_A^{(2)}$, since $\mathbb{E}Z_i$ depends on $\text{Var}Z_i = \sigma^2$, if $r \neq \frac{\sigma^2}{2}$.

On the other hand, we saw in Section 5.6 that by using the kernel $h(x, y) = \frac{x^2 + y^2}{2}$, we can test for changes in the mean and/or variance. In general, the test does not distinguish between changes in the mean or changes in the variance, but it will reject, when at least one of them changes. Fortunately, in our present model, r and Δt are constant, which means that $\mathbb{E}Z_i$ will change if and only if σ^2 changes. Thus, in this specific set-up, the statistics studied in Section 5.6 will reject H_0 when the mean *and* the variance change simultaneously. Namely, on estimating σ^2 via (6.5.1) and $\bar{\sigma}^2$ via a similar procedure, an appropriate test statistic can now be based on

$$\begin{aligned} \sup_{0 < t_1 < t_2 < 1} \frac{1}{n^{3/2}\bar{\sigma}} & \left| \sum_{i=1}^2 (k_{i+1} + k_{i-1} - 2k_i) R(k_i) + k_2 R(n) \right. \\ & \left. - \left\{ \left((r - \frac{\sigma^2}{2}) \sqrt{\Delta t} \right)^2 + \sigma^2 \right\} \sum_{i=1}^2 (k_{i+1} - k_{i-1}) k_i \right| \end{aligned}$$

with $k_0 := 0$, $k_3 := n$, $\bar{\sigma}^2 := \frac{1}{4}\text{Var}(Z_i^2)$, and $R(k) := \sum_{i=1}^k Z_i^2$, to test for at most two change-points in the volatility σ . Under H_0 such a test statistic converges in probability to the sup-functional of the Gaussian process in (4.3.10) and under $H_A^{(2)}$

it converges in probability to ∞ .

Assuming that we detected at most two changes - $H_A^{(2)}$ holds - then it is not clear how to estimate σ in the Black-Scholes formula (cf. (6.4.10)), since we don't know in general where the change-points are. Unfortunately, the maximum of the test statistic is not necessarily taken at the times of change.

In the previous setup, which was essential to test for changes in the variance, we assumed that the stock price $S(t)$ follows a geometric Brownian motion. That the test statistic is not applicable in every model will be clear, if we assume that the stock price, say $S_H(t)$, follows a geometric fractional Brownian motion instead of a geometric Brownian motion. Hence, instead of (6.4.5), we start with a geometric fractional Brownian motion of order $0 < H < 1$ of the form (cf. Csörgő (1999))

$$S_H(t) = S_H(0)e^{(r-H\sigma^2)t^{2H} + \sqrt{2H}\sigma W_H(t)}, \quad 0 \leq t \leq T, \quad (6.5.2)$$

which is a centered Gaussian process $\{W_H(t); 0 \leq t < \infty\}$ with stationary increments and $W_H(0) = 0$, i. e.,

$$W_H(t) - W_H(s) \stackrel{\mathcal{D}}{=} N(0, |t - s|^{2H}), \quad t, s \geq 0.$$

We note that H is the so-called *Hurst constant*. For further references to the fractional Brownian motion in the context of financial modeling we refer to Salopek (1997, Chapter 1) as well as to her references. Obviously for $H = \frac{1}{2}$ we have a geometric Brownian motion. Using now this geometric fractional Brownian motion then we may modify the Black-Scholes formula and calculate the price C of an European call option at $t_0 = 0$ as follows:

$$\begin{aligned} C &= e^{-rT^{2H}} \mathbb{E}(\max(S_H(T) - K, 0)) \\ &= S_H(0)\Phi(d_H) - Ke^{-rT^{2H}}\Phi(d_H - \sqrt{2H}\sigma T^H), \end{aligned}$$

where

$$d_H = \frac{\ln\left(\frac{S_H(0)}{K}\right) + (r + H\sigma^2)T^{2H}}{\sqrt{2H}\sigma T^H}.$$

Similarly, we define

$$\tilde{Z}_{i-1}^H := \log \frac{S_H(t_i)}{S_H(t_{i-1})} \stackrel{\mathcal{D}}{=} N\left((r - H\sigma^2)(t_i^{2H} - t_{i-1}^{2H}), 2H\Delta t\sigma^2\right), \quad 2 \leq i \leq n+1,$$

with $\Delta t = t_i - t_{i-1}$, $1 \leq i \leq n+1$. Hence,

$$Z_i^H := \frac{\tilde{Z}_{i-1}^H}{\sqrt{2H\Delta t}} \stackrel{\mathcal{D}}{=} N\left((r - H\sigma^2)\frac{t_i^{2H} - t_{i-1}^{2H}}{\sqrt{2H\Delta t}}, \sigma^2\right), \quad 1 \leq i \leq n,$$

has variance σ^2 , but the Z_i^H 's fail to be independent, if $H \neq \frac{1}{2}$. Note that under the null hypothesis of no-change $\mathbb{E}Z_i^H$ depends on $t_i^{2H} - t_{i-1}^{2H}$ and

$$\mathbb{E}Z_i^H = \mathbb{E}Z_{i+1}^H \iff t_i^{2H} - t_{i-1}^{2H} = t_{i+1}^{2H} - t_i^{2H} \iff H = \frac{1}{2}, \quad 1 \leq i \leq n-1.$$

Therefore the joint distribution of $Z_1^H, \dots, Z_n^H$ will be different than the product of the corresponding marginal distributions and our proposed test statistic fails, due to the dependence of the observations.

Summarizing, we can test for changes in the volatility when the stock prices are assumed to follow a geometric Brownian motion. This assumption allows us to produce i.i.d. r.v.'s by looking at the log-difference of the stock prices $S(t)$. We suggested a test statistic that will test for at most two changes in the volatility. Using the results from the previous sections, we may also test for more, or also less change-points.

Bibliography

- [1] Billingsley, P. (1986). Probability and Measure. 2nd ed. Wiley, New York.
- [2] Birnbaum, Z.W. and Pyke, R. (1958). *On Some Distributions Related to the Statistic D_n^+* . Ann. Math. Statist., **29**, 179-187.
- [3] Black, F. and Scholes, M. (1973). *The Pricing of Options and Corporate Liabilities*. Journal of Political Economy, **81**, 637-654.
- [4] Brodsky, B.E. and Darkhovsky, B.S. (1993). Nonparametric Methods in Change-Point Problems. Kluwer, Dordrecht.
- [5] Broemeling, L.D. and Tsurumi, H. (1987). Econometrics and Structural Change. Marcel Dekker. New York.
- [6] Casella, G. and Berger, R. (1990). Statistical Inference. Duxbury Press. Belmont.
- [7] Cramér, H. and Leadbetter, M.R. (1967). Stationary and Related Stochastic Processes; Sample Function Properties and their Applications. Wiley, New York.
- [8] Csörgő, M. (1979). *Brownian Motion - Wiener Process*. Canad. Math. Bull. Vol **22** (3), 257-279.
- [9] Csörgő, M. (1983). Quantile Processes with Statistical Applications. Society for Industrial and Applied Mathematics. Philadelphia, Pennsylvania.

- [10] Csörgő, M. (1999). *Random Walking Around Financial Mathematics*. Preprint.
- [11] Csörgő, M., Csörgő, S. and Horváth, L. (1986). *An Asymptotic Theory for Empirical Reliability and Concentration Processes*. Springer-Verlag. Berlin.
- [12] Csörgő, M. and Horváth, L. (1986). *Invariance Principles for Changepoint Problems*. Techn. Rep. Series of the Laboratory for Research in Statistics and Probability, Carleton U. - U. of Ottawa, No.80.
- [13] Csörgő, M. and Horváth, L. (1988a). *Nonparametric Methods for Changepoint Problems*. Handbook of Statistics, Vol. 7 403-425. Elsevier Science Publishers B.V. (Nort-Holland).
- [14] Csörgő, M. and Horváth, L. (1988b). *Invariance Principles for Changepoint Problems*. Journal of Multivariate Analysis 27, 151-168.
- [15] Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*. John Wiley. Chichester.
- [16] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-point Analysis*. John Wiley. Chichester.
- [17] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, New York.
- [18] Donsker, M. (1951). *An Invariance Principle for Certain Probability Limit Theorems*. Mem. Amer. Math. Soc., No. 6.
- [19] Doob, J.L. (1953). *Stochastic Processes*. John Wiley. New York.
- [20] Duffie, D. (1996). *Dynamic Asset Pricing Theory* (2nd edn). Princeton University Press. Princeton, New Jersey.
- [21] Eastwood, B.J. and Eastwood, V.R. (1998). *Tabulating Weighted sup-Norm Functionals of Brownian Bridges via Monte Carlo Simulation*. Asymptotic Methods in Probability and Statistics - a Volume in Honour of Miklós Csörgő (ed. B. Szyszkowicz), 707-719, Elsevier Science B.V., Amsterdam

- [22] Ferger, D. and Stute, W. (1992). *Convergence of Changepoint Estimators*. Stoch. Proc. Appl., **42**, 345-351.
- [23] Gombay, E. (1998). *U-Statistics for Change under Alternatives*. Preprint.
- [24] Gombay, E. and Horváth, L. (1997). *An Application of the Likelihood Method to Change-Point Detection*. Environmetrics, **8**, 459-467.
- [25] Gombay, E., Horváth, L. and Hušková, M. (1996). *Estimators and Tests for Change in the Variance*. Statistics and Decisions, **14**, 145-159.
- [26] Hall, P. (1979). *On the Invariance Principle for U-Statistics*. J. Stoch. Proc. Appl., **9**, 163-174.
- [27] Hoeffding, W. (1948). *A Class of Statistics with Asymptotically Normal Distribution*. Ann. Math. Stat., **19**, 293-325.
- [28] Hoeffding, W. (1961). *The Strong Law of Large Numbers for U-Statistics*. Univ. of North Carolina Institute of Statistics Mimeo Series, No. 302.
- [29] Hull, J. (1993). *Options, Futures, and Other Derivative Securities* (2nd edn). Prentice Hall International Editions. Englewood Cliffs, NJ.
- [30] Huse, V. (1988). *On Some Nonparametric Methods for Changepoint Problems*. PhD Thesis, Carleton University, Ottawa, Canada.
- [31] Janson, S. and Wichura, M.J. (1983). *Invariance Principles for Stochastic Area and Related Stochastic Integrals*. Stoch. Proc. and their Appl., **16**, 71-84.
- [32] Karatzas, I. and Shreve, S. (1988). *Brownian Motion and Stochastic Calculus*. Springer-Verlag. New York, N.Y.
- [33] Kendall, M.G. and Stuart, A. (1963). *The Advanced Theory of Statistics*. Volume 1. Distribution Theory. Charles Griffin & Company Limited, London.
- [34] Kolmogorov, A.N. (1933). *Sulla determinazione empirica di una legge di distribuzione*. Giorn. Inst. Ital. Attuari, **4**, 83-91.

- [35] Lee, C.-B. (1996). *Nonparametric Multiple Change-point Estimators*. Stat. and Prob. Letters, **27**, 295-304.
- [36] Levin, B. and Kline, J. (1985). *The Cusum Test of Homogeneity with an Application in Spontaneous Abortion Epidemiology*. Statist. Med., **4**, 469-488.
- [37] Lombard, F. (1987). *Rank Tests for Changepoint Problems*. Biometrika, **74**, 615-624.
- [38] Lombard, F. and Carroll, R.J. (1992). *Change Point Estimation via Running Cusums*. Technical Report No. 155, 1992, Statistics Department, Texas A&M University.
- [39] Loynes, R.M. (1970). *An Invariance Principle for Reversed Martingales*. Proc. Math. Soc., **25**, 56-64.
- [40] Major, P. (1979). *An Improvement of Strassen's Invariance Principle*. Ann. Probability, **7**, 55-61.
- [41] Miller, R.G., Jr. and Sen, P.K. (1972). *Weak Convergence of U-Statistics and Von Mises' Differentiable Statistical Functions*. Ann. Math. Stat., **43**, 31-41.
- [42] Resnick, S. (1992). *Adventures in Stochastic Processes*. Birkhäuser. Boston.
- [43] Salopek, D.M. (1997). *Tolerance to Arbitrage: Inclusion of Fractional Brownian Motion to Model Stock Price Fluctuations*. PhD Thesis, Carleton University, Ottawa, Canada.
- [44] Sen, P.K. (1977). *Almost Sure Convergence of Generalized U-Statistics*. Ann. Prob., **5**, 287-290.
- [45] Serbinowska, M. (1996). *Consistency of the Estimator of the Number of Change-points in Binomial Observations*. Statist. Probab. Letters, **29**, 337-344.
- [46] Serfling, R. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley. Chichester.

- [47] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- [48] Slutsky, E. (1925). *Über stochastische Asymptoten und Grenzwerte*. Math. Annalen, 5, 93.
- [49] Spanos, A. (1986). *Statistical Foundations of Econometric Modelling*. Cambridge University Press. Cambridge.
- [50] Szyszkowicz, B. (1991). *Weighted Stochastic Processes under Contiguous Alternatives*. C.R.Math.Rep.Acad.Sci. Canada 13, 211-216.
- [51] Szyszkowicz, B. (1992). *Weak Convergence of Stochastic Processes in Weighted Metrics and their Applications to Contiguous Change-point Analysis*. PhD Thesis, Carleton University, Ottawa, Canada.
- [52] Szyszkowicz, B. (1995). *Weighted Sequential Empirical Type Processes with Applications to Change-point Problems*. Techn. Rep. Series of the Laboratory for Research in Statistics and Probability, Carleton U. - U. of Ottawa, No.276.
- [53] Szyszkowicz, B. (1998). *Weighted Sequential Empirical Type Processes with Applications to Change-point Problems*. Handbook of Statistics, Vol. 16 (N. Balakrishnan and C.R. Rao, eds.), 573-630.
- [54] Vostrikova, L. Ju. (1981). *Detecting "Disorder" in Multidimensional Random Processes*. Soviet Mathematics Doklady, 24, 55-59.
- [55] Yao, Q. (1993). *Tests for Change-Points with Epidemic Alternatives*. Biometrika, 80, 179-191.
- [56] Yao, Y.-C. (1988). *Estimating the Number of Change-points via Schwarz's Criterion*. Statist. and Probab. Letters, 6, 181-189.