Section 4.1.	Asymptotic Distributions for Pontograms and Related Processes	
4.1.0.	Improved Results for Kendall-Kendall Pontograms	201
4.1.1.	The Process $M = \{M(x) : x \geq 0\}$	220
4.1.2.	A Result for Generalized Renewal Processes	223
Chapter 5.	Asymptotic Properties for the Sequential CUSUM Procedure	
Section 5.0.	Introduction and Notations	224
Section 5.1.	Results Under the Null Hypothesis	228
Section 5.2.	Results Under the Alternative Hypothesis	242
Appendix A.	Numerical Methods for Computing the Distribution of Quadratic Forms in Normal Random Variables	251
Appendix B.	Differential Equations Related to $\int_0^1 t(1-t)B^2(t)dt$	255
Appendix C.	Miscellaneous Technical Details	258
References		264

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

On Some Nonparametric Methods for Change-point Problems

by

Vera Regine Huse, Dipl. Math.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy

Department of Mathematics and Statistics

Carleton University

Ottawa, Ontario

June 30, 1988

© Copyright

1988, Vera Regine Huse

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-46251-5

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

"On Some Nonparametric Methods for Changepoint Problems"

Submitted by

VERA REGINE HUSE, Dipl. Math.

in partial fulfilment of the requirements
for the degree of Doctor of Philosophy.

Carol W L Bann
.....
Chairman, Department of
Mathematics and Statistics

Miklós Csörgő
.....
Thesis Supervisor

Stefan Zech
.....
External Examiner

Carleton University

Table of Contents

	Page
Acceptance Sheet	ii
Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	ix
List of Symbols	x

Abstract

Various nonparametric models for changepoint problems in the AMOC setting are considered. Sequential as well as fixed sample size (nonsequential) procedures are examined. The statistics for testing a general fixed sample size AMOC problem proposed in Chapter 2 are based on $U_k = \sum_{i=1}^k \sum_{j=k+1}^n h(X_i, X_j) - k(n-k)\theta$. The discussion includes symmetric as well as antisymmetric kernel functions h and relies on results taken from the theory of nondegenerate U-statistics. First U_k is expressed as a linear combination of nondegenerate one-sample U-statistics, then functionals of this linear combination are approximated by appropriate functionals of partial sums of independent identically distributed (under the null hypothesis) random variables. Finally, weak convergence results from the theory of convergence of partial sums are utilized to obtain the limiting distribution under the null hypothesis of no change. In Section 2.3, results under the alternative hypothesis are considered. Approximate upper percentage points for various Cramér-von Mises type statistics are tabulated in Chapter 3.

A different class of fixed sample size AMOC problems related to pontograms and to renewal theory is considered in Chapter 4, while Chapter 5 displays a new CUSUM type procedure for a sequential model.

Acknowledgements

First and foremost, I wish to express my extreme gratitude to Dr. Miklós Csörgő for his dedication, numerous discussions, guidance, and advice freely given during the time in which this thesis was written.

I also wish to thank my M.Sc. Thesis Advisor, Dr. Josef Steinebach of the University of Hannover, F.R.G., for having encouraged me to join Carleton University as a Ph.D. student and for having prepared me to face up to this challenge.

Concerning technical assistance in finalizing this thesis, I wish to acknowledge Paul Bertelmann for his efficient programming of those formulae which led to Tables 5-7 and Gill Murray for the outstanding job she accomplished in typing the manuscript.

Finally, I wish to acknowledge the financial support of the German Academic Exchange Service (DAAD) and the Paterson Centre for International Programmes. Without their support it would have been impossible for me to come to Canada. My special thanks go to Dr. David Farr and Margaret Sheck of the Paterson Centre for their continuing moral support far beyond the call of duty.

Table of Contents

	Page
Chapter 0. Introduction: A Brief History of the Changepoint Problem	1
Chapter 1. Preliminaries: Some Background Material	
Section 1.0. Stochastic Processes	6
Section 1.1. Komlós-Major-Tusnády Approximations	20
Section 1.2. Further Results from Probability Theory	26
Chapter 2. Nonparametric Test Statistics for the Fixed Sample Size Changepoint Problem	
Section 2.0. Introduction	30
Section 2.1. Wilcoxon-Mann-Whitney Type Tests	
2.1.0. Basic Concepts	33
2.1.1. Asymptotic Behaviour of $\hat{U}_k$ in sup-Norm and in Weighted sup-Norms Under the Null Hypothesis	40
2.1.2. Asymptotic Behaviour of U_k in sup-Norm and in Weighted sup-Norms Under the Null Hypothesis	57
2.1.3. Summary	61
Section 2.2. Cramér-von Mises Type Tests	
2.2.0. Auxiliary Results	65
2.2.1. Weak Convergence of $\hat{U}_k$ in L_2 -Norm and in Weighted L_2 -Norms Under the Null Hypothesis	71
2.2.2. Weak Convergence of U_k in L_2 -Norm and in Weighted L_2 -Norms Under the Null Hypothesis	76
2.2.3. A Central Limit Theorem for $\hat{U}_k$ and U_k	78
2.2.4. Summary and Discussion	90

Section 2.3.	Asymptotic Results Under the Alternative Hypothesis	
2.3.0.	Notations and Definitions	96
2.3.1.	Tests Based on Symmetric Kernels	104
2.3.2.	Tests Based on Antisymmetric Kernels	110
2.3.3.	Conclusions	113
Chapter 3.	Limiting Distributions of Certain Cramér-von Mises Type Test Statistics	
Section 3.0.	Introduction	115
Section 3.1.	Some Distribution Results for Functionals of Brownian Bridges and Related Gaussian Processes	
3.1.0.	Notations and Basic Theorems	116
3.1.1.	Distributions of L_2 -Norm Functionals of Brownian Bridges	126
3.1.2.	Distribution of Other Gaussian Processes in L_2 -Norm	135
Section 3.2.	Test Statistics with Degenerate Kernels. An Example	153
Chapter 4.	Kendall-Kendall Pontograms and Some Generalizations to Renewal Theory	
Section 4.0.	Introduction	
4.0.0.	The Kendall-Kendall τ - and τ -Pontograms	166
4.0.1.	Poisson and Generalized Renewal Processes	174
4.0.2.	Komlós-Major-Tusnády Type Approximations	184
4.0.3.	Conclusions	200

Section 4.1.	Asymptotic Distributions for Pontograms and Related Processes	
4.1.0.	Improved Results for Kendall-Kendall Pontograms	201
4.1.1.	The Process $M = \{M(x) : x \geq 0\}$	220
4.1.2.	A Result for Generalized Renewal Processes	223
Chapter 5.	Asymptotic Properties for the Sequential CUSUM Procedure	
Section 5.0.	Introduction and Notations	224
Section 5.1.	Results Under the Null Hypothesis	228
Section 5.2.	Results Under the Alternative Hypothesis	242
Appendix A.	Numerical Methods for Computing the Distribution of Quadratic Forms in Normal Random Variables	251
Appendix B.	Differential Equations Related to $\int_0^1 t(1-t)B^2(t)dt$	255
Appendix C.	Miscellaneous Technical Details	258
References		264

List of Tables

	Page
Table 1. Ranks and Signum of Ranks in a Sample	52
Table 2. Ranks and Signum of Ranks in a Sample	52
Table 3. Ranks and Signum of Ranks in a Sample	53
Table 4. Upper Percentage Points for Test T	112
Table 5. $(1-\alpha) \cdot 100\%$ - Fractiles of $\int_0^1 \sin \pi t B^2(t) dt$	134
Table 6. $(1-\alpha) \cdot 100\%$ - Fractiles of $\int_0^1 w^2(t) dt$	138
Table 7. $(1-\alpha) \cdot 100\%$ - Fractiles of $\int_0^1 \Gamma^2(t) dt$	149

List of Figures

	Page
Figure 1. Light Tail Weight Function	131
Figure 2. $\sin k\pi(1-t) \sin m\pi(1-t)$	151
Figure 3. ϵ -Blunt Triangles	168
Figure 4. Histogram for Maximum Angles in Land's End Trials (taken from D.G. and W.S. Kendall (1980))	170
Figure 5. Theoretical Density for the Maximum Angle on the Gaussian Model (taken from D.G. and W.S. Kendall (1980))	171
Figure 6. t -Pontogram for Land's End Data (taken from D.G. and W.S. Kendall (1980))	173
Figure 7. t -Pontogram for Land's End Data (taken from D.G. and W.S. Kendall (1980))	173
Figure 8. Waiting Times of an Ordinary Renewal Process	177
Figure 9. Increments of the Process M	196

List of Symbols

This list of symbols is presented in chronological order, as the various symbols appear in the text, except for some grouping of notations similar to one another.

$(\Omega, \mathcal{A}, P)$	probability space
$\mathbb{R}$	set of all real numbers
$\mathbb{R}^d, d \geq 1$	d-dimensional Euclidean space
$\mathcal{G}^d$	σ -algebra of the Borel sets in $\mathbb{R}^d, d \geq 1$
$[a, b]$	$\{x \in \mathbb{R} : a \leq x \leq b\}; a, b \in \mathbb{R}, a < b$
$[a, b)$	$\{x \in \mathbb{R} : a \leq x < b\}, -\infty < a < b \leq \infty$
$\mathbb{N}$	set of all positive integers: $\{1, 2, 3, 4, \dots\}$
$\stackrel{D}{=}$	equal in distribution
EX	expected value of the random variable X
$\text{Var } X$	variance of the random variable X
$W = \{W(t) : t \geq 0\}$	Wiener process
$N(a, b^2)$	normal random variable with mean a and variance b^2
$B = \{B(t) : 0 \leq t \leq 1\}$	Brownian bridge
$\mathbb{R}_+^2$	positive half plane: $\{(x, y) : 0 \leq x < \infty, 0 \leq y < \infty\}$
$:=, =:$	equal by definition, the expression that is to be defined, will appear on the dotted side
$I_A(X)$	indicator function: $I_A(X) = \begin{cases} 1 & \text{if } X \in A \\ 0 & \text{if } X \notin A \end{cases}$
$\#\{\dots\}$	number of elements in the set $\{\dots\}$
$\xrightarrow{\text{a.s.}}$	almost sure convergence
$\longrightarrow$	convergence of real numbers

$$B(x) := y_0^{-1/2} K(x, y_0)$$

is a Brownian bridge.

(c) $B_n = \{B_n(x) : 0 \leq x \leq 1\}$ defined pointwise by

$$B_n(x) := K(x, n) - K(x, n-1), \quad n \geq 1,$$

is a member of a sequence $(B_n)_{n \geq 1}$ of independent Brownian bridges.

(d) A Kiefer process K is a two-parameter Gaussian process with

$$EK(x, y) = 0, \quad EK(x_1, y_1)K(x_2, y_2) = (x_1 \wedge x_2 - x_1 x_2)(y_1 \wedge y_2).$$

The latter can also be taken as a definition of a Kiefer process.

(e) The trajectories of K are continuous with probability one.

(f) There are statements similar to the Doob forward and backward transformation connecting now a two-parameter Wiener process to a Kiefer process:

$$W(x, y) = (x+1)K\left(\frac{x}{x+1}, y\right), \quad x \geq 0, y \geq 0,$$

$$K(x, y) = (1-x)W\left(\frac{x}{1-x}, y\right), \quad 0 \leq x < 1, y \geq 0.$$

The Kiefer process will play an important role as the limiting process of CUSUM functionals considered in Chapter 5.

Next we begin an introduction into the area of empirical processes:

Let $X_1, X_2, \dots$ be independent identically distributed random variables on a common probability space $(\Omega, \mathcal{A}, P)$, and let F be their distribution. Then the empirical distribution function F_n of the random sample $X_1, \dots, X_n$ of size n is defined pointwise by:

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i), \quad x \in \mathbb{R},$$

Chapter 0. Introduction: A Brief History of the Changepoint Problem

The area of changepoint problems comprises nowadays a wide variety of different test procedures. Although changepoint problems started originally in the context of statistical quality control (confer also Chapter 2., Section 2.0.), they now have applications in as widely varying fields as econometrics (confer, e.g., J. Segen and A.C. Sanderson (1980)), archaeology and nuclear physics (confer Chapter 4., Subsection 4.0.0.). The different statistical procedures commonly put under the heading of changepoint problems are equally varying. They range from the testing of change in regression parameters (confer, e.g., I.B. MacNeill (1978)) and shift-in-mean problems (confer, e.g., L.A. Gardner (1969)) to nonparametric sequential CUSUM-type tests (confer Chapter 5., Section 5.1.). In the past 35 years, the topic of changepoint problems has been explored from many different viewpoints.

- Tests for a change in mean using a Bayesian approach (confer, e.g., H. Chernoff and S. Zacks (1964); Z. Kander and S. Zacks (1966); A. Sen and M.S. Srivastava (1975))
- nonparametric nonsequential tests for a change in distribution based on weak convergence results for certain functionals of stochastic processes (confer, e.g., M. Csörgö and L. Horváth (1988b))
- tests for change in intensity parameter of Poisson processes (confer, e.g., D.G. and W.S. Kendall (1980)) and for failure rates (confer, e.g., D.E. Matthews, V.T. Farewell and R. Pyke (1985)) as well as generalizations thereof (confer also Chapter 4) involving concepts from geometry and renewal theory

are only three examples utilizing completely different methodologies, from probability theory and statistics.

For a more complete picture of the present state of developments in the field of changepoint problems we refer to two review articles by D.A. Wolfe and E. Schechtman (1984) and M. Csörgö and L. Horváth (1988b) as well as to two bibliographies by S.A. Shaban (1980) and S. Zacks (1983).

In Chapter 1. background material for Chapters 2. - 5. of this thesis is collected. Often used less well known results are presented for easier reference.

Chapter 2. deals with some nonparametric fixed sample size approaches to the changepoint problem. Most of the test statistics considered in Chapter 2. are based on processes of U-statistics. The various subsections of this chapter are organized according to

- Wilcoxon-Mann-Whitney/Cramér-von Mises type tests
- symmetric / antisymmetric kernel functions for the U-statistics
- degenerate / nondegenerate kernel functions.

Concerning introductory articles into the areas of symmetric statistics and degenerate kernel functions we refer to H. Rubin and R.A. Vitale (1980) and G.G. Gregory (1977), respectively. The attempt to develop Cramér-von Mises type tests was inspired by an article by M. Csörgö and L. Horváth (1988a) who considered similar problems for Wilcoxon-Mann-Whitney type tests.

Chapter 3. contains various attempts at finding approximate upper percentage points for the statistics proposed in Chapter 2. First some

already known methods are discussed. Then an approach which seems to have been overlooked in the literature so far is suggested. This approach relies on the Paley-Wiener (1934) representation of Brownian motion and is here used to compute the tables displayed throughout Section 3.1. In Section 3.2. the difference in asymptotic behaviour between tests based on degenerate and on nondegenerate kernel functions is demonstrated by using an example.

Chapter 4. and 5. are independent of the previous chapters, self-contained and represent two additional, completely different aspects of changepoint problems. While in Chapter 3. techniques borrowed from functional analysis and from the theory of ordinary differential equations form the background material, Chapter 4. utilizes results stemming from renewal theory. In Chapter 4. various generalizations of the pontogram-concept of D.G. and W.S. Kendall (1980) are attempted. Sometimes the results are only partial because technical difficulties were encountered in trying to prove the statements for the upper tail-end of the time parameter. The direction in which further research is being conducted is also described in Chapter 4. Using the methodology developed in this chapter to generalize pontograms, similar generalizations can be executed for the tests devised for hazard rates by D.E. Matthews, V.T. Farewell and R. Pyke (1985).

In the final chapter, Chapter 5., a sequential test procedure proposed earlier by the author of this thesis (confer Vera R. Huse (1988)) is discussed in detail. Chapter 5. also contains a sequential almost sure version of a weak convergence theorem by P.K. Bhattacharya

and D. Frierson (1981) under the null hypothesis. Further research concerning a similar result under the alternative hypothesis is being conducted.

In the following a list of the author's own results is provided. Included, with an * in front of the result, are statements which were known but are reproved here because their original proofs either could not be found, were only outlined or contained mistakes. Also included with an * in the front are those results which are reproved in detail because their methodology of proof is essential for the attempted new generalizations. In all cases the proofs represent the author's contribution.

* Generalization (without number) of Theorem 1.1.20.

* Theorem 1.1.25.

* Corollary 4.0.6.

* Lemma 2.1.6.

Corollary 4.0.8.

* (2.1.9.)

Corollary 4.0.8.*

Lemma 2.2.0.

Theorem 4.1.0.

Lemma 2.2.1.

Corollary 4.1.2.

Corollary 2.2.2.

Lemma 4.1.6.

Theorem 2.2.3.

Theorem 4.1.7.

Corollary 2.2.4.

Theorem 4.1.9.

Corollary 2.2.5.

* Lemma 5.1.0.

Theorem 2.2.6.

Theorem 5.1.3.

Theorem 2.2.7.

Corollary 5.1.4.

Corollary 2.2.8.

Theorem 5.1.5.

Lemma 2.2.9. (similar to Theorem 3.1. in M. Csörgő and L. Horváth
(1988c))

Theorem 2.2.10. (similar to Theorem 2.3. in M. Csörgő and L. Horváth
(1988c))

Corollary 2.2.11. (similar to Corollary 2.1. in M. Csörgő and L. Horváth
(1988c))

Formulae (without number) for Table 5; Table 5

(3.1.12.) and Table 6 Corollary 5.1.6.

(3.1.21.) and Table 7 Theorem 5.2.0.

(3.1.22.) and Table 8 Theorem 5.2.2.

all results in Section 3.2. Corollary 4.0.10.

Corollary 4.1.4. (proof extracted from more general case in M. Csörgő
and L. Horváth (1987))

Chapter 1. Preliminaries: Some Background Material

Section 1.0. 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. In addition, some examples of one- and two-parameter stochastic processes will be given. These examples form the background for the limiting processes discussed in Chapters 2-5. For further references on concepts related to stochastic processes compare, e.g., S. Karlin and H.M. Taylor (1975).

Definition 1.1.0.

Let $(\Omega, \mathcal{A}, P)$ be a probability space. Denote by $\mathcal{B}^d$ the σ -algebra of the Borel sets in $\mathbb{R}^d$ ($d \geq 1$ fixed). Let T be an index-set. Then a family (or collection) $X = \{X(t) : t \in T\}$ of random variables $X(t)$ defined on the common probability space $(\Omega, \mathcal{A}, P)$ is called a stochastic process.

The index set T is said to be discrete if it consists of at most countably many elements. In this case the index set T may be regarded as containing discrete units of time. Otherwise the set T is said to be continuous. In the important special case that $T = [0, 1]$ or $[0, \infty)$, $t \in T$ is interpreted as time.

If the index set T is one-dimensional, we speak of a one-parameter stochastic process while if T is k -dimensional, X is said to be a k -parameter stochastic process.

For fixed $t \in T$, $X(t, \cdot)$ is a random variable (where the dotted argument is the ω -value, $\omega \in \Omega$) while for fixed $\omega \in \Omega$ $X(\cdot, \omega)$ (now the

dot represents the variable t) is a measurable function of t , called sample function, sample path or trajectory.

* If $X(t, \omega) \in \mathbb{R}^d$, then X is called a d -dimensional stochastic process. The set of possible values of $X(t)$, here $\mathbb{R}^d$, is called the state space.

Here and in the following only one-dimensional ($d=1$) stochastic processes will be considered. However, one- as well as two-parameter processes will be examined. In Chapters 2-5 we will also discuss sequences of stochastic processes. At this juncture it should be noted that

$$X = \{X_n(t) : t \geq 0, n \geq 1\} = \{X(t, n) : t \geq 0, n \geq 1\}$$

denotes one two-parameter stochastic process X while

$$X_n = \{X_n(t) : t \geq 0\}$$

stands for one member X_n of the sequence $(X_n)_{n \geq 1} = (\{X_n(t) : t \geq 0\})_{n \geq 1}$ of one-parameter stochastic processes X_n .

Definition 1.1.1.

Let $X = \{X(t) : t \in T\}$ be a stochastic process on $(\Omega, \mathcal{A}, P)$. 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 said to be a Gaussian process.

It should be noticed that a Gaussian process is uniquely determined by its mean value and covariance functions. This is true because the expected value and the covariance matrix specify uniquely any multivariate normal distribution.

Definition 1.1.2.

Let $X = \{X(t) : t \in I\}$ be a stochastic process, where $I = [0, 1]$ or $[0, \infty)$. For $s, t \in I$, the random variables $X(t+s) - X(t)$ are called increments of the process X . The stochastic process X is said to possess stationary increments if the distribution of $X(t+s) - X(t)$ depends only on the length s of the time interval under consideration and not on the time t .

$$(1.1.0.) \quad X(t+s) - X(t) \stackrel{D}{=} X(s) - X(0) \quad \text{for all } s, t \in I.$$

A stochastic process X which satisfies (1.1.0.) is often also said to have strictly stationary increments. If the random variables

$X_{t_2} - X_{t_1}, X_{t_3} - X_{t_2}, \dots, X_{t_n} - X_{t_{n-1}}$ are mutually independent for all possible choices of $t_1, t_2, \dots, t_n \in I$, $n \in \mathbb{N}$, satisfying $0 \leq t_1 < t_2 < t_3 < \dots < t_n$, then the process X is said to possess independent increments.

Remark 1.1.3.

If $X = \{X(t) : t \geq 0\}$ is a stochastic process with stationary independent increments satisfying $EX(t) < \infty$ for all $t \geq 0$, then it can be shown (confer, e.g., S. Karlin and H.M. Taylor (1975), p.27-28), that

$$EX(t) = m_0 + m_1 t, \quad \text{where } m_0 = EX(0) \text{ and } m_1 = EX(1) - m_0.$$

Similarly, if $\text{Var } X(t) < \infty$ for every $t \geq 0$, then

$$\text{Var } X(t) = \sigma_0^2 + \sigma_1^2 t, \quad \text{where } \sigma_0^2 = \text{Var } X(0) \text{ and } \sigma_1^2 = \text{Var } X(1) - \sigma_0^2.$$

As a first example of a one-parameter one-dimensional Gaussian process with stationary independent increments we recall the classical example

of Brownian motion. In probability theory Brownian motion is often called a Wiener process.

Definition 1.1.4.

A stochastic process $W = \{W(t) : t \geq 0\}$ is said to be a Wiener process if

- (i) $W(0) = 0$
- (ii) $W(t) - W(s) = N(0, t-s)$ for all $0 \leq s < t < \infty$,
where $N(0, t-s)$ denotes a normal random variable with mean zero and variance $(t-s)$,
- (iii) W is an independent increment process,
- (iv) the sample paths $W(\cdot, \omega)$ are continuous with probability one, i.e. $W(t, \omega)$ as a function of t only is continuous for almost every $\omega \in \Omega$.

Note that (ii) and (iii) in Definition 1.1.4. imply that the covariance structure of a Wiener process is given by

$$\text{Cov}(W(s), W(t)) = EW(s)W(t) = s \wedge t,$$

where $\wedge$ denotes the minimum.

For a constructive proof of the existence of a Wiener process we refer to M. Csörgő and P. Révész (1981), Chapter 1. A completely different approach to the proof of the existence of the Wiener process, using purely measuretheoretic arguments, is given in P. Billingsley (1968), Chapter 2.

As a second example of a Gaussian process we introduce the so-called Brownian bridge (or tied down Brownian motion).

Definition 1.1.5.

A stochastic process $B = \{B(t) : 0 \leq t \leq 1\}$ is called a Brownian bridge if

- (i) $EB(t) = 0$ for every $t \in [0, 1]$,
- (ii) the joint distribution of $B(t_1), B(t_2), \dots, B(t_n)$ is Gaussian for all possible choices of $0 \leq t_1 < t_2 < \dots < t_n \leq 1$, $n \in \mathbb{N}$,
- (iii) the covariance function R of B is given by

$$R(s, t) = EB(s)B(t) = s \wedge t - st \text{ for } 0 \leq s, t \leq 1,$$
 where $\wedge$ again denotes the minimum,
- (iv) the trajectories $B(\cdot, \omega)$ are continuous in t with probability one.

Remark 1.1.6.

Regarding the connection between a Wiener process and a Brownian bridge we have the famous forward and backward transformation of Doob (confer, e.g., Lemma 1.4.1. and Proposition 1.4.2. in M. Csörgő and P. Révész (1981)):

Let $W = \{W(t) : t \geq 0\}$ be a Wiener process. Then

$$(1.1.1.) \quad B(t) := W(t) - tW(1), \quad 0 \leq t \leq 1,$$

defines a Brownian bridge $B = \{B(t) : 0 \leq t \leq 1\}$.

Conversely, let $B = \{B(t) : 0 \leq t \leq 1\}$ be a Brownian bridge, then

$$(1.1.2.) \quad W(t) := (t+1)B\left(\frac{t}{t+1}\right), \quad t \geq 0,$$

defines a Wiener process $W = \{W(t) : t \geq 0\}$.

Another extremely useful fact concerning the Wiener process and the Brownian bridge is their classical Paley-Wiener (1934) representation.

Lemma 1.1.7. (Confer (1.8.4.) and (1.8.5.) in M. Csörgő and P. Révész (1981))

Let (Ω, A, P) be a given probability space with a Wiener process $W = \{W(t) : t \geq 0\}$ and a Brownian bridge $B = \{B(t) : 0 \leq t \leq 1\}$ on it. Then we have the following representations:

$$(1.1.3.) \quad W(t) = Y_0 t + \sqrt{2} \sum_{k=1}^{\infty} Y_k \frac{\sin k\pi t}{k\pi}, \quad t \geq 0, \text{ and}$$

$$(1.1.4.) \quad B(t) = \sqrt{2} \sum_{k=1}^{\infty} Y_k \frac{\sin k\pi t}{k\pi}, \quad 0 \leq t \leq 1,$$

where $(Y_k)_{k \geq 0}$ is a sequence of independent standard normal random variables defined on (Ω, A, P) . The convergence of the infinite series in (1.1.3.) and (1.1.4.) is absolute and uniform in t .

Lemma 1.1.7. will be used in Section 3.1. in order to find limiting distributions for various functionals of Wiener processes and Brownian bridges.

Before turning to the example of the Ornstein-Uhlenbeck process, another property of stochastic processes has to be discussed.

Definition 1.1.8.

A stationary process is a stochastic process $X = \{X(t) : t \in T\}$ with the property that 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 every $h \in T$.

Definition 1.1.8. says that a stationary process is a stochastic process whose finite dimensional distribution laws remain unchanged

through shifts in time. According to S. Karlin and H.M. Taylor (1975), p.443, this "concept captures the very natural notion of a physical system that lacks an inherent time" origin. It follows immediately from Definition 1.1.8. that if $EX(t) = m(t) < \infty$ for all $t \in T$ and if X is a stationary process, then $m(t) = m$ (a constant, independent of the time t) for all $t \in T$. Similarly, if X is a stationary process with two finite moments for each t , then $\text{Var } X(t) = \sigma^2(t) = \sigma^2$ (independent of t) for all $t \in T$. One can also show that the covariance function R of a stationary process X is a function of $|s-t|$ only, namely $R(s,t) := E(X(s)-m)(X(t)-m) = E(X(s-t)-m)(X(0)-m)$.

The last one-parameter Gaussian process to be introduced here is the Ornstein-Uhlenbeck process. Consider the following transformation of a Brownian motion W :

$$V(t) := W(t)/\sqrt{t}, \quad 0 < t < \infty, \quad V(0) := 0.$$

Then $V = \{V(t) : t \geq 0\}$ can be seen to be a Gaussian process with $EV(t) = 0$, $\text{Var } V(t) = 1$ and $\text{Cov}(V(s), V(t)) = EV(s)V(t) = \sqrt{s/t}$ for $0 \leq s < t < \infty$. Hence the process V is an example of a nonstationary Gaussian process.

In order to turn V into a stationary process, consider the transformation

$$(1.1.5.) \quad U_{\alpha}(t) := V(e^{\alpha t}), \quad t \in \mathbb{R}, \quad \text{for } \alpha > 0 \text{ fixed.}$$

The process $U_{\alpha} = \{U_{\alpha}(t) : t \in \mathbb{R}\}$ is a stationary Gaussian process with covariance function given by $EU_{\alpha}(s)U_{\alpha}(t) = e^{-\alpha|s-t|/2}$ for $s, t \in \mathbb{R}$.

Definition 1.1.9.

The stochastic process $U_\alpha = \{U_\alpha(t) : t \in \mathbb{R}\}$ defined via (1.1.5.) is called an Ornstein-Uhlenbeck process with parameter $\alpha > 0$. If $\alpha = 2$, $U_\alpha = U_2$ is often abbreviated by U .

Next the notion of a two-parameter Wiener process is studied. Let $\underline{z} = (x, y) \in \mathbb{R}_+^2$ be a point in the positive half plane. Let $X = \{X(\underline{z}) : \underline{z} \in \mathbb{R}_+^2\}$ be a two-parameter stochastic process. Consider the rectangle $R = [x_1, x_2) \times [y_1, y_2) \subseteq \mathbb{R}_+^2$ in the positive half plane, i.e. R is a rectangle with vertices (x_1, y_1) , (x_2, y_1) , (x_2, y_2) and (x_1, y_2) , where $0 \leq x_1 < x_2 < \infty$ and $0 \leq y_1 < y_2 < \infty$. Then define the "X-measure" $X(R)$ of R as the quantity that is generated by "differencing" X around the four vertices of R as follows:

$$X(R) := X(x_2, y_2) - X(x_1, y_2) - X(x_2, y_1) + X(x_1, y_1).$$

Definition 1.1.10.

A stochastic process $W = \{W(\underline{z}) : \underline{z} \in \mathbb{R}_+^2\}$ is called a two-parameter Wiener process if

- (i) $W(R) \stackrel{D}{=} N(0, \lambda(R))$ for all $R = [x_1, x_2) \times [y_1, y_2) \subseteq \mathbb{R}_+^2$,
where $\lambda(R) := (x_2 - x_1)(y_2 - y_1)$,
- (ii) $W(0, y) = W(x, 0) = 0$ for all $0 \leq x, y < \infty$,
- (iii) W is an independent increment process, i.e. the random variables $W(R_1), W(R_2), \dots, W(R_n)$, $n = 2, 3, \dots$, are independent provided that $R_1, R_2, \dots, R_n$ are disjoint (nonoverlapping) rectangles in $\mathbb{R}_+^2$,
- (iv) the sample function $W(\cdot, \omega)$ is continuous in $\underline{z}$ with probability one.

Remarks 1.1.11.

(a) (i)-(iii) in Definition 1.1.10. imply that the covariance function R of a two-time parameter Wiener process is given by

$$R(\underline{z}_1, \underline{z}_2) = EW(\underline{z}_1)W(\underline{z}_2) = (x_1 \wedge x_2)(y_1 \wedge y_2),$$

where $\underline{z}_1 := (x_1, y_1)$ and $\underline{z}_2 := (x_2, y_2)$.

(b) For $0 < x_0 < \infty$ fixed, the process $\{x_0^{-1/2}W(x_0, y) : y \geq 0\}$ is a one-parameter Wiener process and similarly for fixed $0 < y_0 < \infty$, the process $\{y_0^{-1/2}W(x, y_0) : x \geq 0\}$ is again a one-parameter Wiener process.

(c) A constructive proof of the existence of a two-parameter Wiener process is given in M. Csörgő and P. Révész (1981), pp. 58-59.

Having defined a two-parameter Wiener process, we are now in a position to introduce the last example of this section.

Definition 1.1.12.

Let $W = \{W(x, y) : x, y \geq 0\}$ be a two-parameter Wiener process. A Kiefer process $K = \{K(x, y) : 0 \leq x \leq 1, 0 \leq y < \infty\}$ is defined by

$$(1.1.6.) \quad K(x, y) := W(x, y) - xW(1, y).$$

Definition 1.1.12. implies the following elementary properties of a Kiefer process (confer also M. Csörgő and P. Révész (1981), p.80).

(a) For any $0 < x_0 < 1$ fixed, $W = \{W(y) : y \geq 0\}$ defined pointwise via

$$W(y) := (x_0(1-x_0))^{-1/2}K(x_0, y)$$

is a one-parameter Wiener process.

(b) For any $0 < y_0 < \infty$ fixed, $B = \{B(x) : 0 \leq x \leq 1\}$ defined pointwise via

$$B(x) := y_0^{-1} K(x, y_0)$$

is a Brownian bridge.

(c) $B_n = \{B_n(x) : 0 \leq x \leq 1\}$ defined pointwise by

$$B_n(x) := K(x, n) - K(x, n-1), \quad n \geq 1,$$

is a member of a sequence $(B_n)_{n \geq 1}$ of independent Brownian bridges.

(d) A Kiefer process K is a two-parameter Gaussian process with

$$EK(x, y) = 0, \quad EK(x_1, y_1)K(x_2, y_2) = (x_1 \wedge x_2 - x_1 x_2)(y_1 \wedge y_2).$$

The latter can also be taken as a definition of a Kiefer process.

(e) The trajectories of K are continuous with probability one.

(f) There are statements similar to the Doob forward and backward transformation connecting now a two-parameter Wiener process to a Kiefer process:

$$W(x, y) = (x+1)K\left(\frac{x}{x+1}, y\right), \quad x \geq 0, y \geq 0,$$

$$K(x, y) = (1-x)W\left(\frac{x}{1-x}, y\right), \quad 0 \leq x < 1, y \geq 0.$$

The Kiefer process will play an important role as the limiting process of CUSUM functionals considered in Chapter 5.

Next we begin an introduction into the area of empirical processes:

Let $X_1, X_2, \dots$ be independent identically distributed random variables on a common probability space $(\Omega, \mathcal{A}, P)$, and let F be their distribution. Then the empirical distribution function F_n of the random sample $X_1, \dots, X_n$ of size n is defined pointwise by:

$$F_n(x) := \frac{1}{n} \sum_{i=1}^n I_{(-\infty, x]}(X_i), \quad x \in \mathbb{R},$$

where I denotes the indicator function.

There are two more, often very useful, equivalent ways to define the empirical distribution function:

$$F_n(x) = \frac{1}{n} \# \{1 \leq i \leq n : X_i \leq x\}, \quad x \in \mathbb{R},$$

and in terms of order statistics

$$F_n(x) = \begin{cases} 0 & X_{1,n} > x \\ k/n & X_{k,n} \leq x < X_{k+1,n}, \quad k=1, 2, \dots, n-1 \\ 1 & X_{n,n} \leq x, \end{cases}$$

where $X_{1,n} \leq X_{2,n} \leq \dots \leq X_{n,n}$ denote the n order statistics of the random sample $X_1, \dots, X_n$.

Lemma 1.1.13.

For every fixed $x \in \mathbb{R}$, $F_n(x)$ is the relative frequency of successes in a Bernoulli sequence of trials with $EF_n(x) = F(x)$ and $\text{Var } F_n(x) = \frac{1}{n} F(x)(1-F(x))$.

Applying now Khinchine's version of the strong law of large numbers to the random variables $I_{(-\infty, x]}(X_i)$, $i \geq 1$, we have for fixed $x \in \mathbb{R}$:

$$(1.1.7.) \quad F_n(x) \xrightarrow{\text{a.s.}} F(x) \text{ as } n \rightarrow \infty.$$

In the terminology of statistics, this statement says: for fixed $x \in \mathbb{R}$, the empirical distribution function $F_n(x)$ is an unbiased and strongly consistent estimator for the theoretical distribution $F(x)$.

Looking back again at (1.1.7.) the question whether or not the convergence in (1.1.7.) can be made uniform arises. The answer to this question is given by the famous Glivenko-Cantelli Theorem:

e

$$(1.1.7.*) \quad \sup_{-\infty < x < \infty} |F_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \rightarrow \infty.$$

Definition 1.1.14.

The stochastic process $\beta_n = \{\beta_n(x) : x \in \mathbb{R}\}$, $n \geq 1$, defined pointwise via

$$\beta_n(x) := \sqrt{n}(F_n(x) - F(x)),$$

is called the empirical process.

For further properties and asymptotic behaviour of the empirical process we refer to M. Csörgő and P. Révész (1981), Chapter 4.

A special empirical process, the uniform empirical process, will play a role in Chapter 5:

Definition 1.1.15.

Let U_i , $i \geq 1$, be independent uniform $(0,1)$ -random variables defined on a common probability space and let E_n denote the uniform empirical distribution function of the random sample $U_1, U_2, \dots, U_n$, i.e.

$$E_n(y) := \frac{1}{n} \# \{1 \leq i \leq n : U_i \leq y\} \quad \text{for } 0 \leq y \leq 1.$$

Then the stochastic process α_n defined pointwise by

$$\alpha_n(y) := \sqrt{n}(E_n(y) - y)$$

is said to be a uniform empirical process.

Note that the covariance function of a uniform empirical process coincides with that of a Brownian bridge. For further results on the uniform empirical process we again refer to M. Csörgő and P. Révész (1981), Chapter 4.

In Chapter 2 of this thesis certain weight functions for L_1 - and L_2 -norm functionals of stochastic processes will be discussed. These weight functions are called Chibisov-O'Reilly functions. A detailed summary on Chibisov-O'Reilly functions, including an examination of their relation to Erdős-Feller-Kolmogorov-Petrovsky upper class functions, is contained in M. Csörgő, S. Csörgő, L. Horváth and D.M. Mason (1986). Therefore, we here only give

Definition 1.1.16:

(a) A function $q : (0, 1/2] \rightarrow \mathbb{R}$ is called positive if and only if

$$\inf_{\delta \leq s \leq 1/2} q(s) > 0 \quad \text{for all } 0 < \delta < 1/2.$$

(b) A function $w : (0, 1) \rightarrow \mathbb{R}$ is called positive if and only if

$$\inf_{\delta \leq s \leq 1-\delta} w(s) > 0 \quad \text{for all } 0 < \delta < 1/2.$$

(c) Let q be any positive function defined on $(0, 1/2]$ satisfying:

q is monotonically nondecreasing in a neighbourhood of zero.

Then q is called an Erdős-Feller-Kolmogorov-Petrovsky (EFKP) upper class function of a Brownian bridge $B = \{B(s) : 0 \leq s \leq 1\}$ if and only if

$$(1.1.8.) \quad \limsup_{s \downarrow 0} |B(s)|/q(s) < \infty \quad \text{a.s.}$$

If the limit superior in (1.1.8.) equals zero, then q is said to be a Chibisov-O'Reilly function.

Remarks 1.1.17.

(a) The following integrals play an important role in the characterizations of EFKP upper class and Chibisov-O'Reilly functions:

$$E(q,c) := \int_0^{1/2} s^{-3/2} q(s) \exp\{-cs^{-1}q^2(s)\} ds \text{ and}$$

$$I(q,c) := \int_0^{1/2} s^{-1} \exp\{-cs^{-1}q^2(s)\} ds,$$

where $0 < c < \infty$ is a constant.

(b) Similarly to (1.1.8.) an EFKP upper class function for a Wiener process W is defined by requiring

$$(1.1.8.*) \quad \limsup_{s \downarrow 0} |W(s)|/q(s) < \infty \text{ a.s.}$$

(c) A function w on $(0,1)$ is called an EFKP upper class function if $w(t)$ and $w(1-t)$ define EFKP upper class functions on $(0,1/2)$. The definition of a Chibisov-O'Reilly function can be extended in a similar way.

Section 1.1. Komlós-Major-Tusnády Approximations

At various points throughout the following chapters we will make use of the original Komlós-Major-Tusnády approximation (J. Komlós, P. Major, G. Tusnády (1975) and (1976)) as well as of some slight generalizations thereof. The original approximation consists of a coupling inequality involving partial sums and a Brownian motion (Wiener process). There are two versions of this approximation, one requiring the existence of the moment-generating function of the independent identically distributed random variables underlying the aforementioned partial sums and the other relying on moment-conditions only. Here we will primarily consider the former situation.

Let $X_1, X_2, \dots$ be independent identically distributed random variables with $EX_1 = \mu$, $E(X_1 - \mu)^2 = \text{Var } X_1 = \sigma^2 < \infty$. Put $S_n = X_1 + \dots + X_n$ for $n \geq 1$ and $S_0 = 0$. In this situation the original Komlós-Major-Tusnády theorem says:

Theorem 1.1.18. (Confer Theorem 2.6.2. in M. Csörgö and P. Révész (1981))

Assume that $R(t) = Ee^{tX_1}$ exists in a neighbourhood of $t = 0$.

Then there exists a Wiener process W such that for all real z and every $n \in \mathbb{N}$ we have:

$$(1.1.9.) \quad P\left\{\max_{1 \leq k \leq n} \left| \frac{S_k - k\mu}{\sigma} - W(k) \right| > C \log n + z\right\} < Ke^{-\lambda z},$$

where C, K, λ are positive constants, depending only on the distribution function of X_1 .

Remarks 1.1.19.

(a) The statement "there exists a Wiener process W such that

(1.1.9.) holds" should be understood as follows:

we can construct a new probability space (Ω, A, P) with a Wiener process W and with $\{\tilde{S}_k : k \geq 1\}$ living on it such that $\{\tilde{S}_k : k \geq 1\} \stackrel{D}{=} \{S_k : k \geq 1\}$ and statement (1.1.9.) holds with $\tilde{S}_k$ replacing S_k .

(b) Taking $z = D \log n$ for some suitably large constant $D > 0$,

(1.1.9.) readily implies:

$$(1.1.10.) \quad \max_{1 \leq k \leq n} \left| \frac{S_k - k\mu}{\sigma} - W(k) \right| \stackrel{a.s.}{=} O(\log n) \quad \text{as } n \rightarrow \infty \quad \text{and}$$

$$(1.1.10.*) \quad \left| \frac{S_n - n\mu}{\sigma} - W(n) \right| \stackrel{a.s.}{=} O(\log n) \quad \text{as } n \rightarrow \infty.$$

(c) Theorem 1.1.18. can easily be extended to partial sum processes

$\{Z(t) : t \geq 0\}$ coupled with $\{W(t) : t \geq 0\}$, where $Z(t) := \sum_{i=1}^{[t]} X_i = S_{[t]}$ and where the random variables, X_i are as before, i.e. the X_i 's are independent identically distributed with $EX_1 =: \mu$ and $0 < \text{Var } X_1 =: \sigma^2 < \infty$.

Theorem 1.1.20.

Under the conditions of Theorem 1.1.18. we have:

$$(1.1.11.) \quad P \left\{ \sup_{0 \leq t \leq T} \left| \frac{Z(t) - \mu t}{\sigma} - W(t) \right| > A \log T + z \right\} \leq B e^{-Cz}$$

for every $T \geq 2$ and $z \geq 0$.

Here $A, B, C > 0$ are absolute constants, depending only on the distribution of X_1 .

Proof of Theorem 1.1.20.

The main idea of this proof is to find an upper bound for the probability in statement (1.1.11.) by splitting the set

$$\left\{ \sup_{0 \leq t \leq T} \left| \frac{Z(t) - \mu t}{\sigma} - W(t) \right| > A \log T + z \right\}$$

into a union of four sets whose individual probabilities can then be estimated more easily:

$$\begin{aligned} & P \left\{ \sup_{0 \leq t \leq T} \left| \frac{Z(t) - \mu t}{\sigma} - W(t) \right| > A \log T + z \right\} \\ & \leq P \left\{ \sup_{0 \leq t \leq T} \left| \frac{Z([t]) - \mu[t]}{\sigma} - W([t]) \right| > \frac{1}{4}(A \log T + z) \right\} \\ & \quad + P \left\{ \sup_{0 \leq t \leq T} |Z(t) - Z([t])| > \frac{\sigma}{4}(A \log T + z) \right\} \\ & \quad + P \left\{ \sup_{0 \leq t \leq T} |W(t) - W([t])| > \frac{1}{4}(A \log T + z) \right\} \\ & \quad + P \left\{ \sup_{0 \leq t \leq T} |t - [t]| > \frac{\sigma}{4\mu}(A \log T + z) \right\}. \end{aligned}$$

Now the first of the above four probabilities can be estimated by

(1.1.9.) for $A \geq 4C$ while the stationary increment property of Z and the finiteness of $R(t_0) = E e^{t_0 X_1}$ for some $t_0 > 0$ lead to an appropriate upper bound for the second term. Finally, an application of Lemma 1.2.1 in M. Csörgö and P. Révész (1981), together with the stationary increment property of the Wiener process, gives the desired upper bound for the third probability. Since the last term is equal to zero for sufficiently large $A > 0$, this completes the proof of Theorem 1.1.20.

We note that a proof of Theorem 1.1.20. was given here because we will now summarize its essentials for possible further versions of Theorem 1.1.20. which may be of independent interest.

Let $Z = \{Z(t) : t \geq 0\}$ be an arbitrary real-valued stochastic process. Then we see from the beginning of the proof of Theorem 1.1.20. that this proof essentially hinges on two facts only:

- Z evaluated at any integer k can be represented as a partial sum of independent identically distributed random variables possessing a finite moment-generating function in some nondegenerate interval around zero and
- the sup-norm distance between $Z(t)$ and $Z([t])$ is small enough.

Clearly, the first of these conditions is satisfied by any time-stationary stochastic process with independent increments because we have

$$Z(k) = \sum_{i=1}^k X_i \quad \text{for} \quad X_i := Z(i) - Z(i-1), \quad i \geq 1,$$

provided that $Z(0) = 0$ a.s. and in addition, the stationary independent increment property implies now

- the random variables X_i are independent identically distributed,
- $\sup_{0 \leq t \leq T} |Z(t) - Z([t])| = \sup_{0 \leq x \leq 1} |Z(x)|$,
- $EZ(t) = m_0 + m_1 t$, where $m_0 = EZ(0)$ and $m_1 = EZ(1) - m_0$,
- $\text{Var } Z(t) = \sigma_0^2 + \sigma_1^2 t$, where $\sigma_0^2 = \text{Var } Z(0)$ and $\sigma_1^2 = \text{Var } Z(1) - \sigma_0^2$.

(Compare Remark 1.1.3. for the last two properties.)

$m_0 \neq \sigma_0^2 = 0$ provided that $Z(0) = 0$ almost surely.

Hence we can have a Komlós-Major-Tusnády-type approximation for any stochastic process Z obeying the following conditions:

- (i). Z has stationary independent increments,
- (ii) $Z(0) = 0$ almost surely,
- (iii) the moment-generating function of $Z(t)$ exists near 0,
- (iv) $\sup_{0 \leq x \leq 1} |Z(x)| < \infty$ almost surely with finite moment-generating function.

We note that the supremum in condition (iv) is almost surely finite if $Z(x)$ is almost surely non-decreasing in x as is the case for the Poisson and ordinary renewal processes. Hence (iv) could be relaxed to

(iv)* $Z(x)$ is almost surely non-decreasing as a function of x over $[0,1]$.

because (iv)* implies

$$\sup_{0 \leq x \leq 1} Z(x) = Z(1) \text{ almost surely}$$

and by (iii) the moment-generating function of $Z(1)$ is finite.

More explicitly we have for large enough $A > 0$:

$$\begin{aligned} & P \left\{ \sup_{0 \leq t \leq T} \left| \frac{Z(t) - (m_0 + m_1 t)}{\sqrt{(\sigma_0^2 + \sigma_1^2 t)/t}} - W(t) \right| > A \log T + z \right\} \\ & \leq P \left\{ \sup_{0 \leq t \leq T} \left| \frac{Z([t]) - (m_0 + m_1 t)}{\sqrt{(\sigma_0^2 + \sigma_1^2 t)/t}} - W([t]) \right| > \frac{1}{4}(A \log T + z) \right\} \\ & \quad + P \left\{ \sup_{0 \leq t \leq T} |Z(t) - Z([t])| > \frac{1}{4} \frac{(\sigma_0^2 + \sigma_1^2 t)^{1/2}}{t^{1/2}} (A \log T + z) \right\} \end{aligned}$$

$$\begin{aligned}
& + P \left\{ \sup_{0 \leq t \leq T} (t - [t]) > \frac{(\sigma_0^2 + \sigma_1^2 t)^{\frac{1}{2}}}{4m_1 t^{\frac{1}{2}}} (A \log T + z) \right\} \\
& + P \left\{ \sup_{0 \leq t \leq T} |W(t) - W([t])| > \frac{1}{4} (A \log T + z) \right\} \\
& \leq P \left\{ \max_{1 \leq k \leq T} \left| \frac{Z(k) - m_1 k}{\sigma_1} - W(k) \right| > \frac{1}{4} (A \log T + z) \right\} \\
& + P \left\{ \sup_{0 \leq x \leq 1} |Z(x)| > \frac{\sigma_1}{4} (A \log T + z) \right\} \\
& + 2C^* \exp \left\{ - \left(\frac{A}{24} \log 2 \right) z \right\}
\end{aligned}$$

and so by Theorem 1.1.20. we obtain

$$\leq 2Be^{-Cz} + P \left\{ \sup_{0 \leq x \leq 1} |Z(x)| > \frac{\sigma_1}{4} (A \log T + z) \right\}.$$

From here condition (iv) ensures the correct behaviour of the last term.

Remark 1.1.21.

Imitating the proof of Corollary 4.0.6. in Chapter 4 for a general partial sum process Z (satisfying the assumptions of Theorem 1.1.20.) instead of the Poisson process used there, we get for $\delta > 0$:

$$\sup_{\delta \leq x \leq n/2} x^{-\alpha} \left| \frac{Z(x) - \mu x}{\sigma} - W^{(1)}(x) \right| = O_p(1) \text{ as } n \rightarrow \infty$$

and

$$\sup_{n/2 \leq x \leq n-\delta} (n-x)^{-\alpha} \left| \frac{Z(n-x) - (n-x)\mu}{\sigma} - W^{(2)}(n-x) \right| = O_p(1) \text{ as } n \rightarrow \infty,$$

where $W^{(1)}$ and $W^{(2)}$ are independent Wiener processes. Here $0 < \alpha \leq 1/2$ fixed.

Section 1.2. Further Results from Probability Theory

The first result we will present in this section, is a general version of Kolmogorov's extension theorem. It will be used in particular in Chapter 4. For a proof of Kolmogorov's extension theorem we refer to L. Breiman (1968), Theorem 12.14.; confer also Theorem 2.26. there.

Theorem 1.1.22. (Kolmogorov's extension theorem)

Let I be a compact subinterval of the positive real half-line.

Let $\{F_{t_1 \dots t_n}(x_1, x_2, \dots, x_n) : t_1 < t_2 < \dots < t_n, t_i \in I, x_i \in \mathbb{R}, 1 \leq i \leq n, n \in \mathbb{N}\}$ describe a set of consistent distribution functions, i.e. •

$$\lim_{x_k \uparrow \infty} F_{t_1 \dots t_n}(x_1, \dots, x_n) = F_{t_1 \dots \hat{t}_k \dots t_n}(x_1, \dots, \hat{x}_k, \dots, x_n),$$

where $\hat{}$ denotes missing.

Then there exists a stochastic process $\{X(t)\}_{t \in I}$ such that

$$F_{t_1 \dots t_n}(x_1, \dots, x_n) = P\{X(t_1) < x_1, \dots, X(t_n) < x_n\}.$$

Next we display the famous Donsker Theorem. A proof can be found, e.g., in M. Csörgö and P. Révész (1981), p.90.

Theorem 1.1.23. (Donsker's Theorem)

Let $X_1, X_2, \dots$ be independent identically distributed random variables with $EX_1 = 0$ and $EX_1^2 = 1$. Put $S_0 = 0$, and $S_n = X_1 + \dots + X_n$ for $n \geq 1$. Define

$$S_n(t) := n^{-1/2} \{S_{[nt]} + (nt - [nt])X_{[nt]+1}\} \quad \text{for } 0 \leq t \leq 1.$$

Then we have for every continuous $h : C(0,1) \rightarrow \mathbb{R}$:

$$h(S_n(t)) \xrightarrow{D} h(W(t)) \quad \text{as } n \rightarrow \infty,$$

where W denotes a standard Wiener process.

Note that a member of the stochastic process $\{S_n(t) : 0 \leq t \leq 1\}$ on $C(0,1)$ is often called a Donskerized partial sum.

The following version of the law of iterated logarithm for the Wiener process will be used throughout Chapter 2.

Theorem 1.1.24. (Confer Theorems 1.3.1 and 1.3.1.* in M. Csörgő and P. Révész (1981))

$$\limsup_{T \rightarrow \infty} \frac{|W(T)|}{\sqrt{2T \log \log T}} = \limsup_{T \rightarrow \infty} \sup_{0 \leq t \leq T} \frac{|W(t)|}{\sqrt{2T \log \log T}} \stackrel{\text{a.s.}}{=} 1.$$

In the sequel we present two results that will be needed in order to prove central limit theorems in L_2 -norm for various stochastic processes (confer in particular Section 2.2 of Chapter 2.). Similar results, involving uniform quantile processes and Brownian bridges in lieu of the stochastic processes and Brownian motion considered here, were proven by M. Csörgő and L. Horváth (1986).

Let $X_1, X_2, \dots$ be independent identically distributed random variables with ν finite absolute moments for some $\nu > 2$, i.e.

$$E|X_1|^\nu < \infty \quad \text{for some } \nu > 2. \quad \text{Assume } EX_1 = 0 \quad \text{and } \text{Var } X_1 = EX_1^2 =:$$

$\sigma^2 > 0$. Put $S_0 = 0$ and $S_n = X_1 + \dots + X_n$ for $n \geq 1$. Define

$S(x) := \sum_{i=1}^{[x]} x_i$, where $[x]$ denotes the largest integer less than or equal to x , $x > 0$. Then we have:

Theorem 1.1.25.

Under the aforementioned conditions there exists a Wiener process $W = \{W(x) : x \geq 0\}$ such that the following hold true as $n \rightarrow \infty$:

$$(1.1.12.) \quad n^\tau \sup_{\frac{1}{n} \leq t \leq 1} \frac{|S(nt) - W(nt)|}{\sqrt{n} t^{1/2-\tau}} = O_p(1)$$

whenever $0 \leq \tau < 1/2 - 1/\nu$.

$$(1.1.13.) \quad n^{-1/2} \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|S(nt)|}{q(t)} = O_p(1)$$

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

Proof of Theorem 1.1.25.

Statement (1.1.12.) is an immediate consequence of the Komlós-Major-Tusnády approximation theorem in the case of ν finite absolute moments because in this case we have (confer, e.g., M. Csörgö and P. Révész (1981), Theorem 2.6.3.):

$$|S(x) - W(x)| \stackrel{\text{a.s.}}{=} o(x^{1/\nu}) \quad \text{as } x \rightarrow \infty.$$

Here W is that Wiener process which couples with $S(x)$ according to the Komlós-Major-Tusnády Theorem. Hence we obtain directly:

$$(1.1.14.) \quad \left\{ \begin{array}{l} \sup_{\lambda \leq x < \infty} \frac{|S(x) - W(x)|}{x^{1/2-\tau}} \stackrel{\text{a.s.}}{\leq} \sup_{\lambda \leq x < \infty} O(x^{1/\nu - 1/2 + \tau}) \\ = O(\lambda^{\tau - (1/2 - 1/\nu)}) = O(1) \end{array} \right.$$

for all $\lambda > 0$ fixed, provided that $0 \leq \tau < 1/2 - 1/\nu$.

Substituting $x = nt$ in (1.1.14.) (take $\lambda = 1$), we arrive at

$$\sup_{\frac{1}{n} \leq t \leq 1} \frac{|S(nt) - W(nt)|}{(nt)^{\frac{1}{2}-\tau}} = n^{\tau} \sup_{\frac{1}{n} \leq t \leq 1} \frac{|S(nt) - W(nt)|}{\sqrt{n} t^{\frac{1}{2}-\tau}} \stackrel{\text{a.s.}}{=} O(1).$$

Towards proving (1.1.13.), we apply the following estimations:

$$\begin{aligned} & n^{-\frac{1}{2}} \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|S(nt)|}{q(t)} \\ & \leq n^{-\frac{1}{2}} \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|S(nt) - W(nt)|}{q(t)} + n^{-\frac{1}{2}} \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|W(nt)|}{q(t)} \\ & = n^{\tau} \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|S(nt) - W(nt)|}{\sqrt{n} t^{\frac{1}{2}-\tau}} \cdot \frac{\sqrt{t} (nt)^{-\tau}}{\{t(1-t) \log \log \frac{1}{t(1-t)}\}^{\frac{1}{2}}} \\ & \quad + \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|W(t)|}{q(t)} \\ & \leq \left\{ \frac{1}{2} \log \log 2 \right\}^{-\frac{1}{2}} \cdot \left\{ n^{\tau} \sup_{\frac{1}{n} \leq t \leq 1} \frac{|S(nt) - W(nt)|}{\sqrt{n} t^{\frac{1}{2}-\tau}} \right\} + \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|W(t)|}{q(t)} \\ & = O_p(1) + \sup_{\frac{1}{n} \leq t \leq \frac{1}{2}} \frac{|W(t)|}{q(t)}. \end{aligned}$$

Since q is an EFKP upper-class function (confer (1.1.8.)* for a definition) for a Brownian bridge and hence also for a Wiener process, the remaining sup is bounded in probability which completes the proof of Theorem 1.1.25.

Chapter 2. Nonparametric Test Statistics for the Fixed Sample Size Change-point Problem

Section 2.0. Introduction

A variety of different change-point problems originally arose in the context of quality control. In statistical quality control one typically observes the output of a production line and wishes to signal deviation from an acceptable average output level. Models for change-point problems may be divided into four categories according to whether they follow a parametric or a nonparametric approach and according to whether they are based on a fixed sample size or on a sequential procedure. Furthermore, each of these four categories may be separated into subcategories according to the form of the hypotheses. In most cases the hypotheses will reflect one of the following situations:

- at most one change-point occurs; the so-called AMOC-setting
(confer review article by D.A. Wolfe and E. Schechtman (1984))
- no change-point under H_0 , and exactly two change-points under H_A ;
the epidemic or square wave setting (confer, e.g., D. Siegmund (1985)).

One of the earliest models suggested for the nonparametric fixed sample size AMOC-setting originates from E.S. Page (1954), (1955). Page considers the special situation:

$X_1, X_2, \dots, X_k, \dots, X_n$ are independent continuous random variables such that $X_1, \dots, X_k$ have distribution $F(x)$ and $X_{k+1}, \dots, X_n$ have distribution $F(x-\Delta)$, where $-\infty < \Delta < \infty$.

The unknown Δ is called a shift parameter and represents the magnitude of change. The above introduced integer k is called a

changepoint. The one-sided test

$$H_0 : \Delta = 0$$

versus

$$H_A : \text{exactly one changepoint occurs and } \Delta > 0$$

given that EX_1 is known a priori, is among those tests considered by Page.

After modelling a specific changepoint problem, the next step is to find an appropriate test statistic. The choice of the test statistic is important for the derivation of a decision rule for accepting/rejecting H_0 . Most of the earlier suggestions for test statistics were, like the aforementioned one by Page, based on the assumption that the possible change is a shift in distribution. Sometimes the assumption that the observed data follow a prespecified distribution law (or family of distributions) was used.

Recently, M. Csörgő and L. Horváth (1988a) formulated a nonparametric approach to the fixed sample size AMOC-problem for detecting a possibly more general change in distribution than merely a shift. They propose test statistics that are based on processes of U-statistics to test the hypotheses

$$H_0 : X_1, X_2, \dots, X_n \text{ have the same distribution}$$

against

$$H_A : \text{there exists a } \lambda \in (0,1) \text{ such that}$$

$$P(X_1 \leq t) = \dots = P(X_{[n\lambda]} \leq t) \text{ and}$$

$$P(X_{[n\lambda]+1} \leq t) = \dots = P(X_n \leq t) \text{ for all } t \in \mathbb{R} \text{ and}$$

$$P(X_1 \leq t_0) \neq P(X_n \leq t_0) \text{ for some } t_0 \in \mathbb{R}.$$

Csörgö and Horváth study stochastic processes which lead to test statistics that are of the generalized Wilcoxon-Mann-Whitney type. We will examine these processes in Section 2.1. In Section 2.2 we will develop this approach further to obtain Cramér-von Mises type tests based again on processes of U-statistics. While Sections 2.1 and 2.2 only deal with the null hypothesis of no change, Section 2.3 will be devoted to the discussion of the behaviour of various test statistics under alternative hypotheses.

Section 2.1. Wilcoxon-Mann-Whitney Type Tests

2.1.0. Basic Concepts and Notations

The tests to be considered in this section are related to processes of one and two sample U-statistics. In our terminology and in the notations we will follow R.J. Serfling (1980).

Let $X_1, X_2, \dots, X_n$ be independent observations on a distribution F . Let $\theta = \theta(F)$ be a "parametric function" for which there is an unbiased estimator. Then $\theta(F)$ may be represented as

$$\theta(F) = E_F(h(X_1, \dots, X_m)), \quad 1 \leq m \leq n \text{ fixed,}$$

for some function $h : \mathbb{R}^m \rightarrow \mathbb{R}$. Such a function h is called a kernel. Without loss of generality h is often assumed to be symmetric in all m variables when using h for estimating θ . For, if h is not symmetric, it can be replaced by the symmetric kernel

$$\frac{1}{m!} \sum h(x_{i_1}, x_{i_2}, \dots, x_{i_m}),$$

where the summation extends over the $m!$ permutations $(i_1, \dots, i_m)$ of $\{1, 2, \dots, m\}$. However, we will also consider antisymmetric kernels h (in the context of hypotheses testing).

Definition 2.1.0.

For any kernel h , the corresponding U-statistic formed on the basis of the sample $X_1, X_2, \dots, X_n$ of size $n \geq m$, is obtained by averaging the kernel h symmetrically over the observations:

$$U_n = U(X_1, X_2, \dots, X_n) = \binom{n}{m}^{-1} \sum h(x_{i_1}, \dots, x_{i_m}),$$

where $\sum$ denotes the summation over the $\binom{n}{m}$ combinations of m

distinct elements $i_1, i_2, \dots, i_m$ from $\{1, 2, \dots, n\}$.

Examples 2.1.1.

(a) Let $\theta(F) = \mu(F)$ be the mean of the distribution F . Choosing $h(x) = x$ as the kernel ($m=1$), we obtain

$$U(X_1, X_2, \dots, X_n) = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X};$$

hence the corresponding U-statistic is the sample mean $\bar{X}$.

(b) Let $\theta(F) = \sigma^2(F)$ be the variance of the distribution F . For $m=2$ and $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$, the corresponding U-statistic is

$$U(X_1, X_2, \dots, X_n) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j) = \frac{1}{(n-1)} \left\{ \sum_{i=1}^n X_i^2 - n\bar{X}^2 \right\} = s^2.$$

Thus in this case, the U-statistic is the sample variance s^2 .

(c) To estimate $\theta(F) = P_F(X_1 + X_2 \leq 0)$, take $m=2$ and $h(x_1, x_2) = I(x_1 + x_2 \leq 0)$, where I denotes the indicator function. Then the corresponding U-statistic is the Wilcoxon one-sample test statistic

$$U(X_1, X_2, \dots, X_n) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} I(X_i + X_j \leq 0).$$

When dealing with one-sample U-statistics in Chapters 2 and 3 we will

always choose $m=2$, i.e. we will study kernel functions h that

depend on two (real-valued) arguments only. Now we extend the concept

of one-sample U-statistics to the case of two samples. Let

$\{X_1, X_2, \dots, X_{n_1}\}$ and $\{Y_1, Y_2, \dots, Y_{n_2}\}$ be two independent collections

of independent observations taken from distributions F and G ,

respectively. Denote by $\theta = \theta(F, G)$ a "parametric function" for which

an unbiased estimator exists. Again, we may represent θ by

$$\theta(F, G) = E(h(X_1, \dots, X_{m_1}, Y_1, \dots, Y_{m_2}))$$

using a kernel function $h : \mathbb{R}^{m_1+m_2} \rightarrow \mathbb{R}$ which is assumed to be symmetric within each of the two blocks of arguments when being used for estimation of the parameter(s).

Definition 2.1.2.

The two-sample U-statistic corresponding to the kernel h for the estimation of θ is defined as

$$\begin{aligned} U_n &= U(X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}) \\ &= \binom{n_1}{m_1}^{-1} \binom{n_2}{m_2}^{-1} \sum h(X_{i_{11}}, \dots, X_{i_{1m_1}}, Y_{i_{21}}, \dots, Y_{i_{2m_2}}) \end{aligned}$$

where $n = n_1 + n_2$, provided that $m_1 \leq n_1$ and $m_2 \leq n_2$. Here for $j = 1, 2$ $\{i_{j1}, \dots, i_{jm_j}\}$ denotes a set of m_j distinct elements from $\{1, 2, \dots, n_j\}$.

A standard example for a two-sample U-statistic is the Wilcoxon two-sample test statistic.

Example 2.1.3.

Let $\{X_1, \dots, X_{n_1}\}$ and $\{Y_1, \dots, Y_{n_2}\}$ be as above, but require in addition that the two distributions F and G are continuous. Then an unbiased estimator for $\theta(F, G) = \int F dG = P(X \leq Y)$ is the two-sample U-statistic

$$U_n = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_i \leq Y_j)$$

Next, we turn our attention to the following properties of test statistics: distribution free and nonparametric distribution free. In our terminology we follow R.H. Randles and D.A. Wolfe (1979), Section 2.1. Let $X_1, X_2, \dots, X_n$ be random variables over a common probability space $(\Omega, \mathcal{A}, P)$ with a joint distribution denoted by D , where $D \in \mathcal{D}$ for some collection $\mathcal{D}$ of possible joint distribution functions.

Definition 2.1.4.

Suppose $T = T(X_1, X_2, \dots, X_n)$ is a test statistic based on $X_1, X_2, \dots, X_n$. Then T is said to be distribution free over $\mathcal{D}$, if the distribution of T is the same for every joint distribution in $\mathcal{D}$.

Considering the U-statistics of Examples 2.1.1.(a) and (b) we have

$$U_1 = \sqrt{n}(\bar{X} - \mu_0)/s \stackrel{\mathcal{D}}{=} t \text{ is distribution free over } \mathcal{D}_1,$$

where $\mathcal{D}_1 = \{D \mid D \text{ is joint distribution of } n \text{ independent identically distributed } N(\mu_0, \sigma^2) \text{-random variables for some } 0 < \sigma^2 < \infty\}$;

$$U_2 = \frac{(n-1)}{\sigma_0^2} s^2 \stackrel{\mathcal{D}}{=} \chi_{n-1}^2 \text{ is distribution free on } \mathcal{D}_2,$$

where $\mathcal{D}_2 = \{D \mid D \text{ is joint distribution of } n \text{ independent identically distributed } N(\mu, \sigma_0^2) \text{-random variables for some } \mu \in \mathbb{R}\}$. Thus each of these statistics has a distribution free property over a parametric class of independent identically distributed random variables.

Definition 2.1.5.

A test statistic T is said to be nonparametric distribution free over a class $\mathcal{D}$, if the class $\mathcal{D}$ includes more than one distribution

form.

The two-sample U-statistic of Example 2.1.3 is the Mann-Whitney form of the Wilcoxon rank statistic which is a nonparametric distribution free over $\mathcal{D}_3 = \{D \mid D \text{ is joint distribution of } n_1 + n_2 \text{ independent identically distributed continuous random variables}\}$. (Confer also Example 2.1.11. later on.)

Let now, and in the following, $X_1, X_2, \dots, X_n$ be independent real-valued random variables. In this chapter we wish to test the null hypothesis

$$H_0 : X_1, X_2, \dots, X_n \text{ have the same distribution}$$

against the alternative hypothesis that exactly one changepoint occurs, i.e.

$$H_A : \text{there exists a } \lambda \in (0,1) \text{ such that}$$

$$P(X_1 \leq t) = \dots = P(X_{[n\lambda]} \leq t) \text{ and}$$

$$P(X_{[n\lambda]+1} \leq t) = \dots = P(X_n \leq t) \text{ for all } t \in \mathbb{R} \text{ and}$$

$$\text{there exists a } t_0 \in \mathbb{R} \text{ with } P(X_1 \leq t_0) \neq P(X_n \leq t_0).$$

Here $[x]$ denotes the largest integer less than or equal to x . In this section we are going to consider functionals of

$$(2.1.0.) \quad Z_k = \sum_{1 \leq i \leq k} \sum_{k+1 \leq j \leq n} h(X_i, X_j) \text{ for } 1 \leq k \leq n-1$$

under the null hypothesis for symmetric as well as for antisymmetric kernel functions h . These functionals will be used later on to develop test statistics. Assume

$$(2.1.1) \quad E h^2(X_1, X_2) < \infty,$$

put $\theta = E h(X_1, X_2)$ and for symmetric h define $\tilde{h}(t) = E\{h(t, X_1) - \theta\}$.

For any antisymmetric kernel h , $\theta = 0$. In this case we define

$\hat{h}(t) = E h(t, X_1)$. Condition (2.1.1.) implies that $E \tilde{h}^2(X_1)$ and $E \hat{h}^2(X_1)$ are finite. Let us now assume that

$$(2.1.2.) \quad E \tilde{h}^2(X_1) = \sigma^2 > 0 \quad \text{and} \quad E \hat{h}^2(X_1) = \sigma^2 > 0, \quad \text{respectively.}$$

We wish to investigate

$$U_k = Z_k - k(n-k)\theta \quad \text{for } 1 \leq k \leq n-1$$

in the case of a symmetric kernel h , and

$$\hat{U}_k = Z_k \quad \text{for } 1 \leq k \leq n-1$$

in the case of an antisymmetric kernel h .

We see that for any antisymmetric kernel and for symmetric kernels with $\theta = 0$, the statistics $\hat{U}_k$ and U_k for fixed k have the form of a generalized Wilcoxon-two sample statistic (take $n_1 = k$ and $n_2 = n-k$ in Example 2.1.3.). However, we will not be interested in $\hat{U}_k$ and U_k for fixed k but for varying k . This, in terms of a Wilcoxon-two sample statistic, means that we keep changing the sample sizes. In Subsection 2.1.3. of Chapter 2 we will consider the consequences of this phenomenon. Although U_k and $\hat{U}_k$ are themselves not one-sample U-statistics, they may be represented as linear combinations of nondegenerate one-sample U-statistics provided that (2.1.2.) holds:

$$(2.1.3.) \quad \begin{cases} U_k = U_n^{(3)} - \{U_k^{(1)} + U_k^{(2)}\} & , \text{ where} \\ U_k^{(1)} = \sum_{1 \leq i < j \leq k} h(X_i, X_j) - \binom{k}{2} \theta & , \\ U_k^{(2)} = \sum_{k+1 \leq i < j \leq n} h(X_i, X_j) - \binom{n-k}{2} \theta & \text{ and} \\ U_n^{(3)} = \sum_{1 \leq i < j \leq n} h(X_i, X_j) - \binom{n}{2} \theta & ; \end{cases}$$

$$(2.1.4.) \quad \begin{cases} \hat{U}_k = Z_k = Z_n^{(3)} - \{Z_k^{(1)} + Z_k^{(2)}\} & , \text{ where} \\ Z_k^{(1)} = \sum_{1 \leq i < j \leq k} h(X_i, X_j) & , \\ Z_k^{(2)} = \sum_{k+1 \leq i < j \leq n} h(X_i, X_j) & \text{ and} \\ Z_n^{(3)} = \sum_{1 \leq i < j \leq n} h(X_i, X_j) & . \end{cases}$$

The representations (2.1.3.) and (2.1.4.), which are easily verifiable, will be the starting points for our considerations in subsections 2.1.2. and 2.1.1., respectively. In 2.1.1. we will examine the asymptotic behaviour of L_1 -norm functionals of $\hat{U}_k$ under the null hypothesis while in 2.1.2. we will study similar results for the case of U_k ..

2.1.1. Asymptotic Behaviour of $\hat{U}_k$ in sup-Norm and in Weighted sup-Norms Under the Null Hypothesis

In this subsection we assume that h is an antisymmetric kernel, i.e. $h(x,y) = -h(y,x)$ for all $x,y \in \mathbb{R}$. A well known example for such a kernel is the kernel of the cumulative rank statistic, $h(x,y) = \text{sign}(x-y)$. This particular kernel was already studied by A.N. Pettitt (1979) whose results we will discuss at the end of this subsection. Further examples for antisymmetric kernels are

$$h_1(x,y) = x-y \text{ and } h_2(x,y) = x^2 - y^2.$$

Let us recall that in the case of an antisymmetric kernel we have

$$Eh(X_1, X_2) = 0 ; \quad \hat{h}(t) = Eh(t, X_1) \text{ for } t \in \mathbb{R}.$$

We assumed (2.1.1.) and (2.1.2.)

$$Eh^2(X_1, X_2) < \infty ; \quad E\hat{h}^2(X_1) = \sigma^2 > 0.$$

Considering again h_1 and h_2 as above, we obtain:

$$E\hat{h}_1^2(X_1) = E[E_{X_2}(X_1 - X_2)]^2 = E[X_1 - EX_2]^2 = \text{Var } X_1, \text{ whence } h_1 \text{ satisfies (2.1.1.) and (2.1.2.) if and only if } 0 < \text{Var } X_1 < \infty.$$

$$E\hat{h}_2^2(X_1) = E[E_{X_2}(X_1^2 - X_2^2)]^2 = E[X_1^2 - EX_2^2]^2 = EX_1^4 - (EX_1^2)^2 \text{ and thus } h_2 \text{ satisfies (2.1.1.) and (2.1.2.) if and only if } X_1 \text{ has four finite moments and } EX_1^4 \neq (EX_1^2)^2. \text{ Here } E_{X_2} \text{ denotes the expected value taken only with respect to the random variable } X_2.$$

Now we wish to study the limiting behaviour of

$$\hat{U}_k^* = Z_k = \sum_{1 \leq i \leq k} \sum_{k+1 \leq j \leq n} h(X_i, X_j)$$

under the null hypothesis of no change in various L_1 -norms (sup-norms).

To do this we first utilize representation (2.1.4.) for $\hat{U}_k = Z_k$:

$$\hat{U}_k = Z_k^{(3)} = Z_n^{(3)} - \{Z_k^{(1)} + Z_k^{(2)}\} \quad \text{for } 1 \leq k \leq n-1.$$

Then statements concerning the asymptotic behaviour of $\hat{U}_k$ will be derived from the following "reduction principle" (Lemma 2.1.6.). The basic idea for the proof of this principle originates from S. Janson and M.J. Wichura (1983) (confer in particular Theorem 2.1. there). However, Janson and Wichura state their result in the context of stochastic area integrals and do not present a detailed proof, whence we are going to give an explicit verification in the following. An explicit even though not original treatment of the reduction principle seems to be in order here since Lemma 2.1.6., although indirectly through Corollary 2.1.7., is of vital importance for the proof of Theorem 2.1.8.

Lemma 2.1.6.

Let h , $\hat{U}_k$ and Z_k be as described above. Assume that H_0 is true and define

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

Then the following statements hold as $n \rightarrow \infty$

$$\begin{aligned} \max_{1 \leq k \leq n} |Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i)| &= O_p(n) \\ \max_{1 \leq k \leq n} |Z_k^{(2)} - \sum_{i=k+1}^n (n+k-2i+1) \hat{h}(X_i)| &= O_p(n) \\ |Z_n^{(3)} - \sum_{i=1}^n (n-2i+1) \hat{h}(X_i)| &= O_p(n) \end{aligned}$$

Corollary 2.1.7.

Under the conditions of Lemma 2.1.6. we have as $n \rightarrow \infty$

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

Proof of Corollary 2.1.7.

$$\begin{aligned} & \max_{1 \leq k \leq n} \left| \hat{U}_k - \left\{ n \sum_{i=1}^k \hat{h}(X_i) - k \sum_{i=1}^n \hat{h}(X_i) \right\} \right| \\ & \leq \max_{1 \leq k \leq n} \left| \left[Z_n^{(3)} - \sum_{i=1}^n (n-2i+1) \hat{h}(X_i) \right] - \left[Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i) \right] \right. \\ & \quad \left. - \left[Z_k^{(2)} - \sum_{i=k+1}^n (n+k-2i+1) \hat{h}(X_i) \right] \right| \\ & \quad + \max_{1 \leq k \leq n} \left| \sum_{i=1}^n (n-2i+1) \hat{h}(X_i) - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i) \right. \\ & \quad \left. - \sum_{i=k+1}^n (n+k-2i+1) \hat{h}(X_i) - n \sum_{i=1}^k \hat{h}(X_i) + k \sum_{i=1}^n \hat{h}(X_i) \right| \\ & \leq \left| Z_n^{(3)} - \sum_{i=1}^n (n-2i+1) \hat{h}(X_i) \right| + \max_{1 \leq k \leq n} \left| Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i) \right| \\ & \quad + \max_{1 \leq k \leq n} \left| Z_k^{(2)} - \sum_{i=k+1}^n (n+k-2i+1) \hat{h}(X_i) \right| + 0, \end{aligned}$$

from here Lemma 2.1.6. immediately implies Corollary 2.1.7. ■

Corollary 2.1.7. clearly shows the usefulness of the reduction principle because now we have an approximation of $\hat{U}_k$ by sums of independent identically distributed random variables and for such sums limit theorems are readily available.

Proof of Lemma 2.1.6.

We will only verify the first statement

$$\max_{1 \leq k \leq n} |Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i)| = O_p(n) \text{ as } n \rightarrow \infty.$$

The remaining two statements of Lemma 2.1.6. can be shown in the same vein. First, we will prove that

$$Z_k^* = Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i) \text{ defines a martingale,}$$

then we will use a martingale version of Kolmogorov's inequality to deduce

$$\limsup_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq n} |Z_k^*| \geq nK \right) = 0,$$

which by definition of $O_p(n)$ will imply our assertion.

For the definition and basic properties of martingales as well as for the notations used here we refer to P. Billingsley (1986), Sections 34 and 35. Put $F_k = \sigma(X_1, \dots, X_k)$ for $k \in \mathbb{N}$. Then, because the other requirements for a martingale are immediate, it suffices to show that

$$E[Z_{k+1}^* | F_k] = Z_k^* \text{ for every } k \in \mathbb{N}$$

in order for $(Z_k^*, F_k)_{k \in \mathbb{N}}$ to be a martingale.

$$\begin{aligned} E[Z_{k+1}^* | F_k] &= E \left[\sum_{1 \leq i < j \leq k+1} h(X_i, X_j) - \sum_{j=1}^{k+1} (k+1-2j+1) \hat{h}(X_j) \mid F_k \right] \\ &= E \left[\sum_{j=2}^{k+1} \sum_{i=1}^{j-1} h(X_i, X_j) - \sum_{j=1}^{k+1} (k+1-j) \hat{h}(X_j) + \sum_{j=1}^{k+1} (j-1) \hat{h}(X_j) \mid F_k \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[\sum_{j=2}^k \sum_{i=1}^{j-1} h(X_i, X_j) + \sum_{i=1}^k h(X_i, X_{k+1}) - \sum_{j=2}^{k+1} \sum_{i=1}^{j-1} \hat{h}(X_i) + \sum_{j=2}^{k+1} \sum_{i=1}^{j-1} \hat{h}(X_j) \mid F_k \right] \\
&= E \left[\sum_{j=2}^k \sum_{i=1}^{j-1} h(X_i, X_j) - \sum_{j=2}^k \sum_{i=1}^{j-1} \hat{h}(X_i) + \sum_{j=2}^k \sum_{i=1}^{j-1} \hat{h}(X_j) \mid F_k \right] \\
&\quad + E \left[\sum_{i=1}^k h(X_i, X_{k+1}) - \sum_{i=1}^k \hat{h}(X_i) + k \hat{h}(X_{k+1}) \mid F_k \right] \\
&= E[Z_k^* \mid F_k] + E \left[\sum_{i=1}^k \{h(X_i, X_{k+1}) - \hat{h}(X_i)\} \mid F_k \right] + k E[\hat{h}(X_{k+1}) \mid F_k] \\
&= Z_k^* + E \left[\sum_{i=1}^k \{h(X_i, X_{k+1}) - \hat{h}(X_i)\} \mid F_k \right],
\end{aligned}$$

the latter equality being true because Z_k^* is measurable- F_k , X_{k+1} is independent of F_k and $Eh(X_{k+1}) = 0$. Furthermore,

$$\begin{aligned}
&E \left[\sum_{i=1}^k \{h(X_i, X_{k+1}) - \hat{h}(X_i)\} \mid F_k \right] \\
&= \sum_{i=1}^k \{E[h(X_i, X_{k+1}) \mid F_k] - E[\hat{h}(X_i) \mid F_k]\} \\
&= \sum_{i=1}^k \{E[h(X_i, X_{k+1}) \mid F_k] - \hat{h}(X_i)\} \\
&= k \cdot \{E[h(X_1, X_{k+1}) \mid F_k] - \hat{h}(X_1)\}.
\end{aligned}$$

Since $\hat{h}(X_1(\omega)) = \int_{\Omega} h(X_1(\omega), X_{k+1}(\omega')) dF(X_{k+1}(\omega'))$ for every $\omega \in \Omega$,

assuming that F is the cumulative distribution function of $X_1, X_2, \dots, X_n$, it follows that

$$E \left[\sum_{i=1}^k \{h(X_i, X_{k+1}) - \hat{h}(X_i)\} \mid F_k \right] = 0; \text{ hence}$$

$$E[Z_{k+1}^* | F_k] = Z_k^* \text{ for every } k \in \mathbb{N}.$$

Thus $(Z_k^*, F_k)_{k \in \mathbb{N}}$ is in fact a martingale.

G.R. Shorack and J.A. Wellner (1986) state in Appendix A of their book that the Hájek-Rényi inequality, which contains Kolmogorov's inequality as a special case, is also valid for (sub-)martingales. For a proof we refer to M. Csörgö (1968). Using this and the fact that $(Z_k^*, F_k)_{k \in \mathbb{N}}$ is indeed a martingale, we obtain

$$\begin{aligned} & P \left(\max_{1 \leq k \leq n} \left| Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i) \right| \geq Kn \right) \\ & \leq (Kn)^{-2} \sum_{k=1}^n E \left\{ \left(Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i) \right) - \left(Z_{k-1}^{(1)} - \sum_{i=1}^{k-1} (k-1-2i+1) \hat{h}(X_i) \right) \right\}^2 \\ & = (Kn)^{-2} \sum_{k=1}^n E \left\{ \left(\sum_{j=2}^k \sum_{i=1}^{j-1} h(X_i, X_j) - \sum_{j=2}^{k-1} \sum_{i=1}^{j-1} h(X_i, X_j) \right) \right. \\ & \quad \left. - \left((k-2k+1) \hat{h}(X_k) + \sum_{i=1}^{k-1} \hat{h}(X_i) \right) \right\}^2 \\ & = (Kn)^{-2} \sum_{k=1}^n E \left\{ \sum_{i=1}^{k-1} (h(X_i, X_k) + \hat{h}(X_k) - \hat{h}(X_i)) \right\}^2 \\ & = (Kn)^{-2} \sum_{k=1}^n \left\{ \sum_{i=1}^{k-1} E[h(X_i, X_k) + \hat{h}(X_k) - \hat{h}(X_i)]^2 \right. \\ & \quad \left. + 2 \sum_{i < j} E[h(X_i, X_k) + \hat{h}(X_k) - \hat{h}(X_i)][h(X_j, X_k) + \hat{h}(X_k) - \hat{h}(X_j)] \right\} \\ & = (Kn)^{-2} \sum_{k=1}^n \left\{ (k-1) E(h(X_1, X_k) + \hat{h}(X_k) - \hat{h}(X_1))^2 \right. \\ & \quad \left. + (k-1)(k-2) [Eh(X_1, X_k)h(X_2, X_k) + Eh(X_1, X_k)\hat{h}(X_k) - Eh(X_1, X_k)\hat{h}(X_2)] \right\} \end{aligned}$$

$$+ Eh(X_2, X_k) \hat{h}(X_k) + Eh^2(X_k) - Eh(X_k) Eh(X_2) - Eh(X_2, X_k) Eh(X_1) - Eh(X_1) Eh(X_k) \\ + Eh(X_1) \hat{h}(X_2)]]$$

$$\leq (Kn)^{-2} \{ n(n-1) E(h(X_1, X_k) + \hat{h}(X_k) - \hat{h}(X_1))^2$$

$$+ \sum_{k=1}^n (k-1)(k-2) [Eh(X_1, X_k) h(X_2, X_k) + Eh(X_1, X_k) \hat{h}(X_k) + Eh(X_2, X_k) \hat{h}(X_k) \\ + Eh^2(X_k)]] \}$$

$$\leq K^{-2} E(h(X_1, X_k) + \hat{h}(X_k) - \hat{h}(X_1))^2 \quad \text{using the following argument:}$$

$$Eh(X_1, X_k) h(X_2, X_k) = - Eh(X_1, X_k) h(X_k, X_2) = - Eh(X_1, X_k) \hat{h}(X_k)$$

and similarly

$$Eh(X_2, X_k) \hat{h}(X_k) = - Eh(X_k, X_1) \hat{h}(X_k) = - Eh^2(X_k).$$

Note that this is the critical point of this proof; it is here where the antisymmetry of h comes into the picture. Hence,

$$\limsup_{n \rightarrow \infty} P \left(\max_{1 \leq k \leq n} \left| Z_k^{(1)} - \sum_{i=1}^k (k-2i+1) \hat{h}(X_i) \right| \geq Kn \right) \leq \frac{\text{constant}}{K^2}$$

for each $K > 0$.

This clearly implies the first statement of Lemma 2.1.6. ■

Now we are in a position to study the asymptotic behaviour of $\hat{U}_k$. The following results are due to M. Csörgő and L. Horváth (1988a).

Theorem 2.1.8. (Confer Theorem 4.1. in M. Csörgő and L. Horváth (1988a))

Let h be an antisymmetric kernel, assume that H_0 holds and that (2.1.1.), (2.1.2.) are satisfied. Then we can define a sequence of Brownian bridges $\{B_n(t) : 0 \leq t \leq 1\}_{n \in \mathbb{N}}$ such that as $n \rightarrow \infty$ we have

$$(2.1.5.) \left\{ \begin{array}{l} \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| = o_p(1) \\ \text{and for each fixed } n \in \mathbb{N}, EB_n(t) = 0, EB_n(s)B_n(t) = s \wedge t - st, \\ \text{where } s \wedge t \text{ denotes the minimum of } s \text{ and } t. \end{array} \right.$$

Theorem 2.1.9. (Confer Theorem 4.2. in M. Csörgő and L. Horváth (1988a))

Assume the conditions of Theorem 2.1.8. hold. Suppose in addition that

$$(2.1.6.) \quad E|h(X_1, X_2)|^v < \infty \quad \text{for some } v > 2.$$

Then we can define a sequence of Brownian bridges $\{B_n(t) : 0 \leq t \leq 1\}_{n \in \mathbb{N}}$ such that (2.1.5.) holds and as $n \rightarrow \infty$

$$\frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq \frac{n}{n+1}} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| / (t(1-t))^{1/2} = O_p(1).$$

The importance of Theorem 2.1.9. can be explained from two different viewpoints. In the first place, the statement of Theorem 2.1.9. tells us what is going to happen, if we assume the existence of higher moments. Secondly, one can see that Theorem 2.1.9. gives an improved version of Theorem 2.1.8. in the sense that this weighted sup-norm result highlights the tail end behaviour of $\hat{U}_{[(n+1)t]}$. The weighted version is clearly more sensitive on the tails, i.e. near 0 and 1. In order to investigate what will happen if we choose weight functions that converge more slowly to zero than $\sqrt{t(1-t)}$, we introduce the following class of functions:

$$\begin{aligned} Q^* = \{q : (0,1) \rightarrow (0,\infty) \mid & q \text{ monotone nondecreasing near } 0, \\ & q \text{ monotone nonincreasing near } 1, \\ & \inf_{\delta \leq t \leq 1-\delta} q(t) > 0 \text{ for all } \delta \in (0,1/2)\}. \end{aligned}$$

For $q \in Q^*$ we define the so-called Chibisov-O'Reilly integral $I(q, c)$, which also appears in the characterization of upper class functions for a Wiener process (confer, e.g., M. Csörgö et al. (1986)):

$$I(q, c) = \int_0^1 \frac{1}{t(1-t)} \exp\left\{-\frac{cq^2(t)}{t(1-t)}\right\} dt \quad \text{for } q \in Q^* \text{ and } c > 0.$$

Let $B = \{B(t) : 0 \leq t \leq 1\}$ denote a Brownian bridge. Then we have

Corollary 2.1.10. (Confer Corollary 4.1. in M. Csörgö and L. Horváth (1988a))

Suppose $q \in Q^*$ and the assumptions of Theorem 2.1.9. are satisfied.

$$(a) \quad \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} + B_n(t) \right| / q(t) = o_p(1) \quad \text{as } n \rightarrow \infty$$

if and only if $I(q, c) < \infty$ for each $c > 0$.

$$(b) \quad \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |\hat{U}_{[(n+1)t]}| / q(t) \xrightarrow{D} \sup_{0 < t < 1} |B(t)| / q(t) \quad \text{as } n \rightarrow \infty$$

if and only if $I(q, c) < \infty$ for some $c > 0$.

We are going to give a detailed proof of Theorem 2.1.8. using the ideas and methods outlined by M. Csörgö and L. Horváth (1988a) (confer there in particular the proof of Theorem 4.1.) because the proofs of Corollary 2.2.2, Theorem 2.2.3. and Corollary 2.2.4. (the proofs of the latter two indirectly through Theorem 2.1.8.) will be based on the methodology explained herein. In order to avoid technical difficulties we may assume here and in the future without loss of generality (confer A. De Acosta (1982)) that the underlying probability space is rich enough to accommodate all processes in question.

Proof of Theorem 2.1.8.

The proof will be divided into two parts. First, we reduce the original problem to an easier one via Corollary 2.1.7. and then we make use of Donsker's Theorem to draw the final conclusion.

$$\begin{aligned}
 & \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| \\
 & \leq \sup_{0 < t < 1} \frac{n^{-3/2}}{\sigma} \left| \hat{U}_{[(n+1)t]} - \left\{ n \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - [(n+1)t] \sum_{i=1}^n \hat{h}(X_i) \right\} \right| \\
 & \quad + \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} \left\{ n \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - [(n+1)t] \sum_{i=1}^n \hat{h}(X_i) \right\} - B_n(t) \right| \\
 & = \max_{1 \leq k \leq n} \frac{n^{-3/2}}{\sigma} \left| \hat{U}_k - \left\{ n \sum_{i=1}^k \hat{h}(X_i) - k \sum_{i=1}^n \hat{h}(X_i) \right\} \right| \\
 & \quad + \sup_{0 < t < 1} \left| \frac{n^{-1/2}}{\sigma} \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - \frac{[(n+1)t]/n}{\sqrt{n} \sigma} \sum_{i=1}^n \hat{h}(X_i) - B_n(t) \right|,
 \end{aligned}$$

whence Corollary 2.1.7. implies

$$(2.1.7.) \left\{ \begin{aligned} & \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| \\ & = o_p(1) + \sup_{0 < t < 1} \left| \frac{n^{-1/2}}{\sigma} \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - \frac{[(n+1)t]/n}{\sqrt{n} \sigma} \sum_{i=1}^n \hat{h}(X_i) - B_n(t) \right|. \end{aligned} \right.$$

Next we introduce the notion of Donskerized partial sums. Let $Y_1, Y_2, \dots$ be independent identically distributed random variables with mean 0 and variance $\sigma^2 > 0$. Then we define the Donskerized partial sum $S_n(t)$ as

$$S_n(t) = \frac{n^{-1/2}}{\sigma} \{S_{[nt]} + (nt - [nt])Y_{[nt]+1}\} \quad \text{for } 0 \leq t \leq 1,$$

where $S_k = Y_1 + \dots + Y_k$ denotes the k^{th} partial sum of the Y_i 's.

Now let $Y_i = \hat{h}(X_i)$, then under the null hypothesis of no change, the Y_i 's are independent identically distributed random variables, properly standardized by assumption so that we obtain

$$\begin{aligned}
 & \sup_{0 < t < 1} \left| \frac{n^{-1/2}}{\sigma} \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - \frac{[(n+1)t]/n}{\sqrt{n} \sigma} \sum_{i=1}^n \hat{h}(X_i) - B_n(t) \right| \\
 & \leq \sup_{0 < t < 1} \left| \frac{n^{-1/2}}{\sigma} \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - \frac{n^{-1/2}}{\sigma} \left\{ \sum_{i=1}^{[nt]} \hat{h}(X_i) + (nt - [nt]) \hat{h}(X_{[nt]+1}) \right\} \right| \\
 & \quad + \sup_{0 < t < 1} \left| \left\{ S_n(t) - \frac{[(n+1)t]}{n} S_n(1) \right\} - B_n(t) \right| \\
 & = \sup_{0 < t < 1} \left| \frac{n^{-1/2}}{\sigma} \sum_{i=[nt]+1}^{[(n+1)t]} \hat{h}(X_i) - \frac{nt - [nt]}{\sqrt{n} \sigma} \hat{h}(X_{[nt]+1}) \right| \\
 & \quad + \sup_{0 < t < 1} \left| \left\{ S_n(t) - \frac{[(n+1)t]}{n} S_n(1) \right\} - B_n(t) \right|.
 \end{aligned}$$

Hence we have

$$(2.1.8.) \quad \left\{ \begin{aligned} & \sup_{0 < t < 1} \left| \frac{n^{-1/2}}{\sigma} \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - \frac{[(n+1)t]/n}{\sqrt{n} \sigma} \sum_{i=1}^n \hat{h}(X_i) - B_n(t) \right| \\ & = o_p(1) + \sup_{0 < t < 1} \left| \left\{ S_n(t) - \frac{[(n+1)t]}{n} S_n(1) \right\} - B_n(t) \right|. \end{aligned} \right.$$

Now (2.1.7.) and (2.1.8.) together with Donsker's Theorem (confer Theorem 1.1.23. earlier on) imply Theorem 2.1.8. ■

Proofs of Theorem 2.1.9. and of Corollary 2.1.10. are given in M. Csörgő and L. Horváth (1988a). We will close this subsection with the discussion of an example.

Example 2.1.11.

Let $X_1, X_2, \dots, X_n$ be independent random variables coming from continuous distributions. Assume that H_0 holds. Define $h(x, y) = \text{sgn}(x - y)$, where $\text{sgn}(x) = 1, 0, -1$ if $x > 0, = 0, < 0$, respectively.

Then

$$\hat{U}_k = Z_k = \sum_{i=1}^k \sum_{j=k+1}^n \text{sgn}(X_i - X_j) \quad \text{for } 1 \leq k \leq n-1.$$

Let $R_1, R_2, \dots, R_k$ denote the ranks of the k observations $X_1, X_2, \dots, X_k$ in the complete sample of n observations. Since we require the observations to come from a continuous distribution, there will be no ties among the ranks. Thus $R_1, R_2, \dots, R_k$ are uniquely defined. In this situation A.N. Pettitt (1979) claims

$$(2.1.9.) \quad \hat{U}_k = 2W_k - k(n+1),$$

where $W_k = \sum_{j=1}^k R_j$ is the Wilcoxon-(Mann-Whitney) rank statistic which was mentioned in Subsection 2.1.0. to be nonparametric distribution-free over the broad class of continuous distributions (confer R.H. Randles and D.A. Wolfe (1979), Corollary 2.3.6.). We prove (2.1.9.) by induction over k .

$k = 1$: $\hat{U}_1 = \sum_{j=2}^n \text{sgn}(X_1 - X_j)$ by definition of $\hat{U}_k$. Given the sample $X_1, X_2, \dots, X_n$ we can arrange the $X_1, \dots, X_n$ in order of their magnitude. This leads to the situation represented in the following table.

Observations	X_j	X_1	X_j
Ranks	(R_1-1) -random variables with rank less than R_1	R_1	$(n-R_1)$ -random variables with rank greater than R_1
$\text{sgn}(X_1 - X_j)$	+1	0	-1

Table 1.

Thus we easily see that $\sum_{j=2}^n \text{sgn}(X_1 - X_j) = (+1)(R_1-1) + (-1)(n-R_1) = 2R_1 - (n+1)$ and hence $\hat{U}_1 = 2W_1 - (n+1)$ as claimed.

$k = 2$: Again by definition of $\hat{U}_k$, $\hat{U}_2 = \sum_{j=3}^n \text{sgn}(X_1 - X_j) + \sum_{j=3}^n \text{sgn}(X_2 - X_j)$. Here we must consider two cases, each of which can be treated similarly to the case $k = 1$.

1st case: $X_1 < X_2$. Ordering the sample $X_1, X_2, \dots, X_n$ results in

Observations	X_j	X_1	X_2, X_j
Ranks	(R_1-1) -random variables with rank less than R_1	R_1	$(n-1-R_1)$ -random variables to be counted to 1 st of above $\sum$, discard X_2
$\text{sgn}(X_1 - X_j)$	+1	0	-1

Table 2.

Hence the first of the above sums, $\sum_{j=3}^n \text{sgn}(X_1 - X_j)$, contributes $(+1)(R_1-1) + (-1)(n-1-R_1) = 2R_1 - n$ towards $\hat{U}_2$. For the second sum,

$\sum_{j=3}^n \text{sgn}(x_2 - x_j)$, we obtain

Observations	$x_1 \ x_j$	x_2	x_j
Ranks	(R_2-1-1) -random variables with rank less than R_2	R_2	$(n-R_2)$ -random variables with rank greater than R_2
$\text{sgn}(x_2 - x_j)$	+1	0	-1

Table 3.

This diagram is valid since we do not wish to include x_1 in $\sum_{j=3}^n \text{sgn}(x_2 - x_j)$. Thus the amount the second sum contributes to $\hat{U}_2$ is $(+1)(R_2-2) + (-1)(n-R_2) = 2R_2 - n - 2$. This finally proves that

$$\hat{U}_2 = (2R_1 - n) + (2R_2 - n - 2) = 2W_2 - 2(n+1),$$

whence (2.1.9.) is proven to be true in the first case. The second case,

$x_2 < x_1$, can be dealt with in a similar fashion. Assume now

$$\hat{U}_{k-1} = 2 \sum_{j=1}^{k-1} R_j = (k-1)(n+1) \quad \text{for some } k \geq 2.$$

Then we have

$$\begin{aligned} \hat{U}_k &= \sum_{i=1}^k \sum_{j=k+1}^n \text{sgn}(x_i - x_j) \\ &= \sum_{i=1}^{k-1} \sum_{j=k+1}^n \text{sgn}(x_i - x_j) + \sum_{j=k+1}^n \text{sgn}(x_k - x_j) \\ &= \sum_{i=1}^{k-1} \sum_{j=k}^n \text{sgn}(x_i - x_j) - \sum_{i=1}^{k-1} \text{sgn}(x_i - x_k) + \sum_{j=k+1}^n \text{sgn}(x_k - x_j) \\ &= \hat{U}_{k-1} + \sum_{j=1}^n \text{sgn}(x_k - x_j) \quad \text{and by induction assumption} \end{aligned}$$

$$\begin{aligned}
 &= 2 \sum_{j=1}^{k-1} R_j - (k-1)(n+1) + \sum_{j=1}^n \operatorname{sgn}(X_k - X_j) \\
 &= 2W_k - k(n+1) + \left\{ \sum_{j=1}^n \operatorname{sgn}(X_k - X_j) - [2R_k - (n+1)] \right\},
 \end{aligned}$$

which will yield (2.1.9.) provided that we can show that $\sum_{j=1}^n \operatorname{sgn}(X_k - X_j) = 2R_k - (n+1)$. The latter statement can be verified in the same vein the case $k=1$ was. This completes the proof of (2.1.9.).

Under H_0 we have seen that $\hat{U}_k = \sum_{1 \leq i < k < j \leq n} \operatorname{sgn}(X_i - X_j)$ is non-parametric distribution free; $E(\operatorname{sgn}(X_1 - X_2)) = 0$ and σ^2 can be computed to be $1/12 > 0$ in this special case. Thus the methodology of this section applies to $h(x,y) = \operatorname{sgn}(x-y)$. In particular, Theorem 2.1.8. implies that $(12n^3)^{-1/2} \hat{U}_{[(n+1)t]}$ converges weakly in $D[0,1]$ to a Brownian bridge. This result was suggested by A.N. Pettitt (1979) who states representation (2.1.9.) and the covariance structure of the Wilcoxon rank sum statistic W_k . Pettitt observes that the covariance of W_k converges to that of a Brownian bridge. This amounts to saying that the finite dimensional distributions of $\hat{U}_k$ converge to those of a Brownian bridge which without a proof of tightness does not imply the result of Theorem 2.1.8. (For the concept of tightness and the theorem implicitly quoted here, we refer to P. Billingsley (1968), p.37, Theorems 8.1. and 15.1.) In the following we give a justification for Pettitt's claim that

$$(2.1.10.) \quad \operatorname{Cov}(W_k, W_m) = \frac{k(n+1)(n-m)}{12} \quad \text{for } k \leq m.$$

It is well known from the theory of rank statistics (confer, e.g., R.H. Randles and D.A. Wolfe (1979)), Corollary 2.3.5.) that

$$(2.1.11.) \quad \begin{cases} ER_j = \frac{(n+1)}{2}, & \text{Var } R_j = \frac{(n+1)(n-1)}{12} \quad \text{and} \\ \text{Cov}(R_i, R_j) = -\frac{(n+1)}{12}, & \text{for } i \neq j, \end{cases}$$

provided that the sample comes from a continuous distribution. If that is not the case, a correction factor is needed for the variance. From (2.1.11.) we have immediately

$$EW_k = \frac{k(n+1)}{2} \quad \text{and} \quad \text{Var } W_k = \frac{k(n-k)(n+1)}{12}.$$

Now we compute for $k \leq m$

$$\begin{aligned} \text{Cov}(W_k, W_m) &= EW_k W_m - EW_k EW_m \\ &= E\left[\sum_{j=1}^k R_j \sum_{i=1}^m R_i\right] - \frac{k(n+1)}{2} \cdot \frac{m(n+1)}{2} \\ &= E\left[\sum_{j=1}^k R_j^2\right] + \sum_{j=1}^k \sum_{i=k+1}^m ER_i R_j - \frac{km(n+1)^2}{4} \\ &= \text{Var } W_k + (EW_k)^2 + k(m-k) \{ \text{Cov}(R_1, R_2) + ER_1 ER_2 \} - \frac{km(n+1)^2}{4} \\ &= \frac{k(n-k)(n+1)}{12} + \frac{k^2(n+1)^2}{4} + k(m-k) \left\{ -\frac{(n+1)}{12} + \frac{(n+1)^2}{4} \right\} - \frac{km(n+1)^2}{4} \\ &= \frac{k(n+1)}{12} \{ (n-k) + 3k(n+1) - (m-k) + 3(m-k)(n+1) - 3m(n+1) \} \\ &= \frac{k(n+1)(n-m)}{12} \quad \text{as stated in (2.1.10.).} \end{aligned}$$

Following Pettitt's notation, define $x = \frac{k}{n}$ and $y_n(x) = n^{-1} \{3/(n+1)\}^{1/2} \hat{U}_k$. With this notation we finally show that indeed as Pettitt states, the covariance structure of y_n converges to that of a Brownian bridge.

$$\begin{aligned}
y_n(x) &= \frac{\sqrt{3}}{n(n+1)^{1/2}} \hat{U}_k \\
&= \frac{\sqrt{3}}{n(n+1)^{1/2}} \{2W_k - k(n+1)\} \\
&= \frac{2\sqrt{3}}{n(n+1)^{1/2}} \{W_k - EW_k\} \\
&= \sqrt{\frac{k(n-k)}{n^2}} \cdot \frac{W_k - EW_k}{\sqrt{\text{Var } W_k}} \\
&= \sqrt{x(1-x)} \frac{W_k - EW_k}{\sqrt{\text{Var } W_k}}
\end{aligned}$$

Now let $k \leq m$ and define $\hat{x} = \frac{m}{n}$. Then we obtain

$$\begin{aligned}
\text{Cov}(y_n(x), y_n(\hat{x})) &= E y_n(x) y_n(\hat{x}) - E y_n(x) E y_n(\hat{x}) \\
&= \sqrt{\frac{x(1-x)}{\text{Var } W_k}} \sqrt{\frac{\hat{x}(1-\hat{x})}{\text{Var } W_m}} \text{Cov}(W_k, W_m) \\
&= \frac{k(n-m)}{n^2} \\
&= x(1-\hat{x}).
\end{aligned}$$

Hence the covariance function of y_n converges as $n \rightarrow \infty$ to that of a Brownian bridge.

2.1.2. Asymptotic Behaviour of U_k in sup-Norm and in Weighted sup-Norms Under the Null Hypothesis

In the present subsection we will study results corresponding to those discussed in Subsection 2.1.1. when replacing antisymmetric kernels by symmetric ones. Let h now denote a symmetric function in two real variables and recall the following notations.

$$\begin{aligned} Z_k &= \sum_{1 \leq i \leq k} \sum_{k+1 \leq j \leq n} h(X_i, X_j) && \text{for } 1 \leq k \leq n-1 \\ \theta &= E h(X_1, X_2) \\ \tilde{h}(t) &= E \{h(t, X_1) - \theta\} && \text{for } t \in \mathbb{R} \\ U_k &= Z_k - k(n-k)\theta && \text{for } 1 \leq k \leq n-1. \end{aligned}$$

As mentioned earlier, representation (2.1.3.) of U_k ,

$$U_k = U_n^{(3)} - \{U_k^{(1)} + U_k^{(2)}\}, \quad \text{where}$$

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

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

$$U_n^{(3)} = \sum_{1 \leq i < j \leq n} h(X_i, X_j) - \binom{n}{2} \theta$$

as a linear combination of nondegenerate ($\sigma^2 > 0$!) U-statistics will be the starting point for the considerations here. In order to prove the following theorems, we will need a new lemma in lieu of Lemma 2.1.6. Lemma 2.1.12., which is a consequence of Theorem 1 of P. Hall (1979), will produce the desired results.

Lemma 2.1.12.

Assume H_0 holds. Let $\theta, h, \tilde{h}$, and U_k be as defined above.

Suppose

$$(2.1.1.) \quad E h^2(X_1, X_2) < \infty \quad \text{and}$$

$$(2.1.2.) \quad E \tilde{h}^2(X_1) = \sigma^2 > 0.$$

Then we have as $n \rightarrow \infty$

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

$$\max_{1 \leq k \leq n} \left| U_k^{(2)} - (n-k) \sum_{i=k+1}^n \tilde{h}(X_i) \right| = O_p(n) \quad \text{and}$$

$$\left| U_n^{(3)} - n \sum_{i=1}^n \tilde{h}(X_i) \right| = O_p(n),$$

where $U_k^{(1)}$, $U_k^{(2)}$, and $U_n^{(3)}$ are as in representation (2.1.3.).

Corollary 2.1.13.

Under the assumptions of Lemma 2.1.12.

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

yields as $n \rightarrow \infty$.

The proof of Corollary 2.1.13. follows along the lines of the proof of Corollary 2.1.7.

Here and in the following $\Gamma = \{\Gamma(t) : 0 \leq t \leq 1\}$ will always denote the Gaussian process specified by

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

where $W = \{W(t): t \geq 0\}$ is a standard Wiener process.

The next results in order of their presentation correspond to Theorem 2.1.8., Theorem 2.1.9. and Corollary 2.1.10., respectively, when exchanging antisymmetric kernel functions for symmetric ones. All three results are again due to and proven in M. Csörgö and L. Horváth (1988a).

Theorem 2.1.14. (Confer Theorem 2.1. in M. Csörgö and L. Horváth (1988a))

Under the conditions of Lemma 2.1.12. we can define a sequence of Gaussian processes $\{\Gamma_n(t): 0 \leq t \leq 1\}_{n \in \mathbb{N}}$ such that as $n \rightarrow \infty$

$$(2.1.12.) \quad \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} U_{[(n+1)t]} - \Gamma_n(t) \right| = o_p(1),$$

where for each fixed $n \in \mathbb{N}$ $\{\Gamma_n(t): 0 \leq t \leq 1\} \stackrel{d}{=} \{\Gamma(t): 0 \leq t \leq 1\}$.

Theorem 2.1.15. (Confer Theorem 2.2. in M. Csörgö and L. Horváth (1988a))

Assume the conditions of Lemma 2.1.12. are satisfied. Suppose that

$$(2.1.6.) \quad E|h(X_1, X_2)|^v < \infty \quad \text{for some } v > 2.$$

Then we can define a sequence of Gaussian processes $\{\Gamma_n(t): 0 \leq t \leq 1\}_{n \in \mathbb{N}}$ such that (2.1.12.) holds and

$$\frac{1}{n+1} \sup_{\frac{n}{n+1} \leq t \leq \frac{n}{n+1}} \left| \frac{n^{-3/2}}{\sigma} U_{[(n+1)t]} - \Gamma_n(t) \right| / \sqrt{t(1-t)} = o_p(1)$$

as $n \rightarrow \infty$.

Corollary 2.1.16. (Confer Corollary 2.1. in M. Csörgö and L. Horváth (1988a)).

Let $q \in Q^*$. Given the situation of Theorem 2.1.15. we obtain:

$$(a) \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} U_{[(n+1)t]} - \Gamma_n(t) \right| / q(t) = o_p(1) \text{ as } n \rightarrow \infty$$

if and only if $I(q, c) < \infty$ for all $c > 0$.

$$(b) \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |U_{[(n+1)t]}| / q(t) \xrightarrow{D} \sup_{0 < t < 1} \frac{|\Gamma(t)|}{q(t)} \text{ as } n \rightarrow \infty$$

if and only if $I(q, c) < \infty$ for some $c > 0$.

We wish to have weighted sup-norm results because the statistical tests based on those are more sensitive on the tails. A typical choice for a weight function q whose Chibisov-O'Reilly integral is finite is

$$q(t) = \left\{ t(1-t) \log \log \frac{1}{t(1-t)} \right\}^{1/2} \text{ for } 0 < t < 1.$$

Since the variance of $\Gamma(t)$ equals $t(1-t)$, another sensible choice for q is $\sqrt{t(1-t)}$. However, the Chibisov-O'Reilly integral $I(q, c)$ is for this particular choice of q infinite for every $c > 0$, whence Corollary 2.1.16. does not apply in this case.

Although the variance of Γ coincides with that of a Brownian bridge, Γ itself is not a Brownian bridge as can be seen by checking that

$$\text{Cov}(\Gamma(s), \Gamma(t)) = (1-2s)(1-2t)(s \wedge t) + (3-2s-2t)st.$$

2.1.3. Summary

In this section we studied various weak convergence theorems for functionals of $\hat{U}_k$ and U_k . In all of the theorems and corollaries, we have made the assumption that the variance of the functions $\hat{h}$ and $\tilde{h}$ is positive. This means that in Section 2.1. only the so-called non-degenerate case was treated. The following provides an example of a nonconstant symmetric kernel function in the degenerate case.

Example 2.1.17.

Suppose $X_1, X_2, \dots, X_n$ are independent observations on a random sample with mean zero and some finite variance. Assume that the null hypothesis of no change is true. Let

$$h(x, y) = xy \text{ for } x, y \in \mathbb{R},$$

thus h is a symmetric kernel. Then by the independence of X_1 and X_2 we have

$$\theta = Eh(X_1, X_2) = E(X_1 \cdot X_2) = EX_1 EX_2 = 0,$$

and by the assumption of finite variances of the random variables X_1 and X_2 ,

$$Eh^2(X_1, X_2) = E(X_1 X_2)^2 = E(X_1^2 X_2^2) = EX_1^2 EX_2^2 < \infty.$$

Hence condition (2.1.1.) is satisfied. However,

$$\sigma^2 = \text{Var } \tilde{h}(X_1) = E\tilde{h}^2(X_1) = E(E_{X_2}(X_1 X_2))^2 = E(X_1 EX_2)^2 = 0$$

and moreover $\tilde{h}(X_1) = E_{X_2}(h(X_1, X_2)) = E_{X_2}(X_1 X_2) = X_1 EX_2 = 0$, where E_{X_2} denotes the expected value taken with respect to X_2 only.

Under certain conditions it is possible to establish a weak convergence result for $U_k = \sum_{i=1}^k \sum_{j=k+1}^n X_i X_j$. This will be done in Section 3.2. of Chapter 3.

Reviewing the main results of this section, one can see that two essentially different Gaussian processes (sequences of such processes) have been established as in distribution limits under the null hypothesis of no change:

$$B = \{B(t) : 0 \leq t \leq 1\}$$

$$(B_n = \{B_n(t) : 0 \leq t \leq 1\}, \text{ where for each } n \in \mathbb{N} : B_n \xrightarrow{D} B),$$

where B is a Brownian bridge, i.e. a tied down Brownian motion on $[0,1]$, in the antisymmetric case and

$$\Gamma = \{\Gamma(t) : 0 \leq t \leq 1\}$$

$$(\Gamma_n = \{\Gamma_n(t) : 0 \leq t \leq 1\}, \text{ where for each } n \in \mathbb{N} : \Gamma_n \xrightarrow{D} \Gamma),$$

where $\Gamma(t) = (1-t)W(t) + tW(1) - tW(t)$ for $t \in [0,1]$

with W denoting a standard Brownian motion

in the symmetric case.

The shape of the limiting processes can be read off the auxiliary results in Corollary 2.1.7. (for B) and in Corollary 2.1.13. (for Γ).

Comparison of these two corollaries yields

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

i.e. the dominating partial sum part of $\hat{U}_k$ (antisymmetric kernel)

plus $2k \sum_{i=k+1}^n \tilde{h}(X_i)$ equals the dominating partial sum part of U_k

(symmetric kernel function). The same relationship exists between the limiting processes B and Γ because

$$\begin{aligned} B(t) &= 2t(W(t) - W(1)) \\ &= (W(t) - tW(1)) - 2tW(t) + 2tW(1) \\ &= (1-t)W(t) + tW(1) - tW(t) \\ &= \Gamma(t). \end{aligned}$$

The form of the above limiting processes can be motivated as follows:

$(n-k)\sum_{i=1}^k \tilde{h}(X_i)$ may be associated with $(1-t)W(t)$ through considering $\frac{k}{n} = t$ and proper normalization; similarly

$k\sum_{i=1}^n \tilde{h}(X_i)$ may be associated with $tW(1)$,

$k\sum_{i=1}^k \hat{h}(X_i)$ may be associated with $tW(t)$,

$n\sum_{i=1}^k \tilde{h}(X_i)$ may be associated with $W(t)$, and

$2k\sum_{i=k+1}^n \tilde{h}(X_i)$ may be associated with $-2t(W(t)-W(1))$.

This explains the main idea behind the proofs of the previous section.

The functionals of $\hat{U}_k$ and U_k can now be used to derive test statistics for testing the fixed sample size changepoint problem

$H_0 : X_1, X_2, \dots, X_n$ are identically distributed
versus

$H_A : \text{there exists exactly one changepoint.}$

Here we only give an example:

When working with an antisymmetric kernel function h , Corollary

2.1.10. (b) suggests that a possible class of test statistics contains those statistics defined via

$$T_n := \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} \frac{|\hat{U}_{[(n+1)t]}|}{q(t)},$$

where the weight function $q \in Q^*$ has to be chosen so that $I(q, c)$ is finite for at least one $c > 0$. Using T_n , the null hypothesis of no change would have to be rejected if T_n produces a value which is "too large". Here "too large" will have to be measured with respect to the critical value generated by

$$\sup_{0 < t < 1} \frac{|B(t)|}{q(t)}$$

for given significance level $0 < \alpha < 1$.

The problem of how to actually compute $c = c(\alpha)$ such that

$$P\left\{ \sup_{0 < t < 1} \frac{|B(t)|}{q(t)} \geq c \right\} = \alpha$$

will be discussed in Chapter 3 for certain functionals of B and Γ .

We also mention that all the test statistics involved here produce under the null hypothesis of no change in distribution limits that do not depend on the null distribution in the given sample. We will make use of this property in Section 2.3. when studying power functions of the statistics examined in this section.

Section 2.2. Cramér-von Mises Type Tests

2.2.0. Auxiliary results

As in the previous section, we will discuss the behaviour of certain stochastic processes on which later statistics will be based, separately for antisymmetric and for symmetric kernel functions. The entire section is again devoted to results under the null hypothesis.

The following lemmata are needed for the proof of the central limit theorem in Subsection 2.2.3. This central limit theorem holds for antisymmetric as well as for symmetric kernel functions. Therefore, let here h be a function in two real variables such that either

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

or

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

holds. Let $\theta = Eh(X_1, X_2)$ with $\theta = 0$ in case h is antisymmetric.

Put

$$\tilde{h}(t) = E\{h(t, X_1) - \theta\} \quad \text{and assume}$$

$$(2.1.1.) \quad Eh^2(X_1, X_2) < \infty,$$

$$(2.1.2.) \quad E\tilde{h}^2(X_1) = \sigma^2 > 0.$$

Define

$$S_n^{(1)}(t) = \sigma^{-1} \sum_{i=1}^{[t]} \tilde{h}(X_i) \quad \text{for } 1 \leq t \leq [n/2] \text{ and}$$

$$S_n^{(2)}(t) = \sigma^{-1} \sum_{i=n-[t]+1}^n \tilde{h}(X_i) \quad \text{for } 1 \leq t \leq n-[n/2].$$

Then we have the following lemma.

Lemma 2.2.0.

Assume that H_0 holds and that (2.1.1.), (2.1.2.) as well as

$$(2.1.6.) \quad E|h(X_1, X_2)|^v < \infty \quad \text{for some } v > 2$$

are satisfied. Then, as $n \rightarrow \infty$

$$(2.2.0.) \quad \left| \int_{k_n}^{n-1} \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \cdot \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)} dy \right| = O_p(\log \log n)$$

where $k_n := \frac{n}{2} + \log n$.

Proof of Lemma 2.2.0.

$$\begin{aligned} & \left| \int_{k_n}^{n-1} \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \cdot \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)} dy \right| \\ &= \left| \frac{1}{n} \int_{k_n/n}^{1-\frac{1}{n}} \frac{\sum_{i=1}^{[nt]} \tilde{h}(X_i)}{t} \cdot \frac{\sum_{i=[nt]+1}^n \tilde{h}(X_i)}{(1-t)} dt \right|, \end{aligned}$$

when using the substitution $y = nt$; now let

$$(2.2.1.) \quad q(t) := \left\{ t(1-t) \log \log \frac{1}{t(1-t)} \right\}^{\frac{1}{2}}, \quad 0 < t < 1,$$

$$\begin{aligned} & \leq \left\{ n^{-\frac{1}{2}} \sup_{\frac{1}{2} \leq t \leq 1} \left| \sum_{i=1}^{[nt]} \tilde{h}(X_i) \right| \right\} \cdot \left\{ n^{-\frac{1}{2}} \sup_{\frac{1}{2} \leq t \leq 1} \frac{\left| \sum_{i=[nt]+1}^n \tilde{h}(X_i) \right|}{q(t)} \right\} \\ & \cdot \left\{ \sqrt{\log \log 2n} \int_0^1 \frac{dt}{\sqrt{t(1-t)}} \right\}. \end{aligned}$$

Now let us examine each of the factors in the above product separately.

By Kolmogorov's inequality (compare, e.g., G.R. Shorack and J.A. Wellner, (1986), p.843) we have for the first term:

$$P \left\{ \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \tilde{h}(X_i) \right| \geq \lambda \right\} \leq \frac{nc^2}{\lambda^2} \quad \text{for all } \lambda > 0$$

and consequently for $\lambda = Cn^{\frac{1}{2}}$:

$$\lim_{C \rightarrow \infty} \lim_{n \rightarrow \infty} P \left\{ n^{-1/2} \max_{1 \leq k \leq n} \left| \sum_{i=1}^k \tilde{h}(X_i) \right| \geq C \right\} = 0,$$

whence the first of the above factors is now known to be bounded in probability. Considering the last term in that product we simply note that the integral equals $\beta(1/2, 1/2)$, i.e. π , and is thus a constant. For the remaining middle term we obtain:

$$n^{-1/2} \sup_{1/2 \leq t \leq 1 - \frac{1}{n}} \frac{\left| \sum_{i=[nt]+1}^n \tilde{h}(X_i) \right|}{q(t)} \stackrel{D}{=} n^{-1/2} \sup_{1/2 \leq t \leq 1 - \frac{1}{n}} \frac{\left| \sum_{i=1}^{n-[nt]} \tilde{h}(X_i) \right|}{q(t)},$$

since $n - [nt]$ is an integer, it equals either $[n(1-t)]$ or $[n(1-t)+1]$. Assume without loss of generality now that $n - [nt] = [n(1-t)]$, then we arrive at (put $x = 1-t$)

$$n^{-1/2} \sup_{1/2 \leq t \leq 1 - \frac{1}{n}} \frac{\left| \sum_{i=[nt]+1}^n \tilde{h}(X_i) \right|}{q(t)} \stackrel{D}{=} n^{-1/2} \sup_{1/2 \leq x \leq 1} \frac{\left| \sum_{i=1}^{[nx]} \tilde{h}(X_i) \right|}{q(x)}.$$

Therefore an application of Theorem 1.1.25. completes the proof of Lemma 2.2.0.

Lemma 2.2.1.

Let $Y_1, Y_2, \dots, Y_n$ be independent identically distributed random variables with $EY_1 = 0$, $EY_1^2 = 1$, and $E|Y_1|^\nu < \infty$ for some $\nu > 2$. Define $S_0 = 0$ and $S(t) = Y_1 + \dots + Y_{[t]}$ for $t > 0$. Then we have

$$(4 \log n)^{-1/2} \left\{ \int_1^n \frac{S^2(t)}{t^2} dt - \log n \right\} \xrightarrow{D} N(0,1) \text{ as } n \rightarrow \infty,$$

where $N(0,1)$ denotes a standard normal variable.

We note that Lemma 2.2.1. implies $\int_1^n \frac{S^2(t)}{t^2} dt \xrightarrow{P} \infty$ as $n \rightarrow \infty$.

Proof of Lemma 2.2.1.

Let W denote a Wiener process. The following two statements together immediately yield Lemma 2.2.1.:

$$(2.2.2.) \quad \left| \int_1^n \frac{S^2(t)}{t^2} dt - \int_1^n \frac{W^2(t)}{t^2} dt \right| \stackrel{\text{a.s.}}{=} O(1) \quad \text{as } n \rightarrow \infty,$$

$$(2.2.3.) \quad (4 \log n)^{-1/2} \left\{ \int_1^n \frac{W^2(t)}{t^2} dt - \log n \right\} \xrightarrow{D} N(0,1) \quad \text{as } n \rightarrow \infty.$$

Proof of (2.2.2.)

Towards this end we first use the equality

$$a^2 - b^2 = (a-b)^2 + 2b(a-b), \quad a, b \in \mathbb{R}.$$

Then we obtain for any $0 < \tau < 1/2 \leq 1/3$:

$$\begin{aligned} & \left| \int_1^n \frac{S^2(t)}{t^2} dt - \int_1^n \frac{W^2(t)}{t^2} dt \right| \\ & \leq \int_1^n \frac{(S(t)-W(t))^2}{t^2} dt + 2 \int_1^n |W(t)| \cdot \frac{|S(t)-W(t)|}{t^2} dt \\ & \leq \left\{ \sup_{1 \leq t \leq n} \frac{|S(t)-W(t)|}{t^{1-\tau}} \right\}^2 \cdot \int_1^n t^{-(1+2\tau)} dt + \sup_{1 \leq t \leq n} \frac{|S(t)-W(t)|}{t^{1-\tau}} \\ & \quad \cdot \int_1^n |W(t)| t^{-(\frac{3}{2}+\tau)} dt. \end{aligned}$$

Next, we use the following improvement of the original Komlós-Major-Tusnády Theorem:

$$|S(t) - W(t)| \stackrel{\text{a.s.}}{=} o(t^{1/2}) \quad \text{as } t \rightarrow \infty$$

(compare Theorem 2.6.3. in M. Csörgő and P. Révész (1981)) in order to obtain

$$\sup_{1 \leq t \leq n} \frac{|S(t) - W(t)|}{t^{\frac{1}{2} - \tau}} \stackrel{\text{a.s.}}{\leq} \sup_{1 \leq t \leq n} O\left(\frac{t^{1/\nu}}{t^{\frac{1}{2} - \tau}}\right) = O(1).$$

Hence it now suffices to show that

$$(2.2.4.) \quad \int_1^n |W(t)| \cdot t^{-(\frac{3}{2} + \tau)} dt \stackrel{\text{a.s.}}{=} O(1) \quad \text{as } n \rightarrow \infty.$$

Proof of (2.2.4.)

By the law of iterated logarithm for the Wiener process (confer Theorem 1.1.24. earlier on) we have:

$$\begin{aligned} & \int_1^n |W(t)| \cdot t^{-(\frac{3}{2} + \tau)} dt \\ & \leq \sup_{1 \leq t \leq n} \frac{|W(t)|}{t^{1/2 + \tau/2}} \cdot \int_1^n t^{-(1 + \tau/2)} dt \\ & \stackrel{\text{a.s.}}{\leq} \sup_{1 \leq t \leq n} O\left(\frac{\sqrt{2t \log \log t}}{\sqrt{t^{1+\tau}}}\right) \cdot O(1) \\ & = O(1). \end{aligned}$$

Proof of (2.2.3.)

The key to the proof of (2.2.3.) is a transformation of the integrand into an Ornstein-Uhlenbeck process with parameter $\alpha = 2$.

(For the concept of an Ornstein-Uhlenbeck process confer Definition 1.1.9. in Chapter 1.) Taking $t = e^{2s}$ yields

$$\int_1^n \frac{W^2(t)}{t^2} dt = 2 \int_0^{\frac{1}{2} \log n} \frac{W^2(e^{2s})}{e^{4s}} e^{2s} ds = 2 \int_0^{\frac{1}{2} \log n} V^2(s) ds.$$

From Theorem 4.5. in M. Csörgö and L. Horváth (1985) (confer also Theorem 3.4. in M. Csörgö and L. Horváth (1988c)) we know that

$$(2T)^{-1/2} \left\{ \int_0^T v^2(t) dt - T \right\} \xrightarrow{D} N(0,1) \quad \text{as } T \rightarrow \infty.$$

Hence application of these results to the discrete sequence $\{T_n\}_{n \geq 1}$, where $T_n := \frac{1}{2} \log n$, completes the proof of (2.2.3.) and consequently also the proof of Lemma 2.2.1. ■

2.2.1. Weak Convergence of $\hat{U}_k$ in L_2 -Norm and in Weighted L_2 -Norms
Under the Null Hypothesis

Recall the following notations introduced in Subsection 2.1.0.:

$$h(x, y) = -h(y, x) \quad \text{for } x, y \in \mathbb{R}$$

$$\hat{h}(t) = E h(t, X_1) \quad \text{for } t \in \mathbb{R} \text{ and}$$

$$\hat{U}_k = Z_k = \sum_{1 \leq i \leq k} \sum_{k+1 \leq j \leq n} h(X_i, X_j) \quad \text{for } 1 \leq k \leq n-1.$$

Assume that

$$(2.1.1.) \quad E h^2(X_1, X_2) < \infty$$

and

$$(2.1.2.) \quad E h^2(X_1) = \sigma^2 > 0$$

hold. This is the set up for the present subsection. Under the assumptions of Theorem 2.1.8. we have

Corollary 2.2.2.

$$\frac{n-3}{\sigma^2} \int_0^1 \hat{U}_{[(n+1)t]}^2 dt \xrightarrow{D} \int_0^1 B^2(t) dt \quad \text{as } n \rightarrow \infty,$$

where $B = \{B(t) : 0 \leq t \leq 1\}$ again denotes a Brownian bridge.

Proof of Corollary 2.2.2.

Using the equality $a^2 - b^2 = (a-b)^2 + 2b(a-b)$ and the fact that the square integral of a Brownian bridge is almost surely finite, Theorem 2.1.8. immediately implies Corollary 2.2.2. via Donsker's Theorem (confer Theorem 1.1.23. in Chapter 1.).

Next, we again would like to have weighted versions of this result.

Theorem 2.2.3.

The following equivalence holds true under the assumptions of Corollary 2.1.10.

$$\frac{n^{-3}}{\sigma^2} \int_0^1 \frac{\hat{U}_{[(n+1)t]}^2}{q(t)} dt \xrightarrow{D} \int_0^1 \frac{B_n^2(t)}{q(t)} dt \text{ if and only if } \int_0^1 \frac{t(1-t)}{q(t)} dt < \infty.$$

Proof of Theorem 2.2.3. (Sufficiency)

Let $\varepsilon > 0$ be arbitrary but fixed. Denote by $\{B_n\}_{n \geq 1}$ that sequence of Brownian bridges for which Theorem 2.1.8. holds. Then we have

$$\begin{aligned} & \left| \int_{\varepsilon}^{1-\varepsilon} \frac{\{n^{-3/2} \sigma^{-1} \hat{U}_{[(n+1)t]}\}^2 - B_n^2(t)}{q(t)} dt \right| \\ & \leq \left\{ \sup_{\varepsilon \leq t \leq 1-\varepsilon} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| \right\}^2 \cdot \int_{\varepsilon}^{1-\varepsilon} \frac{dt}{q(t)} \\ & \quad + 2 \left\{ \sup_{\varepsilon \leq t \leq 1-\varepsilon} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| \right\} \cdot \int_{\varepsilon}^{1-\varepsilon} \frac{|B_n(t)|}{q(t)} dt \end{aligned}$$

because $a^2 - b^2 = (a-b)^2 + 2b(a-b)$;

$$\text{since } \infty > \int_0^1 \frac{t(1-t)}{q(t)} dt \geq \int_{\varepsilon}^{1-\varepsilon} \frac{t(1-t)}{q(t)} dt \geq \varepsilon^2 \int_{\varepsilon}^{1-\varepsilon} \frac{dt}{q(t)},$$

$$\begin{aligned} & \leq \left\{ \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| \right\}^2 \cdot O(1) \\ & \quad + 2 \left\{ \sup_{0 < t < 1} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| \right\} \cdot \sup_{0 < t < 1} |B_n(t)| \cdot O(1) \\ & = o_p(1) \text{ as } n \rightarrow \infty \text{ by (2.1.5.).} \end{aligned}$$

Similarly we obtain:

$$\left| \int_{\frac{1}{n+1}}^{\epsilon} \frac{\left\{ \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} \right\}^2 - B_n^2(t)}{q(t)} dt \right|$$

$$\begin{aligned} &\leq \left\{ \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq \epsilon} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| / \sqrt{t(1-t)} \right\}^2 \cdot \int_{\frac{1}{n+1}}^{\epsilon} \frac{t(1-t)}{q(t)} dt \\ &\quad + 2 \left\{ \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq \epsilon} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right| / \sqrt{t(1-t)} \right\} \cdot \\ &\quad \cdot \int_{\frac{1}{n+1}}^{\epsilon} |B_n(t)| \frac{\sqrt{t(1-t)}}{q(t)} dt \end{aligned}$$

and by Theorem 2.1.9.

$$\begin{aligned} &\leq O_p(1) \cdot \int_0^{\epsilon} \frac{t(1-t)}{q(t)} dt + O_p(1) \cdot \int_{\frac{1}{n+1}}^{\epsilon} \left\{ \frac{|B_n(t)|}{\sqrt{q(t)}} \cdot \frac{\sqrt{t(1-t)}}{q(t)} \right\} dt \\ &\leq O_p(1) \cdot \int_0^{\epsilon} \frac{t(1-t)}{q(t)} dt + O_p(1) \left\{ \int_{\frac{1}{n+1}}^{\epsilon} \frac{B_n^2(t)}{q(t)} dt \cdot \int_{\frac{1}{n+1}}^{\epsilon} \frac{t(1-t)}{q(t)} dt \right\}^{1/2} \end{aligned}$$

by Cauchy-Schwarz inequality;

furthermore, an application of (19.2.) in L.A. Shepp (1966) gives:

$$(2.2.5.) \quad \int_0^{\epsilon} \frac{B_n^2(t)}{q(t)} dt < \infty \quad \text{a.s.} \quad \text{if and only if} \quad \int_0^{\epsilon} \frac{t(1-t)}{q(t)} dt < \infty, \quad \text{hence}$$

$$\leq O_p(1) \cdot \int_0^{\epsilon} \frac{t(1-t)}{q(t)} dt + O_p(1) \cdot O(1) \cdot \left\{ \int_0^{\epsilon} \frac{t(1-t)}{q(t)} dt \right\}^{1/2} \xrightarrow{P} 0$$

as $\epsilon \rightarrow 0$.

In the same vein it can be shown that

$$\left| \int_{1-\varepsilon}^{\frac{n}{n+1}} \frac{n^{-3} \sigma^{-2} \hat{U}^2[(n+1)t] - B_n^2(t)}{q(t)} dt \right| \xrightarrow{P} 0 \text{ as } n \rightarrow \infty, \varepsilon \rightarrow 0.$$

In order to complete the sufficiency part of this proof we now have to show that

$$\int_0^{\frac{1}{n+1}} \frac{B^2(t)}{q(t)} dt = o_p(1) \quad \text{and} \quad \int_{\frac{n}{n+1}}^1 \frac{B^2(t)}{q(t)} dt = o_p(1) \quad \text{as } n \rightarrow \infty.$$

But these statements are immediate from (2.2.5.) which holds for any Brownian bridge and any $\varepsilon > 0$.

Proof of Theorem 2.2.3. (Necessity).

Here we assume that

$$\int_0^1 \frac{t(1-t)}{q(t)} dt = \infty.$$

Then statement (19.4.) in L.A. Shepp (1966) implies immediately:

$$\int_0^1 \frac{B^2(t)}{q(t)} dt = \infty \text{ a.s.}$$

which proves the necessity part of Theorem 2.2.3.

Theorem 2.2.3. contains as a special case the following

Corollary 2.2.4.

Under the conditions of Corollary 2.1.10. we have as $n \rightarrow \infty$

$$\frac{n^{-3}}{\sigma^2} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{\hat{U}^2[(n+1)t]}{t(1-t)} dt \xrightarrow{D} \int_0^1 \frac{B^2(t)}{t(1-t)} dt.$$

We note that Shepp's above quoted results also imply that

$$(2.2.6.) \quad \int_0^1 \frac{B^2(t)}{t(1-t)} dt < \infty \quad \text{a.s.},$$

a fact which we implicitly assumed when formulating Corollary 2.2.4.

Statement (2.2.6.) can also be seen to be true by direct calculations involving Theorem 1.4.1. in M. Csörgö and P. Révész (1981) which states that

$$\limsup_{h \downarrow 0} \frac{|B(h)|}{\sqrt{2h \log \log 1/h}} = 1 \quad \text{a.s.}$$

2.2.2. Weak Convergence of U_k in L_2 -Norm and in Weighted L_2 -Norms

Under the Null Hypothesis

Underlying in this subsection is the following situation. Let h be a symmetric kernel function in two real variables. Put

$$\tilde{h}(t) = E\{h(t, X_1) - \theta\} \quad \text{for } t \in \mathbb{R}; \text{ where } \theta = Eh(X_1, X_2).$$

$$(2.1.1.) \quad Eh^2(X_1, X_2) < \infty$$

and

$$(2.1.2.) \quad Eh^2(X_1) = \sigma^2 > 0.$$

We wish to formulate asymptotic results analogous to Theorem 2.2.3., Corollaries 2.2.2. and 2.2.4. for

$$U_k = Z_k - k(n-k)\theta = \sum_{i=1}^k \sum_{j=k+1}^n h(X_i, X_j) - k(n-k)\theta.$$

Here the limiting process will again be the Gaussian process $\Gamma = \{\Gamma(t); 0 \leq t \leq 1\}$, where

$$\Gamma(t) = (1-t)W(t) + tW(1) - tW(t) \quad \text{for } 0 \leq t \leq 1,$$

and where W denotes a Wiener process. We note that

$$W(t) - tW(1) = B(t) \quad \text{for } 0 \leq t \leq 1,$$

where $B = \{B(t); 0 \leq t \leq 1\}$ is a Brownian bridge. Thus

$$\Gamma(t) = 2(1-t)W(t) - B(t) \quad \text{for } 0 \leq t \leq 1 \text{ and}$$

$$\int_0^1 \Gamma^2(t) dt \quad \text{and} \quad \int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt \quad \text{are almost surely finite}$$

because the same holds true for a Wiener process and a Brownian bridge replacing Γ . Utilizing this fact the proofs of the results in the

present subsection can easily be conducted from the proofs in the last subsection.

Corollary 2.2.5.

Under the conditions of Theorem 2.1.14. we have as $n \rightarrow \infty$

$$\frac{n^{-3}}{\sigma^2} \int_0^1 U_{[(n+1)t]}^2 dt \xrightarrow{D} \int_0^1 \Gamma^2(t) dt.$$

Theorem 2.2.6.

Given the assumptions of Corollary 2.1.16., we obtain as $n \rightarrow \infty$

$$\frac{n^{-3}}{\sigma^2} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{U_{[(n+1)t]}^2}{q(t)} dt \xrightarrow{D} \int_0^1 \frac{\Gamma^2(t)}{q(t)} dt$$

if and only if $\int_0^1 \frac{t(1-t)}{q(t)} dt < \infty$.

In particular,

$$\frac{n^{-3}}{\sigma^2} \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{U_{[(n+1)t]}^2}{t(1-t)} dt \xrightarrow{D} \int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt \quad \text{as } n \rightarrow \infty.$$

2.2.3. A Central Limit Theorem for $\hat{U}_k$ and U_k

As explained earlier the main result of the present subsection will be valid for arbitrary kernel functions h of two real variables. Here we will continue our study of weight functions by considering "heavy" weights, i.e. functions that converge to zero as fast as or even faster than $\{t(1-t)\}^2$. The general set up for this subsection is as follows.

Let $h : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ be a kernel, symmetric or antisymmetric. Let $\theta = Eh(X_1, X_2)$ with $\theta = 0$ if h is antisymmetric. Put

$$\tilde{h}(t) = E\{h(t, X_1) - \theta\} \quad \text{for } t \in \mathbb{R},$$

$$S_n^{(1)}(t) = \sigma^{-1} \sum_{i=1}^{[t]} \tilde{h}(X_i) \quad \text{for } 1 \leq t \leq [n/2] \text{ and}$$

$$S_n^{(2)}(t) = \sigma^{-1} \sum_{i=n-[t]+1}^n \tilde{h}(X_i) \quad \text{for } 1 \leq t \leq n - [n/2].$$

Define for $1 \leq y \leq n-1$, y not necessarily integer,

$$U_y = Z_{[y]} - y(n-y)\theta = \sum_{i=1}^{[y]} \sum_{j=[y]+1}^n h(X_i, X_j) - y(n-y)\theta.$$

Then U_y is an interpolation of U_k for $k \leq y < k+1$.

Theorem 2.2.7.

Suppose H_0 holds. Assume

$$(2.1.1.) \quad Eh^2(X_1, X_2) < \infty$$

$$(2.1.2.) \quad E\tilde{h}^2(X_1) = \sigma^2 > 0 \quad \text{and}$$

$$(2.1.6.) \quad E|h(X_1, X_2)|^v < \infty \quad \text{for some } v > 2.$$

Then as $n \rightarrow \infty$,

$$(8 \log n)^{-1/2} \left\{ \sigma^{-2} \int_1^{n-1} \frac{U_y^2}{y^2 (n-y+1)^2} dy - 2 \log n \right\} \xrightarrow{D} N(0,1),$$

where $N(0,1)$ denotes a standard normal random variable.

Corollary 2.2.8.

Under the conditions of Theorem 2.2.7. we have as $n \rightarrow \infty$

$$(8 \log n)^{-1/2} \left\{ \frac{(n+1)^{-3}}{\sigma^2} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{[(n+1)t]}^2}{(t(1-t))^2} dt - 2 \log n \right\} \xrightarrow{D} N(0,1).$$

Corollary 2.2.8. is the main result of this subsection. It represents a central limit theorem for weighted L_2 -test statistics based on U_k and U_k . From the viewpoint of putting more weight on the tails, Corollary 2.2.8. tells us that $(t(1-t))^2$ is too heavy a weight since

$$\int_0^1 \frac{U_{[(n+1)t]}^2}{(t(1-t))^2} dt \text{ is no longer finite,}$$

not even under the null hypothesis. As it is, Corollary 2.2.8. could be used to construct size α tests, using $N(0,1)$ -tables, for testing H_0 versus H_A (as described in the beginning of Section 2.0.).

Proof of Corollary 2.2.8.

Substituting $y = (n+1)t$, $dy = (n+1)dt$, we obtain

$$\begin{aligned} (2.2.7.) \quad \int_1^{n-1} \frac{U_y^2}{y^2 (n-y+1)^2} dy &= \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{(n+1)t}^2}{((n+1)t)^2} \frac{(n+1)}{(n-(n+1)t+1)^2} dt \\ &= (n+1)^{-3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{(n+1)t}^2}{(t(1-t))^2} dt \end{aligned}$$

and furthermore, in case h is antisymmetric,

$$\int_1^{n-1} \frac{U_y^2}{y^2(n-y+1)^2} dy = (n+1)^{-3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{[(n+1)t]}^2}{(t(1-t))^2} dt.$$

Hence for antisymmetric kernel functions h , Corollary 2.2.8. follows immediately from Theorem 2.2.7. In case h is symmetric we now show that

$$(2.2.8.) \left| (n+1)^{-3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{(n+1)t}^2 - U_{[(n+1)t]}^2}{(t(1-t))^2} dt \right| = O_p(\sqrt{\log \log n}) \quad \text{as } n \rightarrow \infty,$$

which together with (2.2.7.) proves the validity of Corollary 2.2.8.

in the case of symmetric kernel functions h because

$$\begin{aligned} & (8 \log n)^{-\frac{1}{2}} \left\{ \frac{(n+1)^{-3}}{\sigma^2} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{(n+1)t}^2}{(t(1-t))^2} dt - 2 \log n \right\} \\ &= (8 \log n)^{-\frac{1}{2}} \left\{ \frac{(n+1)^{-3}}{\sigma^2} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{(n+1)t}^2}{(t(1-t))^2} dt - 2 \log n \right\} \\ &+ (8 \log n)^{-\frac{1}{2}} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{[(n+1)t]}^2 - U_{(n+1)t}^2}{(t(1-t))^2} dt \cdot \frac{(n+1)^{-3}}{\sigma^2} \\ &= (8 \log n)^{-\frac{1}{2}} \left\{ \sigma^{-2} \int_1^{n-1} \frac{U_y^2}{(y(n-y+1))^2} dy - 2 \log n \right\} \\ &+ (8 \log n)^{-\frac{1}{2}} \sqrt{\log \log n} \cdot O_p(1) \quad \text{by (2.2.7.) and (2.2.8.)} \\ &= (8 \log n)^{-\frac{1}{2}} \left\{ \sigma^{-2} \int_1^{n-1} \frac{U_y^2}{(y(n-y+1))^2} dy - 2 \log n \right\} + o_p(1) \xrightarrow{D} N(0,1) \end{aligned}$$

as $n \rightarrow \infty$ by Theorem 2.2.7.

Proof of (2.2.8.)

As before, we again make use of the equality $a^2 - b^2 = (a-b)^2 +$

$2b(a-b)$.

$$\left| (n+1)^{-3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{U_{(n+1)t}^2 - U_{[(n+1)t]}^2}{(t(1-t))^2} dt \right|$$

$$\leq (n+1)^{-3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{|U_{(n+1)t} - U_{[(n+1)t]}|^2}{(t(1-t))^2} dt$$

$$+ 2(n+1)^{-3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} |U_{[(n+1)t]}| \frac{|U_{(n+1)t} - U_{[(n+1)t]}|}{(t(1-t))^2} dt$$

$$= (n+1)^{-3} \theta^2 \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{\{(n+1)t(n-(n+1)t) - [(n+1)t](n-[(n+1)t])\}^2}{(t(1-t))^2} dt$$

$$+ 2(n+1)^{-3} |\theta| \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} |U_{[(n+1)t]}| \frac{|(n+1)t(n-(n+1)t) - [(n+1)t](n-[(n+1)t])|}{(t(1-t))^2} dt$$

and since $|(n+1)t(n-(n+1)t) - [(n+1)t](n-[(n+1)t])|$

$$= |n((n+1)t - [(n+1)t]) - \{((n+1)t)^2 - [(n+1)t]^2\}|$$

$$= |(n+1)t - [(n+1)t]| \cdot |n - \{(n+1)t + [(n+1)t]\}|$$

$$\leq |n - \{(n+1)t + [(n+1)t]\}|$$

for all $0 \leq t \leq 1$

$$\leq \frac{2}{(n+1)^3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{(n - \{(n+1)t + [(n+1)t]\})^2}{(t(1-t))^2} dt$$

$$+ \frac{2|\theta|}{(n+1)^3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} |U_{[(n+1)t]}| \frac{|n - \{(n+1)t + [(n+1)t]\}|}{(t(1-t))^2} dt.$$

The remaining integrals will now be evaluated separately.

$$\text{Since } |n - \{(n+1)t + [(n+1)t]\}| \leq \begin{cases} n-2k & \text{for } \frac{k}{n+1} \leq t < \frac{k+1}{n+1}, k \leq [n/2]-1 \\ 2(k+1)-n & \text{for } \frac{k}{n+1} \leq t < \frac{k+1}{n+1}, k \geq [n/2] \end{cases}$$

(*)

$$\frac{1}{(n+1)^3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{|n - \{(n+1)t + [(n+1)t]\}|^2}{(t(1-t))^2} dt$$

$$\leq \frac{1}{(n+1)^3} \sum_{k=1}^{[n/2]-1} \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} \frac{(n-2k)^2}{(t(1-t))^2} dt + \frac{1}{(n+1)^3} \sum_{k=[n/2]}^{n-2} \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} \frac{(n-2(k+1))^2}{(t(1-t))^2} dt$$

(*)

$$\leq (n+1) \sum_{k=1}^{[n/2]-1} \frac{n^2 - 4nk + 4k^2}{k^2(n-k)^2} \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} \frac{1}{k} dt + (n+1) \sum_{k=[n/2]}^{n-2} \frac{n^2 - 4n(k+1) + 4(k+1)^2}{k^2(n-k)^2} \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} \frac{1}{k} dt$$

$$= \sum_{k=1}^{[n/2]-1} \left\{ \frac{1}{k^2} + \frac{3k^2 - 2nk}{k^2(n-k)^2} \right\} + \sum_{k=[n/2]}^{n-2} \left\{ \frac{1}{k^2} + \frac{4(k+1)^2 - k^2 - 2nk - 4n}{k^2(n-k)^2} \right\}$$

$$\leq 2 \sum_{k=1}^{\infty} \frac{1}{k^2} + 7 \sum_{k=1}^{[n/2]-1} \frac{1}{(n-k)^2}$$

$$\leq 9 \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

Finally, define $q(t) := \sqrt{t(1-t) \log \log \frac{1}{t(1-t)}}$. Then q satisfies the conditions of Corollary 2.1.16. (b) earlier and hence we have

$$\frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} \frac{|U_{[(n+1)t]}|}{q(t)} = O_p(1) \text{ as } n \rightarrow \infty.$$

Consequently, we arrive at:

$$\begin{aligned}
 & \frac{1}{(n+1)^3} \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} |U_{[(n+1)t}]| \frac{|n - \{(n+1)t + [(n+1)t]\}|}{(t(1-t))^2} dt \\
 & \leq \frac{1}{(n+1)^{3/2}} O_p(1) \cdot \int_{\frac{1}{n+1}}^{\frac{n-1}{n+1}} \frac{|n - \{(n+1)t + [(n+1)t]\}|}{(t(1-t))^{3/2}} \cdot \sqrt{\log \log \frac{1}{t(1-t)}} dt \\
 & \leq (n+1)^{-3/2} O_p(1) \cdot \left\{ \sum_{k=1}^{[n/2]-1} \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} \frac{(n-2k)}{\left\{ \frac{k}{n+1} \left(\frac{n-k}{n+1} \right) \right\}^{3/2}} \sqrt{\log \log \frac{1}{\left(\frac{k}{n+1} \right) \left(\frac{n-k}{n+1} \right)}} dt \right. \\
 & \quad \left. + \sum_{k=[n/2]}^{n-2} \int_{\frac{k}{n+1}}^{\frac{k+1}{n+1}} \frac{(2(k+1)-n)}{\left\{ \frac{k}{n+1} \left(\frac{n-k}{n+1} \right) \right\}^{3/2}} \sqrt{\log \log \frac{1}{\left(\frac{k}{n+1} \right) \left(\frac{n-k}{n+1} \right)}} dt \right\} \\
 & \leq O_p(1) \sqrt{\log \log n} \cdot \left\{ \sum_{k=1}^{[n/2]-1} \frac{1}{k^{3/2}} + \sum_{k=[n/2]}^{n-2} \frac{1}{(n-k)^{3/2}} \right\} = O_p(\sqrt{\log \log n})
 \end{aligned}$$

because $\frac{n-k}{n+1} \geq \frac{n/2}{n+1} \geq \frac{1}{4}$ for all $k \leq [n/2]-1$ and $n \geq 1$, $2(k+1)-n = -(n-k)+k+2 \leq k$ for all $k \leq n-2$ and $n \geq 1$, and $\frac{k}{n+1} \geq \frac{(n/2)-1}{n+1} \geq \frac{n-2}{2(n+1)} \geq \frac{1}{4}$ for all $k \geq [n/2]-1$ and $n \geq 3$.

Proof of Theorem 2.2.7.

First we note that by assumption (2.1.1.) in the case of symmetric kernels h we have

$$(2.2.9.) \quad \sup_{1 \leq y \leq n-1} |U_y - U_{[y]}| = O_p(n) \quad \text{as } n \rightarrow \infty,$$

while in the case of an antisymmetric kernel function we have:

$$U_y = U_{[y]} \quad \text{for all } 1 \leq y \leq n-1.$$

Thus, in the following we prove Theorem 2.2.7. in detail for symmetric kernels. The proof for antisymmetric kernel functions follows along the same lines replacing Corollary 2.1.13. by Corollary 2.1.7. where needed. So, until further notice is given, let h denote a symmetric kernel function. Then (2.2.9.) and Corollary 2.1.13. together imply:

$$U_y = (n-y) \sum_{i=1}^{[y]} \tilde{h}(X_i) + y \sum_{i=[y]+1}^n \tilde{h}(X_i) + O_p(n).$$

From here we obtain the following three representations:

$$(2.2.10.) \quad \left\{ \begin{aligned} \int_1^{m_n} \frac{U_y^2}{(y(n-y+1))^2} dy &= \sigma^2 \int_1^{m_n} \left\{ \frac{S_n^{(1)}(y)}{y} \right\}^2 dy + \int_1^{m_n} \left\{ \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)} \right\}^2 dy \\ &+ 2\sigma \int_1^{m_n} \frac{S_n^{(1)}(y)}{y} \cdot \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)} dy \\ &+ O_p(1) \left\{ \int_1^{m_n} \frac{S_n^{(1)}(y)}{y^2} dy \right. \\ &\left. + \int_1^{m_n} \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{y(n-y)} dy + \int_1^{m_n} y^{-2} dy \right\}, \end{aligned} \right.$$

where $m_n = \frac{n}{2} - \log n$ for $n \geq 1$;

$$(2.2.11.) \quad \left\{ \begin{aligned} \int_{m_n}^{k_n} \frac{U_y^2}{(y(n-y+1))^2} dy \\ &= \int_{m_n}^{k_n} \left\{ \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \right\}^2 dy + \int_{m_n}^{k_n} \left\{ \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)} \right\}^2 dy \\ &+ 2 \int_{m_n}^{k_n} \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \cdot \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)} dy \end{aligned} \right.$$

$$+ O_p(1) \left\{ \int_{k_n}^{k_m} \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y^2} dy + \int_{k_n}^{k_m} \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)^2} dy + \int_{k_n}^{k_m} y^{-2} dy \right\},$$

where $k_n = \frac{n}{2} + \log n$ for $n \geq 1$;

$$(2.2.12.) \left\{ \begin{aligned} & \int_{k_n}^{n-1} \frac{U_y^2}{y^2 (n-y+1)^2} dy \\ &= \sigma^2 \int_{k_n}^{n-1} \left\{ \frac{S_n^{(2)}(n-y+1)}{(n-y+1)} \right\}^2 dy + \int_{k_n}^{n-1} \left(\frac{n-y}{n-y+1} \right)^2 \left\{ \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \right\}^2 dy \\ &+ 2\sigma \int_{k_n}^{n-1} \left(\frac{n-y}{n-y+1} \right) \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \cdot \frac{S_n^{(2)}(n-y+1)}{(n-y+1)} dy \\ &+ O_p(1) \left\{ \int_{k_n}^{n-1} \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y(n-y)} dy + \int_{k_n}^{n-1} \frac{S_n^{(2)}(n-y+1)}{(n-y+1)^2} dy \right. \\ &\quad \left. + \int_{k_n}^{n-1} (n-y)^{-2} dy \right\}, \end{aligned} \right.$$

where

$$S_n^{(1)}(t) = \sigma^{-1} \sum_{i=1}^{[t]} \tilde{h}(X_i) \quad \text{for } 1 \leq t \leq [n/2]$$

$$S_n^{(2)}(t) = \sigma^{-1} \sum_{i=n-[t]+1}^{[n]} \tilde{h}(X_i) \quad \text{for } 1 \leq t \leq n - [n/2].$$

Next we use the method employed for the proof of Lemma 2.2.0. to show that each individual integral in the above representations, except the first integral in each, (2.2.10.) and (2.2.12.), respectively, is bounded in probability when divided by $\log \log n$. This means that none of these integrals will contribute to the limit.

$$\left| \int_1^{m_n} \left\{ \frac{\sum_{i=[y]+1}^n \tilde{h}(x_i)}{(n-y)} \right\}^2 dy \right|$$

$$\stackrel{D}{=} \left| \int_{k_n}^{n-1} \left\{ \frac{\sum_{i=1}^{[x]} \tilde{h}(x_i)}{x} \right\}^2 dx \right| \cdot O_p(1)$$

when using the substitution $(n-y) = x$; note that $n-m_n = k_n$; now substituting $x = nt$,

$$= \left| \frac{1}{n} \int_{1/2}^{1 - \frac{1}{n}} \left\{ \frac{\sum_{i=1}^{[nt]} \tilde{h}(x_i)}{t} \right\}^2 dt \right| \cdot O_p(1)$$

$$\leq \left\{ n^{-1/2} \sup_{1/2 \leq t \leq 1} \left| \sum_{i=1}^{[nt]} \tilde{h}(x_i) \right| \right\}^2 \cdot O_p(1) \quad \text{and by Kolmogorov's inequality}$$

$$= O_p(1).$$

It can easily be seen that

$$\left| \int_1^{m_n} \frac{s_n^{(1)}(y)}{y} \cdot \frac{\sum_{i=[y]+1}^n \tilde{h}(x_i)}{(n-y)} dy \right| \stackrel{D}{=} \left| \int_{k_n}^{n-1} \frac{\sum_{i=1}^{[x]} \tilde{h}(x_i)}{x} \cdot \frac{\sum_{i=[x]+1}^n \tilde{h}(x_i)}{(n-x)} dx \right|$$

$$\cdot O_p(1),$$

which is the quantity already examined in (2.2.0.).

$$\left| \int_1^{m_n} \frac{s_n^{(1)}(y)}{y^2} dy \right| = \frac{\sigma^{-1}}{n} \left| \int_{1/n}^{m_n/n} \frac{\sum_{i=1}^{[nt]} \tilde{h}(x_i)}{t^2} dt \right|$$

$$\leq n^{-1/2} \sup_{1/n \leq t \leq m_n/n} \frac{\left| \sum_{i=1}^{[nt]} \tilde{h}(x_i) \right|}{q(t)} \cdot \frac{1}{\sqrt{n\sigma}} \int_{1/n}^{m_n/n} \frac{\{(1-t) \log \log \frac{1}{t(1-t)}\}^{1/2}}{t^{3/2}} dt$$

$$= O_p(\sqrt{\log \log n}) \cdot \frac{1}{\sqrt{n}} \int_{1/n}^{m/n} t^{-3/2} dt \quad \text{by Theorem 1.1.25.}$$

$$= O_p(\sqrt{\log \log n}).$$

For the next integral we have:

$$\begin{aligned} & \left| \int_1^{m/n} \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{y(n-y)} dy \right| \\ & \leq n^{-1/2} \max_{1 \leq k \leq m/n} \left| \sum_{i=k+1}^n \tilde{h}(X_i) \right| \cdot \int_1^{m/n} (y(n-y))^{-1/2} dy \cdot O(1) \\ & = n^{-1/2} \max_{k_n \leq \ell \leq n-1} \left| \sum_{i=1}^{\ell} \tilde{h}(X_i) \right| \cdot \int_{1/n}^{m/n} (t(1-t))^{-1/2} dt \cdot O(1) \\ & = O_p(1) \end{aligned}$$

by Kolmogorov's inequality (confer, e.g., G.R. Shorack and J.A. Wellner (1986), p.843) applied to the maximum of partial sums. Note that the above integrand is the density function of a $\beta(\frac{1}{2}, \frac{1}{2})$ -random variable (modulo the missing normalization). Hence the integral is finite.

Towards the integrals in representation (2.2.11.) we note that

$$\begin{aligned} & \int_{m/n}^{k_n} \left\{ \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \right\}^2 dy \stackrel{D}{=} \int_{m/n}^{k_n} \left\{ \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{n-y} \right\}^2 dy \quad \text{and} \\ (2.2.13.) \quad & \int_{m/n}^{k_n} \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y^2} dy \stackrel{D}{=} \int_{m/n}^{k_n} \frac{\sum_{i=[y]+1}^n \tilde{h}(X_i)}{(n-y)^2} dy \\ & \int_{m/n}^{k_n} \left\{ \frac{\sum_{i=1}^{[y]} \tilde{h}(X_i)}{y} \right\}^2 dy \leq \left\{ n^{-1/2} \max_{m/n \leq k \leq n} \left| \sum_{i=1}^k \tilde{h}(X_i) \right| \right\}^2 \cdot \frac{n}{m/n} \cdot \int_{m/n}^{k_n} y^{-1} dy \\ & \quad \quad \quad = o_p(1) \end{aligned}$$

again by Kolmogorov's inequality since $n \sim 2m_n$ and since

$$\int_{m_n}^{k_n} y^{-1} dy \leq m_n^{-1} \int_{m_n}^{k_n} dy = \frac{k_n - m_n}{m_n} = \frac{2 \log n}{(n/2) - \log n} \rightarrow 0.$$

The integrals in (2.2.13.) can be treated similarly. The remaining integral in (2.2.11.) can be evaluated in the same way (2.2.10.) was.

This completes the discussion of the asymptotic behaviour of the integrals involved in representations (2.2.10.) - (2.2.12.) because each of the integrals in (2.2.12.) equals in distribution to one of the integrals in (2.2.10) and need hence not be examined again. Summarizing what has been proved so far, we arrive at:

$$(2.2.14.) \left\{ \frac{1}{\sqrt{8 \log n}} \cdot \frac{1}{\sigma^2} \int_1^{n-1} \frac{y^2}{(y(n-y))^2} dy = \frac{1}{\sqrt{8 \log n}} \right. \\ \left. \left[\int_1^{m_n} \left\{ \frac{S_n^{(1)}(y)}{y} \right\}^2 dy + \int_{k_n}^{n-1} \left\{ \frac{S_n^{(2)}(n-y+1)}{(n-y+1)} \right\}^2 dy \right] + o_p(1) \text{ as } n \rightarrow \infty. \right.$$

Finally we notice that

$$\int_{k_n}^{n-1} \left\{ \frac{S_n^{(2)}(n-y+1)}{(n-y+1)} \right\}^2 dy = \int_1^{m_n} \left\{ \frac{S_n^{(2)}(t+1)}{(t+1)} \right\}^2 dt = \int_1^{m_n} \left\{ \frac{S_n^{(1)}(t+1)}{(t+1)} \right\}^2 dt.$$

Now we apply Lemma 2.2.1. to $S_n^{(1)}$ and $S_n^{(2)}$, respectively. Here we make use of the fact that we can define two independent Wiener processes to play the role of W in Lemma 2.2.1. From this we conclude that the two limiting standard normal random variables are independent. This together with (2.2.14.) yields Theorem 2.2.7. in the case of a symmetric kernel function h since the relationship between $\log m_n$ and $\log n$ is

as demonstrated below.

For large enough $n \in \mathbb{N}$ we have:

$$\begin{aligned} \frac{\log n - \log 4}{\log n} &= \frac{\log n + \log(\frac{1}{4})}{\log n} < \frac{\log n + \log(\frac{1}{2} - \frac{\log n}{n})}{\log n} \\ &= \frac{\log(\frac{n}{2} - \log n)}{\log n} = \frac{\log m_n}{\log n} \end{aligned}$$

and

$$\frac{\log m_n}{\log n} = \frac{\log(\frac{n}{2} - \log n)}{\log n} < \frac{\log(\frac{n}{2})}{\log n} = \frac{\log n - \log 2}{\log n}$$

Whence $\log m_n$ is asymptotically equivalent to $\log n$, that is

$$\frac{\log m_n}{\log n} \rightarrow 1 \text{ as } n \rightarrow \infty. \text{ In addition,}$$

$$\frac{\log n - \log m_n}{\sqrt{\log n}} = \frac{-\log(\frac{1}{2} - \frac{\log n}{n})}{\sqrt{\log n}} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

2.2.4. Summary and Discussion

In this section we have presented L_2 -norm results corresponding to the L_1 -norm theorems proven in Section 2.1. Furthermore, we have established a central limit theorem for the L_2 -functionals of U_k and $\hat{U}_k$. It is worthwhile mentioning why the break up points for those weight functions that still yield convergence in distribution in Corollary 2.1.10.(b) and in its analog, Theorem 2.2.3., are different. First we notice that the condition of Theorem 2.2.3.

$$\int_0^1 \frac{t(1-t)}{q(t)} dt < \infty$$

is weaker than the condition of Corollary 2.1.10.(b)

$$I(q, c) < \infty \text{ for some } c > 0.$$

This can be seen as follows. Violation of the condition of Theorem 2.2.3. leads to

$$\int_0^{1/2} \frac{t(1-t)}{q(t)} dt = \infty \quad \text{or} \quad \int_{1/2}^1 \frac{t(1-t)}{q(t)} dt = \infty \quad \text{or}$$

both integrals are infinite. Assume without loss of generality that

$$\int_0^{1/2} \frac{t(1-t)}{q(t)} dt = \infty.$$

By definition of the class $\hat{Q}^*$ of weight functions q we have

$$\inf_{\delta \leq t \leq 1-\delta} q(t) > 0 \text{ for all } \delta \in (0, 1/2)$$

and thus for each $0 < \delta < 1/2$ we must have that

$$\int_0^\delta \frac{t}{q(t)} dt \text{ is infinite.}$$

Hence $\frac{t}{q(t)} \rightarrow \infty$ as $t \downarrow 0$ and also $\frac{\sqrt{t}}{q(t)} \rightarrow \infty$ as $t \downarrow 0$. According to Proposition 3.1. in M. Csörgö et al. (1986) this implies that $I(q, c)$ is infinite for all $c > 0$. This means we have just proven that the condition of Corollary 2.1.10.(b) implies the condition of Theorem 2.2.3. The condition of Corollary 2.1.10.(b) was imposed to guarantee the almost sure finiteness of

$$\sup_{0 < t < 1} \frac{|B(t)|}{q(t)}.$$

However,

$$\int_0^1 \frac{B^2(t)}{q(t)} dt$$

may be almost surely finite although the above supremum is almost surely infinite. The same remarks apply to Corollary 2.1.16.(b) and its analog, Theorem 2.2.6.

It is also possible to have general L_p -norm central limit theorems for $\hat{U}_k$ and U_k . We demonstrate this for the case of $\hat{U}_k$ and $1 \leq p < \infty$. Towards this end we first need a lemma and then an analog to Theorem 2.3. in M. Csörgö and L. Horváth (1988c). We claim:

Lemma 2.2.9.

Under the conditions of Theorem 2.1.9. we have for any $0 \leq \alpha < 1/2 - 1/p$ as $n \rightarrow \infty$:

$$n^\alpha \sup_{\frac{1}{n+1} \leq t \leq 1 - \frac{1}{n+1}} \frac{\left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right|}{(t(1-t))^{1-\alpha}} = O_p(1).$$

Proof of Lemma 2.2.9.

According to Corollary 2.1.7. we have (take $k = [(n+1)t]$ there):

$$\frac{n^{-3/2}}{\sigma} \sup_{\frac{1}{n+1} \leq t \leq 1 - \frac{1}{n+1}} \left| \hat{U}_{[(n+1)t]} - \left\{ n \sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - [(n+1)t] \sum_{i=1}^n \hat{h}(X_i) \right\} \right|$$

$$= O_p(n^{-1/2}),$$

where $\hat{h}(t) := E h(t, X_1)$.

Consequently, we obtain

$$n^\alpha \sup_{\frac{1}{n+1} \leq t \leq 1 - \frac{1}{n+1}} \frac{\left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - B_n(t) \right|}{(t(1-t))^{\frac{1}{2}-\alpha}}$$

$$\leq O_p(n^{\alpha-1/2}) \sup_{\frac{1}{n+1} \leq t \leq 1 - \frac{1}{n+1}} (t(1-t))^{\alpha-1/2}$$

$$+ n^\alpha \sup_{\frac{1}{n+1} \leq t \leq 1 - \frac{1}{n+1}} \frac{\left| \left[\sum_{i=1}^{[(n+1)t]} \hat{h}(X_i) - \frac{[(n+1)t]}{n} \sum_{i=1}^n \hat{h}(X_i) \right] \frac{n^{-1/2}}{\sigma} - B_n(t) \right|}{(t(1-t))^{\frac{1}{2}-\alpha}}$$

The last supremum is bounded in probability (provided that

$0 \leq \alpha < 1/2 - 1/\nu$) by Theorem 2.1 in M. Csörgő and L. Horváth (1988c).

Finally, let $f(t) := t(1-t)$, then f is nondecreasing on $(0, 1/2)$, nonincreasing on $(1/2, 1)$, has absolute maximum at $t = 1/2$ and two minima at the end points. Hence

$$\frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq 1 - \frac{1}{n+1}} (t(1-t))^{\alpha-1/2} = \left(\frac{n}{(n+1)^2} \right)^{\alpha-1/2} \sim n^{\frac{1}{2}-\alpha},$$

which completes the proof of Lemma 2.2.9. ■

Theorem 2.2.10.

Under the assumptions of Theorem 2.1.9. we have as $n \rightarrow \infty$

$$\int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]}^p - |B_n(t)|^p \right| \cdot \{t(1-t)\}^{-\frac{p}{2}-1} dt = O_p(1).$$

Proof of Theorem 2.2.10.

It follows from Markov's inequality that

$$(2.2.15.) \quad n^{-\alpha} \int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} |B_n(t)|^{p-1} / \{t(1-t)\}^{p/2+1/2+\alpha} dt = O_p(1) \text{ for } \alpha > 0.$$

Moreover, for any $a, b \in \mathbb{R}$ it is true that

$$(2.2.16.) \quad \left| |a|^p - |b|^p \right| \leq p 2^{p-1} |a-b|^p + p 2^{p-1} |b|^{p-1} |a-b|.$$

Consequently we obtain for $0 < \alpha < 1/2 - 1/p$:

$$\begin{aligned} & \int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]}^p - |B_n(t)|^p \right| \cdot \{t(1-t)\}^{-p/2-1} dt \\ & \leq p 2^{p-1} \int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - |B_n(t)|^p \cdot \{t(1-t)\}^{-p/2-1} dt \\ & \quad + p 2^{p-1} \int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} |B_n(t)|^{p-1} \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - |B_n(t)| \cdot \{t(1-t)\}^{-p/2-1} dt \\ & \quad \text{by (2.2.16.)} \\ & \leq p 2^{p-1} \left\{ \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq 1-\frac{1}{n+1}} n^\alpha \frac{\left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} - |B_n(t)| \right|^p}{\{t(1-t)\}^{1/2-\alpha}} \right\} \cdot n^{-\alpha p} \int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} \frac{dt}{\{t(1-t)\}^{1+\alpha p}} \end{aligned}$$

$$\begin{aligned}
& + \left\{ n^{-\alpha} \int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} |B_n(t)|^{p-1} \cdot \{t(1-t)\}^{-(p/2+1/2+\alpha)} dt \right\} p 2^{p-1} n^{\alpha} \\
& \cdot \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq 1-\frac{1}{n+1}} \frac{\left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]}^{-B_n(t)} \right|}{\{t(1-t)\}^{1/2-\alpha}} \\
& \leq \left\{ \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq 1-\frac{1}{n+1}} n^{\alpha} \frac{\left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]}^{-B_n(t)} \right|}{\{t(1-t)\}^{1/2-\alpha}} \right\}^p p 2^p n^{-\alpha p} \int_{\frac{1}{n+1}}^{1/2} \frac{dt}{\{t(1-t)\}^{1+\alpha p}} \\
& + O_p(1) \cdot \frac{1}{n+1} \sup_{\frac{1}{n+1} \leq t \leq 1-\frac{1}{n+1}} n^{\alpha} \left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]}^{-B_n(t)} \right| / \{t(1-t)\}^{1/2-\alpha} \\
& \text{by (2.2.15.)}
\end{aligned}$$

and by Lemma 2.2.9.

$$= O_p(1) \cdot n^{-\alpha p} \int_{\frac{1}{n+1}}^{1/2} t^{-(1+\alpha p)} dt + O_p(1) = O_p(1).$$

Thus Theorem 2.2.10. is verified. ■

Next we finally present the central limit theorem for $\hat{U}_k$ in L_p -norms, $1 \leq p < \infty$.

Corollary 2.2.11.

Given the conditions of Theorem 2.2.10. we have as $n \rightarrow \infty$:

$$\left\{ 4D \log n \right\}^{-1/2} \cdot \left\{ \int_{\frac{1}{n+1}}^{1-\frac{1}{n+1}} \frac{\left| \frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]} \right|^p}{(t(1-t))^{p/2+1}} dt - 2m \log n \right\} \xrightarrow{D} N(0,1),$$

where $D = D(p)$ is a strictly positive constant, $m = m(p) = E|N(0,1)|^p$ and where $N(0,1)$ denotes a standard normal random variable.

The proof of Corollary 2.2.11. follows from Theorem 2.2.10. in

exactly the same vein as the proof of Corollary 2.1. from Theorem 2.3. did in M. Csörgő and L. Horváth (1988c).

Similar results can easily be obtained for the case of symmetric kernel functions.

Section 2.3. Asymptotic Results Under the Alternative Hypothesis

2.3.0. Notations and Definitions

In this section we will investigate the limiting behaviour of $\hat{U}_{[(n+1)t]}$ and $U_{[(n+1)t]}$ under the alternative hypothesis H_A as well as a number of properties of the power functions involved. We first develop the methodology of power functions and their properties, following closely in notation Chapter 4 of R.H. Randles and D.A. Wolfe (1979).

The model for testing the fixed sample size AMOC-problem considered in this chapter can be rewritten as follows (confer also Subsection 2.1.0.): Let $X_1, X_2, \dots, X_n$ be independent real-valued random variables. Let F and G be arbitrary distribution functions on $\mathbb{R}$ such that $F \neq G$, i.e. there exists $x_0 \in \mathbb{R}$ so that $F(x_0) \neq G(x_0)$.

Test

$H_0 : X_1, X_2, \dots, X_n$ have the same distribution, say F ,

versus

$H_A : \text{there exists } k \in \mathbb{N} \text{ such that}$

$X_1, X_2, \dots, X_k$ have distribution F and

$X_{k+1}, \dots, X_n$ have distribution G .

Consequently, it can be said that the model may be indexed by the vector-valued parameter $\xi = (\Delta, F, G)$, where $\Delta = \Delta_n = I(\sup_{x \in \mathbb{R}} |F_1(x) - F_n(x)| > 0)$ with F_1 and F_n being the cumulative distribution functions of X_1 and X_n , respectively, and with I denoting the indicator function. From this viewpoint the hypotheses may also be represented

in the form of a one-sided test:

$$(2.3.0.) \quad \left\{ \begin{array}{l} H_0 : \Delta = 0 \\ \text{against} \\ H_A : \Delta > 0 \quad (\text{i.e. } \Delta = 1). \end{array} \right.$$

For different possibilities of modelling this problem see Subsection 2.3.3. later on. This particular test has been chosen because of the simplicity of the parameter, ξ although some loss of information from the sample is to be expected since Δ is only based on the distribution of X_1 and X_n .

Definition 2.3.0.

The power function P^T of any test relevant to (2.3.0.) is given by

$$P^T(\xi) = P_{\xi} \{H_0 \text{ is rejected when using } T\}.$$

Since Δ is only capable of taking on the values 0 or 1, where 0 corresponds to H_0 and 1 corresponds to H_A , we can say that

$$P^T(0, F, G) = P^T(0, F) = P_{(0, F)} \{H_0 \text{ is rejected when using } T\} = \alpha$$

and

$$P^T(1, F, G) = P_{(1, F, G)} \{H_A \text{ is accepted when using } T\} = 1 - \beta.$$

Here α and β stand for the probability of making a type I and a type II error, respectively. Before studying properties of power functions, consider an example.

Example 2.3.1.

Suppose h is an antisymmetric kernel function in two real

variables. Then

$$\hat{U}_{[(n+1)t]} = \sum_{i=1}^{[(n+1)t]} \sum_{j=[(n+1)t]+1}^n h(X_i, X_j)$$

is the stochastic process in t on which Theorems 2.1.8 and 2.1.9, as well as Corollary 2.1.10, are based. Put

$$T_n = \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} \frac{|\hat{U}_{[(n+1)t]}|}{q(t)} \quad \text{for } q \in Q^* ;$$

hence the power function P^{T_n} of T_n has the form

$$P^{T_n}(\xi) = P_{\xi} \{H_0 \text{ is rejected when using } T_n\}$$

in general, and under H_0 in particular

$$P^{T_n}(0, F) = P_{(0, F)} \left\{ \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} \frac{|\hat{U}_{[(n+1)t]}|}{q(t)} \leq c \right\},$$

where $c = c_n > 0$ is a critical value chosen such that, if T_n exceeds c , then H_0 is rejected. Under the assumptions of Corollary 2.1.10.

(b), one obtains as $n \rightarrow \infty$

$$P^{T_n}(0, F) \rightarrow P^T(0, F) = P_{(0, F)} \left\{ \sup_{0 < t < 1} \frac{|B(t)|}{q(t)} \leq c \right\},$$

where $T = \sup_{0 < t < 1} \frac{|B(t)|}{q(t)}$. Note that the limiting quantity $P^T(0, F)$

does not depend on the initially underlying distribution F and will

henceforth be denoted by $P^T(0)$. We also mention that the same holds

true for the other Wilcoxon-Mann-Whitney type tests considered in

Section 2.1. The power functions P^{T_n} of

$$T_n = \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |\hat{U}_{[(n+1)t]}|, \quad n \geq 1,$$

and their asymptotic properties will be examined in Subsection*

2.3.2. We now define the first two properties of a power function.

Definition 2.3.2.

The size of a test T with power function P^T is $\sup_{\xi} P^T(\xi)$, where the supremum extends over all those ξ that are admissible under H_0 . The test is said to be of level α , $0 < \alpha < 1$, if its size is less than or equal to α .

Example 2.3.3.

Assume h again to be an antisymmetric kernel in two real variables and put

$$T_n = \frac{n^{-3}}{c^2} \int_0^1 \frac{\hat{U}^2[(n+1)t]}{q(t)} dt \quad \text{for } q \in Q^*.$$

Then

$$P_n^T(\xi) = P_{(\Delta, F, G)} \{H_0 \text{ is rejected when using } T_n\} \quad \text{and}$$

$$P_n^T(0, F) = P_{(0, F)} \left\{ \frac{n^{-3}}{c^2} \int_0^1 \frac{\hat{U}^2[(n+1)t]}{q(t)} dt \geq c \right\},$$

where c is to be selected as described in Example 2.3.1.

Furthermore, as $n \rightarrow \infty$

$$P_n^T(0, F) \rightarrow P^T(0, F) = P \left\{ \int_0^1 \frac{B^2(t)}{q(t)} dt \geq c \right\}$$

provided that q satisfies the conditions of Theorem 2.2.3.

Here $T = \int_0^1 \frac{B^2(t)}{q(t)} dt$. Again $P^T(0, F)$ does not depend on the underlying null distribution F , whence we will in the future suppress F in the notation. The size of the tests T_n is computed to be

$$\sup_{\xi \in H_0} P^{T_n}(\xi) = \sup_{F \in \mathcal{F}} P_{(0,F)} \left\{ \frac{n^{-3}}{\sigma^2} \int_0^1 \frac{\hat{U}^2[(n+1)t]}{q(t)} dt \geq c \right\};$$

in addition we have

$$\sup_{\xi \in H_0} P^T(\xi) = P \left\{ \int_0^1 \frac{B^2(t)}{q(t)} dt \geq c \right\}.$$

Here $\mathcal{F}$ denotes the class of all nondegenerate distributions of real-valued random variables.

Furthermore, we can (in many cases explicitly) determine a critical value $c_\alpha > 0$ such that asymptotically the level of the test T_n is α with $0 < \alpha < 1$, regardless of the underlying distribution in the sample. Choosing $c_\alpha > 0$ so that the level of the test T_n is approximately α , is the same as fixing the probability of a type I error when the sample size is extremely large ($n \rightarrow \infty$). Therefore, we may say that all the tests considered previously in Sections 2.1. and 2.2. have the (asymptotic) property that we can make the probability of a type I error as small as we wish. The decision concerning an optimal level α will depend on the magnitude of the type II error which in general tends to increase as the type I error decreases.

In Subsection 2.3.2. we will compute c_α at different levels of α for

$$T = \int_0^1 B^2(t) dt \quad \text{and} \quad T = \int_0^1 \frac{B^2(t)}{t(1-t)} dt.$$

We now return to the properties of power functions.

Definition 2.3.4.

An α -level test T with power function P^T is said to be unbiased at level α if $P^T(\xi) \geq \alpha$ for all those ξ admissible under H_A . The importance and meaning of the unbiasedness property of a test can be explained as follows. If T is an α -level test, then the probability of making a type I error is less than or equal to α , that is

$$P\{\text{rejecting } H_0 \mid H_0 \text{ is true}\} \leq \alpha.$$

For such a test T the unbiasedness property now states that

$$P\{\text{rejecting } H_0 \mid H_A \text{ is true}\} \geq \alpha,$$

which implies that the probability of rejecting the null hypothesis is greater when the alternative hypothesis is true than it is when the null hypothesis is true.

In order to decide whether or not the test statistics T_n studied in Sections 2.1. and 2.2. possess the property of unbiasedness, we need to compute

$$P_n^T(\xi) = P_n^T(1, F; G) = P_{(1, F, G)}\{H_0 \text{ is rejected when using } T_n\}.$$

(Note again that $\Delta = 1$ is the only possible Δ -value under H_A .)

This amounts to saying that we have to find the distribution of T_n under the alternative hypothesis H_A of exactly one changepoint occurring. Our main result towards this end is Theorem 2.3.7. which is formulated and proven in the next subsection. The last property of power functions that we wish to study is the important concept of consistency.

Definition 2.3.5.

If $\{T_n\}_{n \in \mathbb{N}}$ denotes a sequence of α -level test statistics with power functions P^{T_n} , then the sequence $\{T_n\}_{n \in \mathbb{N}}$ is said to be consistent against the class C_A of parameters in H_A , if

$$\lim_{n \rightarrow \infty} P^{T_n}(\xi) = 1 \quad \text{for every } \xi \in C_A.$$

The idea behind the concept of consistency of a sequence of tests is that with enough sample information (i.e. $n \rightarrow \infty$) we should be able to detect any point in the alternative, C_A .

The following lemma (confer Theorem 4.2.18. in R.H. Randles and D.A. Wolfe (1979)) is needed later on for the proof of Corollary 2.3.8.

Lemma 2.3.6.

Let $\{\hat{T}_n\}_{n \in \mathbb{N}}$ denote a sequence of test statistics for an α -level test of

$$H_0 : \xi \in \omega$$

versus

$$H_A : \xi \in \Omega - \omega$$

such that the test based on $\hat{T}_n$ rejects H_0 if $\hat{T}_n \geq c_n$. Suppose there exists a function $k(\xi)$ such that

$$\hat{T}_n \rightarrow k(\xi) \quad \text{for every } \xi \in \Omega \text{ in probability as } n \rightarrow \infty.$$

If, in addition,

$$k(\xi) = k_0 \quad \text{for all } \xi \in \omega,$$

$$k(\xi) > k_0 \quad \text{for all } \xi \in \Omega - \omega, \text{ and}$$

$$\lim_{n \rightarrow \infty} c_n \leq k_0,$$

then $\{\hat{T}_n\}$ is a consistent sequence of tests for all alternatives in $H_A : \xi \in \Omega - \omega$.

We note that it will be a consequence of Theorem 2.3.7. that for all the tests T_n (Wilcoxon-Mann-Whitney type as well as Cramér-von Mises type tests) which are based on symmetric kernel functions, Lemma 2.3.6. holds for suitably chosen alternatives H_A with $k_0 = k(0) = 0$ and $k(1)$ defined as the in probability limiting constant in Theorem 2.3.7. (confer also Corollary 2.3.8.).

In this section we will again tacitly assume that the kernel functions h are always nondegenerate, i.e. $\text{Var } h > 0$.

2.3.1. Tests Based on Symmetric Kernels

The following theorem, which including an outline of proof, is due to M. Csörgő and L. Horváth (1988a) (confer there Theorem 3.1) represents the main result in this section.

Theorem 2.3.7.

Let $X_1, X_2, \dots, X_n$ be independent real-valued random variables with

$$\theta_1 = E h(X_{[n\lambda]-1}, X_{[n\lambda]})$$

$$\theta_2 = E h(X_{[n\lambda]+1}, X_{[n\lambda]+2})$$

$$\theta_3 = E h(X_{[n\lambda]}, X_{[n\lambda]+1}) \quad \text{for } \lambda \in (0,1) \text{ fixed.}$$

Suppose $h(x,y) = h(y,x)$ for every $x, y \in \mathbb{R}$.

Put $\log^+ x = \log(x \vee 1)$, where $x \vee 1$ denotes the maximum of x and 1.

1. Assume that H_A holds and that

$$(2.3.1.) \quad \begin{cases} E|h(X_{[n\lambda]-1}, X_{[n\lambda]})| < \infty, E|h(X_{[n\lambda]+1}, X_{[n\lambda]+2})| < \infty, \\ E\{|h(X_{[n\lambda]}, X_{[n\lambda]+1})| \log^+(|h(X_{[n\lambda]}, X_{[n\lambda]+1})|)\} < \infty \end{cases}$$

are satisfied. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} U_{[(n+1)t]} = \begin{cases} (\lambda-t)t\theta_1 + (1-\lambda)t\theta_3 & \text{if } 0 < t \leq \lambda \\ (t-\lambda)(1-t)\theta_2 + \lambda(1-t)\theta_3 & \text{if } \lambda \leq t < 1 \end{cases}$$

in probability.

Note that it is a direct consequence of Theorem 2.1.14. that

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} U_{[(n+1)t]} = 0 \quad \text{in probability under the null hypothesis.}$$

Also note that condition (2.3.1⁺) here replaces the stronger condition of square integrability of h which was used earlier on.

It is in the proof of Theorem 2.3.7. that for the first time the theory of generalized (two-sample) U-statistics comes into the picture. This is the reason why we provide a detailed proof of Theorem 2.3.7. although a sketch of the same proof was already given in M. Csörgö and L. Horváth (1988a).

Proof of Theorem 2.3.7.

Let $k = k_n = [n\lambda]$, $1 \leq k < n$, be the single changepoint allowed under H_A . Set $m = [(n+1)t]$ and assume for the time being that $1 \leq m \leq k$. Then the following statements can easily be seen to be true:

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

$$\sum_{i=1}^m \sum_{j=m+1}^k h(X_i, X_j) + \sum_{i=1}^m \sum_{j=k+1}^n h(X_i, X_j) = \sum_{i=1}^m \sum_{j=m+1}^n h(X_i, X_j) = U_m,$$

whence

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

$$= R_n^{(1)} + R_n^{(2)} - [R_n^{(3)} + R_n^{(4)} + R_n^{(5)}].$$

Note that, except for a missing normalizing factor, $R_n^{(1)}$, $R_n^{(3)}$, and $R_n^{(4)}$ are one-sample U-statistics and that they are nondegenerate provided that the variance of h is positive for each X_i , $1 \leq i \leq n$. In this situation the classical strong law of large numbers generalizes to U-statistics (confer, e.g., Theorem A, Section 5.4, R.J. Serfling (1980)) and yields

$$\binom{k}{2}^{-1} \sum_{1 \leq i < j \leq k} h(X_i, X_j) = \frac{2}{[n\lambda]([n\lambda]-1)} \sum_{1 \leq i < j \leq [n\lambda]} h(X_i, X_j) \xrightarrow{\text{a.s.}} \theta_1 \quad \text{as } n \rightarrow \infty,$$

$$\text{implying that } \frac{1}{n^2} R_n^{(1)} \xrightarrow{\text{a.s.}} \frac{\lambda^2 \theta_1}{2} \quad \text{as } n \rightarrow \infty$$

$$\text{because } \frac{[n\lambda]([n\lambda]-1)}{n^2} \rightarrow \lambda^2 \quad \text{as } n \rightarrow \infty.$$

$$\text{In the same way we obtain } \frac{1}{n^2} R_n^{(3)} \xrightarrow{\text{a.s.}} \frac{t^2 \theta_1}{2} \quad \text{as } n \rightarrow \infty.$$

Next, consider $R_n^{(4)}$. Since

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

$$\text{the above cited Theorem A implies: } \frac{1}{n^2} R_n^{(4)} \xrightarrow{P} \frac{(\lambda-t)^2 \theta_1}{2} \quad \text{as } n \rightarrow \infty.$$

The statistic $R_n^{(2)}$ is, except for the proper normalization factor, a generalized (two-sample) U-statistic in the sense of P.K. Sen (1974) and (1977) with $c = 2$, $m_1 = m_2 = 1$ and taking $n_1 = k$ and $n_2 = n-k$ (renumbering necessary in the second sum). Thus Theorem 1 of P.K. Sen (1977) (here condition (2.3.1.) is used) implies

$$\frac{1}{k(n-k)} \sum_{i=1}^k \sum_{j=k+1}^n h(x_i, x_j) \xrightarrow{P} \theta_3 \text{ as } n \rightarrow \infty$$

and therefore $\frac{1}{n^2} R_n^{(2)} \xrightarrow{P} \lambda(1-\lambda)\theta_3$ as $n \rightarrow \infty$.

In the same vein we have $\frac{1}{n^2} R_n^{(5)} \xrightarrow{P} (\lambda-t)(1-\lambda)\theta_3$ as $n \rightarrow \infty$.

Putting all these observations together, we arrive at

$$\begin{aligned} \frac{1}{n^2} U[(n+1)t] &= \frac{1}{n^2} \{ R_n^{(1)} + R_n^{(2)} - [R_n^{(3)} + R_n^{(4)} + R_n^{(5)}] \} \\ &\xrightarrow{P} \left\{ \frac{\lambda^2 \theta_1}{2} + \lambda(1-\lambda)\theta_3 - \left[\frac{t^2 \theta_1}{2} + \frac{(\lambda-t)^2 \theta_1}{2} + (\lambda-t)(1-\lambda)\theta_3 \right] \right\} \\ &= (\lambda-t)t\theta_1 + (1-\lambda)t\theta_3, \end{aligned}$$

which proves Theorem 2.3.7. in the case that $1 \leq m = [(n+1)t] \leq [n\lambda] = k$.

The proof of the case $m \geq k$ is similar and is hence omitted here. ■

We observe that if $\theta_1 = \theta_2 = \theta_3 = \theta$, then

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} U[(n+1)t] \stackrel{P}{=} t(1-t)\theta \text{ for all } t \in (0,1).$$

This means that under the alternative hypothesis of exactly one change in distribution, say from F to G , none of the tests T_n considered in Subsection 2.3.0. will (even as $n \rightarrow \infty$) be able to distinguish between F and G in the case that $\theta = 0$. In more formal language, we can say that if $\theta_1 = \theta_2 = \theta_3 = \theta = 0$, none of the sequences $\{T_n\}_{n \in \mathbb{N}}$ is consistent against any class C_A of alternatives. However, if at least one of the θ_i 's, $i=1,2,3$, is nonzero, then

$$(\lambda-t)t\theta_1 + (1-\lambda)t\theta_3 \text{ or } (t-\lambda)(1-t)\theta_2 + (1-t)\lambda\theta_3 \neq 0,$$

thus the probability that H_0 will be rejected when H_A is true while

using T_n is in fact 1 in the limit (asymptotic unbiasedness of T_n for each level α) and the limits of $\{T_n\}_{n \in \mathbb{N}}$ are different in probability under H_0 and H_A , i.e. $\{T_n\}_{n \in \mathbb{N}}$ is consistent. We summarize this in

Corollary 2.3.8.

Any of the sequences $\{T_n\}_{n \in \mathbb{N}}$ of α -level tests based on $U_{[(n+1)t]}$ considered previously is consistent against the class C_A of alternatives, where

$$C_A = \{\xi \mid \xi = (\Delta, F, G) \text{ and } F, G \text{ such that at least one of } \theta_1, \theta_2, \theta_3 \neq 0\}.$$

For a verification of Corollary 2.3.8. confer Lemma 2.3.6. and the remark immediately following this lemma.

Example 2.3.9.

Let $c_\alpha > 0$ be such that $P_{(0,F)} \left\{ \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |U_{[(n+1)t]}| \geq c_\alpha \right\} = \alpha$ for some fixed, $0 < \alpha < 1$ and for large enough $n \in \mathbb{N}$.

Then application of Theorem 2.3.7. yields:

$$\begin{aligned} P_{(1,F,G)} \{H_0 \text{ is rejected when using } T_n\} \\ = P_{(1,F,G)} \left\{ \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |U_{[(n+1)t]}| \geq c_\alpha \right\} \\ \geq P_{(1,F,G)} \left\{ \frac{1}{n^2} U_{[(n+1)t]} \geq \frac{\sigma c_\alpha}{\sqrt{n}} \right\} \end{aligned}$$

$$\rightarrow 1 \text{ as } n \rightarrow \infty \text{ (pick } t \text{ such that } \frac{U_{[(n+1)t]}}{n^2} \not\rightarrow 0!),$$

whence for large enough $n \in \mathbb{N}$ $T_n = \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |U_{[(n+1)t]}|$ is certainly unbiased at any level α and the sequence $\{T_n\}_{n \in \mathbb{N}}$ is consistent against the class C_A as introduced above.

The same holds true for all the Wilcoxon-Mann-Whitney type and the Cramér-von Mises type statistics with symmetric kernel function h .

The values c_α can always be computed approximately from the corresponding limiting distribution under H_0 . For $T_n = \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |U_{[(n+1)t]}|$ for instance, we determine $c_\alpha > 0$ so that

$$P\left\{\sup_{0 < t < 1} |\Gamma(t)| \geq c_\alpha\right\} = \alpha \text{ for given } 0 < \alpha < 1.$$

For the Cramér-von Mises type tests with symmetric kernel h c_α -values for several different levels of α can be read off Tables 7 and 8.

2.3.2. Tests Based on Antisymmetric Kernels

A remark following the proof of Theorem 2.1. in W.F. Grams and R.J. Serfling (1973) implies that, if we assume the existence of second moments of h , then the strong law of large numbers holds for the generalized (two-sample) U-statistics as defined in P.K. Sen (1974).

From this fact we draw two conclusions:

(a) If h has at least two finite moments, then the result of Theorem 2.3.7. holds almost surely (instead of in probability). Here the kernel function h is again assumed to be symmetric.

(b) Since generalized U-statistics also allow for antisymmetric kernels of two or more blocks of random variables to be used, we immediately have the following theorem because the weak law of large numbers still holds for one-sample U-statistics based on antisymmetric kernel functions.

Theorem 2.3.10.

Let h be an antisymmetric kernel in two real variables. Assume that H_A holds and that the conditions of Theorem 2.3.7. are satisfied. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \hat{U}_{[(n+1)t]} = \begin{cases} (1-\lambda)t\theta_3 & \text{if } 0 < t \leq \lambda \\ \lambda(1-t)\theta_3 & \text{if } \lambda \leq t < 1 \end{cases} \quad \text{in probability.}$$

Realizing that $\theta_1 = Eh(X_{[n\lambda]-1}, X_{[n\lambda]}) = 0$ and $\theta_2 = Eh(X_{[n\lambda]+1}, X_{[n\lambda]+2}) = 0$ for antisymmetric kernels h , the proof of Theorem 2.3.10. is immediate from Remark (b) above when reconsidering the proof of Theorem 2.3.7.

Corollary 2.3.11.

Given the assumptions of Theorem 2.3.10., the sequences $\{T_n^{(i)}\}_{n \in \mathbb{N}}$, $i = 1, 2, 3$, where

$$\begin{aligned} T_n^{(1)} &= \frac{n^{-3/2}}{\sigma} \sup_{0 < t < 1} |\hat{U}_{[(n+1)t]}|, \\ T_n^{(2)} &= \frac{n^{-3}}{\sigma^2} \int_0^1 \hat{U}_{[(n+1)t]}^2 dt \quad \text{and} \\ T_n^{(3)} &= \frac{n^{-3}}{\sigma^2} \int_0^1 \frac{\hat{U}_{[(n+1)t]}^2}{t(1-t)} dt \end{aligned}$$

are consistent against the class C_A of alternatives, where $C_A = \{\xi \mid \xi = (\Delta, F, G) \text{ and } F, G \text{ such that } \theta_3 \neq 0\}$.

Table 4 below (confer E.S. Pearson and H.O. Hartley (1976), Table 54, Case 1) displays the critical values c_α of

$$\begin{aligned} T^{(1)} &= \sup_{0 < t < 1} |B(t)| \\ T^{(2)} &= \int_0^1 B^2(t) dt \quad \text{and} \\ T^{(3)} &= \int_0^1 \frac{B^2(t)}{t(1-t)} dt \end{aligned}$$

for the α -values 0.15, 0.10, 0.05, 0.025 and 0.001, respectively.

These values may be used as the critical values $c_\alpha = c_{n,\alpha}$ for the test statistics in Corollary 2.3.11.

Table 4. Upper Percentage Points for Test T

Test T.	15%	10%	5%	2.5%	1%
$\sup_{0 < t < 1} B(t) $	1.138	1.224	1.358	1.480	1.628
$\int_0^1 B^2(t) dt$	0.284	0.347	0.461	0.581	0.743
$\int_0^1 \frac{B^2(t)}{t(1-t)} dt$	1.61	1.933	2.492	3.020	3.857

2.3.3. Conclusions

In the present section test statistics for the changepoint problem were derived from the functionals of stochastic processes considered previously in this chapter. Then asymptotic properties of the power functions of these tests were studied. In order to discuss power functions and their properties a parameter Δ was introduced. The particular Δ chosen in Subsection 2.3.0. was

$$\Delta = \Delta_n = I \left\{ \sup_{x \in \mathbb{R}} |F_1(x) - F_n(x)| > 0 \right\}.$$

This special choice of Δ was made for simplicity's sake only. There are two ways in which the parameter Δ can be modified to achieve higher power and faster asymptotic unbiasedness and consistency of the tests under consideration. If one wants to keep Δ independent of the choice of the kernel function h , then a more efficient choice for Δ would be $\Delta = \sup_{x \in \mathbb{R}} |F_1(x) - F_n(x)|$. The sup-norm could also be replaced by any other L_p -norm ($p > 1$). Furthermore, Δ could be based on functionals involving not only the distribution of X_1 and X_n but also the distribution of a number of the random variables X_i , $2 \leq i \leq n$. The latter choice of Δ corresponds to an attempt to avoid the influence of outliers. On the other hand the kernel function h itself could be used to construct a variety of Δ 's. Here we only give one example.

Example 2.3.12.

$$\Delta := E \left\{ \sum_{i=3}^n h(X_1, X_i) - (n-2)h(X_1, X_2) \right\},$$

e

where it is assumed without loss of generality that the changepoint k does not occur until the third observation is taken.

Then under H_0 , $\Delta = 0$, while under H_A we obtain:

$$\begin{aligned}\hat{\Delta} &= E\left\{\sum_{i=3}^k h(X_1, X_i) - (k-2)h(X_1, X_2)\right\} + E\left\{\sum_{i=k+1}^n h(X_1, X_i) - (n-k)h(X_1, X_2)\right\} \\ &= (n-k)E\{h(X_1, X_n) - h(X_1, X_2)\} \\ &= (n-k) \cdot \delta\end{aligned}$$

if we assume that $Eh(X_1, X_n) + \delta = Eh(X_1, X_2)$ for some $\delta \neq 0$.

Corresponding to $\delta < 0$ or $\delta \neq 0$, the alternative hypothesis will be one- or two-sided.

Chapter 3. Limiting Distributions for Certain Cramér-von Mises Type

Test Statistics

Section 3.0. Introduction

In Chapter 2 probabilistic methods (theory of stochastic processes, weak convergence in various L_p -norms, Komlós-Major-Tusnády approximations,...) were used to develop nonparametric test statistics for the fixed sample size changepoint problem. In Chapter 3 now we will employ methods borrowed from functional analysis and from the theory of ordinary differential equations to find explicitly the asymptotic distributions of those test statistics under the null hypothesis of no change. The major goal of the present chapter is to find (approximate) upper percentage points for the test statistics discussed in the previous chapter. Methods already established in the literature rely mostly on the computation of eigenvalues of certain differential equations. The applicability of these methods is examined first. Then a simple approach using the classical representation of a Wiener process by Paley and Wiener (1934) is discussed. This approach which leads to surprisingly short and elementary solutions, seems to have been overlooked in the literature thus far. Utilizing the latter approach, tables for upper percentage points were also computed and are displayed throughout this chapter.

Section 3.1. Some Distribution Results for Functionals of Brownian Bridges and Related Gaussian Processes

3.1.0. Notations and Basic Theorems

In Section 2.2. certain random variables were established as the in distribution limits for functionals of $\hat{U}_k$ and U_k under the null hypothesis of no change. In the case of antisymmetric kernel functions h we obtained the following random variables as the limiting functionals (regardless of the choice of h and the null distribution of the random sample $X_1, X_2, \dots, X_n$).

$$(3.1.0.) \left\{ \begin{array}{l} \int_0^1 B^2(t) dt, \quad \int_0^1 \frac{B^2(t)}{t(1-t)} dt \quad \text{and} \\ \int_0^1 \frac{B^2(t)}{q(t)} dt \quad \text{for } q \in Q^* \text{ such that } \int_0^1 \frac{t(1-t)}{q(t)} dt < \infty, \end{array} \right.$$

while in the case of symmetric kernels we arrive at the limiting random variables

$$(3.1.1.) \left\{ \begin{array}{l} \int_0^1 \Gamma^2(t) dt, \quad \int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt \quad \text{and} \\ \int_0^1 \frac{\Gamma^2(t)}{q(t)} dt \quad \text{provided that } \int_0^1 \frac{t(1-t)}{q(t)} dt < \infty. \end{array} \right.$$

In the present section we are interested in finding the distribution (or a close approximation to it) of the random quantities in (3.1.0.) and (3.1.1.). Towards this end we will now introduce some notations and general theorems from the theory of orthogonal decompositions of stochastic processes.

Let $L^2 = L^2[0,1]$ denote the class of all-real-valued Lebesgue-measurable functions g on the unit interval satisfying

$$\int_0^1 g^2(t) dt < \infty ;$$

where the integral is taken to be in the Lebesgue sense. Let

$$\langle g_1, g_2 \rangle = \int_0^1 g_1(t) g_2(t) dt, \quad g_1, g_2 \in L^2[0,1],$$

be the usual inner product on L^2 and let

$$\|g\| = \sqrt{\langle g, g \rangle}$$

denote the induced norm on $L^2[0,1]$. Then L^2 is a Hilbert space, that is a complete normed linear space (confer, e.g., E. Kreyszig (1978), Chapter 3).

Assume that K is a nonidentically zero, symmetric, square integrable function on the unit square $[0,1] \times [0,1]$. That is

$$(3.1.2.) \quad \begin{cases} K \not\equiv 0, & K(s,t) = K(t,s) \text{ for } 0 \leq s, t \leq 1, \text{ and} \\ \int_0^1 \int_0^1 K^2(s,t) ds dt < \infty. \end{cases}$$

Definition 3.1.0.

A function K as described in (3.1.2.) is called a kernel.

This notion of a kernel has no connection with the symmetric or anti-symmetric kernel functions h studied earlier in Chapter 2. The fact that their names coincide is accidental, but cannot be avoided if one wishes to use the notations nowadays common throughout most of the literature of the respective areas.

Definition 3.1.1.

(a) A kernel K as in Definition 3.1.0. is positive (semi-) definite if and only if

$$\int_0^1 \int_0^1 g(s)K(s,t)g(t)dsdt > 0 \quad \text{for all } g \in L^2 \text{ with } g \not\equiv 0. \quad (\geq)$$

(b) If for $\lambda \in \mathbb{R}$ there exists $f \in L^2$ such that

$$\int_0^1 f(s)K(s,t)ds = \lambda f(t), \quad 0 \leq t \leq 1,$$

then λ is called an eigenvalue of the kernel K and f is called an associated or corresponding eigenfunction.

Using these definitions we will now formulate a version of Mercer's Theorem which will prove to be one of the main tools in the examination of the distribution of the random variables in (3.1.0.) and (3.1.1.).

Theorem 3.1.2.

If K is a positive semi-definite kernel, continuous on $[0,1] \times [0,1]$, then K has the representation

$$K(s,t) = \sum_{k=1}^{\infty} \lambda_k f_k(s) f_k(t) \quad \text{for } 0 \leq s, t \leq 1,$$

where the λ_k 's are the eigenvalues of K , the f_k 's are the corresponding eigenfunctions and where the functions f_k are orthonormal, i.e. $\int_0^1 f_k^2(t)dt = 1$ and the f_k 's are orthogonal. The convergence of the above series is both, uniform and absolute.

For further details concerning Mercer's Theorem and related concepts we refer to G.R. Shorack and J.A. Wellner (1986), Chapter 5. The following lemma gives a criterion for testing whether or not the

covariance functions of certain processes are positive, semi-definite kernels. Thus this lemma provides a first link between the aforementioned functional analytic concepts and the area of stochastic processes.

Lemma 3.1.3.

Let $X = \{X(t) : 0 \leq t \leq 1\}$ be a nondegenerate mean zero process whose sample paths are almost surely L^2 -functions. Denote by $K = K_X$ the covariance function of X . Then K is a positive semi-definite kernel if

$$\int_0^1 K(t,t) dt < \infty.$$

A proof of Lemma 3.1.3. can be found in G.R. Shorack and J.A. Wellner (1986), pp. 208-210. Before developing this theory further, we give two examples.

Examples 3.1.4.

(a) Consider $B = \{B(t) : 0 \leq t \leq 1\}$. A Brownian bridge is a nondegenerate Gaussian process with $EB(t) = 0$ for each $0 \leq t \leq 1$ and almost surely continuous sample paths. Furthermore, for almost every $\omega \in \Omega$, $B(\omega)$ is square integrable as a function of t . Hence we can take $X \equiv B$ in Lemma 3.1.3. The covariance function $K = K_B$ of a Brownian bridge is given by $K(s,t) = s \wedge t - st$ for $0 \leq s, t \leq 1$, where $\wedge$ stands for minimum. Since

$$\int_0^1 K_B(t,t) dt = \int_0^1 t(1-t) dt < \infty,$$

Lemma 3.1.3. yields that K_B is a positive semi-definite kernel. In

addition, K_B is continuous over $[0,1] \times [0,1]$, whence Mercer's Theorem implies that K_B has a representation of the form

$$K_B(s,t) = \sum_{k=1}^{\infty} \lambda_k f_k(s) f_k(t) \quad \text{for } 0 \leq s, t \leq 1.$$

Finally, we wish to determine the eigenvalues λ_k and the associated eigenfunctions f_k of K_B . That is, we need to know all possible λ and f such that for $t \in [0,1]$

$$(3.1.3.) \quad \lambda f(t) = \int_0^1 K(s,t) f(s) ds$$

In order to solve (3.1.3.) we employ the following general method.

Assume $\lambda \neq 0$, then differentiate (3.1.3.) with respect to t to obtain

$$\begin{aligned} \lambda f'(t) &= \frac{d}{dt} \left\{ \int_0^1 (s \wedge t) f(s) ds - \int_0^1 s t f(s) ds \right\} \\ &= \frac{d}{dt} \left\{ \int_0^t s f(s) ds + t \int_t^1 f(s) ds - t \int_0^1 s f(s) ds \right\} \\ &= t f(t) + \left\{ \int_t^1 f(s) ds - t f(t) \right\} - \int_0^1 s f(s) ds \\ &= \int_t^1 f(s) ds - \int_0^1 s f(s) ds. \end{aligned}$$

Differentiating again yields $\lambda f''(t) = -f(t)$.

By plugging 0 and 1 into (3.1.3.) it can easily be verified that we need $f(0) = f(1) = 0$. Thus we require a solution of the boundary value problem

$$\lambda y'' + y = 0, \quad y(0) = y(1) = 0.$$

Furthermore, we have to observe the normalizing condition $\int_0^1 f^2(t) dt = 1$

because we wish our eigenfunctions to be orthonormal. The well known solutions to this second-order homogeneous differential equation are (confer E. Kamke (1951), C2.9.)

$$\lambda_k = (k\pi)^{-2} \quad \text{and} \quad f_k(t) = \sqrt{2} \sin k\pi t \quad \text{for } 0 \leq t \leq 1, k \in \mathbb{N}.$$

This, in turn, gives

$$K_B(s, t) = 2 \sum_{k=1}^{\infty} \frac{\sin k\pi s \sin k\pi t}{(k\pi)^2} \quad \text{uniformly on } [0, 1] \times [0, 1].$$

(b) Consider the nondegenerate mean zero Gaussian process $\Gamma = \{\Gamma(t) : 0 \leq t \leq 1\}$ defined in the beginning of Subsection 2.1.2. Almost every sample path of Γ is again continuous and square integrable. The covariance function $K = K_\Gamma$ of Γ is for $0 \leq s, t \leq 1$

$$\begin{aligned} K_\Gamma(s, t) &= E\Gamma(s)\Gamma(t) \\ &= E((1-s)W(s) + sW(1) - sW(s))((1-t)W(t) + tW(1) - tW(t)) \\ &= (1-2s)(1-2t)EW(s)W(t) + (1-2s)tEW(s)W(1) \\ &\quad + s(1-2t)EW(1)W(t) + stEW(1)W(1) \\ &= (1-2s)(1-2t)(s \wedge t) + (3-2s-2t)st \end{aligned}$$

since $E\Gamma(s) = E\Gamma(t) = 0$ and $EW(s)W(t) = s \wedge t$. Hence,

$$\int_0^1 K_\Gamma(t, t) dt = \int_0^1 \{(1-2t)^2 t + (3-4t)t^2\} dt < \infty$$

and thus Lemma 3.1.3. and also Mercer's Theorem, because K_Γ is continuous in both variables, apply to $K = K_\Gamma$. We note at this point that it will be one of the main goals of Subsection 3.1.2. to determine the eigenvalues and the eigenfunctions of K_Γ .

Our last task in the present subsection is now to establish a theorem that allows us to actually compute the distribution of $\int_0^1 X^2(t)dt$ for a large enough class of stochastic processes X . Under the conditions of Mercer's Theorem this can be done readily. However, we do not wish to make the continuity assumption of Mercer's Theorem since the processes X defined via

$$X(t) = \frac{B(t)}{\sqrt{t(1-t)}} \quad \text{and} \quad X(t) = \frac{F(t)}{\sqrt{t(1-t)}}$$

violate this condition. Hence we would like to work under more general assumptions. For this we first need the following

Definition 3.1.5.

An orthonormal set $\{f_1, f_2, \dots\}$ in L^2 is called complete if and only if any one of the following four equivalent conditions holds:

- (i) $\{f_1, f_2, \dots\}$ is a maximal orthonormal set.
- (ii) The set of all finite linear combinations of f_k 's is dense (with respect to $\|\cdot\|$) in $L^2[0,1]$.
- (iii) For all $g \in L^2[0,1]$ we have: $\|g\|^2 = \sum_{k=1}^{\infty} \langle g, f_k \rangle^2$.
- (iv) For each $g, h \in L^2$ we have: $\langle g, h \rangle = \sum_{k=1}^{\infty} \langle g, f_k \rangle \langle h, f_k \rangle$.

The next theorem is due to M. Kac and A.J.F. Siegert. An easily accessible proof is given in G.R. Shorack and J.A. Wellner (1986), pp.210-212.

Theorem 3.1.6.

Assume that $K = K_X$ is the covariance function of a nondegenerate mean zero stochastic process $X = \{X(t) : 0 \leq t \leq 1\}$ whose sample

paths are almost surely in a subspace L of $L^2[0,1]$. Suppose that

K satisfies

$$\int_0^1 K(t,t) dt < \infty.$$

Let $\lambda_1 \geq \lambda_2 \geq \dots > 0$ represent the entire spectrum of eigenvalues of K and assume that the associated orthonormal eigenfunctions form a complete set for the subspace L . Furthermore, suppose that the Kac-Siebert decomposition

$$(3.1.4.) \quad K(s,t) = \sum_{k=1}^{\infty} \lambda_k f_k(s) f_k(t) \text{ is valid for } 0 < s, t < 1.$$

Then

$$(3.1.5.) \quad T = T_X = \int_0^1 X^2(t) dt \stackrel{a.s.}{=} \sum_{k=1}^{\infty} \lambda_k Z_k^2,$$

where $(Z_k)_{k \in \mathbb{N}}$ is a sequence of uncorrelated standard normal random variables,

$$E \left| X(t) - \sum_{k=1}^m \sqrt{\lambda_k} Z_k f_k(t) \right|^2 \rightarrow 0 \text{ as } m \rightarrow \infty \text{ for each fixed } t \in (0,1)$$

and

$$X \stackrel{D}{=} \sum_{k=1}^{\infty} \sqrt{\lambda_k} Z_k f_k.$$

Example 3.1.7.

We apply Theorem 3.1.6. to $X = B$. From Example 3.1.4.(a) we know that $\lambda_k = (k\pi)^{-2}$ and $f_k(t) = \sqrt{2} \sin k\pi t$; thus (3.1.5.) reads in this particular case:

$$T_B = \int_0^1 B^2(t) dt \stackrel{a.s.}{=} \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} Z_k^2$$

for uncorrelated $N(0,1)$ -random variables Z_k .

A question that should be answered at this point is the following: How can we use representation (3.1.5.) to determine the distribution of $\int_0^1 X^2(t) dt$. Various methods, among which the ones mentioned below play a prominent role, have been introduced to solve the problem.

METHOD 1. Use characteristic functions (moment generating functions etc.)

Observe that the random variable T in representation (3.1.5.) has the characteristic function

$$\phi_T(t) = E e^{itT} = \prod_{k=1}^{\infty} (1 - 2\lambda_k t^2)^{-1/2},$$

if the Z_k 's are independent. In any given situation it may or may not be possible to invert this characteristic function to obtain the exact distribution of T .

METHOD 2. Use truncation techniques.

Truncate the infinite series in (3.1.5.) and assume the Z_k 's are not only uncorrelated but independent random variables. Then the remaining finite sum is a sum of weighted chi-square random variables, independent, central and with one degree of freedom; that is, the remaining sum is a quadratic form in independent standard normal random variables that can be dealt with using numerical methods (confer, e.g., S. Kotz, N.L. Johnson and D.W. Boyd (1967)).

METHOD 3. Use Monte-Carlo techniques.

It is obvious that methods 2 and 3 provide only numerical approximations to the distribution of T while method 1, when available, gives the exact distribution of T . All three methods will be demonstrated in

in the sequel. Moreover, we will establish a method based on the Paley-Wiener representation of a Wiener process which does not appear to have been investigated in this context before. The advantage of this last method is that it automatically leads to independent $N(0,1)$ -random variables in (3.1.5.).

3.1.1. Distribution of L_2 -Norm Functionals of Brownian Bridges

In the present subsection we are interested in finding the distribution, exactly or approximately, of the random quantities in (3.1.0.). We start with the classical Paley-Wiener representation of a standard Wiener process $W = \{W(t): t \geq 0\}$ (confer Lemma 1.1.7. of Chapter 1)

$$(3.1.6.) \quad W(t) = y_0 t + \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin k\pi t}{k\pi} y_k \quad \text{for } 0 \leq t \leq 1.$$

Here the convergence of the series is uniform and absolute in $[0,1]$ and $(y_k)_{k \geq 0}$ denotes a sequence of independent identically distributed standard normal random variables. Since for a Brownian bridge $B = \{B(t): 0 \leq t \leq 1\}$ we know that (confer Remark 1.1.6., in particular (1.1.1.))

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

(3.1.6.) directly implies

$$(3.1.7.) \quad B(t) = \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin k\pi t}{k\pi} y_k \quad \text{for } 0 \leq t \leq 1.$$

From (3.1.7.) $\int_0^1 B^2(t) dt$ can be evaluated immediately. We have

$$(3.1.8.) \quad \begin{cases} \int_0^1 B^2(t) dt = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{y_k}{k\pi} \frac{y_m}{m\pi} 2 \int_0^1 \sin k\pi t \sin m\pi t dt \\ = \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} y_k^2 \end{cases}$$

because $2 \int_0^1 \sin k\pi t \sin m\pi t dt = \delta_{km}$, where $\delta_{km} = \begin{cases} 1 & \text{if } k = m \\ 0 & \text{if } k \neq m \end{cases}$ is the Kronecker symbol.

Therefore,

$$\int_0^1 B^2(t) dt = \lim_{K \rightarrow \infty} \sum_{k=1}^K \frac{1}{(k\pi)^2} Y_k^2,$$

which agrees with the result found earlier in Example 3.1.7. except for the fact that the Y_k 's are independent random variables while the Z_k 's in Example 3.1.7. were only uncorrelated. I.e. we have just seen that the distribution of the square integral of a Brownian bridge is equal to the limit of the distributions of sums of independent identically distributed central chi-square random variables with one degree of freedom, weighted by the eigenvalues $(k\pi)^{-2}$ of the operator induced by the Hilbert-Schmidt kernel $K_{B_1}(s, t) = s \wedge t - st$, $0 \leq s, t \leq 1$. Thus, truncating

$$\sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} Y_k^2$$

at a suitably large value K , we obtain a quadratic form in independent identically distributed standard Gaussian variates as an arbitrarily close approximation for $\int_0^1 B^2(t) dt$. Easily accessible methods for the computation of

$$P\left(\sum_{k=1}^K \frac{Y_k^2}{(k\pi)^2} \leq y\right)$$

may be found in S. Kotz, N.L. Johnson and D.W. Boyd (1967). Tabulated values for various $0.2480 \leq y \leq 1.16786$ are contained in Table 1 of T.W. Anderson and D.A. Darling (1952) (see also Table 4 in Subsection 2.3.2. earlier).

The following theorem which gives the distribution of $\int_0^1 \frac{B^2(t)}{t(1-t)} dt$ is also due to T.W. Anderson and D.A. Darling (1952), confer in particular Example 2 there.

Theorem 3.1.8.

$$\int_0^1 \frac{B^2(t)}{t(1-t)} dt = \sum_{k=1}^{\infty} \frac{1}{k(k+1)} Z_k^2,$$

where the Z_k 's are independent identically distributed $N(0,1)$ -random variables. The corresponding eigenfunctions f_k are given by

$$f_k(t) = 2 \sqrt{\frac{2k+1}{k(k+1)}} \sqrt{t(1-t)} p_k'(2t-1),$$

where p_k is the Legendre polynomial of degree k and p_k' is its first derivative. The cumulative distribution function $a_2(z) =$

$P\left(\int_0^1 B^2(t)(t(1-t))^{-1} dt \leq z\right)$ is given by

$$a_2(z) = \frac{\sqrt{2\pi}}{z} \sum_{k=0}^{\infty} \binom{-1/2}{k} (4k+1) e^{-(4k+1)^2 (\pi^2)/(8z)} \\ \cdot \int_0^{\infty} e^{z/(8(\omega^2+1)) - ((4k+1)^2 \pi^2 \omega^2)/(8z)} d\omega.$$

We recall that the Legendre polynomial p_k of degree k is defined by Rodrigues' formula as

$$p_k(t) = \frac{1}{2^k k!} \frac{d^k}{dt^k} [(t^2-1)^k] \text{ for } k=1,2,\dots \text{ and } 0 \leq t \leq 1;$$

confer also E. Kreyszig (1978) pp. 176-180.

T.W. Anderson and D.A. Darling (1954) compute the significance points for the three levels 0.10, 0.05, and 0.01 for $\int_0^1 \frac{B^2(t)}{t(1-t)} dt$ (see also Table 4 in Subsection 2.3.2.). So far we have determined the distributions of $\int_0^1 B^2(t) dt$ and of $\int_0^1 \frac{B^2(t)}{t(1-t)} dt$. Next, we consider the distribution of the following general kind of functionals:

$$\int_0^1 X^2(t) dt, \text{ where } X(t) = \sqrt{\psi(t)} B(t) \text{ for } t \in [0,1].$$

Here $K_X(s,t) = \text{Cov}(X(s), X(t)) = \sqrt{\psi(s)\psi(t)} (s \wedge t - st)$ for $0 \leq s, t \leq 1$.

In order to find the Karhnen-Love expansion (3.1.5.) for $X =$

$\{X(t): 0 \leq t \leq 1\}$, we must solve the integral equation

$$(3.1.9.) \quad \lambda f(t) = \int_0^1 \sqrt{\psi(s)\psi(t)} (s \wedge t - st) f(s) ds \text{ for } t \in [0,1]$$

subject to $f(0) = f(1) = 0$.

Put $f(t) = \sqrt{\psi(t)} h(t)$ for $h \in L^2[0,1]$, then differentiation with respect to t yields

$$\begin{aligned} \lambda h'(t) &= \frac{d}{dt} \left\{ \int_0^t s\psi(s)h(s)ds + t \int_t^1 \psi(s)h(s)ds - t \int_0^1 s\psi(s)h(s)ds \right\} \\ &= \int_t^1 \psi(s)h(s)ds - \int_0^t s\psi(s)h(s)ds \end{aligned} \quad \text{and thus}$$

$$(3.1.10.) \quad \lambda h''(t) = -\psi(t)h(t). \quad \circ$$

Since for representation (3.1.5.) the eigenfunctions f_k of (3.1.9.) are of no particular interest, it suffices to solve (3.1.10.) in order to obtain the eigenvalues $\lambda_k > 0$ of (3.1.9.). Unfortunately in general this knowledge is not of much help for finding the λ_k 's as we shall see now. The second order linear differential equation (3.1.10.) is closely related to the Riccati differential equation

$$y'(t) + y^2(t) = -\frac{\psi(t)}{\lambda}, \quad 0 \leq t \leq 1,$$

because taking $gh'(t) = h(t)y(t)$ we have

$$h''(t) = h'(t)y(t) + h(t)y'(t),$$

$\lambda h''(t) + \psi(t)h(t) = \lambda h'(t)y(t) + \lambda h(t)y'(t) + \psi(t)h(t) = \{\lambda y^2(t) + \lambda y'(t) + \psi(t)\}h(t)$, and thus (3.1.10.) turns into

$$(3.1.11.) \quad y'(t) + y^2(t) = -\frac{\psi(t)}{\lambda}$$

If $\psi(t) = ct^\alpha$ for some $\alpha > -2$ and $c \in \mathbb{R}$, then the solutions of (3.1.11.) can either be represented by Bessel functions or they do not have a close form solution at all (confer E. Kamke (1951), A.48.).

We note that the condition $\alpha > -2$ originates from the requirement that $\int_0^1 \psi(t)B^2(t)dt$ be almost surely finite. For the definition of Bessel functions (of the first kind) we refer to Subsection 3.1.2.

For other functions ψ , there is little hope of finding any closed form solution of (3.1.11.) at all. Going back to (3.1.10.) itself, we can say that most reasonable functions ψ , if they produce close form solutions at all, lead to extremely involved solutions such as Whittaker or Matthieu functions or functions of Bessel functions. This is demonstrated in Appendix B for the choice $\psi(t) = t(1-t)$.

However, the distribution of $\int_0^1 x^2(t)dt$ may in many cases still be found by integrating a properly weighted representation (3.1.7.) because in the present situation

$$\int_0^1 x^2(t)dt = \int_0^1 \psi(t)B^2(t)dt = \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} 2 \int_0^1 \psi(t) \sin k\pi t \sin m\pi t dt$$

and hence real numbers a_{km} , $k, m = 1, 2, \dots$, can be determined by elementary integral calculus such that

$$\int_0^1 x^2(t)dt = \lim_{K \rightarrow \infty} \sum_{k,m=1}^K a_{km} Y_k Y_m$$

From here tables for $P(\int_0^1 X^2(t) dt \leq y)$ may be computed by letting K get larger and larger, up to a point of stability of y . (Confer also Appendix A.) In the following example the above described method is applied.

Example 3.1.9.

Suppose we wish to neglect the tail behaviour of our test statistic while at the same time stressing the behaviour in the middle section. One way of doing so is to choose $\psi(t) = c \sin \pi t$ for $t \in [0,1]$ and $c > 1$ a suitable constant.

Figure 1.

According to I.S. Gradshteyn and I.M. Ryzhik (1965), formula 2.533.2 with $a = \pi$, $b = k\pi$, and $c = m\pi$, we have

$$\begin{aligned} & \int_0^1 \sin \pi t \sin k\pi t \sin m\pi t \, dt \\ &= -\frac{1}{4} \left[\frac{\cos(1-k+m)\pi}{(1-k+m)\pi} + \frac{\cos(k+m-1)\pi}{(k+m-1)\pi} + \frac{\cos(1+k-m)\pi}{(1+k-m)\pi} - \frac{\cos(1+k+m)\pi}{(1+k+m)\pi} \right]_0^1 \\ &= -\frac{1}{4\pi} \left[\frac{(-1)^{1-(k-m)} - 1}{(1-k+m)} + \frac{(-1)^{(k+m)-1} - 1}{(k+m-1)} + \frac{(-1)^{-1-(k-m)} - 1}{(1+k-m)} + \frac{(-1)^{k+m} - 1}{(1+k+m)} \right] \end{aligned}$$

$$= \frac{1}{2\pi} \begin{cases} 0 & \left\{ \begin{array}{l} \text{if } k, m \text{ odd, } m \text{ even or} \\ \text{if } k \text{ even, } m \text{ odd} \end{array} \right. \\ \frac{2m(1+k-m)(1+k+m) + 2m(1-k+m)(k+m-1)}{(1-k+m)(k+m-1)(1+k-m)(1+k+m)} & \left\{ \begin{array}{l} \text{if } k, m \text{ odd} \\ \text{if } k, m \text{ even} \end{array} \right. \end{cases}$$

$$= \frac{4}{\pi} \begin{cases} 0 & \left\{ \begin{array}{l} \text{if } k \text{ odd, } m \text{ even or} \\ \text{if } k \text{ even, } m \text{ odd} \end{array} \right. \\ \frac{k\pi}{(1-k+m)(k+m-1)(1+k-m)(1+k+m)} & \left\{ \begin{array}{l} \text{if } k, m \text{ odd} \\ \text{if } k, m \text{ even} \end{array} \right\} \end{cases}$$

whenever the denominator is not equal to zero, i.e. whenever $m \neq k+1$.

Therefore, we obtain

$$\int_0^1 x^2(t) dt = 2c \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \int_0^1 \sin \pi t \sin k\pi t \sin m\pi t dt$$

$$= \left(\frac{2}{\pi}\right)^3 c \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k Y_m}{(1-k+m)(k+m-1)(1+k-m)(1+k+m)}$$

$$+ 4c \sum_{k=2}^{\infty} \frac{Y_k}{k\pi} \frac{Y_{k-1}}{(k-1)\pi} \int_0^1 \sin \pi t \sin k\pi t \sin(k-1)\pi t dt$$

Next, we evaluate $\int_0^1 \sin \pi t \sin k\pi t \sin(k-1)\pi t dt$ through integration by parts:

$$\int_0^1 \sin \pi t [\sin k\pi t \sin(k-1)\pi t] dt$$

$$= \sin \pi t \left\{ \frac{\sin \pi t}{2\pi} - \frac{\sin(2k-1)\pi t}{2(2k-1)\pi} \right\} \Big|_0^1 - \pi \int_0^1 \cos \pi t \left\{ \frac{\sin \pi t}{2\pi} - \frac{\sin(2k-1)\pi t}{2(2k-1)\pi} \right\} dt$$

$$= \frac{1}{2(2k-1)} \int_0^1 \sin(2k-1)\pi t \cos \pi t dt - \frac{1}{2} \int_0^1 \sin \pi t \cos \pi t dt.$$

Now apply formula 2.532.2, p.140, of Gradshteyn and Ryzhik with $b=d=0$

and $a = (2k-1)\pi, c = \pi$ to $\int_0^1 \sin(2k-1)\pi t \cos \pi t dt$:

$$\begin{aligned} \int_0^1 \sin(2k-1)\pi t \cos \pi t dt &= -\frac{\cos 2(k-1)\pi t}{4(k-1)\pi} - \frac{\cos 2k\pi t}{4k\pi} \Big|_0^1 \\ &= -\frac{(-1)^{2(k-1)} - 1}{4(k-1)\pi} - \frac{(-1)^{2k} - 1}{4k\pi} = 0. \end{aligned}$$

Applying 2.532.5, p.140, Gradshteyn and Ryzhik with $b=d=0$ and

$a=c=\pi$ to $\int_0^1 \sin \pi t \cos \pi t dt$ yields

$$\int_0^1 \sin \pi t \cos \pi t dt = \frac{t}{2} \sin 0 - \frac{\cos 2\pi t}{4\pi} \Big|_0^1 = 0,$$

whence we have

$$c \int_0^1 \sin \pi t B^2(t) dt = \left(\frac{2}{\pi}\right)^3 c \sum_{\substack{k=1 \\ k,m \text{ odd} \\ \text{or } k,m \text{ even}}}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k Y_m}{(1-k+m)(k+m-1)(1+k-m)(1+k+m)}.$$

Therefore,

$$c \int_0^1 \sin \pi t B^2(t) dt = \lim_{K \rightarrow \infty} \sum_{k,m=1}^K a_{km} Y_k Y_m, \text{ where}$$

$$a_{km} = \begin{cases} \frac{(8c)/\pi^3}{(1-k+m)(k+m-1)(1+k-m)(1+k+m)} & \text{if } k,m \text{ odd or } k,m \text{ even} \\ 0 & \text{otherwise.} \end{cases}$$

Table 5. below displays upper percentage points for various values

α , $0.001 \leq \alpha \leq 0.01$. Here the constant c was taken to be 1.

α	0.1	0.075	0.05	0.025	0.01	0.005	0.001
$x_n(50)$	0.27989899	0.31971859	0.37702342	0.47743749	0.61356736	0.71837460	0.96585925
$x_n(90)$	0.28046368	0.32028327	0.37758808	0.47800209	0.6143185	0.71893900	0.96642341

Table 5. $(1-\alpha) \cdot 100\%$ - Fractiles of $\int_0^1 \sin \pi t B^2(t) dt$.

Table 5 was computed using the following facts, notations and subroutines from the NAG library:

- $A := (a_{km})_{k,m=1}^K$, where a_{km} is as defined on page 133;
- find eigenvalues of A , for given K , via FOZAAF subroutine of NAG library;
- sort eigenvalues (all positive) by magnitude and call them $\lambda_1, \lambda_2, \dots, \lambda_K$;
- $Q := \sum_{i=1}^K \lambda_i Z_i^2$, where $(Z_i)_{i=1}^K$ is a sequence of independent identically distributed $N(0,1)$ -random variables; then

$$P\left\{\sum_{k,m=1}^K a_{km} Y_k Y_m \geq x\right\} = P\left\{\sum_{i=1}^K \lambda_i Z_i^2 \geq x\right\};$$
- $P\{Q \geq x\} = P\{\chi_{h'}^2 \geq y\}$ (confer J.P. Imhof (1961), formula (4.1)), where $h' = c_2^3/c_3^2$ and $y = (x - c_1)h'(c_2)^{1/2} + h'$. Here c_1 denotes the 1th cumulant of Q ;
- use NAG library subroutine GO1BCF and linear interpolation to find y above for given values of α and for $K = 50$ and $K = 90$;
- the main program was written in PASCAL.

3.1.2. Distribution of Other Gaussian Processes in L_2 -Norm

The main goal of this subsection is to find the distributions of $\int_0^1 w^2(t)dt$ and of the random functionals in (3.1.1.). We will also consider the distribution of

$$\int_0^1 w^2(t)\psi(t)dt, \text{ where } \psi(t) = at^\alpha \text{ for constant } a \in \mathbb{R}, \alpha > -2.$$

The functions ψ for $\alpha \geq -1$ provide examples of the weight functions g considered in (3.1.1.). We begin with the distribution of the functionals $\int_0^1 w^2(t)\psi(t)dt$. To find the eigenvalues λ_k required in representation (3.1.5.), we will use a technique established by I.B. MacNeill (1974) which is based on the solution of a modified differential equation. In addition, for the case $\alpha = 0$, we will first present the approach discussed in the previous subsection which utilizes the Paley-Wiener representation of a Wiener process and direct integration. This approach is again displayed here for two reasons:

- (a) to show the extreme shortness of the solution
- (b) to provide a different representation of the solution.

While the "Paley-Wiener-approach" has the advantage of a rather short proof, the beauty of the solution suggested by MacNeill lies in the fact that it creates a unified method for the treatment of a whole class of problems. Beginning with the distribution of $\int_0^1 w^2(t)dt$, recall

$$(3.1.6.) \quad w(t) = y_0 t + \sqrt{2} \sum_{k=1}^{\infty} \frac{\sin k\pi t}{k\pi} y_k \quad \text{for } 0 \leq t \leq 1.$$

Therefore,

$$w^2(t) = (Y_0 t)^2 + 2^{3/2} Y_0 t \sum_{k=1}^{\infty} \frac{\sin k\pi t}{k\pi} Y_k + 2 \left(\sum_{k=1}^{\infty} \frac{\sin k\pi t}{k\pi} Y_k \right)^2.$$

Thus, by (3.1.8.) we obtain

$$\int_0^1 w^2(t) dt = \frac{1}{3} Y_0^2 + 2^{3/2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \int_0^1 t \sin k\pi t dt + \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} Y_k^2,$$

whence

$$(3.1.12.) \quad \int_0^1 w^2(t) dt = \frac{1}{3} Y_0^2 + 2^{3/2} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{(k\pi)^2} (Y_0 Y_k) + \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} Y_k^2.$$

This can again be expressed as a limit of quadratic forms in independent $N(0,1)$ -random variables:

$$\int_0^1 w^2(t) dt = \lim_{K \rightarrow \infty} \sum_{k,m=0}^K a_{km} Y_k Y_m, \quad \text{where}$$

$$a_{km} = \begin{cases} 1/3 & \text{if } k = m = 0 \\ (k\pi)^{-2} & \text{if } k = m \geq 1 \\ 0 & \text{if } k, m \geq 1 \text{ and } k \neq m \\ \frac{(-1)^{k-1}}{(k\pi)^2} \sqrt{2} & \text{if } k \geq 1 \text{ and } m = 0 \\ \frac{(-1)^{m-1}}{(m\pi)^2} \sqrt{2} & \text{if } m \geq 1 \text{ and } k = 0. \end{cases}$$

Using the PASCAL program described on page 134, now for a_{km} as above, upper percentage points were found for the random variable in (3.1.12.). They are displayed in Table 6 below (see page 138).

In the following we will show that

$$E \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) \rightarrow E \left(\int_0^1 w^2(t) dt \right) = \frac{1}{2} \quad \text{as } K \rightarrow \infty$$

and

$$\text{Var} \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) \rightarrow \text{Var} \left(\int_0^1 W^2(t) dt \right) = \frac{1}{3} \quad \text{as } K \rightarrow \infty.$$

G.R. Shorack and J.A. Wellner (1986), pp.212-213, provide us with the following general principle. Let again $T_X = \int_0^1 X^2(t) dt$, where $X = \{X(t): 0 \leq t \leq 1\}$ is a stochastic process with square-integrable covariance function K_X such that $\int_0^1 K_X(t,t) dt < \infty$. Suppose $T_X = \sum_{k=1}^{\infty} \lambda_k Z_k^2$ as in (3.1.5.). Then

$$(3.1.13.) \quad \begin{cases} ET = \int_0^1 K_X(t,t) dt = \sum_{k=1}^{\infty} \lambda_k \\ \text{Var } T = 2 \int_0^1 \int_0^1 K_X^2(s,t) ds dt = 2 \sum_{k=1}^{\infty} \lambda_k^2 \end{cases} \quad \text{and}$$

Application of (3.1.13.) to $X(t) = W(t)$ yields

$$E \int_0^1 W^2(t) dt = \int_0^1 K_W(t,t) dt = \int_0^1 t dt = \frac{1}{2} \quad \text{and}$$

$$\begin{aligned} \text{Var} \int_0^1 W^2(t) dt &= 2 \int_0^1 \int_0^1 K_W^2(s,t) ds dt = 2 \int_0^1 \left\{ \int_0^t s^2 ds + t^2 \int_t^1 ds \right\} dt \\ &= 2 \int_0^1 \left(t^2 - \frac{2}{3} t^3 \right) dt = \frac{1}{3}. \end{aligned}$$

In addition, we have

$$(3.1.14.) \quad \begin{cases} E \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) = \sum_{k=0}^K \sum_{m=0}^K a_{km} E(Y_k Y_m) = \sum_{k=0}^K a_{kk} \\ \text{Var} \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) = 2 \sum_{k,m=0}^K a_{km}^2 \end{cases} \quad \text{and}$$

(1- α).100% - Fractiles of $\int_0^1 w^2(t)dt$.

α	0.1	0.075	0.05	0.025	0.01	0.005	0.001
$\hat{x}_1(90)$	1.19607904	1.38373053	1.6541815	2.13098687	2.77908387	3.27926320	4.46302669
x_1	1.196	-	1.656	2.135	2.78	-	-

Table 6.

The values $\hat{x}_1(90)$ given in Table 6 above were generated using the computer routine described on page 134. The reference values x_1 are taken from E.D. Rothman and M. Woodroffe (1972) who also conducted a Monte Carlo simulation study to confirm their results. Table 6 clearly demonstrates the accuracy of our Paley-Wiener approach.

It should be noted at this point that the accuracy of the x_1 values in Table 6 is remarkable considering that the computer program which generated these values is based on a simple three-moment-central-chi-square approximation of a quadratic form. Presumably an even better approximation could be obtained using a four-moment Pearson curve fit. However, this cannot be done easily since Pearson curve charts are only available (confer E.S. Pearson and H.O. Hartley (1976)) for $\sqrt{\bar{\epsilon}_1} = 2.00$ while for the random variable considered in Table 6 $\sqrt{\bar{\epsilon}_1} = 2.7$. Here $\sigma_1 = \mu_3^2/\mu_2^3$, where μ_i is the i^{th} central moment of the quadratic form in question.

For the last statement see T. DeWet and J.H. Venter (1973) who consider a similar formula for variances in a more general case. From here we obtain

$$E \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) = \sum_{k=0}^K a_{kk} = \frac{1}{3} + \frac{1}{\pi^2} \sum_{k=1}^K \frac{1}{k^2}$$

and hence

$$E \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) \rightarrow \frac{1}{3} + \frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} \quad \text{as } K \rightarrow \infty.$$

Now let $\zeta(z)$ denote Riemann's zeta function (for $\text{Re } z > 1$), i.e.

$$\zeta(z) = \sum_{k=1}^{\infty} k^{-z} \quad \text{for } z \in \mathbb{C} \text{ and } \text{Re } z > 1, \text{ and let } B_m \text{ be the } m^{\text{th}}$$

Bernoulli number, then

$$\frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{1}{\pi^2} \zeta(2) = \frac{1}{\pi^2} \left(\frac{2\pi^2}{2!} |B_2| \right) = |B_2| = \frac{1}{6}.$$

(Confer I.S. Gradshteyn and I.M. Ryzhik (1965), 9.542.1 and 9.71.)

Therefore, $\frac{1}{3} + \frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{1}{3} + \frac{1}{6} = \frac{1}{2}$, implying that

$$\lim_{K \rightarrow \infty} E \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) = E \int_0^1 W^2(t) dt = \frac{1}{2}.$$

Similarly, we obtain (again using Gradshteyn and Ryzhik where indicated by (GR)):

$$\begin{aligned} E \left(\sum_{k,m=0}^K a_{km} Y_k Y_m \right) &= 2 \sum_{k=0}^K \sum_{m=0}^K a_{km}^2 \\ &= 2 \left\{ a_{00}^2 + \sum_{m=1}^K a_{0m}^2 + \sum_{k=1}^K a_{k0}^2 + \sum_{k=1}^K \sum_{m=1}^K a_{km}^2 \right\} \\ &= \frac{2}{9} + 4 \sum_{k=1}^K \left(\sqrt{2} \frac{(-1)^{k-1}}{(k\pi)^2} \right)^2 + 2 \sum_{k=1}^K \left(\frac{1}{(k\pi)^2} \right)^2 \\ &= \frac{2}{9} + \frac{10}{\pi^4} \sum_{k=1}^K \frac{1}{k^4} \end{aligned}$$

$$\begin{aligned}
& - \left[\frac{(-1)^{k-m}}{((k-m)\pi)^2} - \frac{(-1)^{k+m}}{((k+m)\pi)^2} \right] \Bigg\} \\
& - 8 \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \left\{ \frac{1}{4} - \frac{1}{6} \right\} \\
& = \frac{2}{15} Y_0^2 + 2^{5/2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \left\{ 6 \left(\frac{(-1)^k + 1}{(k\pi)^3} \right) \right\} + \frac{1}{3} \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \\
& - 8 \sum_{k=1}^{\infty} \sum_{\substack{m=1 \\ k \neq m}}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \left\{ \frac{(-1)^{k+m} + 1}{2(k+m)^2 \pi^2} - \frac{(-1)^{k-m} + 1}{2(k-m)^2 \pi^2} \right\} \\
& = \frac{2}{15} Y_0^2 + 48\sqrt{2} Y_0 \sum_{k=1}^{\infty} \frac{Y_{2k}}{(2k\pi)^4} + \frac{1}{3} \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \\
& + 32 \sum_{k=1}^{\infty} \sum_{\substack{m=1 \\ k \neq m \\ (k+m) \text{ even}}}^{\infty} \frac{Y_k Y_m}{((k-m)^2 \pi^2)^2}
\end{aligned}$$

Therefore, we can write even the functional $\int_0^1 \dot{r}^2(t) dt$ again as a limit of quadratic forms in independent $N(0,1)$ -random variables, namely as

$$(3.1.21.) \quad \int_0^1 \dot{r}^2(t) dt = \lim_{K \rightarrow \infty} \sum_{k,m=0}^K a_{km} Y_k Y_m, \quad \text{where}$$

$$a_{km} = \begin{cases} \frac{2}{15} & \text{if } k = m = 0 \\ \frac{1}{3(k\pi)^2} & \text{if } k = m \geq 1 \\ \frac{24\sqrt{2}}{(k\pi)^4} & \text{if } k \geq 1, k \text{ even}, m = 0 \\ \frac{24\sqrt{2}}{(m\pi)^4} & \text{if } m \geq 1, m \text{ even}, k = 0 \end{cases}$$

and differentiating again gives

$$\lambda h''(t) = -\psi(t)h(t).$$

In order to write down the solutions of the class of differential equations

$$(3.1.16.) \quad h''(t) + \left(\frac{a}{\lambda}\right)t^\alpha \cdot h(t) = 0 \quad (\text{subject to the constraints } h(0)=h'(1)=0)$$

we first introduce the notion of Bessel functions: A function $J(\nu; z)$ is called a Bessel function of the first kind if it satisfies the differential equation

$$\frac{d^2 z_\nu}{dz^2} + \frac{1}{z} \frac{dz_\nu}{dz} + \left(1 - \frac{\nu^2}{z^2}\right) z_\nu = 0.$$

Both arguments of J may be complex. If the second argument of J satisfies $|\arg z| < \pi$, then $J(\nu; z)$ has the series representation

$$J(\nu; z) = \frac{z^\nu}{2^\nu} \sum_{k=0}^{\infty} (-1)^k \frac{z^{2k}}{2^{2k} k! \Gamma(\nu+k+1)}.$$

This and the following representations for special cases of Bessel functions are due to I.S. Gradshteyn and I.M. Ryzhik (1965), Sections 8.4 and 8.5. I.B. MacNeill (1974) states that the solutions of (3.1.16.), i.e. the eigenfunctions of (3.1.15.), are

$$h_k(t) = A\sqrt{t} J\left(\frac{1}{\alpha+2}; 2\left(\frac{a}{\lambda_k}\right)^{1/2} t^{(\alpha+2)/2} (\alpha+2)^{-1}\right),$$

where A is a normalizing constant chosen such that $\|h_k\| = 1$ and where

$$(3.1.17.) \quad \lambda_k = \frac{4a}{(\alpha+2)^2} j^{-2}\left(-\frac{\alpha+1}{\alpha+2}; k\right)$$

is the k^{th} largest eigenvalue corresponding to h_k with $j(v, k)$ denoting the k^{th} zero of $J(v; z)$.

For the special cases $\alpha = 0, -\frac{4}{3}$, and $-\frac{8}{5}$ the eigenfunctions h_k take on the following simple forms:

$$\alpha = 0, \quad h_k(t) = A\sqrt{t} J\left(\frac{1}{2}; 2\left(\frac{a}{\lambda_k}\right)^{1/2} t\right) = \{A \cdot \sqrt{\frac{2}{\pi}} \sin \sqrt{\frac{a}{\lambda_k}} t\} / \left(\frac{a}{\lambda_k}\right)^{1/4};$$

$$\alpha = -\frac{4}{3}, \quad h_k(t) = A\sqrt{t} J\left(\frac{3}{2}; 3\left(\frac{a}{\lambda_k}\right)^{1/2} t^{1/3}\right) = A\sqrt{\frac{2}{\pi}} \left(\frac{\lambda_k}{a}\right)^{1/4} \left\{ \frac{\sin z}{z} - \cos z \right\};$$

$$\alpha = -\frac{8}{5}, \quad h_k(t) = A\sqrt{t} J\left(\frac{5}{2}; 5\left(\frac{a}{\lambda_k}\right)^{1/2} t^{1/3}\right) = A\sqrt{\frac{2}{\pi}} \left(\frac{3}{z^2} - 1\right) (\sin z - \frac{3}{z} \cos z),$$

$$\text{where } z = \sqrt{\frac{a}{\lambda_k}} t.$$

Summarizing the results shown so far in this section we can say that the distribution of

$$\int_0^1 w^2(t) \psi(t) dt = a \int_0^1 w^2(t) t^\alpha dt$$

has been established to be

$$\sum_{k=1}^{\infty} \lambda_k z_k^2 = \frac{4a}{(\alpha+2)^2} \sum_{k=1}^{\infty} j^{-2}\left(-\frac{\alpha+1}{\alpha+2}; k\right) z_k^2.$$

Tables for selected percentage points of $P\left(\int_0^1 (at^\alpha) w^2(t) dt \leq y\right)$ are contained in I.B. MacNeill (1974) for various values of α as well as in I.B. MacNeill (1971) and in E.D. Rothman and M. Woodroffe (1972) for $\alpha = 0$. Corollary 2 and Theorem 3 of the latter article are summarized in the following theorem.

Theorem 3.1.10.

$$(a) \int_0^1 w^2(t) dt = \sum_{k=1}^{\infty} \frac{4}{(2k-1)^2 \pi^2} z_k^2 \text{ in distribution, where}$$

$z_1, z_2, \dots$ are independent $N(0,1)$ -random variables.

(b) The distribution function H of $\int_0^1 w^2(t) dt$ is given as

$$H(x) = 2^{3/2} \sum_{j=0}^{\infty} \binom{-1/2}{j} (1 - \Phi(\frac{4j+1}{2\sqrt{x}})) \text{ for } x > 0, \text{ where } \Phi \text{ denotes}$$

the cumulative distribution function of a standard normal variable.

Our next aim is to find the (approximate) distribution of

$$\int_0^1 \Gamma^2(t) dt, \text{ where}$$

$$\Gamma(t) = (1-t)W(t) + tW(1) - tW(t).$$

The covariance kernel K_{Γ} of Γ was computed in Example 3.1.4.(b)

to be $K(s,t) = K_{\Gamma}(s,t) = (1-2s)(1-2t)(s \wedge t) + (3-2s-2t)st$. Hence we wish to solve the integral equation

$$(3.1.18.) \quad \lambda f(t) = \int_0^1 \{ (1-2s)(1-2t)(s \wedge t) + (3-2s-2t)st \} f(s) ds \text{ for } 0 \leq t \leq 1$$

in order to find the eigenvalues $\lambda_k > 0$ for representation (3.1.15.)

applied to $X = \Gamma$. Differentiation of (3.1.18.) with respect to t yields

$$\begin{aligned} \lambda f'(t) &= \frac{d}{dt} \left\{ (1-2t) \int_0^t (1-2s) s f(s) ds + (1-2t)t \int_t^1 (1-2s) f(s) ds \right\} \\ &+ \int_0^1 s(3-2s) f(s) ds - 4t \int_0^1 s f(s) ds \\ &= -2 \int_0^t (1-2s) s f(s) ds + (1-4t) \int_t^1 (1-2s) f(s) ds \\ &+ \int_0^1 (3-2s) s f(s) ds - 4t \int_0^1 s f(s) ds; \end{aligned}$$

and differentiating again

$$\begin{aligned}\lambda f''(t) &= -2(1-2t)tf(t) - 4\int_t^1 (1-2s)f(s)ds - (1-4t)(1-2t)f(t) - 4\int_0^1 sf(s)ds \\ &= -(1-2t)^2 f(t) - 4\int_t^1 (1-2s)f(s)ds - 4\int_0^1 sf(s)ds.\end{aligned}$$

This shows that (3.1.18.) cannot be transformed into a linear differential equation of order 2. We note that the last equation does not depend on how we split up the integrals before starting the differentiations. Differentiating a third time, we obtain

$$\lambda f'''(t) = 8(1-2t)f(t) - (1-2t)^2 f'(t).$$

One attempt to find the exact solutions λ of (3.1.18.) is shown in the following. Consider the boundary value problem

$$(3.1.19.) \quad \lambda f'''(t) + (1-2t)^2 f'(t) - 8(1-2t)f(t) = 0, \quad f(0) = f(1) = 0.$$

Assume a solution f of the form $f(t) = \sum_{n=0}^{\infty} a_n t^n$ exists. It follows from the general theory of linear differential equations of order n that such solutions do in fact exist. Then the boundary condition $f(0) = 0$ implies $a_0 = 0$. If $f(t) = \sum_{n=1}^{\infty} a_n t^n$, then $f'(t) = \sum_{n=1}^{\infty} n a_n t^{n-1}$, $f''(t) = \sum_{n=2}^{\infty} n(n-1) a_n t^{n-2}$, and $f'''(t) = \sum_{n=3}^{\infty} n(n-1)(n-2) a_n t^{n-3}$. Therefore, (3.1.19.) holds if

$$\lambda \sum_{n=3}^{\infty} n(n-1)(n-2) a_n t^{n-3} + (1-4t+4t^2) \sum_{n=1}^{\infty} n a_n t^{n-1} - 8(1-2t) \sum_{n=1}^{\infty} a_n t^n = 0$$

and the latter statement is equivalent to

$$\begin{aligned}\sum_{n=2}^{\infty} \{ \lambda(n+3)(n+2)(n+1) a_{n+3} + (n+1) a_{n+1} - 4(n+2) a_n + 4(n+3) a_{n-1} \} t^n \\ + (6\lambda a_3 + a_1) + 2(12\lambda a_4 + a_2 - 6a_1)t = 0,\end{aligned}$$

which can be written as a system of equations by setting the coefficients

of the different powers of t equal to zero:

$$(3.1.20.) \begin{cases} 6\lambda a_3 + a_1 = 0. \\ 12\lambda a_4 + a_2 - 6a_1 = 0 \\ \lambda(n+3)(n+2)(n+1)a_{n+3} + (n+1)a_{n+1} - 4(n+2)a_n + 4(n+3)a_{n-1} = 0 \end{cases}$$

for each $n \geq 2$. From here we obtain the following statements

$$\lambda a_3 = -\frac{a_1}{6}$$

$$\lambda a_4 = \frac{a_1}{2} - \frac{a_2}{12}$$

$$\lambda a_5 = \frac{1}{60} \left\{ 16a_2 + \left(\frac{1}{2\lambda} - 20 \right) a_1 \right\}$$

$$\lambda a_6 = \frac{1}{120} \left\{ \left(\frac{1}{3\lambda} - 24 \right) a_2 - \frac{16}{3\lambda} a_1 \right\}$$

$$\lambda a_7 = \frac{1}{210} \left\{ -\left(\frac{4}{3} + \frac{2}{\lambda} \right) a_2 + \left(-\frac{1}{24\lambda^2} + \frac{55}{3\lambda} \right) a_1 \right\}$$

$$\lambda a_8 = \frac{1}{336} \left\{ \left(-\frac{1}{60\lambda^2} + \frac{34}{3\lambda} \right) a_2 + \left(\frac{1}{2\lambda^2} - \frac{76}{3\lambda} \right) a_1 \right\}$$

⋮

(3.1.20.) defines a recursion formula for the coefficients a_n but since it involves four different a_n 's simultaneously, it is practically impossible to obtain further information about the a_n 's from here without knowing a_1 and a_2 . Careful examination of the remaining boundary condition and of f , f' , f'' , and f''' for the arguments $0, \frac{1}{2}$, and 1 does not prove to be of any help.

Since the previous attempt to determine the eigenvalues λ in (3.1.18.) does not seem to lead to any concrete conclusions, we now

turn back to the Paley-Wiener representation (3.1.6.) of a Brownian motion $W = \{W(t) : 0 \leq t < \infty\}$,

$$(3.1.6.) \quad W(t) = Y_0 t + \sqrt{2} \sum_{k=1}^{\infty} Y_k \frac{\sin k\pi t}{k\pi},$$

which is valid uniformly for $0 \leq t \leq 1$, where $\{Y_k\}_{k \geq 1}$ is a sequence of independent identically distributed $N(0,1)$ -random variables.

Recall that $\Gamma(t) = (1-2t)W(t) + tW(1)$; whence

$$\Gamma(t) = 2t(1-t)Y_0 + \sqrt{2}(1-2t) \sum_{k=1}^{\infty} Y_k \frac{\sin k\pi t}{k\pi}$$

and

$$\int_0^1 \Gamma^2(t) dt = 4Y_0^2 \int_0^1 t^2(1-t)^2 dt + 4\sqrt{2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \int_0^1 t(1-t)(1-2t) \sin k\pi t dt$$

$$+ 2 \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \int_0^1 (1-2t)^2 \sin k\pi t \sin m\pi t dt$$

$$= \frac{2}{15} Y_0^2 + 2^{5/2} \sum_{k=1}^{\infty} \frac{Y_0 Y_k}{k\pi} \int_0^1 (t-3t^2+2t^3) \sin k\pi t dt + \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2}$$

$$- 8 \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \int_0^1 (t-t^2) \sin k\pi t \sin m\pi t dt$$

$$- 8 \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \int_0^1 (t-t^2) (\sin k\pi t)^2 dt$$

$$= \frac{2}{15} Y_0^2 + 2^{5/2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \left\{ \frac{(-1)^{k-1}}{k\pi} - \frac{1}{3} \left(\frac{(-1)^{k-1}}{k\pi} + 2 \frac{[(-1)^{k-1}]}{(k\pi)^3} \right) \right. \\ \left. + 2 \left(\frac{(-1)^{k-1}}{k\pi} + 6 \frac{(-1)^k}{(k\pi)^3} \right) \right\} + \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2}$$

$$- 8 \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \left\{ \left[\frac{(-1)^{k-m}-1}{2(k-m)^2 \pi^2} - \frac{(-1)^{k+m}-1}{2(k+m)^2 \pi^2} \right] \right. \\ \left. - k \neq m \right\}$$

$$\begin{aligned}
& - \left[\frac{(-1)^{k-m}}{((k-m)\pi)^2} - \frac{(-1)^{k+m}}{((k+m)\pi)^2} \right] \\
& - 8 \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \left\{ \frac{1}{4} - \frac{1}{6} \right\} \\
& = \frac{2}{15} Y_0^2 + 2^{5/2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \left\{ 6 \left(\frac{(-1)^k + 1}{(k\pi)^3} \right) \right\} + \frac{1}{3} \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \\
& - 8 \sum_{k=1}^{\infty} \sum_{\substack{m=1 \\ k \neq m}}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \left\{ \frac{(-1)^{k+m} + 1}{2(k+m)^2 \pi^2} - \frac{(-1)^{k-m} + 1}{2(k-m)^2 \pi^2} \right\} \\
& = \frac{2}{15} Y_0^2 + 48\sqrt{2} Y_0 \sum_{k=1}^{\infty} \frac{Y_{2k}}{(2k\pi)^4} + \frac{1}{3} \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \\
& + 32 \sum_{k=1}^{\infty} \sum_{\substack{m=1 \\ k \neq m \\ (k+m) \text{ even}}}^{\infty} \frac{Y_k Y_m}{((k^2 - m^2)\pi^2)^2}
\end{aligned}$$

Therefore, we can write even the functional $\int_0^1 \dot{r}^2(t) dt$ again as a limit of quadratic forms in independent $N(0,1)$ -random variables, namely as

$$(3.1.21.) \quad \int_0^1 \dot{r}^2(t) dt = \lim_{K \rightarrow \infty} \sum_{k,m=0}^K a_{km} Y_k Y_m, \quad \text{where}$$

$$a_{km} = \begin{cases} \frac{2}{15} & \text{if } k = m = 0 \\ \frac{1}{3(k\pi)^2} & \text{if } k, m \geq 1 \\ \frac{24\sqrt{2}}{(k\pi)^4} & \text{if } k \geq 1, k \text{ even}, m = 0 \\ \frac{24\sqrt{2}}{(m\pi)^4} & \text{if } m \geq 1, m \text{ even}, k = 0 \end{cases}$$

$$\begin{cases} \frac{32}{((k^2 - m^2)\pi^2)^2} & \text{if } k, m \geq 1, k \neq m, (k+m) \text{ even} \\ 0 & \text{otherwise.} \end{cases}$$

Tabulated values of $P(\int_0^1 \Gamma^2(t) dt \leq y)$ are given in Table 7 below.

The expected value and the variance of $\int_0^1 \Gamma^2(t) dt$ are easily accessible quantities. They can be computed using formula (3.1.13.) from the beginning of this subsection again. We obtain

$$ET = \int_0^1 \int_0^1 K_{\Gamma}(t, t) dt = \int_0^1 [(1-2t)^2 t + (3-4t)t^2] dt = \int_0^1 t(1-t) dt = \frac{1}{6}$$

and notice in passing that this equals the mean of $\int_0^1 B^2(t) dt$;

$$\begin{aligned} \text{Var } T &= 2 \int_0^1 \int_0^1 K_{\Gamma}^2(s, t) ds dt \\ &= 2 \int_0^1 \int_0^1 (1-2s)^2 (1-2t)^2 (s \wedge t)^2 ds dt \\ &\quad + 4 \int_0^1 \int_0^1 (1-2s)(1-2t)(s \wedge t)(3-2s-2t)(st) ds dt \\ &\quad + 2 \int_0^1 \int_0^1 (3-2s-2t)^2 (st)^2 ds dt, \end{aligned}$$

which is calculated in Appendix C to be $\frac{4}{105}$; thus, $\text{Var } T = \frac{4}{105}$.

, G.R. Shorack and J.A. Wellner (1986), pp.212-213, also present a formula for the computation of the m^{th} semi-invariant of T , namely

$$\kappa_m = 2^{m-1} (m-1)! \int_0^1 K_m(t, t) dt = 2^{m-1} (m-1)! \sum_{k=1}^{\infty} \lambda_k^m,$$

where $K_m(s, t) = \int_0^1 K_{m-1}(s, r) K(r, t) dr$ and where κ_m , the m^{th} semi-invariant of T , is given by the coefficient of $\frac{(i\lambda)^m}{m!}$ in the power

(1- α)-100% - Fractile of $\int_0^1 \Gamma^2(t) dt$.

α	0.1	0.075	0.05	0.025	0.01	0.005	0.001
$x_{(50)}$	0.43148164	0.49466527	0.58551704	0.74455948	0.95996144	1.12569053	1.51678733
$x_{(90)}$	0.43178660	0.49497022	0.5	0.74486441	0.96026635	1.12599541	1.51709215

Table 7.

Table 7 was again obtained using the program described on page 134. It should be recorded here that the original intent was to utilize the methods displayed in Appendix A to generate Tables 5-7. However, our computer routines constructed according to Appendix A indicate that this method only works with acceptable accuracy when the quadratic form under consideration is of dimension 2, 3 or at most 4. Hence the technology of Appendix A could not be applied here.

series expansion of $\log \phi(\lambda)$. Here ϕ denotes the characteristic function of T , i.e. $\phi(\lambda) = E \exp(i\lambda T)$. If the random variables Z_k in (3.1.5.) are independent and identically distributed, then

$$\phi(\lambda) = E e^{i\lambda T} = \prod_{k=1}^{\infty} (1 - 2i\lambda_k \lambda)^{-1/2}.$$

In addition to giving the mean and the variance of T , formula (3.1.13.) also provides us with the information that

$$\sum_{k=1}^{\infty} \lambda_k = \frac{1}{6} \quad \text{and} \quad \sum_{k=1}^{\infty} \lambda_k^2 = \frac{2}{105}.$$

which may be used to obtain a first impression of how good certain numerical approximations to the λ_k 's are.

Finally, we are interested in the distribution of $\int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt$.

Finding the Karhunen-Loève expansion of this functional is practically impossible because the covariance function $K(s, t) = \{s(1-s)t(1-t)\}^{-1/2}$.

$((1-2s)(1-2t)(s \wedge t) + (3-2s-2t)st)$ regarded as a kernel for an integral operator, creates an extremely difficult integral equation which is again of a mixed Volterra-Fredholm type (see Appendix C for details.)

We therefore work with the Paley-Wiener representation (3.1.6.) for a Brownian motion in the same vein as before.

$$\Gamma(t) = 2t(1-t)Y_0 + \sqrt{2}(1-2t) \sum_{k=1}^{\infty} Y_k \frac{\sin k\pi t}{k\pi} \quad \text{uniformly in } t, t \in [0, 1];$$

$$\Gamma^2(t) = 4t^2(1-t)^2 Y_0^2 + 2^{5/2} t(1-t)(1-2t) Y_0 \sum_{k=1}^{\infty} Y_k \frac{\sin k\pi t}{k\pi}$$

$$+ 2(1-2t)^2 \left(\sum_{k=1}^{\infty} Y_k \frac{\sin k\pi t}{k\pi} \right)^2 \quad \text{and hence}$$

$$\begin{aligned}
\int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt &= 4Y_0^2 \int_0^1 t(1-t) dt + 2^{5/2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \int_0^1 (1-2t) \sin k\pi t dt \\
&+ 2 \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \int_0^1 \frac{(1-2t)^2}{t(1-t)} \sin k\pi t \sin m\pi t dt \\
&= \frac{2}{3} Y_0^2 + 2^{5/2} Y_0 \sum_{k=1}^{\infty} \frac{Y_k}{k\pi} \left\{ \frac{(-1)^{k-1} + 1}{k\pi} - 2 \frac{(-1)^{k-1}}{k\pi} \right\} \\
&+ 2 \sum_{k=1}^{\infty} \sum_{m=1}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \left\{ \int_0^1 \frac{\sin k\pi t \sin m\pi t}{t} dt + \int_0^1 \frac{\sin k\pi t \sin m\pi t}{1-t} dt \right\} \\
&+ 2 \sum_{k=1}^{\infty} \frac{1}{(k\pi)^2} \left\{ \int_0^1 \frac{\sin^2 k\pi t}{t} dt + \int_0^1 \frac{\sin^2 k\pi t}{1-t} dt - 4 \int_0^1 \sin^2 k\pi t dt \right\};
\end{aligned}$$

comparing $\sin k\pi t$ and $\sin k\pi(1-t)$ we see that

$$\sin k\pi(1-t) \sin m\pi(1-t) = \begin{cases} \sin k\pi t \sin m\pi t & \text{if } k, m \text{ even or } k, m \text{ odd} \\ -\sin k\pi t \sin m\pi t & \text{if } k \text{ even, } m \text{ odd or } k \text{ odd, } m \text{ even} \end{cases}$$

as illustrated below.

Figure 2.

Thus we obtain

$$\int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt = \frac{2}{3} Y_0^2 + 2^{7/2} Y_0 \sum_{k=1}^{\infty} \frac{Y_{2k}}{(2k\pi)^2} + 4 \sum_{k=1}^{\infty} \sum_{\substack{m=1 \\ k \neq m; k, m \text{ even} \\ \text{or } k, m \text{ odd}}}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \int_0^1 \frac{\sin k\pi t \sin m\pi t}{t} dt$$

$$+ 4 \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} \left\{ \int_0^{k\pi} \frac{\sin^2 y}{y} dy - 1 \right\}.$$

The last two integrals are evaluated separately in Appendix C:

$$\int_0^1 \frac{\sin k\pi t \sin m\pi t}{t} dt = \frac{1}{4} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \frac{(k-m)^{2n} - (k+m)^{2n}}{(2n)!} \pi^{2n}$$

and

$$\int_0^{k\pi} \frac{\sin^2 y}{y} dy = \frac{1}{4} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \frac{(2k\pi)^{2n}}{(2n)!}.$$

Therefore, we finally have

$$(3.1.22.) \quad \int_0^1 \frac{\Gamma^2(t)}{t(1-t)} dt = \frac{2}{3} Y_0^2 - 4 \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2} + 2^{7/2} \sum_{k=1}^{\infty} \frac{Y_0 Y_{2k}}{(2k\pi)^2}$$

$$+ \sum_{k=1}^{\infty} \sum_{\substack{m=1 \\ k, m \text{ even or} \\ k, m \text{ odd}}}^{\infty} \frac{Y_k}{k\pi} \frac{Y_m}{m\pi} \left(\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \frac{(k-m)^{2n} - (k+m)^{2n}}{(2n)!} \pi^{2n} \right)$$

Section 3.2. Test Statistics With Degenerate Kernels. An Example

In this section we will examine the previously introduced Example 2.1.17. more closely and we will prove a weak convergence result for $U_k = \sum_{i=1}^k \sum_{j=k+1}^n h(X_i, X_j) = \sum_{i=1}^k \sum_{j=k+1}^n X_i X_j$. We note that Theorem 3.2.0. cannot, at least in no obvious way, be generalized to hold for other possible degenerate symmetric kernels since the main idea of the proof is to express $U_n^{(3)}$, $U_k^{(1)}$, and $U_k^{(2)}$ as squares of appropriate partial sums of the X_i 's plus a negligible remainder term. Consider the decomposition of U_k into three U-statistics

$$\begin{aligned}
 U_k &= U_n^{(3)} - U_k^{(1)} - U_k^{(2)}, & \text{where} \\
 U_n^{(3)} &= \sum_{1 \leq i < j \leq n} X_i X_j, \\
 U_k^{(1)} &= \sum_{1 \leq i < j \leq k} X_i X_j & \text{for } 2 \leq k \leq n, \quad \text{and} \\
 U_k^{(2)} &= \sum_{k+1 \leq i < j \leq n} X_i X_j & \text{for } 1 \leq k \leq n-1.
 \end{aligned}$$

This decomposition agrees formally with the previously used representation (2.1.3.) except for the fact that the U-statistics here are degenerate. Let $X_1, X_2, \dots, X_n$ be as in Example 2.1.17. and require additionally that either $\text{Var } X_1 = \sigma_1^2 > 0$ is known a priori or that there is a consistent estimator $\hat{\sigma}_1^2$ for $\sigma_1^2 > 0$. The given random sample $X_1, \dots, X_n$ may be used to construct $\hat{\sigma}_1^2$. We may assume additionally without loss of generality by Slutsky's Theorem that X_1 is such that $\text{Var } X_1 = \frac{1}{3}$.

Define a bounded linear operator H via

$$H : L^2[0,1] \rightarrow L^2[0,1]$$

$$(Hf)(x) = E(h(x, X_1)f(X_1)) = xE(X_1 f(X_1)) \quad \text{for } x \in [0,1].$$

We wish to find the at most countably many eigenvalues λ_i and the corresponding orthonormal eigenfunctions f_i of H such that

$$(3.2.0.) \quad E[X_1 X_2 - \sum_{i=1}^n \lambda_i f_i(X_1) f_i(X_2)]^2 \rightarrow 0 \text{ as } n \rightarrow \infty.$$

That is, we must solve the integral equation

$$xE(X_1 f(X_1)) = \lambda f(x) \text{ for each } x \in [0,1].$$

As previously in Section 3.1., we will solve this integral equation by studying the associated differential equation assuming that $\lambda > 0$.

Differentiation with respect to x gives

$$\frac{d}{dx} [xE(X_1 f(X_1))] = E(X_1 f(X_1)) = \lambda f'(x).$$

Differentiating again, we obtain

$$\frac{d^2}{dx^2} [xE(X_1 f(X_1))] = 0 = \lambda f''(x).$$

Hence we conclude that $f(x) = ax+b$ for suitable constants $a, b \in \mathbb{R}$.

Thus

$$\begin{aligned} xE(X_1 f(X_1)) &= xE(X_1 (aX_1 + b)) \\ &= axEX_1^2 + bxEX_1 \\ &= ax \text{ Var } X_1 \quad \text{since } EX_1 = 0 \text{ by assumption.} \end{aligned}$$

Therefore $ax \text{ Var } X_1 = \lambda(ax+b)$ which implies that $b=0$. Consequently, we found the eigenfunctions f of H to be of the form $f(x) = ax$ for $a \geq 0$ with the constraint that $\text{Var } X_1 = \lambda$. Furthermore, any eigenfunction of H is desired to be normed to one. We compute the norm of $f(x) = ax$ as

$$\|f(x)\| = \|ax\| = |a| \cdot \|x\| = |a| \cdot \sqrt{\langle x, x \rangle} = |a| \cdot \left(\int_0^1 x^2 dx \right)^{1/2} = \frac{|a|}{\sqrt{3}}.$$

implying that $f(x) = \sqrt{3} x$ for each $x \in [0,1]$. In addition, (3.2.0.) forces us to assume that $\lambda = \text{Var } X_1 = \frac{1}{3}$ which now explains the seemingly strange assumption $\text{Var } X_1 = \frac{1}{3}$ in the beginning of this subsection. We summarize these results by stating that $\lambda = \frac{1}{3}$ is the only eigenvalue of H and its associated eigenfunction is $f(x) = \sqrt{3} x$.

Our next goal is to find a sequence of nondegenerate processes, say $\{G_n(t): 0 \leq t \leq 1\}$, $n \geq 1$, such that $U_{[nt]} - G_n(t)$, properly normalized, converges in sup-norm to zero (in probability). Towards this end we define

$$G(t) = W^2(1) - W^2(t) - W^2(1-t) \quad \text{for } 0 \leq t \leq 1,$$

where $W = \{W(t): 0 \leq t < \infty\}$ denotes a standard Wiener process. Then we have the following theorem.

Theorem 3.2.0.

Let $X_1, X_2, \dots, X_n$ be independent identically distributed random variables with $EX_1 = 0$ and $\text{Var } X_1 = \frac{1}{3}$. Put

$$h(x,y) = xy \quad \text{for } x,y \in \mathbb{R} \text{ and}$$

$$U_k = \sum_{i=1}^k \sum_{j=k+1}^n h(X_i, X_j) = \sum_{i=1}^k \sum_{j=k+1}^n X_i X_j \quad \text{for } 1 \leq k \leq n-1.$$

Then there exists a sequence of processes $\{G_n(t): 0 \leq t \leq 1\}$, $n \geq 1$, such that

$$\{G_n(t): 0 \leq t \leq 1\} \stackrel{D}{=} \{G(t): 0 \leq t \leq 1\} \quad \text{for each } n \in \mathbb{N} \text{ and}$$

$$\sup_{0 \leq t \leq 1} \left| \frac{1}{n} U_{[nt]} - G_n(t) \right| = o_p(1) \quad \text{as } n \rightarrow \infty.$$

Note that Theorem 3.2.0. is a formal analogue of Theorem 2.1.14. for the degenerate case. Concerning the variance assumption recall the remarks made at the beginning of this subsection. Furthermore, if the variance of X_1 is unknown a priori, but assumed to be finite, and if $(\hat{\sigma}_n^2)_{n \geq 2}$ is a sequence of consistent estimators for $\sigma_1^2 = \text{Var } X_1 > 0$, then by Slutsky's Theorem (confer e.g., R.H. Randles and D.A. Wolfe (1979), Theorem 3.2.8.) we have the following corollary.

Corollary 3.2.1.

$$\sup_{0 \leq t \leq 1} \left| \frac{2}{\hat{\sigma}_n^2 \cdot n} U_{[nt]} - G_n(t) \right| = o_p(1) \text{ as } n \rightarrow \infty.$$

The next lemma is needed for the proof of Theorem 3.2.0.

Lemma 3.2.2.

Given the assumptions of Theorem 3.2.0. we have

$$(3.2.1.) \quad \left| \frac{6}{n} U_n^{(3)} - (W_n^2(1) - 1) \right| = o_p(1) \text{ as } n \rightarrow \infty,$$

$$(3.2.2.) \quad \sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(1)} - (W_n^2(t) - t) \right| = o_p(1) \text{ as } n \rightarrow \infty, \text{ and}$$

$$(3.2.3.) \quad \sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(2)} - (W_n^2(1-t) - (1-t)) \right| = o_p(1) \text{ as } n \rightarrow \infty,$$

where $\{W_n\}_{n \geq 1}$ is a sequence of standard Wiener processes.

Proof of Lemma 3.2.2.

Put $S_0 = 0$ and $S_n = X_1 + X_2 + \dots + X_n$ for $n \geq 1$. Define $S_n(t)$ as $S_n(t) = \frac{\sqrt{3}}{\sqrt{n}} S_{[nt]}$, $0 \leq t \leq 1$. Then

$$S_n^2 = \left(\sum_{i=1}^n X_i \right) \left(\sum_{j=1}^n X_j \right) = \sum_{i=1}^n X_i^2 + 2 \sum_{1 \leq i < j \leq n} X_i X_j$$

and therefore

$$S_n^2(1) = \left(\frac{\sqrt{3}}{\sqrt{n}} S_n \right)^2 = \frac{3}{n} \sum_{i=1}^n X_i^2 + \frac{6}{n} U_n^{(3)}, \text{ whence}$$

$$\begin{aligned} & \left| \frac{6}{n} U_n^{(3)} - (W_n^2(1) - 1) \right| \\ &= \left| S_n^2(1) - \frac{3}{n} \sum_{i=1}^n X_i^2 - (W_n^2(1) - 1) \right| \\ &\leq \left| S_n^2(1) - W_n^2(1) \right| + \left| \frac{3}{n} \sum_{i=1}^n X_i^2 - 1 \right| \\ &= \left| S_n^2(1) - W_n^2(1) \right| + o(1) \text{ a.s.} \end{aligned}$$

by the strong law of large numbers for $Y_i = X_i^2$ since $EY_i = EX_i^2 = \text{Var } X_1 = \frac{1}{3}$. Thus (3.2.1.) will be proven if we can show that

$$\left| S_n^2(1) - W_n^2(1) \right| = o_p(1) \text{ as } n \rightarrow \infty.$$

$$\begin{aligned} & \left| S_n^2(1) - W_n^2(1) \right| \\ &= \left| \frac{3}{n} S_n^2 - W_n^2(1) \right| \\ &\stackrel{D}{=} \left| \frac{3}{n} S_n^2 - \frac{1}{n} W^2(n) \right| \\ &\leq \frac{1}{n} (\sqrt{3} S_n - W(n))^2 + 2 \frac{|W(n)|}{\sqrt{n}} \cdot \frac{|\sqrt{3} S_n - W(n)|}{\sqrt{n}} \\ &= (o_p(1))^2 + 2 \frac{|W(n)|}{\sqrt{n}} \cdot o_p(1) \\ &\stackrel{D}{=} o_p(1) + |W(1)| \cdot o_p(1) \\ &= o_p(1) \end{aligned}$$

as $n \rightarrow \infty$.

Here we used the supplementary remarks to Section 2.2. in M. Csörgő and P. Révész (1981) in order to obtain

$$|\sqrt{3} S_n - W(n)| = o_p(\sqrt{n}) \text{ as } n \rightarrow \infty.$$

This completes the proof of (3.2.1.).

Next we show (3.2.2.). Note that $2U_{[nt]}^{(1)} = 2 \sum_{1 \leq i < j \leq [nt]} x_i x_j =$
 $\left[\sum_{1 \leq i \neq j \leq [nt]} x_i x_j + \sum_{i=1}^{[nt]} x_i^2 \right] - \sum_{i=1}^{[nt]} x_i^2$, whence
 $2U_{[nt]}^{(1)} = S_{[nt]}^2 - \sum_{i=1}^{[nt]} x_i^2$, and

$$\frac{6}{n} U_{[nt]}^{(1)} = \frac{3}{n} S_{[nt]}^2 - \frac{3}{n} \sum_{i=1}^{[nt]} x_i^2 = S_n^2(t) - \frac{3}{n} \sum_{i=1}^{[nt]} x_i^2.$$

Thus

$$\sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(1)} - (W_n^2(t) - t) \right| \leq \sup_{0 \leq t \leq 1} |S_n^2(t) - W_n^2(t)| + \sup_{0 \leq t \leq 1} \left| \frac{3}{n} \sum_{i=1}^{[nt]} x_i^2 - t \right|.$$

At this stage we again use the strong law of large numbers to obtain

$$\frac{3}{n} \sum_{i=1}^{[nt]} x_i^2 = \left(\frac{3}{[nt]} \sum_{i=1}^{[nt]} x_i^2 \right) \frac{[nt]}{n} \xrightarrow{\text{a.s.}} (3 \text{ Var } X_1) t = t$$

uniformly in $t \in [0, 1]$ as $n \rightarrow \infty$.

Consequently, $\sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(1)} - (W_n^2(t) - t) \right| \leq \sup_{0 \leq t \leq 1} |S_n^2(t) - W_n^2(t)| + o(1)$. Next, consider $\sup_{0 \leq t \leq 1} |S_n^2(t) - W_n^2(t)|$.

$$\begin{aligned} \sup_{0 \leq t \leq 1} |S_n^2(t) - W_n^2(t)| &\leq \sup_{0 \leq t \leq 1} (S_n(t) - W_n(t))^2 + 2 \sup_{0 \leq t \leq 1} |S_n(t) - W_n(t)| |W_n(t)| \\ &= o_p(1) + o_p(1) \cdot \sup_{0 \leq t \leq 1} |W_n(t)| \quad (*) \\ &\stackrel{D}{=} o_p(1) + o_p(1) \cdot \sup_{0 \leq t \leq 1} |W(t)| \end{aligned}$$

Since $\sup_{0 \leq t \leq 1} |W(t)|$ is bounded in probability it follows that

$$\sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(1)} - (W_n^2(t) - t) \right| = \sup_{0 \leq t \leq 1} |S_n^2(t) - W_n^2(t)| + o_p(1) = o_p(1) \text{ as } n \rightarrow \infty.$$

Where indicated by (*) we again made use of the results on pp.112-113

in M. Csörgő and P. Révész (1981). This completes the proof of (3.2.2.).

Finally, we prove (3.2.3.). We notice that

$$\begin{aligned} 2U_{[nt]}^{(2)} &= \left(\sum_{[nt]+1 \leq i \neq j \leq n} x_i x_j + \sum_{i=[nt]+1}^n x_i^2 \right) - \sum_{i=[nt]+1}^n x_i^2 \\ &= (S_n - S_{[nt]})^2 - \sum_{i=[nt]+1}^n x_i^2, \quad \text{whence} \end{aligned}$$

$$\begin{aligned} \sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(2)} - (W_n^2(1-t) - (1-t)) \right| \\ = \sup_{0 \leq t \leq 1} \left| \frac{3}{n} (S_n - S_{[nt]})^2 - W_n^2(1-t) \right| + \sup_{0 \leq t \leq 1} \left| \frac{3}{n} \sum_{i=[nt]+1}^n x_i^2 - (1-t) \right| \end{aligned}$$

and with the same arguments as used before, $\frac{3}{n} \sum_{i=[nt]+1}^n x_i^2 \xrightarrow{\text{a.s.}} 1-t$

uniformly in $t \in [0,1]$ as $n \rightarrow \infty$. Therefore,

$$\sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(2)} - (W_n^2(1-t) - (1-t)) \right| \stackrel{\text{a.s.}}{\leq} \sup_{0 \leq t \leq 1} \left| \frac{3}{n} (S_n - S_{[nt]})^2 - W_n^2(1-t) \right| + o(1).$$

Next we show that

$$\sup_{0 \leq t \leq 1} \left| \frac{3}{n} S_{[n(1-t)]}^2 - W_n^2(1-t) \right| = o_p(1) \text{ as } n \rightarrow \infty.$$

Let now $y = 1-t$ for $0 \leq t \leq 1$. Then

$$\sup_{0 \leq t \leq 1} \left| \frac{3}{n} S_{[n(1-t)]}^2 - W_n^2(1-t) \right| = \sup_{0 \leq y \leq 1} \left| \frac{3}{n} S_{[ny]}^2 - W_n^2(y) \right|$$

and thus the methods of the proof of (3.2.2.) can be employed to obtain

(3.2.3.).

Proof of Theorem 3.2.0.

Let again $G_n(t) = W_n^2(1) - W_n^2(t) - W_n^2(1-t)$. Then $\{G_n(t): 0 \leq t \leq 1\}$
 $\stackrel{D^*}{=} \{G(t): 0 \leq t \leq 1\}$ for each $n \in \mathbb{N}$ and

$$\begin{aligned} & \sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]} - G_n(t) \right| \\ &= \sup_{0 \leq t \leq 1} \left| \left(\frac{6}{n} U_n^{(3)} - \frac{6}{n} U_{[nt]}^{(1)} - \frac{6}{n} U_{[nt]}^{(2)} \right) - \{ (W_n^2(1) - 1) - (W_n^2(t) - t) - (W_n^2(1-t) - (1-t)) \} \right| \\ &\leq \sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_n^{(3)} - (W_n^2(1) - 1) \right| + \sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(1)} - (W_n^2(t) - t) \right| \\ &\quad + \sup_{0 \leq t \leq 1} \left| \frac{6}{n} U_{[nt]}^{(2)} - (W_n^2(1-t) - (1-t)) \right| \\ &= o_p(1) \text{ by Lemma 3.2.2.} \end{aligned}$$

Next we compute the expected value, variance and covariance structure of the limiting process $G = \{G(t): 0 \leq t \leq 1\}$.

$$EG(t) = EW^2(1) - EW^2(t) - EW^2(1-t) = \text{Var } W(1) - \text{Var } W(t) - \text{Var } W(1-t) = 0$$

for $t \in [0, 1]$.

Now we compute the covariance function of G . Let $s \leq t$, then

$$\begin{aligned} \text{Cov}(G(s), G(t)) &= EG(s)G(t) \\ &= E(W^2(1) - W^2(s) - W^2(1-s))(W^2(1) - W^2(t) - W^2(1-t)) \\ &= EW^4(1) - EW^2(1)W^2(t) - EW^2(1)W^2(1-t) - EW^2(1)W^2(s) + EW^2(s)W^2(t) \\ &\quad + EW^2(s)W^2(1-t) - EW^2(1)W^2(1-s) + EW^2(1-s)W^2(t) + EW^2(1-s)W^2(1-t) \\ &= 3 - (t+2t^2) - (1-t+2(1-t)^2) - (s+2s^2) + (st+2s^2) \\ &\quad + (s(1-t)+2s^2) - (1-s+2(1-s)^2) + (t(1-s)+2t^2) \\ &\quad + ((1-s)(1-t)+2(1-t)^2) \end{aligned}$$

$$= 4s \quad \text{provided that } s \leq 1-t.$$

Similarly, if $s \leq t$ and $s \geq 1-t$, then we have

$$\begin{aligned} \text{Cov}(G(s), G(t)) &= 3-(t+2t^2)-(1-t+2(1-t)^2)-(s+2s^2)+(st+2s^2) \\ &\quad +(s(1-t)+2(1-t)^2)-(1-s+2(1-s)^2)+(t(1-s)+2(1-s)^2) \\ &\quad +((1-s)(1-t)+2(1-t)^2) \\ &= 4(1-t). \end{aligned}$$

Therefore, $\text{Cov}(G(s), G(t)) = 4\{s \wedge (1-t)\}$ for $0 \leq s \leq t \leq 1$, which can also be expressed as

$$\text{Cov}(G(s), G(t)) = \begin{cases} 4(s \wedge t) & \text{if } 0 \leq t \leq \frac{1}{2} \\ 4(s \wedge (1-t)) & \text{if } \frac{1}{2} \leq t \leq 1 \end{cases} \quad \text{for } 0 \leq s \leq t \leq 1.$$

I.e. G behaves like four times a regular standard Wiener process between 0 and $1/2$, and like this with a reversed time scale between $1/2$ and 1. We also have $\text{Var } G(t) = 4\{t \wedge (1-t)\}$ for $t \in [0, 1]$.

We close this section with a corollary in which we collect miscellaneous results concerning the processes appearing in Theorem 3.2.0.

Corollary 3.2.3.

Under the assumptions of Theorem 3.2.0. the following statements hold.

$$(a) \quad \sup_{0 \leq t \leq 1} \left| \frac{U_{[nt]}}{n} \right| = O_p(1) \quad \text{as } n \rightarrow \infty.$$

$$(b) \quad \sup_{0 \leq t \leq 1} |G(t)| < \infty \quad \text{a.s.}$$

$$(c) \frac{6}{n} \sup_{0 \leq t \leq 1} |U_{[nt]}| \xrightarrow{D} \sup_{0 \leq t \leq 1} |G(t)| \text{ as } n \rightarrow \infty.$$

$$(d) \frac{6}{n} \int_0^1 U_{[nt]} dt \xrightarrow{D} \left\{ \lim_{K \rightarrow \infty} \sum_{i,j=0}^K a_{ij} Y_i Y_j \right\} \text{ as } n \rightarrow \infty \text{ for suitable}$$

$a_{ij} \in \mathbb{R}$, where $(Y_i)_{i \geq 1}$ is a sequence of independent $N(0,1)$ -random variables.

Proof of Corollary 3.2.3.

Proof of (a).

(a) is implied by (b) and Theorem 3.2.0.

Proof of (b).

Part (b) is obvious from the definition of G and the corresponding properties of W .

Proof of (c).

(c) follows immediately from Theorem 3.2.0. and part (b).

Proof of (d).

Theorem 3.2.0. implies $\frac{6}{n} \int_0^1 U_{[nt]} dt = \int_0^1 \left(\frac{6}{n} U_{[nt]} - G_n(t) \right) dt + \int_0^1 G_n(t) dt$
 $\xrightarrow{D} \int_0^1 G(t) dt$. In addition to this, we have: $\int_0^1 G(t) dt =$

$W^2(1) - 2 \int_0^1 W^2(t) dt$, whence (3.1.12.) implies

$$\int_0^1 G(t) dt = \frac{1}{3} Y_0^2 + 2^{5/2} \sum_{k=1}^{\infty} \frac{(-1)^k}{(k\pi)^2} (Y_0 Y_k) - \sum_{k=1}^{\infty} \frac{Y_k^2}{(k\pi)^2}$$

which can again be expressed as a limit of quadratic forms of independent $N(0,1)$ -random variables:

$$\int_0^1 G(t) dt = \lim_{K \rightarrow \infty} \sum_{i,j=0}^K a_{ij} Y_i Y_j, \text{ where}$$

$$a_{ij} = \begin{cases} 1/3 & \text{if } i = j = 0 \\ -\frac{1}{(i\pi)^2} & \text{if } i = j \geq 1 \\ 0 & \text{if } i \neq j \geq 1 \\ 2^{3/2} \frac{(-1)^i}{(i\pi)^2} & \text{if } i \geq 1, j = 0 \\ 2^{3/2} \frac{(-1)^j}{(j\pi)^2} & \text{if } j \geq 1, i = 0. \end{cases}$$

We note that any test statistic based on

$$U_k = \sum_{i=1}^k \sum_{j=k+1}^n h(X_i, X_j) = \sum_{i=1}^k \sum_{j=k+1}^n X_i X_j$$

will be distribution free. This is not true for every symmetric kernel h in the degenerate case as the next example shows.

Example 3.2.4. (Confer also Example A, p.194, in R.J. Serfling (1980).)

Consider the kernel $h(x_1, x_2) = \frac{1}{2}(x_1 - x_2)^2$ of Example 2.1.1.(b).

In this case, the parameter θ is the theoretical (population) variance of the random variable X_1 , i.e.

$$\theta = \sigma_1^2 = \text{Var } X_1 > 0.$$

If we assume that

$$\mu_1 = EX_1 = 0 \quad \text{and} \quad \mu_4 = EX_1^4 = \sigma_1^4 < \infty,$$

then h is a degenerate kernel since

$$\begin{aligned} \text{Var } \tilde{h}(X_1) &= E(\tilde{h}(X_1) - \sigma_1^2)^2 \\ &= E(E_{X_2} h(X_1, X_2) - \sigma_1^2)^2 \end{aligned}$$

$$= E\left[\frac{1}{2}(X_1^2 + EX_2^2) - \sigma_1^2\right]^2$$

$$= \frac{1}{4}(\mu_4 - \sigma_1^4)$$

Furthermore,

$$\text{Var } h(X_1, X_2) = E(h(X_1, X_2) - \sigma_1^2)^2$$

$$= \frac{1}{4}E(X_1^2 + X_2^2 - 2X_1X_2)^2 - \sigma_1^2 E(X_1^2 + X_2^2 - 2X_1X_2) + \sigma_1^4$$

$$= \frac{1}{4}[E(X_1^2 + X_2^2)^2 + 4EX_1^2X_2^2] - \sigma_1^4$$

$$= 2\sigma_1^4 - \sigma_1^4 = \sigma_1^4 > 0.$$

Hence we may apply Theorem (without number, p. 194, R.J. Serfling (1980))

with $m = 2$, $\zeta_0 = \zeta_1 = 0$, $\zeta_2 = \sigma_1^4 > 0$ in order to obtain the asymptotic distribution of $U_n^{(3)}$. It is a consequence of the aforementioned theorem that

$$\frac{1}{n} \sum_{1 \leq i < j \leq n} (X_i - X_j)^2 - n\sigma_1^2 \xrightarrow{D} Y \text{ as } n \rightarrow \infty, \text{ where}$$

Y is a random variable of the form $Y = \sum_{k=1}^{\infty} \lambda_k (Y_k^2 - 1)$ (with the $\sum$ possibly being finite). Here the random variables Y_k are independent identically distributed Gaussian variates. The λ_k 's are the (not necessarily distinct) eigenvalues of the Hilbert space operator A defined via $Ag(x) = \int_{\mathbb{R}} (h(x, y) - \sigma_1^2) g(y) dF(y)$, where F denotes the cumulative distribution function of X_1 . Through solving this integral equation we shall now see that the eigenvalues λ_k strongly depend on the given distribution F of X_1 .

$$\begin{aligned}
 \lambda g(x) &= \int_{\mathbb{R}} \left[\frac{1}{2}(x-y)^2 - \sigma_1^2 \right] g(y) dF(y) \\
 &= \left(\frac{1}{2} \int_{\mathbb{R}} g(y) dF(y) \right) x^2 + \left(- \int_{\mathbb{R}} y g(y) dF(y) \right) x \\
 &\quad + \left(\frac{1}{2} \int_{\mathbb{R}} y^2 g(y) dF(y) - \sigma_1^2 \int_{\mathbb{R}} g(y) dF(y) \right) x^0,
 \end{aligned}$$

whence g must be a polynomial of degree at most two.

Put $g(x) = ax^2 + bx + c$ for some $a, b, c \in \mathbb{R}$, then comparison of corresponding coefficients of the different powers of x leads to the following system of equations.

$$2\lambda a = \int_{\mathbb{R}} g(y) dF(y)$$

$$-\lambda b = \int_{\mathbb{R}} y g(y) dF(y)$$

$$\lambda c = \frac{1}{2} \int_{\mathbb{R}} y^2 g(y) dF(y) - \sigma_1^2 \int_{\mathbb{R}} g(y) dF(y),$$

which clearly shows the dependency between λ and the distribution F of X_1 . R.J. Serfling (1980), Example A, pp. 194-195 solves the above system of equations for the case of a binomial distribution with parameters $n = 1$ and $p = 0.5$.

Chapter 4. Kendall-Kendall Pontograms and Some Generalizations to Renewal Theory

Section 4.0. Introduction

4.0.0. The Kendall-Kendall t - and T -Pontograms

In the present part of the introduction we give a short discussion of the mathematics contained in the paper "Alignments in two-dimensional random sets of points" by D.G. and W.S. Kendall (1980). This is done in order to demonstrate the practical statistical problem from which Theorem 4.1.3. originally arose. Before we can describe this statistical problem we need the following

Definition 4.0.0.

Working in two-dimensional Euclidean space, i.e. in the plane $\mathbb{R}^2$, we call any set of 3 points in $\mathbb{R}^2$ a triad of points. Furthermore, we say a given triad of points constitutes an ε -blunt triangle, ($0 < \varepsilon < 10$ fixed), if its largest angle is greater than or equal to $(180-\varepsilon)^\circ$. The real number $\varepsilon \in (0,10)$ is called the tolerance.

Globally speaking, the data-analytic problem which D.G. and W.S. Kendall (1980) desire to solve, is the problem of testing whether an empirical set of n points ($n \in \mathbb{N}$ fixed) could be considered to contain "too many" such ε -blunt triangles in a situation where the generating mechanism of the n points is unknown and where all that can be said about the tolerance ε is that it must be allowed to take values in some given interval $[T_0, T_1]$ with $0 < T_0 < T_1 < 10$. For the solution

of this problem two plots, called t - and τ -pontograms, are introduced. Regarding the region $[T_0, T_1]$ in which the tolerance ε is allowed to vary, it is worthwhile noting that T_0 corresponds to δ/n in Theorem 4.1.3. and that it is immediate from the proof of this theorem that δ cannot be chosen to be zero. Taking $T_0 = 0$ of course has to be interpreted in the context of the detection of alignments of infinite precision. D.G. and W.S. Kendall (1980) state that in practice one cannot specify the value of ε but since they were not in a position to prove their results with T_0 going to zero, they reach a "general agreement" on the choice of "sensible" lower and upper bounds for ε , T_0 and T_1 , respectively. It is here that one reason for importance of Theorem 4.1.3. is anchored because this theorem prepares the ground for the consideration of pontograms over regions whose lower bounds converge to zero. In the discussion (confer J.R. Statist. Soc. A 143, 109-140) following the presentation of S. Broadbent's (1980) paper on "Simulating the Ley Hunter" D.G. Kendall describes his new testing device, the pontogram: "We started with Broadbent's uniform-rectangular model, and extended that to one involving i.i.d. sampling from an arbitrary compact convex set. For this we found the expected number $N(\varepsilon)$ of ε -collinearities in a sample of given size n , and showed how one could compute the variance of $N(\varepsilon)$ so as to permit a significance test. ... We then turned to a model based on i.i.d. sampling from an arbitrary distribution, ... We overcame the difficulty that the tolerance-angle ε , far from being given, is actually an unknown nuisance-parameter by devising a test using a construction that we call the Pontogram. ... Asymptotically the pontogram curve is a realization

of an Ornstein-Uhlenbeck process, "... ."

Carefully restated, the hypotheses in D.G. and W.S. Kendall (1980) are:

H_0 : the given data set of n points does not contain planned alignments constructed with a tolerance of $\epsilon \in [T_0, T_1]$

versus

H_A : there exist planned alignments constructed with a tolerance $\epsilon \in [T_0, T_1]$.

Here "planned alignments constructed with a tolerance ϵ " are meant to be ϵ -blunt triangles (compare also Figure 3. below), while perfect alignments ($\epsilon = 0$) are understood to be triads of points which are

ϵ -Blunt Triangles

Figure 3.

collinear. If we include the case of perfect alignments (i.e. the case of collapsed or degenerate triangles), then there are $\binom{n}{3}$ triangles with vertices in a given data set containing n points. Corresponding to these $\binom{n}{3}$ triangles we have $\binom{n}{3}$ maximal angles θ and the same number of defects, where the defect T is defined as $T = 180^\circ - \theta$. Now we model the defect T as a random variable and consider its order statistics. Since the binomial coefficient $\binom{n}{3}$ is large as compared to the number of points n , the above order statistics can be thought of as describing the events of a point process on the positive half line which might be approximated by a non-uniform Poisson process N , i.e. by a Poisson process whose intensity parameter $\mu > 0$ depends on time, $\mu = \mu(x)$, $x > 0$. Since for testing the hypothesis H_0 against H_A the information arising from triangles whose defect T falls outside the $(0^\circ, 10^\circ)$ -region is irrelevant, we may now ignore everything that happens outside this interval and scale the whole problem so that the $0 < x < 1$ section of the aforementioned Poisson process N with intensity parameter $\mu = \mu(x) > 0$ represents all relevant data points. In terms of the stochastic process N the hypotheses can now be reformulated as

H_0^* : $N = \{N(x): 0 < x < 1\}$ possesses a constant but unknown intensity parameter $\mu > 0$

H_A^* : $N = \{N(x): 0 < x < 1\}$ has a constant parameter $\mu = \mu_0$ over $(\varepsilon, 1)$ and over $(0, \varepsilon)$ a parameter $\mu = \mu_1 > \mu_0$.

The fact that we can specify μ_1 to be larger than μ_0 (as opposed to simply saying $\mu_1 \neq \mu_0$) is immediate from the original alternative

hypothesis H_A , since H_A assumes the influence of planning (resulting in a large number of ϵ -blunt triangles) to be present and this is nothing but the existence of a large intensity parameter of N in $(0, \epsilon)$. Initially, D.G. and W.S. Kendall (1980) test H_0^* against H_A^* with fixed value of ϵ and then subsequently with ϵ free to vary inbetween T_0 and T_1 .

Turning now to the important example of the Land's end data, D.G. and W.S. Kendall (1980) observe that the defect T appears to have an approximately constant density over the extreme left-hand interval which in degrees is $(0^\circ, 10^\circ)$. (Compare also Figures 4. and 5. below.)

Figure 4

The histogram for the maximum angles of the 22100 Land's End triads (top of each column shown by *). The curve shows the theoretical density for the maximum angle of a triangle with independent skew-gaussian vertices ($s = 1.661$)

(Taken from D.G. and W.S. Kendall (1980).)

Figure 5
The theoretical density for the maximum angle on the gaussian model for $s = 1.0(0.5)5.0$. Note that the ordinates on the extreme right increase as s increases

(Taken from D.G. and W.S. Kendall (1980).)

In terms of the above introduced Poisson process N this means that the intensity parameter $\mu = \mu(x)$ seems to be independent of x in the corresponding interval. To test the hypothesis H_0^* versus H_A^* they chose $T_0 = 0.01$ and $T_1 = 0.1$.

The 'old stones of Land's end', or as it often is referred to, the 'Land's end data problem', has been examined by various, predominantly British, authors (confer in particular S. Broadbent (1980) and further references there). The data material consists of 52 sites exhibiting old stone configurations in Cornwall, U.K. These sites as well as other remains of neolithic cultures have been studied by archaeologists who were interested in finding out whether or not the objects were deliberately put on straight lines. Although the example of the Land's end data, is partly responsible for the creation of the

pontogram, it is by no means the only example for an application of the new pontogram technique as D.G. Kendall explained in the aforementioned discussion of S. Broadbent's paper: "It should be stressed that the problem is important for several reasons. Quite apart from the application to archaeology which prompted the work, there are substantial applications to geomorphology, to oil technology, and to nuclear physics, ...".

Going back to the general problem where realistically ε has to be assumed to be an unknown quantity in $[T_0, T_1]$, D.G. and W.S. Kendall (1980) define their test statistic to be (denote from now on ε by t):

$$\sup_{T_0 \leq t \leq T_1} |Z(t)|, \text{ where } Z(t) = \left(\frac{N(t)}{t} - \frac{R-N(t)}{1-t} \right) \sqrt{\frac{t(1-t)}{R}}, \quad 0 < t < 1,$$

and $Z(0) = Z(1) = 0$.

Here $N(t)$ denotes the number of Poisson events (i.e. defects) in the standardized range $(0, t)$ and R stands for the number of Poisson events in the whole section $(0, 1)$.

We observe that

$$Z(t) = \frac{1}{\sqrt{R}} \cdot \frac{N(t) - tR}{\sqrt{t(1-t)}}, \quad 0 < t < 1.$$

It is this representation of Z that inspired the notion of the process $K_n = \{K_n(t) : 0 \leq t \leq 1\}$ in Theorem 4.1.3.

Next D.G. and W.S. Kendall (1980) introduce the t -pontogram as the plot of $Z(t)$ against t over the full range $(0, 1)$ and similarly the τ -pontogram as a blown up version of the t -pontogram, created by making the "time"-transformation $\tau = \log \frac{t}{1-t}$. Figure 8. shows the

4.0.1. Poisson and Generalized Renewal Processes

At this point we will undertake an excursion into renewal theory in order to establish the background material necessary for the proofs of Theorems 4.1.3. and 4.1.9., respectively. We agree on the following definitions, notations and abbreviations to be valid in all of Chapter 4.

Let $X_1, X_2, \dots$ denote independent identically distributed random variables obeying an exponential distribution law with parameter

$\mu > 0$, i.e.

$$(4.0.0.) \quad P(X_1 \leq x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\mu x} & x \geq 0 \end{cases}, \quad i = 1, 2, \dots$$

Put

$$(4.0.1.) \quad S_0 = 0 \text{ and } S_x = \sum_{i=1}^{[x]} X_i \text{ for } x \in (0, \infty)$$

Here $[x]$ again denotes the largest integer less than or equal to x .

Definition 4.0.1.

The stochastic process $N = \{N(x) : x \geq 0\}$, where

$$N(x) = \max \{n \in \mathbb{N}_0 : S_n \leq x\},$$

is said to be a Poisson process with (time-independent) intensity parameter $\mu > 0$. $S_1(\omega), S_2(\omega), \dots, \omega \in \Omega$ fixed, i.e. the points of discontinuity of $N(x, \omega)$ regarded as a function of x , are called the arrival times.

Note that using (4.0.1.), $N(x)$ may also be written as $N(x) = \sup \{y \geq 0 : S_y \leq x\}$. In addition, for each fixed x , we have the

relationships

$$\begin{aligned}
 N(x)+1 &= \max \{n \in \mathbb{N}_0 : S_n \leq x\} + 1 \\
 &= \min \{n \in \mathbb{N}_0 : S_n > x\} =: L(x) \quad \text{and} \\
 N(x)+1 &= \sup \{y \geq 0 : S_y \leq x\} + 1 \\
 &= \inf \{y \geq 0 : S_y > x\} =: L(x).
 \end{aligned}$$

For obvious reasons we will either say that $L(x)$ is the first passage time out of the interval $(-\infty, x]$ for the sequence $\{S_n\}_{n \geq 1}$ (confer C.C. Heyde (1967)) or that $L(x)$ is the first passage time from $[0, x]$ of the random walk $\{S_n\}_{n \geq 1}$ (confer J. Steinebach (1986)). These notions will be used in Theorem 4.1.3. whereas the following conventions are made with Theorem 4.1.9. in mind.

Let $\{Y_i\}_{i \geq 1}$ be a sequence of independent identically distributed random variables satisfying

$$(4.0.2.) \quad EY_1 = m > 0 \quad \text{and} \quad \text{Var } Y_1 = E(Y_1 - m)^2 = \sigma^2 > 0.$$

Set

$$(4.0.3.) \quad \begin{cases} T_0 = 0 \quad \text{and} \quad T_x = \sum_{i=1}^{[x]} Y_i & \text{for } 0 < x < \infty, \\ R_n = \max \{T_0, T_1, \dots, T_n\} & \text{for each } n \geq 0, \\ \tilde{N}(x) = \max \{n \in \mathbb{N}_0 : R_n \leq x\} & \text{for } x \in [0, \infty), \quad \text{and} \\ \tilde{L}(x) = \min \{n \in \mathbb{N}_0 : T_n > x\} & \text{for } x \in [0, \infty). \end{cases}$$

Definition 4.0.2.

The stochastic process $\tilde{N} = \{\tilde{N}(x) : x \geq 0\}$ is called generalized renewal process (confer C.C. Heyde (1967) for this nomenclature). The

random variable $\tilde{L}(x)$ is again said to be the first passage time.

Using $R_y := \max_{0 \leq x \leq y} T_x$, $y \geq 0$ where appropriate, the following facts are immediate from the above definitions:

$$(4.0.4.) \left\{ \begin{array}{ll} \text{(i)} & \tilde{N}(x) = \sup\{y \geq 0: R_y \leq x\}, \quad x \geq 0 \\ \text{(ii)} & \tilde{L}(x) = \inf\{y \geq 0: T_y > x\} = \tilde{N}(x) + 1, \quad x \geq 0, \\ \text{(iii)} & P\{\tilde{N}(x) \geq n\} = P\{R_n \leq x\} \quad \text{for each } x \in (0, \infty) \text{ and } n \geq 1, \\ \text{(iv)} & P\{N(x) \geq n\} = P\{S_n \leq x\} \quad \text{for each } x \in (0, \infty) \text{ and } n \geq 1, \\ \text{(v)} & P\{N(x) = n\} = P\{S_n \leq x, S_{n+1} > x\}, \quad 0 < x < \infty \text{ and } n \geq 1, \\ \text{(vi)} & P\{\tilde{N}(x) = n\} = P\{S_0, S_1, \dots, S_n \leq x, S_{n+1} > x\}, \quad x \in (0, \infty), n \geq 1, \\ \text{(vii)} & \tilde{N}(x) \text{ is monotonically nondecreasing as a function of } x. \end{array} \right.$$

We see that (iii) and (vi) will be of the same form as (iv) and (v), respectively, if the random variables $Y_1, 1 \geq 1$, are assumed to be almost surely nonnegative. In the latter case $\tilde{N}$ reduces to an ordinary renewal process. We recall that the prime example for the class of ordinary renewal processes is the Poisson process itself. Here an ordinary renewal process is meant to be a generalized renewal process (confer Definition 4.0.2. above) whose underlying sequence of random variables $(Y_n)_{n \geq 1}$ consists of nonnegative independent identically distributed random variables. (Confer also W. Feller (1966), p.181ff., for the definition and for examples of ordinary renewal processes.) An ordinary renewal process (or as it is often also called in the literature, a counting process) may be interpreted to be a nonnegative integer valued stochastic process which registers the successive occurrences of an "event" during the time period $(0, x)$. The underlying random variables Y_1 then represent the time that elapses between the

occurrence of the $(i-1)^{\text{st}}$ and the i^{th} "event". In this setup we refer to $T_n = Y_1 + \dots + Y_n$ as the waiting time until the occurrence of the n^{th} "event". The prototype physical experiment which can be modelled by an ordinary renewal process is the following (confer also Figure 8. below):

A light bulb is put into service at time $x=0$,

it fails at time Y_1 ,

then it is immediately replaced by a new bulb,

the second light bulb fails at time $Y_1 + Y_2$,

is then exchanged for a fresh bulb, and so on.

In general, the k^{th} light bulb burns out at time $\sum_{i=1}^k Y_i$,

and is immediately replaced by a $(k+1)^{\text{st}}$ light bulb.

Figure 8

For this and further examples regarding the ordinary renewal process, confer, e.g., S. Karlin and H.M. Taylor (1975), Chapter 5.

The following result concerning the asymptotic behaviour of the ordinary and the generalized renewal processes is due to J.L. Doob (1948) (confer in particular Theorem 9) and C.C. Heyde (1966), respectively.

Lemma 4.0.3.

Given the assumptions (4.0.2.) we have (using the notations (4.0.3.)),

$$\frac{\tilde{N}(x)}{x} \xrightarrow{\text{a.s.}} 1/m \text{ as } x \rightarrow \infty.$$

At this point we remark upon the fact that in general neither the ordinary nor the generalized renewal process possesses either independent or stationary increments (confer also J. Steinebach (1979), p. 13). However, the Poisson process and the process $M = \{M(x) : x \geq 0\}$ of Subsection 4.1.1. share the stationary independent increment property. We note that the proofs of Theorems 4.1.3. and 4.1.7. heavily lean on this property. Furthermore, it is worthwhile noticing at this point that J. Steinebach ((1978), Lemma 4.3., for the ordinary renewal process and (1979), Lemma 4.1., for the generalized renewal process) introduced the so-called method of random epochs (Methode der zufälligen Erneuerungszeitpunkte) in order to provide the renewal process with a kind of asymptotic stationary independent increment property.

We close this part of the introduction with an example of an ordinary renewal process that does not possess stationary increments.

Example 4.0.4.

Let $Y_1, Y_2, \dots$ be independent identically distributed random variables with density function f , where

$$f(y) = \begin{cases} ye^{-y} & y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

This means that Y_1 follows a Gamma-distribution with parameters $\alpha = 2$ and $\beta = 1$. We first show that

$$(4.0.5.) \quad P\{N^*(t) = n\} = \frac{t^{2n+1} e^{-t}}{(2n+1)!} + \frac{t^{2n} e^{-t}}{(2n)!} \quad \text{for each } n=0,1,2,\dots \text{ and } t \geq 0.$$

Here N^* denotes the renewal process based on $Y_1, Y_2, \dots$.

Proof of (4.0.5.).

Since $P\{N^*(t) = n\} = P\{N^*(t) \geq n\} - P\{N^*(t) \geq n+1\}$, we proceed in two steps, namely we first compute $P\{N^*(t) \geq k\}$ for any $k \geq 0$ and then we use the above identity to evaluate (4.0.5.).

Step 1. Computation of $P\{N^*(t) \geq k\}$.

$$P\{N^*(t) \geq k\} = P\{T_k \leq t\} = P\left\{\sum_{i=1}^k Y_i \leq t\right\}.$$

The moment generating function M_{Y_1} of Y_1 is given by

$$M_{Y_1}(t) = (1-t)^{-2} \quad \text{for } t < 1$$

and so, because the Y_i 's are independent and identically distributed, we arrive at

$$M_{\sum_{i=1}^k Y_i}(t) = [M_{Y_1}(t)]^k = (1-t)^{-(2k)}, \quad t < 1:$$

Consequently $\sum_{i=1}^k Y_i \stackrel{D}{=} \Gamma(2k, 1)$, where $\Gamma(2k, 1)$ denotes a random variable with Gamma-distribution with parameters $\alpha = 2k$ and $\beta = 1$. Applying this, we have

$$P\{N^*(t) \geq k\} = \frac{1}{\Gamma(2k)} \int_0^t y^{2k-1} e^{-y} dy = 1 - \sum_{i=0}^{2k-1} \frac{t^i e^{-t}}{i!},$$

where the latter equality sign holds by Exercise 3.32. in R. Hogg and A. Craig (1978). $\square$

Step 2. Evaluation of (4.0.5.).

$$\begin{aligned} P\{N^*(t) = n\} &= - \sum_{i=0}^{2n-1} \frac{t^i e^{-t}}{i!} + \sum_{j=0}^{2n+1} \frac{t^j e^{-t}}{j!} \\ &= \frac{t^{2n} e^{-t}}{(2n)!} + \frac{t^{2n+1} e^{-t}}{(2n+1)!} \end{aligned}$$

as required. $\square$

Next, in order to derive the expected value and the variance of $N^*(t)$, we compute the moment generating function.

$$\begin{aligned} Ee^{sN^*(t)} &= \sum_{k=0}^{\infty} e^{sk} P\{N^*(t) = k\} \quad (\text{all terms are strictly positive}) \\ &= \sum_{k=0}^{\infty} e^{sk} \frac{t^{2k+1} e^{-t}}{(2k+1)!} + \sum_{k=0}^{\infty} e^{sk} \frac{t^{2k} e^{-t}}{(2k)!} \quad \text{by (4.0.5.)} \end{aligned}$$

$$\text{and so } M_{N^*(t)}(s) = Ee^{sN^*(t)} = e^{-t} [e^{-s/2} \sinh(te^{s/2}) + \cosh(te^{s/2})],$$

where this representation is valid for each $s \in \mathbb{R}$.

Taking the first derivative with respect to s gives

$$\frac{d}{ds} Ee^{sN^*(t)} = \frac{1}{2} e^{-t} [-e^{-s/2} \sinh(te^{s/2}) + t \cosh(te^{s/2}) + te^{s/2} \sinh(te^{s/2})];$$

hence

3

of/de

3

$$\begin{aligned}
 EN^*(t) &= \left. \frac{d}{ds} Ee^{sN^*(t)} \right|_{s=0} \\
 &= \frac{e^{-t}}{2} \left[-\frac{e^t - e^{-t}}{2} + te^t \right] = \frac{t}{2} + \frac{1}{4}(e^{-2t} - 1);
 \end{aligned}$$

all identities used for the hyperbolic functions here and in the following can be found in Section 7.4. of L. Leithold (1986). Taking the second derivative of the moment generating function leads to

$$\begin{aligned}
 \frac{d^2}{ds^2} Ee^{sN^*(t)} &= \frac{e^{-t}}{4} \{ e^{-s/2} \sinh(te^{s/2}) - t \cosh(te^{s/2}) + t^2 e^{s/2} \sinh(te^{s/2}) \\
 &\quad + te^{s/2} \sinh(te^{s/2}) + t^2 e^s \cosh(te^{s/2}) \}
 \end{aligned}$$

and therefore

$$\begin{aligned}
 E(N^*(t))^2 &= \left. \frac{d^2}{ds^2} Ee^{sN^*(t)} \right|_{s=0} \\
 &= \frac{e^{-t}}{4} \{ \sinh t - te^{-t} + t^2 e^t \} \\
 &= \frac{1}{4} \left(\frac{1-e^{-2t}}{2} - te^{-2t} + t^2 \right), \text{ whence} \\
 \text{Var } N^*(t) &= \frac{1}{4} \left(\frac{1-e^{-2t}}{2} - te^{-2t} + t^2 \right) - \left\{ \frac{t}{2} + \frac{1}{4}(e^{-2t} - 1) \right\}^2 \\
 &= \frac{t}{4} - \frac{t}{2} e^{-2t} - \frac{1}{16} (e^{-4t} - 1).
 \end{aligned}$$

Our computations are in agreement with the elementary renewal theorem (confer, e.g., H. Taylor and S. Karlin (1984), p.285) which states that

$$\lim_{t \rightarrow \infty} \frac{1}{t} EN^*(t) = m^{-1}$$

because $E\Gamma(2,1) = 2 = m$ and $\lim_{t \rightarrow \infty} \frac{1}{t} EN^*(t) = \lim_{t \rightarrow \infty} \left[\frac{1}{2} + \frac{1}{4} \left(\frac{e^{-2t} - 1}{t} \right) \right] = \frac{1}{2}$.

Considering a refinement of the renewal theorem, namely the inclusion of a second term in the asymptotic expansion, we know (confer again H. Taylor and S. Karlin above, p.289):

$$(4.0.6.) \quad \lim_{t \rightarrow \infty} [EN^*(t) - \frac{t}{m}] = \frac{\sigma^2 - m^2}{2m^2}$$

provided that $\text{Var } Y_1 = \sigma^2 < \infty$.

Since $\text{Var } \Gamma(2,1) = 2 = \sigma^2$, we obtain

$$\frac{\sigma^2 - m^2}{2m^2} = -\frac{1}{4} \quad \text{and} \quad \lim_{t \rightarrow \infty} [EN^*(t) - \frac{t}{m}] = \frac{1}{4} \lim_{t \rightarrow \infty} (e^{-2t} - 1) = -\frac{1}{4}.$$

Lastly, we have under the condition of (4.0.6.) (confer H. Taylor and S. Karlin above, p.290):

$$\lim_{t \rightarrow \infty} \frac{\text{Var } N^*(t)}{t} = \frac{\sigma^2}{m^3},$$

which in our case is seen to be true since

$$\frac{\sigma^2}{m^3} = \frac{1}{4} \quad \text{and} \quad \lim_{t \rightarrow \infty} \frac{\text{Var } N^*(t)}{t} = \frac{1}{4} - \lim_{t \rightarrow \infty} \left[\frac{1}{2} e^{-2t} + \frac{e^{-4t} - 1}{16t} \right] = \frac{1}{4}.$$

We conclude our considerations with an application of the basic renewal theorem (confer, e.g., Theorem 5.1. in S. Karlin and H. Taylor (1975)) which will imply that N^* cannot have stationary increments. If the underlying random variables Y_1 follow a non-arithmetic distribution, then the basic renewal theorem immediately implies that

$$(4.0.7.) \quad \lim_{t \rightarrow \infty} [EN^*(t+s) - EN^*(t)] = \frac{s}{m} \quad \text{for each fixed } s > 0.$$

So, assume for the time being that N^* does indeed possess stationary increments. Then (4.0.7.) leads to

$$\frac{s}{2} = \lim_{t \rightarrow \infty} [EN^*(t+s) - EN^*(t)] = EN^*(s), \text{ for each fixed } s > 0.$$

But the latter statement is in contradiction with our computation

$$EN^*(s) = \frac{s}{2} + \frac{1}{4}(e^{-2s} - 1).$$

Hence N^* cannot have stationary increments. We note at this point

that (4.0.6.) and (4.0.7.) combined show that an ordinary renewal

process with two finite moments can only then be stationary if

$\text{Var } Y_1 = \sigma^2 = m^2 = (EY_1)^2$. This immediately implies that if the random variables Y_1 that underlie the renewal process N^* are given to follow any Gamma-distribution, say $\Gamma(\alpha, \beta)$, then N^* is either the Poisson process or N^* does not have stationary increments. This can be seen as follows:

$$E\Gamma(\alpha, \beta) = \alpha\beta = m \quad \text{and} \quad \text{Var } \Gamma(\alpha, \beta) = \alpha\beta^2 = \sigma^2,$$

whence $m^2 = \sigma^2$ if and only if $\alpha = 1$.

But $\Gamma(1, \beta)$ is an exponential distribution with parameter $\mu = 1/\beta$.

4.0.2. Komlós-Major-Tusnády Type Approximations

In order of their presentation, the following three lemmata provide Komlós-Major-Tusnády type approximations for the Poisson process, for the process $M = \{M(x) : x \geq 0\}$ of Theorem 4.1.7., and for the generalized renewal process. Their corollaries will be crucial in the proofs of Theorems 4.1.3., 4.1.7., and 4.1.9., respectively. In the light of Lemma 4.0.5. below, we first recall the classical Komlós-Major-Tusnády approximation of partial sums by a suitably constructed Wiener process which was displayed in Section 1.1. of Chapter 1.

Let N be a Poisson process with (uniform parameter) $\mu > 0$.

Put

$$K_i^{(1)} := N(i) - N(i-1) \quad \text{for } 1 \leq i \leq n$$

and

$$K_i^{(2)} := K_{i,n}^{(2)} := N(n-i+1) - N(n-i) \quad \text{for } 1 \leq i \leq n.$$

Then the random variables $K_i^{(1)}$, $1 \leq i \leq [n/2]$ (and similarly the random variables $K_i^{(2)}$, $1 \leq i \leq [n/2] - 1$) are independent Poisson variables with parameter $\mu > 0$. (Confer, e.g., P. Billingsley (1986), p.309ff.) Hence $EK_i^{(1)} = EK_j^{(2)} = EN(1) = \mu$ and $\text{Var } K_i^{(1)} = \text{Var } K_j^{(2)} = \text{Var } N(1) = \mu$ for each $1 \leq i \leq [n/2]$ and $1 \leq j \leq [n/2] - 1$. Since the moment generating function M of any Poisson variable is finite over the whole real line, the Komlós-Major-Tusnády approximation holds for partial sums of Poisson random variables. Therefore, there exists a Wiener process $W = W^{(1)}$ approximating $\sum_{i=1}^k K_i^{(1)}$ such that we have for each $z \in \mathbb{R}$ and for every $n \geq 1$:

$$(4.0.8.) \quad P\left\{ \max_{1 \leq k \leq n} \left| \sum_{i=1}^k (K_i^{(1)} - \mu) - \mu W^{(1)}(k) \right| > C \log n + z \right\} \leq K e^{-\lambda z}$$

for some absolute constants $C, K, \lambda > 0$.

Using a telescopic sum argument, we obtain

$$(4.0.9.) \quad N(k) - \mu k = \sum_{i=1}^k (K_i^{(1)} - \mu) \quad \text{for } 1 \leq k \leq n$$

since $N(0) = 0$. Hence (4.0.8.) may be rephrased as

$$(4.0.8.*) \quad P\left\{ \max_{1 \leq k \leq n} |N(k) - \mu k - \mu W^{(1)}(k)| > C \log n + z \right\} \leq K e^{-\lambda z}$$

As mentioned before, the random variables $K_i^{(2)}$, $1 \leq i \leq n$, are also

independent Poisson random variables with parameter $\mu > 0$. Hence

another application of the Komlós-Major-Tusnády Theorem yields: there

exists a Wiener process $W = W^{(2)}$, which couples with the partial

sums $\sum_{i=1}^{n-k} K_i^{(2)}$, such that for every $z \in \mathbb{R}$ and for each $n \geq 1$ we have:

$$(4.0.10.) \quad P\left\{ \max_{0 \leq k \leq n} \left| \sum_{j=1}^{n-k} (K_j^{(2)} - \mu) - \mu W^{(2)}(n-k) \right| > C \log n + z \right\} \leq K e^{-\lambda z}$$

Again using a telescopic sum argument, we obtain

$$(4.0.9.*) \quad N(n) - N(k) = \sum_{j=1}^{n-k} K_j^{(2)} \quad \text{for } 0 \leq k \leq n,$$

whence we may rewrite (4.0.10.) as

$$(4.0.10.*) \quad P\left\{ \max_{0 \leq k \leq n} |N(n) - N(k) - (n-k)\mu - \mu W^{(2)}(n-k)| > C \log n + z \right\} \leq K e^{-\lambda z}$$

Since we are only interested in the nonoverlapping maxima

$$\max_{1 \leq k \leq n/2} \sum_{i=1}^k (K_i^{(1)} - \mu) \quad \text{and} \quad \max_{n/2 \leq k \leq n} \sum_{j=1}^{n-k} (K_j^{(2)} - \mu),$$

we will indicate this in the approximating Wiener processes $W^{(1)}$ and $W^{(2)}$ by writing:

$$\begin{aligned} W_n^{(1)}(x) &:= W^{(1)}(x) && \text{for all } 0 \leq x \leq n/2 \text{ and} \\ W_n^{(2)}(n-x) &:= W^{(2)}(n-x) && \text{for all } n/2 \leq x \leq n. \end{aligned}$$

With these notations we immediately obtain the second half of the following lemma.

Lemma 4.0.5.

Let $N = \{N(x) : x \geq 0\}$ denote a Poisson process with intensity parameter $\mu > 0$. Then there exist two sequences of Wiener processes,

$$\{W_n^{(1)}\}_{n \geq 1} \text{ and } \{W_n^{(2)}\}_{n \geq 1} \text{ such that}$$

$$(4.0.11.) \quad W_n^{(1)} \text{ and } W_n^{(2)} \text{ are stochastically independent for each } n \geq 2$$

and

$$(4.0.12.) \quad \left\{ \begin{aligned} &P \left\{ \max_{1 \leq k \leq n/2} \left| \sum_{i=1}^k (K_i^{(1)} - \mu) - \mu W_n^{(1)}(k) \right| > A_1 \log n + z \right\} \leq B_1 e^{-C_1 z} \\ &P \left\{ \max_{\frac{n}{2} \leq k \leq n} \left| \sum_{i=1}^{n-k} (K_i^{(2)} - \mu) - \mu W_n^{(2)}(n-k) \right| > A_1^* \log n + z \right\} \leq B_1^* e^{-C_1^* z} \end{aligned} \right\} \text{ as well as}$$

holds for every $n \geq 2$ and $z \in \mathbb{R}$.

Here $A_1, B_1, C_1, A_1^*, B_1^*,$ and $C_1^* > 0$ are absolute constants.

The first statement of Lemma 4.0.5. follows from the independence of $\sum_{i=1}^k K_i^{(1)}$ and $\sum_{i=1}^{n-k} K_i^{(2)}$, which is in turn implied by the independent increment property of the Poisson process N . In the light of (4.0.9.) and (4.0.9.*) we may reformulate Lemma 4.0.5. by saying

Lemma 4.0.5. *

(Let $N = \{N(x) : x \geq 0\}$ denote a Poisson process with parameter $\mu > 0$. Then there exist two sequences of Wiener processes, $\{W_n^{(1)}\}_{n \geq 1}$ and $\{W_n^{(2)}\}_{n \geq 1}$ such that (4.0.8.) holds and

$$(4.0.12.*) \left\{ \begin{array}{l} P \left\{ \max_{1 \leq k \leq n/2} | (N(k) - \mu k) - \mu^{1/2} W_n^{(1)}(k) | > A_1 \log n + z \right\} \leq B_1 e^{-C_1 z} \\ \text{as well as} \\ P \left\{ \max_{n/2 \leq k \leq n} | (N(n) - N(k) - (n-k)\mu) - \mu^{1/2} W_n^{(2)}(n-k) | > A_1^* \log n + z \right\} \leq B_1^* e^{-C_1^* z} \end{array} \right.$$

is satisfied for each $n \geq 2$ and $z \in \mathbb{R}$.

Corollary 4.0.6.

Under the assumptions of Lemma 4.0.5. we have as $n \rightarrow \infty$,

$$(4.0.13.) \quad \sup_{\delta \leq x \leq n/2} x^{v-1/2} |N(x) - \mu x - \mu^{1/2} W_n^{(1)}(x)| = O_p(1) \quad \text{and}$$

$$(4.0.14.) \quad \sup_{n/2 \leq x \leq n-\delta} (n-x)^{v-1/2} |N(n) - N(x) - (n-x)\mu - \mu^{1/2} W_n^{(2)}(n-x)| = O_p(1)$$

for each $\delta > 0$ and $0 \leq v < 1/2$.

Proof of Corollary 4.0.6.

This proof will be divided into two main parts according to the format of Corollary 4.0.6. For the proof of (4.0.13.) we will first establish a kind of universal nearness of $N(y)$ and $N([y])$; then we will represent N , evaluated at an integer point, as a partial sum of appropriately chosen independent identically distributed random variables (confer also (4.0.9.)); finally we will conclude the proof

with an application of the Komlós-Major-Tusnády approximation (4.0.8.*).

The verification of (4.0.13.) as well as that of (4.0.14.) will be heavily based upon the stationary increment property of the Poisson process.

Proof of Statement (4.0.13.).

$$\begin{aligned}
 & \sup_{\delta \leq x \leq n/2} x^{v-1/2} |N(x) - \mu x - \mu^{1/2} W_n^{(1)}(x)| \\
 & \leq \sup_{\delta \leq x \leq n/2} x^{v-1/2} \{ |N(x) - N([x])| + \mu(x - [x]) + \mu^{1/2} |W_n^{(1)}(x) - W_n^{(1)}([x])| \\
 & \quad + |N([x]) - \mu[x] - \mu^{1/2} W_n^{(1)}([x])| \} \\
 & \leq \sup_{\delta \leq x \leq n/2} x^{v-1/2} |N(x) - N([x])| + \mu^{1/2} \sup_{\delta \leq x \leq n/2} x^{v-1/2} |W_n^{(1)}(x) - W_n^{(1)}([x])| \\
 & \quad + \max_{1 \leq k \leq n/2} k^{v-1/2} |N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| + O(1) \\
 & \leq \sup_{\delta \leq x \leq n/2} x^{v-1/2} (N([x]+1) - N([x])) + \mu^{1/2} \max_{1 \leq k \leq n/2} k^{v-1/2} \sup_{0 \leq h \leq 1} |W_n^{(1)}(k+h) - W_n^{(1)}(k)| \\
 & \quad + \mu^{1/2} \sup_{\delta \leq x \leq 1} x^{v-1/2} |W_n^{(1)}(x)| \\
 & \quad + \max_{1 \leq k \leq n/2} k^{v-1/2} |N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| + O(1)
 \end{aligned}$$

by (4.0.4.) (vii) and (4.0.5.)

and for any standard Wiener process $W = \{W(x) : x \geq 0\}$

$$\begin{aligned}
 & \stackrel{D}{=} N(1) \cdot \sup_{\delta \leq x \leq n/2} x^{v-1/2} + \mu^{1/2} \max_{1 \leq k \leq n/2} k^{v-1/2} \sup_{0 \leq h \leq 1} |W(k+h) - W(k)| + \mu^{1/2} \sup_{\delta \leq x \leq 1} x^{v-1/2} |W(x)| \\
 & (*) \quad + \max_{1 \leq k \leq n/2} k^{v-1/2} |N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| + O(1)
 \end{aligned}$$

$$= \mu^{1/2} \max_{1 \leq k \leq n/2} k^{v-1/2} \sup_{0 \leq h \leq 1} |W(k+h) - W(k)| + \max_{1 \leq k \leq n/2} k^{v-1/2} |N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| + O_p(1).$$

Where indicated by (*) above, we used the stationary increment property of a Poisson process. It is now sufficient to show that

$$(4.0.16.) \quad \max_{1 \leq k \leq n/2} k^{v-1/2} \sup_{0 \leq h \leq 1} |W(k+h) - W(k)| = O_p(1) \quad \text{as } n \rightarrow \infty \quad \text{and}$$

$$(4.0.17.) \quad \max_{1 \leq k \leq n/2} k^{v-1/2} |N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| = O_p(1) \quad \text{as } n \rightarrow \infty$$

in order to complete the proof of (4.0.13.).

Proof of (4.0.16.).

Let $K > 3$ be a fixed positive constant, then

$$\begin{aligned} & P \left\{ \max_{1 \leq k \leq n/2} k^{v-1/2} \sup_{0 \leq h \leq 1} |W(k+h) - W(k)| > \sqrt{2K} \right\} \\ & \leq \sum_{k=1}^n P \left\{ \sup_{0 \leq h \leq 1} |W(k+h) - W(k)| > \sqrt{2K} \cdot \log k \right\} \\ & \leq \sum_{k=1}^n P \left\{ \sup_{0 \leq s \leq k-1} \sup_{0 \leq h \leq 1} |W(s+h) - W(s)| > \sqrt{2K} \cdot \sqrt{\log k} \right\} \end{aligned}$$

and by Lemma 1.2.1. in M. Csörgő and P. Révész (1981) (take $T = k$, $h=1$, $\varepsilon=1$, $v = \sqrt{\log k}$ there)

$$\leq C \cdot \sum_{k=1}^n k \cdot \exp \left\{ -\frac{2K}{3} \cdot \log k \right\} \quad \text{for some constant } C > 0$$

$$\leq C \cdot \sum_{k=1}^{\infty} k^{1-2K/3} < \infty \quad \text{if } K > 3.$$

Thus, we arrive at

$$\limsup_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} P \left\{ \max_{1 \leq k \leq n/2} k^{\nu-1/2} \sup_{0 \leq h \leq 1} |W(k+h) - W(k)| > K \right\} = 0,$$

whence the definition of boundedness in probability immediately implies statement (4.0.16.).

Proof of (4.0.17.)

With i taking on the role of k and then with k in lieu of n , (4.0.8.*) implies immediately:

$$|N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| \stackrel{\text{a.s.}}{=} O(\log k) \quad (k \rightarrow \infty).$$

And so we have as $n \rightarrow \infty$

$$\begin{aligned} & \max_{1 \leq k \leq n/2} k^{\nu-1/2} |N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| \\ &= \max_{1 \leq k \leq n/2} k^{\nu-1/2} |N(k) - \mu k - \mu^{1/2} W_n^{(1)}(k)| \quad (\text{by definition of } W_n^{(1)}) \\ &\stackrel{\text{a.s.}}{=} O(1) \cdot \max_{1 \leq k \leq n/2} (\log k) \cdot k^{\nu-1/2} \\ &= O(1). \end{aligned}$$

Proof of (4.0.14.)

Let $\delta > 0$ and $0 \leq \nu < 1/2$, then in the same vein as before:

$$\begin{aligned} & \max_{n/2 \leq x \leq n-\delta} (n-x)^{\nu-1/2} |N(n) - N(x) - (n-x)\mu - \mu^{1/2} W_n^{(2)}(n-x)| \\ &\leq \max_{n/2 \leq k \leq n-1} (n-k)^{\nu-1/2} |N(n) - N(k) - (n-k)\mu - \mu^{1/2} W_n^{(2)}(n-k)| \cdot \sup_{n/2 \leq x \leq n-\delta} \left(\frac{n-x}{n-[x]} \right)^{\nu-1/2} \\ &\quad + \delta^{\nu-1/2} \max_{n/2 \leq x \leq n-\delta} |N(x) - N([x])| + \mu \delta^{\nu-1/2} + \mu^{1/2} \delta^{\nu-1/2} \sup_{\delta \leq x \leq n} |W_n^{(2)}(x) - W_n^{(2)}([x])|. \end{aligned}$$

Moreover, for any constant $K > 0$,

$$P\left\{\sup_{n/2 \leq x \leq n-\delta} |N(x) - N([x])| > K\right\} = P\{|N(1)| > K\}$$

by the stationary increment property of N and hence

$$\lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty} P\left\{\sup_{n/2 \leq x \leq n-\delta} |N(x) - N([x])| > K\right\} = 0;$$

$$\text{Similarly, } \sup_{\delta \leq x \leq n} |w_n^{(2)}(x) - w_n^{(2)}([x])| = O_p(1)$$

by the stationary increment property of $w_n^{(2)}$ and the modulus of nondifferentiability of a Wiener process (confer, e.g., Theorem 1.6.1.

in M. Csörgő and P. Révész (1981)). Putting these results together, we obtain (4.0.14.) provided that we can show:

$$(4.0.14.*) \quad \max_{n/2 \leq k \leq n-1} (n-k)^{\nu-1/2} |N(n) - N(k) - (n-k)\mu - \mu^{1/2} w_n^{(2)}(n-k)| = O_p(1).$$

Proof of (4.0.14.*)

First, rewrite statement (4.0.10.) with i instead of k and $n-k$ in place of n :

$$(4.0.10.***) \quad P\left\{\max_{0 \leq i \leq n-k} \left| \sum_{j=1}^{n-k-i} (K_j^{(2)} - \mu) - \mu^{1/2} w^{(2)}(n-k-i) \right| > C \log(n-k) + z \right\} \leq Ke^{-\lambda z}$$

for every $1 \leq k \leq n-1$, $z \in \mathbb{R}$.

Then we have

$$\begin{aligned} & P\{|N(n) - N(k) - (n-k)\mu - \mu^{1/2} w^{(2)}(n-k)| > C \log(n-k) + z\} \\ &= P\left\{\left| \sum_{j=1}^{n-k} (K_j^{(2)} - \mu) - \mu^{1/2} w^{(2)}(n-k) \right| > C \log(n-k) + z\right\} \\ &\leq P\left\{\max_{0 \leq i \leq n-k} \left| \sum_{j=1}^{n-k-i} (K_j^{(2)} - \mu) - \mu^{1/2} w^{(2)}(n-k-i) \right| > C \log(n-k) + z\right\} \end{aligned}$$

$\leq K e^{-\lambda z}$ by (4.0.10.**).

Now, choosing $z = O(\log(n-k))$, we have immediately

$$|N(n) - N(k) - (n-k)\mu - \mu^{(2)} W^{(2)}(n-k)| \stackrel{a.s.}{=} O(\log(n-k)) \text{ as } (n-k) \rightarrow \infty.$$

Hence, an argument similar to that ending the proof of (4.0.17.) will also complete this part of the proof of Corollary 4.0.6. We note in passing that the proofs of (4.0.17.) as well as (4.0.14.*) actually verify stronger results than originally claimed, i.e. almost sure convergence instead of weak convergence is shown. ■

Regarding Corollary 4.0.6. (and similarly any statement of the same form in this subsection), it should be noted again that we went through the effort of creating the two approximating sequences not because sequences (rather than one single Wiener process) are necessary for the Komlós-Major-Tusnády type approximations, but because we need two independent Wiener processes approximating N for $0 \leq x \leq n/2$ and $n/2 \leq x \leq n$, respectively. This fact will become evident from the proofs in the next subsection.

The same holds for the processes M and $\tilde{N}$ which will be considered now. Before turning to the next of the aforementioned three lemmata we wish to acknowledge that the main idea for the proof of Corollary 4.0.6. stems from M. Csörgö and L. Horváth (1987) (confer here in particular the proof of Theorem 1.1.). The detailed proof was given here for reasons explained in the beginning of Subsection 4.1.0.

Lemma 4.0.,7.

Let $W = \{W(t) : t \geq 0\}$ be an arbitrary standard Wiener process,

$\lambda > 0$ a constant and $M(t) := \inf \{x \geq 0: \lambda W(t) = t - x\}$, $0 \leq t < \infty$. Then we can define another standard Wiener process $\bar{W} = \{\bar{W}(t): t \geq 0\}$ such that for all $T > 0$ and $z \in \mathbb{R}$ we have

$$(4.0.18.) \quad P\left\{\sup_{0 \leq t \leq T} \left| \frac{M(t) - t}{\lambda} - \bar{W}(t) \right| > A_2 \log T + z\right\} \leq B_2 e^{-C_2 z^2},$$

where A_2 , B_2 , and C_2 are strictly positive constants.

For a proof of Lemma 4.0.7. we refer to M. Csörgő, L. Horváth and J. Steinebach (1987) (confer here in particular Theorem 2.1.). Lemma 4.0.7. clearly represents a Komlós-Major-Tusnády type approximation for the stochastic process $M = \{M(t): t \geq 0\}$ whose significance is discussed in detail in the following section. At this point, we only note that (4.0.19.) holds:

- the stochastic process M has stationary independent increments;
- the random variable $M(t)$, $t \geq 0$ fixed, possesses a finite moment generating function in an open nonempty neighbourhood of zero.

For a proof of these facts we again refer to Theorem 2.1. in the aforementioned publication.

Corollary 4.0.8.

Given the conditions and notations of Lemma 4.0.7., there exist two collections of Wiener processes, say $\{\bar{W}_T^{(1)}\}_{T \geq 0}$ and $\{\bar{W}_T^{(2)}\}_{T \geq 0}$ such that simultaneously

$$(4.0.20.) \quad P\left\{\sup_{0 \leq x \leq T/2} \left| \frac{M(x) - x}{\lambda} - \bar{W}_T^{(1)}(x) \right| > A_2 \log T + z\right\} \leq B_2 e^{-C_2 z^2}$$

$$(4.0.21.) \quad P\left\{\sup_{T/2 \leq x \leq T} \left| \frac{M(T) - M(x) - (T - x)}{\lambda} - \bar{W}_T^{(2)}(T - x) \right| > A_2 \log T + z\right\} \leq B_2 e^{-C_2 z^2}$$

and

(4.0.22.) $\bar{W}_T^{(1)}$ and $\bar{W}_T^{(2)}$ are independent Wiener processes for each $T \geq 0$ hold.

Proof of Corollary 4.0.8.

(4.0.18.) immediately implies

$$P\left\{\sup_{0 \leq s \leq 1} \left| \frac{M(sT) - sT}{\lambda} - \bar{W}(sT) \right| > A_2 \log T + z\right\} \leq B_2 e^{-C_2 z}.$$

When now changing the notation by introducing $x = sT$, we do not wish to suppress the dependency of $\bar{W}$ on T . Thus we define $\bar{W}_T^{(1)} = \{\bar{W}_T^{(1)}(x) : 0 \leq x \leq T/2\}$ to be that Wiener process which, according to Lemma 4.0.7., satisfies in particular (4.0.20.). (Note that restricting the region over which a supremum is taken is a monotonically nonincreasing operation.)

In the same vein Lemma 4.0.7. also ensures that we may define standard Wiener processes $\bar{W}^{(3)} = \{\bar{W}^{(3)}(x) : x \geq 0\}$ such that

$$(4.0.21.*) \quad P\left\{\sup_{T/2 \leq x \leq T} \left| \frac{M(T-x) - (T-x)}{\lambda} - \bar{W}^{(3)}(T-x) \right| > A_2 \log T + z\right\} \leq B_2 e^{-C_2 z}$$

holds. But in the light of (4.0.19.) (stationary increment property of M) we also have:

$$\{M(T-x) : 0 \leq x \leq T\} \stackrel{D}{=} \{M(T) - M(x) : 0 \leq x \leq T\} \quad \text{for each } T > 0.$$

For the time being let us assume that T is integer valued; then we will write n instead of T and we have

$$\{M(n-x) : 0 \leq x \leq n\} \stackrel{D}{=} \{M(n) - M(x) : 0 \leq x \leq n\} \quad \text{for each } n \geq 1.$$

Next, we will apply Lemma 4.4.4. in M. Csörgő and P. Révész (1981)

(confer also, for a more general setting, M. Csörgö (1983) Lemma 3.1.1. and Lemma 3.1.2.) to obtain an inequality similar to (4.0.21.*) but with $M(n)-M(x)$ in lieu of $M(n-x)$:

in the aforementioned Lemma 4.4.4. we choose:

$\mu = \mu_n$ to be the joint distribution of $M(n-x)$ and $\bar{W}_n^{(3)}(n-x)$,

$\xi = \xi_n = M(n)-M(\cdot)$, and

$\eta = \eta_n$ to be an appropriate Wiener process, for each $n \geq 1$, say $\eta_n = \bar{W}_n^{(2)}$.

Then Lemma 4.4.4. above and Kolmogorov's extension theorem (confer

Theorem 1.1.22.) imply:

$$(4.0.21.**)\ P\left\{\sup_{n/2 \leq x \leq n} \left| \frac{M(n)-M(x)-(n-x)}{\lambda} - \bar{W}_n^{(2)}(n-x) \right| > A_2 \log n + z\right\} \leq B_2 e^{-C_2 z^2}$$

holds for each $n \geq 2$.

In order to obtain (4.0.21.) from here, it is enough to repeat the arguments exhibited in the end of the proof of Theorem 2.1. in M. Csörgö, L. Horváth, and J. Steinebach (1987) (confer especially (2.6.) through (2.8.) here).

Since, again by (4.0.19.), M also possesses independent increments, we conclude that for each $0 \leq x_0 \leq T/2$ and $T/2 \leq x_1 \leq T$ ($T \geq 0$ fixed), the random variables $M(x_0)$ and $(M(T)-M(x_1))$ are independent. Figure 9. below illustrates this fact. Now the independence of $M(x_0)$ and $(M(T)-M(x_1))$ implies the independence of $\bar{W}_T^{(1)}$ and $\bar{W}_T^{(2)}$ by construction of these two Wiener processes. ■

Unfortunately, for the purposes of the next section, the coupling result in Corollary 4.0.8. is not strong enough. There we will need the following improved version of Corollary 4.0.8.

$M(T) - M(x_1)$ only depends on what happens after time x_1 up to time T .

$M(x_0)$ only depends on what happens up to time x_0

Figure 9.

Corollary 4.0.8.*

Under the assumptions of Corollary 4.0.8., we have

$$(4.0.23.) \quad \sup_{\delta \leq x \leq n/2} x^{\nu-1/2} \left| \frac{M(x) - x}{\lambda} - \bar{W}_n^{(1)}(x) \right| = O_p(1) \quad \text{as } n \rightarrow \infty,$$

$$(4.0.24.) \quad \sup_{\frac{n}{2} \leq x \leq n-\delta} (n-x)^{\nu-1/2} \left| \frac{M(n) - M(x) - (n-x)}{\lambda} - \bar{W}_n^{(4)}(n-x) \right| = O_p(1) \quad \text{as } n \rightarrow \infty,$$

and (4.0.22.) with $\bar{W}_n^{(4)}$ instead of $\bar{W}_n^{(2)}$ hold simultaneously. Here $\delta > 0$ and $0 \leq \nu < 1/2$.

Proof of Corollary 4.0.8.*

Lemma 4.0.7. immediately implies

$$\left| (M(x) - x)/\lambda - \bar{W}(x) \right| \stackrel{a.s.}{=} O(\log x) \quad \text{as } x \rightarrow \infty.$$

This statement in turn implies (4.0.23.) when taking into account that

$(\log x)/x^\alpha \rightarrow 0$ as $x \rightarrow \infty$ for each $\alpha > 0$ fixed and that

$$\{\bar{W}(x) : 0 \leq x \leq n/2\} = \{\bar{W}_n^{(1)}(x) : 0 \leq x \leq n/2\}.$$

On the other hand, (4.0.18.) with $n-x$ in lieu of T , says

$$P\left\{\sup_{0 \leq t \leq n-x} \left| \frac{M(t)-t}{\lambda} - \bar{W}(t) \right| > A_2 \log(n-x) + z\right\} \leq B_2 e^{-C_2 z},$$

hence

$$P\left\{\left| (M(n-x) - (n-x))/\lambda - \bar{W}(n-x) \right| > A_2 \log(n-x) + z\right\} \leq B_2 e^{-C_2 z}.$$

Now applying the same methodology that was used in the proof of

Corollary 4.0.8., we can define an appropriate Wiener process $\bar{W}_n^{(4)}$ such that

$$P\left\{\left| (M(n) - M(x) - (n-x))/\lambda - \bar{W}_n^{(4)}(n-x) \right| > A_2 \log(n-x) + z\right\} \leq B_2 e^{-C_2 z}.$$

Consequently, taking $z = \text{constant} \cdot \log(n-x)$ for an appropriately large constant, we obtain as $(n-x) \rightarrow \infty$

$$\left| (M(n) - M(x) - (n-x))/\lambda - \bar{W}_n^{(4)}(n-x) \right| \stackrel{\text{a.s.}}{=} O(\log(n-x)).$$

From here (4.0.24.) is immediate. ■

Before we turn our attention to generalized renewal processes, a remark on an alternative way of proving Corollary 4.0.8.* seems to be called for. Realizing that

$$M(n) - M(k) = \sum_{i=1}^{n-k} (M(n-i+1) - M(n-i)),$$

where $K_1 := M(n-i+1) - M(n-i)$ are independent identically distributed random variables with distribution $M(1)$ and hence with finite moment generating function in a neighbourhood of 0, Corollary 4.0.8.* could also be verified by imitating the proof of Corollary 4.0.6. It is for this reason that previously the proof of Corollary 4.0.6. was discussed in such great detail.

Lemma 4.0.9. (Confer Corollary 4.2. in M. Csörgő, L. Horváth and J. Steinebach (1987))

Let $\{Y_i\}_{i \in \mathbb{N}}$ be a sequence of independent identically distributed random variables satisfying

$$(4.0.2.) \quad EY_1 =: m > 0 \quad \text{and} \quad 0 < \text{Var } Y_1 =: \sigma^2 < \infty.$$

Furthermore, let $\tilde{N} = \{\tilde{N}(x) : x \geq 0\}$ denote the generalized renewal process based on $\{Y_i\}_{i \in \mathbb{N}}$ as defined in (4.0.4.). Assume that $Ee^{tY_1} < \infty$ for $|t| < t_0$, $t_0 > 0$. Then there exists a standard Wiener process $\tilde{W} = \{\tilde{W}(t) : t \geq 0\}$ such that for every $z > 0$

$$(4.0.25.) \quad P\left\{\sup_{0 \leq t \leq T} \left| \frac{\tilde{N}(t) - t/m}{\sigma/m^{3/2}} - \tilde{W}(t) \right| > A_3 \log T + z\right\} \leq B_3 e^{-C_3 z}$$

where A_3 , B_3 , and C_3 are positive constants.

At this point the original aim of this subsection was to draw a corollary from Lemma 4.0.9. in the same way Corollary 4.0.8.* was concluded from Lemma 4.0.7. Unfortunately, major difficulties were encountered in trying to verify that

$$(4.0.26.) \quad \sup_{\frac{n}{2} \leq x \leq n-\delta} (n-x)^{v-1/2} \left| \frac{\tilde{N}(n) - \tilde{N}(x) - (n-x)/m}{\sigma/m^{3/2}} - \tilde{W}_n^{(2)}(n-x) \right| = O_p(1) \quad (n \rightarrow \infty).$$

It appears to be reasonable to believe that assertion (4.0.26.) is true. We can at this time only present a proof of the "lower" half of the desired corollary.

Corollary 4.0.10.

Under the conditions of Lemma 4.0.9. we have:

there exists a sequence $\{\tilde{W}_n^{(1)}\}_{n \geq 1}$ of Wiener processes satisfying

$$(4.0.27.) \quad \sup_{\delta \leq x \leq n/2} x^{\nu-1} \left| \frac{N(x) - x/m}{\sigma/m^{3/2}} - \tilde{W}_n^{(1)}(x) \right| = O_p(1) \quad \text{as } n \rightarrow \infty.$$

Here again $\delta > 0$ and $0 \leq \nu < 1$ are fixed constants.

Proof of Corollary 4.0.10.

Imitating the applicable half of the proof of Corollary 4.0.6.,
the proof of Corollary 4.0.10. is immediate now. ■

4.0.3. Conclusions

The present section of Chapter 4 was originally intended to provide the basic background and tools necessary to extend the concept of a pontogram. It was thought that as an intermediary step results should be proved for the process M , the nonhomogeneously shifted first exit time of a Wiener process. Then these results would be carried over to generalized renewal processes by virtue of M and a general renewal process being "close" to another. This programme could not be carried out. At this point we only suggest one possible starting point for further research. A technique similar to J. Steinebach's "random epoch"-method could be utilized. The idea behind using the random epochs would be as follows: although a general renewal process does not normally possess either independent or stationary increments, it could be stopped at random times in order to create at least asymptotically stationary and independent increments. The main problem in doing so is going to be the problem of being able to stop the process often enough (though still at random times). Here often enough would have to be approximately once inbetween each two consecutive integers.

Section 4.1. Asymptotic Distributions for Pontograms and Related Processes

4.1.0. Improved Results for Kendall-Kendall Pontograms

In Subsection 4.0.0. we introduced the Kendall-Kendall t -pontogram as the plot of $Z(t)$ against t , $0 \leq t \leq 1$, where

$$Z(t) = \frac{1}{\sqrt{R}} \cdot \frac{N(t) - tR}{\sqrt{t(1-t)}}, \quad 0 < t < 1, \quad Z(0) = Z(1) = 0.$$

Using the name pontogram also for the stochastic process $Z = \{Z(t): 0 \leq t \leq 1\}$ itself, M. Csörgő and L. Horváth (1987) introduce a similar pontogram by defining

$$K_n(t) := n^{-1/2} \{N(nt) - tN(n)\} \quad \text{for } 0 \leq t \leq 1 \text{ and } n \geq 1.$$

Note that the normalizing factor $(t(1-t))^{-1/2}$ will be absorbed into the weighted sup-norm rather than kept in the definition of the stochastic process $K_n = \{K_n(t): 0 \leq t \leq 1\}$. Consequently we have for $R = N(n)$

$$Z(t) = \left(\frac{n}{N(n)}\right)^{1/2} \cdot \frac{K_n(t)}{\sqrt{t(1-t)}} \quad \text{for } 0 \leq t \leq 1.$$

We observe that, since a Poisson process with intensity parameter $\mu > 0$ evaluated at time x has the mean μx ,

$$EZ(t) = EK_n(t) = 0 \quad \text{for each } t \in [0, 1].$$

Kendall and Kendall examine the asymptotic behaviour of the t -pontogram (asymptotic in n) for t -values ranging from ε_0 to $1 - \varepsilon_1$, where $\varepsilon_0, \varepsilon_1 > 0$ are fixed constants. According to Kendall and Kendall, it

would be of practical statistical interest to be able to remove this restriction that the t-pontogram has to be bounded away from zero and one. This is the point where the aforementioned paper of M. Csörgő and L. Horváth (1987) comes into the picture since the authors extend their examination of the t-pontogram to intervals of the form $[\delta/n, 1-\delta/n]$ for arbitrary positive fixed $\delta > 0$. Both, the Kendall-Kendall and the Csörgő-Horváth publication, provide an interesting anchor point for further research. As far as the original Kendall-Kendall paper is concerned, the question that arises from there is: Is it possible to obtain asymptotic results in intervals bounded away from the endpoints for pontograms which are based on generalized or ordinary renewal processes instead of Poisson processes? This question will be answered positively in the following theorem. Given that this kind of an extension of the original pontogram concept is possible, one might then ask the question: which of the Csörgő-Horváth results can be generalized in a similar way? An answer to the last question will also be attempted in this section.

Theorem 4.1.0.

Denote by $\tilde{N} = \{\tilde{N}(x) : x \geq 0\}$ a generalized renewal process as introduced in Definition 4.0.2. Let $\tilde{N}$ satisfy (4.0.2.) and (4.0.3.). Then we have for each $\varepsilon > 0$ fixed and $0 \leq v < 1/2$: the original probability space on which $\tilde{N}$ is defined can be extended so that there exists a sequence of Brownian bridges $\tilde{B}_n$ with

$$(4.1.0.) \quad \sup_{\varepsilon t \leq 1-\varepsilon} n^v \frac{|n^{-1/2}(\tilde{N}(nt) - t\tilde{N}(n)) / (\sigma n^{-3/2}) - \tilde{B}_n(t)|}{(t(1-t))^{1/2-v}} \text{ a.s. } o(1) \text{ as } n \rightarrow \infty$$

Proof of Theorem 4.1.0.

It is a consequence of Lemma 4.0.9. that

$$\left| \frac{\tilde{N}(nt) - (nt)/m}{\sigma m^{-3/2}} - \tilde{W}(nt) \right| \stackrel{a.s.}{=} O(\log(nt)) \quad \text{as } n \rightarrow \infty.$$

Noticing now that $\tilde{B}_n(t) := n^{-1/2} \{ \tilde{W}(nt) - t\tilde{W}(n) \}$ defines a Brownian bridge for each $n \geq 1$, we obtain

$$\begin{aligned} & \sup_{\varepsilon \leq t \leq 1-\varepsilon} n^{\nu} \frac{|n^{-1/2} (\tilde{N}(nt) - t\tilde{N}(n)) / (\sigma m^{-3/2}) - \tilde{B}_n(t)|}{(t(1-t))^{\frac{1}{2}-\nu}} \\ & \leq \sup_{\varepsilon \leq t \leq 1-\varepsilon} \frac{n^{\nu-1/2}}{(t(1-t))^{\frac{1}{2}-\nu}} \left| \frac{\tilde{N}(nt) - (nt)/m}{\sigma m^{-3/2}} - \tilde{W}(nt) \right| \\ & \quad + \sup_{\varepsilon \leq t \leq 1-\varepsilon} \frac{n^{\nu-1/2}}{(t(1-t))^{\frac{1}{2}-\nu}} t \left| \frac{\tilde{N}(n) - n/m}{\sigma m^{-3/2}} - \tilde{W}(n) \right| \\ & \stackrel{a.s.}{=} \sup_{\varepsilon \leq t \leq 1-\varepsilon} \frac{O(\log(nt))}{(nt)^{\frac{1}{2}-\nu}} : (1-t)^{\nu-1/2} + \frac{O(\log n)}{n^{\frac{1}{2}-\nu}} \cdot \sup_{\varepsilon \leq t \leq 1-\varepsilon} t^{\nu+1/2} (1-t)^{\nu-1/2} \\ & = o(1). \end{aligned}$$

Remarks 4.1.1.

(a) It is clear from the proof of Theorem 4.1.0. that the lower bound ε can be taken to be $\varepsilon_1 > 0$ and the upper bound $1-\varepsilon$ can be replaced by $1-\varepsilon_2$, where $0 < \varepsilon_1 < 1-\varepsilon_2 < 1$. That is to say that there is no need for "symmetric" bounds.

(b) A careful examination of the second last equality in the proof of Theorem 4.1.0. also shows that we can choose the lower bound ε to be $\varepsilon_n = \delta/n$ for any given constant $\delta > 0$ provided we replace the $o(1)$ by a $O(1)$ in (4.1.0.). I.e. as far as the lower bound of the t -interval is concerned we already have answered even the second of the two

questions posed in the beginning of the present section.

(c) Having just extended the left-hand side of the t -interval as described in (b), the impending next question is: can a similar extension be derived for the right-hand side of the t -interval? Working towards an answer of this question, we closely examine the last two suprema in the proof of Theorem 4.1.0.

Let $f_n(t) := \log(nt) \cdot (nt(1-t))^{-\alpha}$ for $n \geq 1$ and $0 < t < 1$, where $\alpha := \frac{1}{2} - \nu > 0$. In order to find the extreme points of f_n for a fixed given n , we consider the derivative of f_n with respect to t :

$$\begin{aligned} \frac{d}{dt} f_n(t) &= t^{-1} (nt(1-t))^{-\alpha} - \alpha n^{-\alpha} (t(1-t))^{-(1+\alpha)} \log(nt) \cdot (1-2t) \\ &= n^{-\alpha} (t(1-t))^{-\alpha-1} \{ (1-t) + \alpha(2t-1) \log(nt) \}, \end{aligned}$$

whence $\frac{d}{dt} f_n(t) > 0$ for all $t \in [1/2, 1)$ provided that $n \geq 2$. Thus, f_n as a function of t only is monotonically increasing over $[1/2, 1)$.

Consequently, we have:

$$(*) \quad \sup_{\frac{1}{2} \leq t \leq 1 - \epsilon'_n} f_n(t) = \log(n(1 - \epsilon'_n)) (n(1 - \epsilon'_n) \epsilon'_n)^{-\alpha},$$

where $(\epsilon'_n)_{n \geq 2}$ denotes a still unspecified sequence of positive numbers.

We will now show that the numbers ϵ'_n that satisfy (*) remains finite, are exactly those numbers that keep the second supremum of the proof of Theorem 4.1.0. finite.

Turning now to the other supremum involved, we obtain for

$$g(t) := t^{\nu + \frac{1}{2}} (1-t)^{\nu - \frac{1}{2}} :$$

$$g'(t) = (t(1-t))^{\nu - \frac{1}{2}} \{ (\nu + \frac{1}{2}) - (\nu - \frac{1}{2}) t(1-t)^{-1} \}$$

and so $g'(t) = 0$ if and only if $\nu + \frac{1}{2} = (\nu - \frac{1}{2}) t(1-t)^{-1}$, which cannot happen for any $t \in (0, 1)$ because the left-hand side of the last equality

is always positive while the right-hand side is always negative.

We hence conclude that the supremum (maximum) will occur at one end-

point. Since $g(\epsilon_n) = (\epsilon_n)^{v+\frac{1}{2}}(1-\epsilon_n)^{v-\frac{1}{2}} \leq (\frac{1}{2})^{v+\frac{1}{2}}(\frac{1}{2})^{v-\frac{1}{2}} = (\frac{1}{2})^{2v}$ for

all $\epsilon_n \leq 1/2$ and since for all $\epsilon'_n \leq 1/2$

$$g(1-\epsilon'_n) = (1-\epsilon'_n)^{v+\frac{1}{2}}(\epsilon'_n)^{v-\frac{1}{2}} \geq (\frac{1}{2})^{v+\frac{1}{2}}(\epsilon'_n)^{v-\frac{1}{2}} \rightarrow \infty \text{ as } \epsilon'_n \rightarrow 0,$$

we now know that $\sup_{\epsilon_n \leq t \leq 1-\epsilon_n} t^{v+\frac{1}{2}}(1-t)^{v-\frac{1}{2}} = (1-\epsilon'_n)^{v+\frac{1}{2}}(\epsilon'_n)^{v-\frac{1}{2}}$.

And so our overall conclusion is that $1-\epsilon$ in (4.1.0.) can be replaced

by $1-\epsilon'_n$ when changing simultaneously $o(1)$ to $O(1)$ provided we

choose ϵ'_n so that

$$(\log n)(n\epsilon'_n)^{v-\frac{1}{2}} \leq c \text{ for all } n \geq 1,$$

where $c > 0$ is a constant.

But that is to say that $\epsilon'_n := \frac{(\log n)^{\frac{1}{\frac{1}{2}-v}}}{n}$ will satisfy the requirements.

(d) The only open problem remaining after the discussion in (c) is the

question if ϵ'_n similarly to ϵ_n can be chosen even smaller, namely

$$\epsilon'_n = \epsilon_n = \frac{\delta}{n} \text{ for a constant } \delta > 0.$$

We summarize the results of the above remarks in a corollary.

Corollary 4.1.2.

Under the conditions of Theorem 4.1.0. we have:

$$(a) \quad (4.1.1.) \quad \sup_{\epsilon_1 \leq t \leq 1-\epsilon_2} n^v \frac{|n^{-\frac{1}{2}}(\tilde{N}(nt) - t\tilde{N}(n)) / (\sigma n^{-3/2}) - \tilde{B}_n(t)|}{(t(1-t))^{\frac{1}{2}-v}} \stackrel{a.s.}{=} o(1) \text{ as } n \rightarrow \infty$$

for $0 < \epsilon_1 < 1-\epsilon_2 \leq 1$; $\epsilon_1, \epsilon_2 > 0$ fixed and $0 \leq v < \frac{1}{2}$.

(b)

$$(4.1.2.) \sup_{\delta/n \leq t \leq 1-\epsilon} n^v \frac{|n^{-1/2}(\tilde{N}(nt) - t\tilde{N}(n))/(\sigma_m^{-3/2}) - \tilde{B}_n(t)|}{(t(1-t))^{1/2-v}} \stackrel{a.s.}{=} O(1) \text{ as } n \rightarrow \infty$$

for $\epsilon, \delta > 0$ fixed and $0 \leq v < 1/2$.

(c)

$$(4.1.2.*) \sup_{\delta/n \leq t \leq 1-\epsilon'_n} n^v \frac{|n^{-1/2}(\tilde{N}(nt) - t\tilde{N}(n))/(\sigma_m^{-3/2}) - \tilde{B}_n(t)|}{(t(1-t))^{1/2-v}} \stackrel{a.s.}{=} O(1) \text{ as } n \rightarrow \infty$$

for $\delta > 0$, $0 \leq v < 1/2$ fixed and $\epsilon'_n = \frac{(\log n)^\alpha}{n}$, where $\alpha = (1/2 - v)^{-1} > 0$.

In their 1987 publication M. Csörgő and L. Horváth prove a number of theorems concerning the asymptotic distribution of the new pontograms in weighted sup-norm as well as weighted L_2 -norm. All of these results are based on the following

Theorem 4.1.3. (Confer Theorem 1.1. in M. Csörgő and L. Horváth (1987))

It is possible to extend the given probability space of the Poisson process N so that with a sequence of Brownian bridges $B_n = \{B_n(t) : 0 \leq t \leq 1\}$ living on this new probability space we have

$$(4.1.3.) \sup_{\delta/n \leq t \leq 1-\delta/n} n^v \frac{|K_n(t) - \mu_n^{1/2} B_n(t)|}{(t(1-t))^{1/2-v}} = O_p(1) \text{ as } n \rightarrow \infty$$

for every $\delta > 0$ and $0 \leq v < 1/2$.

We will present a more detailed version of the proof than given by M. Csörgő and L. Horváth (1987) in order to familiarize ourselves with its methodology. This is desirable because the proofs of the analogous theorems, which we are going to formulate for the process M and for the generalized renewal process (confer Theorem 4.1.7. and

Theorem 4.1.9.), will mostly use similar techniques. However, some adjustments will be necessary as for instance the proof of Theorem 4.1.3. is heavily based on the fact that the Poisson process N is a stationary independent increment process, a fact which in general is no longer true for the renewal process $\tilde{N}$. We again note that the extension of (4.1.3.) to a general renewal process is right now only available for the lower half $[\delta/n, 1/2]$.

Proof of Theorem 4.1.3.

Using the results of A. De Acosta (1982) we can enlarge the given probability space so that all processes and random variables introduced in the sequel can be accommodated on the new space.

Let now $\{W_n^{(1)}\}_{n \geq 1}$ and $\{W_n^{(2)}\}_{n \geq 1}$ be the two sequences of standard Wiener processes specified in Corollary 4.0.6. Then we claim that

$$(4.1.4.) \quad W_n(x) = \begin{cases} W_n^{(1)}(x) & 0 \leq x \leq n/2 \\ W_n^{(1)}(n/2) + W_n^{(2)}(n/2) - W_n^{(2)}(n-x) & n/2 \leq x \leq n \end{cases}$$

defines a Wiener process $W_n = \{W_n(x) : x \geq 0\}$ for each $n \geq 1$.

Proof of Statement (4.1.4.)

One only needs to check that the covariance structure of W_n coincides with the covariance structure of a standard Wiener process.

Let $0 \leq x \leq n/2 \leq y \leq n$, then

$$\begin{aligned} \text{Cov}(W_n(x), W_n(y)) &= E[W_n(x) - EW_n(x)][W_n(y) - EW_n(y)] \\ &= E[W_n^{(1)}(x) - EW_n^{(1)}(x)][W_n^{(1)}(n/2) + W_n^{(2)}(n/2) - W_n^{(2)}(n-y) \\ &\quad - E(W_n^{(1)}(n/2) + W_n^{(2)}(n/2) - W_n^{(2)}(n-y))] \end{aligned}$$

and because $W_n^{(1)}$ and $W_n^{(2)}$ are known to be Wiener processes,

$$= E[W_n^{(1)}(x)][W_n^{(1)}(n/2) + W_n^{(2)}(n/2) - W_n^{(2)}(n-y)]$$

by the linearity of the expected value

$$= E[W_n^{(1)}(x)W_n^{(1)}(n/2)] + E[W_n^{(1)}(x)W_n^{(2)}(n/2)] - E[W_n^{(1)}(x)W_n^{(2)}(n-y)]$$

and again $W_n^{(1)}$ is known to be a Wiener process

$$= x \wedge \frac{n}{2} + E[W_n^{(1)}(x)W_n^{(2)}(n/2)] - E[W_n^{(1)}(x)W_n^{(2)}(n-y)]$$

where $\wedge$ denotes the minimum and by the independence of $W_n^{(1)}$ and $W_n^{(2)}$

$$= x + E[W_n^{(1)}(x) \cdot E[W_n^{(2)}(n/2)] - E[W_n^{(1)}(x) \cdot E[W_n^{(2)}(n-y)]]$$

$$= x \wedge y.$$

By symmetry we also have, $\text{Cov}(W_n(x), W_n(y)) = y = x \wedge y$ for $0 \leq y \leq n/2 \leq x \leq n$. Hence, in the case that exactly one of x and y is less than or equal to $n/2$ while the other is greater than or equal to $n/2$, the covariance function of W_n agrees with that of a standard Wiener process. In the case that either $n/2 \leq x \leq y \leq n$ or $n/2 \leq y \leq x \leq n$, similar arguments show the validity of (4.1.4.). The case $0 \leq x, y \leq n/2$ is trivial by definition of W_n . ■

We wish to emphasize at this point that W_n could only be proven to be a Wiener process because the defining processes $W_n^{(1)}$ and $W_n^{(2)}$ were independent. Next, put

$$B_n(t) = n^{-1/2} \{W_n(nt) - tW_n(n)\} \quad \text{for } 0 \leq t \leq 1 \text{ and } n \geq 1.$$

Then $B_n = \{B_n(t) : 0 \leq t \leq 1\}$ is a Brownian bridge for every $n \geq 1$ and we get

$$\sup_{\delta/n \leq t \leq 1-\delta/n} \frac{|X_n(t) - \mu_n(t)|}{(t(1-t))^{\nu-1/2}}$$

$$\sup_{\delta/n \leq t \leq 1/2} n^{\nu-1/2} (t(1-t))^{\nu-1/2} |N(nt) - tN(n) - \mu^{1/2} \{W_n(nt) - tW_n(n)\}|$$

$$+ \sup_{1/2 \leq t \leq 1-\delta/n} n^{\nu-1/2} (t(1-t))^{\nu-1/2} |N(nt) - tN(n) - \mu^{1/2} \{W_n(nt) - tW_n(n)\}|$$

and setting $x = nt$

$$= \sup_{\delta \leq x \leq n/2} n^{\nu-1/2} \left(\frac{x}{n}(1-\frac{x}{n})\right)^{\nu-1/2} |N(x) - \frac{x}{n}N(n) - \mu^{1/2} \{W_n^{(1)}(x) - \frac{x}{n}W_n^{(1)}(n/2)$$

$$+ W_n^{(2)}(n/2) - W_n^{(2)}(0)\}|$$

$$+ \sup_{n/2 \leq x \leq n-\delta} n^{\nu-1/2} \left(\frac{x}{n}(1-\frac{x}{n})\right)^{\nu-1/2} |N(x) - \frac{x}{n}N(n) - \mu^{1/2} \{W_n^{(1)}(n/2) + W_n^{(2)}(n/2)$$

$$- W_n^{(2)}(n-x) - \frac{x}{n}W_n^{(1)}(n/2) + W_n^{(2)}(n/2)\}|$$

$$\leq \sup_{\delta \leq x \leq n/2} x^{\nu-1/2} \left(1-\frac{x}{n}\right)^{\nu-1/2} |N(x) - \mu x - \mu^{1/2} W_n^{(1)}(x)|$$

$$+ \sup_{\delta \leq x \leq n/2} n^{\nu-1/2} \left(\frac{x}{n}(1-\frac{x}{n})\right)^{\nu-1/2} \frac{x}{n} |N(n) - \mu n - \mu^{1/2} \{W_n^{(1)}(n/2) + W_n^{(2)}(n/2)\}|$$

$$+ \sup_{n/2 \leq x \leq n-\delta} (n-x)^{\nu-1/2} \left(\frac{x}{n}\right)^{\nu-1/2} |N(n) - N(x) - (n-x)\mu - \mu^{1/2} W_n^{(2)}(n-x)|$$

$$+ \sup_{n/2 \leq x \leq n-\delta} n^{\nu-1/2} \left(\frac{x}{n}(1-\frac{x}{n})\right)^{\nu-1/2} \left(1-\frac{x}{n}\right) |N(n) - \mu n - \mu^{1/2} \{W_n^{(1)}(n/2) + W_n^{(2)}(n/2)\}|$$

$$\leq \sqrt{2} \sup_{\delta \leq x \leq n/2} x^{\nu-1/2} |N(x) - \mu x - \mu^{1/2} W_n^{(1)}(x)|$$

$$+ n^{\nu-1/2} |N(n) - N(n/2) - \mu n/2 - \mu^{1/2} W_n^{(2)}(n/2)| \cdot \sup_{\delta \leq x \leq n/2} \left(\frac{x}{n}\right)^{\nu+1/2} \left(1-\frac{x}{n}\right)^{\nu-1/2}$$

$$+ n^{\nu-1/2} |N(n/2) - \mu n/2 - \mu^{1/2} W_n^{(1)}(n/2)| \cdot \sup_{\delta \leq x \leq n/2} \left(\frac{x}{n}\right)^{\nu+1/2} \left(1-\frac{x}{n}\right)^{\nu-1/2}$$

$$\begin{aligned}
& + \sqrt{2} \cdot \sup_{n/2 \leq x \leq n-\delta} (n-x)^{v-1/2} |N(n) - N(x) - (n-x)\mu - \mu^{1/2} W_n^{(2)}(n-x)| \\
& + n^{v-1/2} |N(n) - N(n/2) - \mu n/2 - \mu^{1/2} W_n^{(2)}(n/2)| \cdot \sup_{n/2 \leq x \leq n-\delta} \left(\frac{x}{n}\right)^{v-1/2} \left(1 - \frac{x}{n}\right)^{v+1/2} \\
& + n^{v-1/2} |N(n/2) - \mu n/2 - \mu^{1/2} W_n^{(1)}(n/2)| \cdot \sup_{n/2 \leq x \leq n-\delta} \left(\frac{x}{n}\right)^{v-1/2} \left(1 - \frac{x}{n}\right)^{v+1/2} \\
& = O_p(1) + 2(1/2)^{2v} n^{v-1/2} |N(n/2) - \mu n/2 - \mu^{1/2} W_n^{(1)}(n/2)| \quad \text{by (4.0.13.), (4.0.14.)} \\
& + 2(1/2)^{2v} n^{v-1/2} |N(n) - N(n/2) - \mu n/2 - \mu^{1/2} W_n^{(2)}(n/2)| \\
& = O_p(1) \quad \text{by (4.0.13.), (4.0.14.).}
\end{aligned}$$

This verifies the statement of Theorem 4.1.3. ■

It should be noted that Theorem 4.1.3. is a formal analog to Theorem 2.1.9. when taking $\delta = 1$ and $v = 0$ and replacing $\frac{n^{-3/2}}{\sigma} \hat{U}_{[(n+1)t]}$ by the pontogram $K_n(t)$. This of course means that we interpret the pontogram as another test statistic for still unspecified hypotheses H_0 and H_A and that Theorem 4.1.3. is valid under the null hypothesis H_0 .

Our next aim is to find a formal analog of Corollary 2.1.10. for the pontogram.

Corollary 4.1.4.

Let $q \in Q^*$ and let $\{B_n\}_{n \geq 1}$ be that sequence of Brownian bridges for which (4.1.3.) holds. Then, as $n \rightarrow \infty$,

e

$$(4.1.5.) \quad \left\{ \begin{array}{l} \sup_{0 < t < 1} \frac{|\mu^{-1/2} K_n(t) - B_n(t)|}{q(t)} = o_p(1) \\ \text{if and only if } I(q, c) < \infty \text{ for each } c > 0; \end{array} \right.$$

$$(4.1.6.) \quad \left\{ \begin{array}{l} \sup_{0 < t < 1} \frac{|\mu^{-1/2} K_n(t)|}{q(t)} \xrightarrow{D} \sup_{0 < t < 1} \frac{|B(t)|}{q(t)} \\ \text{if and only if } I(q, c) < \infty \text{ for some } c > 0. \end{array} \right.$$

Here B stands for a Brownian bridge.

Corollary 4.1.4. was presented in M. Csörgő and L. Horváth (1988b) as Corollary 2.8. without proof. A proof of Corollary 4.1.4. may be extracted out of the proof of the more general Theorem 2.1. in M. Csörgő and L. Horváth (1987) as follows.

Proof of Corollary 4.1.4.

We first prove the sufficiency part of statement (4.1.5.). It is a consequence of Theorem 4.1.3. (take $\delta = 1$ and $\nu = 0$) that for every fixed $\varepsilon > 0$ we obtain

$$(4.1.7.) \quad \sup_{\varepsilon \leq t \leq 1-\varepsilon} \frac{|\mu^{-1/2} K_n(t) - B_n(t)|}{q(t)} = o_p(1) \text{ as } n \rightarrow \infty \text{ and}$$

$$(4.1.8.) \quad \sup_{1/n \leq t \leq \varepsilon} \frac{|\mu^{-1/2} K_n(t) - B_n(t)|}{q(t)} = O_p(1) \cdot \sup_{0 < t \leq \varepsilon} \frac{\sqrt{t(1-t)}}{q(t)} \text{ as } n \rightarrow \infty.$$

(Statement (4.1.7.) can be seen to be valid by closely examining the proof of Theorem 4.1.3.; recall that $\frac{\log n}{\sqrt{n}} \rightarrow 0$ ($n \rightarrow \infty$).) But by Theorem 3.3. (characterization of Erdős-Feller-Kolmogorov-Petrovski upper class functions) in M. Csörgő et al. (1986) we know that

$I(q, c) < \infty$ for some $c > 0$ implies

$$(4.1.9.) \quad \lim_{t \rightarrow 0} \frac{q(t)}{t} = \infty.$$

Therefore we have immediately

$$(4.1.10.) \quad \sup_{0 < t < \varepsilon} \frac{\sqrt{t(1-t)}}{q(t)} \rightarrow 0 \text{ as } \varepsilon \rightarrow 0 \text{ as well as}$$

$$(4.1.10.*) \quad \sup_{1-\varepsilon < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} \rightarrow 0 \text{ as } \varepsilon \rightarrow 0.$$

Also, again by Theorem 4.1.3. with $\delta = 1$ and $\nu = 0$,

$$(4.1.8.*) \quad \sup_{1-\varepsilon \leq t \leq 1 - \frac{1}{n}} \frac{|\mu_n^{-1/2} K_n(t) - B_n(t)|}{q(t)} = O(1) \cdot \sup_{1-\varepsilon < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} \text{ as } n \rightarrow \infty.$$

Finally, we consider those t for which $0 < t < 1/n$ and $1 - \frac{1}{n} < t < 1$, respectively. As noted previously (confer beginning of Subsection 4.0.1.), $S_1(\omega), S_2(\omega), \dots$ are the (only) jump points of the Poisson process N . In addition, $S_1(\omega)$ is the first such point and so $N(x) = 0$ for all $0 \leq x < S_1(\omega)$. Thus we arrive at

$$\begin{aligned} \sup_{0 < t < \frac{1}{n}} \mu_n^{-1/2} \frac{K_n(t)}{\sqrt{t(1-t)}} &= O(1) \cdot \sup_{0 < x < S_1} n^{-1/2} \left(\frac{x}{n}\right)^{-1/2} |N(x) - \frac{x}{n} N(n)| \\ &= O(1) \cdot \frac{N(n)}{n} \cdot \sup_{0 < x < S_1} \sqrt{x} \\ &= O(1) \cdot \frac{N(n)}{n} \cdot \sqrt{S_1} \end{aligned}$$

Since by the strong law of large numbers for independent identically distributed random variables with finite mean, $N(n)/n \xrightarrow{\text{a.s.}} \mu^{-1}$, and since S_1 , being an exponential random variable, is bounded in probability, we get

$$(4.1.11.) \quad \sup_{0 < t < \frac{1}{n}} S \mu^{-\frac{1}{2}} \frac{|K_n(t)|}{\sqrt{t(1-t)}} = O_p(1) \text{ as } n \rightarrow \infty.$$

Next, we consider

$$\begin{aligned} S \sup_{\frac{1}{n} \leq t \leq \frac{1}{n}} \mu^{-\frac{1}{2}} \frac{|K_n(t)|}{\sqrt{t(1-t)}} &= \sup_{S_1 \leq x \leq 1} n^{-\frac{1}{2}} \left(\frac{x}{n} \left(1 - \frac{x}{n} \right) \right)^{-\frac{1}{2}} \left| N(x) - \frac{x}{n} N(n) \right| \\ &= \sup_{S_1 \leq x \leq 1} x^{-\frac{1}{2}} \left(1 - \frac{x}{n} \right)^{-\frac{1}{2}} x \left| \frac{N(x)}{x} - \frac{N(n)}{n} \right| \\ &\leq \sup_{S_1 \leq x \leq 1} x^{\frac{1}{2}} \left(1 - \frac{x}{n} \right)^{-\frac{1}{2}} \left\{ \left| \frac{N(x)}{x} \right| + \left| \frac{N(n)}{n} \right| \right\} \\ &\leq \left(1 - \frac{1}{n} \right)^{-\frac{1}{2}} \left\{ \frac{N(1)}{S_1} + \frac{N(n)}{n} \right\}, \end{aligned}$$

where the last inequality holds because N is monotonically nondecreasing in x . Since in addition $N(n)/n \xrightarrow{\text{a.s.}} \mu^{-1}$ as $n \rightarrow \infty$, it suffices to show:

$$N(1)/S_1 = O_p(1).$$

Towards this end let $C > 0$ be an arbitrary constant. Then we have

$$\begin{aligned} P\{N(1)/S_1 > C\} &= P\{N(1) > CX_1\} && (S_1 = X_1, \text{ by definition}) \\ &= \int_0^\infty P\{N(1) > Cx\} \mu e^{-\mu x} dx && (\text{by (4.0.0.)}) \\ &= \int_0^\infty P\left\{e^{t_0(N(1) - Cx)} > 1\right\} \mu e^{-\mu x} dx && (t_0 > 0 \text{ fixed}) \end{aligned}$$

$$\begin{aligned}
&\leq \mu \int_0^\infty \int_{\Omega} e^{t_0(N(1)-Cx)} e^{-\mu x} dP dx \\
&= \mu \int_0^\infty e^{-(t_0 C + \mu)x} \left(\int_{\Omega} e^{t_0 N(1)} dP \right) dx
\end{aligned}$$

and since the moment generating function $M = M_{N(1)}$ of a Poisson random variable is finite over the whole real line,

$$= \mu M(t_0) \int_0^\infty e^{-(t_0 C + \mu)x} dx = \mu M(t_0) (t_0 C + \mu)^{-1},$$

whence $\lim_{C \rightarrow \infty} P\{N(1)/S_1 > C\} = 0$.

Thus, we have just verified

$$(4.1.12.) \quad \sup_{\frac{1}{n} \leq t \leq \frac{1}{n}} \frac{\mu^{-1/2} |K_n(t)|}{\sqrt{t(1-t)}} = O_p(1) \text{ as } n \rightarrow \infty.$$

An application of Theorem 3.4. (characterization of Chibisov-O'Reilly functions) in M. Csörgő et al. (1986), gives

$$(4.1.13.) \quad \sup_{0 \leq t \leq 1/n} \frac{|B_n(t)|}{q(t)} = \sup_{0 \leq t \leq 1/n} \frac{|B(t)|}{q(t)} = o_p(1) \text{ as } n \rightarrow \infty$$

provided that $I(q, c) < \infty$ for all $c > 0$, where B denotes a standard Brownian bridge.

This concludes our considerations of the 'lower tail' of the problem.

Turning now to the upper tail end, i.e. to those t close to 1, we have similarly to (4.1.13.):

$$(4.1.13.*) \quad \sup_{1 - \frac{1}{n} \leq t < 1} \frac{|B_n(t)|}{q(t)} = o_p(1) \text{ as } n \rightarrow \infty.$$

Moreover, we have

$$(4.1.11.*) \quad \sup_{S_{N(n)}/n < t < 1} \mu^{-1/2} \frac{|K_n(t)|}{\sqrt{t(1-t)}} = O_p(1) \quad \text{as } n \rightarrow \infty$$

because

$$\begin{aligned} \sup_{S_{N(n)}/n < t < 1} \mu^{-1/2} \frac{|K_n(t)|}{\sqrt{t(1-t)}} &= O(1) \cdot \sup_{S_{N(n)} < x < n} n^{-1/2} \left(1 - \frac{x}{n}\right)^{-1/2} \left|N(x) - \frac{x}{n} N(n)\right| \\ &= O(1) \cdot \sup_{S_{N(n)} < x < n} n^{-1/2} \left(1 - \frac{x}{n}\right)^{-1/2} \left|N(n) - \frac{x}{n} N(n)\right| \\ &= O(1) \cdot \frac{N(n)}{\sqrt{n}} \sup_{S_{N(n)} < x < n} \left(1 - \frac{x}{n}\right)^{1/2} \\ &= O(1) \cdot \frac{N(n)}{\sqrt{n}} \left\{1 - S_{N(n)}/n\right\}^{1/2} \\ &= O(1) \cdot \frac{N(n)}{n} \left\{n - S_{N(n)}\right\}^{1/2} \end{aligned}$$

$\stackrel{a.s.}{=} O(1) \left\{n - S_{N(n)}\right\}^{1/2}$ by the strong law of large

numbers and $\lim_{n \rightarrow \infty} P\{n - S_{N(n)} \leq x\} = 1 - e^{-\mu x}$, $x > 0$;

the latter follows from W. Feller (1966), formula (3.14.), p.356, which is not only valid for the Poisson process but also for any ordinary renewal process.

Finally, we claim that

$$(4.1.12.*) \quad \sup_{1 - \frac{1}{n} \leq t \leq S_{N(n)}/n} \frac{\mu^{-1/2} |K_n(t)|}{\sqrt{t(1-t)}} = O_p(1) \quad \text{as } n \rightarrow \infty.$$

$$\sup_{1-\frac{1}{n} \leq t \leq S_{N(n)}/n} \frac{|K_n(t)|}{\sqrt{t(1-t)}} = \sup_{n-1 \leq x \leq S_{N(n)}} n^{-\frac{1}{2}} \left(\frac{x}{n} \left(1 - \frac{x}{n}\right)\right)^{-\frac{1}{2}} |N(x) - \frac{x}{n} N(n)|$$

$$\leq \left(\frac{n-1}{n}\right)^{-\frac{1}{2}} (n - S_{N(n)})^{-\frac{1}{2}} \left\{ \sup_{n-1 \leq x \leq S_{N(n)}} |N(x) - N(n)| + \sup_{n-1 \leq x \leq S_{N(n)}} \left(1 - \frac{x}{n}\right) |N(n)| \right\}$$

and by (4.0.4a) (vii)

$$\leq \left(\frac{n-1}{n}\right)^{-\frac{1}{2}} (n - S_{N(n)})^{-\frac{1}{2}} \left\{ [N(n) - N(n-1)] + \left(1 - \frac{n-1}{n}\right) N(n) \right\}$$

$$\leq \left(\frac{n-1}{n}\right)^{-\frac{1}{2}} (n - S_{N(n)})^{-\frac{1}{2}} \left\{ N(1) + \frac{N(n)}{n} \right\}.$$

Now applying the aforementioned formula (3.14.) of W. Feller (1966) again, we immediately arrive at (4.1.12.*).

Putting all of the above results together, we obtain

$$\sup_{0 < t < 1} \frac{|\mu^{-\frac{1}{2}} K_n(t) - B_n(t)|}{q(t)}$$

$$\leq \left\{ \sup_{0 < t < S_1/n} \frac{\mu^{-\frac{1}{2}} |K_n(t)|}{\sqrt{t(1-t)}} \cdot \sup_{0 < t < S_1/n} \frac{\sqrt{t(1-t)}}{q(t)} + \sup_{0 < t < S_1/n} \frac{|B_n(t)|}{q(t)} \right\}$$

$$\vee \left\{ \sup_{S_1/n \leq t \leq 1/n} \frac{\mu^{-\frac{1}{2}} |K_n(t)|}{\sqrt{t(1-t)}} \cdot \sup_{S_1/n \leq t \leq 1/n} \frac{\sqrt{t(1-t)}}{q(t)} + \sup_{S_1/n \leq t \leq 1/n} \frac{|B_n(t)|}{q(t)} \right\}$$

$$\vee \sup_{1/n \leq t \leq \varepsilon} \frac{|\mu^{-\frac{1}{2}} K_n(t) - B_n(t)|}{\sqrt{t(1-t)}} \cdot \sup_{1/n \leq t \leq \varepsilon} \frac{\sqrt{t(1-t)}}{q(t)}$$

$$\vee \sup_{\varepsilon \leq t \leq 1-\varepsilon} \frac{|\mu^{-\frac{1}{2}} K_n(t) - B_n(t)|}{q(t)}$$

$$\vee \sup_{1-\varepsilon \leq t \leq 1-\frac{1}{n}} \frac{|\mu^{-\frac{1}{2}} K_n(t) - B_n(t)|}{\sqrt{t(1-t)}} \cdot \sup_{1-\varepsilon \leq t \leq 1-\frac{1}{n}} \frac{\sqrt{t(1-t)}}{q(t)}$$

$$v \left\{ 1 - \frac{1}{n} \sup_{\frac{1}{n} \leq t \leq \frac{S_{N(n)}}{n}} \frac{\mu^{-1/2} |K_n(t)|}{\sqrt{t(1-t)}} \cdot 1 - \frac{1}{n} \sup_{\frac{1}{n} \leq t \leq \frac{S_{N(n)}}{n}} \frac{\sqrt{t(1-t)}}{q(t)} + 1 - \frac{1}{n} \sup_{\frac{1}{n} \leq t \leq \frac{S_{N(n)}}{n}} \frac{|B_n(t)|}{q(t)} \right\}$$

$$v \left[\sup_{\frac{S_{N(n)}}{n} < t < 1} \mu^{-1/2} \frac{|K_n(t)|}{\sqrt{t(1-t)}} \cdot \sup_{\frac{S_{N(n)}}{n} < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} + \sup_{\frac{S_{N(n)}}{n} < t < 1} \frac{|B_n(t)|}{q(t)} \right]$$

$$P \left\{ O_P(1) \cdot \sup_{0 < t < \epsilon} \frac{\sqrt{t(1-t)}}{q(t)} + \sup_{0 < t < \frac{1}{n}} \frac{|B_n(t)|}{q(t)} \right\} \quad \text{by (4.1.11.)}$$

$$v \left\{ O_P(1) \cdot \sup_{0 < t < \epsilon} \frac{\sqrt{t(1-t)}}{q(t)} + \sup_{0 < t < \frac{1}{n}} \frac{|B_n(t)|}{q(t)} \right\} \quad \text{by (4.1.12.)}$$

$$v O_P(1) \cdot \sup_{0 < t < \epsilon} \frac{\sqrt{t(1-t)}}{q(t)} \quad \text{by (4.1.8.)}$$

$$v o_P(1) \quad \text{by (4.1.7.)}$$

$$v O_P(1) \cdot \sup_{1-\epsilon < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} \quad \text{by (4.1.8.*)}$$

$$v \left\{ O_P(1) \cdot \sup_{1-\epsilon < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} + 1 - \frac{1}{n} \sup_{\frac{1}{n} \leq t < 1} \frac{|B(t)|}{q(t)} \right\} \quad \text{by (4.1.12.*)}$$

$$v \left\{ O_P(1) \cdot \sup_{1-\epsilon < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} + 1 - \frac{1}{n} \sup_{\frac{1}{n} \leq t < 1} \frac{|B(t)|}{q(t)} \right\} \quad \text{by (4.1.11.*)}$$

$$= \left\{ O_P(1) \cdot \sup_{0 < t < \epsilon} \frac{\sqrt{t(1-t)}}{q(t)} + o_P(1) \right\} \cdot v O_P(1) \cdot \sup_{0 < t < \epsilon} \frac{\sqrt{t(1-t)}}{q(t)} \cdot v o_P(1)$$

$$v O_P(1) \cdot \sup_{1-\epsilon < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} \cdot v \left\{ O_P(1) \cdot \sup_{1-\epsilon < t < 1} \frac{\sqrt{t(1-t)}}{q(t)} + o_P(1) \right\} \quad \text{by (4.1.13.), (4.1.13.*)}$$

Now, letting $\varepsilon \rightarrow 0$, (4.1.10.) and (4.1.10.*) imply the sufficiency part of statement (4.1.5.). Turning to the necessity part of (4.1.5.), we have to assume now

$$(*) \quad \sup_{0 < t < 1} \frac{|\mu^{-1/2} K_n(t) - B_n(t)|}{q(t)} = o_p(1) \quad \text{as } n \rightarrow \infty.$$

Since

$$\sup_{0 < t \leq \frac{1}{n}} \frac{\mu^{-1/2} |K_n(t)|}{\sqrt{t(1-t)}} = O_p(1) \quad \text{as } n \rightarrow \infty \quad \text{by (4.1.11.) and (4.0.12.)}$$

implying that

$$\sup_{0 < t \leq \frac{1}{n}} \mu^{-1/2} |K_n(t)| = o_p(1) \quad \text{as } n \rightarrow \infty,$$

and since

$$\sup_{0 < t < 1} |B(t)| = O_p(1) \quad \text{as } n \rightarrow \infty,$$

implying that

$$\sup_{0 < t \leq \frac{1}{n}} |B(t)| = o_p(1) \quad \text{as } n \rightarrow \infty,$$

we must also have (in the light of (*))

$$(4.1.14.) \quad \sup_{0 < t \leq \frac{1}{n}} \frac{|B(t)|}{q(t)} = o_p(1) \quad \text{as } n \rightarrow \infty.$$

Similarly, we obtain

$$(4.1.14.*) \quad \sup_{1 - \frac{1}{n} \leq t < 1} \frac{|B(t)|}{q(t)} = o_p(1) \quad \text{as } n \rightarrow \infty.$$

But according to Theorem 3.4. in M. Csörgö et al. (1986) (4.1.14.)

and (4.1.14.*) together are equivalent to the statement:

$I(q, c) < \infty$ for all constants $c > 0$,

which completes the proof of the necessity part of (4.1.5.).

Proof of Statement (4.1.6.).

Again we verify first the direct half. As established previously (confer the beginning of the proof of (4.1.5.)), statement (4.1.9.) together with Theorem 4.1.3. ($\delta = 1, v = 0$) implies

$$(4.1.15.) \quad \sup_{\frac{1}{n} \leq t \leq 1 - \frac{1}{n}} \frac{|\mu^{-\frac{1}{2}} K_n(t) - B_n(t)|}{q(t)} = o_p(1) \quad \text{as } n \rightarrow \infty.$$

Hence it is sufficient to show that

$$\sup_{\frac{1}{n} \leq t \leq 1 - \frac{1}{n}} \frac{|B(t)|}{q(t)} \xrightarrow{D} \sup_{0 < t < 1} \frac{|B(t)|}{q(t)} \quad \text{as } n \rightarrow \infty,$$

which is a direct consequence of Theorem 3.3. in M. Csörgö et al. (1986)

under the assumption $I(q, c) < \infty$ for some $c > 0$.

As to the indirect half of (4.1.6.) we have

$$\sup_{0 < t < 1} \frac{\mu^{-\frac{1}{2}} |K_n(t)|}{q(t)} \xrightarrow{D} \sup_{0 < t < 1} \frac{|B(t)|}{q(t)} \quad \text{as } n \rightarrow \infty.$$

This statement implicitly contains the requirement that the limiting random variable be finite almost surely, whence another application of Theorem 3.3. above yields the indirect half of (4.1.6.).

Note that the proof of Corollary 4.1.4. follows the same general methodology as the proof of Corollary 2.1.10.

4.1.1. The Process $M = \{M(x) : x \geq 0\}$

The present subsection is devoted to the discussion of the asymptotic behaviour of the stochastic process M whose significance was discussed in the previous subsection.

Definition 4.1.5.

Let $W = \{W(x) : 0 \leq x < \infty\}$ be a given standard Wiener process and let $\lambda > 0$ be a fixed positive constant. Then the stochastic process $M = \{M(x) : 0 \leq x < \infty\}$ is defined pointwise via

$$M(x) := \inf \{y \geq 0 : \lambda W(y) = x - y\}, \quad 0 \leq x < \infty.$$

The next lemma lists a number of elementary properties of the just defined process M .

Lemma 4.1.6.

Let M be the stochastic process of Definition 4.1.5. Then we have

(a) M is a stationary independent increment process, i.e. for each $0 \leq x_1 < x_2 < \infty$ fixed $M(x_2) - M(x_1) \stackrel{D}{=} M(x_2 - x_1)$ and $\{M(x_2) - M(x_1)\}$ is independent of any event in the σ -algebra generated by $\{M(s) : 0 \leq s \leq x_1\}$.

(b)
$$\sup_{k < x \leq k+1} |M(x) - M(k)| \stackrel{D}{=} M(1) \quad \text{for each } k \geq 1,$$

in particular, M is a monotonically nondecreasing function of x (for ω fixed).

Proof of Lemma 4.1.6.

A proof for part (a) is given in M. Csörgő, L. Horváth and J. Steinebach (1987) inside the proof of Theorem 2.1. Applying now part (a), the remainder is immediate.

From the definition of $M(x)$ in terms of a given standard Wiener process we see that $M(x)$ is a kind of inverse of a stretched (shortened) and shifted Wiener process. However, we have just seen that $M(\cdot, \omega)$ as a function in x is still monotonically nondecreasing. This means that the "zick-zack" motion of the original Wiener process must be entirely compensated by the shift. This of course is only possible because the shift is not constant but rather time dependent.

The following theorem is a formal analogue of Theorem 4.1.3. for the process M playing the role of the Poisson process N .

Theorem 4.1.7.

Suppose W is a given standard Wiener process and put

$$M(x) = \inf \{y \geq 0: \lambda W(y) = x-y\} \quad \text{for } 0 \leq x < \infty,$$

where λ is a strictly positive constant.

Then we can construct a sequence $\{\bar{B}_n\}_{n \geq 1}$ of Brownian bridges such that

$$(4.1.3.*) \quad \sup_{\delta/n \leq t \leq 1-\delta/n} n^v \left| \frac{M(nt) - tM(n)}{\lambda \sqrt{n}} - \bar{B}_n(t) \right| / (t(1-t))^{1/2-v} = O_p(1)$$

holds as $n \rightarrow \infty$ for every $\delta > 0$ and $0 \leq v < 1/2$.

Remark 4.1.8.

It should be understood that Theorem 4.1.7. has to be interpreted in the spirit of Theorem 4.1.3. This is to say that Theorem 4.1.7. holds on a new, rich enough probability space on which all processes and random variables in question can be accommodated. The proof of Theorem 4.1.7. follows along the lines of the proof of Theorem 4.1.3. and is hence omitted here. It is nevertheless worthwhile to remark

at this juncture that the central fact on which the generalizations of the pontogram concept where thought of relying on is the "surprising" observation of J. Steinebach (1986) that the generalized renewal process $\tilde{N}$ can be nearer to a renewal process based on $W(x)$ (this process is of course the previously introduced stochastic process M) than $\tilde{N}$ is to $W(x)$ itself. Here naturally "surprising" is meant to be surprising in the light of the Bahadur-Kiefer phenomenon. This phenomenon (confer in particular R.R. Bahadur (1966), J. Kiefer (1970), M. Csörgö and P. Révész (1978)) tells us that, if we wish to approximate a partial sum process and its renewal process simultaneously by the same Wiener process, then the best possible rate of such a joint approximation is of order $(n \log \log n)^{1/2} \cdot (\log n)^{1/2}$. (Compare also M. Csörgö (1983), Chapter 3.)

4.1.2. A Result for Generalized Renewal Processes

In this subsection we introduce a more general pontogram, namely a pontogram where the generalized renewal process takes over the role of the Poisson process, and prove that the main result of Subsection 4.1.0. still holds for the more general pontogram, at least in the lower half of the t -interval. Towards this end we define, using the notations of Subsection 4.0.1.:

$$\tilde{K}_n(t) := n^{-1/2} \{ \tilde{N}(nt) - t\tilde{N}(n) \} \quad \text{for } 0 \leq t \leq 1 \text{ and } n \geq 1.$$

Then we have the following result for the generalized pontogram

$$\tilde{K}_n = \{ \tilde{K}_n(t) : 0 \leq t \leq 1 \}.$$

Theorem 4.1.9.

Assume that the moment generating function of Y_1 is finite in some nondegenerate neighbourhood of zero. Then there exists a new probability space with a sequence $\{\tilde{B}_n\}_{n \geq 1}$ of Brownian bridges of it such that as $n \rightarrow \infty$

$$\sup_{\delta/n \leq t \leq 1/2} n^v \frac{|\tilde{K}_n(t)/(\sigma n^{-3/2}) - \tilde{B}_n(t)|}{(t(1-t))^{1/2-v}} = O_p(1)$$

holds for every $\delta > 0$ and $0 \leq v < 1/2$.

Proof of Theorem 4.1.9.

Define $\tilde{B}_n(t) = n^{-1/2} \{ \tilde{W}_n^{(1)}(nt) - t\tilde{W}_n^{(1)}(n) \}$ for $0 \leq t \leq 1/2$ and $n \geq 1$, where $\tilde{W}_n^{(1)}$ is the Wiener process constructed in Corollary 4.0.10.

With the notation $t = \frac{x}{n}$, Theorem 4.1.9. is then immediate from Corollary 4.0.10. (Compare also Corollary 4.1.2.)

Chapter 5. Asymptotic Properties for the Sequential CUSUM Procedure

Section 5.0. Introduction and Notations

As mentioned earlier, in statistical quality control one observes the output of a production line and one would ideally wish to detect the presence of out-of-control conditions in a repetitive process as soon as possible after a defective product is produced. Clearly, early detection of a change in quality is a prime objective. Hence one would want a test procedure to be designed in such a way that an "off-target" signal is triggered as soon as possible after an actual change has occurred. Nevertheless, we do not wish to neglect the cost factor. In terms of a test procedure this means that we would like the probability of ever stopping the production line while it is still "on-target" to be less than any preassigned positive small number.

In this chapter only nonparametric, distribution free sequential test procedures will be considered. The hypotheses discussed here are of the form

H_0 : $X_1, X_2, \dots$ are identically distributed (i.e. no changepoint occurs)

versus

H_A : there exists an integer $N \geq 2$ such that

$X_1, X_2, \dots, X_N$ have distribution F ,

$X_{N+1}, X_{N+2}, \dots$ have distribution G ,

and $F \neq G$

(i.e. exactly one changepoint, namely at time N , occurs).

Here $(X_n)_{n \geq 1}$ stands for a sequence of independent random variables

representing the observations. In the above $F \neq G$ abbreviates the fact that there exists at least one $x_0 \in \mathbb{R}$ such that $F(x_0) \neq P\{X_1 \leq x_0\} \neq P\{X_{N+1} \leq x_0\} = G(x_0)$.

For the derivation of our test statistics and their (asymptotic) properties the following notations will prove useful.

Let $W = \{W(t): 0 \leq t < \infty\}$ denote a standard Wiener process (Brownian motion). Let $K = \{K(s, t): 0 \leq s \leq 1, 0 \leq t < \infty\}$ denote a Kiefer process, i.e. a two-time parameter Gaussian process with mean zero and covariance structure $EK(s, t)K(u, w) = (t \wedge w)(s \wedge u - su)$, where $\wedge$ stands for minimum. Throughout this chapter we will, without loss of generality, assume that our probability space is rich enough to accommodate all processes and random variables under discussion.

For the given random variables X_j let

$$(5.0.0.) \quad F_{i-1}(t) := \frac{1}{i} \#\{1 \leq j \leq i-1: X_j \leq t\} \quad i \geq 2,$$

be their empirical distribution function.

Next, we define the sequential rank of X_i among $X_1, X_2, \dots, X_i$ via

$$(5.0.1.) \quad \xi_i := F_{i-1}(X_i), \quad i \geq 2.$$

For these random variables ξ_i P.K. Bhattacharya and P. Frierson

(1981) (confer also E.A. Parent (1965)) proved the following representation theorem:

Lemma 5.0.0. (Compare Lemma 2.1. (and the remark thereafter) in P.K. Bhattacharya and D. Frierson (1981)).

Under the null hypothesis H_0 the random variables ξ_i are independent with distribution $P\{\xi_i = j\} = \frac{1}{i}, \quad 0 \leq j \leq i-1$.

Consequently, under H_0 there exist independent identically distributed random variables $U_i, i \geq 1$, with uniform distribution on $(0,1)$, i.e. $U_i \stackrel{D}{=} U(0,1)$, such that

$$(5.0.2.) \quad \xi_i = \frac{[iU_i]}{i}, \text{ holds for every } i \geq 2.$$

Applying (5.0.2.), one can easily verify that

$$E\xi_i \sim \frac{1}{2} \text{ and } \text{Var } \xi_i \sim \frac{1}{12} \text{ as } i \rightarrow \infty,$$

where $\sim$ symbolizes asymptotic equivalence.

Let

$$(5.0.3.) \quad \tilde{F}_k(t) := \frac{1}{k} \# \{2 \leq i \leq k: \xi_i \leq t\}, \quad k \geq 2,$$

be the empirical distribution function of the sequential ranks ξ_i , $2 \leq i \leq k$, and put

$$(5.0.4.) \quad L_t := \sqrt{12} \sum_{i=2}^{[t]} (\xi_i - \frac{1}{2}) \quad \text{for } t \in [2, \infty).$$

Since under H_0 the mean of L_t is of order $\log t$ while EL_t clearly explodes for t much larger than N under the alternative hypothesis, the shape of the distribution of L_t will certainly change in the presence of a disorder (changepoint) and so it is reasonable to use appropriate functionals of L_t as statistics for testing H_0 versus H_A . Note that the above argument confirms the intuition that functionals of L_t should lead to good test statistics even for the detection of small changes.

In agreement with commonly used nomenclature in statistics, where any sum of residuals is called a CUSUM (cumulative sum) (confer, e.g., P.R. Krishnaiah and P.K. Sen (1984), p.703), we say that L_t is the CUSUM of the sequential ranks up to time t .

Finally, define a one-sided stopping rule by

$$(5.0.5.) \quad \tau_1(m) := \inf \{k : k \geq m, L_k > \sqrt{m} g(k/m)\}$$

and a two-sided stopping rule by

$$(5.0.5.*) \quad \tau_2(m) := \inf \{k : k \geq m, |L_k| > \sqrt{m} g(k/m)\},$$

where m is an initial value and g is a positive function (weight function) which will be specified in the sequel.

Section 5.1. Results Under the Null Hypothesis

In the present section we assume the null hypothesis of no change to be true. Thus the random variables $X_1, X_2, \dots$ representing the observed data are independent identically distributed with distribution function $F = G$. Two different types of test statistics and their asymptotic behaviour under the null hypothesis will be presented here. These test statistics are based on functionals of $k(\tilde{F}_k(t) - t)$, where $\tilde{F}_k$ denotes the empirical distribution function of the sequential ranks. Hence functionals of $k(\tilde{F}_k(t) - t)$ will act like the corresponding functionals of empirical processes blown up by $k^{1/2}$ (for the concept of empirical processes compare Section 1.0.). The following lemma establishes the connection between $k(\tilde{F}_k(t) - t)$ and the previously introduced CUSUM L_k .

Lemma 5.1.0.

Let $\tilde{F}_k$ and L_k be as defined in (5.0.3.) and (5.0.4.), respectively. Then we have for each $k \geq 2$:

$$-k \int_0^1 (\tilde{F}_k(t) - t) dt = \frac{1}{\sqrt{12}} L_k + \frac{1}{2}.$$

Proof of Lemma 5.1.0.

By the definition of an empirical distribution function we have:

$$\tilde{F}_k(t) := \frac{1}{k} \# \{2 \leq i \leq k: \xi_i \leq t\} = \frac{1}{k} \sum_{i=2}^k I_{[0,t]}(\xi_i),$$

where $I_{[0,t]}$ denotes the indicator function on $[0,t]$. Thus,

$$k \int_0^1 \tilde{F}_k dt = \int_0^1 \sum_{i=2}^k I_{[0,t]}(\xi_i) dt = \sum_{i=2}^k \int_0^1 I_{[0,t]}(\xi_i) dt$$

$$= \sum_{i=2}^k (1 - \xi_i) = - \sum_{i=2}^k (\xi_i - \frac{1}{2}) + \frac{k}{2} - \frac{1}{2},$$

and therefore we arrive at:

$$\begin{aligned} -k \int_0^1 (\tilde{F}_k(t) - t) dt &= -k \int_0^1 \tilde{F}_k(t) dt + k \int_0^1 t dt \\ &= \sum_{i=2}^k (\xi_i - \frac{1}{2}) - \frac{k}{2} + \frac{1}{2} + k \left. \frac{t^2}{2} \right|_0^1 \\ &= \frac{1}{\sqrt{12}} L_k + \frac{1}{2}. \end{aligned}$$

Recently, M. Csörgő and L. Horváth (1987a) accomplished a strong approximation of $k(\tilde{F}_k(t) - t)$ which involves the Kiefer process defined and discussed in Section 1.0.

Theorem 5.1.1.

We can define a Kiefer process $K = \{K(t, s) : 0 \leq t \leq 1, s \geq 0\}$ such that

$$P\left\{\max_{2 \leq k \leq n} \sup_{0 \leq t \leq 1} |k(\tilde{F}_k(t) - t) - K(t, k)| > A(\log n)^2\right\} \leq Bn^{-\epsilon}$$

for all $\epsilon > 0$ fixed, where $A = A(\epsilon)$ and B are absolute positive constants.

A proof of Theorem 5.1.1. is given in M. Csörgő and L. Horváth (1987a) (confer there the proof of Theorem 2.1.). This theorem is then used in the aforementioned publication to deduce the following

Corollary 5.1.2. (Compare Theorem 2.2. in M. Csörgő and L. Horváth (1987a))

For positive functions w and g we assume that

$w(t) \leq D(t(1-t))^{-\nu}$ with some constants $D > 0$, $0 < \nu < 1/2$,

and g is continuous on $[1, \infty)$ satisfying

$$\limsup_{x \rightarrow \infty} (x \log \log x)^{1/2} / g(x) < \infty.$$

Then we have as $m \rightarrow \infty$.

$$(5.1.0.) \quad \sup_{k \geq m} \sup_{\frac{1}{k} \leq t \leq 1 - \frac{1}{k}} k \frac{|\tilde{F}_k(t) - t| w(t)}{\sqrt{m} g(k/m)} \xrightarrow{D} \sup_{x \geq 1} \sup_{0 < t < 1} \frac{|K(t, x)| w(t)}{g(x)}.$$

Remarks

(a) Imitating the proof of Corollary 5.1.2. one can also show that under the conditions of that corollary

$$(5.1.0.*) \quad \sup_{k \geq m} \sup_{\frac{1}{k} \leq t \leq 1 - \frac{1}{k}} k \frac{(\tilde{F}_k(t) - t) w(t)}{\sqrt{m} g(k/m)} \xrightarrow{D} \sup_{x \geq 1} \sup_{0 < t < 1} \frac{K(t, x) w(t)}{g(x)}$$

holds as $m \rightarrow \infty$.

(b) Applying Remark (a) following Definition 1.1.12. in Chapter 1 to (5.1.0.*), one immediately obtains:

$$(t_0(1-t_0))^{+1/2} \sup_{k \geq m} k \frac{(\tilde{F}_k(t_0) - t_0)}{\sqrt{m} g(k/m)} \xrightarrow{D} \sup_{x \geq 1} \frac{w(x)}{g(x)}$$

as $m \rightarrow \infty$, for each fixed $0 < t_0 < 1$.

(c) Define a two-sided stopping rule by

$$\tau(m) := \inf \left\{ k: k \geq m, \sup_{\frac{1}{k} \leq t \leq 1 - \frac{1}{k}} k |\tilde{F}_k(t) - t| w(t) > \sqrt{m} g(k/m) \right\},$$

where m is an initial value and where g and w satisfy the assumptions of Corollary 5.1.2. Then we obtain from (5.1.0.): the probability of ever stopping the production, given that the null hypothesis of no

change is true, is

$$(5.1.1.) \quad \lim_{m \rightarrow \infty} P\{\tau(m) < \infty\} = P\left\{\sup_{x \geq 1} \sup_{0 < t < 1} \frac{|K(t, x)| w(t)}{g(x)} > 1\right\}.$$

If one now defines a function $h : [1, \infty) \rightarrow (0, \infty)$ such that $g(x) =: ch(x)$ for all $x \in [1, \infty)$, then one can (in principle at least) determine a critical value $c = c(\alpha)$ so that

$$(5.1.2.) \quad P\left\{\sup_{x \geq 1} \sup_{0 < t < 1} \frac{|K(t, x)| w(t)}{h(x)} \geq c\right\} = \alpha$$

for any given $\alpha \in (0, 1)$.

Consequently, putting (5.1.1.) and (5.1.2.) together, we see that for large enough initial values m the probability of stopping the production line when actually no change has occurred is approximately α .

(d) Similar results can be obtained for a one-sided stopping rule provided (5.1.0.*) is used in lieu of (5.1.0.).

After these introductory remarks special attention will now be given to the previously introduced CUSUM process L_t . The main goals are

- to formulate and prove a strong approximation result,
- to find the limiting distribution of certain weighted functionals of L_t ,
- to (approximately) compute the probability of stopping the production line when in reality no change had occurred.

Theorem 5.1.3. (Confer also Theorem 1. in V.R. Huse (1988))

Assuming that H_0 holds, we can define a Wiener process $W = \{W(t) : t \geq 0\}$ such that for $x \geq 2$ and for $n \geq n_0$ we have

$$(5.1.3.) \quad P\left\{ \sup_{2 \leq t \leq n} |L_t - W(t)| > A \log n + x \right\} \leq B e^{-\gamma x} \quad \text{and}$$

$$(5.1.3.*) \quad P\left\{ \max_{2 \leq k \leq n} \left| \int_0^1 k(t - \tilde{F}_k(t)) dt - \frac{1}{\sqrt{12}} W(t) \right| > C \log n + x \right\} \leq K e^{-\lambda x},$$

where $A, B, C, K, \gamma, \lambda$ and n_0 are absolute positive constants.

Proof of Theorem 5.1.3:

First we prove statement (5.1.3.). The uniform random variables U_i , $i \geq 2$, in representation (5.0.2.) satisfy the assumptions of the Komlós-Major-Tusnády Theorem 1.1.18. Consequently we have for every $n \geq 1$ and $y \in \mathbb{R}$:

$$(5.1.4.) \quad P\left\{ \sup_{2 \leq t \leq n} \left| \sqrt{12} \sum_{i=2}^{[t]} \left(U_i - \frac{1}{2} \right) - W([t]) \right| > A_1 \log n + y \right\} \leq B_1 e^{-\gamma_1 y},$$

where A_1, B_1, γ_1 are positive constants. (Recall that $EU_1 = \frac{1}{2}$ and $\text{Var } U_1 = \frac{1}{12}$.) Furthermore, (5.0.2.) implies that for every $\omega \in \Omega$ and for each $i \geq 2$:

$$0 \leq U_i(\omega) - \xi_i(\omega) \leq i^{-1}.$$

Hence we arrive at

$$(5.1.5.) \quad \sup_{2 \leq t \leq n} \left| \sqrt{12} \sum_{i=2}^{[t]} (U_i - \xi_i) \right| \leq \sqrt{12} \sum_{i=2}^{[t]} i^{-1} \leq \sqrt{12} \log n.$$

So, letting $A = 2A_1 + \sqrt{12}$, where $A_1 \geq 2$, and $x = 2y \geq 2$, we obtain

$$\begin{aligned} & P\left\{ \sup_{2 \leq t \leq n} |L_t - W(t)| > A \log n + x \right\} \\ & \leq P\left\{ \sup_{2 \leq t \leq n} \left| \sqrt{12} \sum_{i=2}^{[t]} \left(U_i - \frac{1}{2} \right) - W([t]) \right| + \sup_{2 \leq t \leq n} \left| \sqrt{12} \sum_{i=2}^{[t]} (U_i - \xi_i) \right| \right. \\ & \quad \left. + \sup_{2 \leq t \leq n} |W(t) - W([t])| > (2A_1 + \sqrt{12}) \log n + 2y \right\} \quad \text{by (5.0.4.)} \end{aligned}$$

$$\begin{aligned}
&\leq P\left\{ \sup_{2 \leq t \leq n} \left| \sqrt{12} \sum_{i=2}^{[t]} \left(U_i - \frac{1}{2} \right) - W([t]) \right| > A_1 \log n + y \right\} \\
&\quad + P\left\{ \sup_{2 \leq t \leq n} |W(t) - W([t])| > A_1 \log n + y \right\} \quad \text{by (5.1.5.)} \\
&\leq B_1 e^{-\gamma_1 y} + P\left\{ \sup_{2 \leq t \leq n} |W(t) - W([t])| > A_1 \log n + y \right\} \quad \text{by (5.1.4.).}
\end{aligned}$$

Summarizing these estimations, we have just proven that

$$(5.1.6.) \quad \left\{ \begin{aligned} &P\left\{ \sup_{2 \leq t \leq n} |L_t - W(t)| > A \log n + x \right\} \\ &\leq B_1 e^{-\gamma_1 y} + P\left\{ \sup_{2 \leq t \leq n} |W(t) - W([t])| > A_1 \log n + y \right\}. \end{aligned} \right.$$

With regard to the probability on the right-hand side of (5.1.6.) we have

$$\begin{aligned}
&P\left\{ \sup_{2 \leq t \leq n} |W(t) - W([t])| > A_1 \log n + y \right\} \\
&\leq P\left\{ \sup_{0 \leq s \leq n} \sup_{0 \leq t \leq 1} |W(s+t) - W(s)| > A_1 \log n + y \right\}
\end{aligned}$$

and by Lemma 1.2.1. in M. Csörgö and P. Révész (1981) (take there $h=1$, $\epsilon=2$, $T=n+1$, and $v=A_1 \log n + x$) we obtain now

$$\begin{aligned}
&\leq K_1 \cdot (n+1) \exp\left\{-\frac{1}{4}(A_1 \log n + x)^2\right\} \quad (\text{for some } K_1 > 0) \\
&\leq K_1 \cdot (n+1) \exp\left(-\frac{A_1^2}{4} \log n\right) \cdot \exp\left(-\frac{x^2}{4}\right) \\
&\leq K_1 \cdot (n+1) n^{-(A_1/2)^2} \exp\left(-\frac{x}{4}\right) \quad (\text{since } x \geq 1) \\
&\leq B_2 e^{-x/4} \quad \text{for some } B_2 > 0 \quad (\text{since } A_1 \geq 2).
\end{aligned}$$

Thus (5.1.3.) follows with $B = 2 \cdot \max\{B_1, B_2\}$ and $\gamma = \min\{\gamma, \frac{1}{4}\}$. ■

We conclude (5.1.3.*) immediately from (5.1.3.) and Lemma 5.1.0. ■

An application of the first Borel-Cantelli Lemma to (5.1.3.) and (5.1.3.*) gives:

Corollary 5.1.4.

Under the conditions of Theorem 5.1.3. we have as $n \rightarrow \infty$

$$\sup_{2 \leq t \leq n} |L_t - W(t)| = O(\log n) \quad \text{a.s.}$$

$$\max_{2 \leq k \leq n} \left| \int_0^1 k(t - F_k(t)) dt - (12)^{-1} W(k) \right| = O(\log n) \quad \text{a.s.}$$

Our next aim is to find the limiting distribution of a wide enough class of weighted functionals of L_t . Towards this end the following theorem is proposed.

Theorem 5.1.5. (Confer also Theorem 2. in V.R. Huse (1988))

Assume that H_0 holds. Let g be a strictly positive, continuous function on $[1, \infty)$. Suppose there exists a $T \geq 1$ such that g is monotonically nondecreasing on $[T, \infty)$. Furthermore assume

$$(5.1.7.) \quad \int_1^{\infty} t^{-1} \exp\left\{-c \frac{g^2(t)}{t}\right\} dt < \infty \quad \text{for some } c > 0.$$

Then, as $m \rightarrow \infty$

$$(5.1.8.) \quad \max_{k \geq m} \frac{L_k}{\sqrt{m} g(k/m)} \xrightarrow{D} \sup_{t \geq 1} \frac{W(t)}{g(t)},$$

where $W = \{W(t): t \geq 0\}$ denotes a Wiener process.

Proof of Theorem 5.1.5.

Before going into the proof of the convergence result itself, one has to ensure that the limiting random variable $\sup_{t \geq 1} W(t)/g(t)$ is almost surely finite. Hence the behaviour of the weight function g has to be examined more closely. Condition (5.1.7.) implies immediately

$$\exp \left\{ -c \frac{g^2(t)}{t} \right\} \rightarrow 0 \quad \text{as } t \rightarrow \infty$$

or equivalently

$$(5.1.9.) \quad \sqrt{t}/g(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty.$$

(5.1.9.) ensures that

$$(a) \quad \sup_{t \geq 1} W(t)/g(t) < \infty \quad \text{a.s.} \quad \text{and}$$

$$(b) \quad \text{there exists a } T^* \geq 1 \text{ such that } \frac{\sqrt{t}}{g(t)} \leq 1 \text{ for all } t \geq T^*.$$

Furthermore, the function f defined pointwise by

$$f(t) := \sqrt{t}/g(t) \quad \text{for } t \in [1, \infty),$$

is continuous on the compact interval $[1, T^*]$. Thus

$$M^* := \max_{1 \leq t \leq T^*} f(t) < \infty$$

exists. This together with Corollary 5.1.4. yields

$$\max_{k \geq m} \frac{|L_k - W(k)|}{\sqrt{m} g(k/m)} \leq \max_{k \geq m} \sup_{2 \leq t \leq k} \frac{|L_t - W(t)|}{\sqrt{m} g(k/m)}$$

$$\stackrel{\text{a.s.}}{=} \max_{k \geq m} \frac{O(\log k)}{\sqrt{m} g(k/m)}$$

$$(5.1.10.) \quad \leq \max_{k \geq m} \frac{O(\log k)}{\sqrt{k}} \cdot \max_{k \geq m} \frac{\sqrt{k/m}}{g(k/m)}$$

On examining the function h defined via

$$h(x) := x^{-1/2} \log x, \quad x > 0,$$

we see that

$$h'(x) = x^{-3/2} (1 - \frac{1}{2} \log x)$$

and consequently that h is monotonically nondecreasing over $(0, e^2)$, assumes its maximum at $x = e^2$, and is monotonically nonincreasing over (e^2, ∞) . Therefore, (5.1.10.) implies for $m \geq e^2$:

$$\max_{k \geq m} \frac{|L_k - W(k)|}{\sqrt{m} g(k/m)} \stackrel{\text{a.s.}}{\leq} \frac{O(\log m)}{\sqrt{m}} \cdot \left\{ \max_{m \leq k \leq mT^*} \frac{\sqrt{k/m}}{g(k/m)} \vee \max_{k \geq mT^*} \frac{\sqrt{k/m}}{g(k/m)} \right\}$$

where $\vee$ stands for maximum

$$\leq \frac{O(\log m)}{\sqrt{m}} \cdot \{M^* \vee 1\},$$

whence we have just verified that

$$(5.1.11.) \quad \max_{k \geq m} \frac{|L_k - W(k)|}{\sqrt{m} g(k/m)} \stackrel{\text{a.s.}}{=} o(1) \quad \text{as } m \rightarrow \infty.$$

Moreover, we know that $W(k) \stackrel{D}{=} \sqrt{m} W(k/m)$ for each $m, k \geq 1$. Therefore,

$$\frac{W(k)}{\sqrt{m} g(k/m)} \stackrel{D}{=} \frac{W(k/m)}{g(k/m)} \quad \text{for each } m, k \geq 1.$$

Using the fact that we can always choose an almost surely continuous version of the stochastic process $\{ \frac{W(t)}{g(t)} : t \geq 1 \}$, we obtain:

$$\sup_{k \geq m} \frac{W(k/m)}{g(k/m)} \stackrel{\text{a.s.}}{\rightarrow} \sup_{t \geq 1} \frac{W(t)}{g(t)} \quad \text{as } m \rightarrow \infty.$$

This together with (5.1.11.) immediately implies Theorem 5.1.5. ■

There are two different ways in which we can use Theorem 5.1.5. in order to draw conclusions concerning the testing of the changepoint hypotheses. On the one hand (5.1.8.) tells us that for large enough initial values m we have

$$P\left\{\max_{k \geq m} \frac{L_k}{\sqrt{m} g(k/m)} \geq x\right\} \doteq P\left\{\sup_{t \geq 1} \frac{W(t)}{g(t)} \geq x\right\}$$

for every $x \in \mathbb{R}$, where $\doteq$ denotes "approximately equal". Therefore, upon fixing a significance level $0 < \alpha < 1$, we can in principle compute a critical value $c = c_\alpha$ such that

$$P\left\{\sup_{t \geq 1} \frac{W(t)}{g(t)} \geq c\right\} = \alpha.$$

(The practical problem of actually computing $c(\alpha)$ will be discussed in Remark 5.1.7.) Consequently, one arrives at

$$P\left\{\max_{k \geq m} \frac{L_k}{\sqrt{m} g(k/m)} \geq c\right\} \doteq \alpha \text{ for sufficiently large } m.$$

And so a decision rule can be derived as follows: If the test statistic $T_m := \max_{k \geq m} \frac{L_k}{\sqrt{m} g(k/m)}$ based on the sample observations comes up with an experimental (observed) value c_e and if $c_e > c_\alpha$ (c_α being the theoretical critical value), then the null hypothesis of no change in the distribution is rejected at the (approximate) α -level of significance, i.e. we reject H_0 if T_m is too large, given H_0 . Note that one can obtain an analog result for a two-sided test because (5.1.8.) immediately implies:

$$(5.1.8.*) \quad \max_{k \geq m} \frac{|L_k|}{\sqrt{m} g(k/m)} \xrightarrow{D} \sup_{t \geq 1} \frac{|W(t)|}{g(t)} \quad \text{as } m \rightarrow \infty.$$

Properties of both the one- and the two-sided test will be discussed in Section 5.2.

A different way of looking at Theorem 5.1.5. results in:

Corollary 5.1.6. (Confer also Corollaries 2. and 3. in V.R. Huse (1988))

Under the conditions of Theorem 5.1.5.:

$$(5.1.12.) \quad \lim_{m \rightarrow \infty} P\{\tau_1(m) < \infty\} = P\left\{\sup_{t \geq 1} \frac{W(t)}{g(t)} > 1\right\},$$

where the stopping time $\tau_1(m)$ is as defined in (5.0.5.), and

$$(5.1.12.*) \quad \lim_{m \rightarrow \infty} P\{\tau_2(m) < \infty\} = P\left\{\sup_{t \geq 1} \frac{|W(t)|}{g(t)} > 1\right\}.$$

Here $\tau_2(m)$ is as defined in (5.0.5.).

Proof of Corollary 5.1.6.

Due to the similarity of statements (5.1.12.) and (5.1.12.*) only the former will be proven here.

Proof of (5.1.12.).

From Theorem 5.1.5. we conclude that

$$\lim_{m \rightarrow \infty} P\left\{\max_{k \geq m} \frac{L_k}{\sqrt{m} g(k/m)} > 1\right\} = P\left\{\sup_{t \geq 1} \frac{W(t)}{g(t)} > 1\right\},$$

whence it will suffice to show that

$$(5.1.13.) \quad \lim_{m \rightarrow \infty} P\{\tau_1(m) < \infty\} = \lim_{m \rightarrow \infty} P\left\{\max_{k \geq m} \frac{L_k}{\sqrt{m} g(k/m)} > 1\right\}.$$

Proof of (5.1.13.)

By definition of $\tau_1(m)$ in (5.0.5.) we have

$$\begin{aligned}
& \lim_{m \rightarrow \infty} P\{\tau_1(m) < \infty\} \\
&= \lim_{m \rightarrow \infty} P\{\inf\{k \geq m: \frac{L_k}{\sqrt{m} g(k/m)} > 1\} < \infty\} \\
&= 1 - \lim_{m \rightarrow \infty} P\{\inf\{k \geq m: \frac{L_k}{\sqrt{m} g(k/m)} > 1\} = \infty\} \\
&= 1 - \lim_{m \rightarrow \infty} P\left\{\frac{L_k}{\sqrt{m} g(k/m)} \leq 1 \text{ for all } k \geq m\right\} \\
&= \lim_{m \rightarrow \infty} P\left\{\frac{L_k}{\sqrt{m} g(k/m)} > 1 \text{ for some } k \geq m\right\} = \lim_{m \rightarrow \infty} P\left\{\max_{k \geq m} \frac{L_k}{\sqrt{m} g(k/m)} > 1\right\}.
\end{aligned}$$

Thus we have just proven (5.1.13.) and hence also (5.1.12.) ■

Remark 5.1.7.

(a) The tables available through the Wiener-pack package (confer C.F. Chung (1987)) provide calculated values for

$$P\{W(t) \leq c+dt: a \leq t \leq b, c > 0, c+db > 0, a \geq 0, b < \infty\}$$

as well as for

$$P\{|W(t)| \leq c+dt: a \leq t \leq b, c > 0, c+db > 0, a \geq 0, b < \infty\}$$

(Compare Routines GWPAB and GWPSAB of Wiener-pack, respectively.)

Taking $a=1$ one can hence compute

$$(5.1.14.) \quad P\left\{\sup_{1 \leq t \leq b} \frac{W(t)}{c+dt} > 1\right\} \quad \text{and}$$

$$(5.1.14.*) \quad P\left\{\sup_{1 \leq t \leq b} \frac{|W(t)|}{c+dt} > 1\right\} \quad \text{for } c, d > 0 \text{ specified.}$$

It is clear that (5.1.14) and (5.1.14.*) with b large provide good

approximations to

$$P\left\{\sup_{t \geq 1} \frac{W(t)}{c+dt} > 1\right\} \quad \text{and}$$

$$P\left\{\sup_{t \geq 1} \frac{|W(t)|}{c+dt} > 1\right\}, \quad \text{respectively.}$$

Hence we can approximately compute via the aforementioned routines of Wiener-pack:

$$P\left\{\sup_{t \geq 1} \frac{W(t)}{g(t)} \geq k\right\} \quad \text{and}$$

$$P\left\{\sup_{t \geq 1} \frac{|W(t)|}{g(t)} \geq k\right\}$$

for $g(t) := \frac{1}{k}(c+dt)$, $t \geq 1$. We note that this particular choice of g satisfies (5.1.7.). This solves the problem of the actual computation of the (approximate) critical values for linear weight functions g .

(b) For some smaller weight functions (in particular those of order $(t \log t)^{1/2}$ and $(t \log \log t)^{1/2}$) exact probabilities in Corollary 5.1.6. may be derived from those computed by H. Robbins and D. Siegmund (1970). Robbins and Siegmund obtain some of their results from the following statement (confer Theorem 1 in their publication):

$$(5.1.15.) \quad \begin{cases} P\{W(t) \geq A(t+h, \varepsilon) - b \text{ for some } t \geq \tau\} \\ = 1 - \Phi\left(\frac{A(\tau+h, \varepsilon) - b}{\sqrt{\tau}}\right) + \frac{1}{\varepsilon} \int_0^\infty \exp\{-by - y^2 h/2\} \Phi\left(\frac{A(\tau+h, \varepsilon) - b}{\sqrt{\tau}} - y\tau^{1/2}\right) dF(y), \end{cases}$$

where

Φ denotes the cumulative distribution function of a standard normal random variable,

F denotes any measure on $(0, \infty)$ which is finite on bounded intervals,

for $x, t \in \mathbb{R}$ and $0 < \varepsilon < \infty$:

$$0 \leq f(x, t) := \int_0^\infty \exp\{xy - y^2 t/2\} dF(y) \leq \infty, \quad \text{and}$$

$$-\infty \leq A(t, \varepsilon) := \inf \{x: f(x, t) \geq \varepsilon\} < \infty,$$

and b, h, τ denote positive constants.

Taking $b=0$, $\tau=1$, $a>0$, $\varepsilon = e^{a^2/2}$, $dF(y) = (2\pi)^{-1/2} dy$, and

$$g(t) = (t(a^2 + \log t))^{1/2} \quad \text{for } t \geq 1,$$

we have the following elegant example:

Example 5.1.8.

$$\begin{aligned} & \lim_{m \rightarrow \infty} P\{\tau_2(m) < \infty\} \\ &= \lim_{m \rightarrow \infty} P\{\inf \{k: k \geq m, |L_k| > \sqrt{m} g(k/m)\} < \infty\} \\ &= \lim_{m \rightarrow \infty} P\{\inf \{k: k \geq m, |L_k| > \sqrt{k}(a^2 + \log \frac{k}{m})^{1/2}\} < \infty\} \\ &= P\left\{\sup_{t \geq 1} \frac{|W(t)|}{(t(a^2 + \log t))^{1/2}} > 1\right\} \\ &= 2[1 - \Phi(a) + a\phi(a)], \end{aligned}$$

where ϕ denotes the density function of a standard Gaussian random variable.

Section 5.2. Results Under the Alternative Hypothesis

Throughout this section we assume that the alternative hypothesis of exactly one changepoint occurring is true and that the change happens at (unknown) time N . Furthermore, assume that $X_1, \dots, X_N$ have distribution F where F is continuous and similarly that $X_{N+1}, X_{N+2}, \dots$ possess distribution G with G continuous. In this setup $F \neq G$ in the sense that there exists an $x \in \mathbb{R}$ such that $F(x) \neq G(x)$. Then we have

Theorem 5.2.0.

Assume that H_A holds. Suppose that $EG(X_1) \neq 1/2$. Then there exists a positive constant c_0 such that

$$(5.2.0.) \quad \liminf_{N \rightarrow \infty} (k-N)^{-1} \left| \int_0^1 k(t - \tilde{F}_k(t)) dt \right| \geq c_0 \text{ a.s.}$$

whenever $N+1 \leq k \leq 2N$ and $(k-N) \cdot (N \log \log N)^{-1/2} \rightarrow \infty$ as $N \rightarrow \infty$.

Remark 5.2.1.

In case we are only interested in convergence in probability instead of in almost sure convergence, we can weaken the condition of Theorem 5.2.0. as follows:

Theorem 5.2.2.

Under the assumptions of Theorem 5.2.0. there exists a constant $c_1 > 0$ such that

$$(5.2.1.) \quad \lim_{N \rightarrow \infty} P \left\{ (k-N)^{-1} \left| \int_0^1 k(t - \tilde{F}_k(t)) dt \right| \geq c_1 \right\} = 0$$

whenever $N+1 \leq k \leq 2N$ and $(k-N) \cdot N^{-1/2} \rightarrow \infty$ as $N \rightarrow \infty$.

Remarks 5.2.3.

(a) Lemma 5.1.0. and P. Lévy's law of iterated logarithm for the Wiener process jointly imply that the two-sided test based on L_k is asymptotically consistent against the class of alternative hypotheses:

$$H_A : EG(X_1) \neq 1/2; F, G \text{ continuous.}$$

(b) It follows from (a) that the one-sided test is also asymptotically consistent against

$$H_A^* : EG(X_1) = \frac{1}{2} + \delta, \quad \delta < 0 \text{ or } \delta > 0.$$

(c) Choosing the special weight function

$$g(x) = c \sqrt{x \log(\log x \vee 3)}, \quad x \geq 1,$$

where $\vee$ stands for maximum, $c > 0$ constant, Theorem 5.2.2. implies for the two-sided stopping rule

$$\tau_3(m) := \inf \left\{ n : n \geq m, \left| \int_0^1 n(t - \tilde{F}_n(t)) dt \right| > \sqrt{m} g(n/m) \right\}:$$

$$\lim_{m \rightarrow \infty} P\{\tau_3(m) \leq m \vee N + c\sqrt{N \log \log N}\} = 1.$$

Thus, for large enough initial value m , the test procedure will approximately stop either at m or $c\sqrt{N \log \log N}$ steps after the change has occurred.

(d) The next lemma is needed for the proof of Theorem 5.2.0. A proof of Lemma 5.2.4. can be found in M. Csörgö and L. Horváth (1987a), confer in particular statement (3.3.) and its proof there.

Lemma 5.2.4.

Assume that H_A holds, then

$$\limsup_{N \rightarrow \infty} \max_{N+1 \leq j \leq 2N} \sup_{x \in \mathbb{R}} |F_j(x) - H_{j,N}(x)| \sqrt{N} \cdot (\log N)^{-\frac{1}{2}} \leq A \text{ a.s.}$$

for some positive constant A .

Here F_j is again the empirical distribution function of X_1 and

$$H_{j,N}(x) := \frac{N+1}{j+1} F(x) + \frac{j-N}{j+1} G(x)$$

for $x \in \mathbb{R}$ and $N+1 \leq j \leq 2N$.

Proof of Theorem 5.2.0.

Introduce the following notations:

$$F_k(x) := \frac{1}{k+1} \# \{1 \leq j \leq k: X_j \leq x\} \quad \text{for } x \in \mathbb{R}, k \geq 1,$$

represents the empirical distribution function of $X_1, X_2, \dots, X_k$;

$$\tilde{F}_k(t) := \frac{1}{k} \# \{2 \leq j \leq k: \xi_j \leq t\} \quad \text{for } t \in (0,1), k \geq 2,$$

where ξ_j is the sequential rank defined in (5.0.1.);

$$G_{k,N}(x) := \frac{1}{k-N} \# \{N+1 \leq j \leq k: X_j \leq x\} \quad \text{for } x \in \mathbb{R}, N+1 \leq k \leq 2N,$$

so $G_{k,N}$ is the empirical distribution function after the change;

$$\text{and } \tilde{G}_{k,N}(t) := \frac{1}{k-N} \# \{N+1 \leq j \leq k: \xi_j \leq t\} \quad \text{for } t \in (0,1), N+1 \leq k \leq 2N.$$

Then we have:

$$\begin{aligned} \tilde{F}_k(t) &= \frac{1}{k} \# \{2 \leq j \leq k: \xi_j \leq t\} \\ &= \frac{1}{k} \# \{2 \leq j \leq N: \xi_j \leq t\} + \frac{1}{k} \# \{N+1 \leq j \leq k: \xi_j \leq t\} \\ &= \frac{N}{k} \tilde{F}_N(t) + \frac{k-N}{k} \tilde{G}_{k,N}(t), \end{aligned}$$

and consequently

$$(5.2.2.) \quad \int_0^1 k(t - \tilde{F}_k(t)) dt = \int_0^1 (k-N)(t - \tilde{G}_{k,N}(t)) dt - \int_0^1 N(\tilde{F}_N(t) - t) dt.$$

Next, applications of Theorem 2.1. in M. Csörgő and L. Horváth (1987a) and of the law of iterated logarithm for the Kiefer process (confer, e.g., Corollary 1.15.1. in M. Csörgő and P. Révész (1981)) yield:

$$(5.2.3.) \quad \limsup_{N \rightarrow \infty} \frac{1}{\sqrt{N \log \log N}} \sup_{0 \leq t \leq 1} N |\tilde{F}_N(t) - t| = 2^{-1/2} \text{ a.s.}$$

Since (5.2.2.) and (5.2.3.) together imply

$$\liminf_{N \rightarrow \infty} \frac{1}{k-N} \left| \int_0^1 k(t - \tilde{F}_k(t)) dt \right| \geq \liminf_{N \rightarrow \infty} \left| \int_0^1 (t - \tilde{G}_{k,N}(t)) dt \right| \text{ a.s.}$$

(also recall the condition that $(k-N) \cdot (N \log \log N)^{-1/2} \rightarrow \infty$ as $N \rightarrow \infty$),

Theorem 5.2.0. will be proven if it can be shown that

$$(5.2.0.*) \quad \liminf_{N \rightarrow \infty} \left| \int_0^1 (t - \tilde{G}_{k,N}(t)) dt \right| \geq c_0 \text{ a.s.}$$

holds for some constant $c_0 > 0$.

Towards this end notice that Lemma 5.2.4. implies that there exists a random variable $N_0 = N_0(\omega)$ such that for all $N \geq N_0$ we have

$$-AN^{-1/2} \sqrt{\log N} \leq F_j(x) - H_{j,N}(x) \leq AN^{-1/2} \sqrt{\log N},$$

and so in particular:

$$(5.2.4.) \quad F_j(x) \leq H_{j,N}(x) + AN^{-1/2} \sqrt{\log N} \quad \text{for all } N \geq N_0 \text{ and}$$

$$(5.2.4.*) \quad F_j(x) \geq H_{j,N}(x) - AN^{-1/2} \sqrt{\log N} \quad \text{for all } N \geq N_0.$$

With these inequalities we obtain:

$$\begin{aligned}
& \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t - AN^{-1/2} \sqrt{\log N}))\} \\
&= \frac{1}{k-N} \sum_{j=N+1}^k I\{G(X_j) \leq G(H_{j-1,N}^{-1}(t - AN^{-1/2} \sqrt{\log N}))\} \\
&= \frac{1}{k-N} \sum_{j=N+1}^k I\{X_j \leq H_{j-1,N}^{-1}(t - AN^{-1/2} \sqrt{\log N})\} \\
&= \frac{1}{k-N} \sum_{j=N+1}^k I\{H_{j-1,N}(X_j) + AN^{-1/2} \sqrt{\log N} \leq t\} \\
&\leq \frac{1}{k-N} \sum_{j=N+1}^k I\{F_{j-1}(X_j) \leq t\} \quad \text{by (5.2.4.) } (*) \\
&\leq \frac{1}{k-N} \sum_{j=N+1}^k I\{H_{j-1,N}(X_j) - AN^{-1/2} \sqrt{\log N} \leq t\} \quad \text{by (5.2.4.*)}
\end{aligned}$$

and retracing the previous steps

$$= \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t + AN^{-1/2} \sqrt{\log N}))\}.$$

But we also have in (*)

$$\begin{aligned}
& \frac{1}{k-N} \sum_{j=N+1}^k I\{F_{j-1}(X_j) \leq t\} \\
&= \frac{1}{k-N} \# \{N+1 \leq j \leq k: F_{j-1}(X_j) \leq t\} \\
&= \frac{1}{k-N} \# \{N+1 \leq j \leq k: \xi_j \leq t\} \quad \text{by (5.0.1.)} \\
&= \tilde{G}_{k,N}(t) \quad \text{by definition in the beginning of this proof.}
\end{aligned}$$

Hence we have currently shown that

$$\begin{aligned}
& \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t - AN^{-1/2} \sqrt{\log N}))\} \\
&\leq \tilde{G}_{k,N}(t)
\end{aligned}$$

$$\leq \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t + AN^{-1/2} \sqrt{\log N}))\}.$$

Since monotonicity arguments also yield

$$\frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t - AN^{-1/2} \sqrt{\log N}))\}$$

$$\leq \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t))\}$$

$$\leq \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t + AN^{-1/2} \sqrt{\log N}))\},$$

we have now accomplished the proof of

$$\begin{aligned} (5.2.5) \quad & \left| \tilde{G}_{k,N}(t) - \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t))\} \right| \\ & \leq \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(H_{j-1,N}^{-1}(t - AN^{-1/2} \sqrt{\log N})) \leq G(X_j) \\ & \quad \leq G(H_{j-1,N}^{-1}(t + AN^{-1/2} \sqrt{\log N}))\} \\ & = \frac{1}{k-N} \# \{N+1 \leq j \leq k: t - AN^{-1/2} \sqrt{\log N} \leq H_{j-1,N}(X_j) \leq t + AN^{-1/2} \sqrt{\log N}\} \\ & = \frac{1}{k-N} \# \{N+1 \leq j \leq k: t - AN^{-1/2} \sqrt{\log N} \leq \frac{N+1}{j} F(X_j) + \frac{j-1-N}{j} G(X_j) \\ & \quad \leq t + AN^{-1/2} \sqrt{\log N}\}. \end{aligned}$$

Since for each fixed $t \in (0,1)$ by the assumed continuity of F and G the last expression converges to zero as $N \rightarrow \infty$, Lebesgue's dominated convergence theorem implies together with (5.2.5.)

$$\lim_{N \rightarrow \infty} \left| \int_0^1 (\tilde{G}_{k,N}(t) - \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t))\}) dt \right| \stackrel{a.s.}{=} 0$$

and so it suffices to show (5.2.6.):

$$\lim_{N \rightarrow \infty} \inf_0 \left| \int_0^1 [t - \frac{1}{k-N} \# \{N+1 \leq j \leq k: G(X_j) \leq G(H_{j-1,N}^{-1}(t))\}] dt \right| \stackrel{a.s.}{=} c_0$$

in lieu of (5.2.0.*) earlier by the reverse triangle inequality. In order to verify (5.2.6.) we first perform the necessary integration to arrive at:

$$\begin{aligned}
 & \int_0^1 \left[t - \frac{1}{k-N} \# \{ N+1 \leq j \leq k : G(X_j) \leq G(H_{j-1,N}^{-1}(t)) \} \right] dt \\
 &= \frac{1}{2} - \frac{1}{k-N} \int_0^1 \# \{ N+1 \leq j \leq k : \frac{N+1}{j} F(X_j) + \frac{j-1-N}{j} G(X_j) \leq t \} dt \\
 &= \frac{1}{2} - \frac{1}{k-N} \sum_{j=N+1}^k \int_0^1 I \left\{ \frac{N+1}{j} F(X_j) + \frac{j-1-N}{j} G(X_j) \leq t \right\} dt \\
 &= \frac{1}{2} - \frac{1}{k-N} \sum_{j=N+1}^k \left\{ 1 - \frac{N+1}{j} F(X_j) - \frac{j-1-N}{j} G(X_j) \right\} \\
 &= \frac{1}{2} - 1 + \frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} F(X_j) + \frac{1}{k-N} \sum_{j=N+1}^k G(X_j) - \frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} G(X_j) \\
 &= \left\{ -\frac{1}{2} + \frac{1}{k-N} \sum_{j=N+1}^k G(X_j) \right\} + \left[\frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} F(X_j) - \frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} G(X_j) \right]
 \end{aligned}$$

and finally since by the strong law of large numbers

$$\frac{1}{k-N} \sum_{j=N+1}^k G(X_j) \xrightarrow{\text{a.s.}} \frac{1}{2} \quad \text{as } (k-N) \rightarrow \infty,$$

we have reduced the proof of (5.2.6.) to

$$(5.2.6.*) \quad \liminf_{N \rightarrow \infty} \left| \frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} F(X_j) - \frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} G(X_j) \right| \stackrel{\text{a.s.}}{\geq} c_0.$$

Proof of (5.2.6.*)

Both of the above sums converge absolutely because

$$0 \leq \frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} F(X_j) \leq \frac{1}{k-N} \sum_{j=N+1}^k \frac{N+1}{j} \leq 1 \quad \text{and}$$

$$0 \leq \frac{1}{k-N} \sum_{j=N+1}^k \left\{ \frac{N+1}{j} G(X_j) \right\} \leq \frac{1}{k-N} \sum_{j=N+1}^k G(X_j) \rightarrow \frac{1}{2} \text{ as } (k-N) \rightarrow \infty$$

by the general convergence test for series of positive terms (compare, e.g., statement (12.22.) of A. Mizrahi and M. Sullivan (1982)). Since both sums are of the same structure, each will converge to the same proportion of the respective mean. Hence the assumed condition $EF(X_{N+1}) \neq 1/2 = EG(X_{N+1})$ is sufficient to guarantee the existence of a positive constant c_0 in (5.2.6.*). This completes the proof of Theorem 5.2.0. ■

Proof of Theorem 5.2.2.

$$\begin{aligned} & P\left\{ \left| \int_0^1 (t - \tilde{G}_{k,N}(t)) dt \right| \geq c_1 \right\} \\ &= P\left\{ \frac{1}{k-N} \left| \left[\int_0^1 (k-N)(t - \tilde{G}_{k,N}(t)) dt - \int_0^1 N(\tilde{F}_N(t) - t) dt \right] + \int_0^1 N(\tilde{F}_N(t) - t) dt \right| \geq c_1 \right\} \\ &\leq P\left\{ \frac{1}{k-N} \left| \int_0^1 (k-N)(t - \tilde{G}_{k,N}(t)) dt - \int_0^1 N(\tilde{F}_N(t) - t) dt \right| \geq c_1/2 \right\} \\ &\quad + P\left\{ \frac{1}{k-N} \left| \int_0^1 N(\tilde{F}_N(t) - t) dt \right| \geq c_1/2 \right\} \\ &\leq P\left\{ \frac{1}{k-N} \left| \int_0^1 k(t - \tilde{F}_k(t)) dt \right| \geq c_1/2 \right\} \\ &\quad + P\left\{ \sup_{0 \leq t \leq 1} \sqrt{N} |\tilde{F}_N(t) - t| \geq \frac{c_1}{2} \cdot \frac{k-N}{\sqrt{N}} \right\} \quad \text{by (5.2.2.)} \end{aligned}$$

and by the Dvoretzky-Kiefer-Wolfowitz inequality (compare, e.g., Theorem 4.1.3. in M. Csörgő and P. Révész (1981), noting that we are here dealing with a uniform empirical process since the empirical distribution

$\tilde{F}_N$ of the sequential ranks is based solely on the data collected before the change in distribution)

$$\leq P\left\{\frac{1}{k-N} \left| \int_0^1 k(t - \tilde{F}_k(t)) dt \right| \geq c_1/2\right\} + c_2 \exp\left\{-\frac{c_1^2}{2} \frac{(k-N)^2}{N}\right\}.$$

By assumption we know that $\frac{k-N}{\sqrt{N}} \rightarrow \infty$ as $N \rightarrow \infty$ and therefore it suffices to show

$$(5.2.7.) \quad P\left\{\left| \int_0^1 (t - \tilde{G}_{k,N}(t)) dt \right| \geq 2c_1\right\} \rightarrow 1 \text{ as } N \rightarrow \infty.$$

In the same vein the proof of (5.2.0.*) was reduced to the verification of (5.2.6.) earlier in the proof of Théorem 5.2.0., (5.2.7.) can be reduced to

$$(5.2.7.*) \quad P\left\{\left| \int_0^1 \left(t - \frac{1}{k-N} \# \{N+1 \leq j \leq k : \frac{N+1}{j} F(X_j) + \frac{j-1-N}{j} G(X_j) \leq t\} \right) dt \right| \geq c_1\right\} \rightarrow 1$$

as $N \rightarrow \infty$.

The remainder of the proof follows along the same lines as the proof of the previous theorem and is hence omitted here. ■

Appendix A. Numerical Methods for Computing the Distribution of Quadratic Forms in Normal Random Variables

The aim of this appendix is to describe one unified approach to the numerical evaluation of

$$P\left\{\int_0^1 X^2(t)dt \leq y\right\},$$

where X denotes a stochastic process. The method described here was applied to obtain those tables in the text that display probabilities of the above form for Brownian Bridges, the process Γ and weighted versions thereof in lieu of X . The discussion in this appendix is based on a paper by S. Kotz, N.L. Johnson and D.W. Boyd (1967) to which we will refer as KJB in the sequel.

As the starting point we take the series expansion

$$(A1) \quad \int_0^1 X^2(t)dt \stackrel{a.s.}{=} \sum_{k,m=1}^{\infty} a_{km} Y_k Y_m,$$

where $(Y_k)_{k \geq 1}$ stands for a sequence of independent identically distributed $N(0,1)$ -random variables. Let $A_{K \times K} = A$ denote the square $K \times K$ -matrix of the first K^2 coefficients in (A1), i.e.

$$A = (a_{km})_{k,m=1}^K.$$

If K is chosen suitably large (how to arrive at a particular choice of K will be discussed later on), then we have

$$P\left\{\int_0^1 X^2(t)dt \leq y\right\} \doteq P\left\{\sum_{k,m=1}^K a_{km} Y_k Y_m \leq y\right\} \quad \text{for all } y,$$

where $\doteq$ symbolizes approximate equality.

Hence the problem is now reduced to finding the distribution of a

quadratic form in independent identically distributed standard normal random variables. But this situation is a special case of what KJB consider. By (3) of KJB we then have:

$$P\left\{\sum_{k,m=1}^K a_{km} Y_k Y_m \leq y\right\} = P\left\{\sum_{i=1}^K \alpha_i Z_i^2 \leq y\right\} \quad \text{for all } y, \quad (A2)$$

where $(Z_i)_{i \geq 1}$ denotes a sequence of independent identically distributed $N(0,1)$ -random variables and where

$$\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_K$$

are the eigenvalues of A .

In the following we will now summarize two approaches by KJB to finding series representations for the right hand side of (A2). The first approach is based on a power series expansion and is conceptually easier than the second approach which gives a Laguerre expansion.

According to KJB the Laguerre expansion is preferable for constructing computer routines. For this reason we have used the Laguerre expansion in our computations of the distribution of the aforementioned Gaussian functionals. The results were displayed in the tables of Sections 2.3. and 3.1.

Power series expansion

$$\text{Put } G_K(\underline{\alpha}; y) := G_K(\alpha_1, \dots, \alpha_K; y) := P\left\{\sum_{i=1}^K \alpha_i Z_i^2 \leq y\right\}.$$

For the eigenvalues α_j of A define:

$$C_K := C := \prod_{j=1}^K \alpha_j^{-1/2}.$$

Let $c_0^P := C$ and define recursively for $n \geq 1$:

$$c_n^P := \frac{1}{n} \sum_{r=0}^{n-1} d_{n-r}^P c_r^P, \quad \text{where}$$

$$d_n^P := \sum_{j=1}^K \alpha_j^{-n}, \quad n \geq 1.$$

Using these notations, KJB (compare statement (40) there) obtain:

$$(A3) \quad G_K(\underline{\alpha}; y) = (y/2)^{K/2} \sum_{n=0}^{\infty} c_n^P (-1)^n (y/2)^n / \Gamma(\frac{K}{2} + n + 1).$$

If this representation is used to construct a computer algorithm, then the infinite series in (A3) has to be truncated. For this purpose KJB give the following error bound (confer their statements (16) and (45)):

$$E_N^P(y) := \left| \sum_{n=N+1}^{\infty} c_n^P (-1)^n (y/2)^{(K/2)+n} / \Gamma(\frac{K}{2} + n + 1) \right| \\ \leq \left(\frac{y}{2\alpha_K} \right)^{(K/2)+N+1} \cdot \{ (N+1)! \left(\frac{K}{2} + N + 1 \right) \Gamma(\frac{K}{2}) \}^{-1}.$$

Laguerre expansion.

Let again $G_K(\underline{\alpha}; y) := P\{\sum_{i=1}^K \alpha_i Z_i^2 \leq y\}$ and denote by $G(n; y)$ the cumulative distribution function (at argument y) of a $\chi_n^2 = \chi_n^2$ random variable with n degrees of freedom. Set $\beta := \beta(K) := \frac{1}{2}(\alpha_1 + \alpha_K)$; where α_1 is the largest and α_K the smallest eigenvalue of $A = A_{K \times K}$. Recall that for $a > 1$ the generalized Laguerre polynomial $L_K^{(a)}(x)$ of degree k is defined by Rodrigues' formula as:

$$L_K^{(a)}(x) := \frac{1}{k!} e^{x-a} \frac{d^k}{dx^k} [e^{-x} x^{k+a}].$$

Using these Laguerre polynomials, KJB arrive at (confer their statement (71)):

$$(A4) \quad G_K(\underline{\alpha}; y) = G(K; y/\beta) + \sum_{n=1}^{\infty} \frac{c_n^L (n-1)!}{\Gamma(\frac{K}{2} + n)} \left(\frac{y}{2\beta} \right)^{K/2} e^{-y/2} L_{n-1}^{(K/2)}(y/(2\beta)),$$

where the coefficients c_n^L are to be determined recursively via

$$c_0^L := 1$$

$$c_n^L := \frac{1}{n} \sum_{r=0}^{n-1} d_{n-r}^L c_r^L \quad \text{for } n \geq 1, \text{ where}$$

$$d_k^L := \frac{1}{2} \operatorname{tr} \left(I - \frac{1}{\beta} A \right)^k.$$

Here $\operatorname{tr}(B)$ denotes the trace, i.e. the sum of the diagonal elements, of the matrix $B := \left(I - \frac{1}{\beta} A \right)^k$.

Finally KJB suggest the following formula to be used to control the accuracy of the computer routines.

$$(A5) \quad E_N^L(y) \leq \frac{K}{N+1} g(K+2; y/\beta) \frac{e^{y/(4\beta)} \varepsilon^{\frac{1}{2}(N+2)}}{(1-\sqrt{\varepsilon})^{K/2+2}},$$

$$\text{where } \varepsilon = \frac{\alpha_1 - \alpha_K}{\alpha_1 + \alpha_K}.$$

Appendix B. Differential Equations Related to $\int_0^1 t(1-t)B^2(t)dt$

Here we will demonstrate the initial steps necessary to determine the distribution of

$$\int_0^1 t(1-t)B^2(t)dt.$$

using the Karhnen-Love method. The functional $\int_0^1 t(1-t)B^2(t)dt$ would play a role in a test situation where one wishes to 'play down' the importance of the tail behaviour of $\hat{U}_k$ under the null hypothesis of no change by using a corresponding test statistic based on the latter process. Here we 'smooth down' the tails as opposed to highlighting them as we did when considering the distribution of

$$\int_0^1 \frac{B^2(t)}{t(1-t)} dt.$$

Define $X(t) = \sqrt{t(1-t)} B(t)$, then $K_X(s, t) = \text{Cov}(X(s), X(t)) = \sqrt{s(1-s)t(1-t)} (s \wedge t - st)$ for $0 \leq s, t \leq 1$. To find the eigenvalues $\lambda_k > 0$ for representation (3.1.5.) we have to solve the integral equation

$$\lambda f(t) = \int_0^1 \sqrt{s(1-s)t(1-t)} (s \wedge t - st) f(s) ds \quad \text{subject to } f(0) = f(1) = 0$$

for $0 \leq t \leq 1$.

This leads to the second order homogeneous differential equation

$$\lambda h''(t) = -t(1-t)h(t) \quad \text{for } t \in [0, 1],$$

where $f(t) = \sqrt{t(1-t)} h(t)$, as shown in Chapter 3, Subsection 3.1.1.

(take there $\psi(t) = t(1-t)$ in (3.1.10.)). The above differential equation is of the general form

$$y'' + (ax+b)y' + (\alpha x^2 + \beta x + \gamma)y = 0 \quad \text{with } a = b = \gamma = 0, \alpha = -\frac{1}{\lambda}, \text{ and } \beta = \frac{1}{\lambda}.$$

Let s be a root of $4s^2 + \alpha = 0$, i.e. $s = \pm \frac{1}{2\sqrt{\lambda}}$, and put $y(x) = u(x)e^{sx^2}$, then

$$y'(x) = u'(x)e^{sx^2} + (2sx)u(x)e^{sx^2} \quad \text{and}$$

$$y''(x) = u''(x)e^{sx^2} + (4sx)u'(x)e^{sx^2} + ((2sx)^2 + 2s)u(x)e^{sx^2}.$$

Therefore, the following statements are equivalent:

$$y'' + (\alpha x^2 + \beta x)y = 0$$

$$u''(x)e^{sx^2} + (4sx)u'(x)e^{sx^2} + ((2sx)^2 + 2s)u(x)e^{sx^2} + (\alpha x^2 + \beta x)u(x)e^{sx^2} = 0,$$

$$u''(x) + (4sx)u'(x) + ((2sx)^2 + 2s + \alpha x^2 + \beta x)u(x) = 0,$$

$$u''(x) + (4sx)u'(x) + (\beta x + 2s)u(x) = 0.$$

Thus, we must now solve the easier differential equation

$$u''(x) + (4sx)u'(x) + (\beta x + 2s)u(x) = 0.$$

Set $u(x) = v(\xi)\exp(-\frac{\beta}{4s}x)$ with $\xi = \sqrt{4s}\left(x - \frac{2\beta}{4s}\right)$, then $\frac{d\xi}{dx} = \sqrt{4s}$ and thus

$$u'(x) = (\sqrt{4s} v'(\xi) - \frac{\beta}{4s} v(\xi)) \exp(-\frac{\beta}{4s}x) \quad \text{and}$$

$$u''(x) = (4s v''(\xi) + \frac{2\beta}{\sqrt{4s}} v'(\xi) + \frac{\beta^2}{(4s)^2} v(\xi)) \exp(-\frac{\beta}{4s}x), \text{ whence}$$

$$u''(x) + (4sx)u'(x) + (\beta x + 2s)u(x) = 0 \quad \text{if and only if}$$

$$4s v''(\xi) + \left\{ (4sx) \sqrt{4s} + \frac{2\beta}{\sqrt{4s}} \right\} v'(\xi) + \left\{ \frac{\beta^2}{(4s)^2} + 2s \right\} v(\xi) = 0$$

or equivalently

$$4s v''(\xi) + \left[4s \sqrt{4s} \left[\frac{\xi}{\sqrt{4s}} + \frac{2\beta}{(4s)^2} \right] + \frac{2\beta}{\sqrt{4s}} \right] v'(\xi) + \left[\frac{\beta^2}{(4s)^2} + 2s \right] v(\xi) = 0$$

$$v(\xi) = 0$$

or equivalently

$$v''(\xi) \pm \xi v'(\xi) \pm (4s)^{-3} (\beta^2 + 32s^3) v(\xi) = 0,$$

where the upper sign applies when $s > 0$ and the lower sign occurs if $s < 0$. Consequently, we now have to solve each of the following differential equations:

$$v''(\xi) + \xi v'(\xi) + \frac{1+4\sqrt{\lambda}}{2\sqrt{\lambda}} v(\xi) = 0,$$

$$v''(\xi) - \xi v'(\xi) + \frac{1-4\sqrt{\lambda}}{2\sqrt{\lambda}} v(\xi) = 0.$$

The general solutions of these equations are

$$v(\xi) = \frac{1}{\sqrt{x}} e^{-\frac{x^2}{4}} v\left(\frac{1+4\sqrt{\lambda}}{4\sqrt{\lambda}} - \frac{1}{4}, \frac{1}{4}, \frac{x^2}{2}\right) \quad \text{and}$$

$$v(\xi) = \frac{1}{\sqrt{\xi}} e^{\frac{1}{4}\xi^2} v\left(-\frac{1-4\sqrt{\lambda}}{4\sqrt{\lambda}} - \frac{1}{4}, \frac{1}{4}, -\frac{\xi^2}{2}\right), \quad \text{respectively,}$$

where $v(k, m, \xi)$ is a solution of $\xi v'' + (2m+1-\xi)v' + (k-m-\frac{1}{2})v = 0$.

Since $2m$ here is not an integer, we know that

$$v = c_1 M_{k,m}(\xi) + c_2 M_{k,m}(\xi) \quad \text{with } M_{k,m} \text{ being a Whittaker function}$$

is the form of our solutions. (Confer E. Kamke (1951), C.2.273.)

Appendix C. Miscellaneous Technical Details

For Chapter 3, Subsection 3.1.2.

$$\begin{aligned}
 \int_0^1 \int_0^1 K_T^2(s, t) ds dt &= \int_0^1 \int_0^1 \{ (1-2s)(1-2t)(s \wedge t) + (3-2s-2t)st \}^2 ds dt \\
 &= \int_0^1 \left\{ (1-2t)^2 \int_0^1 (1-2s)^2 (s \wedge t)^2 ds \right. \\
 &\quad + 2t(1-2t) \int_0^1 (1-2s)(s \wedge t)(3-2s-2t)s ds \\
 &\quad \left. + t^2 \int_0^1 (3-2s-2t)^2 s^2 ds \right\} dt.
 \end{aligned}$$

Next we evaluate the three inner integrals separately one by one.

$$\begin{aligned}
 \int_0^1 (1-2s)^2 (s \wedge t)^2 ds &= \int_0^t (s^2 - 4s^3 + 4s^4) ds + t^2 \int_t^1 (1-4s+4s^2) ds \\
 &= \frac{t^2}{3} - \frac{2}{3} t^3 + t^4 - \frac{8}{15} t^5;
 \end{aligned}$$

$$\begin{aligned}
 \int_0^1 (1-2s)(s \wedge t)(3-2s-2t)s ds &= \int_0^t (1-2s)s^2(3-2s-2t) ds \\
 &\quad + t \int_t^1 (1-2s)(3-2s-2t)s ds \\
 &= \int_0^t (3s^2 - 8s^3 + 4s^4 - 2s^2 t + 4s^3 t) ds \\
 &\quad + t \int_t^1 (3s - 8s^2 + 4s^3 - 2st + 4s^2 t) ds \\
 &= -\frac{t}{6} + \frac{t^2}{3} - \frac{t^3}{2} + t^4 - \frac{8}{15} t^5;
 \end{aligned}$$

$$\int_0^1 (3-2s-2t)^2 s^2 ds = \int_0^1 (9s^2 - 12s^3 - 12s^2 t + 4s^4 + 4s^2 t^2 + 8s^3 t) ds = \frac{4}{5} - 2t + \frac{4}{3} t^2.$$

Hence

$$\begin{aligned}
\int_0^1 \int_0^1 K_T^2(s, t) ds dt &= \int_0^1 \left\{ (1-2t)^2 \left(\frac{t^2}{3} - \frac{2}{3}t^3 + t^4 - \frac{8}{15}t^5 \right) \right. \\
&\quad \left. + 2t(1-2t) \left(-\frac{t}{6} + \frac{t^2}{3} - \frac{t^3}{2} + t^4 - \frac{8}{15}t^5 \right) \right\} dt \\
&\quad + \int_0^1 t^2 \left(\frac{4}{5} - 2t + \frac{4}{3}t^2 \right) dt \\
&= \int_0^1 (1-2t) \left(-\frac{2}{3}t^3 + \frac{4}{3}t^4 - \frac{8}{15}t^5 \right) dt + \frac{1}{10} \\
&= \int_0^1 \left(-\frac{2}{3}t^3 + \frac{8}{3}t^4 - \frac{48}{15}t^5 + \frac{16}{15}t^6 \right) dt + \frac{1}{30} \\
&= -\frac{1}{6} + \frac{16}{15 \cdot 7} + \frac{1}{30} = \frac{2}{105}
\end{aligned}$$

Therefore, we obtain $\int_0^1 \int_0^1 K_T^2(s, t) ds dt = \frac{2}{105}$ and thus $\text{Var } T = \frac{4}{105}$.

For Chapter 3, Subsection 3.1.2.

Consider $\lambda f(t) = \int_0^1 \frac{K_T(s,t)}{\sqrt{s(1-s)t(1-t)}} f(s) ds$ and find λ .

$$\int_0^1 \frac{K_T(s,t)}{\sqrt{s(1-s)t(1-t)}} f(s) ds = \int_0^1 \frac{(1-2s)(1-2t)(s\Delta t) + (3-2s-2t)st}{\sqrt{s(1-s)t(1-t)}} f(s) ds.$$

Put $f(t) = \sqrt{t(1-t)} g(t)$ for some $g \in L^2[0,1]$, then

$$\begin{aligned} \lambda t(1-t)g(t) &= (1-2t) \int_0^t (1-2s)sg(s)ds + (1-2t)t \int_t^1 (1-2s)g(s)ds \\ &\quad + t \int_0^1 (3-2s)sg(s)ds - 2t^2 \int_0^1 sg(s)ds. \end{aligned}$$

Analogously to what was shown in Subsection 3.1.2. when considering

$\lambda f(t) = \int_0^1 K_T(s,t)f(s)ds$ we now have

$$\begin{aligned} \frac{d}{dt} \{ \lambda t(1-t)g(t) \} &= -2 \int_0^t (1-2s)sg(s)ds + (1-4t) \int_t^1 (1-2s)g(s)ds + \int_0^1 (3-2s)sg(s)ds \\ &\quad - 4t \int_0^1 sg(s)ds, \end{aligned}$$

$$\frac{d^2}{dt^2} \{ \lambda t(1-t)g(t) \} = -(1-2t)^2 g(t) - 4 \int_t^1 (1-2s)g(s)ds - 4 \int_0^1 sg(s)ds \quad \text{and}$$

$$\frac{d^3}{dt^3} \{ \lambda t(1-t)g(t) \} = -(1-2t)^2 g'(t) + 8(1-2t)g(t).$$

Furthermore,

$$\{ t(1-t)g(t) \}' = (1-2t)g(t) + t(1-t)g'(t)$$

$$\{ t(1-t)g(t) \}'' = -2g(t) + 2(1-2t)g'(t) + t(1-t)g''(t)$$

$$\{ t(1-t)g(t) \}''' = -6g'(t) + 3(1-2t)g''(t) + t(1-t)g'''(t)$$

Hence, in order to solve the above integral equation one has to solve the following third order differential equation:

$$\lambda g'''(t) = \frac{3\lambda(1-2t)}{t(1-t)} g''(t) + \frac{6\lambda-(1-2t)^2}{t(1-t)} g'(t) + \frac{8(1-2t)}{t(1-t)} g(t).$$

For Chapter 3, Subsection 3.1.2.

Consider $\int_0^{k\pi} \frac{\sin^2 y}{y} dy$:

$$\int_0^{k\pi} \frac{\sin^2 y}{y} dy = \frac{1}{2} \int_0^{2k\pi} \frac{1 - \cos x}{x} dx, \text{ use } \sin^2 y = \frac{1}{2}(1 - \cos 2y) \text{ and}$$

substitute $x = 2y$

$$= \frac{1}{2} \int_0^{2k\pi} \sum_{n=1}^{\infty} \left\{ \frac{(-1)^{n+1}}{(2n)!} x^{2n-1} \right\} dx \text{ by Taylor series expansion of } \cos x$$

$$= \frac{1}{4} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \cdot \frac{(2k\pi)^{2n}}{(2n)!} \text{ since } \sum \text{ uniformly convergent over } \mathbb{R}.$$

Therefore, $\int_0^{k\pi} \frac{\sin^2 y}{y} dy = \frac{1}{4} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \frac{(2k\pi)^{2n}}{(2n)!} \text{ for each } k \geq 1.$

Consider $\int_0^1 \frac{\sin k\pi t \sin m\pi t}{t} dt$ for $k, m \geq 1$; $k \neq m$; k, m even or k, m odd:

Recall that $\cos(x+y) = \cos x \cos y - \sin x \sin y$ for each $x, y \in \mathbb{R}$

and $\cos(x-y) = \cos x \cos y + \sin x \sin y$ for each $x, y \in \mathbb{R}$.

Hence, subtracting the first of the above equations from the second yields

$$\sin x \sin y = \frac{1}{2}(\cos(x-y) - \cos(x+y)) \text{ for each } x, y \in \mathbb{R}.$$

Therefore,

$$\begin{aligned} \int_0^1 \frac{\sin k\pi t \sin m\pi t}{t} dt &= \frac{1}{2} \int_0^1 \frac{\cos(k-m)\pi t - \cos(k+m)\pi t}{t} dt \\ &= \frac{1}{2} \int_0^1 \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} \left[((k-m)\pi)^{2n} - ((k+m)\pi)^{2n} \right] t^{2n-1} dt \end{aligned}$$

$$= \frac{1}{4} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \cdot \frac{(k-m)^{2n} - (k+m)^{2n}}{(2n)!} \pi^{2n}$$

We conclude that

$$\int_0^1 \frac{\sin k\pi t \sin m\pi t}{t} dt = \frac{1}{4} \sum_{n=1}^{\infty} \frac{(-1)^n}{n} \cdot \frac{(k-m)^{2n} - (k+m)^{2n}}{(2n)!} \pi^{2n}$$

References

- Anderson, T.W. and Darling, D.A. (1954). "A test of goodness of fit",
JASA, 49, 765-769.
- Anderson, T.W. and Darling, D.A. (1952). "Asymptotic theory of
certain 'goodness of fit' criteria based on stochastic processes",
Ann. Mathem. Statist. 23, 193-212.
- Bahadur, R.R. (1966). "A note on quantiles in large samples", Ann.
Mathem. Statist. 37, 577-580.
- Bhattacharya, P.K. and Frierson, Dargan (1981). "A nonparametric
control chart for detecting small disorders", Ann. Statist. 9,
544-554.
- Billingsley, Patrick (1986). Probability and Measure, 2nd ed., Wiley,
New York.
- Billingsley, Patrick (1968). Convergence of Probability Measures,
Wiley, New York.
- Breiman, Leo (1968). Probability, Addison-Wesley, Reading, Massachusetts.
- Broadbent, Simon (1980). "Simulating the Ley Hunter", J.R. Statist.
Soc. A-143, 109-140.
- Chernoff, H. and Zacks, S. (1964). "Estimating the current mean of
a normal distribution which is subject to change in time", Ann.
Mathem. Statist. 35, 999-1018.
- Chung, C.F. (1987). Wiener Pack: A Subroutine Package for Computing
Probabilities Associated with Wiener and Brownian Bridge Processes,
Paper 87-12, Geological Survey of Canada, Ottawa, Ont.
- Csörgő, Miklós (1983). Quantile processes with statistical applications,
SIAM 42, Philadelphia, Pennsylvania.

- Csörgő, Miklós (1968). "On the strong law of large numbers and the central limit theorem for martingales", Trans. AMS 131, 259-275.
- Csörgő, Miklós, Csörgő, Sándor, Horváth, Lajos, and Mason, David M. (1986). "Weighted empirical and quantile processes", Ann. Prob. 14, 31-85.
- Csörgő, Miklós and Horváth, Lajos (1988a). "Invariance principles for changepoint problems", J. Mult. Anal., Memorial Volume for P.R. Krishnaiah, to appear.
- Csörgő, Miklós and Horváth, Lajos (1988b). "Nonparametric Methods for Changepoint Problems", Handbook of Statistics, Vol. 7, Elsevier Science Publishers B.V. (North-Holland), to appear.
- Csörgő, Miklós and Horváth, Lajos (1988c). "On the distributions of L_p norms of weighted uniform empirical and quantile processes", Ann. Prob. 16, 142-161.
- Csörgő, Miklós and Horváth, Lajos (1987). "Asymptotic distributions of pontograms", Math. Proc. Camb. Phil. Soc. 101, 131-139.
- Csörgő, Miklós and Horváth, Lajos (1987a). "Detecting change in a random sequence", J. Mult. Anal. 23, 119-130.
- Csörgő, Miklós and Horváth, Lajos (1987b). "Nonparametric tests for the changepoint problem", J. Statist. Plann. Inference 17, 1-9.
- Csörgő, Miklós and Horváth, Lajos (1986). "Approximations of weighted empirical and quantile processes", Statist. & Prob. Letters 4, 275-280.
- Csörgő, Miklós and Horváth, Lajos (1985). "On the distributions of L_1 and L_2 norms of weighted uniform empirical and quantile processes", Tech. Rep. Ser. Lab. Res. Statist. Probab., Carleton University, Ottawa, Ont., Canada, No. 60.

- Csörgő, Miklós, Horváth, Lajos and Steinebach, Josef (1987). "Invariance principles for renewal processes", Ann. Prob. 15, 1441-1460.
- Csörgő, Miklós and Révész, Pal (1981). Strong Approximations in Probability and Statistics, Academic Press, New York.
- Csörgő, Miklós and Révész, Pal (1978). "Strong approximations of the quantile process", Ann. Statist. 6, 882-894.
- De Acosta, A. (1982). "Invariance principles in probability for triangular arrays of B-valued random vectors and some applications", Ann. Prob. 10, 346-373.
- De Wet, T. (1987). "Degenerate U- and V-statistics", Preprint of invited review paper for South African Statist. J., to appear.
- De Wet, T. and Venter, J.H. (1973). "Asymptotic distributions for quadratic forms with applications to tests of fit", Ann. Statist. 1, 380-387.
- Doob, J.L. (1948). "Renewal theory from the point of view of the theory of probability", Trans. Amer. Mathem. Soc. 63, 422-438.
- Feller, William (1966). An Introduction to Probability Theory and its Applications, Vol. 2, Wiley, New York.
- Gardner, L.A. (1969). "On detecting changes in the mean of normal variates", Ann. Mathem. Statist. 40, 116-126.
- Gradshteyn, I.S. and Ryzhik, I.M. (1965). Tables of Integrals, Series and Products, Academic Press, New York.
- Grams, William F. and Serfling, Robert J. (1973). "Convergence rates for U-statistics and related statistics", Ann. Statist. 1, 153-160.

Gregory, Gavin G. (1977). "Large sample theory for U-statistics and tests of fit", Ann. Statist, 5, 110-123.

Hall, Peter (1979). "On the invariance principle for U-statistics", Stoch. Proc. Appl. 9, 163-174.

Heyde, C.C. (1967). "Asymptotic Renewal Results for a Natural Generalization of Classical Renewal Theory", J.R. Statist. Soc. B 29, 141-150.

Heyde, C.C. (1966). "Some renewal theorems with application to a first passage problem", Ann. Mathem. Statist. 37, 699-710.

Hogg, Robert V. and Craig, Allen T. (1978). Introduction to mathematical statistics, Macmillan, New York.

Huse, Vera R. (1988). "Asymptotic properties for the sequential CUSUM procedure", Statist. & Probab. Letters, to appear.

Imhof, J.P. (1961). "Computing the Distribution of Quadratic Forms in Normal Variables", Biometrika 48, 419-426.

James, Barry, James, Kang L. and Siegmund, David (1987). "Tests for a change-point", Biometrika 74, 71-83.

Janson, Svante and Wichura, Michael J. (1983). "Invariance principles for stochastic area and related stochastic integrals", Stoch. Proc. Appl. 16, 71-84.

Kamke, E. (1951). Differentialgleichungen, Lösungsmethoden und Lösungen, Band 1, 4. verb. Auflage, Akadem. Verlagsges., Leipzig, G.D.R.

Kander, Z. and Zacks, S. (1966). "Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points", Ann. Mathem. Statist. 37, 1196-1210.

Karlin, Samuel and Taylor, Howard M. (1975). A First Course in Stochastic Processes, 2nd ed., Academic Press, New York.

Kendall, David G. and Kendall, Wilfrid S. (1980). "Alignments in two-dimensional random sets of points", Adv. Appl. Prob. 12, 380-424.

Kiefer, J. (1970). "Deviations between the sample quantile process and the sample D.F.", Nonparam. Tech. in Statist. Inf., M. Puri, ed., Cambridge Univ. Press, London, 299-319.

Komlós, J., Major, P., and Tusnády, G. (1976). "An Approximation of Partial Sums of Independent RV'-s, and the Sample DF. II", Z. Wahrsch. verw. Geb. 34, 33-58.

Komlós, J., Major, P., and Tusnády, G. (1975). "An Approximation of Partial Sums of Independent RV'-s, and the Sample DF. I", Z. Wahrsch. verw. Geb. 32, 111-131.

Kotz, Samuel, Johnson, N.L. and Boyd, D.W. (1967). "Series representations of distributions of quadratic forms in normal variables I. Central case", Ann. Mathem. Statist. 38, 823-837.

Kreyszig, Erwin (1978). Introductory functional analysis with applications, Wiley, New York.

Krishnaiah, P.R. and Sen, P.K. (1984). Nonparametric Methods, Handbook of Statistics, Vol. 4, North-Holland, Amsterdam, the Netherlands.

Leithold, Louis (1986). The calculus of a single variable with analytic geometry, Harper & Row, New York.

Lombard, F. (1987). "Rank tests for changepoint problems", Biometrika, to appear.

Lombard, F. (1983). "Asymptotic distributions of rank statistics in the change-point problem", South African Statist. J. 17, 88-105.

MacNeill, Ian B. (1978). "Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times", *Ann. Statist.* 6, 422-433.

MacNeill, Ian B. (1974). "Tests for change of parameter at unknown times and distributions of some related functionals on Brownian motion", *Ann. Statist.* 2, 950-962.

MacNeill, Ian B. (1971). "Limit processes for co-spectral and quadrature spectral distribution functions", *Ann. Mathem. Statist.* 42, 81-96.

Matthews, D.E., Farewell, V.T., and Pyke, R. (1985). "Asymptotic score-statistic processes and tests for constant hazard against a change-point alternative", *Ann. Statist.* 13, 583-591.

Mizrahi, Abe and Sullivan, Michael (1982). *Calculus and Analytic Geometry*, Wadsworth, Belmont, California.

Page, E.S. (1955). "A test for a change in a parameter occurring at an unknown point", *Biometrika* 42, 523-526.

Page, E.S. (1954). "Continuous inspection schemes" *Biometrika* 41, 100-115.

Parent, E.A. (1965). "Sequential ranking procedures", *Doctoral Dissertation*, Stanford University, Stanford, California.

Pearson, E.S. and Hartley, H.O. (1976). *Biometrika Tables for Statisticians*, Vol. II, Lowe and Brydone, Norfolk, U.K.

Pettitt, A.N. (1979). "A nonparametric approach to the changepoint problem", *Appl. Statist.* 28, 126-135.

Randles, Ronald H. and Wolfe, Douglas A. (1979). *Introduction to the Theory of Nonparametric Statistics*, Wiley, New York.

- Robbins, Herbert and Siegmund, David (1970). "Boundary crossing probabilities for the Wiener process and sample sums", Ann. Mathem. Statist. 41, 1410-1429.
- Rothman, E.D. and Woodroffe, Michael (1972). "A Cramér-von Mises type statistic for testing symmetry", Ann. Mathem. Statist. 43, 2035-2038.
- Rubin, H. and Vitale, R.A. (1980). "Asymptotic distribution of symmetric statistics", Ann. Statist. 8, 165-170.
- Segen, Jakub and Sanderson, Arthur C. (1980). "Detecting Change in a Time Series", IEEE Transactions on Information Theory IT-26, 249-255.
- Sen, Ashish and Srivastava, Muni S. (1975). "On tests for detecting change in mean", Ann. Statist. 3, 98-108.
- Sen, Pranab Kumar (1977). "Almost sure convergence of generalized U-statistics", Ann. Prob. 5, 287-290.
- Sen, Pranab Kumar (1974). "Weak convergence of generalized U-statistics", Ann. Prob. 2, 90-102.
- Serfling, Robert J. (1980). Approximation Theorems of Mathematical Statistics, Wiley, New York.
- Shaban, S.A. (1980). "Change-point problem and two-phase regression: An annotated bibliography", Int. Statist. Rev. 48, 83-93.
- Shepp, L.A. (1966). "Radon-Nikodym derivatives of Gaussian measures", Ann. Math. Statist. 37, 321-354.
- Shorack, Galen R. and Wellner, Jon A. (1986). Empirical Processes with Applications to Statistics, Wiley, New York.
- Siegmund, David (1985). "Boundary crossing probabilities and statistical applications", Tech. Rep. No. 32, Department of Statistics, Stanford University, Stanford, California.

Steinebach, Josef (1986). "Improved Erdős-Rényi and strong approximation laws for increments of renewal processes", Ann. Probab. 14, 547-559.

Steinebach, Josef (1979). "Erdős-Rényi-Zuwächse bei Erneuerungsprozessen und Partialsummen auf Gittern", Habilitationsschrift, Universität Düsseldorf, Düsseldorf, F.R.G.

Steinebach, Josef (1978). "A strong law of Erdős-Rényi type for cumulative processes in renewal theory", J. Appl. Prob. 15, 96-111.

Taylor, Howard M. and Karlin, Samuel (1984). An introduction to stochastic modeling, Academic Press, Orlando, Florida.

Wolfe, Douglas A. and Schechtman, Edna (1984). "Nonparametric statistical procedures for the changepoint problem", J. Statist. Plann. Inference 9, 389-396.

Zacks, S. (1983). "Survey of classical and Bayesian approaches to the change-point problem: fixed sample and sequential procedures of testing and estimation", Recent Adv. Statist., 245-269.

END

06|07|89|

FIN