CZECH TECHNICAL UNIVERSITY

FACULTY OF CIVIL ENGINEERING

CHANGE-POINT DETECTION IN TEMPERATURE SERIES

PhD. Thesis

MONIKA RENCOVÁ

Supervisor: Prof. RNDr. Daniela Jarušková, CSc.

Department of Mathematics
Thákurova 7, Prague

2009

Abstract

This work is devoted to some change-point detection problems in temperature series. Results of this thesis are based on working with real data. The submitted work presents suggestions on how the change-point methods may be applied to detect changes in annual maximal, resp. minimal temperatures and to detect changes in occurrences of unusually hot, resp. cold days. Solving these practical examples we came across some theoretical problems, we tried to work out in this thesis. In the first problem we apply the change-point theory and we will be looking for a change in parameters in a large class of independent random variables with a GEV distribution not satisfying regularity conditions. In the second problem we will focus on dependent variables and show how the change-point theory might be extended from linear processes to strong-mixing sequences.

Acknowledgements

It is a real pleasure to express my thanks to Prof. RNDr. Daniela Jarušková, CSc., my thesis supervisor, who introduced me into a field of change-point detection, encouraged me and spent a lot of time in discussions on this topic. I would like to thank the Management and the Staff of Department of Mathematics for providing a good environment for my work on the thesis.

However, first of all I thank to my family for supporting me during my work, especially, I would like to appreciate my husband Jirka for his patience.

Contents

Chapter 1

Preface

This work is devoted to some change-point detection problems in temperature series. Results of this thesis are based on working with real data.

The broadly accepted hypothesis of global warming stimulated an interest in studying long temperature series. Some scientists assume that changes do not necessarily occur in the mean of the series but rather in some other characteristics, e.g. appearance of some extreme events or increase of difference between summer and winter temperatures etc. This raises an interest in studying statistical properties of extremes of random sequences, see e.g. Embrechts et al. [10], Leadbetter et al. [22]. Our paper presents suggestions on how the change-point methods may be applied to detect changes in annual maximal, resp. minimal temperatures and to detect changes in occurrences of unusually hot, resp. cold days. Solving these practical examples we came across some theoretical problems, we tried to work out in this thesis.

The world is filled with changes. We encounter them in economics, medicine, meteorology, climatology etc. A change-point analysis is a statistical method allowing to decide whether an observed stochastic process follows one model or whether the model changes. In the case of a change, we might be interested in following problems: when a change was detected and how many changes have occurred.

The change-point detection is formulated in terms of hypotheses testing. The null hypothesis claims that the series is stationary, usually it means that the parameters of the model do not change, while the alternative hypothesis claims that at an unknown time point the model changes. The decision rule for rejecting the null hypothesis is based on test statistics.

The earliest change-point studies go back to the 1950s, where they arose in the context of quality control. We observe an output of a manufacturing process and assume that a certain characteristic varies around a certain *in-control* constant a_0. Sometimes, for example due to a failure of the production device, this constant starts to vary around another *out-of-control* constant $a_1 \neq a_0$ and we want to know if and when such a change occurred. Statistical procedures in change-point analysis can be divided into two categories: "on-line" and "off-line" procedures. The "on-line" approach, coming from the manufacturing process, is based on the idea that after each observation we apply a new test and hope to be warned that the change occurred. In this thesis we will work with the "off-line" analysis when we already have all the observations and we apply a test for the whole data to decide whether and when the change occurred.

This historically first change-point problem can be formulated as follows. For simplicity we assume that the starting value a_0 and variance σ^2 are known and the observations are independent and distributed according to the normal distribution. Moreover we standardize the observations and obtain variables $Y_i, i = 1 \ldots n$ with a zero mean and unit variance at the beginning. Then change-point problem formulated by hypotheses testing is:

$$\begin{aligned} &H : Y_i = e_i, && i = 1, \ldots, n, \\ &A : \text{there exists} \quad k \in \{0, \ldots, n-1\} \quad \text{such that} \\ &\quad Y_i = e_i, && i = 1, \ldots, k, \\ &\quad Y_i = a + e_i, && i = k+1, \ldots, n, \end{aligned} \tag{1.1}$$

where $a \neq 0$. Using likelihood ratio method we obtain a so-called maximum-type statistic

$$\max_{1 \le k \le n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\}. \tag{1.2}$$

However, in practice the distribution of this statistic is very complex, so that it can be computed only for small sample sizes. Therefore, for n large, the asymptotic behavior of the statistic (1.2) is of interest. The maximum-type statistic goes to infinity as $n \to \infty$ a.s., however we can approximate this statistic by a maximum of a standardized Wiener process satisfying

$$\left| \max_{1 \le k \le n} \frac{\sum_{i=1}^{n} |Y_i|}{\sqrt{k}} - \sup_{1/n \le t \le 1} \frac{|W(t)|}{\sqrt{t}} \right| = o_p\left(\frac{1}{\sqrt{2 \log\log n}} \right).$$

The approximate critical values can be calculated from the asymptotic behavior of the probabilities under H:

$$P\left(\max_{0 \le k \le n-1} \left\{ \left| \frac{1}{\sqrt{n-k}} \sum_{i=k+1}^{n} Y_i \right| \right\} > \frac{x + b_n}{a_n} \right) \approx 1 - \exp\left\{ -e^{-x} \right\}, \quad x \in \mathbb{R}, \tag{1.3}$$

where

$$\begin{aligned} a_n &= \sqrt{2 \log\, \log\, n}, \\ b_n &= 2 \log\, \log\, n + \frac{1}{2} \log\log\, \log\, n - \frac{1}{2} \log\, \pi. \end{aligned}$$

This approximation was derived by Darling and Erdös [8] in 1956.

However, it often happens that the starting value a_0 is unknown. In such case we test the following null hypothesis H against the alternative A:

$$\begin{aligned} &H : Y_i = a + e_i, && i = 1, \ldots, n, \\ &A : \text{there exists} \quad k \in \{0, \ldots, n-1\} \quad \text{such that} \\ &\quad Y_i = a + e_i, && i = 1, \ldots, k, \\ &\quad Y_i = a + \delta + e_i, && i = k+1, \ldots, n \end{aligned} \tag{1.4}$$

with $\delta \neq 0$. Again for i.i.d. random variables distributed according to the normal distribution $N(0, \sigma^2)$ with σ^2 known we obtain the maximum-type statistic of a form

$$\max_{1 \le k \le n-1} \left\{ \frac{1}{\sigma} \left| \sqrt{\frac{n}{k\,(n-k)}} \sum_{i=1}^{k} (Y_i - \overline{Y}_n) \right| \right\}. \tag{1.5}$$

For n large we may approximate the statistic (1.5) by the maximum of a standardized Brownian bridges

$$\left| \max_{1 \le k \le n-1} \frac{1}{\sigma} \sqrt{\frac{n}{k\,(n-k)}} \left| \sum_{i=1}^{k} (Y_i - \overline{Y}_n) \right| - \sup_{\frac{1}{n} \le t \le 1-\frac{1}{n}} \frac{|B(t)|}{\sqrt{t\,(1-t)}} \right| = o_p \left(\frac{1}{\sqrt{2\,\log\log n}} \right).$$

Yao and Davis [28] proved similar approximation as in (1.3) for a sequence of independent normal variables. For $x \in \mathbb{R}$ it holds

$$P\left(\max_{1 \le k \le n-1} \left\{ \frac{1}{\sigma} \left| \frac{n}{\sqrt{k\,(n-k)}} \sum_{i=1}^{k} \left(Y_i - \overline{Y}_n\right) \right| \right\} > \frac{x + b_n}{a_n} \right) \approx 1 - \exp\left\{ -2\, e^{-x} \right\}.$$

For a quite extensive survey on change-point detection we refer to Csörgő– Horváth [7]. They used the log-likelihood ratio for the general model working with a sequence of independent random vectors $X_1, X_2, \ldots, X_n$ with distribution functions $F(x; \theta_1), \ldots, F(x; \theta_n)$, respectively, where $\theta_i \in \Theta \subseteq \mathbb{R}^d$ for $i = 1, \ldots, n$ are parameters of the distribution functions and are assumed to change at unknown time. The general problem tested in Csörgő–Horváth [7] has a form:

$$\begin{aligned} &H : \varphi_1 = \varphi_2 = \ldots = \varphi_n \\ &A : \text{there exists} \quad k \in \{0, \ldots, n-1\} \quad \text{such that} \\ &\qquad \varphi_1 = \ldots = \varphi_k \neq \varphi_{k+1} = \ldots = \varphi_n. \end{aligned}$$

The asymptotics of a testing statistic

$$\max_{1 \le k \le n} 2 \log(\Lambda_k)$$

is under null hypothesis again

$$\lim_{n \to \infty} P\left(A(\log n) \left(\max_{0 \le k \le n-1} 2 \log(\Lambda_k) \right)^{1/2} \le t + D_d(\log n) \right) = \exp(-2e^{-t})$$

for all $t \in \mathbb{R}$, where

$$A(x) = \sqrt{2\,\log x}$$

and

$$D_d(x) = 2 \log x + (d/2) \log\log x - \log \Gamma(d/2),$$

where $\Gamma(t)$ is the gamma function defined

$$\Gamma(t) = \int_0^\infty y^{t-1} \exp(-y)\,dy.$$

In Chapter 4 of Csörgő– Horváth [7] it is also shown that the limit results remain true for a large class of dependent observations.

Many articles have been published on change-point detection, see e.g. Antoch et al. [1], Antoch J. and Hušková M. [2], Gombay E. and Horváth, L. [11]. For application in climatology we refer to publications of Jarušková [18] and Jandhyala [**?**].

In this thesis we will apply the change-point methods for two special cases. In the first problem we apply the change-point theory and we will be looking for a change in parameters in a large class of independent random variables with the GEV distribution. In the second problem we will focus on dependent variables and show how the change-point theory might be extended from linear processes to strong-mixing sequences. The material is divided into chapters, sections and paragraphs.

The studied data are presented in the second chapter. We summarize the origin of the data sets and provide some statistical characteristics of the observations.

The results of the third chapter were obtained while solving change-point detection in annual maximal and minimal temperatures. We are looking for a change in parameters in a large class of random variables with the GEV distribution. The general Theorem 1.3.1. presented by Csörgő– Horváth [7], confer Appendix–Theorem A.1.1, can not be applied here directly, as extremal distributions do not satisfy conditions on regularity. Since the density function $h(x;\mu,\psi,\xi)$ is defined on the set $\{x; 1+\xi(x-\mu)/\psi > 0\}$, the classical regularity conditions for maximum likelihood estimators are not satisfied. The next problem is caused by the conditions C.4 and C.5 of Theorem A.1.1, since they require the continuity of third derivatives. This can be weakened by Smith's theorem, see Appendix–Theorem A.3.1, and we will show that for $\xi > -\frac{1}{2}$ there exists a sequence $(\widehat{\mu}_n, \widehat{\psi}_n, \widehat{\xi}_n)$ of solutions of the likelihood equations such that $\sqrt{n}\,(\widehat{\mu}_n - \mu_0, \widehat{\psi}_n - \psi_0, \widehat{\xi}_n - \xi_0)$ converge in distribution to a zero mean normal vector with a variance−covariance matrix $\mathbf{M}^{-1}$ ($\mathbf{M}$ is a Fisher information matrix) and hence $\xi > -\frac{1}{2}$ is still a regular case. From here an idea comes that the assertion of Theorem A.1.1 is still valid. The results on temperature series are presented at the end of the chapter.

Many articles have been published on independent observations. Clearly, working with real temperature series, we can not expect that the condition of independency is fulfilled. This problem we encountered solving the second example with occurrences of unusually hot or cold days. While for the annual maximal and minimal temperatures from the first example we might assume that the data form i.i.d. sequence, for occurrences of unusually hot or cold days we have a strong correlation between the temperature values measured at subsequent days, the value of correlation coefficient is for all series very close to 0.8. In the fourth chapter we define the problem by a model working with data forming strong-mixing processes. We generalize the theory for dependent data presented in Csörgő and Horváth [7], and show that the asymptotic distribution of the testing statistic $T_n(t)$ is valid not only for linear processes but for strong-mixing sequences as well. In the context, it is an important question how to estimate σ^2. We can replace σ^2 with an estimator, where the rate of convergency to σ^2 must be at least $o_p((\log\log n)^{-1})$, which is fulfilled by:

$$\widehat{\sigma^2} = \widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i), \tag{1.6}$$

where $\widehat{R}(j) = \frac{1}{n}\sum_{i=1}^{n-j}\left(Y_i - \overline{Y_n}\right)\left(Y_{i+j} - \overline{Y_n}\right)$, $\overline{Y_n} = \frac{1}{n}\sum_{1\le j\le n} Y_j$ and $\psi(n)$ tends to infinity with a certain speed. The estimator $\widehat{\sigma^2}$ is a simplified version of the Bartlett log window estimator

$$\sigma_n^2(L) = \hat{R}(0) + 2\sum_{i=1}^{L}\left(1 - \frac{i}{L}\right)\hat{R}(i).$$

For more information about this estimator we refer to Antoch et al. [1].

It is wide known that the rate of convergency to distributional asymptotics under the null hypothesis is very slow, therefore in the fifth chapter we propose a permutation principle for obtaining the corresponding critical values. We generalize the theory presented by Kirch [20] from linear processes to strong-mixing processes. We show that the estimator of variance

$$\widehat{\sigma}^2_{LK} = \frac{1}{KL}\sum_{l=0}^{L-1}\left[\sum_{k=1}^{K}(Y(Kl+k) - \overline{Y}_n)\right]^2,$$

where $\overline{Y_n} = \frac{1}{n}\sum_{1\le j\le n} Y_j$, satisfies necessary condition on the rate of convergency. Critical values obtained from our data using the permutation test are listed at the end of the chapter.

In Appendix we summarize the known theory for extremal distributions, the theorems concerning change-point analysis for i.i.d. data, Smith's theorem and some results on strong-mixing sequences and rank statistics.

Chapter 2

Data

The data sets we have studied were taken from a CD-ROM that was a part of the book edited by Camuffo and Jones [6]. The book sums up results of EU research project IMPROVE. One of the main goals of the project was to produce seven highly reliable daily series (Brussels, Cadiz, Milan, Padua, St. Petersburg, Stockholm, Uppsala), extending over more than two centuries, by correcting errors and inhomogeneities caused by changes in measurement style etc. We add one more data set, which attracted our attention the most – Prague temperature series obtained from *http://eca.knmi.ne*. Of course, statisticians enjoy analyzing such long natural series but the length of the series also brings problems. Temperature often started to be measured at famous universities which are now mostly situated in city centers with their "heat island effect". The climatologists who analyzed the Milan and Stockholm series tried to remove this effect by comparing the series with the measurements taken in nearby observatories, while the authors of the other series were not able to do it. This is not the only reason why the properties of the studied series are difficult to compare. The other reason is that the way in which the daily averages were calculated differs from place to place.

In spite of the effort of the climatologists participating in the project, the series are not complete. Table 1 and 2 show the periods of measurement.

	period	missing data	number of obser.
Brussels	1795 – 1998	none	204
Cadiz	1817 – 2000	1851 – 1852; 1989 – 2000	170
Milan	1763 – 1998	none	236
Padua	1777 – 1992	1850; 1855; 1865; 1907; 1947	212
St. Petersburg	1744 – 1996	1745 – 1751; 1763 – 1766; 1787 – 1788; 1793; 1795 – 1797; 1800 – 1805; 1846	229
Stockholm	1756 – 2000	none	245
Uppsala	1725 – 2000	none	266
Prague	1775 – 2004	none	230

Table 1. Periods of observations together with missing data for annual minima.

	period	missing data	number of obser.
Brussels	1795 – 1998	none	204
Cadiz	1817 – 2000	1852; 1873; 1964; 1988; 1990; 1992 – 1994	176
Milan	1763 – 1998	none	236
Padua	1777 – 1997	1912; 1921 –1922; 1947; 1954; 1994; 1996	214
St. Petersburg	1744 – 1996	1745 – 1753; 1784; 1787; 1793; 1795 – 1797; 1800 – 1804	235
Stockholm	1756 – 2000	none	245
Uppsala	1725 – 2000	none	266
Prague	1775 – 2004	none	230

Table 2. Periods of observations together with missing data for annual maxima.

The following tables summarize basic descriptive statistics of the data.

		maxima	
	$\overline{x}$	σ_{n-1}	sk
Brussels	23.60	1.92	0.17
Cadiz	29.36	1.44	0.10
Milan	28.09	1.33	0.26
Padua	27.71	1.28	0.42
St. Petersburg	23.80	1.82	-0.01
Stockholm	22.39	1.94	0.11
Uppsala	22.13	1.95	0.10
Prague	26.15	1.71	0.26

Table 3. Descriptive statistics for annual maximal temperatures.

		minima	
	$\overline{x}$	σ_{n-1}	sk
Brussels	-6.80	2.39	-0.24
Cadiz	5.63	4.87	-0.89
Milan	-4.43	3.09	-0.61
Padua	-3.94	3.08	-0.79
St. Petersburg	-23.09	2.87	-0.18
Stockholm	-15.02	2.81	-0.43
Uppsala	-17.81	2.53	-0.06
Prague	-11.66	4.84	-0.38

Table 4. Descriptive statistics for annual minimal temperatures.

Comparing Tables 3 and 4 we can see that skewness for minimal temperatures is negative and its absolute values are larger, while skewness for maximal temperatures is positive with smaller absolute values. It might suggest that maximal temperatures might be modelled by the normal distribution. For minimal temperatures, the three parameter

Weibull distribution fits better, see Rencová [23].
The following figures show the behavior of the series under study.

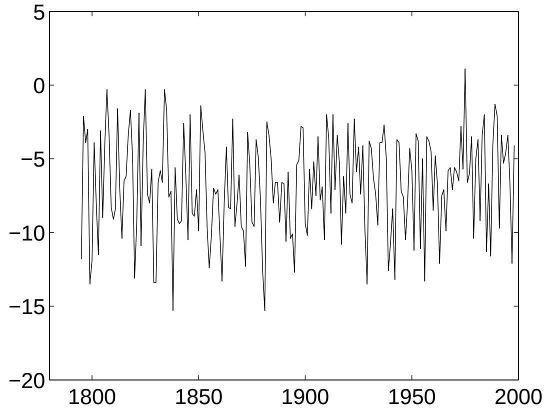

Figure 1. Annual minimal temperatures (in °C) in Brussels.

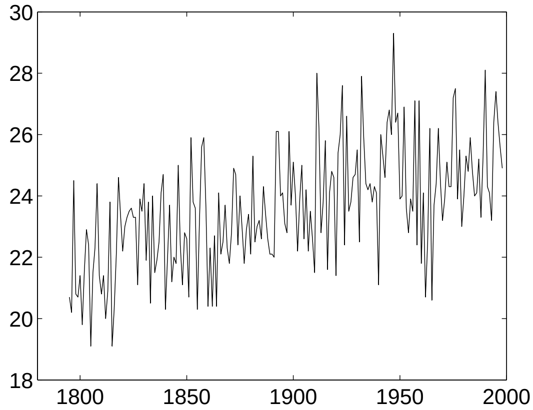

Figure 2. Annual maximal temperatures (in °C) in Brussels.

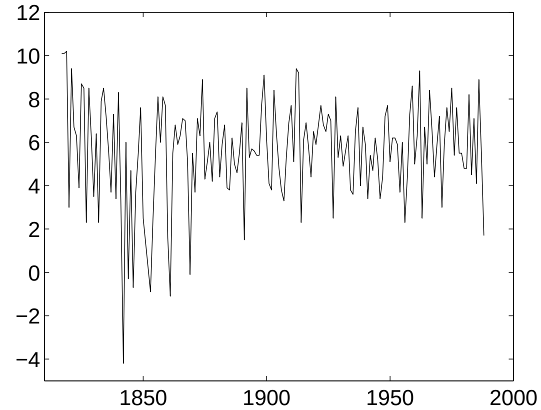

Figure 3. Annual minimal temperatures (in °C) in Cadiz.

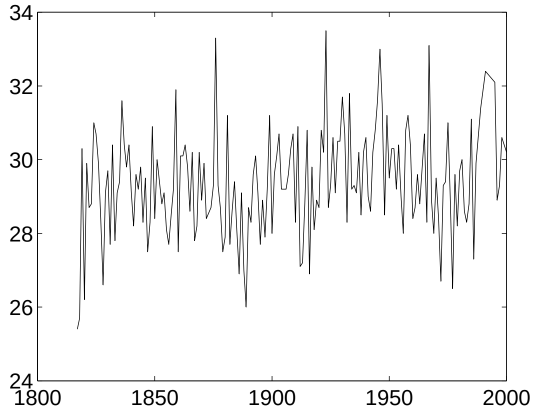

Figure 4. Annual maximal temperatures (in °C) in Cadiz.

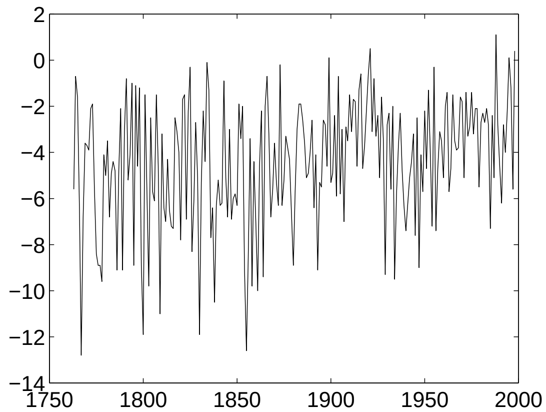

Figure 5. Annual minimal temperatures (in °C) in Milan.

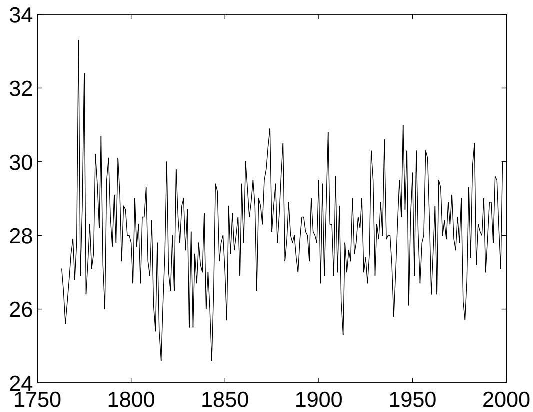

Figure 6. Annual maximal temperatures (in °C) in Milan.

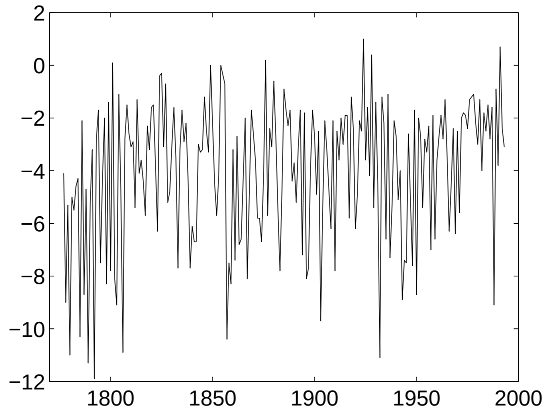

Figure 7. Annual minimal temperatures (in °C) in Padua.

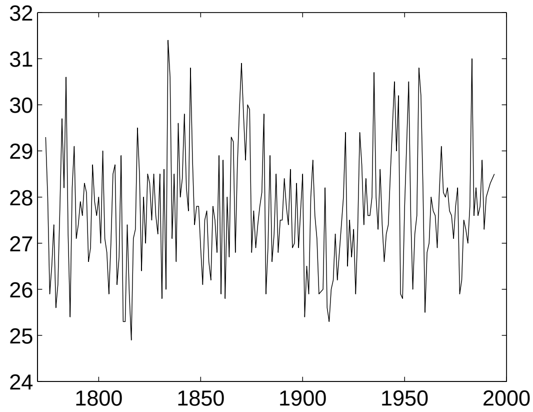

Figure 8. Annual maximal temperatures (in °C) in Padua.

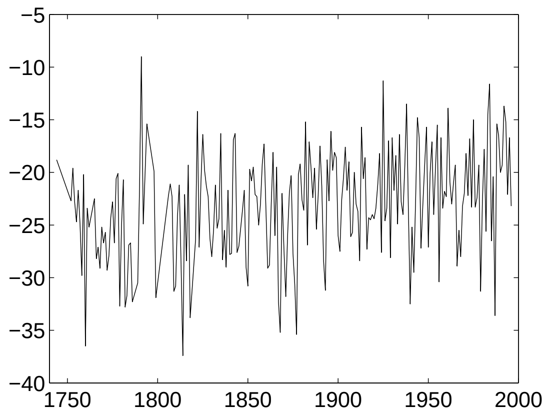

Figure 9. Annual minimal temperatures (in °C) in St. Petersburg.

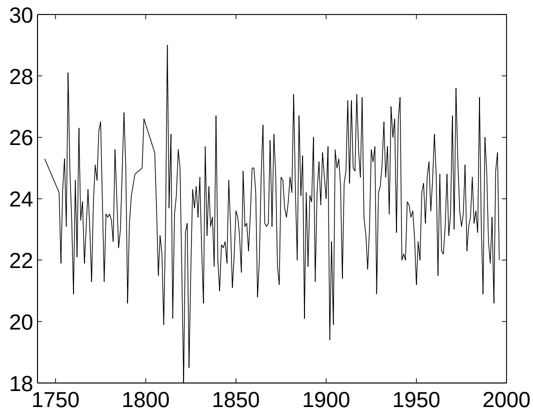

Figure 10. Annual maximal temperatures (in °C) in St. Petersburg.

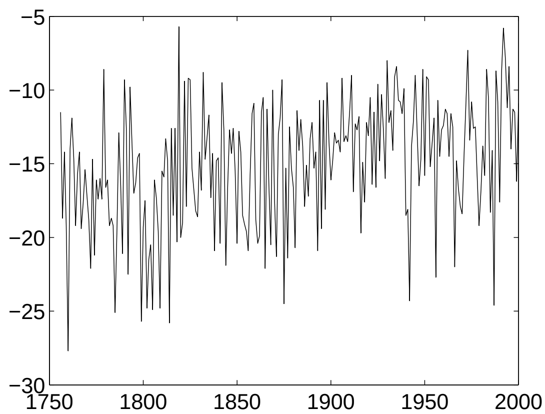

Figure 11. Annual minimal temperatures (in °C) in Stockholm.

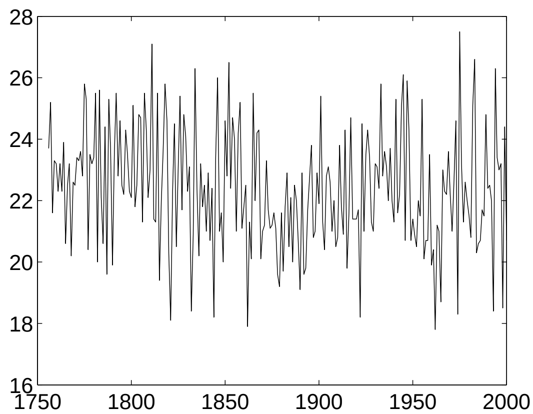

Figure 12. Annual maximal temperatures (in °C) in Stockholm.

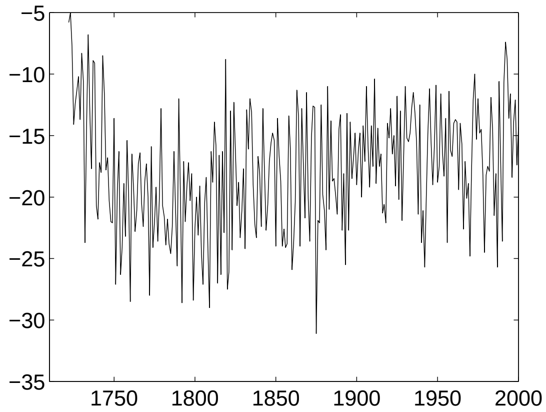

Figure 13. Annual minimal temperatures (in °C) in Uppsala.

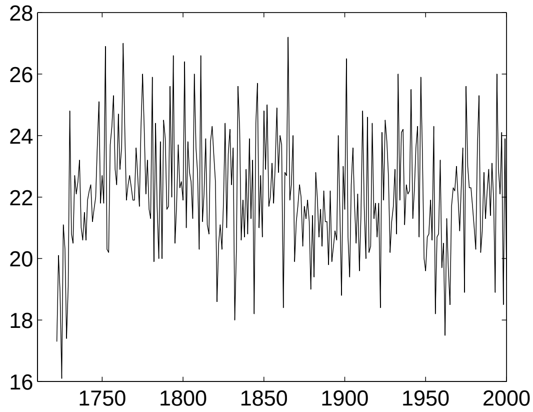

Figure 14. Annual maximal temperatures (in °C) in Uppsala.

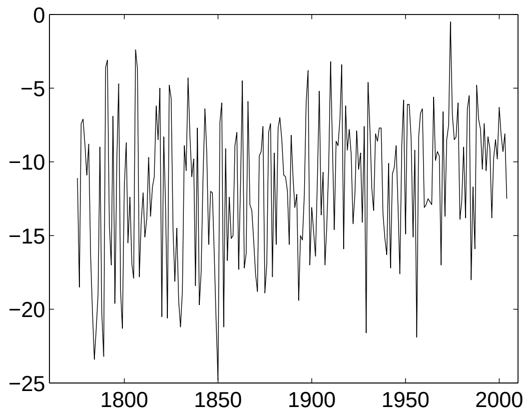

Figure 15. Annual minimal temperatures (in °C) in Prague.

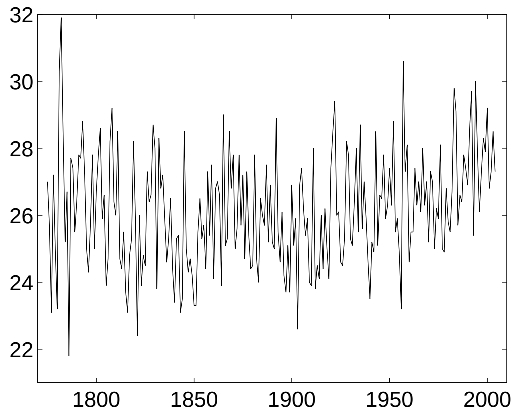

Figure 16. Annual maximal temperatures (in °C) in Prague.

Figures 17−32 describe the second problem - changes in occurrences of unusually hot, resp. cold days. We provide graphs of sums of exceedances over a chosen level $h = 2.5$ and under a chosen level $c = -2.5$ for standardized daily series of studied data sets, see Section 4.2.

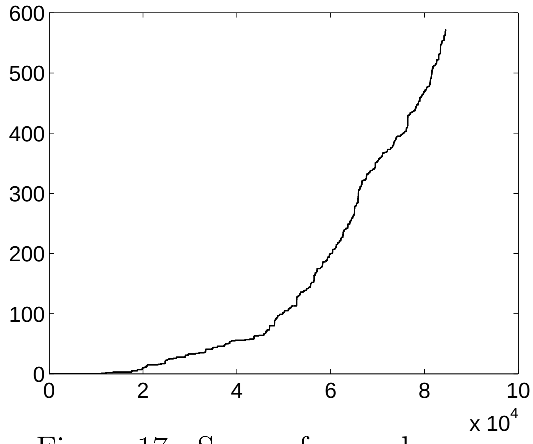

Figure 17. Sums of exceedances over a chosen level $h = 2.5$ in Brussels.

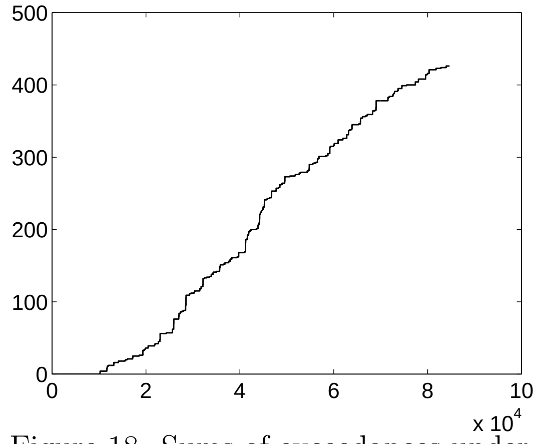

Figure 18. Sums of exceedances under a chosen level $c = -2.5$ in Brussels.

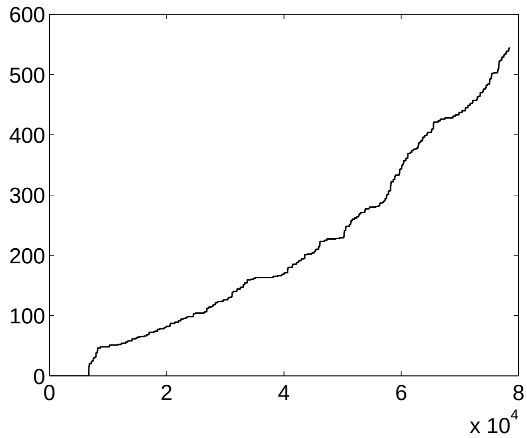

Figure 19. Sums of exceedances over a chosen level $h = 2.5$ in Cadiz.

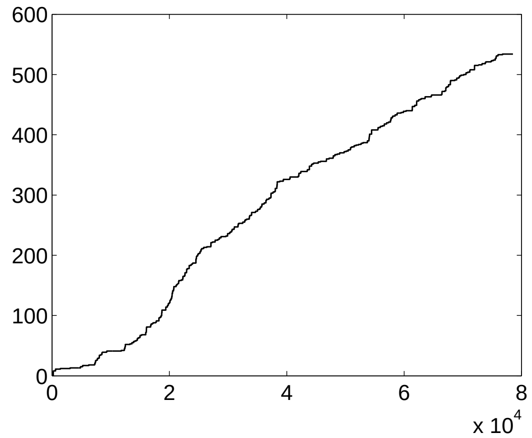

Figure 20. Sums of exceedances under a chosen level $c = -2.5$ in Cadiz.

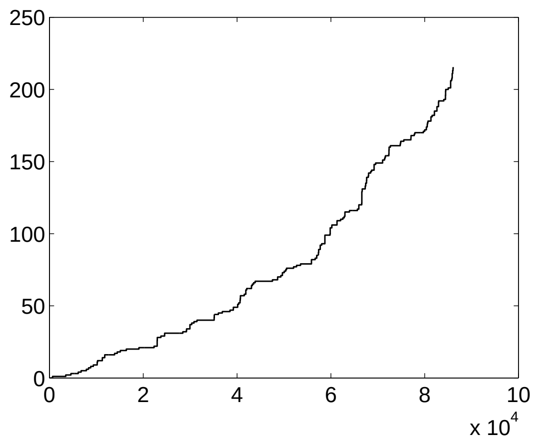

Figure 21. Sums of exceedances over a chosen level $h = 2.5$ in Milan.

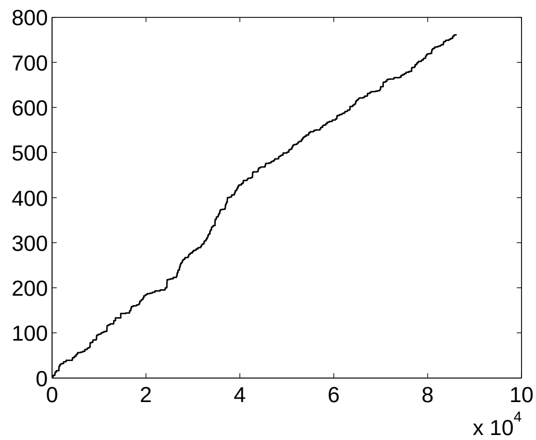

Figure 22. Sums of exceedances under a chosen level $c = -2.5$ in Milan.

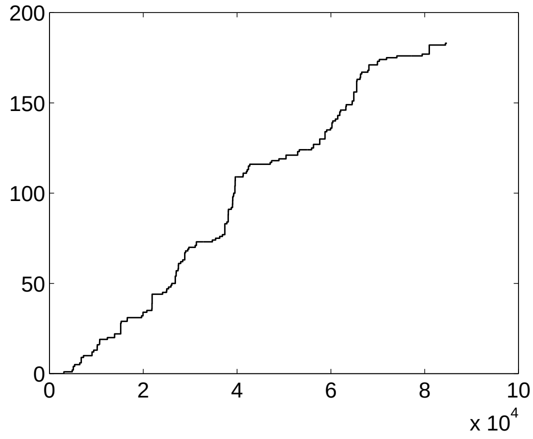

Figure 23. Sums of exceedances over a chosen level $h = 2.5$ in Padua.

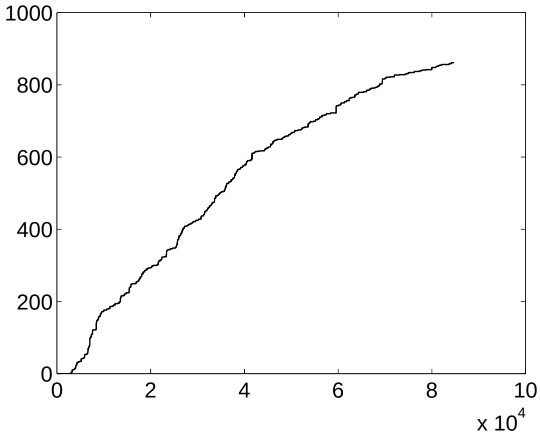

Figure 24. Sums of exceedances under a chosen level $c = -2.5$ in Padua.

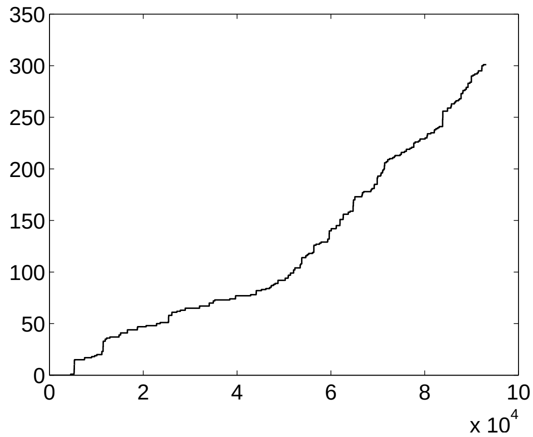

Figure 25. Sums of exceedances over a chosen level $h = 2.5$ in St. Petersburg.

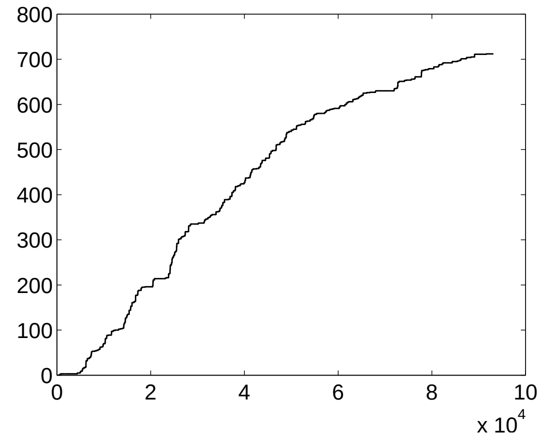

Figure 26. Sums of exceedances under a chosen level $c = -2.5$ in St. Petersburg.

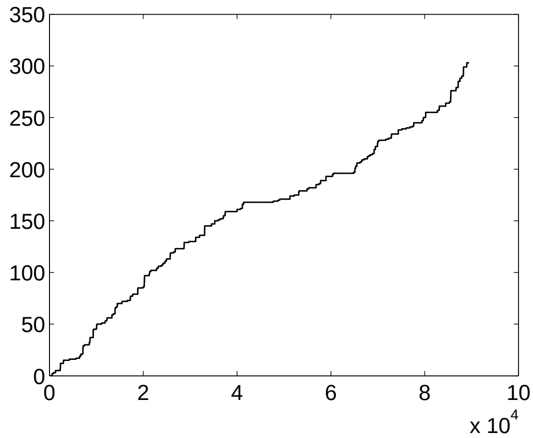

Figure 27. Sums of exceedances over a chosen level $h = 2.5$ in Stockholm.

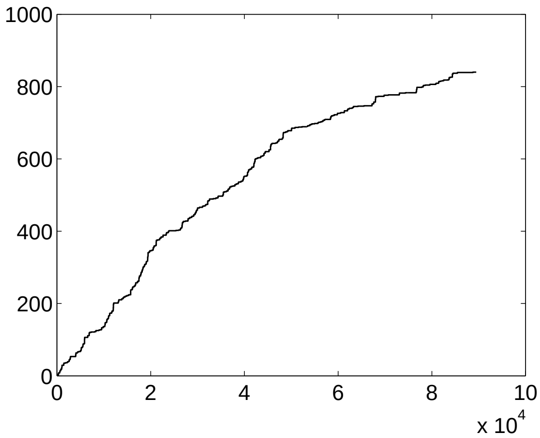

Figure 28. Sums of exceedances under a chosen level $c = -2.5$ in Stockholm.

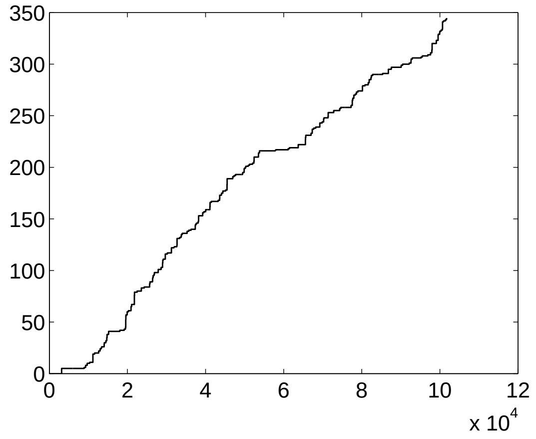

Figure 29. Sums of exceedances over a chosen level $h = 2.5$ in Uppsala.

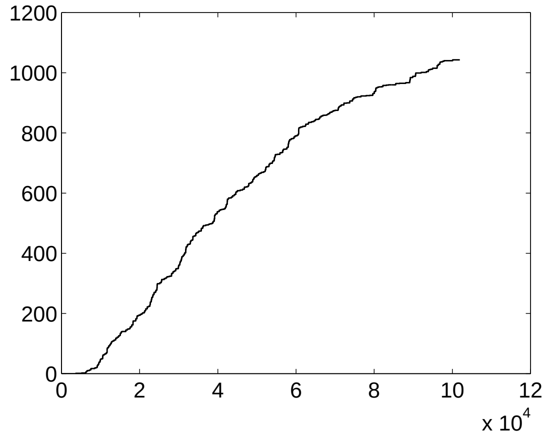

Figure 30. Sums of exceedances under a chosen level $c = -2.5$ in Uppsala.

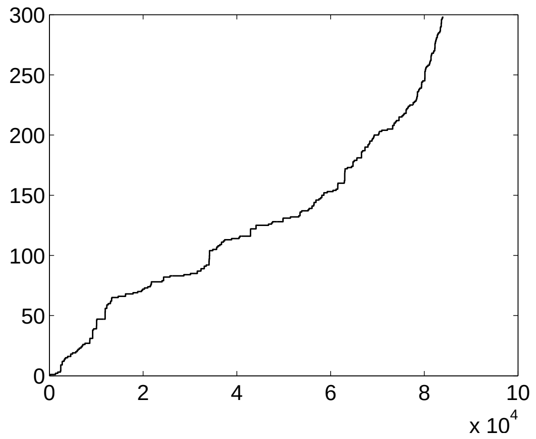

Figure 31. Sums of exceedances over a chosen level $h = 2.5$ in Prague.

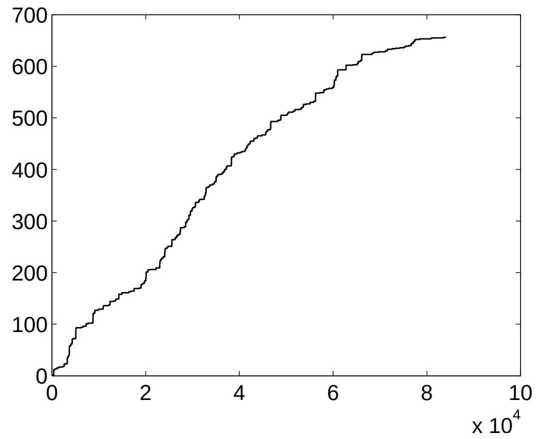

Figure 32. Sums of exceedances under a chosen level $c = -2.5$ in Prague.

Figure 33 shows that there is a strong correlation between the temperature values measured at subsequent days, its value is for all series very close to 0.8, see the upper stationary graph in Figure 33. The following graphs, from the top to the bottom, depict correlation coefficients between two days with lag equal to 2, 3 and 4, e.g. the first value in the upper graph is the value of the correlation coefficient between 1^{st} January and 2^{nd} January, the first value in the second graph from the top is the value of the correlation coefficient between 1^{st} January and 3^{rd} January etc.

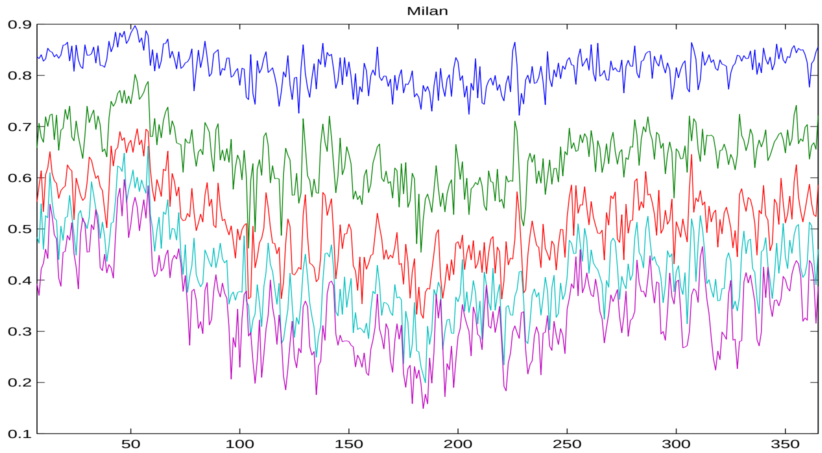

Figure 33. The autocorrelation coefficients between subsequent days in Milan.

We can notice that the autocorrelation coefficients for lag equal to 1 oscillate about the value 0.8 during the whole year, while the autocorrelation coefficients for larger lags are smaller in summer than in winter, see Figure 33.

Problem 1

Application of change-point detection for annual maxima and minima

Chapter 3

The change-point detection for the GEV distributions

In the first part of the thesis we study annual maximal and minimal temperatures. Figures 1 − 16 and Tables 3, 4 show behavior of the series under study.

Extremes of random sequences are modelled by the GEV distribution, for details confer Appendix, Section A.2. Our goal is applying the general Csörgő and Horváth theory for detecting a sudden change (Appendix, Section A.1) of the GEV distribution $H(x;\mu,\psi,\xi)$ with a density function

$$h(x;\mu,\psi,\xi) = \frac{1}{\psi}\left(1+\xi\,\frac{x-\mu}{\psi}\right)^{-\frac{1}{\xi}-1} \exp\left\{-\left(1+\xi\,\frac{x-\mu}{\psi}\right)^{-\frac{1}{\xi}}\right\}, \tag{3.1}$$

provided $1+\xi(x-\mu)/\psi > 0$. Notice that the support of the density function $h(x;\mu,\psi,\xi)$ depends on the parameters μ,ψ,ξ. Suppose that $X_1,\dots,X_n$ are independent random variables, we are to test the null hypothesis H_0 against the alternative A_1:

$$\begin{aligned}
&H_0 : X_i \sim GEV\,(\mu_0,\psi_0,\xi_0), && i=1,\dots,n, \\
&A_1 : \text{there exists} \quad k \in \{\,0,\dots,n-n_0\} \quad \text{such that} && \\
&\qquad X_i \sim GEV\,(\mu_0,\psi_0,\xi_0), && i=1,\dots,k, \\
&\qquad X_i \sim GEV\,(\mu,\psi,\xi), && i=k+1,\dots,n,
\end{aligned} \tag{3.2}$$

where the parameters (μ_0,ψ_0,ξ_0) before the change point are known while $(\mu,\psi,\xi) \neq (\mu_0,\psi_0,\xi_0)$ are unknown or to test the null hypothesis H_0 against the alternative A_2:

$$\begin{aligned}
&A_2 : \text{there exists} \quad k \in \{\,n_0,\dots,n-n_0\} \quad \text{such that} && \\
&\qquad X_i \sim GEV\,(\mu_1,\psi_1,\xi_1), && i=1,\dots,k, \\
&\qquad X_i \sim GEV\,(\mu_2,\psi_2,\xi_2), && i=k+1,\dots,n,
\end{aligned} \tag{3.3}$$

where neither the parameters before nor after the change point are known and $\big((\mu_1,\psi_1,\xi_1) \neq (\mu_2,\psi_2,\xi_2)\big)$. The constant n_0 may be any fixed integer larger than three. However, to obtain a good estimates of all three parameters we need to have enough observations.

For testing the problem (3.3) we may use the twice log–likelihood ratio

$$\max_{0\le k\le n-1} 2\log(\Lambda_k) = 2\left[L_k(\widehat{\varphi}_k) + L_k^*(\varphi_k^*) - L_n(\widehat{\varphi}_n)\right],$$

while testing the problem (3.2) yields in a simplified version

$$\max_{0\le k\le n-1}\left(2\log(\Lambda_k^{(0)})\right) = \max_{0\le k\le n-1} 2\left[L_k^*(\widehat{\varphi}_k^*) - L_k^*(\varphi_0)\right].$$

for more details we refer to Appendix, Remark A.1.2.

To find critical values we have to find distribution of the test statistics under H_0. As the exact distribution of $\max_{0\le k\le n-1} 2\log(\Lambda_k)$, resp. $\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})$, under H_0 are very complex, the approximate critical values can be found using the limit behavior of $\max_{0\le k\le n-1} 2\log(\Lambda_k)$, resp. $\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})$, see Appendix, Csörgő – Horváth theorem. The conditions of Theorem A.1.1 are classical regularity conditions for the existence of the maximum likelihood estimator. Since the density function (3.1) is defined on the set $\{x; 1+\xi(x-\mu)/\psi > 0\}$, the classical regularity conditions for maximum likelihood estimators are not satisfied. The next problem is caused by the conditions C.4 and C.5 of Theorem A.1.1, since they require the continuity of third derivatives. This can be weakened by Smith's Theorem A.3.1 and we will show that for $\xi > -\frac{1}{2}$ there exists a sequence $(\widehat{\mu}_n, \widehat{\psi}_n, \widehat{\xi}_n)$ of solutions of the likelihood equations such that $\sqrt{n}\,(\widehat{\mu}_n - \mu_0, \widehat{\psi}_n - \psi_0, \widehat{\xi}_n - \xi_0)$ converge in distribution to a zero mean normal vector with a variance – covariance matrix $\mathbf{M}^{-1}$ ($\mathbf{M}$ is a Fisher information matrix) and hence $\xi > -\frac{1}{2}$ is still a regular case.

We will proceed in two steps. At first we show theory for the three parameter Weibull distributions using the results of Smith's theorem, see Appendix–Theorem A.3.1, and then we will focus on the Fréchet distributions. For those purposes we can use the following reparameterization of the GEV distribution.

For $\xi < 0$, substituting $\theta = \mu - \frac{\psi}{\xi}, \beta = \left(-\frac{\xi}{\psi}\right)^{-\frac{1}{\xi}}, \alpha = -\frac{1}{\xi}$ we obtain

$$\begin{aligned} h(x;\theta,\alpha,\beta) &= \alpha\beta(-x+\theta)^{\alpha-1}\exp\{-\beta(-x+\theta)^{\alpha}\} && \text{for} \quad -x \ge -\theta, \\ &= 0 && \text{for} \quad -x < -\theta. \end{aligned} \tag{3.4}$$

It is the three parameter Weibull distribution $Weib(\theta, \alpha, \beta)$ of a random variable $-x$ (not the Weibull distribution as a limit distribution for maxima concentrated on $(-\infty, -\theta)$ from the Fisher – Tippet theorem, see Appendix–Theorem A.2.1.)

For $\xi > 0$, substituting $\theta = \mu - \frac{\psi}{\xi}, \beta = \left(\frac{\xi}{\psi}\right)^{-\frac{1}{\xi}}, \alpha = \frac{1}{\xi}$, we obtain a reparameterization

$$\begin{aligned} h(x;\theta,\alpha,\beta) &= \alpha\beta(x-\theta)^{-\alpha-1}\exp\{-\beta(x-\theta)^{-\alpha}\} && \text{for} \quad x \ge \theta, \\ &= 0 && \text{for} \quad x < \theta, \end{aligned} \tag{3.5}$$

which corresponds to the Fréchet distribution $Fr\acute{e}ch\,(\theta, \alpha, \beta)$.

Remark 3.0.1. *Results for the Gumbel distribution from Fisher – Tippett theorem A.2.1 corresponding to a case* $\xi = 0$ *are obtained by* $\xi \to 0$ *in* (3.1).

3.1 The change-point detection for the Weibull distributions

The general theory presented in Csörgő and Horváth [7] was applied by Jandhyala et al. [?] to develop a test for detecting a sudden change in the two parameter Weibull distribution. However, in the case we use as a model the three parameter Weibull distribution $Weib\,(\theta, \alpha, \beta)$ with a density function

$$\begin{aligned} f(x;\theta,\alpha,\beta) &= (x-\theta)^{\alpha-1}\alpha\beta\exp\{-\beta(x-\theta)^{\alpha}\} && \text{for} \quad x \geq \theta, \\ &= 0 && \text{for} \quad x < \theta, \end{aligned} \tag{3.6}$$

it seems more natural to look for a change in all three parameters. Notice that the support of the density function f is given by the parameter θ. Suppose that $X_1, \ldots, X_n$ are independent random variables, we are to test the null hypothesis H_0 against the alternative A_1:

$$\begin{aligned} &H_0 : X_i \sim Weib\,(\theta_0, \alpha_0, \beta_0), && i = 1, \ldots, n, \\ &A_1 : \text{there exists} \quad k \in \{0, \ldots, n-n_0\} \quad \text{such that} && \\ &\qquad X_i \sim Weib\,(\theta_0, \alpha_0, \beta_0), && i = 1, \ldots, k, \\ &\qquad X_i \sim Weib\,(\theta, \alpha, \beta), && i = k+1, \ldots, n, \end{aligned} \tag{3.7}$$

where the parameters $(\theta_0, \alpha_0, \beta_0)$ before the change point are known while $(\theta, \alpha, \beta) \neq (\theta_0, \alpha_0, \beta_0)$ are unknown or to test the null hypothesis H_0 against the alternative A_2:

$$\begin{aligned} &A_2 : \text{there exists} \quad k \in \{n_0, \ldots, n-n_0\} \quad \text{such that} && \\ &\qquad X_i \sim Weib\,(\theta_1, \alpha_1, \beta_1), && i = 1, \ldots, k, \\ &\qquad X_i \sim Weib\,(\theta_2, \alpha_2, \beta_2), && i = k+1, \ldots, n, \end{aligned} \tag{3.8}$$

where neither the parameters before nor after the change point are known and $(\theta_1, \alpha_1, \beta_1) \neq (\theta_2, \alpha_2, \beta_2)$. The constant n_0 may be any fixed integer larger than three.

For testing the problem (3.8) we may use the twice log–likelihood ratio

$$\max_{0 \leq k \leq n-1} 2\log(\Lambda_k) = 2\left[L_k(\widehat{\varphi}_k) + L_k^*(\varphi_k^*) - L_n(\widehat{\varphi}_n)\right],$$

while testing the problem (3.7) yields in a simplified version

$$\max_{0 \leq k \leq n-1} \left(2\log(\Lambda_k^{(0)})\right) = \max_{0 \leq k \leq n-1} 2\left[L_k^*(\widehat{\varphi}_k^*) - L_k^*(\varphi_0)\right],$$

for more details we refer to Appendix. As the exact distribution of $\max_{0\le k\le n-1} 2\log(\Lambda_k)$, resp. $\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})$, under H_0 are very complex, the approximate critical values can be found using the limit behavior of test statistics $\max_{0\le k\le n-1} 2\log(\Lambda_k)$, resp. $\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})$, see Appendix, Csörgő – Horváth Theorem A.1.1. Let $(\theta_0, \alpha_0, \beta_0)$ be the true values of the parameters under H_0. We will assume that $\theta_0 \in R^1$, $\alpha_0 > 2$ and $\beta_0 > 0$. The assumptions C.4. and C.5. of Csörgő and Horváth are not satisfied and Theorem A.1.1 cannot be applied directly to get a limit distribution of $\max_{0\le k\le n-1} 2\log(\Lambda_k)$, resp. $\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})$. On the other hand, Smith showed, confer Appendix–Theorem A.3.1, that for $\alpha_0 > 2$ there exists a sequence $(\widehat{\theta}_n, \widehat{\alpha}_n, \widehat{\beta}_n)$ of solutions of the likelihood equations such that $\sqrt{n}\,(\widehat{\theta}_n - \theta_0, \widehat{\alpha}_n - \alpha_0, \widehat{\beta}_n - \beta_0)$ converge in distribution to a zero mean normal vector with a variance – covariance matrix $\mathbf{M}^{-1}$ ($\mathbf{M}$ is a Fisher information matrix) and hence $\alpha_0 > 2$ is still a regular case. From here an idea comes that the assertion of Csörgő and Horváth theorem A.1.1 is still valid for $\alpha_0 > 2$.

3.2 Main results for the Weibull distributions

Our main results concern the asymptotic distribution of the statistic $\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})$ under H_0 for testing a change in all three parameters, when the parameters before a change point are known while after it they are unknown as well as the statistic $\max_{0\le k\le n-1} 2\log(\Lambda_k)$ under H_0 for testing a change in all three parameters, when the parameters both before and after a change point are unknown.

We start with the characteristics of the log–likelihood function L_k. The first and second derivatives of $L_k(\theta, \alpha, \beta)$ may be expressed as follows:

$$\frac{\partial L_k}{\partial \theta} = \sum_{i=1}^{k} \left[-\frac{(\alpha-1)}{(X_i-\theta)} + \alpha\beta(X_i-\theta)^{\alpha-1} \right], \tag{3.9}$$

$$\frac{\partial L_k}{\partial \alpha} = \sum_{i=1}^{k} \left[\log(X_i-\theta) + \frac{1}{\alpha} - \beta(X_i-\theta)^{\alpha}\log(X_i-\theta) \right], \tag{3.10}$$

$$\frac{\partial L_k}{\partial \beta} = \sum_{i=1}^{k} \left[\frac{1}{\beta} - (X_i-\theta)^{\alpha} \right], \tag{3.11}$$

$$\frac{\partial^2 L_k}{\partial \theta^2} = \sum_{i=1}^{k} \left[-\frac{(\alpha-1)}{(X_i-\theta)^2} - \alpha(\alpha-1)\beta(X_i-\theta)^{\alpha-2} \right], \tag{3.12}$$

$$\frac{\partial^2 L_k}{\partial \theta \partial \alpha} = \sum_{i=1}^{k} \left[-\frac{1}{(X_i-\theta)} + \beta(X_i-\theta)^{\alpha-1} + \alpha\beta(X_i-\theta)^{\alpha-1}\log(X_i-\theta) \right], \tag{3.13}$$

$$\frac{\partial^2 L_k}{\partial\theta\partial\beta} = \sum_{i=1}^{k}\left[\alpha(X_i-\theta)^{\alpha-1}\right], \tag{3.14}$$

$$\frac{\partial^2 L_k}{\partial\alpha^2} = \sum_{i=1}^{k}\left[-\frac{1}{\alpha^2}-\beta(X_i-\theta)^{\alpha}\log^2(X_i-\theta)\right], \tag{3.15}$$

$$\frac{\partial^2 L_k}{\partial\alpha\partial\beta} = \sum_{i=1}^{k}\left[-(X_i-\theta)^{\alpha}\log(X_i-\theta)\right], \tag{3.16}$$

$$\frac{\partial^2 L_k}{\partial\beta^2} = \sum_{i=1}^{k}\left[-\frac{1}{\beta^2}\right]. \tag{3.17}$$

For simplicity we denote the true value of parameter

$$\varphi_{\mathbf{0}} = (\theta_0, \alpha_0, \beta_0),$$

a maximum likelihood estimator based on $X_1, \ldots, X_k$ (when it exists) by

$$\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k).$$

It holds

$$E\Big(\frac{\partial}{\partial\theta}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big) = 0, \tag{3.18}$$
$$E\Big(\frac{\partial}{\partial\alpha}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big) = 0,$$
$$E\Big(\frac{\partial}{\partial\beta}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big) = 0.$$

Let's denote a Fisher information matrix $\mathbf{M}$ on a parameter $\varphi_{\mathbf{0}} = (\theta_0, \alpha_0, \beta_0)$ with elements

$$\mathbf{M} = \begin{pmatrix} m_{\theta\theta} & m_{\theta\alpha} & m_{\theta\beta} \\ m_{\alpha\theta} & m_{\alpha\alpha} & m_{\alpha\beta} \\ m_{\beta\theta} & m_{\beta\alpha} & m_{\beta\beta} \end{pmatrix},$$

where

$$\begin{aligned} m_{\theta\theta} &= E\{\frac{\partial}{\partial\theta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\theta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\} \\ &= -E\{\frac{\partial^2}{\partial\theta^2}\log(f(X_i;\,\varphi_{\mathbf{0}}))\}, \\ m_{\alpha\alpha} &= E\{\frac{\partial}{\partial\alpha}\log(f(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\alpha}\log(f(X_i;\,\varphi_{\mathbf{0}}))\} \\ &= -E\{\frac{\partial^2}{\partial\alpha^2}\log(f(X_i;\,\varphi_{\mathbf{0}}))\}, \\ m_{\beta\beta} &= E\{\frac{\partial}{\partial\beta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\} \\ &= -E\{\frac{\partial^2}{\partial\beta^2}\log(f(X_i;\,\varphi_{\mathbf{0}}))\}, \end{aligned}$$

$$\begin{aligned}
m_{\theta\alpha} = m_{\alpha\theta} &= E\{\frac{\partial}{\partial\theta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\alpha}\log(f(X_i;\,\varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\theta\partial\alpha}\log(f(X_i;\,\varphi_{\mathbf{0}}))\}, \\
m_{\theta\beta} = m_{\beta\theta} &= E\{\frac{\partial}{\partial\theta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\theta\partial\beta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\}, \\
m_{\alpha\beta} = m_{\beta\alpha} &= E\{\frac{\partial}{\partial\alpha}\log(f(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\alpha\partial\beta}\log(f(X_i;\,\varphi_{\mathbf{0}}))\}.
\end{aligned} \tag{3.19}$$

A maximum likelihood estimator $\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)$ satisfies

$$\frac{\partial L_k}{\partial\theta}(\widehat{\varphi}_{\mathbf{k}}) = 0, \quad \frac{\partial L_k}{\partial\alpha}(\widehat{\varphi}_{\mathbf{k}}) = 0, \quad \frac{\partial L_k}{\partial\beta}(\widehat{\varphi}_{\mathbf{k}}) = 0. \tag{3.20}$$

The existence of $\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)$ is guaranteed for $\alpha > 2$ by Theorem A.3.1.

If we consider positive moments only, then

$$E\Big(\frac{\partial}{\partial\theta}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^r < \infty \quad \text{if and only if} \quad r < \alpha$$

and for the same r we also have

$$\begin{aligned}
E\Big(\frac{\partial}{\partial\alpha}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^r &< \infty, \\
E\Big(\frac{\partial}{\partial\beta}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^r &< \infty.
\end{aligned}$$

For the second derivatives

$$E\Big(\frac{\partial^2}{\partial\theta^2}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^s < \infty \quad \text{if and only if} \quad 2\,s < \alpha$$

and for the same s is also

$$\begin{aligned}
E\Big(\frac{\partial^2}{\partial\theta\partial\alpha}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\theta\partial\beta}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\alpha^2}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\alpha\partial\beta}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\beta^2}\big(\log f(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty.
\end{aligned}$$

M is a positive definite matrix. According to the Marcinkiewicz-Zygmund law, (Appendix - Theorem A.3.2), for any τ such that $0 < \tau < 1 - 2/\alpha$

$$\begin{aligned}
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\theta^2}L_k(\varphi_{\mathbf{0}}) + m_{\theta\theta}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\theta\partial\alpha}L_k(\varphi_{\mathbf{0}}) + m_{\theta\alpha}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\theta\partial\beta}L_k(\varphi_{\mathbf{0}}) + m_{\theta\beta}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\alpha^2}L_k(\varphi_{\mathbf{0}}) + m_{\alpha\alpha}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\alpha\partial\beta}L_k(\varphi_{\mathbf{0}}) + m_{\alpha\beta}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\beta^2}L_k(\varphi_{\mathbf{0}}) + m_{\beta\beta}\Big) &= 0 \quad a.s.
\end{aligned} \tag{3.21}$$

By the law of the iterated logarithm, confer Appendix–Theorem A.3.3

$$\begin{aligned}
\limsup_{k\to\infty} \frac{\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}})}{\sqrt{k\log\log k}} &= O(1) \quad a.s.,\\
\limsup_{k\to\infty} \frac{\frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}})}{\sqrt{k\log\log k}} &= O(1) \quad a.s.,\\
\limsup_{k\to\infty} \frac{\frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})}{\sqrt{k\log\log k}} &= O(1) \quad a.s.
\end{aligned} \tag{3.22}$$

We start with several technical lemmas on the three parameter Weibull distribution $Weib\,(\theta, \alpha, \beta)$.

Lemma 3.2.1. *Let $X_i \sim Weib\,(\theta, \alpha, \beta)$, then*
a) for $\theta \in \mathbb{R}$, $\beta > 0$ and $\alpha > 3$

$$E\,\frac{1}{X_i - \theta} < \infty, \quad E\,\frac{1}{(X_i - \theta)^2} < \infty, \quad E\,\frac{1}{(X_i - \theta)^3} < \infty$$

b) for $\theta \in \mathbb{R}$, $\beta > 0$ and $2 < \alpha \le 3$

$$E\,\frac{1}{X_i - \theta} < \infty, \quad E\,\frac{1}{(X_i - \theta)^2} < \infty,$$

$$\sum_{i=1}^{k} \frac{1}{(X_i - \theta)^3} = o\big(k^{3/\alpha}(\log k)^{3(1/\alpha+\Delta)}\big) \quad a.s.\ for\ some \quad \Delta > 0. \tag{3.23}$$

Proof. See Feller [12]. □

Lemma 3.2.2. *Let $X_1, X_2, \dots, X_n$ are i.i.d. random variables, $X_i \sim Weib(\theta, \alpha, \beta)$. Under H_0 for any $\Delta > 0$, the minimum X_{k1} and the second minimum X_{k2} satisfy*

$$\frac{1}{\Big(1 - \frac{X_{k1}-\theta}{X_{k2}-\theta}\Big)} = o\Big(\log k\Big)^{1+\Delta} \quad a.s. \tag{3.24}$$

Proof. See Jarušková [18]. □

Further, we prove the following lemma.

We denote for any $\delta_k > 0$

$$I_{\delta_k} = \Big\{\widehat{\theta} \in \mathbb{R}, \widehat{\alpha} \in \mathbb{R}, \widehat{\beta} \in \mathbb{R}; |\widehat{\theta} - \theta_0| < \delta_k, |\widehat{\alpha} - \alpha_0| < \delta_k, |\widehat{\beta} - \beta_0| < \delta_k\Big\}.$$

Lemma 3.2.3. *For any sequence $\{\delta_k\}$ satisfying $\delta_k\, k^{1/\alpha+\delta} \to 0$ for some $\delta > 0$ and for any τ such that $0 < \tau < 1 - 2/\alpha$*

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\left|\frac{\partial^2}{\partial\theta^2}L_k(\widehat{\varphi}) - \frac{\partial^2}{\partial\theta^2}L_k(\varphi_{\mathbf{0}})\right|\right) = 0 \quad a.s., \tag{3.25}$$

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\left|\frac{\partial^2}{\partial\theta\partial\alpha}L_k(\widehat{\varphi}) - \frac{\partial^2}{\partial\theta\partial\alpha}L_k(\varphi_{\mathbf{0}})\right|\right) = 0 \quad a.s., \tag{3.26}$$

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\left|\frac{\partial^2}{\partial\theta\partial\beta}L_k(\widehat{\varphi}) - \frac{\partial^2}{\partial\theta\partial\beta}L_k(\varphi_{\mathbf{0}})\right|\right) = 0 \quad a.s., \tag{3.27}$$

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\left|\frac{\partial^2}{\partial\alpha^2}L_k(\widehat{\varphi}) - \frac{\partial^2}{\partial\alpha^2}L_k(\varphi_{\mathbf{0}})\right|\right) = 0 \quad a.s., \tag{3.28}$$

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\left|\frac{\partial^2}{\partial\alpha\partial\beta}L_k(\widehat{\varphi}) - \frac{\partial^2}{\partial\alpha\partial\beta}L_k(\varphi_{\mathbf{0}})\right|\right) = 0 \quad a.s., \tag{3.29}$$

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\left|\frac{\partial^2}{\partial\beta^2}L_k(\widehat{\varphi}) - \frac{\partial^2}{\partial\beta^2}L_k(\varphi_{\mathbf{0}})\right|\right) = 0 \quad a.s. \tag{3.30}$$

Proof. The proof is divided into two parts: the first part, rather lengthy, corresponds to the condition $2 < \alpha \leq 3$ and the second part corresponds to the condition $\alpha > 3$.

Let's suppose $2 < \alpha \leq 3$. We consider only terms in second derivatives, which are discontinuous at $X_i = \theta$.

We start with proving (3.25). Substituting (3.12) into (3.25) and concentrating on terms discontinuous at $X_i = \theta$ we obtain a following assertion to be examined:

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\sum_{i=1}^{k}\Big(\frac{1}{(X_i - \widehat{\theta})^2} - \frac{1}{(X_i - \theta_0)^2}\Big)\right) = 0 \quad a.s., \quad 0 < \tau < 1 - 2/\alpha. \tag{3.31}$$

The proof can be found in Jarušková [18], Lemma 3.

Next we look into the assertion (3.26). Substituting (3.13) into (3.26) we get following assertions to be proved

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\sum_{i=1}^{k}\Big(\frac{1}{X_i-\widehat{\theta}}-\frac{1}{X_i-\theta_0}\Big)\right)=0 \quad a.s. \tag{3.32}$$

and

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\sum_{i=1}^{k}\Big(\widehat{\alpha}\widehat{\beta}(X_i-\widehat{\theta})^{(\widehat{\alpha}-1)}\log(X_i-\widehat{\theta}) - \alpha_0\beta_0(X_i-\theta_0)^{(\alpha_0-1)}\log(X_i-\theta_0)\Big)\right)=0 \quad a.s. \tag{3.33}$$

We start with (3.32). Similarly as in Smith [26] and Jarušková [18] we write

$$\frac{1}{k}\sum_{i=1}^{k}\Big(\frac{1}{X_i-\widehat{\theta}}-\frac{1}{X_i-\theta_0}\Big)=\frac{1}{k(X_{k1}-\widehat{\theta})}-\frac{1}{k(X_{k1}-\theta_0)} + \frac{1}{k}\sum_{i=2}^{k}\Big(\frac{1}{X_{ki}-\widehat{\theta}}-\frac{1}{X_{ki}-\theta_0}\Big). \tag{3.34}$$

First, let X has the three parameter Weibull distribution with the density function (3.6). Then $Z=(X-\theta)\,\beta^{1/\alpha}$ has the Weibull distribution with parameters $\theta=0$, $\beta=1$ and α. Random variable $Y=\frac{1}{Z}$ has the density

$$\begin{aligned} f(y) &= (1/y)^{\alpha+1}\alpha\exp\{-(1/y)^{\alpha}\} \quad \text{for} \quad y\ge 0,\\ &=0 \quad \text{for} \quad y<0, \end{aligned} \tag{3.35}$$

with a finite moment $\mathsf{E}Y^r<\infty$ for any $r<\alpha$. Then, according to Theorem A.3.4, we get $1/\left(k^{1/r}(X_{k1}-\theta_0)\right)\to 0 \quad a.s.$ for any $r<\alpha$, i.e. the second term on the right side of (3.34) satisfies $1/\left(k\left(X_{k1}-\theta_0\right)\right)=o(k^{-1+\frac{1}{r}}) \quad a.s.$ for any $r<\alpha$.

Second, choose r' satisfying $1/\alpha<1/r'<1/\alpha+\delta$, then we can write the first term on the right side of (3.34) as follows:

$$\frac{1}{X_{k1}-\widehat{\theta}}=\frac{1}{(X_{k1}-\theta_0)\left(1-\frac{\widehat{\theta}-\theta_0}{(X_{k1}-\theta_0)}\right)}.$$

Recall that $\delta_k=o(\frac{1}{k^{1/\alpha+\delta}})$ then

$$\sup_{|\widehat{\theta}-\theta_0|<\delta_k}\frac{|\widehat{\theta}-\theta_0|}{(X_{k1}-\theta_0)}=o(k^{1/r'}/k^{1/\alpha+\delta})=o(1) \quad a.s. \tag{3.36}$$

In the other words, for two first terms in (3.34)

$$\sup_{|\widehat{\theta}-\theta_0|<\delta_k}\left(\frac{1}{k(X_{k1}-\widehat{\theta})}-\frac{1}{k(X_{k1}-\theta_0)}\right)=o(k^{-1+\frac{1}{r}})\quad a.s. \tag{3.37}$$

for any $r<\alpha$.

Further, for the third term in (3.34), again as in Smith [26] and Jarušková [18], using the Taylor expansion, there exists $\tilde{\theta}$ that $|\tilde{\theta}-\theta_0|<|\widehat{\theta}-\theta_0|$ and

$$\frac{1}{k}\sum_{i=2}^{k}\Big(\frac{1}{X_{ki}-\widehat{\theta}}-\frac{1}{X_{ki}-\theta_0}\Big)=\frac{|\widehat{\theta}-\theta_0|}{k}\sum_{i=2}^{k}\frac{1}{(X_{ki}-\tilde{\theta})^2}. \tag{3.38}$$

We have to distinguish between two cases: i) $\theta_0-\hat{\theta}>0$ and ii) $\theta_0-\hat{\theta}<0$.

We start with i). For $\tilde{\theta}$ satisfying $\hat{\theta}<\tilde{\theta}<\theta_0$ we have in (3.38)

$$\frac{|\widehat{\theta}-\theta_0|}{k}\sum_{i=2}^{k}\frac{1}{(X_{ki}-\tilde{\theta})^2}\leq\frac{|\widehat{\theta}-\theta_0|}{k}\sum_{i=2}^{k}\frac{1}{(X_{ki}-\theta_0)^2}.$$

Applying Lemma 3.2.1 for the right side of the above inequality, we obtain

$$\sup_{I_{\delta_k}}\frac{1}{k}\sum_{i=2}^{k}\Big(\frac{1}{X_i-\widehat{\theta}}-\frac{1}{X_i-\theta_0}\Big)=o(k^{-(1/\alpha+\delta)})\,O(1)=o(k^{-\upsilon})\quad a.s.$$

for any $0<\upsilon<\frac{1}{\alpha}$.

For the proof of ii) we use a characteristic of the first minimum that $X_{ki}-\tilde{\theta}\leq X_{ki}-X_{k1}$ for every $i=1,\ldots,n$. We obtain following inequalities

$$\begin{aligned}\frac{|\widehat{\theta}-\theta_0|}{k}\sum_{i=2}^{k}\frac{1}{(X_{ki}-\tilde{\theta})^2}&\leq\frac{|\widehat{\theta}-\theta_0|}{k}\sum_{i=2}^{k}\frac{1}{(X_{ki}-X_{k1})^2}\\&=\frac{|\widehat{\theta}-\theta_0|}{k}\sum_{i=2}^{k}\frac{1}{\Big((X_{ki}-\theta_0)-(X_{k1}-\theta_0)\Big)^2}\\&=\frac{|\widehat{\theta}-\theta_0|}{k}\sum_{i=2}^{k}\frac{1}{(X_{ki}-\theta_0)^2\Big(1-\frac{(X_{k1}-\theta_0)}{(X_{ki}-\theta_0)}\Big)^2}\\&\leq|\widehat{\theta}-\theta_0|\frac{1}{\big(1-\frac{X_{k1}-\theta_0}{X_{k2}-\theta_0}\big)^2}\Big(\frac{1}{k}\sum_{i=2}^{k}\frac{1}{(X_{ki}-\theta_0)^2}\Big).\end{aligned}$$

Applying Lemma 3.2.1 and Lemma 3.2.2 for the factors on the right side of the above inequality, we obtain similarly as in the case i)

$$\begin{aligned}\sup_{I_{\delta_k}} \frac{1}{k}\sum_{i=2}^{k}\Big(\frac{1}{X_i-\widehat{\theta}}-\frac{1}{X_i-\theta_0}\Big) &= o(k^{-(1/\alpha+\delta)})\, o((\log k)^{2(1+\Delta)})\, O(1)\\ &= o(k^{-\upsilon}) \quad a.s.\end{aligned}$$

for any $0<\upsilon<\frac{1}{\alpha}$.

Since we assume $2<\alpha\leq 3$, we have $1-\frac{2}{\alpha}\leq\frac{1}{\alpha}$ implying for any $0<\tau<1-\frac{2}{\alpha}$ the rate of convergency of the third term of (3.34)

$$\sup_{I_{\delta_k}} \frac{1}{k}\sum_{i=2}^{k}\Big(\frac{1}{X_i-\widehat{\theta}}-\frac{1}{X_i-\theta_0}\Big) = o(k^{-\tau}) \quad a.s.$$

and with the result (3.37) we have the assertion of (3.32).

Now we prove the assertion (3.33). The Taylor expansion implies that there exists $(\tilde\theta,\tilde\alpha,\tilde\beta)$, such that $|\tilde\theta-\theta_0|<|\widehat\theta-\theta_0|, |\tilde\alpha-\alpha_0|<|\widehat\alpha-\alpha_0|, |\tilde\beta-\beta_0|<|\widehat\beta-\beta_0|$ and

$$\begin{aligned}&\frac{1}{k}\sum_{i=1}^{k}\Big(\widehat\alpha\widehat\beta(X_i-\widehat\theta)^{(\widehat\alpha-1)}\log(X_i-\widehat\theta)\\ &\qquad -\alpha_0\beta_0(X_i-\theta_0)^{(\alpha_0-1)}\log(X_i-\theta_0)\Big)\end{aligned}$$

$$\begin{aligned}=\frac{1}{k}\sum_{i=1}^{k}\Bigg(&\tilde\beta(X_i-\tilde\theta)^{(\tilde\alpha-1)}\Big(\log(X_i-\tilde\theta)\Big)(\widehat\alpha-\alpha_0)\\ &+\tilde\alpha\tilde\beta(X_i-\tilde\theta)^{(\tilde\alpha-1)}\Big(\log^2(X_i-\tilde\theta)\Big)(\widehat\alpha-\alpha_0)\\ &+\tilde\alpha(X_i-\tilde\theta)^{(\tilde\alpha-1)}\Big(\log(X_i-\tilde\theta)\Big)(\widehat\beta-\beta_0)\\ &+\tilde\alpha(\tilde\alpha-1)\tilde\beta(X_i-\tilde\theta)^{(\tilde\alpha-2)}\Big(\log(X_i-\tilde\theta)\Big)(\widehat\theta-\theta_0)\\ &+\tilde\alpha\tilde\beta(X_i-\tilde\theta)^{(\tilde\alpha-2)}(\widehat\theta-\theta_0)\Bigg).\end{aligned} \tag{3.39}$$

The first term on the right side of (3.39) can be rewritten

$$
\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k}\tilde{\beta}(X_i-\tilde{\theta})^{(\tilde{\alpha}-1)}\left(\log(X_i-\tilde{\theta})\right)(\widehat{\alpha}-\alpha_0)\\
&=\frac{1}{k}\sum_{i=1}^{k}\tilde{\beta}(X_i-\theta_0)^{(\tilde{\alpha}-1)}\log\left((X_i-\theta_0)(\frac{X_i-\tilde{\theta}}{X_i-\theta_0})\right)\left(\frac{X_i-\tilde{\theta}}{X_i-\theta_0}\right)^{(\tilde{\alpha}-1)}(\widehat{\alpha}-\alpha_0)\\
&=\frac{1}{k}\sum_{i=1}^{k}\left(\tilde{\beta}(X_i-\theta_0)^{(\tilde{\alpha}-1)}\left(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\right)^{\tilde{\alpha}-1}\log(X_i-\theta_0)\right.\\
&\quad\left.+\tilde{\beta}(X_i-\theta_0)^{(\tilde{\alpha}-1)}\left(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\right)^{\tilde{\alpha}-1}\log\left(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\right)\right)(\widehat{\alpha}-\alpha_0). \qquad (3.40)
\end{aligned}
$$

From inequality $\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}<\frac{\theta_0-\tilde{\theta}}{X_{k1}-\theta_0}$ for every $i=1,2.\dots$ and (3.36) we have

$$\sup_{|\widehat{\theta}-\theta_0|<\delta_k}\frac{|\widehat{\theta}-\theta_0|}{(X_i-\theta_0)}=o(1)\quad a.s.$$

Then

$$\sup_{|\widehat{\theta}-\theta_0|<\delta_k}\left(1+\frac{|\widehat{\theta}-\theta_0|}{(X_i-\theta_0)}\right)^{(\tilde{\alpha}-1)}=O(1)\quad a.s. \qquad (3.41)$$

Similarly

$$\log\left(1+\frac{|\widehat{\theta}-\theta_0|}{(X_i-\theta_0)}\right)\le\log\left(1+\frac{|\widehat{\theta}-\theta_0|}{(X_{k1}-\theta_0)}\right)\le K\frac{|\widehat{\theta}-\theta_0|}{(X_{k1}-\theta_0)}$$

for some $K\in\mathbb{R}$ a.s. and then

$$\sup_{|\widehat{\theta}-\theta_0|<\delta_k}\log\left(1+\frac{|\widehat{\theta}-\theta_0|}{(X_i-\theta_0)}\right)=o(1)\quad a.s. \qquad (3.42)$$

We can suppose that for sufficiently large k, is $\tilde{\alpha}>2$ since we have $\alpha_0>2$ and then

$$E(X_i-\theta_0)^{(\tilde{\alpha}-1)}\log(X_i-\theta_0)<\infty,\quad E(X_i-\theta_0)^{(\tilde{\alpha}-2)}<\infty.$$

Therefore the first term in (3.39) satisfies

$$\sup_{I_{\delta_k}}\frac{1}{k}\sum_{i=1}^{k}\tilde{\beta}(X_i-\tilde{\theta})^{(\tilde{\alpha}-1)}\left(\log(X_i-\tilde{\theta})\right)(\widehat{\alpha}-\alpha_0)=o(k^{-(\frac{1}{\alpha}+\delta)}).$$

The second term on the right side of (3.39) can be rewritten

$$
\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k}\tilde{\alpha}\tilde{\beta}(X_i-\tilde{\theta})^{(\tilde{\alpha}-1)}\left(\log^2(X_i-\tilde{\theta})\right)(\widehat{\alpha}-\alpha_0)\\
&=\frac{1}{k}\sum_{i=1}^{k}\tilde{\alpha}\tilde{\beta}(X_i-\theta_0)^{(\tilde{\alpha}-1)}\Big(\frac{X_i-\tilde{\theta}}{X_i-\theta_0}\Big)^{(\tilde{\alpha}-1)}\left(\log^2\left((X_i-\theta_0)(\frac{X_i-\tilde{\theta}}{X_i-\theta_0})\right)\right)(\widehat{\alpha}-\alpha_0)\\
&=\frac{1}{k}\sum_{i=1}^{k}\Bigg[\tilde{\alpha}\tilde{\beta}(X_i-\theta_0)^{(\tilde{\alpha}-1)}\Big(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\Big)^{\tilde{\alpha}-1}\log^2(X_i-\theta_0)\\
&\quad+\tilde{\alpha}\tilde{\beta}(X_i-\theta_0)^{(\tilde{\alpha}-1)}\Big(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\Big)^{\tilde{\alpha}-1}2\log(X_i-\theta_0)\log\left(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\right)\\
&\quad+\tilde{\alpha}\tilde{\beta}(X_i-\theta_0)^{(\tilde{\alpha}-1)}\Big(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\Big)^{\tilde{\alpha}-1}\log^2\left(1+\frac{\theta_0-\tilde{\theta}}{X_i-\theta_0}\right)\Bigg](\widehat{\alpha}-\alpha_0).
\end{aligned}
\tag{3.43}
$$

Using inequalities (3.41), (3.42) and a property

$$E(X_i-\theta_0)^{(\tilde{\alpha}-2)}\log^2(X_i-\theta_0)<\infty \quad \text{for} \quad \tilde{\alpha}>2$$

we get

$$\sup_{I_{\delta_k}}\frac{1}{k}\sum_{i=1}^{k}\tilde{\alpha}\tilde{\beta}(X_i-\tilde{\theta})^{(\tilde{\alpha}-1)}\left(\log^2(X_i-\tilde{\theta})\right)(\widehat{\alpha}-\alpha_0)=o(k^{-(\frac{1}{\alpha}+\delta)}).$$

Similarly we can rewrite the other terms on the right side of (3.39) for which, supposing again $\tilde{\alpha}>2$, it holds

$$
\begin{aligned}
E(X_i-\theta_0)^{(\tilde{\alpha}-1)}\log(X_i-\theta_0)&<\infty,\\
E(X_i-\theta_0)^{(\tilde{\alpha}-2)}\log(X_i-\theta_0)&<\infty,\\
E(X_i-\theta_0)^{(\tilde{\alpha}-2)}&<\infty,
\end{aligned}
\tag{3.44}
$$

therefore we have for all the terms in (3.39)

$$
\begin{aligned}
&\sup_{I_{\delta_k}}\frac{1}{k}\sum_{i=1}^{k}\left(\widehat{\alpha}\widehat{\beta}(X_i-\widehat{\theta})^{(\widehat{\alpha}-1)}\log(X_i-\widehat{\theta})-\alpha_0\beta_0(X_i-\theta_0)^{(\alpha_0-1)}\log(X_i-\theta_0)\right)\\
&\qquad=o(k^{-\upsilon})\quad a.s.
\end{aligned}
$$

for any $0<\upsilon<\frac{1}{\alpha}$.

Since we assume $2<\alpha\leq 3$, we have $1-\frac{2}{\alpha}\leq\frac{1}{\alpha}$ implying for any $0<\tau<1-\frac{2}{\alpha}$ the rate of convergency in (3.33)

$$
\begin{aligned}
&\sup_{I_{\delta_k}}\frac{1}{k}\sum_{i=1}^{k}\left(\widehat{\alpha}\widehat{\beta}(X_i-\widehat{\theta})^{(\widehat{\alpha}-1)}\log(X_i-\widehat{\theta})-\alpha_0\beta_0(X_i-\theta_0)^{(\alpha_0-1)}\log(X_i-\theta_0)\right)\\
&\qquad=o(k^{-\tau})\quad a.s.
\end{aligned}
$$

Now we investigate the assertion (3.28). Substituting (3.15) into (3.28) we have to prove

$$\lim_{k\to\infty} k^{\tau}\left(\sup_{I_{\delta_k}} \frac{1}{k}\sum_{i=1}^{k}\Big(-\widehat{\beta}\log^2(X_i-\widehat{\theta})(X_i-\widehat{\theta})^{\widehat{\alpha}} + \beta_0\log^2(X_i-\theta_0)(X_i-\theta_0)^{\alpha_0}\Big)\right)$$
$$= 0 \quad a.s. \tag{3.45}$$

From the Taylor expansion we get that there exists $(\tilde{\theta},\tilde{\alpha},\tilde{\beta})$, such that $|\tilde{\theta}-\theta_0| < |\widehat{\theta}-\theta_0|, |\tilde{\alpha}-\alpha_0| < |\widehat{\alpha}-\alpha_0|, |\tilde{\beta}-\beta_0| < |\widehat{\beta}-\beta_0|$ and

$$-\frac{1}{k}\sum_{i=1}^{k}\left(\widehat{\beta}(X_i-\widehat{\theta})^{\widehat{\alpha}}\log^2(X_i-\widehat{\theta}) - \beta_0(X_i-\theta_0)^{\alpha_0}\log^2(X_i-\theta_0)\right)$$
$$= -\frac{1}{k}\sum_{i=1}^{k}\Bigg[\tilde{\beta}(X_i-\tilde{\theta})^{\tilde{\alpha}}\left(\log^3(X_i-\tilde{\theta})\right)(\widehat{\alpha}-\alpha_0)$$
$$+ (X_i-\tilde{\theta})^{\tilde{\alpha}}\left(\log^2(X_i-\tilde{\theta})\right)(\widehat{\beta}-\beta_0)$$
$$+ 2\tilde{\beta}(X_i-\tilde{\theta})^{(\tilde{\alpha}-1)}\left(\log(X_i-\tilde{\theta})\right)(\widehat{\theta}-\theta_0)$$
$$+ \tilde{\alpha}\tilde{\beta}(X_i-\tilde{\theta})^{(\tilde{\alpha}-1)}\left(\log^2(X_i-\tilde{\theta})\right)(\widehat{\theta}-\theta_0)\Bigg]. \tag{3.46}$$

Similarly as in the second term of (3.39) we have

$$\sup_{I_{\delta_k}}\frac{1}{k}\sum_{i=1}^{k}\sup_{I_{\delta_k}}\left(-\widehat{\beta}(X_i-\widehat{\theta})^{\widehat{\alpha}}\log^2(X_i-\widehat{\theta}) + \beta_0(X_i-\theta_0)^{\alpha_0}\log^2(X_i-\theta_0)\right)$$
$$= o(k^{-\upsilon}) \quad a.s.$$

for any $0<\upsilon<\frac{1}{\alpha}$. Since for $2<\alpha\le 3$ it holds $1-\frac{2}{\alpha}\le\frac{1}{\alpha}$ and then

$$\sup_{I_{\delta_k}}\frac{1}{k}\sum_{i=1}^{k}\sup_{I_{\delta_k}}\left(-\widehat{\beta}(X_i-\widehat{\theta})^{\widehat{\alpha}}\log^2(X_i-\widehat{\theta}) + \beta_0(X_i-\theta_0)^{\alpha_0}\log^2(X_i-\theta_0)\right)$$
$$= o(k^{-\tau}) \quad a.s.$$

for any $0<\tau<1-\frac{2}{\alpha}$.

That were all the terms in (3.25), (3.26), (3.27), (3.28), (3.29), (3.30) discontinuous at $X_i=\theta$. The proof for the condition $\alpha>3$ is trivial, as according to Lemma 3.2.1, all the terms of Taylor expansions have finite expectations. □

The next theorem gives the convergency of the proposed maximum likelihood estimators.

Theorem 3.2.4. *There exists a sequence of real number $\{\delta_k\}$, such that*

$$\delta_k \sqrt{k/\log\log k} \to \infty \quad \textit{and} \quad \delta_k\, k^{(1/\alpha)+\delta} \to 0 \quad \textit{for some} \quad \delta > 0,$$

and there exists a set A with $P(A) = 1$, such that for any $\omega \in A$ we can find $k_0(\omega)$, such that for all $k \geq k_0$ there exists a local maximum of $L_k(\theta, \alpha, \beta)$ denoted by $\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)$ satisfying

$$\frac{\partial}{\partial\theta} L_k((\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)) = 0, \quad \frac{\partial}{\partial\alpha} L_k((\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)) = 0, \quad \frac{\partial}{\partial\beta} L_k((\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)) = 0$$

and

$$\frac{|\widehat{\theta}_k - \theta_0|}{\delta_k} \leq 1, \quad \frac{|\widehat{\alpha}_k - \alpha_0|}{\delta_k} \leq 1, \quad \frac{|\widehat{\beta}_k - \beta_0|}{\delta_k} \leq 1.$$

Proof. Similarly as in Smith [26], for any sequence $\{\delta_k\}$ satisfying assumptions of Theorem 3.2.4 we define for $t \in \mathbb{R}$, $x \in \mathbb{R}$, $y \in \mathbb{R}$ the function

$$f_k(t, x, y) = \frac{1}{\delta_k^2\, k} L_k(\theta_0 + \delta_k t, \alpha_0 + \delta_k x, \beta_o + \delta_k y).$$

The Taylor expansion for any $t \in \mathbb{R}$, $x \in \mathbb{R}$, $y \in \mathbb{R}$ satisfying $t^2 + x^2 + y^2 \leq 1$ implies that there exist $|\tilde{t}_\theta| < 1$, $|\tilde{x}_\theta| < 1$, $|\tilde{y}_\theta| < 1$ such that

$$\begin{aligned}
\frac{\partial f_k}{\partial t}(t, x, y) = {} & \frac{\partial f_k}{\partial t}(0, 0, 0) + \frac{\partial^2 f_k}{\partial t^2}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta)\, t + \frac{\partial^2 f_k}{\partial t \partial x}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta)\, x + \\
& + \frac{\partial^2 f_k}{\partial t \partial y}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta)\, y = \\
= {} & \frac{1}{\delta_k\, k} \frac{\partial L_k}{\partial \theta}(\varphi_{\mathbf{0}}) + \frac{t}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\widetilde{\varphi}_\theta) + \frac{x}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\widetilde{\varphi}_\theta) + \frac{y}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\widetilde{\varphi}_\theta) \\
= {} & \frac{1}{\delta_k\, k} \frac{\partial L_k}{\partial \theta}(\varphi_{\mathbf{0}}) + \frac{t}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\widetilde{\varphi}_\theta) + \frac{x}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\widetilde{\varphi}_\theta) + \frac{y}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\widetilde{\varphi}_\theta) \\
& + \frac{t}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\varphi_{\mathbf{0}}) + \frac{x}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\varphi_{\mathbf{0}}) + \frac{y}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\varphi_{\mathbf{0}}) \\
& - \frac{t}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\varphi_{\mathbf{0}}) - \frac{x}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\varphi_{\mathbf{0}}) - \frac{y}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\varphi_{\mathbf{0}}) \\
& + t\, m_{\theta\theta} - t\, m_{\theta\theta} + x\, m_{\theta\alpha} - x\, m_{\theta\alpha} + y\, m_{\theta\beta} - y\, m_{\theta\beta},
\end{aligned} \tag{3.47}$$

where $m_{\theta\theta}$, $m_{\theta\alpha}$ and $m_{\theta\beta}$ are the elements of the Fisher information matrix $\mathbf{M}$, see the definition (3.19).

For $|\tilde{t}_\theta| < 1$, $|\tilde{x}_\theta| < 1$, $|\tilde{y}_\theta| < 1$ we denote $\widetilde{\varphi}_\theta = (\widetilde{\theta}_\theta, \widetilde{\alpha}_\theta, \widetilde{\beta}_\theta)$ such as $\widetilde{\theta}_\theta = \theta_0 + \tilde{t}_\theta\, \delta_k$, $\widetilde{\alpha}_\theta = \alpha_0 + \tilde{x}_\theta\, \delta_k$, $\widetilde{\beta}_\theta = \beta_0 + \tilde{y}_\theta\, \delta_k$ satisfying $|\widetilde{\theta}_\theta - \theta_0| < \delta_k$, $|\widetilde{\alpha}_\theta - \alpha_0| < \delta_k$, $|\widetilde{\beta}_\theta - \beta_0| < \delta_k$

and we denote

$$\begin{aligned}
\varepsilon_{k,\theta}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta) = {} & \frac{1}{\delta_k\, k} \frac{\partial L_k}{\partial \theta}(\varphi_{\mathbf{0}}) + \frac{t}{k}\left(\frac{\partial^2 L_k}{\partial \theta^2}(\widetilde{\varphi}_\theta) - \frac{\partial^2 L_k}{\partial \theta^2}(\varphi_{\mathbf{0}}) \right) \\
& + \frac{x}{k}\left(\frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\widetilde{\varphi}_\theta) - \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\varphi_{\mathbf{0}}) \right) \\
& + \frac{y}{k}\left(\frac{\partial^2 L_k}{\partial \theta \partial \beta}(\widetilde{\varphi}_\theta) - \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\varphi_{\mathbf{0}}) \right) \\
& + t\left(\frac{1}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\varphi_{\mathbf{0}}) + m_{\theta\theta} \right) \\
& + x\left(\frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\varphi_{\mathbf{0}}) + m_{\theta\alpha} \right) \\
& + y\left(\frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\varphi_{\mathbf{0}}) + m_{\theta\beta} \right).
\end{aligned}$$

We can then rewrite (3.47) in a form

$$\frac{\partial f_k}{\partial t}(t, x, y) = -t\, m_{\theta\theta} - x\, m_{\theta\alpha} - y\, m_{\theta\beta} + \varepsilon_{k,\theta}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta),$$

For the term $\varepsilon_{k,\theta}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta)$ it holds

$$\begin{aligned}
\varepsilon_{k,\theta}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta) \le {} & \left| \frac{1}{\delta_k\, k} \frac{\partial L_k}{\partial \theta}(\varphi_{\mathbf{0}}) \right| + |t| \sup_{I_{\delta_k}} \left| \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\widetilde{\varphi}_\theta) - \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\varphi_{\mathbf{0}}) \right| \\
& + |x| \sup_{I_{\delta_k}} \left| \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\widetilde{\varphi}_\theta) - \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\varphi_{\mathbf{0}}) \right| \\
& + |y| \sup_{I_{\delta_k}} \left| \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\widetilde{\varphi}_\theta) - \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\varphi_{\mathbf{0}}) \right| \\
& + |t| \left| \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta^2}(\varphi_{\mathbf{0}}) + m_{\theta\theta} \right| \\
& + |x| \left| \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \alpha}(\varphi_{\mathbf{0}}) + m_{\theta\alpha} \right| \\
& + |y| \left| \frac{1}{k} \frac{\partial^2 L_k}{\partial \theta \partial \beta}(\varphi_{\mathbf{0}}) + m_{\theta\beta} \right|. \qquad (3.48)
\end{aligned}$$

From the law of the iterated logarithm (3.22) and the characteristics of the sequence δ_k that $\frac{\sqrt{\log\log k}}{\delta_k \sqrt{k}} \to 0$, we get

$$\left| \frac{1}{\delta_k\, k} \frac{\partial L_k}{\partial \theta}(\varphi_{\mathbf{0}}) \right| \to 0.$$

Then, combining (3.25), (3.26), (3.27) with (3.21), we get that also all the next terms in (3.48) tend to 0 and so we obtain

$$\varepsilon_{k,\theta}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta) \to 0 \quad a.s.$$

Similarly,

$$\frac{\partial f_k}{\partial x}(t,x,y) = \frac{1}{\delta_k\, k}\frac{\partial L_k}{\partial \alpha}(\varphi_{\mathbf{0}}) + \frac{t}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\theta}(\widetilde{\varphi}_\alpha) + \frac{x}{k}\frac{\partial^2 L_k}{\partial\alpha^2}(\widetilde{\varphi}_\alpha) + \frac{y}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\widetilde{\varphi}_\alpha)$$

$$= -t\, m_{\theta\alpha} - x\, m_{\alpha\alpha} - y\, m_{\alpha\beta} + \varepsilon_{k,\alpha}(\tilde{t}_\alpha, \tilde{x}_\alpha, \tilde{y}_\alpha),$$

where $m_{\theta\alpha}$, $m_{\alpha\alpha}$ and $m_{\alpha\beta}$ are the elements of the Fisher information matrix $\mathbf{M}$, see the definition (3.19).

For $|\tilde{t}_\alpha| < 1$, $|\tilde{x}_\alpha| < 1$, $|\tilde{y}_\alpha| < 1$ we denote $\widetilde{\varphi}_\alpha = (\widetilde{\theta}_\alpha, \widetilde{\alpha}_\alpha, \widetilde{\beta}_\alpha)$, where $\tilde{\theta}_\alpha = \theta_0 + \tilde{t}_\alpha\, \delta_k$, $\tilde{\alpha}_\alpha = \alpha_0 + \tilde{x}_\alpha\, \delta_k$, $\tilde{\beta}_\alpha = \beta_0 + \tilde{y}_\alpha\, \delta_k$ satisfying $|\tilde{\theta}_\alpha - \theta_0| < \delta_k$, $|\tilde{\alpha}_\alpha - \alpha_0| < \delta_k$, $|\tilde{\beta}_\alpha - \beta_0| < \delta_k$. For $\varepsilon_{k,\alpha}(\tilde{t}_\alpha, \tilde{x}_\alpha, \tilde{y}_\alpha)$ we obtain following inequalities.

$$\begin{aligned} \varepsilon_{k,\alpha}(\tilde{t}_\alpha, \tilde{x}_\alpha, \tilde{y}_\alpha) \le & \left|\frac{1}{\delta_k\, k}\frac{\partial L_k}{\partial\alpha}(\varphi_{\mathbf{0}})\right| + |t| \sup_{I_{\delta_k}} \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\theta\partial\alpha}(\widetilde{\varphi}_\alpha) - \frac{1}{k}\frac{\partial^2 L_k}{\partial\theta\partial\alpha}(\varphi_{\mathbf{0}})\right| \\ & + |x| \sup_{I_{\delta_k}} \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha^2}(\widetilde{\varphi}_\alpha) - \frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha^2}(\varphi_{\mathbf{0}})\right| \\ & + |y| \sup_{I_{\delta_k}} \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\widetilde{\varphi}_\alpha) - \frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\varphi_{\mathbf{0}})\right| \\ & + |t| \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\theta\partial\alpha}(\varphi_{\mathbf{0}}) + m_{\theta\alpha}\right| \\ & + |x| \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha^2}(\varphi_{\mathbf{0}}) + m_{\alpha\alpha}\right| \\ & + |y| \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\varphi_{\mathbf{0}}) + m_{\alpha\beta}\right|. \end{aligned} \tag{3.49}$$

From the law of the iterated logarithm (3.22) and the characteristics of the sequence δ_k that $\frac{\sqrt{\log\log k}}{\delta_k \sqrt{k}} \to 0$, we get

$$\left|\frac{1}{\delta_k\, k}\frac{\partial L_k}{\partial\alpha}(\varphi_{\mathbf{0}})\right| \to 0.$$

Then, combining (3.28), (3.26), (3.29) with (3.21), we get that also all the next terms in (3.49) tend to 0 and so we obtain

$$\varepsilon_{k,\alpha}(\tilde{t}_\alpha, \tilde{x}_\alpha, \tilde{y}_\alpha) \to 0 \quad a.s.$$

Further,

$$\frac{\partial f_k}{\partial y}(t,x,y) = \frac{1}{\delta_k\, k}\frac{\partial L_k}{\partial \beta}(\varphi_{\mathbf{0}}) + \frac{t}{k}\frac{\partial^2 L_k}{\partial\beta\partial\theta}(\widetilde{\varphi}_\beta) + \frac{x}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\widetilde{\varphi}_\beta) + \frac{y}{k}\frac{\partial^2 L_k}{\partial\beta^2}(\widetilde{\varphi}_\beta)$$

$$= -t\, m_{\theta\beta} - x\, m_{\alpha\beta} - y\, m_{\beta\beta} + \varepsilon_{k,\beta}(\tilde{t}_\beta, \tilde{x}_\beta, \tilde{y}_\beta),$$

where $m_{\theta\beta}$, $m_{\alpha\beta}$ and $m_{\beta\beta}$ are the elements of the Fisher information matrix $\mathbf{M}$, see the definition (3.19).

For $|\tilde{t}_\beta| < 1$, $|\tilde{x}_\beta| < 1$, $|\tilde{y}_\beta| < 1$ we denote $\widetilde{\varphi}_\beta = (\tilde{\theta}_\beta, \tilde{\alpha}_\beta, \tilde{\beta}_\beta)$, where $\tilde{\theta}_\beta = \theta_0 + \tilde{t}_\beta\,\delta_k$, $\tilde{\alpha}_\beta = \alpha_0 + \tilde{x}_\beta\,\delta_k$, $\tilde{\beta}_\beta = \beta_0 + \tilde{y}_\beta\,\delta_k$ satisfying $|\tilde{\theta}_\beta - \theta_0| < \delta_k$, $|\tilde{\alpha}_\beta - \alpha_0| < \delta_k$, $|\tilde{\beta}_\beta - \beta_0| < \delta_k$.

For $\varepsilon_{k,\beta}(\tilde{t}_\beta, \tilde{x}_\beta, \tilde{y}_\beta)$ we have following inequalities

$$
\begin{aligned}
\varepsilon_{k,\beta}(\tilde{t}_\beta, \tilde{x}_\beta, \tilde{y}_\beta) \le & \left|\frac{1}{\delta_k\,k}\frac{\partial L_k}{\partial\beta}(\varphi_{\mathbf{0}})\right| + |t| \sup_{I_{\delta_k}} \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\beta\partial\theta}(\widetilde{\varphi}_\beta) - \frac{1}{k}\frac{\partial^2 L_k}{\partial\beta\partial\theta}(\varphi_{\mathbf{0}})\right| \\
& + |x| \sup_{I_{\delta_k}} \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\widetilde{\varphi}_\beta) - \frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\varphi_{\mathbf{0}})\right| \\
& + |y| \sup_{I_{\delta_k}} \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\beta^2}(\widetilde{\varphi}_\beta) - \frac{1}{k}\frac{\partial^2 L_k}{\partial\beta^2}(\varphi_{\mathbf{0}})\right| \\
& + |t| \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\theta\partial\beta}(\varphi_{\mathbf{0}}) + m_{\theta\beta}\right| \\
& + |x| \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\alpha\partial\beta}(\varphi_{\mathbf{0}}) + m_{\alpha\beta}\right| \\
& + |y| \left|\frac{1}{k}\frac{\partial^2 L_k}{\partial\beta^2}(\varphi_{\mathbf{0}}) + m_{\beta\beta}\right|. \qquad (3.50)
\end{aligned}
$$

From the law of the iterated logarithm (3.22) and the characteristics of the sequence δ_k that $\frac{\sqrt{\log\log k}}{\delta_k\sqrt{k}} \to 0$, we get

$$
\left|\frac{1}{\delta_k\,k}\frac{\partial L_k}{\partial\beta}(\varphi_{\mathbf{0}})\right| \to 0.
$$

Then, combining (3.27), (3.29), (3.30) with (3.21), we get that also all the next terms in (3.50) tend to 0 and so we obtain

$$
\varepsilon_{k,\beta}(\tilde{t}_\beta, \tilde{x}_\beta, \tilde{y}_\beta) \to 0 \quad a.s.
$$

Let $t^2 + x^2 + y^2 = 1$. Then we have

$$
\begin{aligned}
t\frac{\partial f_k}{\partial t} + x\frac{\partial f_k}{\partial x} + y\frac{\partial f_k}{\partial y} = & -t^2\,m_{\theta\theta} - x^2\,m_{\alpha\alpha} - y^2\,m_{\beta\beta} \\
& - 2\,(x\,t\,m_{\theta\alpha} + y\,t\,m_{\theta\beta} + x\,y\,m_{\alpha\beta}) \\
& + t\,\varepsilon_{k,\theta}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta) + x\,\varepsilon_{k,\alpha}(\tilde{t}_\alpha, \tilde{x}_\alpha, \tilde{y}_\alpha) + y\,\varepsilon_{k,\beta}(\tilde{t}_\beta, \tilde{x}_\beta, \tilde{y}_\beta) \qquad (3.51)
\end{aligned}
$$

Since we proved that $\varepsilon_{k,\theta}(\tilde{t}_\theta, \tilde{x}_\theta, \tilde{y}_\theta) \to 0 \quad a.s.$, $\varepsilon_{k,\alpha}(\tilde{t}_\alpha, \tilde{x}_\alpha, \tilde{y}_\alpha) \to 0 \quad a.s.$, $\varepsilon_{k,\beta}(\tilde{t}_\beta, \tilde{x}_\beta, \tilde{y}_\beta) \to 0 \quad a.s.$, the expression (3.51) is as $k \to \infty$ strictly negative by the assumed positive-definiteness of $\mathbf{M}$. Hence Lemma A.3.5 shows that f_k has a local maximum in the range $t^2 + x^2 + y^2 < 1$ as $k \to \infty$.

Clearly, there exists a sequence $\{\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k\}$ and a set A with $P(A) = 1$ such that for any $\omega \in A$ there exists $k_0(\omega)$ such that for all $k \geq k_0$, $\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k$ is a local maximum of $L_k(\theta, \alpha, \beta)$ satisfying

$$\frac{\partial}{\partial\theta} L_k(\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k) = 0, \quad \frac{\partial}{\partial\alpha} L_k(\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k) = 0, \quad \frac{\partial}{\partial\beta} L_k(\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k) = 0$$

and

$$\frac{|\widehat{\theta}_k - \theta_0|}{\delta_k} \leq 1, \quad \frac{|\widehat{\alpha}_k - \alpha_0|}{\delta_k} \leq 1, \quad \frac{|\widehat{\beta}_k - \beta_0|}{\delta_k} \leq 1.$$

□

The next lemma shows that, in the limit, the proposed maximum likelihood estimators behave as if they were partial sums of random vectors.

Lemma 3.2.5. *For any τ such that $0 < \tau < 1 - 2/\alpha$ and for $k \to \infty$ it holds*

$$\lim_{k\to\infty} k^{\tau} \frac{1}{\sqrt{k \log\log k}} \left(\begin{pmatrix} \frac{\partial}{\partial\theta} L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\alpha} L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\beta} L_k(\varphi_{\mathbf{0}}) \end{pmatrix} - k\, \boldsymbol{M} \begin{pmatrix} \widehat{\theta}_k - \theta_0 \\ \widehat{\alpha}_k - \alpha_0 \\ \widehat{\beta}_k - \beta_0 \end{pmatrix} \right) = 0 \quad a.s. \tag{3.52}$$

Proof. Now, for instance we choose the sequence $\widetilde{\delta}_k = \log\log\log k \sqrt{\log\log k / k}\}$ satisfying conditions from Lemma 3.2.3. For maximum likelihood estimators $\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)$ we have

$$\begin{aligned}
0 =& \frac{\partial}{\partial\theta} L_k(\widehat{\varphi}_{\mathbf{k}}) = \\
=& \frac{\partial}{\partial\theta} L_k(\varphi_{\mathbf{0}}) - k(\widehat{\theta}_k - \theta_0) m_{\theta\theta} \\
&- k(\widehat{\alpha}_k - \alpha_0) m_{\theta\alpha} - k(\widehat{\beta}_k - \beta_0) m_{\theta\beta} \\
&+ \left(\frac{\partial^2}{\partial\theta^2} L_k(\widetilde{\varphi}_{\mathbf{k}}) - \frac{\partial^2}{\partial\theta^2} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\theta}_k - \theta_0) \\
&+ \left(\frac{\partial^2}{\partial\theta^2} L_k(\widetilde{\varphi}_{\mathbf{k}}) - \frac{\partial^2}{\partial\theta\partial\alpha} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\alpha}_k - \alpha_0) \\
&+ \left(\frac{\partial^2}{\partial\theta\partial\beta} L_k(\widetilde{\varphi}_{\mathbf{k}}) - \frac{\partial^2}{\partial\theta\partial\beta} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\beta}_k - \beta_0) \\
&+ \left(\frac{\partial^2}{\partial\theta^2} L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\theta} \right) (\widehat{\theta}_k - \theta_0) \\
&+ \left(\frac{\partial^2}{\partial\theta\partial\alpha} L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\alpha} \right) (\widehat{\alpha}_k - \alpha_0) \\
&+ \left(\frac{\partial^2}{\partial\theta\partial\beta} L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\beta} \right) (\widehat{\beta}_k - \beta_0)
\end{aligned} \tag{3.53}$$

for $\widetilde{\varphi}_\theta = (\widetilde{\theta}_\theta, \widetilde{\alpha}_\theta, \widetilde{\beta}_\theta)$ satisfying $|\widetilde{\theta}_k - \theta_0| < |\widehat{\theta}_k - \theta_0|, |\widetilde{\alpha}_k - \alpha_0| < |\widehat{\alpha}_k - \alpha_0|, |\widetilde{\beta}_k - \beta_0| < |\widehat{\beta}_k - \beta_0|$. Then

$$\begin{aligned}
&\frac{1}{\sqrt{k \log\log k}}\Big(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}) - k(\widehat{\theta}_k - \theta_0)m_{\theta\theta} - k(\widehat{\alpha}_k - \alpha_0)m_{\theta\alpha} - k(\widehat{\beta}_k - \beta_0)m_{\theta\beta}\Big)\\
&\quad = -\frac{\sqrt{k}}{\sqrt{\log\log k}}\frac{1}{k}\Big(\frac{\partial^2}{\partial\theta^2}L_k(\widetilde{\varphi}_{\mathbf{k}}) - \frac{\partial^2}{\partial\theta^2}L_k(\varphi_{\mathbf{0}})\Big)(\widehat{\theta}_k - \theta_0)\\
&\quad - \frac{\sqrt{k}}{\sqrt{\log\log k}}\frac{1}{k}\Big(\frac{\partial^2}{\partial\theta^2}L_k(\widetilde{\varphi}_{\mathbf{k}}) - \frac{\partial^2}{\partial\theta\partial\alpha}L_k(\varphi_{\mathbf{0}})\Big)(\widehat{\alpha}_k - \alpha_0)\\
&\quad - \frac{\sqrt{k}}{\sqrt{\log\log k}}\frac{1}{k}\Big(\frac{\partial^2}{\partial\theta\partial\beta}L_k(\widetilde{\varphi}_{\mathbf{k}}) - \frac{\partial^2}{\partial\theta\partial\beta}L_k(\varphi_{\mathbf{0}})\Big)(\widehat{\beta}_k - \beta_0)\\
&\quad - \frac{\sqrt{k}}{\sqrt{\log\log k}}\frac{1}{k}\Big(\frac{\partial^2}{\partial\theta^2}L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\theta}\Big)(\widehat{\theta}_k - \theta_0)\\
&\quad - \frac{\sqrt{k}}{\sqrt{\log\log k}}\frac{1}{k}\Big(\frac{\partial^2}{\partial\theta\partial\alpha}L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\alpha}\Big)(\widehat{\alpha}_k - \alpha_0)\\
&\quad - \frac{\sqrt{k}}{\sqrt{\log\log k}}\frac{1}{k}\Big(\frac{\partial^2}{\partial\theta\partial\beta}L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\beta}\Big)(\widehat{\beta}_k - \beta_0). \qquad (3.54)
\end{aligned}$$

Using the Marcinkiewicz-Zygmund law (3.21), characteristics (3.25), (3.26), (3.27) and the characteristics of the sequence $\{\widetilde{\delta}_k\}$ that

$$\begin{aligned}
|\widehat{\theta}_k - \theta_0| &\le (\log\log\log k)\sqrt{\frac{\log\log k}{k}},\\
|\widehat{\alpha}_k - \alpha_0| &\le (\log\log\log k)\sqrt{\frac{\log\log k}{k}},\\
|\widehat{\beta}_k - \beta_0| &\le (\log\log\log k)\sqrt{\frac{\log\log k}{k}} \qquad (3.55)
\end{aligned}$$

for the right side of (3.54) we get

$$\begin{aligned}
&\lim_{k\to\infty} k^\tau \frac{1}{\sqrt{k\log\log k}}\Big(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}) - k\ m_{\theta\theta}(\widehat{\theta}_k - \theta_0) - k\, m_{\theta\alpha}(\widehat{\alpha}_k - \alpha_0) - k\, m_{\theta\beta}(\widehat{\beta}_k - \beta_0)\Big)\\
&\qquad = 0 \quad a.s. \qquad (3.56)
\end{aligned}$$

for any τ satisfying $0 < \tau < 1 - 2/\alpha$, where the coefficients $m_{\theta\theta}$, $m_{\theta\alpha}$, $m_{\theta\beta}$ are the elements of the first line of the matrix $\mathbf{M}$. From the similar expressions for derivatives $\frac{\partial}{\partial\alpha}L_k(\widehat{\varphi}_{\mathbf{k}})$ and $\frac{\partial}{\partial\beta}L_k(\widehat{\varphi}_{\mathbf{k}})$ we obtain

$$\begin{aligned}
&\lim_{k\to\infty} k^\tau \frac{1}{\sqrt{k\log\log k}}\Big(\frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}) - km_{\alpha\theta}(\widehat{\theta}_k - \theta_0) - km_{\alpha\alpha}(\widehat{\alpha}_k - \alpha_0) - km_{\alpha\beta}(\widehat{\beta}_k - \beta_0)\Big)\\
&\qquad = 0 \quad a.s. \qquad (3.57)
\end{aligned}$$

$$\lim_{k\to\infty} k^{\tau}\frac{1}{\sqrt{k\log\log k}}\left(\frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}}) - km_{\beta\theta}(\widehat{\theta}_k-\theta_0) - km_{\beta\alpha}(\widehat{\alpha}_k-\alpha_0) - km_{\beta\beta}(\widehat{\beta}_k-\beta_0)\right)$$
$$= 0 \quad a.s., \tag{3.58}$$

where coefficients $m_{\alpha\theta}$, $m_{\alpha\alpha}$, $m_{\alpha\beta}$ are the elements of the second line of the matrix $\mathbf{M}$ and where coefficients $m_{\beta\theta}$, $m_{\beta\alpha}$, $m_{\beta\beta}$ are the elements of the third line of the matrix $\mathbf{M}$. From (3.56), (3.57), (3.58) we get the assertion (3.52). □

The next corollary gives the rate of convergency of the proposed maximum likelihood estimators.

Corollary 3.2.6. *The sequence of the proposed maximum likelihood estimators* $\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)$ *from Theorem 3.2.4 satisfies*

$$\begin{aligned}
\limsup_{k\to\infty}\frac{\sqrt{k}}{\sqrt{\log\log k}}|\widehat{\alpha}_k-\alpha_0| &= O(1) \quad a.s.,\\
\limsup_{k\to\infty}\frac{\sqrt{k}}{\sqrt{\log\log k}}|\widehat{\beta}_k-\beta_0| &= O(1) \quad a.s.,\\
\limsup_{k\to\infty}\frac{\sqrt{k}}{\sqrt{\log\log k}}|\widehat{\theta}_k-\theta_0| &= O(1) \quad a.s.
\end{aligned} \tag{3.59}$$

Proof. The proof is an easy consequence of (3.52) and the law of iterated logarithm (3.22). □

Corollary 3.2.7. *For any* τ *such that* $0<\tau<1-2/\alpha$ *it holds*

$$\lim_{k\to\infty} k^{\tau}\left(\frac{1}{k}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\boldsymbol{M}^{-1}\begin{pmatrix}\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\end{pmatrix}\right.$$
$$\left.-k\left(\widehat{\theta}_k-\theta_0, \widehat{\alpha}_k-\alpha_0, \widehat{\beta}_k-\beta_0\right)\boldsymbol{M}\begin{pmatrix}\widehat{\theta}_k-\theta_0\\ \widehat{\alpha}_k-\alpha_0\\ \widehat{\beta}_k-\beta_0\end{pmatrix}\right) = 0 \quad a.s. \tag{3.60}$$

Proof. For a matrix $\mathbf{P}$, such that $\mathbf{P}^T\mathbf{P}=\mathbf{M}$, we can write equation (3.52) as follows:

$$\lim_{k\to\infty} k^{\tau}\frac{1}{\sqrt{\log\log k}}\left(\frac{1}{\sqrt{k}}(\mathbf{P}^T)^{-1}\begin{pmatrix}\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\end{pmatrix} - \sqrt{k}\mathbf{P}\begin{pmatrix}\widehat{\theta}_k-\theta_0\\ \widehat{\alpha}_k-\alpha_0\\ \widehat{\beta}_k-\beta_0\end{pmatrix}\right) = 0 \quad a.s. \tag{3.61}$$

For $k\to\infty$ we have

$$\frac{1}{\sqrt{k}}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\mathbf{P}^{-1} + \sqrt{k}\quad\left(\widehat{\theta}_k-\theta_0, \widehat{\alpha}_k-\alpha_0, \widehat{\beta}_k-\beta_0\right)\mathbf{P}^T$$
$$= O(\sqrt{\log\log k}) \quad a.s. \tag{3.62}$$

and combining equations (3.61) and (3.62) we obtain the assertion of Corollary 3.2.7. □

Theorem 3.2.8. *For any τ such that $0 < \tau < 1 - 2/\alpha$ and for $k \to \infty$*

$$k^{\tau}\Bigg(2\left(L_k(\widehat{\varphi}_{\mathbf{k}}) - L_k(\varphi_{\mathbf{0}})\right) - \frac{1}{k}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\boldsymbol{M}^{-1}\begin{pmatrix} \frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}}) \end{pmatrix}\Bigg) \to 0 \quad a.s.$$

Proof. The Taylor expansion

$$2\left(L_k(\widehat{\varphi}_k) - L_k(\varphi_{\mathbf{0}})\right) = 2D_{(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})}L_k(\widehat{\varphi}_{\mathbf{k}}) + D^2_{(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})}L_k(\tilde{\varphi}), \tag{3.63}$$

where $D_{(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})}L_k(\widehat{\varphi}_{\mathbf{k}})$ is the first differential at the point $(\widehat{\varphi}_{\mathbf{k}})$ in the direction $(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})$, $D^2_{(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})}L_k(\tilde{\varphi})$ is the second differential at the point $(\tilde{\varphi})$ in the direction $(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})$ and

$$|\tilde{\theta} - \theta_0| < |\widehat{\theta}_k - \theta_0|, \quad |\tilde{\alpha} - \alpha_0| < |\widehat{\alpha}_k - \alpha_0|, \quad |\tilde{\beta} - \beta_0| < |\widehat{\beta}_k - \beta_0|.$$

We can rewrite (3.63)

$$\begin{aligned}
2\left(L_k(\widehat{\varphi}_{\mathbf{k}}) - L_k(\varphi_{\mathbf{0}})\right) &- \frac{1}{k}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\mathbf{M}^{-1}\begin{pmatrix} \frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}}) \end{pmatrix} \\
&= D^2_{(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})}L_k(\tilde{\varphi}) - D^2_{(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})}L_k(\varphi_{\mathbf{0}}) + D^2_{(\widehat{\varphi_{\mathbf{k}}} - \varphi_{\mathbf{0}})}L_k(\varphi_{\mathbf{0}})) \\
&\quad - k\left(\widehat{\theta}_k - \theta_0, \widehat{\alpha}_k - \alpha_0, \widehat{\beta}_k - \beta_0\right)\mathbf{M}\begin{pmatrix} \widehat{\theta}_k - \theta_0 \\ \widehat{\alpha}_k - \alpha_0 \\ \widehat{\beta}_k - \beta_0 \end{pmatrix} \\
&\quad + k\left(\widehat{\theta}_k - \theta_0, \widehat{\alpha}_k - \alpha_0, \widehat{\beta}_k - \beta_0\right)\mathbf{M}\begin{pmatrix} \widehat{\theta}_k - \theta_0 \\ \widehat{\alpha}_k - \alpha_0 \\ \widehat{\beta}_k - \beta_0 \end{pmatrix} \\
&\quad - \frac{1}{k}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\mathbf{M}^{-1}\begin{pmatrix} \frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}) \\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}}) \end{pmatrix}.
\end{aligned} \tag{3.64}$$

For the differences on the right side of (3.64) we obtain:

- The difference of second differentials is

$$
\begin{aligned}
D^2_{(\widehat{\varphi_{\mathbf{k}}}-\varphi_{\mathbf{0}})} L_k(\tilde{\varphi}) - D^2_{(\widehat{\varphi_{\mathbf{k}}}-\varphi_{\mathbf{0}})} L_k(\varphi_{\mathbf{0}}) \\
= \left(\frac{\partial^2}{\partial\theta^2} L_k(\tilde{\varphi}) - \frac{\partial^2}{\partial\theta^2} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\theta}_k - \theta_0)^2 \\
+ \left(\frac{\partial^2}{\partial\alpha^2} L_k(\tilde{\varphi}) - \frac{\partial^2}{\partial\alpha^2} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\alpha}_k - \alpha_0)^2 \\
+ \left(\frac{\partial^2}{\partial\beta^2} L_k(\tilde{\varphi}) - \frac{\partial^2}{\partial\beta^2} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\beta}_k - \beta_0)^2 \\
+ 2 \left(\frac{\partial^2}{\partial\theta\partial\alpha} L_k(\tilde{\varphi}) - \frac{\partial^2}{\partial\theta\partial\alpha} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\theta}_k - \theta_0)(\widehat{\alpha}_k - \alpha_0) \\
+ 2 \left(\frac{\partial^2}{\partial\theta\partial\beta} L_k(\tilde{\varphi}) - \frac{\partial^2}{\partial\theta\partial\beta} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\theta}_k - \theta_0)(\widehat{\beta}_k - \beta_0) \\
+ 2 \left(\frac{\partial^2}{\partial\alpha\partial\beta} L_k(\tilde{\varphi}) - \frac{\partial^2}{\partial\alpha\partial\beta} L_k(\varphi_{\mathbf{0}}) \right) (\widehat{\alpha}_k - \alpha_0)(\widehat{\beta}_k - \beta_0)
\end{aligned}
$$

and its elements are $o(k^{-\tau})$ according to Lemma 3.2.5 and Corollary 3.2.6.

- Similarly for the difference

$$
\begin{aligned}
D^2_{(\widehat{\varphi_{\mathbf{k}}}-\varphi_{\mathbf{0}})} L_k(\varphi_{\mathbf{0}}) - k \left(\widehat{\theta}_k - \theta_0, \widehat{\alpha}_k - \alpha_0, \widehat{\beta}_k - \beta_0 \right) \mathbf{M} \begin{pmatrix} \widehat{\theta}_k - \theta_0 \\ \widehat{\alpha}_k - \alpha_0 \\ \widehat{\beta}_k - \beta_0 \end{pmatrix} \\
= \left(\frac{\partial^2}{\partial\theta^2} L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\theta} \right) (\widehat{\theta}_k - \theta_0)^2 \\
+ \left(\frac{\partial^2}{\partial\alpha^2} L_k(\varphi_{\mathbf{0}}) + k\, m_{\alpha\alpha} \right) (\widehat{\alpha}_k - \alpha_0)^2 \\
+ \left(\frac{\partial^2}{\partial\beta^2} L_k(\varphi_{\mathbf{0}}) + k\, m_{\beta\beta} \right) (\widehat{\beta}_k - \beta_0)^2 \\
+ 2 \left(\frac{\partial^2}{\partial\theta\partial\alpha} L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\alpha} \right) (\widehat{\theta}_k - \theta_0)(\widehat{\alpha}_k - \alpha_0) \\
+ 2 \left(\frac{\partial^2}{\partial\theta\partial\beta} L_k(\varphi_{\mathbf{0}}) + k\, m_{\theta\beta} \right) (\widehat{\theta}_k - \theta_0)(\widehat{\beta}_k - \beta_0) \\
+ 2 \left(\frac{\partial^2}{\partial\alpha\partial\beta} L_k(\varphi_{\mathbf{0}}) + k\, m_{\alpha\beta} \right) (\widehat{\alpha}_k - \alpha_0)(\widehat{\beta}_k - \beta_0)
\end{aligned}
$$

and its elements are $o(k^{-\tau})$ according to the Marcinkiewicz-Zygmund law (3.21) and Corollary 3.2.6.

- The difference

$$k\left(\widehat{\theta}_k-\theta_0,\widehat{\alpha}_k-\alpha_0,\widehat{\beta}_k-\beta_0\right)\mathbf{M}\begin{pmatrix}\widehat{\theta}_k-\theta_0\\ \widehat{\alpha}_k-\alpha_0\\ \widehat{\beta}_k-\beta_0\end{pmatrix}$$

$$-\frac{1}{k}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}),\frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}),\frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\mathbf{M}^{-1}\begin{pmatrix}\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\end{pmatrix}$$

is $o(k^{-\tau})$ according to the Corollary 3.2.7.

Summarizing these three results we obtain that the right side of (3.64) is $o(k^{-\tau})$ a.s. □

We introduce $A(x)=\sqrt{2\log x}$ and $D_d(x)=2\log x+(d/2)\log\log x-\log\Gamma(d/2)$, similarly as in Theorem A.1.1.

Theorem 3.2.9. *The asymptotic distribution of the maximum likelihood statistic for testing the problem* (3.7) *under* H_0 *provided* $\alpha_0>2$ *is given by*

$$\lim_{n\to\infty}P\left(A(\log n)\left(\max_{0\le k\le n-1}2\log(\Lambda_k^{(0)})\right)^{1/2}\le t+D_3(\log(n))\right)=\exp(-e^{-t})$$

and for the maximum likelihood statistic for testing the problem (3.8) *we have*

$$\lim_{n\to\infty}P\left(A(\log n)\left(\max_{0\le k\le n-1}2\log(\Lambda_k)\right)^{1/2}\le t+D_3(\log(n))\right)=\exp(-2e^{-t})$$

for all $t\in\mathbb{R}$.

Proof. Using Theorem 3.2.8 we can similarly as in Csörgő and Horváth [7] prove that

$$\Bigg|\max_{1\le k\le n}\left(2\left(L_k(\widehat{\varphi}_{\mathbf{k}})-L_k(\varphi_{\mathbf{0}})\right)\right)$$

$$-\max_{1\le k\le n}\frac{1}{k}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}),\frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}),\frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\mathbf{M}^{-1}\begin{pmatrix}\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\end{pmatrix}\Bigg|$$

$$=o_P(\log\log n)$$

and the assertion of Theorem 3.2.9 is an easy consequence. □

Now, coming back to the GEV distribution and using parameters μ,ψ,ξ we can write the Theorem 3.2.9 as follows.

Theorem 3.2.10. *Provided* $-\frac{1}{2} < \xi < 0$, *the asymptotic distribution of the maximum likelihood statistic for testing the problem* (3.2) *under* H_0 *is given by*

$$\lim_{n\to\infty} P\Big(A(\log n)\left(\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})\right)^{1/2} \le t + D_3(\log(n)\Big) = \exp(-e^{-t})$$

and for the maximum likelihood statistic for testing the problem (3.3) *we have*

$$\lim_{n\to\infty} P\left(A(\log n)\left(\max_{0\le k\le n-1} 2\log(\Lambda_k)\right)^{1/2} \le t + D_3(\log(n))\right) = \exp(-2e^{-t})$$

for all $t \in \mathbb{R}$.

Proof. An easy consequence of the inequality $\alpha > 2$. □

3.3 The change-point detection for the Fréchet distributions

Now we concentrate on proving a similar theorem as Theorem 3.2.10 for parameter $\xi > 0$ corresponding to the Fréchet distribution $Fréch\,(\theta, \alpha, \beta)$ with the density function (3.5)

$$\begin{aligned} h(x;\theta,\alpha,\beta) &= \alpha\beta(x-\theta)^{-\alpha-1}\exp\{-\beta(x-\theta)^{-\alpha}\} && \text{for} \quad x \ge \theta, \\ &= 0 && \text{for} \quad x < \theta. \end{aligned}$$

Suppose that $X_1, \ldots, X_n$ are independent random variables, we are to test the null hypothesis H_0 against the alternative A_1:

$$\begin{aligned} &H_0 : X_i \sim Fréch\,(\theta_0, \alpha_0, \beta_0), && i = 1, \ldots, n, \\ &A_1 : \text{there exists} \quad k \in \{0, \ldots, n-n_0\} \quad \text{such that} && \\ &\qquad X_i \sim Fréch\,(\theta_0, \alpha_0, \beta_0), && i = 1, \ldots, k, \\ &\qquad X_i \sim Fréch\,(\theta, \alpha, \beta), && i = k+1, \ldots, n, \end{aligned} \tag{3.65}$$

where the parameters $(\theta_0, \alpha_0, \beta_0)$ before the change point are known while $(\theta, \alpha, \beta) \neq (\theta_0, \alpha_0, \beta_0)$ are unknown or to test the null hypothesis H_0 against the alternative A_2:

$$\begin{aligned} &A_2 : \text{there exists} \quad k \in \{n_0, \ldots, n-n_0\} \quad \text{such that} && \\ &\qquad X_i \sim Fréch\,(\theta_1, \alpha_1, \beta_1), && i = 1, \ldots, k, \\ &\qquad X_i \sim Fréch\,(\theta_2, \alpha_2, \beta_2), && i = k+1, \ldots, n, \end{aligned} \tag{3.66}$$

where neither the parameters before nor after the change point are known and $(\theta_1, \alpha_1, \beta_1) \neq (\theta_2, \alpha_2, \beta_2)$. The constant n_0 may be any fixed integer larger than three, $\alpha > 0$ is an unknown shape parameter, $\beta > 0$ is an unknown scale parameter and $\theta \in \mathbb{R}$ is an unknown location parameter.

Our goal is to find the limit distribution of $\max_{0\le k\le n-1} 2\log(\Lambda_k)$ for the problem (3.66), resp. $\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})$ for the problem (3.65). At first we show an important characteristic that the right tail of the density function $h(x;\theta,\alpha,\beta)$ (defined in (3.5)) decreases faster than any power of $(x-\theta)$.

Lemma 3.3.1. *For every $m \in \mathbb{R}$*

$$\lim_{x\to\infty} (x-\theta)^m \, h(x;\theta,\alpha,\beta) \to 0 \quad \textit{for} \quad x \ge \theta$$

Proof. It is an easy consequence of a limit

$$\lim_{y\to\infty} \frac{y^p}{e^y} = 0 \quad \text{for every} \quad p \in \mathbb{R}.$$

□

The log likelihood of (3.5) is given by

$$L_k(\theta,\alpha,\beta) = k\log\alpha + k\log\beta + (-\alpha-1)\sum_{i=1}^{k} \log\left(X_i-\theta\right) - \sum_{i=1}^{k} \beta\left(X_i-\theta\right)^{-\alpha}. \quad (3.67)$$

First and second derivatives of $L_k(\theta,\alpha,\beta)$ are:

$$\begin{aligned}
\frac{\partial L_k}{\partial\theta} &= \sum_{i=1}^{k}\left[\frac{(\alpha+1)}{(X_i-\theta)} - \alpha\beta(X_i-\theta)^{-\alpha-1}\right], \\
\frac{\partial L_k}{\partial\alpha} &= \sum_{i=1}^{k}\left[-\log(X_i-\theta) + \frac{1}{\alpha} - \beta(X_i-\theta)^{-\alpha}\log(X_i-\theta)\right], \\
\frac{\partial L_k}{\partial\beta} &= \sum_{i=1}^{k}\left[\frac{1}{\beta} - (X_i-\theta)^{-\alpha}\right], \\
\frac{\partial^2 L_k}{\partial\theta^2} &= \sum_{i=1}^{k}\left[\frac{(\alpha+1)}{(X_i-\theta)^2} - \alpha(\alpha+1)\beta(X_i-\theta)^{-\alpha-2}\right], \\
\frac{\partial^2 L_k}{\partial\theta\partial\alpha} &= \sum_{i=1}^{k}\left[\frac{1}{(X_i-\theta)} - \beta(X_i-\theta)^{-\alpha-1} - \alpha\beta(X_i-\theta)^{-\alpha-1}\log(X_i-\theta)\right], \\
\frac{\partial^2 L_k}{\partial\theta\partial\beta} &= \sum_{i=1}^{k}\left[-\alpha(X_i-\theta)^{-\alpha-1}\right], \\
\frac{\partial^2 L_k}{\partial\alpha^2} &= \sum_{i=1}^{k}\left[-\frac{1}{\alpha^2} - \beta(X_i-\theta)^{-\alpha}\log^2(X_i-\theta)\right],
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 L_k}{\partial\alpha\partial\beta} &= \sum_{i=1}^{k}\left[-(X_i-\theta)^{-\alpha}\log(X_i-\theta)\right], \\
\frac{\partial^2 L_k}{\partial\beta^2} &= \sum_{i=1}^{k}\left[-\frac{1}{\beta^2}\right].
\end{aligned} \tag{3.68}$$

It holds

$$\begin{aligned}
E\Big(\frac{\partial}{\partial\theta}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big) &= 0, \\
E\Big(\frac{\partial}{\partial\alpha}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big) &= 0, \\
E\Big(\frac{\partial}{\partial\beta}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big) &= 0.
\end{aligned} \tag{3.69}$$

According to Lemma 3.3.1, for every $s \in \mathbb{R}$ it holds

$$\begin{aligned}
E\Big(\frac{\partial^2}{\partial\theta^2}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\theta\partial\alpha}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\theta\partial\beta}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\alpha^2}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\alpha\partial\beta}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty, \\
E\Big(\frac{\partial^2}{\partial\beta^2}\big(\log h(X_i;\varphi_{\mathbf{0}})\big)\Big)^s &< \infty.
\end{aligned}$$

Let's denote a Fisher information matrix $\mathbf{M}$ on a parameter $\varphi_{\mathbf{0}} = (\theta_0, \alpha_0, \beta_0)$ with elements

$$\mathbf{M} = \begin{pmatrix} m_{\theta\theta} & m_{\theta\alpha} & m_{\theta\beta} \\ m_{\alpha\theta} & m_{\alpha\alpha} & m_{\alpha\beta} \\ m_{\beta\theta} & m_{\beta\alpha} & m_{\beta\beta} \end{pmatrix},$$

where

$$\begin{aligned}
m_{\theta\theta} &= E\{\frac{\partial}{\partial\theta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\theta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\theta^2}\log(h(X_i;\,\varphi_{\mathbf{0}}))\}, \\
m_{\alpha\alpha} &= E\{\frac{\partial}{\partial\alpha}\log(h(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\alpha}\log(h(X_i;\,\varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\alpha^2}\log(h(X_i;\,\varphi_{\mathbf{0}}))\},
\end{aligned}$$

$$\begin{aligned}
m_{\beta\beta} &= E\{\frac{\partial}{\partial\beta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\}\\
&= -E\{\frac{\partial^2}{\partial\beta^2}\log(h(X_i;\,\varphi_{\mathbf{0}}))\},\\
m_{\theta\alpha} = m_{\alpha\theta} &= E\{\frac{\partial}{\partial\theta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\alpha}\log(h(X_i;\,\varphi_{\mathbf{0}}))\}\\
&= -E\{\frac{\partial^2}{\partial\theta\partial\alpha}\log(h(X_i;\,\varphi_{\mathbf{0}}))\},
\end{aligned}$$

$$\begin{aligned}
m_{\theta\beta} = m_{\beta\theta} &= E\{\frac{\partial}{\partial\theta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\}\\
&= -E\{\frac{\partial^2}{\partial\theta\partial\beta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\},\\
m_{\alpha\beta} = m_{\beta\alpha} &= E\{\frac{\partial}{\partial\alpha}\log(h(X_i;\,\varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\}\\
&= -E\{\frac{\partial^2}{\partial\alpha\partial\beta}\log(h(X_i;\,\varphi_{\mathbf{0}}))\}. \qquad (3.70)
\end{aligned}$$

A maximum likelihood estimator based on $X_1,\ldots,X_k$ (when it exists) will be denoted by $\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)$ and satisfies

$$\frac{\partial L_k}{\partial\theta}(\widehat{\varphi}_{\mathbf{k}}) = 0, \quad \frac{\partial L_k}{\partial\alpha}(\widehat{\varphi}_{\mathbf{k}}) = 0, \quad \frac{\partial L_k}{\partial\beta}(\widehat{\varphi}_{\mathbf{k}}) = 0. \qquad (3.71)$$

The existence of $\widehat{\varphi}_{\mathbf{k}} = (\widehat{\theta}_k, \widehat{\alpha}_k, \widehat{\beta}_k)$ is guaranteed by Theorem A.3.1, see Appendix.

M is a positive definite matrix. According to the Marcinkiewicz-Zygmund law, (see Appendix - Theorem A.3.2) we obtain for the Fréchet distribution following relations: for any τ such that $0 < \tau < \frac{1}{2}$

$$\begin{aligned}
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\theta^2}L_k(\varphi_{\mathbf{0}}) + m_{\theta\theta}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\theta\partial\alpha}L_k(\varphi_{\mathbf{0}}) + m_{\theta\alpha}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\theta\partial\beta}L_k(\varphi_{\mathbf{0}}) + m_{\theta\beta}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\alpha^2}L_k(\varphi_{\mathbf{0}}) + m_{\alpha\alpha}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\alpha\partial\beta}L_k(\varphi_{\mathbf{0}}) + m_{\alpha\beta}\Big) &= 0 \quad a.s.,\\
\lim_{k\to\infty} k^{\tau}\Big(\frac{1}{k}\frac{\partial^2}{\partial\beta^2}L_k(\varphi_{\mathbf{0}}) + m_{\beta\beta}\Big) &= 0 \quad a.s.
\end{aligned}$$

$$(3.72)$$

Similarly, applying the law of the iterated logarithm (Appendix - Theorem A.3.3) we get:

$$\begin{aligned}\limsup_{k\to\infty} \frac{\frac{\partial}{\partial\theta} L_k(\varphi_{\mathbf{0}})}{\sqrt{k \log\log k}} &= O(1) \quad a.s.,\\ \limsup_{k\to\infty} \frac{\frac{\partial}{\partial\alpha} L_k(\varphi_{\mathbf{0}})}{\sqrt{k \log\log k}} &= O(1) \quad a.s.,\\ \limsup_{k\to\infty} \frac{\frac{\partial}{\partial\beta} L_k(\varphi_{\mathbf{0}})}{\sqrt{k \log\log k}} &= O(1) \quad a.s.\end{aligned} \tag{3.73}$$

For the reparameterization (3.5) we can prove similar lemma as for the Weibull distribution.

Lemma 3.3.2. *For any sequence* $\{\delta_k\}$ *satisfying* $\delta_k\, k^{1/\tau} \to 0$ *and for any* τ *such that* $0 < \tau < \frac{1}{2}$

$$\begin{aligned}\lim_{k\to\infty} k^\tau \left(\sup_{I_{\delta_k}} \frac{1}{k} \left| \frac{\partial^2}{\partial\theta^2} L_k(\varphi) - \frac{\partial^2}{\partial\theta^2} L_k(\varphi_{\mathbf{0}}) \right| \right) &= 0 \quad a.s.,\\ \lim_{k\to\infty} k^\tau \left(\sup_{I_{\delta_k}} \frac{1}{k} \left| \frac{\partial^2}{\partial\theta\partial\alpha} L_k(\varphi) - \frac{\partial^2}{\partial\theta\partial\alpha} L_k(\varphi_{\mathbf{0}}) \right| \right) &= 0 \quad a.s.,\\ \lim_{k\to\infty} k^\tau \left(\sup_{I_{\delta_k}} \frac{1}{k} \left| \frac{\partial^2}{\partial\theta\partial\beta} L_k(\varphi) - \frac{\partial^2}{\partial\theta\partial\beta} L_k(\varphi_{\mathbf{0}}) \right| \right) &= 0 \quad a.s.,\\ \lim_{k\to\infty} k^\tau \left(\sup_{I_{\delta_k}} \frac{1}{k} \left| \frac{\partial^2}{\partial\alpha^2} L_k(\varphi) - \frac{\partial^2}{\partial\alpha^2} L_k(\varphi_{\mathbf{0}}) \right| \right) &= 0 \quad a.s.,\\ \lim_{k\to\infty} k^\tau \left(\sup_{I_{\delta_k}} \frac{1}{k} \left| \frac{\partial^2}{\partial\alpha\partial\beta} L_k(\varphi) - \frac{\partial^2}{\partial\alpha\partial\beta} L_k(\varphi_{\mathbf{0}}) \right| \right) &= 0 \quad a.s.,\\ \lim_{k\to\infty} k^\tau \left(\sup_{I_{\delta_k}} \frac{1}{k} \left| \frac{\partial^2}{\partial\beta^2} L_k(\varphi) - \frac{\partial^2}{\partial\beta^2} L_k(\varphi_{\mathbf{0}}) \right| \right) &= 0 \quad a.s.\end{aligned} \tag{3.74}$$

Proof. We can use similar arguments for the terms in second derivatives as in the proof of Lemma 3.2.3 for the three parameter Weibull distribution from previous section. Substituting second derivatives from (3.68) to the differences of the second derivatives (3.74) results in the differences of terms, which are discontinuous at the points $X_i = \theta$. Using the Taylor expansion for these differences, we obtain terms of the type

$$\frac{1}{(X_i - \theta)^m} \log^p(X_i - \theta), \quad \text{where} \quad m > 0, \quad p \geq 0. \tag{3.75}$$

Applying Lemma 3.3.1, we get that all the expectations of the terms (3.75) are finite and then, similarly as for the three parameter Weibull distribution, we get Lemma 3.3.2. □

Theorem 3.3.3. *The asymptotic distribution of the maximum likelihood statistic for testing the problem* (3.65) *under* H_0 *provided* $\xi > 0$ *is given by*

$$\lim_{n\to\infty} P\Big(A(\log n)\left(\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})\right)^{1/2} \le t + D_3(\log(n))\Big) = \exp(-e^{-t})$$

and for the maximum likelihood statistic for testing the problem (3.66) *we have*

$$\lim_{n\to\infty} P\Big(A(\log n)\left(\max_{0\le k\le n-1} 2\log(\Lambda_k)\right)^{1/2} \le t + D_3(\log(n))\Big) = \exp(-2e^{-t})$$

for all t.

Proof. Similarly to the Weibull distribution, it can be proved

$$\Bigg|\max_{1\le k\le n}\left(2\left(L_k(\widehat{\varphi}_{\mathbf{k}}) - L_k(\varphi_{\mathbf{0}})\right)\right) - \max_{1\le k\le n}\frac{1}{k}\left(\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}}), \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\right)\mathbf{M}^{-1}\begin{pmatrix}\frac{\partial}{\partial\theta}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\alpha}L_k(\varphi_{\mathbf{0}})\\ \frac{\partial}{\partial\beta}L_k(\varphi_{\mathbf{0}})\end{pmatrix}\Bigg| = o_P(\log\log n)$$

and the assertion of Theorem 3.3.3 is an easy consequence. □

Using the facts from the Theorem 3.2.10 and Theorem 3.3.3, we get the asymptotic distribution of the GEV distribution.

Theorem 3.3.4. *The asymptotic distribution of the maximum likelihood statistic for testing the problem* (3.2) *under* H_0 *provided* $\xi > -\frac{1}{2}$ *is given by*

$$\lim_{n\to\infty} P\Big(A(\log n)\left(\max_{0\le k\le n-1} 2\log(\Lambda_k^{(0)})\right)^{1/2} \le t + D_3(\log(n))\Big) = \exp(-e^{-t})$$

and for the maximum likelihood statistic for testing the problem (3.3) *we have*

$$\lim_{n\to\infty} P\Big(A(\log n)\left(\max_{0\le k\le n-1} 2\log(\Lambda_k)\right)^{1/2} \le t + D_3(\log(n))\Big) = \exp(-2e^{-t})$$

for all t.

Asymptotic critical values of the testing statistic $\max_{0\le k\le n-1} 2\log(\Lambda_k)$ calculated for different values n according to Theorem 3.3.4 are listed in the following table.

	$\alpha = 0.05$	$\alpha = 0.01$
n=200	18.65	27.15
n=250	18.75	27.17
n=500	19.05	27.21
n=1000	19.3	27.28

Table 6. Asymptotic critical values of the testing statistic $\max_{0\le k\le n-1} 2\log(\Lambda_k)$ for different values n.

3.4 Results

We tried to fit a GEV distribution to the minus annual minima as well as to the annual maxima and test for a change in all three parameters. The null hypothesis H and the alternative hypothesis A for a change in all three parameters may be set as in (3.3):

$$
\begin{aligned}
&H : Y_i \sim GEV(\mu,\, \psi,\, \xi), && i = 1, \ldots, n,\\
&A : \exists\, k \in \{1, \ldots, n-1\} \quad \text{such that} &&\\
&\qquad Y_i \sim GEV(\mu_1,\, \psi_1,\, \xi_1), && i = 1, \ldots, k,\\
&\qquad Y_i \sim GEV(\mu_2,\, \psi_2,\, \xi_2), && i = k+1, \ldots, n,
\end{aligned}
$$

where the parameters $(\mu_1, \psi_1, \xi_1) \neq (\mu_2, \psi_2, \xi_2)$ are unknown both before as well as after the change point. Testing of the problem above may be based on the likelihood ratio Λ_k, more specifically

$$\max_{3 \le k \le n-3} 2 \log \Lambda_k = \max_{3 \le k \le n-3} 2\big(L_k(\hat{\mu}_k, \hat{\psi}_k, \hat{\xi}_k) + L_k^*(\hat{\mu}_k^*, \hat{\psi}_k^*, \hat{\xi}_k^*) - L_n(\hat{\mu}_n, \hat{\psi}_n, \hat{\xi}_n)\big),$$

where

$$L_k(\hat{\mu}_k, \hat{\psi}_k, \hat{\xi}_k) = \sum_{i=1}^{k} \log h(x; \hat{\mu}_k, \hat{\psi}_k, \hat{\xi}_k),\ L_k^*(\hat{\mu}_k^*, \hat{\psi}_k^*, \hat{\xi}_k^*) = \sum_{i=k+1}^{n} \log h(x; \hat{\mu}_k^*, \hat{\psi}_k^*, \hat{\xi}_k^*)$$

and $\hat{\mu}_k,\ \hat{\psi}_k,\ \hat{\xi}_k$ are the maximum likelihood estimators of the parameters based on the first k observations, while $\hat{\mu}_k^*,\ \hat{\psi}_k^*,\ \hat{\xi}_k^*$ are the maximum likelihood estimators based on the last $n-k$ observations. We recall that for every k under H the statistic $2 \log \Lambda_k$ is asymptotically distributed according to a χ^2 distribution with 3 degrees of freedom. Theoretically, the approximate critical values may be calculated using the asymptotic behavior of $\big(\max_{3 \le k \le n-3} 2 \log \Lambda_k)^{1/2}$. However, to obtain maximum of all log-likelihood ratios we have to calculate maximum-likelihood estimates for all possible splits, i.e. for all time points $k = 3, \ldots, n-3$. According to our experience good estimates of the parameters are obtained only if they are calculated from 50 observations at least. That is why we recommend to use a test statistic

$$TT_3 = \big(\max_{50 \le k \le n-50} 2 \log \Lambda_k\big)^{1/2}, \quad \alpha \in (0, 1).$$

The limit distribution of the statistic TT_3 is given by the asymptotics:

$$P\left(\big(\max_{k_0 \le k \le n-k_0} 2 \log \Lambda_k\big)^{1/2} > \frac{x + b_{n3}}{a_{n3}}\right) \approx 1 - \exp\left\{-2\, e^{-x}\right\}, \tag{3.76}$$

where

$$
\begin{aligned}
a_{n3} &= \sqrt{2 \log \log n},\\
b_{n3} &= 2 \log \log n + (3/2) \log\log \log n - \log \Gamma\big(3/2\big).
\end{aligned}
$$

We tried to fit a GEV distribution to the minus annual minima as well as to the annual maxima and test for a change in all three parameters using the test statistic TT_3. Table 7 shows values of TT_3^2 for annual minima as well as for annual maxima. The numbers in red denote significant values, compare with the asymptotical critical values in Table 6 for $n = 250$.

	TT_3^2		TT_3^2
Brussels min	7.62	Brussels max	70.4
Cadiz min	26.8	Cadiz max	14.2
Milan min	25.0	Milan max	27.2
Padua min	7.3	Padua max	5.4
St. Peter. min	26.0	St. Peter. max	9.4
Stockholm min	38.2	Stockholm max	17.8
Uppsala min	32.4	Uppsala max	13.0
Prague min	24.4	Prague max	22.0

Table 7. The values of the statistic TT_3^2.

As the number of the trimmed portion is too large, we propose another way of the asymptotics. Under H the following approximation holds true for large values of u^2:

$$P\Big(\max_{\lceil\beta n\rceil \le k \le n-\lceil\beta n\rceil} 2\log\Lambda_k > u^2\Big) \approx 2\log\big(\tfrac{1-\beta}{\beta}\big)\frac{u^3 e^{-u^2/2}}{2^{3/2}\Gamma(3/2)}. \qquad (3.77)$$

The asymptotic 5% critical value for the statistic TT_3^2 obtained by (3.77) with $\beta = 0.2$ is equal to 11.14. We fit again the GEV distribution to the minus annual minima as well as to the annual maxima and test for a change in all three parameters using the test statistic TT_3^2 with $\beta = 50/n \doteq 0.2$. Table 8 presents the results for a change in all three parameters of the GEV distribution for the minimal values with significant values of the testing statistic denoted in red , Table 9 presents the same for the minimal values.

	μ	ψ	ξ	TT_3^2
Brussels min	6.02 5.08	3.56 2.78	-0.29 -0.15	7.62
Cadiz min	-6.73 -6.42	2.57 1.68	0.00 -0.25	26.8
Milan min	4.04 2.72	2.66 1.94	-0.15 -0.14	25.0
Padua min	3.22 2.50	2.38 1.85	-0.06 0.00	7.3
St. Peter. min	22.9 19.78	5.10 4.37	-0.28 -0.22	26.0
Stockholm min	14.94 11.92	4.19 3.07	-0.25 -0.09	38.2
Uppsala min	17.04 14.68	5.36 3.61	-0.35 -0.18	32.4
Prague min	10.95 8.53	5.07 3.44	-0.29 -0.12	24.4

Table 8. Change in all three parameters of GEV for minus annual minima using TT_3^2 with $\beta = [50/n]$.

	μ	ψ	ξ	TT_3^2
Brussels max	22.0 24.0	1.51 1.69	-0.30 -0.25	70.4
Cadiz max	28.5 29.3	1.37 1.34	-0.24 -0.21	14.2
Padua max	27.0 27.2	1.21 1.15	-0.26 -0.12	5.4
Milan max	27.1 27.9	1.35 1.14	-0.11 -0.28	27.2
St. Peter. max	22.8 23.5	1.86 1.80	-0.24 -0.36	9.4
Stockholm max	22.7 21.4	1.71 1.89	-0.49 -0.21	17.8
Uppsala max	21.8 21.4	2.08 1.89	-0.31 -0.21	13.0
Prague max	25.2 26.3	1.56 1.39	-0.15 -0.21	22.0

Table 9. Change in all three parameters of GEV for annual maxima using TT_3^2 with $\beta = [50/n]$.

3.5 Conclusion

The GEV distribution fit was in all the cases far from being perfect. It worked better for the minus minimal values than for the maximal values, where surprisingly a normal distribution seems to be a better model than a GEV distribution as the annual maxima are almost symmetric. Moreover, to find numerically the maximum likelihood estimates for all three parameters together with the corresponding value of the log-likelihood function is a difficult task. We recommend to use this approach only if the analyzed series is very long. It seems that even 200 is not enough.

More specifically, in the case of the annual maxima the change in location was evidently present in the Brussels series and it was slightly less evident for the Cadiz series. It seems that the Stockholm annual maximal series decreases. Furthermore, shortly after the beginning the Milan annual maxima series contains a several very high temperatures which appear never again later. This is mainly the reason why when applying the GEV distribution for modelling the Milan maxima, the null hypothesis H of stationarity is rejected (there is a strong decrease in the shape parameter). If we omit the first 30 observations, the value TT_3 goes down. Nevertheless, it seems that a small increase in location is present here as well. Figures 34 − 41 presenting densities before the estimated change point (solid line) and after the change point (dashed line) accompanied by Table 10 presenting 5%, 50% and 95% quantizes of the estimated GEV distributions suggest that there might be a slight increase in distribution of maxima of almost all series (with an exception of the Stockholm series), nevertheless it is very small. The hypothesis expressed by climatologists that due to climate change the occurrence of extremal high temperatures becomes more frequent seems to be false.

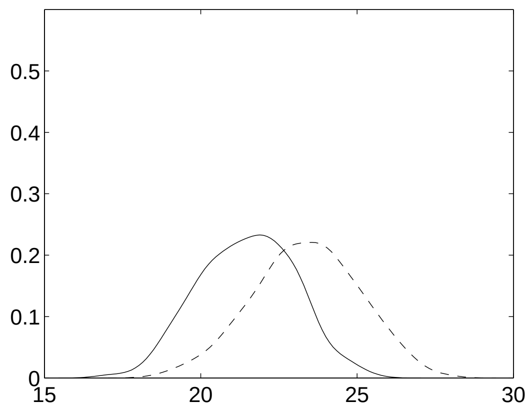

Figure 34. Brussels (annual maxima in °C) - fitted GEV densities.

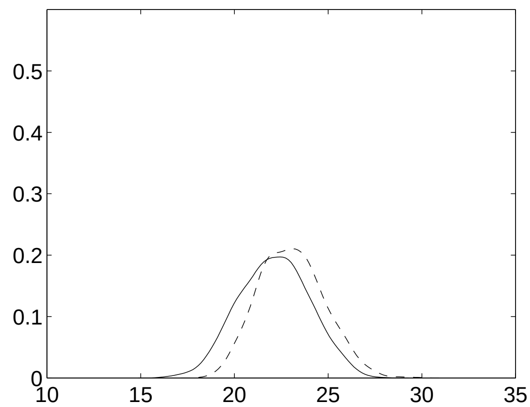

Figure 35. St. Petersburg (annual maxima in °C) - fitted GEV densities.

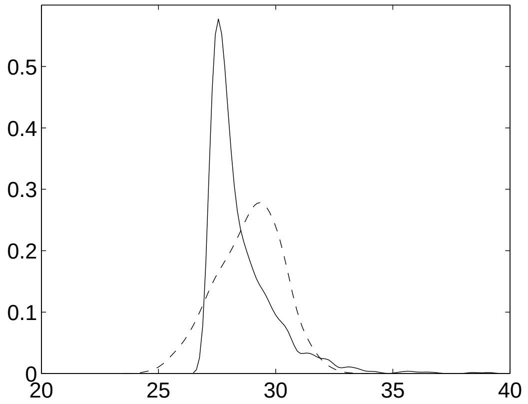

Figure 36. Cadiz (annual maxima in °C) - fitted GEV densities.

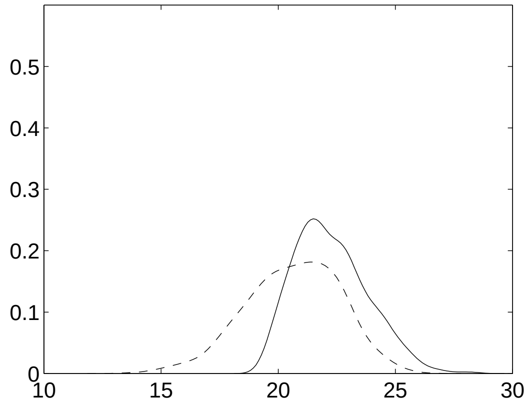

Figure 37. Stockholm (annual maxima in °C) - fitted GEV densities.

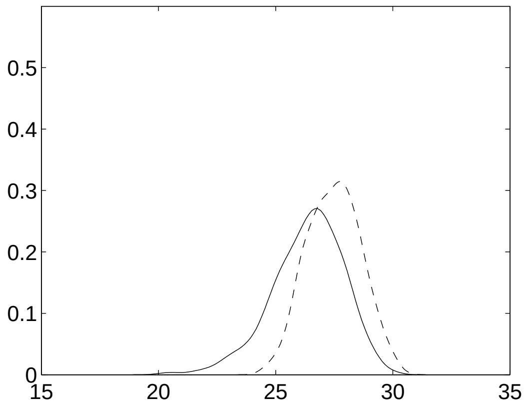

Figure 38. Milan (annual maxima in °C) - fitted GEV densities.

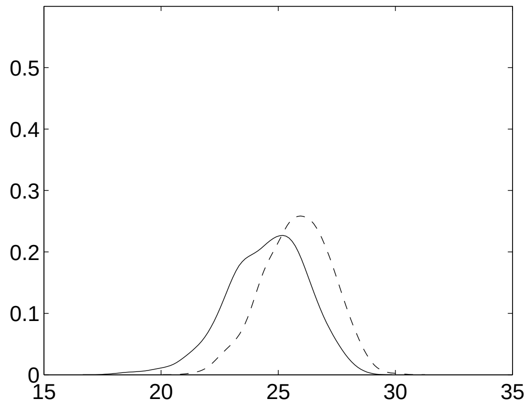

Figure 39. Prague (annual maxima in °C) - fitted GEV densities.

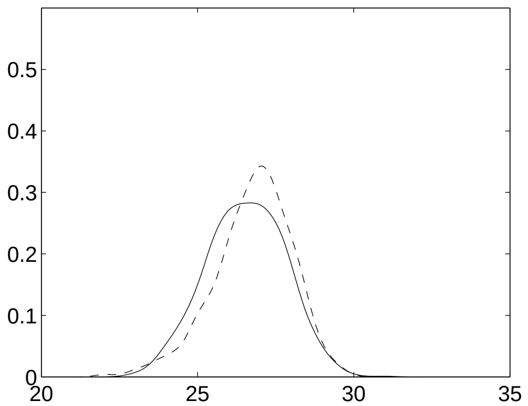

Figure 40. Padua (annual maxima in °C) - fitted GEV densities.

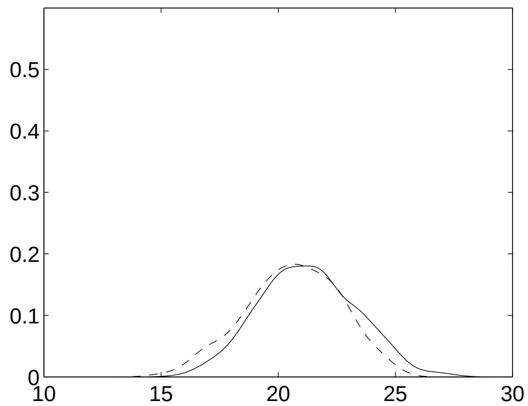

Figure 41. Uppsala (annual maxima in °C) - fitted GEV densities.

quant.	Brussels first part	Brussels second part		quant.	St. Petersburg first part	St. Petersburg second part
5 %	20.0	21.9		5 %	20.5	21.1
50 %	22.5	24.6		50 %	23.5	24.1
95 %	25.0	27.5		95 %	26.8	26.8
quant.	Cadiz first part	Cadiz second part		quant.	Stockholm first part	Stockholm second part
5 %	26.8	27.6		5 %	20.2	19.1
50 %	29.0	29.8		50 %	23.3	22.1
95 %	31.4	32.3		95 %	25.4	25.6
quant.	Milan first part	Milan second part		quant.	Padua first part	Padua second part
5 %	25.5	26.4		5 %	25.5	25.9
50 %	27.6	28.3		50 %	27.4	27.7
95 %	30.5	30.2		95 %	29.5	30.1
quant.	Upsala first part	Uppsala second part		quant.	Prague first part	Prague second part
5 %	19.0	19.0		5 %	23.3	24.6
50 %	22.5	21.7		50 %	25.7	26.8
95 %	25.8	24.9		95 %	28.9	39.4

Table 10. Several quantizes of estimated GEV distribution of annual maxima before and after a change point.

When analyzing the annual minimal series, the tests confirm a clear increase in the Cadiz, Milan, St. Petersburg, Stockholm, Uppsala and Prague series, while the increase in Brussels and Padua annual minimal temperatures was not significant. Moreover, if we look at the values of the parameters of a GEV distribution before and after the change point more closely, we see that a change in the location parameter is more striking and a shift in the location parameter is accompanied by a decrease of the scale parameter and an increase of the parameter of asymmetry. Figures 42 – 49 presenting densities before the estimated change point (solid line) and after the change point (dashed line) accompanied by Table 11 presenting 5%, 50% and 95% quantizes of the estimated GEV distributions give us an idea what are the main futures of change in distribution. The hypothesis that winters become milder as extremely cold days appear less frequent seems to be correct.

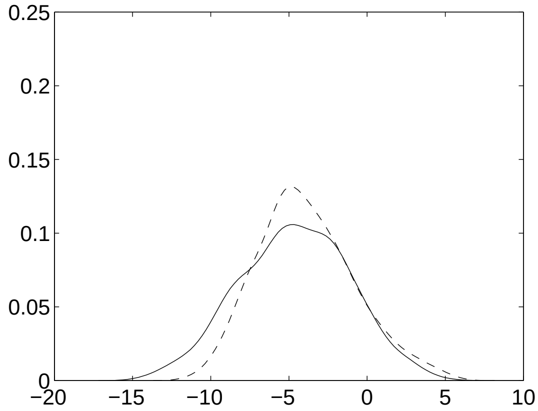

Figure 42. Brussels (annual minima in °C) - fitted GEV densities.

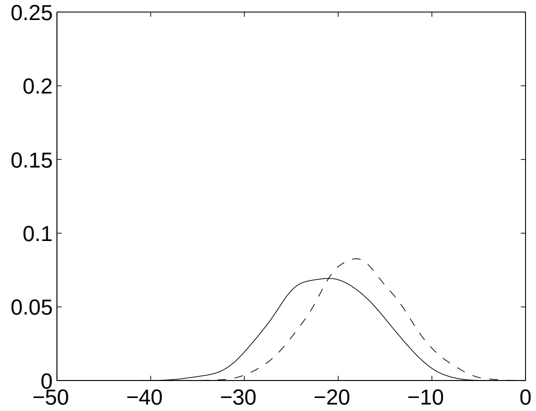

Figure 43. St. Petersburg (annual minima in °C) - fitted GEV densities.

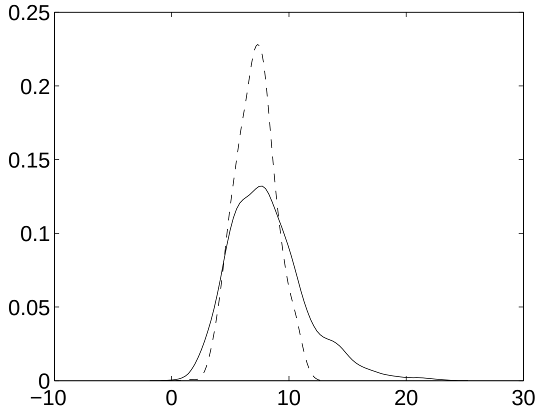

Figure 44. Cadiz (annual minima in °C) - fitted GEV densities.

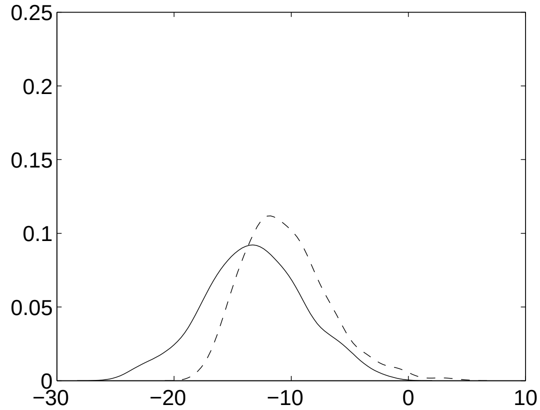

Figure 45. Stockholm (annual minima in °C) - fitted GEV densities.

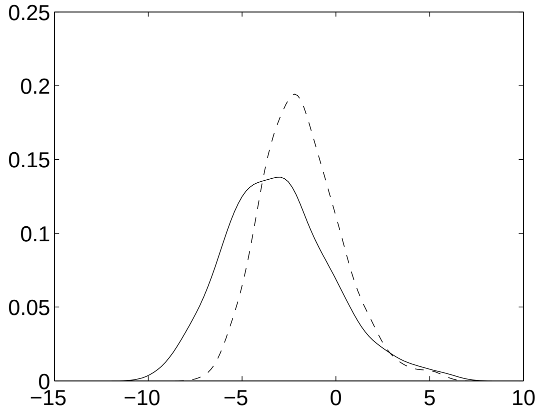

Figure 46. Milan (annual minima in °C) - fitted GEV densities.

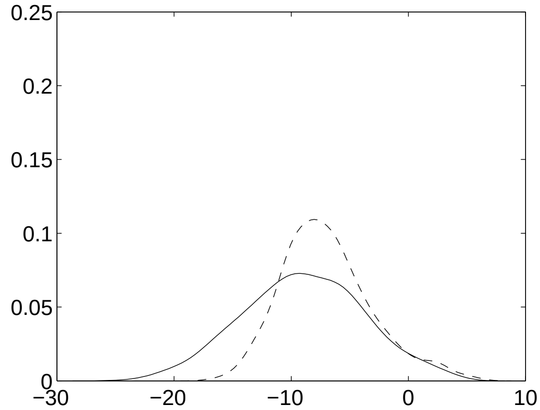

Figure 47. Prague (annual minima in °C) - fitted GEV densities.

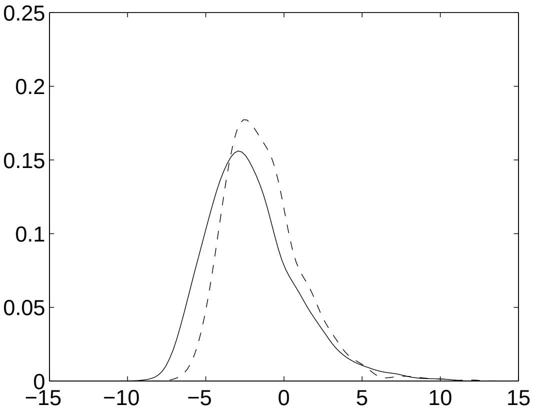

Figure 48. Padua (annual minima in °C) - fitted GEV densities.

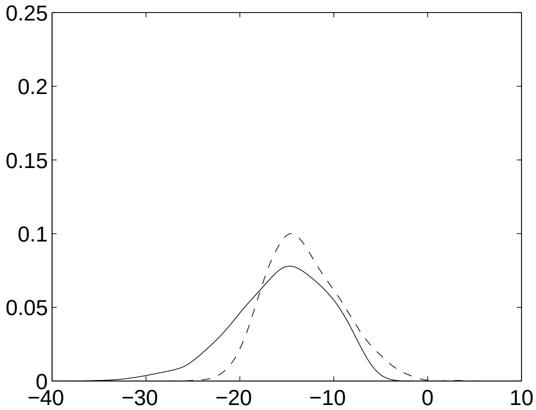

Figure 49. Uppsala (annual minima in °C) - fitted GEV densities.

quant.	Brussels first part	Brussels second part		quant.	St. Petersburg first part	St. Petersburg second part
5 %	-13.1	-11.8		5 %	-33.2	-29.3
50 %	-7.3	-6.1		50 %	-24.7	-21.3
95 %	-1.5	-1.8		95 %	-16.4	-14.4
quant.	Cadiz first part	Cadiz second part		quant.	Stockholm first part	Stockholm second part
5 %	-0.9	2.9		5 %	-23.7	-19.9
50 %	5.8	5.8		50 %	-16.4	-13.0
95 %	9.5	8.5		95 %	-9.7	-8.4
quant.	Milan first part	Milan second part		quant.	Padua first part	Padua second part
5 %	-10.4	-7.4		5 %	-9.7	-8.0
50 %	-5.0	-3.4		50 %	-4.1	-3.2
95 %	-0.9	-0.4		95 %	-0.5	-0.5
quant.	Uppsala first part	Uppsala second part		quant.	Prague first part	Prague second part
5 %	-26.8	-24.7		5 %	-21.1	-17.1
50 %	-19.2	-16.6		50 %	-12.7	-9.8
95 %	-9.3	-10.6		95 %	-4.4	-4.4

Table 11. Several quantizes of estimated GEV distribution of annual minima before and after a change point.

Using the results of our statistical inference it is possible to say that the increase in the annual minimal temperatures is usually more pronounced than the increase in the annual maximal temperatures. (We even discovered a decrease in the Stockholm annual maximal temperatures.) This is in agreement with the remark of Camuffo and Jones [6] who claim: Analysis of the distribution of extreme events, undertaken using the results from IMPROVE, has shown that for most of the study sites the recent warming is characterized more by a decrease in frequency of the coldest days than by increase in frequency of the warmest.

Problem 2

Application of change-point detection for occurrences of unusually hot, resp. cold days

Chapter 4

The change-point detection for dependent data

Clearly, working with real temperature series, we can not expect that the condition of independency is fulfilled, especially when the measurements are very close in time. The study of our series shows that there is a strong correlation between daily temperature values, the correlation between two subsequent days is for all series very close to 0.8, see Figure 33. To obtain similar result concerning the distribution of the test statistic under the hypothesis of stationarity for dependent variables, we use the almost sure approximation of the partial sums of random variables, satisfying a strong-mixing condition, by a suitable Brownian motion and we show that the analogue of Csörgő and Horváth theorem holds.

4.1 The change-point detection for strong-mixing sequences

We consider the following assumptions.

Let $\{X(i), i = 0, 1, 2, \ldots, n\}$, $\{X^{(1)}(i) \cdot i = 0, 1, 2, \ldots, n\}$, $\{X^{(2)}(i), i = 0, 1, 2, \ldots, n\}$ form

$$\textit{strictly stationary, strong-mixing sequences with mixing coefficients} \quad \alpha(k) = O(r_0^{-k}) \tag{4.1}$$

satisfying

$$\begin{aligned} \mathsf{E}X(i) = \mu, \quad \mathsf{E}X^{(1)}(i) = \mu_1, \quad \mathsf{E}X^{(2)}(i) = \mu_2, \quad d := \mu_2 - \mu_1 \neq 0, \\ \mathsf{E}|X(i)|^\nu < \infty, \quad \mathsf{E}|X^{(1)}(i)|^\nu < \infty, \quad \mathsf{E}|X^{(2)}(i)|^\nu < \infty \\ \textit{with} \quad \nu > 4 \quad \textit{for all} \quad i = 0, 1, 2, \ldots, n. \end{aligned} \tag{4.2}$$

We consider a sequence $\{Y(i), i = 0, 1, 2, \ldots, n\}$ and the hypotheses testing may be set as follows:

$$\begin{aligned}
&H_0 : Y(i) = X(i), && i = 1, \ldots, n \qquad (4.3)\\
&H_A : \text{there exists} \quad m^* \in \{1, \ldots, n-1\} \quad \text{such that} && \\
&\qquad Y(i) = X^{(1)}(i), && i = 1, \ldots, m^*,\\
&\qquad Y(i) = X^{(2)}(i), && i = m^* + 1, \ldots, n.
\end{aligned}$$

The test for H_0 against H_A will be based on functionals of

$$T_n(t) = S_n(t) - tS_n(1), \tag{4.4}$$

where

$$S_n(t) = \begin{cases} n^{-1/2} \sum\limits_{1 \le i \le (n+1)t} Y_i & 0 \le t < 1, \\ n^{-1/2} \sum\limits_{1 \le i \le n} Y_i & t = 1. \end{cases}$$

Similarly as for linear processes in Theorem 4.1.3 of Csörgő and Horváth [7] we obtain the following theorem for strong-mixing sequences.

Theorem 4.1.1. *Assume conditions* (4.1), (4.2) *and* H_0 *hold then we have for all* $x \in \mathbb{R}$

$$\lim_{n\to\infty} P\big(A(\log n) \frac{1}{\sigma} \sup_{0<t<1} \frac{|T_n(t)|}{\sqrt{t(1-t)}} \le x + D(\log n)\big) = \exp(-2e^{-x}),$$

where

$$\sigma^2 = \mathsf{E}(Y_0 - \mu)^2 + 2\sum_{i=1}^{\infty} \mathsf{E}(Y_0 - \mu)(Y_i - \mu), \tag{4.5}$$

$$A(x) = \sqrt{2 \, \log x}, \tag{4.6}$$

$$D(x) = 2 \log x + \frac{1}{2} \log\log x - \frac{1}{2} \log \pi. \tag{4.7}$$

Proof. For simplicity denote $\{e_i = (Y_i - \mu), i = 0, 1, 2, \ldots, n\}$. Without loss of generality we can assume that $\mathsf{E}Y_1 = \mathsf{E}Y_2 = \ldots = \mathsf{E}Y_n = \mu = 0$. Then we can put $Y_i = e_i, \quad i = 0, 1, 2, \ldots, n$. We use Theorem 4 of Kuelbs and Philipp [21], see Appendix–Theorem A.4.8, who proved for strong-mixing processes that we can redefine the sequence $\{e_i\}$ on a new probability space together with a Brownian motion $W(k)$ such that

$$\left| \sum_{1 \le i \le k} e_i - \sigma W(k) \right| = o(k^{\frac{1}{2} - \beta}) \quad \text{a.s.}$$

with some $\beta > 0$. Then similarly as in Yao and Davis [28] and in the proof of Theorem 4.1.3 in Csörgő and Horváth [7]

$$|S_n(1)| = O_p(1), \tag{4.8}$$

$$\lim_{n\to\infty} P\Big((2\,\sigma^2 \log\log n)^{-1/2} \sup_{0<t\le 1/2} |T_n(t)|/(t(1-t))^{1/2}\Big) =$$
$$\lim_{n\to\infty} P\Big((2\,\sigma^2 \log\log n)^{-1/2} \sup_{0<t\le 1/\log n} |T_n(t)|/(t(1-t))^{1/2}\Big) = 1. \tag{4.9}$$

By the stationarity of $\{e_i, -\infty < i < \infty\}$ we get

$$\lim_{n\to\infty} P\Big((2\,\sigma^2 \log\log n)^{-1/2} \sup_{1/2\le t<1} |T_n(t)|/(t(1-t))^{1/2}\Big) =$$
$$\lim_{n\to\infty} P\Big((2\,\sigma^2 \log\log n)^{-1/2} \sup_{1-1/\log n\le t<1} |T_n(t)|/(t(1-t))^{1/2}\Big) = 1. \tag{4.10}$$

From (4.8), (4.9), (4.10) we conclude

$$\sup_{0<t\le 1/\log n} |T_n(t)|/(t(1-t))^{1/2}$$
$$= \max_{1\le k\le n/\log n} \frac{1}{k^{1/2}} \left|\sum_{i=1}^{k} e_i\right| + O_p((\log\log n)^{1/2}/\log n) \tag{4.11}$$

and

$$\sup_{1-1/\log n<t<1} |T_n(t)|/(t(1-t))^{1/2}$$
$$= \max_{n-n/\log n\le k<n} \frac{1}{(n-k)^{1/2}} \left|\sum_{i=k+1}^{n} e_i\right| + O_p((\log\log n)^{1/2}/\log n). \tag{4.12}$$

Using the limit theorem for standardized partial sums we obtain

$$\lim_{n\to\infty} P\big(A(\log n)\,\frac{1}{\sigma} \max_{1\le k\le n/\log n} \frac{1}{k^{1/2}} \left|\sum_{i=1}^{k} e_i\right| \le u + D(\log n)\big) = \exp(-e^{-u}), \tag{4.13}$$

$$\lim_{n\to\infty} P\big(A(\log n)\,\frac{1}{\sigma} \max_{n-n/\log n\le k<n} \frac{1}{(n-k)^{1/2}} \left|\sum_{i=k+1}^{n} e_i\right| \le s + D(\log n)\big) = \exp(-e^{-s}), \tag{4.14}$$

for all real s and u.
We need to prove that the random variables $\max\limits_{1\le k\le n/\log n} \frac{1}{k^{1/2}} \left|\sum_{i=1}^{k} e_i\right|$ and $\max\limits_{n-n/\log n\le k<n} \frac{1}{(n-k)^{1/2}} \left|\sum_{i=k+1}^{n} e_i\right|$ in (4.13) and (4.14) are asymptotically independent.
But it is easy to see that for strong-mixing processes

$$P\left(\max_{1\le k\le n/\log n} \frac{1}{\sqrt{k}} \left|\sum_{1\le i\le k} e_i\right| < a_n \bigcap \max_{n-n/\log n\le k<n} \frac{1}{\sqrt{n-k}} \left|\sum_{k+1\le i\le n} e_i\right| < a_n\right) -$$
$$-P\left(\max_{1\le k\le n/\log n} \frac{1}{\sqrt{k}} \left|\sum_{1\le i\le k} e_i\right| < a_n\right) P\left(\max_{n-n/\log n\le k<n} \frac{1}{\sqrt{n-k}} \left|\sum_{k+1\le i\le n} e_i\right| < a_n\right) =$$
$$= r_o^{(n-2\frac{n}{\log n})} \to 0 \quad \text{as} \quad n\to\infty \quad \text{for} \quad 0<r_0<1.$$

Hence (4.13) and (4.14) imply

$$\lim_{n\to\infty} P\Big(A(\log n)\,\frac{1}{\sigma}\max_{1\le k\le n/\log n}\frac{1}{k^{1/2}}\left|\sum_{i=1}^{k} e_i\right| \le t + D(\log n),$$
$$A(\log n)\,\frac{1}{\sigma}\max_{n-n/\log n\le k<n}\frac{1}{(n-k)^{1/2}}\left|\sum_{i=k+1}^{n} e_i\right| \le s + D(\log n)\Big)$$
$$= \exp(-e^{-u} - e^{-s}), \tag{4.15}$$

for all real u and s. □

In practice we do not know the value of σ^2 and the question of how to estimate the variance σ^2 in Theorem 4.1.1 is important. We can replace σ^2 with an estimator, where the rate of convergency to σ^2 must be at least $o_p((\log\log n)^{-1})$, which is, according to following lemmas, fulfilled by an estimator:

$$\widehat{\sigma^2} = \widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i), \tag{4.16}$$

where $\widehat{R}(j) = \frac{1}{n}\sum_{i=1}^{n-j}\left(Y_i - \overline{Y_n}\right)\left(Y_{i+j} - \overline{Y_n}\right)$ and $\overline{Y_n} = \frac{1}{n}\sum_{1\le j\le n} Y_j$ and $\psi(n)$ tends to infinity with a certain speed.

First of all we show some lemmas on the estimators $\{\widehat{R}(j), j = 0,1,2,\ldots\}$. In order to do this, it is simpler to work with the functions

$$\widetilde{R}(j) = \frac{1}{n}\sum_{i=1}^{n-j}(Y_i-\mu)(Y_{i+j}-\mu) = \frac{1}{n}\sum_{i=1}^{n-j} e_i\, e_{i+j} \quad j = 0,1,2,\ldots \tag{4.17}$$

which, according to a following lemma, has the same asymptotic properties as the sample autocovariance function $\widehat{R}(j)$.

Lemma 4.1.2.

$$\widehat{R}(0) - \widetilde{R}(0) = O_p(1/n)$$
$$\widehat{R}(j) - \widetilde{R}(j) = O_p(j/n), \quad j = 1,2,\ldots$$

Proof. We can write

$$\widehat{R}(0) = \frac{1}{n}\sum_{i=1}^{n} e_i^2 - \overline{e_n}^2$$

and for each $j = 1,2,\ldots,n$

$$\widehat{R}(j) = \frac{1}{n}\sum_{i=1}^{n-j} e_i\, e_{i+j} - \overline{e_n}^2 + \overline{e_n}\left[\frac{1}{n}\sum_{i=n-j+1}^{n} e_i + \frac{1}{n}\sum_{i=1}^{j} e_i - \frac{j}{n}\overline{e_n}\right].$$

Since

$$\overline{e_n} = O_p(1/\sqrt{n}) \tag{4.18}$$

and

$$(e_i + \ldots + e_{i+j})/n = O_p(1/\sqrt{n}) \quad \text{for} \quad 1 \le i \le n, \quad 1 \le i+j \le n, \tag{4.19}$$

we conclude

$$\begin{aligned} \widehat{R}(0) &= \widetilde{R}(0) + O_p(1/n) \quad \text{as} \quad n \to \infty, \\ \widehat{R}(j) &= \widetilde{R}(j) + O_p(j/n) \quad \text{as} \quad n \to \infty. \end{aligned}$$

□

For random variables $\{e_i = (Y_i - \mu), i = 0, 1, 2, \ldots\}$ we denote their correlation coefficients

$$\rho(j) := \mathsf{E} e_i e_{i+j}$$

and applying Corollary A.4.5 – a result for moment inequalities for strong-mixing random sequences – we obtain a following corollary.

Corollary 4.1.3. *Let $e_m e_{m+i}$ and $e_{m+i+k} e_{m+i+k+j}$ are bounded random variables, then for all $m, i, j \in \mathbb{Z}$ and $k \ge 0$*

$$|\mathsf{E} e_m e_{m+i} e_{m+i+k} e_{m+i+k+j} - \rho_i \, \rho_j| \le K\alpha(k). \tag{4.20}$$

Next lemma asses magnitude of bias for $\widetilde{R}(0), \widetilde{R}(j)$ and $\widetilde{R}(i)\widetilde{R}(j)$.

Lemma 4.1.4. *Assume random variables $\{e_i, i = 0, 1, 2, \ldots, n\}$ satisfy conditions* (4.1), (4.2)*. Then*

$$\begin{aligned} \mathsf{E}\{\widetilde{R}(0)\}^2 &= (\rho(0))^2 + O\left(\frac{1}{n}\right), && (4.21) \\ \mathsf{E}\{\widetilde{R}(j)\}^2 &= (\rho(j))^2 + O\left(\frac{j}{n}\right) \quad j = 1, 2, \ldots && (4.22) \\ \mathsf{E}\{\widetilde{R}(i)\widetilde{R}(j)\} &= \rho(i)\rho(j) + O\left(\frac{i+j}{n}\right) \quad i = 0, 1, 2, \ldots, \quad j > i. && (4.23) \end{aligned}$$

Proof. We start with proving (4.22), as relation (4.21) is in fact a part of (4.22). At first we study the estimators

$$\widetilde{R}(j)^2 = \frac{1}{n^2} \left(e_1 e_{j+1} + \ldots + e_{n-j} e_n\right)^2. \tag{4.24}$$

The proof is based on the fact that the dependance between the variables decreases with their distance in time exponentially fast. Therefore, the variables that are distant behave "almost" as independent. We divide the terms of relation (4.24) into two parts, the first part will contain only "a little" terms written as a product of four factors $e_p e_{p+j} e_q e_{q+j}$, where the third index q satisfies $q \in \{p, \dots, p+j\}$, while for the second part, with majority of terms written as a product of factors $e_p e_{p+j} e_q e_{q+j}$ with the third index q satisfying $q \notin \{p, \dots, p+j\}$ and this part will be estimated according to inequality (4.20). According to this description we write then

$$(e_1 e_{j+1} + e_2 e_{j+2} + \ldots + e_{n-j} e_n)^2 = P_1 + P_2, \tag{4.25}$$

where

$$P_1 = \sum_{p=1}^{n-j} e_p^2 e_{p+j}^2 + 2 \sum_{p=1}^{n-j-1} \sum_{q=p+1}^{\min(p+j,n-j)} e_p e_{p+j} e_q e_{q+j},$$

constrained by the condition that the third index $q \in \{p, \dots, p+j\}$ and

$$P_2 = 2 \sum_{p=1}^{n-2j} \sum_{q=p+j+1}^{n-j} e_p e_{p+j} e_q e_{q+j},$$

with majority of terms with the third index $q \notin \{p, \dots, p+j\}$.
Part P_1 contains $(n-j) + 2j(n-2j) + (j-1)j$ terms $e_p e_{p+j} e_q e_{q+j}$.
According to the Schwarz inequality and the assumption that $\mathsf{E}|e_i|^4 \le K, i = 0, 1, 2, \dots, n$ it holds

$$E|e_p e_{p+i} e_q e_{q+j}| < \infty \quad \text{for every} \quad p, i, j, q \in \mathbb{N}$$

and it implies

$$\frac{\mathsf{E}(P_1)}{n^2} = O(1/n). \tag{4.26}$$

Part P_2 contains $(n-2j)(n-2j-1)$ terms which can be written as a sum

$$\begin{aligned}
&2\, e_1 e_{j+1} e_{j+2} e_{2j+2} + \ldots + 2\, e_1 e_{j+1} e_{n-j-1} e_{n-1} + 2\, e_1 e_{j+1} e_{n-j} e_n + \\
&2\, e_2 e_{j+2} e_{j+3} e_{2j+3} + \ldots + 2\, e_2 e_{j+2} e_{n-j} e_n + \\
&\vdots \\
&2\, e_{n-2j-1} e_{n-j-1} e_{n-j} e_n.
\end{aligned}$$

We can notice that the first column contains $(n-2j-1)$ terms $e_p e_{p+j} e_q e_{q+j}$, where $p = 1, \dots, n-2j-1, \quad q = p+j+1$ and applying (4.20) for each term in the first column we obtain

$$2\, |\mathsf{E} e_p e_{p+j} e_{p+j+1} e_{p+2j+1} - (\rho(j))^2\, | \le 2K\alpha(1).$$

Similarly the second column contains $(n-2j-2)$ terms $e_p e_{p+j} e_{p+j+2} e_{p+2j+2}$, where $p = 1, \dots, n-2j-2$ and for each term in the second column we have

$$2\, |\mathsf{E} e_p e_{p+j} e_{p+j+2} e_{p+2j+2} - (\rho(j))^2\, | \le 2K\alpha(2).$$

The last column contains one term $2 e_1 e_{j+1} e_{n-j} e_n$ for which

$$2\, |\mathsf{E} e_1 e_{j+1} e_{n-j} e_n - (\rho(j))^2\, | \le 2K\alpha(n-2j-1).$$

We sum all the terms in all the columns and obtain $(n-2j)(n-2j-1)$ terms of part P_2, for which

$$\left|\mathsf{E}P_2 - (n-2j)(n-2j-1)\left(\rho(j)\right)^2\right| \leq$$
$$\leq 2K \sum_{k=1}^{n-2j-1} \alpha(k)(n-2j-k) = 2K \sum_{k=1}^{n-2j-1} r_0^k(n-2j-k) = O(n).$$

implying

$$\frac{\mathsf{E}P_2}{n^2} = \rho(j)^2 + \left(\frac{4j^2+2j}{n^2} - \frac{4j+1}{n}\right)\left(\rho(j)\right)^2 + O(\frac{1}{n}) = \left(\rho(j)\right)^2 + O(\frac{j}{n}). \tag{4.27}$$

Substituting (4.26) and (4.27) into (4.25) gives

$$\frac{1}{n^2}\big(\mathsf{E}(P_1+P_2)\big) = \left(\rho(j)\right)^2 + O(\frac{j}{n})$$

and the expectation $\mathsf{E}\{\widetilde{R}(j)\}^2$ fulfills (4.22).

The proof of (4.23) is very similar. For simplicity suppose $i < j$. We again divide the terms of

$$\widetilde{R}_i\widetilde{R}_j = \frac{1}{n^2}\left(e_1e_{i+1} + \ldots + e_{n-i}e_n\right)\left(e_1e_{j+1} + \ldots + e_{n-j}e_n\right) \tag{4.28}$$

into two parts. The first part containing "a little" terms written as a product of four factors $e_pe_{p+i}e_qe_{q+j}$, where the third index q satisfies $q \in \{p,\ldots,p+i\}$ or where the first index p satisfies $p \in \{q,\ldots,q+j\}$. While for the second part, with majority of terms written as a product of factors $e_pe_{p+i}e_qe_{q+j}$, the third index q satisfies $q \notin \{p,\ldots,p+i\}$ nor the first index p satisfies $p \notin \{q,\ldots,q+j\}$, and this part will be estimated according to inequality (4.20).
According to this description we write then

$$\left(e_1e_{i+1} + \ldots + e_{n-i}e_n\right)\left(e_1e_{j+1} + \ldots + e_{n-j}e_n\right) = Q_1 + Q_2, \tag{4.29}$$

where

$$Q_1 = \sum_{p=1}^{n-j} e_pe_{p+i}e_pe_{p+j} + \sum_{p=1}^{n-j-1}\sum_{q=p+1}^{\min(p+i,n-j)} e_pe_{p+i}e_qe_{q+j} +$$
$$+\sum_{p=1}^{n-i}\sum_{q=\max(1,p-j)}^{p-1} e_pe_{p+i}e_qe_{q+j}$$

with the third index satisfying $q \in \{p,\ldots,p+i\}$ (this is fulfilled for the first two sums on the right) or the first index $p \in \{q,\ldots,q+j\}$ (it holds true for the third sum on the right)
and

$$Q_2 = \sum_{p=1}^{n-j-1}\sum_{q=p+j+1}^{n-j} e_pe_{p+i}e_qe_{q+j} + \sum_{p=j+2}^{n-i}\sum_{q=1}^{p-j-1} e_pe_{p+i}e_qe_{q+j},$$

for which the third index $q \notin \{p, \dots, p+i\}$ (the first sum on the right) or the first index $p \notin \{q, \dots, q+j\}$ (the second sum on the right).
Part Q_1 contains $(n-j) + j(n-2j) + \frac{(j-1)j}{2} + \frac{(j+i+1)(j-i)}{2} + i(n-i-j) + \frac{(i-1)i}{2}$ terms $e_p e_{p+i} e_q e_{q+j}$. According to the Schwarz inequality and the assumption that $\mathsf{E}|e_i|^4 \leq K, i = 0, 1, 2, \dots, n$ these terms satisfy

$$E|e_p e_{p+j} e_q e_{q+j}| < \infty \quad \text{for every} \quad p, i, j, q \in \mathbb{N}$$

and it implies

$$\frac{\mathsf{E}Q_1}{n^2} = O(\frac{1}{n}). \tag{4.30}$$

Part Q_2 contains $(n-i-j)(n-i-j-1)$ terms which can be written as a sum

$$\begin{aligned}
&e_1 e_{i+1} e_{i+2} e_{i+j+2} + e_1 e_{i+1} e_{i+3} e_{i+j+3} + \dots + e_1 e_{i+1} e_{n-j} e_n + \\
&e_2 e_{i+2} e_{i+3} e_{i+j+3} + e_2 e_{i+2} e_{i+4} e_{i+j+4} + \dots + e_2 e_{i+2} e_{n-j} e_n + \\
&\vdots \\
&e_{n-i-j-1} e_{n-j-1} e_{n-j} e_n,
\end{aligned}$$

and

$$\begin{array}{cccc}
 & & & e_1 e_{1+j} e_{j+2} e_{j+2+i} + \\
 & & & \vdots \\
 & + e_1 e_{1+j} e_{n-i-1} e_{n-1} & + \dots + & e_{n-i-j-2} e_{n-i-2} e_{n-i-1} e_{n-1} + \\
+ e_1 e_{1+j} e_{n-i} e_n & + e_2 e_{2+j} e_{n-i} e_n & + \dots + & e_{n-i-j-1} e_{n-i-1} e_{n-i} e_n.
\end{array}$$

We can notice that in the first part the first column contains $(n-i-j-1)$ terms $e_p e_{p+i} e_q e_{q+j}$ with the third index $q = p+i+1$ and in the second part the last column contains $(n-i-j-1)$ terms $e_p e_{p+j} e_q e_{q+i}$ with the third index $q = p+j+1$. Applying (4.20) for each term in the first column in the first part we obtain

$$|\mathsf{E}e_p e_{p+i} e_{p+i+1} e_{p+i+1+j} - \rho(i)\rho(j)| \leq K\alpha(1)$$

and for each term in the last column in the second part

$$|\mathsf{E}e_p e_{p+j} e_{p+j+1} e_{p+j+1+i} - \rho(i)\rho(j)| \leq K\alpha(1).$$

Similarly the second column in the first part contains $(n-i-j-2)$ terms $e_p e_{p+i} e_{p+i+2} e_{p+i+2+j}$ and in the second part the last but one column contains $(n-i-j-2)$ terms $e_p e_{p+j} e_{p+j+2} e_{p+j+2+i}$. Applying (4.20) for each term in the second column of the first part we obtain

$$|\mathsf{E}e_p e_{p+i} e_{p+i+2} e_{p+i+2+j} - \rho(i)\rho(j)| \leq K\alpha(2)$$

and for each term in the last but one column of the second part

$$|\mathsf{E}e_p e_{p+j} e_{p+j+2} e_{p+j+2+i} - \rho(i)\rho(j)| \leq K\alpha(2).$$

The last column in the first part contains one term $e_1e_{i+1}e_{n-j}e_n$ and the first column in the second part contains one term $e_1e_{1+j}e_{n-i}e_n$ for which

$$\mathsf{E}|e_1e_{1+i}e_{n-j}e_n - \rho(i)\rho(j)| \leq K\alpha(n-i-j-1),$$
$$\mathsf{E}|e_1e_{1+j}e_{n-i}e_n - \rho(i)\rho(j)| \leq K\alpha(n-i-j-1).$$

We sum all the terms in all the columns and obtain $(n-i-j)(n-i-j-1)$ terms of part Q_2, for which

$$\begin{aligned}&|\mathsf{E}(Q_2) - (n-i-j)(n-i-j-1)\rho(i)\rho(j)|\\ &\qquad\leq 2K\sum_{k=1}^{n-i-j-1}\alpha(k)(n-i-j-k) = 2K\sum_{k=1}^{n-i-j-1} r_0^k(n-i-j-k) = O(n),\end{aligned}$$

implying

$$\begin{aligned}\frac{\mathsf{E}(Q_2)}{n^2} &= \rho(i)\rho(j) - \left(\frac{i^2+2ij+j^2+i+j}{n^2} - \frac{2i+2j+1}{n}\right)\\ &= \rho(i)\rho(j) + O(\frac{i+j}{n}).\end{aligned} \tag{4.31}$$

Substituting (4.30) and (4.31) into (4.29) gives

$$\frac{1}{n^2}\big(\mathsf{E}(Q_1+Q_2)\big) = \rho(i)\rho(j) + O(\frac{i+j}{n})$$

and the expectation $\mathsf{E}\{\widetilde{R}(i)\widetilde{R}(j)\}$ fulfills (4.23). □

Summarizing results of Lemma 4.1.2 and Lemma 4.1.4, we obtain a following corollary for the estimator of the variance $\widehat{\sigma_n^2} = \widehat{R}(0) + 2\sum_{i=1}^{\psi(n)}\widehat{R}(i)$.

Corollary 4.1.5. *For any sequence $\{\psi(n)\}$, $\psi(n) \in \mathbb{N}$, $\psi(n) \leq n$ it holds*

$$\mathsf{E}\left(\widehat{R}(0) + 2\sum_{i=1}^{\psi(n)}\widehat{R}(i)\right) - \left(\rho(0) + 2\sum_{i=1}^{\psi(n)}\rho(i)\right) = O(1/n), \tag{4.32}$$

$$\mathsf{E}\left(\widehat{R}(0) + 2\sum_{i=1}^{\psi(n)}\widehat{R}(i)\right)^2 - \left(\rho(0) + 2\sum_{i=1}^{\psi(n)}\rho(i)\right)^2 = O\left(\frac{(\psi(n))^3}{n}\right). \tag{4.33}$$

Proof. First we prove (4.32). For $i = 0, 1, \dots$ it holds

$$\widehat{R}(i) \leq |\widehat{R}(i) - \widetilde{R}(i)| + |\widetilde{R}(i) - \rho(i)| + \rho(i). \tag{4.34}$$

For the first term on the right of (4.34) we have, according to Lemma 4.1.2

$$\mathsf{E}|\widehat{R}(0) - \widetilde{R}(0)| = O(1/n)$$
$$\mathsf{E}|\widehat{R}(j) - \widetilde{R}(j)| = O(j/n), \quad j = 1, 2, \dots$$

For the second term on the right of (4.34)

$$\mathsf{E}\widetilde{R}(0) = \rho(0) + O(1/n),$$

and for $i = 1, 2 \dots$

$$\mathsf{E}\widetilde{R}(i) = \frac{n-i}{n}\rho(i) = \rho(i) + O(i\rho(i)/n).$$

Then

$$\mathsf{E}\left(\widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i)\right) = \left(\rho(0) + 2\sum_{i=1}^{\psi(n)} \rho(i)\right) + O(\sum_{i=1}^{\psi(n)} \frac{i\rho(i)}{n})$$

Assumption (4.1) on correlation coefficients implies convergency of a series $\sum_{i=1}^{\psi(n)} i\rho(i)$ yielding in relation (4.32).
Now we shall prove (4.33). Substituting equations (4.21), (4.22), (4.23) into the expectation

$$\mathsf{E}\left(\widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i)\right)^2$$

we obtain (4.33) as

$$\mathsf{E}\left(\widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i)\right)^2 = \mathsf{E}\left(\widehat{R}(0)\right)^2 + 4\sum_{i=1}^{\psi(n)} \mathsf{E}\left(\widehat{R}(0)\widehat{R}(i)\right)$$
$$+4\sum_{i=1}^{\psi(n)} \mathsf{E}\left(\widehat{R}(i)\right)^2 + 8\sum_{i=1}^{\psi(n)}\sum_{j>i}^{\psi(n)} \mathsf{E}\left(\widehat{R}(i)\widehat{R}(j)\right) =$$
$$= \left(\rho(0) + 2\sum_{i=1}^{\psi(n)} \rho(i)\right)^2 + O\left(\frac{(\psi(n))^3}{n}\right).$$

□

Now we are ready to prove a theorem which gives the rate of convergency of the proposed estimators $\widehat{\sigma_n^2} = \widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i)$ to the variance σ^2. For simplicity we introduce

$$\widetilde{\sigma_n^2} = \rho(0) + 2\sum_{i=1}^{\psi(n)} \rho(i).$$

Theorem 4.1.6. *For any sequence* $\psi(n) \longrightarrow \infty$ *such that* $\frac{(\psi(n))^3 (\log\log n)^2}{n} \longrightarrow 0$ *it holds*

$$\left| \left(\widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i) \right) - \left(\mathsf{E}(Y_0-\mu)^2 + 2\sum_{i=1}^{\infty} \mathsf{E}(Y_0-\mu)(Y_i-\mu) \right) \right| =$$
$$= o_p((\log\log n)^{-1}).$$

Proof. Assumption (4.1) yields $\sum_{i=1}^{\infty} \rho(i) \leq \sum_{i=1}^{\infty} C r_0^i < \infty$ and it implies it is sufficient to prove

$$\left| \left(\widehat{R}(0) + 2\sum_{i=1}^{\psi(n)} \widehat{R}(i) \right) - \left(\rho(0) + 2\sum_{i=1}^{\psi(n)} \rho(i) \right) \right| = o_p((\log\log n)^{-1}).$$

We have

$$P\left(\log\log n \left(\widehat{\sigma_n^2} - \widetilde{\sigma_n^2} \right) \geq \varepsilon \right) \leq \frac{\log\log n}{\varepsilon^2} \mathsf{E} \left(\widehat{\sigma_n^2} - \widetilde{\sigma_n^2} \right)^2 \leq$$
$$\leq \frac{(\log\log n)^2}{\varepsilon^2} 2 \left[\mathsf{E}\left(\widehat{\sigma_n^2} - \mathsf{E}\widehat{\sigma_n^2} \right)^2 + [\mathsf{E}\widehat{\sigma_n^2} - \widetilde{\sigma_n^2}]^2 \right] \leq$$
$$\leq \frac{(\log\log n)^2}{\varepsilon^2} \left[O\left(\frac{(\psi(n))^3}{n} \right) + O\left(\frac{1}{n^2} \right) \right],$$

as $\psi(n)$ fulfills $\frac{(\psi(n))^3 (\log\log n)^2}{n} \longrightarrow 0$ we obtain the assertion of Theorem 4.1.6. □

4.2 Application

As was mentioned in Introduction, the aim of the second part of this thesis is to suggest and apply methods for a change/s detection in appearance of unusually hot, resp. cold days. More exactly, we create standardized daily series and count how often our series exceed some high, resp. low levels. The exceedance over high, resp. low level means an appearance of unusually warm, resp. cold temperature for the corresponding calender day. Applying the change-point analysis for dependent data (strong-mixing processes), we try to decide whether the frequency of such days changed.

Figure 33 shows that in summer the autocorrelation coefficients for larger lags are smaller than in winter. However, the difference is not extremely large and we simplify the situation and suppose that the data form a stationary sequence.

For our purposes we produced standardized temperature series. If we denote our data $\{X_i, 1 \leq i \leq n\}$, then the standardized series $\{X_i^s, 1 \leq i \leq n\}$ is obtained as

$$X_i^s = \frac{X_i - \overline{X}_{year,i}}{std(X_{year,i})}, \quad 1 \leq i \leq n, \tag{4.35}$$

where $\overline{X}_{year,i}$, $std(X_{year,i})$ denote a mean and a standard deviation of calender days corresponding to the i^{th} measurement. For example,

$$X_{366}^s = \frac{X_{366} - \overline{X}_{year,366}}{std(X_{year,366})},$$

is the standardized value of 1^{st} January obtained from the value X_{366} by extracting the mean temperature of all 1^{st} January temperatures during all years of observation and this difference is divided by the standard deviation of all values of 1^{st} January.

To define exceedances over thresholds we produce two new time series $Y_1^H, Y_2^H, \dots, Y_n^H$ and $Y_1^C, Y_2^C, \dots, Y_n^C$, where for $i = 1, 2, \dots, n$ we denote

- unusually hot days

$$\begin{aligned} Y_i^H &= 1, \quad X_i^s > h, \\ &= 0, \quad X_i^s \leq h, \end{aligned} \tag{4.36}$$

- unusually cold days

$$\begin{aligned} Y_i^C &= 1, \quad X_i^s < c, \\ &= 0, \quad X_i^s \geq c. \end{aligned} \tag{4.37}$$

Levels h, c are suitably chosen constants.

Assume that standardized data $X_1^s, X_2^s, \dots, X_n^s$ form a stationary ARMA sequence

$$X_i^s = \rho_1 X_{i-1}^s + \rho_2 X_{i-2}^s + \ldots + \rho_p X_{i-p}^s + \theta_1 \epsilon_{i-q} + \ldots + \theta_q \epsilon_{i-1} + \epsilon_i, \tag{4.38}$$

where ϵ_i are i.i.d. random variables with $\mathsf{E}\epsilon_i = 0$, $\mathsf{E}\epsilon_i^2 = \sigma_\epsilon^2$ satisfying conditions (1), (2), (5), (11) from Withers [27], see Appendix – conditions (A.5), (A.6), (A.9), (A.11). Then $\{X_i^s, i = 1, 2, \dots, n\}$ is a linear process which can be represented as

$$X_i^s = \sum_{j=0}^{\infty} g_j \epsilon_{i-j},$$

where

$$g_k = O(k^p r^k),$$

parameter r is defined in condition (11) in Withers [27], see Appendix – condition (A.11). According to Withers [27], see Appendix Corollary A.4.3, X_i^s is also a strong-mixing sequence with mixing coefficients

$$\alpha(k) = O(r_0^{\lambda k}), \quad \text{where} \quad \lambda = \frac{\delta}{1+\delta}, \quad 1 > r_0 > r,$$

and the parameter δ is an exponent in condition (5) in Withers [27], confer Appendix. Under these conditions exceedances over thresholds defined by random variables $\{Y_i^H, i = 1, 2, \dots, n\}$ or $\{Y_i^C i = 1, 2, \dots, n\}$ form strong-mixing sequences with the same mixing coefficients $\alpha(k) = O(r_0^{\lambda k})$ for $1 > r_0 > r$. The testing statistic of our problem has the form

$$T_n(t) = \sup_{0<t<T} \left\{ \frac{\frac{1}{\sqrt{T}} \left| N(t) - \frac{t}{T} N(T) \right|}{\widehat{\sigma} \sqrt{\frac{t}{T}\left(1 - \frac{t}{T}\right)}} \right\}, \tag{4.39}$$

where $N(\cdot)$ is the sum of variables $\sum_{1=1}^{[(n+1)t]} Y_i^H$, resp. $\sum_{1=1}^{[(n+1)t]} Y_i^C$, i.e. the number of exceedances over the level H, resp. C during the period $[0, T]$ and $\widehat{\sigma}^2$ is the estimator of the variance defined at the beginning of this chapter as

$$\widehat{\sigma^2} = \widehat{R}(0) + 2 \sum_{i=1}^{\psi(n)} \widehat{R}(i),$$

where $\widehat{R}(j) = \frac{1}{n} \sum_{i=1}^{n-j} \left(Y_i - \overline{Y_n}\right)\left(Y_{i+j} - \overline{Y_n}\right)$ and $\overline{Y_n} = \frac{1}{n} \sum_{1 \le j \le n} Y_j$ and $\psi(n)$ tends to infinity with a certain speed.
The conditions of Theorem 4.1.1 are fulfilled and as a result the approximate critical values can be obtained by the limit distribution of $T_n(t)$ under H_0.

As an example, we show the results for the standardized Milan series. The other observatories give similar results. To the given standardized data set we find an autoregressive process $AR(20)$. Figure 50 represents the autocorrelation function for the standardized Milan series. The values of the autocorrelation function for small lags decrease exponentially. That suggest that an autoregressive sequence might be a good model. However, for larger lags the autocorrelation function does not die out. This is a typical feature when the sequence has some trend or there is a change (or changes).

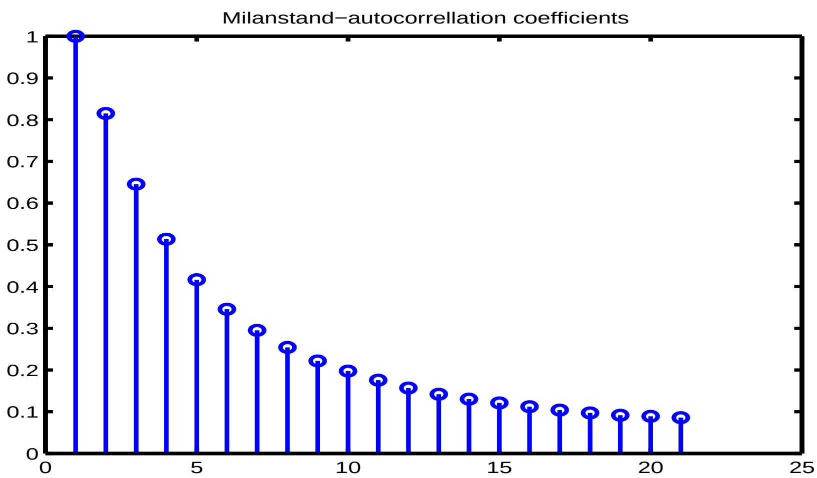

Figure 50. Autocorrelation function for standardized data - Milan.

The estimator of variance was obtained by (4.16), the number of summands was chosen equal to $\psi(n) = 30$.

The levels $h = 2.5$, $c = -2.5$ from (4.36) and (4.37) are symmetric, the values were chosen on purpose to assure sufficient number of exceedances.

For demonstration of the test results we provide two pairs of figures. A pair of graphs in Figure 51 illustrates the case when we proved the change. The left figure shows the plot of the statistic $T_n(t)$ with a noticeable maximum, the right figure shows the sums of exceedances - the vertical line shows the estimated point of the change. A pair of graphs in Figure 52 show similar graphs for the case when H_0 is not rejected, with smaller values of the statistic $T_n(t)$ than in Figure 51.

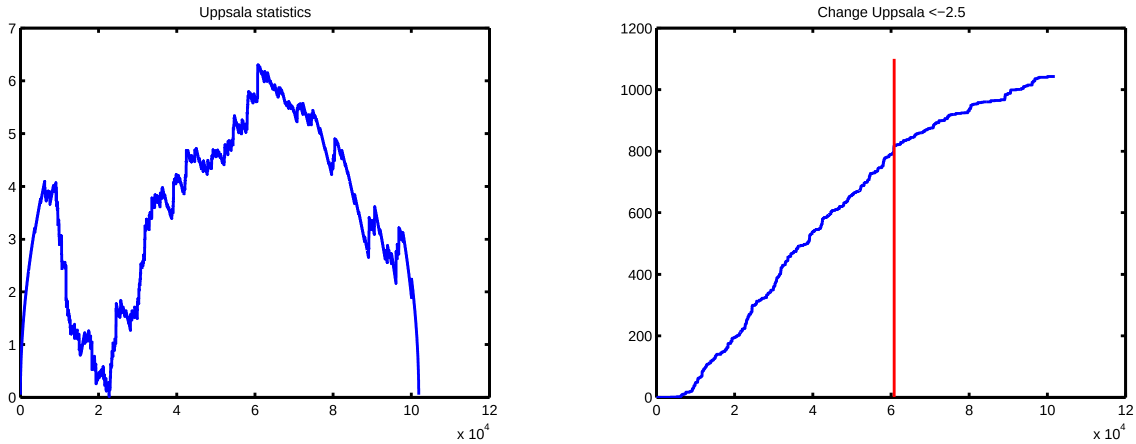

Figure 51. Significant change. Left - the statistic $T_n(t)$ with noticeable maximum, right - the sum of exceedances, the vertical line shows the estimated point of the change.

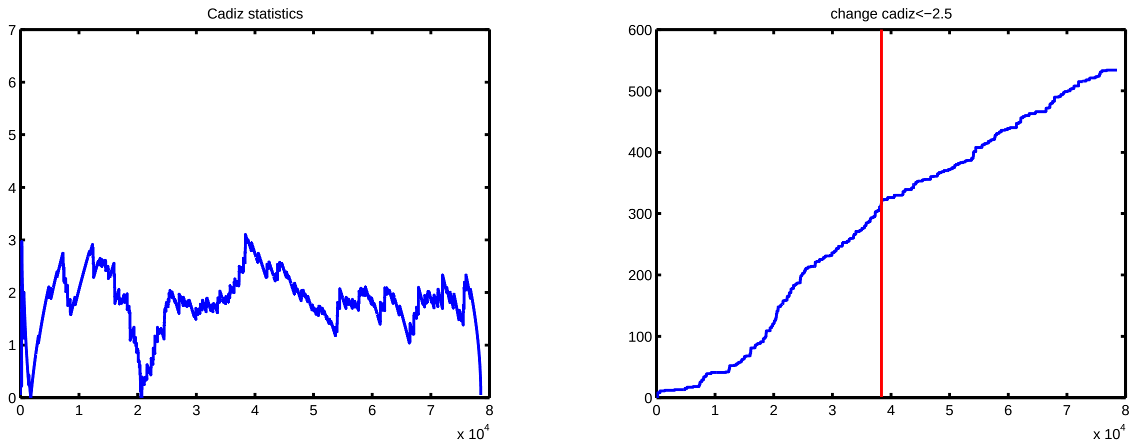

Figure 52. Insignificant change. Left - the statistic $T_n(t)$, right - the sum of exceedances, the vertical line shows the estimated point of the change.

Tables 12 and 13 provide results for appearance of unusually hot days. The values of the statistic $T_n(t)$ and the estimated date of change can be found in Table 12, significant values are marked red. For observatories with significant values of the statistic $T_n(t)$ we also counted the frequency of occurrences of unusually hot days, which can be found in Table 13. We can notice that the values in the first column corresponding to the frequency before the change point are approximately three times smaller than the values in the second column with the estimated frequency of occurrences of unusually hot days after the change point.

	$T_n(t)$	change point
Brussels over 2.5	11.26	21.7.1911
Cadiz over 2.5	5.97	5.6.1923
Milan over 2.5	6.12	29.4.1997
Padua over 2.5	2.67	20.8.1952
St.Petersburg over 2.5	4.53	2.6.1882
Stockholm over 2.5	3.85	17.4.1990
Uppsala over 2.5	3.04	16.6.1989
Prague over 2.5	6.82	11.4.1990

Table 12. First column - values of the statistic $T_n(t)$, second column - estimated date of change.

	first part	second part
Brussels	0.0021	0.014
Cadiz	0.0046	0.0111
Milan	0.0024	0.0233
St. Petersburg	0.0018	0.0049
Stockholm	0.0031	0.0096
Prague	0.0029	0.0119

Table 13. Estimated frequency of occurrences of unusually hot days. First column - before the change point, second column - after the change point.

Tables 14 and 15 provide results for appearance of unusually cold days. The values of the statistic $T_n(t)$ and the estimated date of change can be found in Table 14, significant values are marked red. Table 15 provides for observatories with significant values of the statistic $T_n(t)$ estimated frequencies of occurrences of unusually cold days. We can notice that the values in the first column corresponding to the frequencies before the change point are approximately three times higher than the values in the second column with the estimated frequencies of occurrences of unusually cold days after the change point.

	$T_n(t)$	change point
Brussels under -2.5	3.75	8.12.1819
Cadiz under -2.5	3.10	15.2.1891
Milan under -2.5	3.19	14.1.1880
Padua under -2.5	6.20	18.1.1880
St.Petersburg under -2.5	6.16	27.6.1877
Stockholm under -2.5	7.40	2.8.1888
Uppsala under -2.5	6.30	10.8.1888
Prague under -2.5	4.39	13.3.1943

Table 14. First column - values of the statistic $T_n(t)$, second column - estimated date of change.

	first part	second part
Stockholm	0.0139	0.0041
Uppsala	0.0134	0.0055
Padua	0.0146	0.0058
St. Petersburg	0.0110	0.0040
Prague	0.0097	0.0027

Table 15. Estimated frequency of occurrences of unusually cold days. First column - before the change point, second column - after the change point.

We performed also one comparison – we studied the distribution of unusually hot, resp. cold days within a year. Unfortunately, we were limited to work only with the complete data, i.e. Brussels, Milan, Stockholm, Uppsala and Prague series. Figures 53–58 show differently distributed occurrences of unusually cold days before and after the change point. We studied Prague, Stockholm and Uppsala series with significant values of the testing statistic $T_n(t)$ and we can notice decreasing frequency of unusually cold days in winter months and increasing frequency of unusually cold days in the summer period.

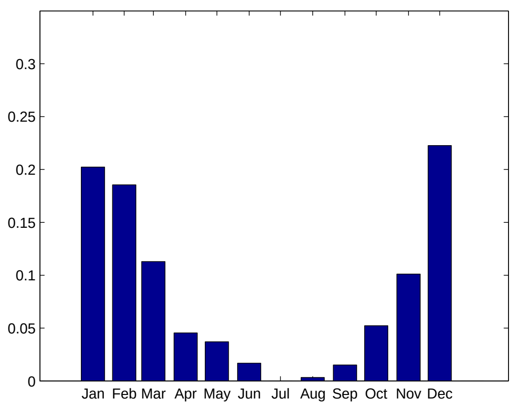

Figure 53. Distribution of unusually cold days in Prague before the change during a year.

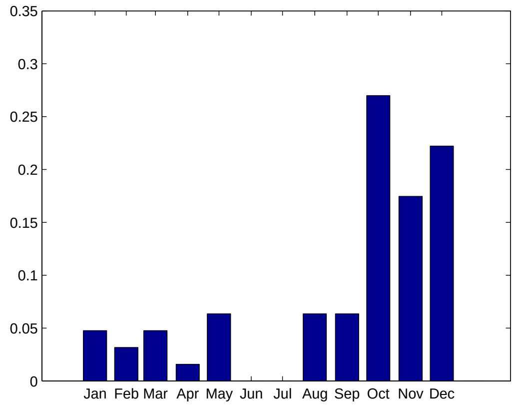

Figure 54. Distribution of unusually cold days in Prague after the change during a year.

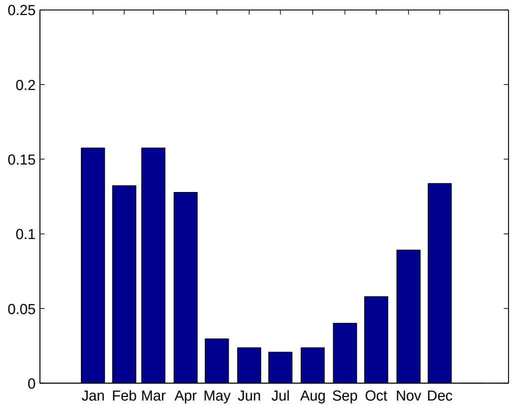

Figure 55. Distribution of unusually cold days in Stockholm before the change during a year.

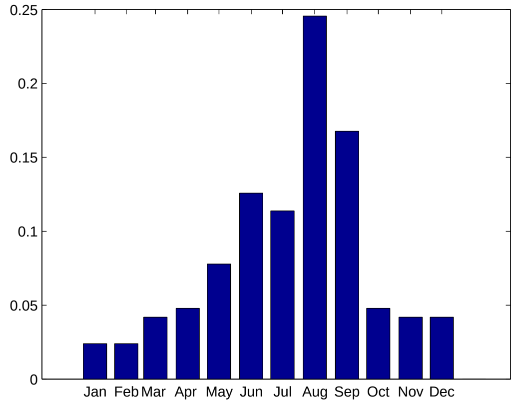

Figure 56. Distribution of unusually cold days in Stockholm after the change during a year.

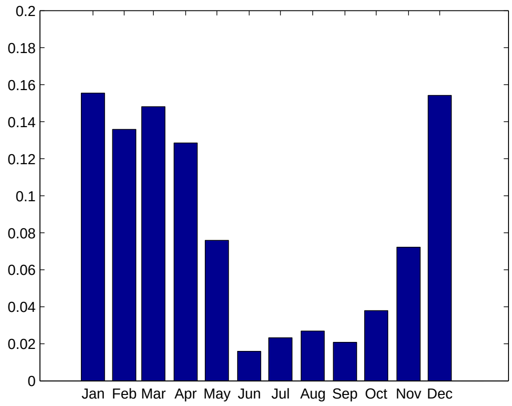

Figure 57. Distribution of unusually cold days in Uppsala before the change during a year.

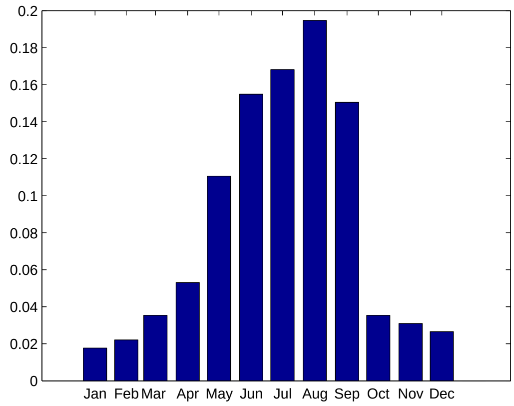

Figure 58. Distribution of unusually cold days in Uppsala after the change during a year.

Figures 59–66 show the distribution of unusually hot days for Brussels, Prague, Milan and Stockholm series with significant values of the statistic $T_n(t)$. We can notice a decreasing frequency of unusually hot days in summer months and an increasing frequency of unusually hot days in winter.

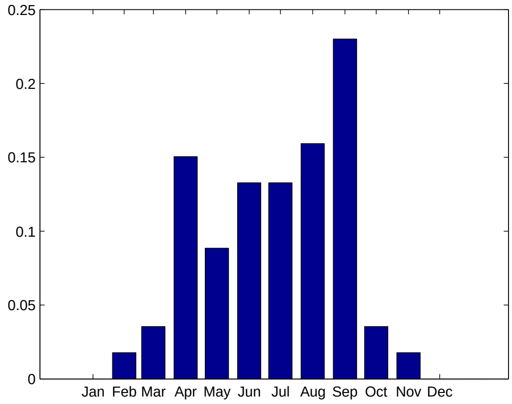

Figure 59. Distribution of unusually hot days in Brussels before the change during a year.

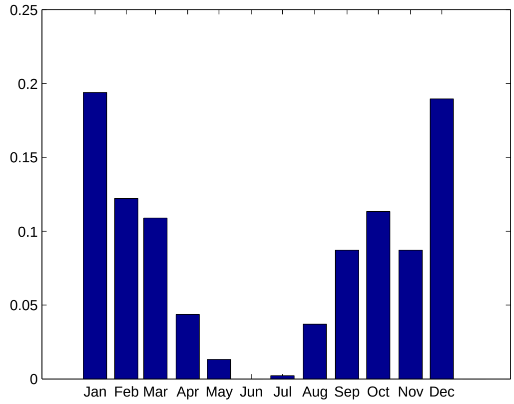

Figure 60. Distribution of unusually hot days in Brussels after the change during a year.

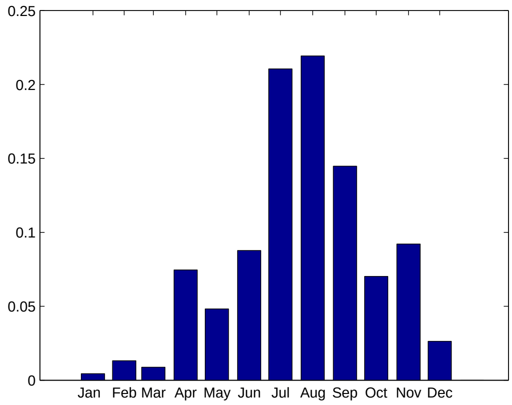

Figure 61. Distribution of unusually hot days in Prague before the change during a year.

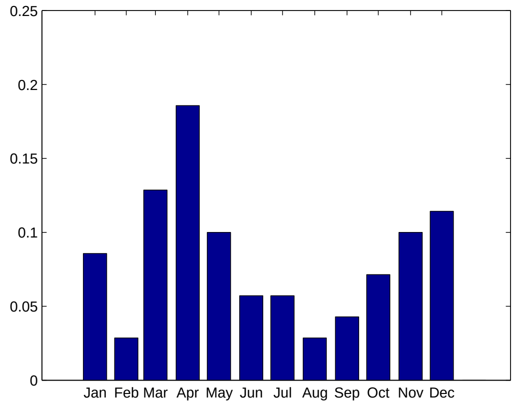

Figure 62. Distribution of unusually hot days in Prague after the change during a year.

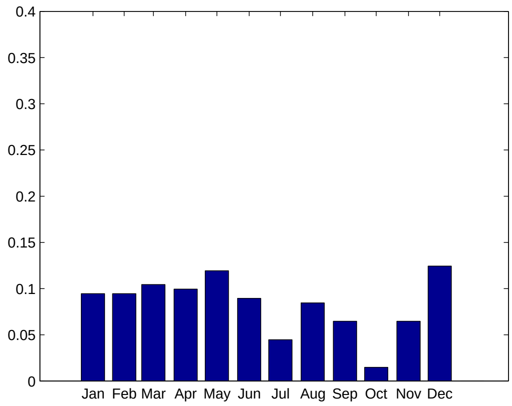

Figure 63. Distribution of unusually hot days in Milan before the change during a year.

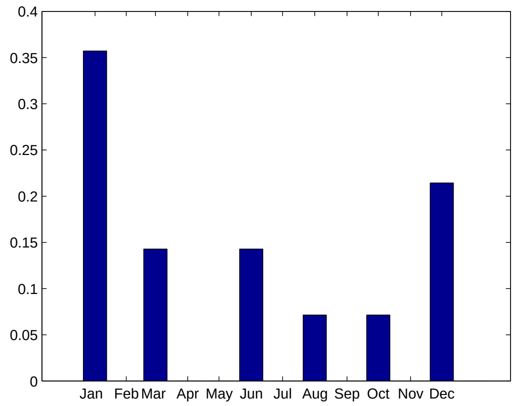

Figure 64. Distribution of unusually hot days in Milan after the change during a year.

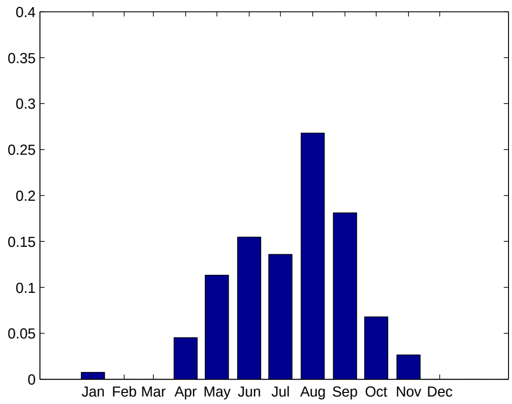

Figure 65. Distribution of unusually hot days in Stockholm before the change during a year.

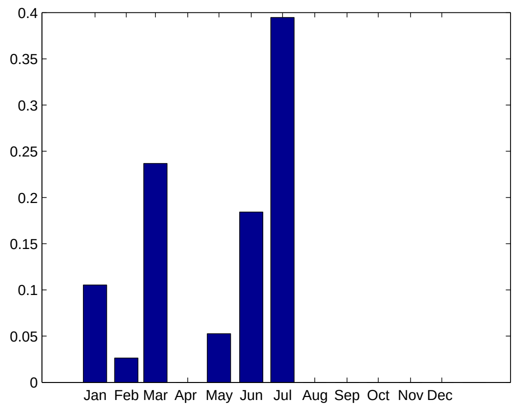

Figure 66. Distribution of unusually hot days in Stockholm after the change during a year.

The changing histograms of unusually hot, resp. cold days within a year raised another question – *Does the number of unusually hot, resp. cold days change within every particular month in a year?* Figures 67–78 provide a detailed numbers of unusually cold days for particular months in Prague series during 230 years of measurement, figures 79–90 show the same for unusually hot days in Brussels series during 232 years of measurement. The red vertical lines are the estimated points of the change.

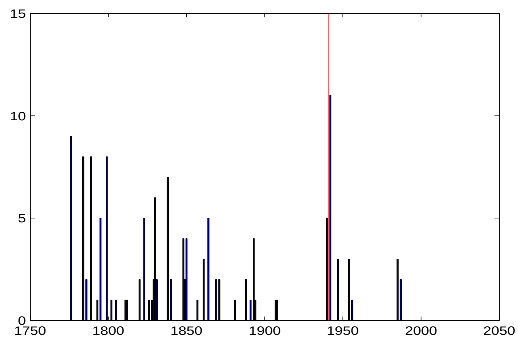

Figure 67. January. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

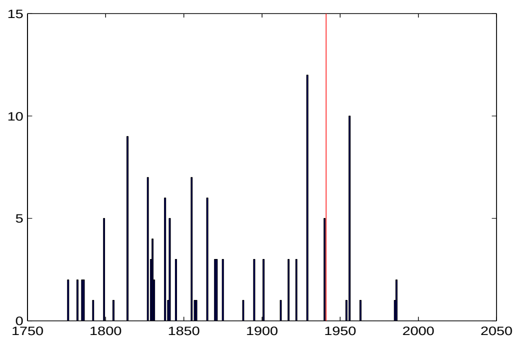

Figure 68. February. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

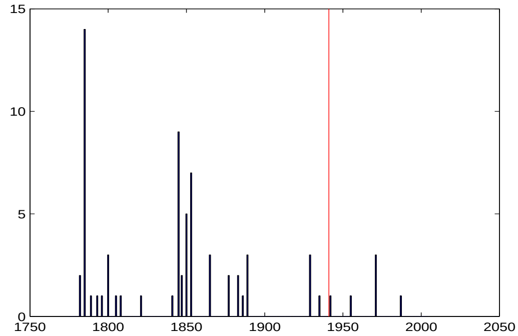

Figure 69. March. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

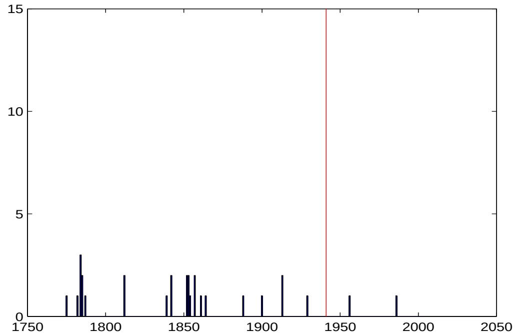

Figure 70. April. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

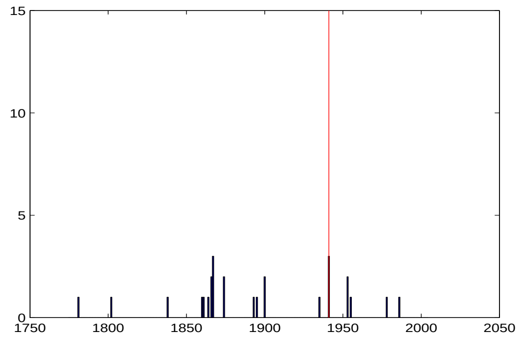

Figure 71. May. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

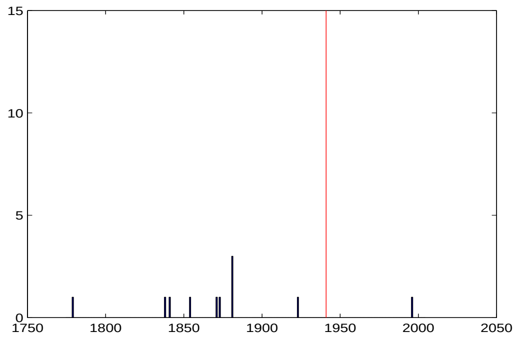

Figure 72. June. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

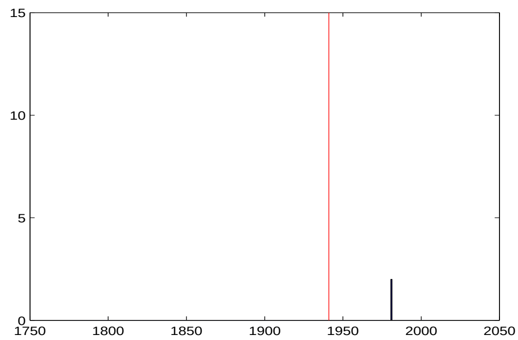

Figure 73. July. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

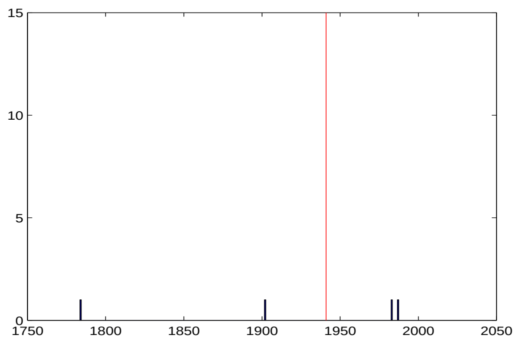

Figure 74. August. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

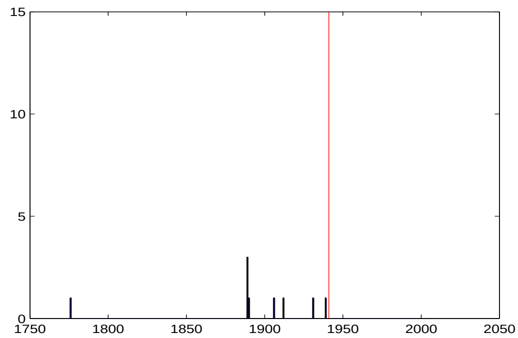

Figure 75. September. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

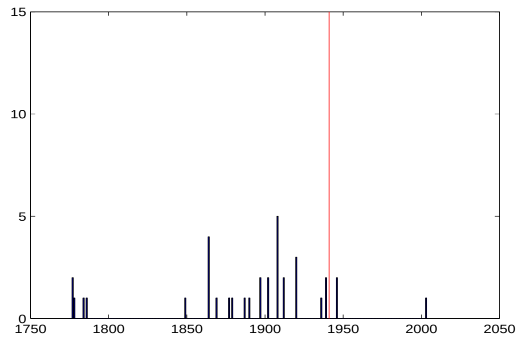

Figure 76. October. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

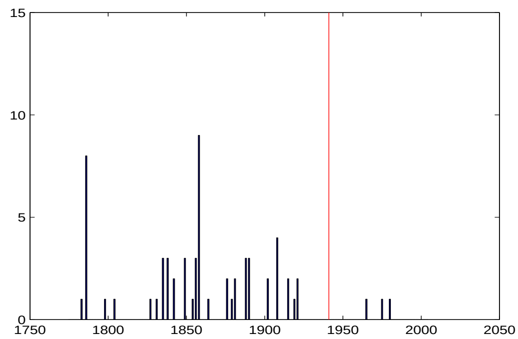

Figure 77. November. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

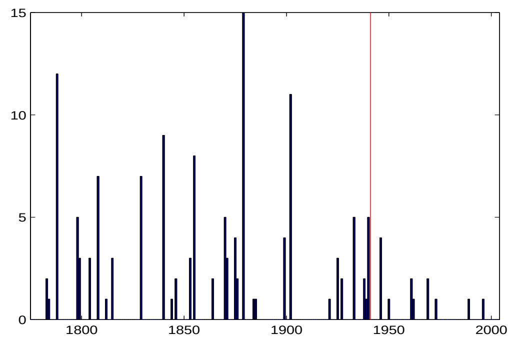

Figure 78. December. The number of unusually cold days in Prague. The red vertical line shows the estimated change point.

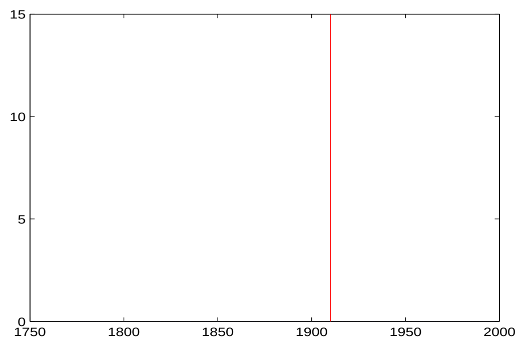

Figure 79. January. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

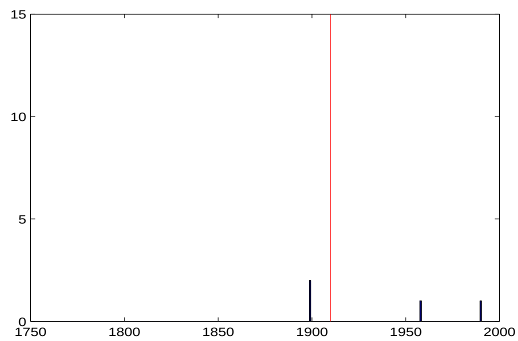

Figure 80. February. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

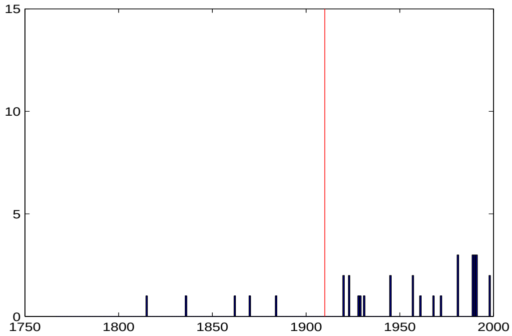

Figure 81. March. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

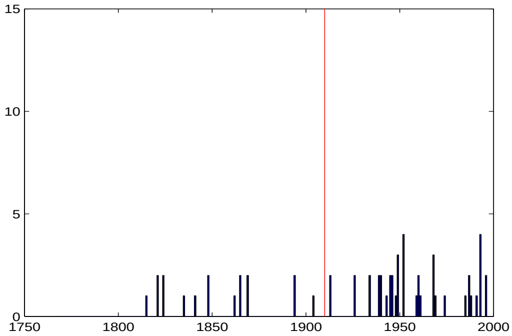

Figure 82. April. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

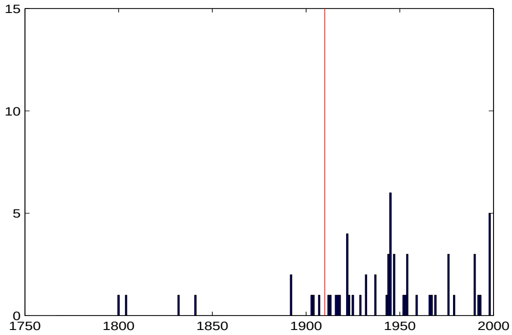

Figure 83. May. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

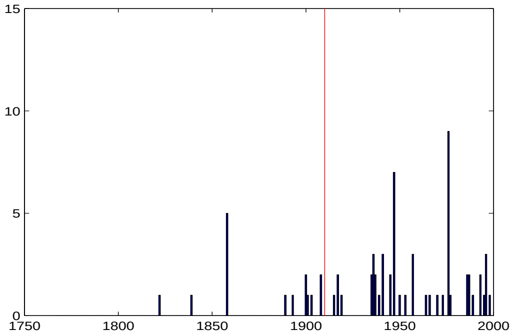

Figure 84. June. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

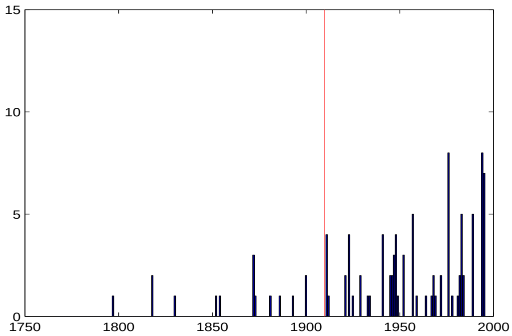

Figure 85. July. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

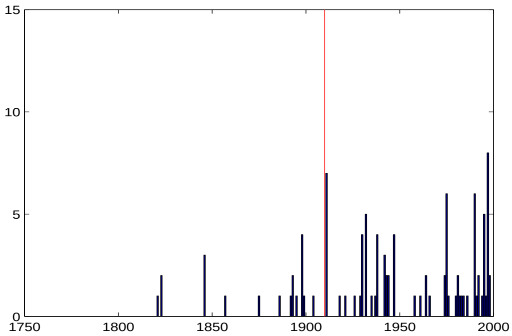

Figure 86. August. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

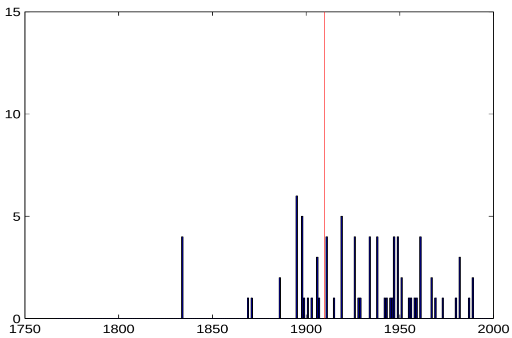

Figure 87. September. The number of unusually Brussels days in Brussels. The red vertical line shows the estimated change point.

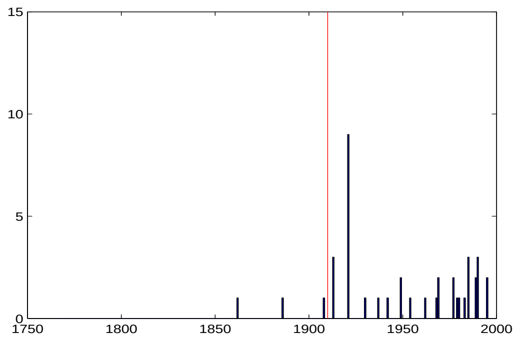

Figure 88. October. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

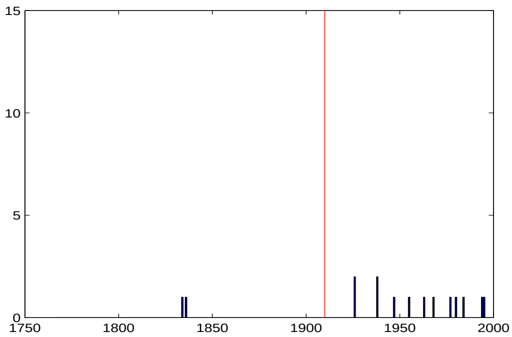

Figure 89. November. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

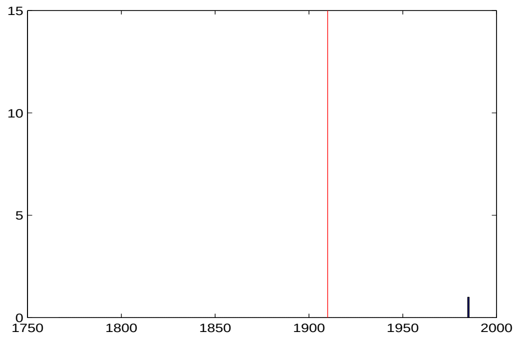

Figure 90. December. The number of unusually hot days in Brussels. The red vertical line shows the estimated change point.

The previous figures suggest decreasing number of unusually cold days, especially for winter months, and an increasing number of unusually hot days during the whole year. For better comparison we counted the mean number of unusually cold days in Prague and the mean number of unusually hot days in Brussels for particular months before and after the change and we obtain a following Table 16 confirming in Prague series a decreasing number of unusually cold days in winter months and an increasing number of unusually hot days within the whole year for Brussels. Similar results were obtained for the other observatories – Milan, Stockholm and Uppsala.

	Prague first part	Prague second part	Brussels first part	Brussels second part
January	0.6707	0.3651	0.0000	0.0000
February	0.6587	0.2381	0.0139	0.0227
March	0.3832	0.0952	0.0347	0.3182
April	0.1617	0.0317	0.1181	0.4886
May	0.1257	0.0794	0.0625	0.5909
June	0.0599	0.0159	0.1042	0.6250
July	0.0000	0.0317	0.1181	0.9886
August	0.0120	0.0317	0.1319	0.9545
September	0.0539	0.0000	0.0486	0.6477
October	0.1916	0.0476	0.0208	0.4205
November	0.3593	0.0476	0.0000	0.0114
December	0.8024	0.2063	0.0097	0.0027

Table 16. The mean numbers of occurrences of unusually cold days in Prague before and after the change in particular months. The mean numbers of occurrences of unusually hot days in Brussels before and after the change in particular months.

4.3 Conclusion

The broadly accepted hypothesis of global warming stimulated an interest for temperature series. Some scientists assume that the change does not necessarily occur in the mean of the series but rather in some other characteristics, e.g. appearance of some extreme events or increase of difference between summer and winter temperatures etc. In the second part of the thesis we concentrated on studying appearances of unusually hot, resp. cold days. More precisely, we were looking for a change in time series of indicators of an event that the standardized value exceeds a certain level. An analogue of Csörgő and Horváth theorem was proved for strong-mixing sequences providing critical values of the limit distribution of T_n under H_0.

When analyzing the exceedances over the level 2.5, the tests confirm a clear increase in the Brussels, Cadiz, Milan, St. Petersburg, Prague and Stockholm series with frequencies of these occurrences three times higher in the second part than before the estimated change, while the increase in Padua and Uppsala occurrences of unusually hot days was not significant. We were also trying to estimate the change point and found out that the change occurred at the end of 19^{th} century or at the beginning of 20^{th} century. There are two exceptions - Milan and Stockholm series, with the estimated change at the end of

20^{th} century. This might be due to different kind of data, as climatologists who analyzed the Milan and Stockholm series tried to remove "heat island effect" in these series, while the authors of the other series were not able to do it, see Camuffo and Jones [6]. We tried to find the second changes in those two series and the significant second change occurred for the Milan series again at the beginning of 20^{th} century. The second change for the Stockholm series was not significant.

For the exceedances under the level -2.5, the tests confirm a clear decrease in the Uppsala, Padua, St. Petersburg, Prague and Stockholm series with frequencies about three times smaller in the second part than before the estimated change, while the decrease in Brussels, Cadiz and Milan occurrences of unusually cold days was not significant. The estimated change point occurred at the end of 19^{th} century.

We also studied the distributions of unusually hot, resp cold days within a year. For observatories with significant values of the statistic $T_n(t)$ we detected also a change in their distributions suggesting for unusually cold days a decreasing frequency in winter months and an increasing frequency in the summer. Distribution of unusually hot days during a year suggest a decreasing frequency in the summer and an increasing frequency in winter months. What concerns the mean numbers of unusually cold days, we detected an increasing mean numbers of unusually hot days in Brussels during the whole year and a decreasing number of unusually cold days in winter months for Prague series.

Although our results might suggest confirmation of the hypothesis that the increased mean of temperature observed since the end of 19^{th} century and a decreasing variability of temperature series is related to the fact that extremely cold days appear less frequent and extremal high temperatures become more frequent, we have to admit several problems which might have influenced our results:

- the number of data. Although we were working with long temperature series, in fact only about 200, resp. 800 observations satisfy our definition of unusually hot, resp. cold day.
- a disproportion between the numbers of unusually hot, resp. cold days is caused by a distinctive negative skewness of minimal temperatures, see Tables 3 and 4.
- heat island effect causing milder winters in city centers.

We hope that in the future the proposed change-point methods should be applied to the series that are not affected by the heat island effect, that are divided into a summer and winter periods and a climatologic definition of an unusually hot, resp. cold day should be taken into account as *a day whose maximum, resp. minimum temperature is within the lowest 5th centile of the daily temperature series for each observatory* and this way we obtain more reasonable results.

Chapter 5

Block permutation

Throughout this chapter it will be more convenient to denote the sequences of random variables by $\{Y(i), i = 1, 2, \dots\}$.

The other way of obtaining critical values of change-point tests is to use the permutation principle. This method was first suggested by Antoch and Hušková [2]. The main tool in deriving this method are limit theorems for rank statistics, see the work by Hušková [15] or Appendix A.5. The sequence $\{Y(i)\}$ is permuted randomly many times and for every permutation the value of $T_n(t)$ is calculated. Then the $(1-\alpha)\,100\%$ empirical quantile of all $\{T_n(t)\}$ serves as $\alpha\,100\%$ critical value. With the number of observations n increasing, the obtained approximate critical values are getting closer and closer to the exact critical values regardless our observations follow the null hypothesis or an alternative. So far, this method has mostly been dealt with independent observations. Kirch [20] considered block permutation principles for dependent data with errors forming a linear process. We examine the behavior of this statistic for strong-mixing sequences.

5.1 Block permutation for strong-mixing sequences

The idea of block permutation is to split the observation sequence of length n into L sequences of length K (i.e. $n = K\,L$), where the block contains enough information about the dependency structure. K and L depend on n and converge to infinity with n. Instead of permuting observations $\{Y(i)\}$, we permute the blocks $Y(Kl+1), \dots, Y(K(l+1))$, $\quad l = 0, \dots, L-1$, and compute the statistics using the permuted blocks.

We consider the following assumptions.

Let $\{X(i), i = \dots, -2, -1, 0, 1, 2, \dots\}$, $\{X^{(1)}(i), i = \dots, -2, -1, 0, 1, 2, \dots\}$, $\{X^{(2)}(i), i = \dots, -2, -1, 0, 1, 2, \dots\}$ form

strictly stationary, strong-mixing sequences with mixing coefficients $\quad \alpha(k) = O(r_0^{-k})$ (5.1) satisfying

$$\begin{aligned} \mathsf{E}X(0) = \mu, \quad \mathsf{E}X^{(1)}(0) &= \mu_1, \quad \mathsf{E}X^{(2)}(0) = \mu_2, \quad d := \mu_2 - \mu_1 \neq 0, \\ \mathsf{E}|X(i)|^\nu &< \infty, \quad \mathsf{E}|X^{(1)}(i)|^\nu < \infty, \quad \mathsf{E}|X^{(2)}(i)|^\nu < \infty \\ \textit{with} \quad \nu &> 4 \quad \textit{for all} \quad i. \end{aligned} \tag{5.2}$$

We consider a sequence $\{Y(i), i = 0, 1, 2, \ldots, n\}$. The hypotheses testing problem may be set as follows:

$$\begin{aligned} &H_0 : Y(i) = X(i), && i = 1, \ldots, n, \qquad (5.3) \\ &H_A : \text{there exists} \quad m^* \in \{1, \ldots, n-1\} \quad \text{such that} && \\ &\quad Y(i) = X^{(1)}(i), && i = 1, \ldots, m^*, \\ &\quad Y(i) = X^{(2)}(i), && i = m^* + 1, \ldots, n. \end{aligned}$$

Remark 5.1.1. *According to Corollary A.4.5, the autocorrelation functions of* $\{X(i)\}$, $\{X^{(1)}(i)\}$ *and* $\{X^{(2)}(i)\}$ *decrease exponentially, i.e.*

$$\begin{aligned} \rho(j) = \mathsf{E}(X(i) - \mu)(X(i+j) - \mu) &= \mathsf{E}(X(0) - \mu)(X(j) - \mu) \\ &\leq Cr_0^j. \end{aligned} \tag{5.4}$$

The same is true for $\{X^{(1)}(i)\}$ *and* $\{X^{(2)}(i)\}$.

We assume that $L \to \infty, K = K(L) \to \infty, n = n(L) = KL$ and $K/L = O(1)$.
Let $\mathbf{R} = (R_1, \ldots, R_L)$ is a random permutation of $(1, \ldots, L)$ independent of $\{Y(\cdot)\}$ chosen with probability $P(\mathbf{R} = r) = \frac{1}{L!}$ for all permutations $r = (r_1, \ldots, r_L)$.

We are interested in the permutation statistic

$$T_{L,K}(Y) := \max_{2 \leq l \leq L-1} \max_{1 \leq k \leq K} \sqrt{\frac{LK}{(K(l-1)+k)\,(LK - K(l-1) - k)}} |S_{L,K}(l,k)|, \tag{5.5}$$

where

$$\begin{aligned} S_{L,K}(l,k) &= \sum_{i=1}^{l-1} \sum_{j=1}^{K} (Y(K(R_i - 1) + j) - \overline{Y}_n) + \sum_{j=1}^{k} (Y(K(R_l - 1) + j) - \overline{Y}_n) \\ \text{and} \quad \overline{Y}_n &= \frac{1}{n} \sum_{i=1}^{n} Y(i). \end{aligned}$$

For the permutation result to hold true we standardize $T_{L,K}(Y)$ using the variance of the block statistic

$$\widehat{\sigma}^2_{LK} = \frac{1}{KL} \sum_{l=0}^{L-1} \left[\sum_{k=1}^{K} (Y(Kl + k) - \overline{Y}_n) \right]^2 .$$

Its advantage is that it does not depend on the permutations, thus the outcome of the permutation test is in fact independent of the actual value of the estimator. We prove that under H_0 the estimator of variance $\widehat{\sigma}^2_{LK}$ converges in probability to

$$\sigma^2 = E(Y(0)^2) + 2\sum_{j=1}^{\infty} \mathsf{E}(Y(0)-\mu)(Y(j)-\mu).$$

Similarly as in the previous chapter, we show some characteristics of the proposed estimator. First of all we have

$$\frac{1}{K\,L}\sum_{l=0}^{L-1}\Big[\sum_{k=1}^{K}\big(Y(Kl+k)-\overline{Y}_n\big)\Big]^2 =$$

$$= \frac{1}{K\,L}\sum_{l=0}^{L-1}\left[\sum_{k=1}^{K}\big(Y(Kl+k)-\mu\big)\right]^2 - K\,(\overline{Y}_n-\mu)^2. \tag{5.6}$$

For the second term in (5.6) it holds

$$K\,(\overline{Y}_n-\mu)^2 = O_p(1/L) \quad \text{as} \quad L\to\infty. \tag{5.7}$$

Therefore we concentrate only on the first part of (5.6), which we denote

$$\widetilde{\sigma}^2_{LK} = \frac{1}{KL}\sum_{l=0}^{L-1}\left[\sum_{k=1}^{K}\big(Y(Kl+k)-\mu\big)\right]^2.$$

If we denote $D_l = \frac{1}{K}\left[\sum_{k=1}^{K}(Y(Kl+k)-\mu)\right]^2$, we can write the estimator $\widetilde{\sigma}^2_{LK}$ as an arithmetic mean of random variables D_l

$$\widetilde{\sigma}^2_{LK} = \frac{1}{L}\sum_{l=0}^{L-1} D_l.$$

For the moments of $\widetilde{\sigma}^2_{LK}$ we have then

$$E\widetilde{\sigma}^2_{LK} = \frac{1}{L}\sum_{l=0}^{L-1}\mathsf{E}D_l, \tag{5.8}$$

$$E(\widetilde{\sigma}^2_{LK})^2 = \frac{1}{L^2}\mathsf{E}\bigg(\sum_{l=0}^{L-1} D_l\bigg)^2. \tag{5.9}$$

The following lemma assesses moments of random variables D_l.

Lemma 5.1.2. *For any $l, l' = 0, 1, \ldots, L-1$, $l \neq l'$ and $K\to\infty$, it holds*

$$\mathsf{E}D_l = \sigma^2 + O(1/K) + O(r_o^K), \tag{5.10}$$

$$\mathsf{E}D_l^2 = O(1), \tag{5.11}$$

$$\mathsf{cov}(D_l, D_{l'}) = O(r_0^{K\,|l-l'-1|}). \tag{5.12}$$

Proof. We start with relation (5.10).

$$\mathsf{E}D_l = \frac{1}{K}\Big(K\rho(0) + 2\,(K-1)\rho(1) + \ldots + 2\,\rho(K-1)\Big) =$$
$$= \left(\rho(0) + 2\sum_{i=1}^{K-1}\rho(i)\right) - \frac{2}{K}\sum_{i=1}^{K-1} i\rho(i) = \sigma^2 + O(r_o^K) + O(1/K),$$

where the last relation is a consequence of convergency of the series $\sum_{i=1}^{\infty} i\,\rho(i)$ with $\rho(i) = r_0^i, \quad 0 < r_0 < 1$.
The proof of relation (5.11) is trivial and follows from Corollary A.4.6 (see Appendix), as

$$\mathsf{E}D_l^2 = \frac{1}{K^2}\mathsf{E}\left(\sum_{k=1}^{K}(Y(Kl+k) - \mu)\right)^4 = O(1).$$

We show the proof of relation (5.12) for $l' = l+1$. The general case $l' \neq l,\, l' = 0, 1, \ldots, L-1$ is then proved similarly. We denote summands of a covariance $\mathsf{cov}(D_l, D_{l+1})$ by

$$S_{i,j,p,q} = \mathsf{E}Y(i)Y(j)Y(p)Y(q) - \big(\mathsf{E}Y(i)Y(j)\big)\big(\mathsf{E}Y(p)Y(q)\big), \qquad \begin{array}{r} i,j = 1,\ldots,K, \\ p,q = K+1,\ldots,2K. \end{array}$$

According to Corollary A.4.5 (see Appendix) we obtain

$$S_{i,j,p,q} \leq M\,\alpha(p-j) = O(r_0^{|p-j|}), \qquad \begin{array}{r} i,j = 1,\ldots,K, \\ p,q = K+1,\ldots,2K. \end{array} \tag{5.13}$$

The idea is to count the number of summands with the mixing coefficient $\alpha(1)$, the number of summands with the mixing coefficient $\alpha(2)$, etc. Therefore we first sort all the K^4 summands $S_{i,j,p,q}, \quad i,j = 1\ldots,K, \quad p,q = K+1\ldots,2K$ into a table of $K^2 \times K^2$ elements and show characteristics of this layout for the index structure.

We create a table containing K^2 rows denoted by pair indexes $(i,j), 1 \leq i \leq K, 1 \leq j \leq K, i \leq j$. The inequality $i \leq j$ is important for the layout, therefore we, for example, transpose the pair $(2,1)$ to the pair $(1,2)$. This way we obtain a sequence of indicated rows:

$$\{(1,1),(1,2),\ldots,(1,K),(1,2),(2,2),\ldots,(2,K),\ldots,(1,K),(2,K),\ldots,(K,K)\}. \tag{5.14}$$

Similarly, K^2 columns we denote by pair indexes $(p,q), K+1 \leq p \leq 2K, K+1 \leq q \leq 2K, p \leq q$ creating a sequence (with a respect to the layout defined by the inequality $p \leq q$)

$$\{(K+1,K+1),(K+1,K+2),\ldots,(K+1,2K),\ldots,$$
$$\ldots,(K+1,2K),(K+2,2K),\ldots,(2K,2K)\}. \tag{5.15}$$

In the sequence of indicated rows (5.14) we have

$$\begin{array}{llll} (2K-1) & \text{indexes of a type} & (i,K), & \text{for} \quad i=1,\ldots,K, \\ (2K-3) & \text{indexes of a type} & (i,K-1), & \text{for} \quad i=1,\ldots,K-1, \\ & \vdots & & \\ 1 & \text{index of a type} & (1,1) & \end{array}$$

and similarly for the sequence of indicated columns (5.15) we get

$$\begin{array}{llll} (2K-1) & \text{indexes of a type} & (K+1,q), & q=K+1,\ldots,2K, \\ (2K-3) & \text{indexes of a type} & (K+2,q), & q=K+2,\ldots,2K, \\ & \vdots & & \\ 1 & \text{index of a type} & (2K,2K). & \end{array}$$

Now we are ready to sort and count the summands $S_{i,j,p,q}$ according to their mixing coefficients. Applying (5.13), we get that certain combinations of rows (i,j) and columns (p,q) result in different summands $S_{i,j,p,q}$ corresponding to the mixing coefficients $\alpha(p-j)$. For example, summands corresponding to the mixing coefficient

$$\begin{array}{lll} \alpha(1) & \text{are of the type} & S_{i,K,K+1,q}, \\ \alpha(2) & \text{are of the type} & S_{i,K,K+2,q}, S_{i,K-1,K+1,q}, \\ \alpha(3) & \text{are of the type} & S_{i,K,K+3,q}, S_{i,K-1,K+2,q}, S_{i,K-2,K+1,q}, \\ & \vdots & \\ \alpha(2K-2) & \text{are of the type} & S_{1,2,2K,2K}, S_{1,1,2K-1,2K}, \\ \alpha(2K-1) & \text{are of the type} & S_{1,1,2K,2K}. \end{array}$$

Using the numbers of rows and the numbers of columns, we get
$(2K-1)\,(2K-1)$ summands $S_{i,K,K+1,q}$, $i=1,\ldots,K$, $q=K+1,\ldots,2K$ with a mixing coefficient $\alpha(1)$,
$(2K-1)\,(2K-3)$ summands $S_{i,K,K+2,q}$, $i=1,\ldots,K$, $q=K+2,\ldots,2K$ with a mixing coefficient $\alpha(2)$,
$(2K-1)\,(2K-5)$ summands $S_{i,K,K+3,q}$, $i=1,\ldots,K$, $q=K+3,\ldots,2K$ with a mixing coefficient $\alpha(3)$,
$\vdots$
$(2K-1)\,1$ summands $S_{i,K,2K,2K}$, $i=1,\ldots,K$ with a mixing coefficient $\alpha(K)$.

Similarly,

$(2K-3)\,(2K-1)$ summands $S_{i,K-1,K+1,q}$, $i=1,\ldots,K-1$, $q=K+1,\ldots,2K$ with

a mixing coefficient $\alpha(2)$,
$(2K-3)\,(2K-3)$summands $S_{i,K-1,K+2,q}$, $i=1,\dots,K-1$, $q=K+2,\dots,2K$ with a mixing coefficient $\alpha(3)$,
$(2K-3)\,(2K-5)$ summands $S_{i,K-1,K+3,q}$, $i=1,\dots,K-1$, $q=K+3,\dots,2K$ with a mixing coefficient $\alpha(4)$,
$\vdots$
$(2K-3)\,1$ summands $S_{i,K-1,2K,2K}$, $i=1,\dots,K-1$ with a mixing coefficient $\alpha(K+1)$.

And finally

$(2K-1)$ summands $S_{1,1,K+1,q}$, $q=K+1,\dots,2K$ with a mixing coefficient $\alpha(K)$,
$(2K-3)$ summands $S_{1,1,K+2,q}$, $q=K+2,\dots,2K$ with a mixing coefficient $\alpha(K+1)$,
$(2K-5)$ summands $S_{1,1,K+3,q}$, $q=K+3,\dots,2K$ with a mixing coefficient $\alpha(K+2)$,
$\vdots$
1 summand $S_{1,1,2K,2K}$, with a mixing coefficient $\alpha(2K-1)$.

For the sum of all these elements we obtain using (5.13) for $l=0,\dots,L-2$

$$\begin{aligned}
\mathsf{cov}(D_l,D_{l+1}) &\le \frac{1}{K^2}\sum_{i=1}^{2K-1}\sum_{j=1}^{i}[2K-(2j-1)]\,\{2K-[2i-(2j-1)]\}M\,\alpha(i)\\
&\le \frac{1}{K^2}\sum_{i=1}^{2K-1}\sum_{j=1}^{i}[2K-(2j-1)]\,\{2K-[2i-(2j-1)]\}M\,r_0^i\\
&\le \frac{1}{K^2}\,M\,4\,K^2\sum_{i=1}^{\infty} i\,r_0^i = O(1),
\end{aligned}$$

which is a consequence of convergency of the series $\sum_{i=1}^{\infty} i\,r_0^i$ with $0<r_0<1$.

The previous procedure can be generalized for l,t such that $0\le l\le L-1$, $0\le l+t\le L-1$, then

$$\begin{aligned}
&\mathsf{cov}(D_l,D_{l+t}) \le\\
&\le \frac{1}{K^2}\sum_{i=1}^{2K-1}\sum_{j=1}^{i}[2K-(2j-1)]\,\{2K-[2i-(2j-1)]\}M\,\alpha(i+(t-1)K)\\
&\le \frac{1}{K^2}\sum_{i=1}^{2K-1}\sum_{j=1}^{i}[2K-(2j-1)]\,\{2K-[2i-(2j-1)]\}M\,r_0^{i+(t-1)K}\\
&\le \frac{1}{K^2}\,M\,4\,K^2\,r_0^{(t-1)K}\sum_{i=1}^{\infty} i\,r_0^i = O\Big(r_0^{(t-1)K}\Big),
\end{aligned}$$

which is the assertion (5.12). □

Corollary 5.1.3. *For K, L such that $L \to \infty, K = K(L) \to \infty, n = n(L) = KL$, $K/L = O(1)$, it holds*

$$\begin{aligned} \mathsf{E}(\widehat{\sigma}^2_{LK}) &= \sigma^2 + O\left(1/\min(K,L)\right), && (5.16) \\ \mathsf{E}(\widehat{\sigma}^2_{LK})^2 &= O(1/L). && (5.17) \end{aligned}$$

Proof. Relation (5.16) follows immediately from (5.6), (5.7), (5.8), (5.10), as

$$\mathsf{E}(\widehat{\sigma}^2_{LK}) = \sigma^2 + O\left(1/K + 1/L + r_o^K\right) = \sigma^2 + O\left(1/\min(K,L)\right).$$

For $\mathsf{E}(\widehat{\sigma}^2_{LK})^2$ we have for $L \to \infty$ according to (5.6), (5.7), (5.9), (5.11), (5.12):

$$\begin{aligned} \mathsf{E}(\widehat{\sigma}^2_{LK})^2 &= \frac{1}{L^2}\sum_{l=0}^{L-1}\left(\mathsf{E}(D_l)^2 + 2\sum_{l'>l}^{L-1}\mathsf{cov}(D_l, D_{l'})\right) \\ &= \frac{1}{L^2}O\left(L + 2\left[L + (L-1)r_0^{1\,K} + (L-2)r_0^{2\,K} + \ldots + r_0^{(L-1)\,K}\right]\right) \\ &= O(\frac{1}{L}). \end{aligned}$$

□

Theorem 5.1.4. *For any K, L such that $L \to \infty, K = K(L) \to \infty, n = n(L) = KL$, $K/L = O(1)$, $K = O((\log n)^\gamma)$ for some $\gamma > 0$ satisfying $\frac{(\log\log n)^2}{\min(K,L)} \to 0$, it holds*

$$\widehat{\sigma}^2_{LK} - \sigma^2 = o_p((\log\log n)^{-1}).$$

Proof. We have according to the Markov inequality and Corollary 5.1.3

$$\begin{aligned} P\left(\log\log n\left(\widehat{\sigma}^2_{LK} - \sigma^2\right) \geq \varepsilon\right) &\leq \frac{(\log\log n)^2}{\varepsilon^2}\mathsf{E}\left(\widehat{\sigma}^2_{LK} - \sigma^2\right)^2 \\ &\leq \frac{(\log\log n)^2}{\varepsilon^2}2\left[\mathsf{E}\left(\widehat{\sigma}^2_{LK} - \mathsf{E}\widehat{\sigma}^2_{LK}\right)^2 + [\mathsf{E}\widehat{\sigma}^2_{LK} - \sigma^2]^2\right] \\ &\leq \frac{(\log\log n)^2}{\varepsilon^2}\left[O\left(\frac{1}{\min(K,L)}\right) + O\left(\frac{1}{L^2}\right)\right] \\ &= \frac{(\log\log n)^2}{\varepsilon^2}\left[O\left(\frac{1}{\min(K,L)}\right)\right]. \end{aligned}$$

According to our assumption, minimum $\min(K,L)$ fulfills $\frac{(\log\log n)^2}{\min(K,L)} \to 0$ as $L \to \infty, \quad K = K(L) \to \infty$ and we obtain the assertion of Theorem 5.1.4. □

In the following theorem we show that the block permutation method gives asymptotically correct critical values, i.e we prove that the quantiles of the permutation statistic $T_{L,K}(Y)$ conditioned on the observations $Y(\cdot)$ approximate the correct critical values as the number of observations tends to infinity. This is true not only when the observations follow the null hypothesis but even when they follow the alternative hypothesis.

Theorem 5.1.5. *$\{X(i), i = 0,1,2,\ldots,n\}$, $\{X^{(1)}(i), i = 0,1,2,\ldots,n\}$, $\{X^{(2)}(i), i = 0,1,2,\ldots,n\}$ fulfill the assumptions (5.1), (5.2). Let $0 < \widetilde{\delta} < (\nu-4)/2$. Under the alternative let either*

(i) $K^{(2+\widetilde{\delta})/2}|d|^{2+\widetilde{\delta}}\min(\frac{m^*}{n}, \frac{n-m^*}{n}) = O(1)$ *and* $\frac{d^2K}{L} = o(1)$ *or*

(ii) $\min(\frac{m^*}{n}, \frac{n-m^*}{n}) \geq \epsilon > 0$.

Let $A(x)$, $D(x)$ be as in Theorem A.5.1. If $K = O((\log n)^\gamma)$ for some $\gamma > 0$ and $K/L = O(1)$ then for all $t \in \mathbb{R}$ as $L \to \infty$

$$\lim_{n\to\infty} P\left(A(\log n)\,\frac{T_{L,K}(\pi, Y)}{\widehat{\sigma}_{LK}} - D(\log n) \leq t | Y(1), \ldots, Y(n)\right) = \exp(-2e^{-t}) \quad a.s.$$

Remark 5.1.6. *The assumptions (i) and (ii) come from Theorem 3.1 in Kirch [20].*

Proof. The idea of the proof is similar to Kirch's proof of Theorem 3.1 in [20] for errors forming a linear sequence, i.e. to apply asymptotics for the rank statistic, confer Appendix – Theorem A.5.1, with special scores (for (i) we choose $a_n(i) := Y(i)$ and for (ii) $a_n(i) := Y(i)/\sqrt{d^2K}$). In a first step we prove that the assumptions (A.13), (A.14) are fulfilled for the error sequence $\{e(i)\} = \{Y(i)\} - \mu$, in a second step we conclude that they are also fulfilled a.s. for $\{Y(i)\}$.

We start with the error sequence $\{e(i)\} = \{Y(i)\} - \mu$ and prove (A.13).
We denote $l^* = \lceil m^*/K \rceil \quad (m^* \neq n)$, where $\lceil x \rceil$ denotes the integer part of x. Then for the block where the change occurs it holds

$$\begin{aligned}
\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl^*+j)\right|^4 &= \left(\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl^*+j)\right|\right)^4 \\
&\leq 8\Bigg\{\left(\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e^{(1)}(Kl^*+j)\right|\right)^4 \\
&\quad + \left(\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e^{(2)}(Kl^*+j)\right|\right)^4\Bigg\},
\end{aligned}$$

which is a consequence of inequality $(a+b)^4 \leq 8(a^4+b^4)$. This gives

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl+j)\right|^4 \\
&\quad\leq 8\Bigg\{\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e^{(1)}(Kl+j)\right|^4 \\
&\qquad + \frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e^{(2)}(Kl+j)\right|^4\Bigg\}.
\end{aligned} \tag{5.18}$$

Sequences $\{\frac{1}{K}\sum_{k=1}^{K} e^{(1)}(Kl+k), l \geq 0\}$, $\{\frac{1}{K}\sum_{k=1}^{K} e^{(2)}(Kl+k), l \geq 0\}$ satisfy assumptions of Theorem A.4.7, confer Appendix, as these sequences remain stationary and their mixing coefficient are smaller than the one of $\{e^{(1)}(i), i = 1, \dots, n\}$, resp. $\{e^{(2)}(i), i = 1, \dots, n\}$ for all K, therefore applying strong law of large numbers we obtain

$$\begin{aligned}
\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\dots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} e^{(1)}(Kl+j) \right|^4 &\leq D_1^{(1)} \quad a.s., \\
\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\dots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} e^{(2)}(Kl+j) \right|^4 &\leq D_1^{(2)} \quad a.s.
\end{aligned}$$

Substituting these inequalities into (5.18) we obtain

$$\frac{1}{L}\sum_{l=0}^{L-1} \max_{k=0,\dots,K-1} \left| \frac{1}{\sqrt{K}} \sum_{j=k+1}^{K} e(Kl+j) \right|^4 < D_1 \tag{5.19}$$

which gives (A.13) for the error sequence $\{e(\cdot)\}$.

Concerning assumption (A.14) for the error sequence $\{e(\cdot)\}$ we have following inequalities:

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1} \left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} e(Kl+j) \right)^2 \\
&\quad \geq \frac{1}{L}\left\{ \sum_{l=0}^{l^*-1} \left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} e^{(1)}(Kl+j) \right)^2 + \sum_{l=l^*+1}^{L-1} \left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} e^{(2)}(Kl+j) \right)^2 \right\} \\
&\quad = \frac{l^*}{L}\frac{1}{l^*}\sum_{l=0}^{l^*-1} \left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} e^{(1)}(Kl+j) \right)^2 \\
&\qquad + \frac{L-l^*-1}{L}\frac{1}{L-l^*-1}\sum_{l=l^*+1}^{L-1} \left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} e^{(2)}(Kl+j) \right)^2 \\
&\quad \geq \max\left\{ \frac{l^*}{L}\frac{1}{l^*}\sum_{l=0}^{l^*-1} \left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} e^{(1)}(Kl+j) \right)^2 ; \right. \\
&\qquad\qquad \left. \frac{L-l^*-1}{L}\frac{1}{L-l^*-1}\sum_{l=l^*+1}^{L-1} \left(\frac{1}{\sqrt{K}} \sum_{j=1}^{K} e^{(2)}(Kl+j) \right)^2 \right\}.
\end{aligned}$$

Expressions $\frac{1}{l^*}\sum_{l=0}^{l^*-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(1)}(Kl+j)\right)^2$ and

$\frac{1}{L-l^*-1}\sum_{l=l^*+1}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(2)}(Kl+j)\right)^2$ are positive, since for $t=1,2$ and any $q\in\mathbb{N}$

$$\frac{1}{q}\sum_{l=0}^{q-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(t)}(Kl+j)\right)^2 = \mathsf{Var}(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(t)}(j))$$
$$+\frac{1}{q}\sum_{l=0}^{q-1}\left[\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(t)}(Kl+j)\right)^2 - \mathsf{Var}(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(t)}(j))\right] \longrightarrow C>0 \quad a.s.,$$

where we either use results of Theorem 5.1.4 ($K\to\infty$) or the fact that $\mathsf{Var}(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(t)}(j)) \geq c>0$, if K is bounded. It is obvious that both limits $\frac{l^*}{L}\to 0$ and $\frac{L-l^*-1}{L}\to 0$ can not converge simultaneously, therefore

$$\max\left\{\frac{l^*+1}{L}\frac{1}{l^*+1}\sum_{l=0}^{l^*}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(1)}(Kl+j)\right)^2;\right.$$
$$\left.\frac{L-l^*-1}{L}\frac{1}{L-l^*-1}\sum_{l=l^*}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e^{(2)}(Kl+j)\right)^2\right\} \geq D_2$$

and the error sequence $\{e(\cdot)\}$ fulfills

$$\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{j=1}^{K} e(Kl+j)\right)^2 \geq D_2, \tag{5.20}$$

which is the assumption (A.14).

Before we start with verification of assumptions (A.13), (A.14) almost surely for $\{Y(\cdot)\}$, we present some strong laws of large numbers for strong-mixing processes $\{e(\cdot)\}$, confer Kirch [20] and Appendix A.4. According to Theorem A.4.7 it holds as $L\to\infty$

$$\frac{1}{n}\left|\sum_{j=1}^{n} e(j)\right| = O\left(\sqrt{\frac{\log n}{n}}\right) \quad a.s. \tag{5.21}$$

To prove (A.13) under alternative we denote $l^* := \lceil m^*/K\rceil$ and derive following laws of large number

$$\frac{1}{K}\left|\sum_{k=1}^{K} e(Kl^*+k)\right| = O\left(\sqrt{\frac{\log K}{K}}\right), \quad a.s. \tag{5.22}$$

as $K\to\infty$. From Markov inequality

$$P\left(\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e(Kl^*+k)\right| \geq \epsilon\right) \leq \frac{K^\nu}{\epsilon^\nu} L^{-\nu/2}\mathsf{E}|e(0)|^\nu$$

Because $\sum_{L=1}^{\infty} L^{-\nu/2}<\infty$ with $\nu>4$, it holds as $L\to\infty$

$$\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e(Kl^*+k)\right| = O\left(\sqrt{\log K}\right) \quad a.s. \tag{5.23}$$

for $K \to \infty$ as well as K bounded. Similarly we deduce

$$\frac{1}{\sqrt{L(n-m^*)}} \left| \sum_{j=Kl^*+1}^{n} e(j) \right| = o(1) \quad a.s. \tag{5.24}$$

Now we are ready to verify the assumptions (A.13), (A.14) almost surely for

$$\begin{aligned} Y(i) &= \mu_1 + e^{(1)}(i), \qquad & i = 1, \dots, m^*, \\ &= \mu_1 + d + e^{(2)}(i), \qquad & i = m^* + 1, \dots, n. \end{aligned}$$

First we consider the alternative (i) and choose the scores $a_n(i) := Y(i)$. Without loss of generality we can assume the $\mu_1 = 0$. We begin with the assumption (A.14). We can write

$$\frac{1}{L}\sum_{l=0}^{L-1} \left(\frac{1}{\sqrt{K}} \sum_{k=1}^{K} (Y(Kl+k) - \overline{Y}_n) \right)^2 = \frac{1}{K\,L} \sum_{l=0}^{L-1} \left(\sum_{k=1}^{K} (Y(Kl+k) \right)^2 - K\,\overline{Y}_n^2. \tag{5.25}$$

Using equation (5.21) for $\overline{e^{(1)}} = \frac{1}{m^*}\sum_{i=1}^{m^*} e^{(1)}(i)$ and $\overline{e^{(2)}} = \frac{1}{n-m^*}\sum_{i=m^*+1}^{n} e^{(2)}(i)$ in $\overline{Y}_n = d\,\frac{n-m^*}{n} + \frac{m^*}{n}\overline{e^{(1)}} + \frac{n-m^*}{n}\overline{e^{(2)}}$ we obtain

$$\begin{aligned} K\,\overline{Y}_n^2 \;&=\; K\,d^2 \left(\frac{n-m^*}{n}\right)^2 + K\left(\frac{m^*}{n}\right)^2 \overline{e^{(1)}}^2 + K\left(\frac{n-m^*}{n}\right)^2 \overline{e^{(2)}}^2 \\ &\quad +2\sqrt{K}\,d\,\frac{(n-m^*)\,m^*}{n^2}\sqrt{K}\,\overline{e^{(1)}} + 2\sqrt{K}\,d\left(\frac{n-m^*}{n}\right)^2 \sqrt{K}\,\overline{e^{(2)}} \\ &\quad +2\,\frac{(n-m^*)\,m^*}{n^2}\sqrt{K}\,\overline{e^{(1)}}\sqrt{K}\,\overline{e^{(2)}} \\ &= K\,d^2\left(\frac{n-m^*}{n}\right)^2 + o\left(\left(\frac{n-m^*}{n}\right)^2\right) + o\left(\frac{(n-m^*)\,m^*}{n^2}\right) + o\left((\frac{m^*}{n})^2\right) \\ &\quad +o\left(\sqrt{K}\,d\,\frac{(n-m^*)\,m^*}{n^2}\right) + o\left(\sqrt{K}\,d\left(\frac{n-m^*}{n}\right)^2\right) + o(1) \quad a.s. \end{aligned} \tag{5.26}$$

and for the first term on the right of equation (5.25) we have as $L \to \infty$

$$\begin{aligned} &\frac{1}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}(d\,1_{\{Kl+k>m^*\}} + e(Kl+k))\right)^2 \\ &= \frac{1}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} e(Kl+k)\right)^2 + \frac{1}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} d\,1_{\{Kl+k>m^*\}}\right)^2 \\ &\quad + \frac{2}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} d\,1_{\{Kl+k>m^*\}}\right)\left(\sum_{k=1}^{K} e(Kl+k)\right) \\ &= \frac{1}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} e(Kl+k)\right)^2 + Kd^2\frac{n-m^*}{n} + o\left(\sqrt{|d|^2 K\frac{n-m^*}{n}}\right) + o(1) \quad a.s., \end{aligned} \tag{5.27}$$

as from the assumption $\frac{d^2 K}{L} = o(1)$ we get for $\frac{1}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} d\,1_{\{Kl+k>m^*\}}\right)^2$

$$\frac{1}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} d\,1_{\{Kl+k>m^*\}}\right)^2 = Kd^2\frac{n-m^*}{n} + o(1) \quad a.s.$$

and equations (5.23), (5.24) imply for

$$\begin{aligned}
&\frac{2}{K\,L}\sum_{l=0}^{L-1}\left(\sum_{j=1}^{K} d\,1_{\{Kl+j>m^*\}}\right)\left(\sum_{k=1}^{K} e(Kl+k)\right)\\
&\leq |d|\,K\frac{1}{n}\left|\sum_{j=Kl^*+1}^{n} e^{(1)}(j)\right| + |d|\,K\frac{1}{n}\left|\sum_{j=Kl^*+1}^{n} e^{(2)}(j)\right|\\
&\quad + K\,|d|\frac{1}{n}\left|\sum_{k=1}^{K} e^{(1)}(Kl^*+k)\right| + K\,|d|\frac{1}{n}\left|\sum_{k=1}^{K} e^{(2)}(Kl^*+k)\right|\\
&\leq |d|\sqrt{K\frac{n-m^*}{n}}\frac{1}{\sqrt{L(n-m^*)}}\left|\sum_{j=Kl^*+1}^{n} e^{(1)}(j)\right|\\
&\quad + |d|\sqrt{K\frac{n-m^*}{n}}\frac{1}{\sqrt{L(n-m^*)}}\left|\sum_{j=Kl^*+1}^{n} e^{(2)}(j)\right|\\
&\quad + \sqrt{K}\,|d|\min\left(\frac{n-m^*}{n},\frac{m^*}{n}\right)^{\frac{1}{2+\tilde{\delta}}}\frac{1}{n^{\frac{\tilde{\delta}}{2(2+\tilde{\delta})}}}\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e^{(1)}(Kl^*+k)\right|\\
&\quad + \sqrt{K}\,|d|\min\left(\frac{n-m^*}{n},\frac{m^*}{n}\right)^{\frac{1}{2+\tilde{\delta}}}\frac{1}{n^{\frac{\tilde{\delta}}{2(2+\tilde{\delta})}}}\frac{1}{\sqrt{L}}\left|\sum_{k=1}^{K} e^{(2)}(Kl^*+k)\right|\\
&= o\left(\sqrt{|d|^2\,K\frac{n-m^*}{n}}\right) + o(1) \quad a.s.
\end{aligned}$$

Results (5.26), (5.27) and the result (5.20) for the error sequence $\{e(\cdot)\}$ imply for $L \to \infty$

$$\begin{aligned}
&\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(Y(Kl+k)-\overline{Y}_n)\right)^2\\
&= D_2 + K\,d^2\frac{m^*}{n} - K\,d^2\left(\frac{m^*}{n}\right)^2 + o\left(\sqrt{|d|^2\,K\frac{m^*}{n}}\right) + o(1)\\
&= D_2 + K\,d^2\frac{(n-m^*)\,m^*}{n^2} + o\left(\sqrt{|d|^2\,K\frac{n-m^*}{n}}\right) + o(1) \quad a.s.
\end{aligned}$$

Instead of representation $\{Y(i) = \mu_1 + d\,1_{\{i>m*\}} + e^{(1)}(i)\,1_{\{i\leq m*\}} + e^{(2)}(i)\,1_{\{i>m*\}}, i = 1,\ldots,n\}$, we can write

$$\{Y(i) = (\mu_1 + d) - d\,1_{\{i>m*\}} + e^{(1)}(i)\,1_{\{i\leq m*\}} + e^{(2)}(i)\,1_{\{i>m*\}}, i = 1,\ldots,n\}$$

and an analogous calculation gives

$$\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(Y(Kl+k)-\overline{Y}_n)\right)^2$$
$$= D_2 + K\,d^2\frac{(n-m^*)\,m^*}{n^2} + o\left(\sqrt{|d|^2\,K\frac{m^*}{n}}\right) + o(1) \quad a.s.$$

and putting these two results together gives for $L\to\infty$

$$\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(Y(Kl+k)-\overline{Y}_n)\right)^2$$
$$= D_2 + K\,d^2\frac{(n-m^*)\,m^*}{n^2} + o\left(\sqrt{|d|^2\,K\min\left(\frac{m^*}{n},\frac{n-m^*}{n}\right)}\right) + o(1)$$
$$\geq D_2 + o(1) \quad a.s.,$$

which is (A.14) for $\{Y(\cdot)\}$.

Now we prove that (A.13) holds almost surely for $\{Y(\cdot)\}$. Using equations (5.19) and (5.22) we get

$$\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}\left(Y(Kl+j)-\overline{Y_n}\right)\right|^4$$
$$\leq \frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}e(Kl+j)\right|^4 + \left(\frac{m^*}{n}\right)^4\,|\sqrt{K}\,\overline{e^{(1)}}|^4$$
$$+ \left(\frac{n-m^*}{n}\right)^4|\sqrt{K}\,\overline{e^{(2)}}|^4 + K^2\,d^4\left(\frac{n-m^*}{n}\right)^4 + K^2\,d^4\,\frac{n-m^*}{n}$$
$$\leq D + K^2\,d^4\,\frac{n-m^*}{n} \quad a.s.$$

As above we get using the representation $\{Y(i) = (\mu_1+d) - d\,1_{\{i>m*\}} + e^{(1)}(i)\,1_{\{i\leq m*\}} + e^{(2)}(i)\,1_{\{i>m*\}}$

$$\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}Y(Kl+j)-\overline{Y_n}\right|^4 \leq D + K^2\,d^4\,\frac{m^*}{n} \quad a.s.,$$

which gives

$$\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\ldots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K}Y(Kl+j)-\overline{Y_n}\right|^4$$
$$\leq D + K^2\,d^4\,\min\left(\frac{n-m^*}{n},\frac{m^*}{n}\right)$$
$$\leq D_1 \quad a.s.$$

For the proof of the alternative (ii) we distinguish two main cases $K\,d_n^2 = O(1)$ and $\frac{1}{K\,d_n^2} = O(1)$. The first one is included in (i), so let us assume that $\frac{1}{K\,d^2} = O(1)$. We choose the scores $a_n(i) := Y(i)/\sqrt{d^2\,K}$. Without loss of generality assume again $\mu_1 = 0$. Similarly as in the case (i) we have

$$\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{|d|\,K}\sum_{k=1}^{K}(Y(Kl+k)-\overline{Y}_n)\right)^2 = \frac{1}{d^2\,K^2\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}(Y(Kl+k)\right)^2 - \frac{1}{d^2}\,\overline{Y}_n^2.$$

Using equation (5.21) for $\overline{e^{(1)}} = \frac{1}{m^*}\sum_{i=1}^{m^*} e^{(1)}(i)$ and $\overline{e^{(2)}} = \frac{1}{n-m^*}\sum_{i=m^*+1}^{n} e^{(2)}(i)$ in $\overline{Y}_n = d\,\frac{n-m^*}{n} + \frac{m^*}{n}\overline{e^{(1)}} + \frac{n-m^*}{n}\overline{e^{(2)}}$ we obtain

$$\begin{aligned}
\frac{1}{d^2}\overline{Y_n}^2 &= \left(\frac{n-m^*}{n}\right)^2 + \frac{1}{d^2}\left(\frac{m^*}{n}\right)^2\overline{e^{(1)}}^2 + \frac{1}{d^2}\left(\frac{n-m^*}{n}\right)^2\overline{e^{(2)}}^2 \\
&\quad + \frac{2}{d\sqrt{K}}\frac{(n-m^*)m^*}{n^2}\sqrt{K}\,\overline{e^{(1)}} + \frac{2}{d\sqrt{K}}\left(\frac{(n-m^*)}{n}\right)^2\sqrt{K}\,\overline{e^{(2)}} \\
&\quad + \frac{2}{d^2\,K}\frac{(n-m^*)m^*}{n^2}\sqrt{K}\,\overline{e^{(1)}}\sqrt{K}\,\overline{e^{(2)}} \\
&= \left(\frac{n-m^*}{n}\right)^2 + o(1) \quad a.s.
\end{aligned}$$

Furthermore equation (5.21) yields as $L \to \infty$

$$\begin{aligned}
&\frac{1}{d^2\,K^2\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K}(d\,1_{\{Kl+k>m^*\}} + e(Kl+k))\right)^2 \\
&\qquad \geq \frac{1}{d^2\,K^2\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} d\,1_{\{Kl+k>m^*\}}\right)^2 \\
&\qquad\quad + \frac{2}{d^2\,K^2\,L}\sum_{l=0}^{L-1}\left(\sum_{j=1}^{K} d\,1_{\{Kl+j>m^*\}}\right)\left(\sum_{k=1}^{K} e(Kl+k)\right) \\
&\qquad = \frac{n-m^*}{n} + o(1) \quad a.s.,
\end{aligned}$$

since

$$\frac{1}{d^2\,K^2\,L}\sum_{l=0}^{L-1}\left(\sum_{k=1}^{K} d\,1_{\{Kl+k>m^*\}}\right)^2 = \frac{1}{K^2\,L}K^2\frac{n-m^*}{K} + o(1)$$

and

$$\frac{2}{d^2K^2L}\sum_{l=0}^{L-1}\left(\sum_{j=1}^{K} d\,1_{\{Kl+j>m^*\}}\right)\left(\sum_{k=1}^{K} e(Kl+k)\right)$$
$$\le \frac{1}{\sqrt{K}\,|d|}\sqrt{K}\frac{1}{n}\left|\sum_{k=1}^{K} e^{(1)}(Kl^*+k)\right| + \frac{1}{\sqrt{K}\,|d|}\sqrt{K}\frac{1}{n}\left|\sum_{j=Kl^*+1}^{n} e^{(1)}(j)\right|$$
$$+ \frac{1}{\sqrt{K}\,|d|}\sqrt{K}\frac{1}{n}\left|\sum_{k=1}^{K} e^{(2)}(Kl^*+k)\right| + \frac{1}{\sqrt{K}\,|d|}\sqrt{K}\frac{1}{n}\left|\sum_{j=Kl^*+1}^{n} e^{(2)}(j)\right| \to 0 \quad a.s.,$$

using equations (5.23), (5.24) and assumption $\frac{1}{K\,d^2} = O(1)$. Putting everything together, we have as $L \to \infty$

$$\begin{aligned}
\frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{|d|\,K}\sum_{k=1}^{K}(Y(Kl+k)-\overline{Y}_n)\right)^2 &\ge \left(\frac{n-m^*}{n}\right) - \left(\frac{n-m^*}{n}\right)^2 + o(1) \quad a.s. \\
&\ge \left(\frac{n-m^*}{n}\right)\left(\frac{m^*}{n}\right) + o(1) \quad a.s. \\
&\ge \min\left(\left(\frac{n-m^*}{n}\right)^2, \left(\frac{m^*}{n}\right)^2\right) + o(1) \quad a.s. \\
&\ge \epsilon^2 + o(1) \quad a.s.,
\end{aligned}$$

which is again the assumption (A.14).

Concerning (A.13), we have using equations (5.21) and (5.19)

$$\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\dots,K-1}\left|\frac{1}{|d|\,K}\sum_{j=k+1}^{K} Y(Kl+j) - \overline{Y_n}\right|^4$$
$$\le \frac{1}{d^4\,K^2}\frac{1}{L}\sum_{l=0}^{L-1}\max_{k=0,\dots,K-1}\left|\frac{1}{\sqrt{K}}\sum_{j=k+1}^{K} e(Kl+j)\right|^4$$
$$+ \frac{1}{d^4\,K^2}|\sqrt{K}\,\overline{e^{(1)}}|^4 + \frac{1}{d^4\,K^2}|\sqrt{K}\,\overline{e^{(2)}}|^4 + \left(\frac{n-m^*}{n}\right)^4 + \frac{n-m^*}{n}$$
$$\le D_1 \quad a.s.$$

In the case, where $d = d_n$ follows neither of the above possibilities, we have infinitely many n with $K_n\,d_n^2 \le 1$ and also infinitely many with $K_n\,d_n^2 > 1$. Then we choose the scores

$$a_n(i) = \begin{cases} X(i) & K_n\,d_n^2 \le 1, \\ X(i)/\sqrt{K\,d^2} & K_n\,d_n^2 > 1. \end{cases}$$

As above, the assumptions of Theorem A.5.1 are fulfilled for both sequences, hence also for the complete sequence.

□

5.2 Application

We applied the permutation principle to the occurrences of unusually hot or cold days. For the details concerning our data we refer to Chapter 2, Figures 17–33 and the application part of Chapter 4, where we described the obtained time series of indicators signaling an event that the standardized value exceeds a certain level. As we mentioned earlier, we can not assume that our real data form independent series, see Figure 33, but the block permutation principle provides a technique for studying changes also for strong -mixing sequences.

We tried several lengths of the block $K = 5, 10, 20, 30$ giving almost similar results. Tables 16 and 17 present our results for the length of the block $K = 5$. We hope that $K = 5$ is big enough to capture the dependence. According to Section 5.1, we divided the series into blocks of length $K = 5$, permuted the blocks randomly 10 000 times and for every permutation the value of $T_{L,5}(t)$ was calculated. Then the $(1 - \alpha)\,100\%$ empirical quantile of all $T_{L,5}(t)$ served as $\alpha\,100\%$ critical value. The values of the test statistics $T_{L,5}(t)$ together with the approximate critical values obtained by permutation principle for exceedances over level 2.5 can be found in Table 17. Table 18 presents the same for the exceedances under level -2.5.

	$T_{L,5}(t)$	5 % crit.v.
Brussels over 2.5	15.24	5.66
Cadiz over 2.5	7.55	5.54
Milan over 2.5	7.33	6.15
Padua over 2.5	3.17	6.17
St.Petersburg over 2.5	5.98	5.67
Stockholm over 2.5	4.89	5.54
Uppsala over 2.5	3.95	5.79
Prague over 2.5	8.75	5.76

Table 17. Values of $T_{L,5}(t)$ together with the corresponding approximate critical values obtained by permutation principle.

	$T_{L,5}(t)$	5 % crit.v.
Brussels under -2.5	5.34	5.92
Cadiz under -2.5	3.91	5.49
Milan under -2.5	4.24	5.38
Padua under -2.5	9.03	5.53
St.Petersburg under -2.5	8.57	5.59
Stockholm under -2.5	10.74	5.47
Uppsala under -2.5	8.97	5.31
Prague under -2.5	6.69	5.91

Table 18. Values of $T_{L,5}(t)$ together with the corresponding approximate critical values obtained by permutation principle.

In the case of unusually hot days we reject the null hypothesis for Brussels, Cadiz, Milan, St. Petersburg and Prague series, while in the case of unusually cold days we reject the null hypothesis for Padua, Uppsala, Stockholm St. Petersburg and Prague series.

5.3 Comparison

If we compare the results obtained from the asymptotic theory with the results from the block permutation tests, see Tables 19 and 20, at the 5% significance level the test statistics $T_n(t)$ and $T_{L,5}$ rejected the null hypothesis H_0 in almost the same cases – red numbers denote significant values. The only difference is in occurrences of unusually hot days for Stockholm series.

	$T_n(t)$	$T_{L,5}(t)$
Brussels over 2.5	11.26	15.24
Cadiz over 2.5	5.97	7.55
Milan over 2.5	6.12	7.33
Padua over 2.5	2.67	3.17
St.Petersburg over 2.5	4.53	5.98
Stockholm over 2.5	3.85	4.89
Uppsala over 2.5	3.04	3.95
Prague over 2.5	6.82	8.75

Table 19. Results – statistics $T_n(t)$, $T_{L,5}(t)$, red numbers denote significant values.

	$T_n(t)$	$T_{L,5}(t)$
Brussels under -2.5	3.75	5.34
Cadiz under -2.5	3.10	3.91
Milan under -2.5	3.19	4.24
Padua under -2.5	6.20	9.03
St.Petersburg under -2.5	6.16	8.57
Stockholm under -2.5	7.40	10.74
Uppsala under -2.5	6.30	8.97
Prague under -2.5	4.39	6.69

Table 20. Results – statistics $T_n(t)$, $T_{L,5}(t)$, red numbers denote significant values.

5.4 Results

In the second part of this thesis we studied one phenomena of weather behavior – occurrences of unusually hot, resp. cold days. We were working with long temperature series, the temperature values measured at subsequent days were strongly correlated, the correlation coefficients were for all series very close to 0.8. We defined the problem by a model working with data forming strong-mixing processes and in the fourth chapter we showed that the asymptotic distribution of the testing statistic $T_n(t)$ is valid not only for linear processes but for strong-mixing sequences as well. In the fifth chapter we examined the behavior of the block permutation statistic $T_{L,K}$ and again the theory was generalized from linear processes to strong-mixing sequences.

We applied these two methods for the standardized data describing the exceedance over high, resp. low level characterizing an appearance of unusually warm, resp. cold day, for more details confer Section 4.2. We showed that both methods give similar results, rejecting the null hypothesis for the same observatories.

When analyzing the exceedances over the level 2.5, the tests confirm a clear increase in the Brussels, Cadiz, Milan, St. Petersburg and Prague series with frequencies of these occurrences three times higher in the second part than before the estimated change, while the increase in Padua and Uppsala occurrences of unusually hot days was not significant.

For the exceedances under the level -2.5, the tests confirm a clear decrease in the Uppsala, Padua, St. Petersburg and Stockholm series with frequencies about three times smaller in the second part than before the estimated change, while the decrease in Brussels, Cadiz and Milan occurrences of unusually cold days was not significant.

The advantage of the asymptotic method is that it provides the estimated change-point as well. We might notice two characteristic periods in the estimated date of change – the end of 19^{th} century the end of 20^{th} century.

Although our results might suggest confirmation of the hypothesis that the increased mean of temperature observed since the end of 19^{th} century and a decreasing variability of temperature series is related to the fact that extremely cold days appear less frequent and extremal high temperatures become more frequent, we have to admit several problems which might have influenced our results:

- the number of data. Although we were working with long temperature series, in fact only about 200, resp. 800 observations satisfy our definition of unusually hot, resp. cold day.
- a disproportion between the numbers of unusually hot and cold days is caused by a distinctive negative skewness of minimal temperatures, see Tables 3 and 4.
- heat island effect causing milder winters in city centers.

We hope that in the future the proposed change-point methods should be applied to the series that are not affected by the heat island effect, that are divided into a summer and winter periods and a climatologic definition of an unusually hot, resp. cold day should be taken into account as *a day whose maximum, resp. minimum temperature is within the lowest 5th centile of the daily temperature series for each observatory* and this way we obtain more reasonable results.

Appendix

Appendix A

Some useful theorems and inequalities

A.1 The test statistic for a change-point detection

For detecting change(s) in the behavior of a series, change-point methods may be applied, especially the methods based on the log−likelihood ratio became very popular. The general theory was presented in Csörgő and Horváth [7] (pages 1−34).

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent random variables with the distribution functions $F(x;\varphi_1), \ldots, F(x;\varphi_n)$, respectively, where φ_i are parameters of the distribution functions such that $\varphi_i \in \Phi \subseteq \mathbb{R}^d$ for $i = 1, \ldots, n$.

We are interesting in testing the null hypothesis

$$H_0 : \varphi_1 = \varphi_2 = \ldots = \varphi_n$$

versus the alternative

$$H_A : \varphi_1 = \ldots = \varphi_k \neq \varphi_{k+1} = \ldots = \varphi_n.$$

We have then a two-sample problem and we can apply the likelihood ratio test. The null hypothesis will be rejected for large values of the test statistic

$$\Lambda_k = \frac{\sup_{(\varphi,\tau)\in\Phi\times\Phi} \prod_{1\le i\le k} f(\mathbf{X}_i;\varphi) \prod_{k<i\le n} f(\mathbf{X}_i;\tau)}{\sup_{(\varphi)\in\Phi} \prod_{1\le i\le n} f(\mathbf{X}_i;\varphi)}.$$

We suppose now, that the time of change k is unknown. The null hypothesis and the alternative have the form:

$$\begin{aligned} &H_0 : \varphi_1 = \varphi_2 = \ldots = \varphi_n \\ &H_A : \text{there exists} \quad k \in \{0, \ldots, n-1\} \quad \text{such that} \\ &\qquad \varphi_1 = \ldots = \varphi_k \neq \varphi_{k+1} = \ldots = \varphi_n, \end{aligned}$$

where parameters $\varphi_1 = \ldots = \varphi_k$ before the change as well as parameters $\varphi_{k+1} = \ldots = \varphi_n$ are unknown. The null hypothesis will be rejected for large values of the maximally selected likelihood ratio statistic

$$\max_{0\le k\le n-1}\Lambda_k=\max_{0\le k\le n-1}\frac{\sup_{(\varphi,\tau)\in\Phi\times\Phi}\prod_{1\le i\le k}f(\mathbf{X}_i;\varphi)\prod_{k<i\le n}f(\mathbf{X}_i;\tau)}{\sup_{(\varphi)\in\Phi}\prod_{1\le i\le n}f(\mathbf{X}_i;\varphi)}.$$

If we denote

$$L_k(\varphi)=\sum_{1\le i\le k}\log f(X_i;\varphi),$$
$$L_k^*(\varphi)=\sum_{k<i\le n}\log f(X_i;\varphi)$$

and the true value of the parameters under H_0 by φ_0, then the logarithm of likelihood ratio can be written as

$$\log(\Lambda_k)=\left[L_k(\widehat{\varphi}_k)+L_k^*(\widehat{\varphi}_k^*)-L_n(\widehat{\varphi}_n)\right],$$

where $\widehat{\varphi}_n$ is the maximum likelihood estimator of parameter φ_0 based on the observations $\varphi_1,\varphi_2,\ldots,\varphi_n$, $\widehat{\varphi}_k$ is the maximum likelihood estimator of parameter φ_0 based on the first k observations, resp. $\widehat{\varphi}_k^*$ is the maximum likelihood estimator of parameter φ_0 based on the last $n-k$ observations.

We reject H_0 for large values of the maximum-type statistic

$$\max_{0\le k\le n-1}\left(2\log(\Lambda_k)\right). \tag{A.1}$$

The asymptotic distribution of $\left(\max_{0\le k\le n-1}\left(2\log(\Lambda_k)\right)\right)^{\frac{1}{2}}$ is given by Csörgő and Horváth theorem.

Theorem A.1.1. (Csörgő and Horváth theorem) *Let $\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n$ are independent random vectors in $\mathbb{R}^m$ with the distribution functions $F(\mathbf{x};\varphi_1),\ldots,F(\mathbf{x};\varphi_n)$, where $\varphi_i\in\Phi\subset\mathbb{R}^d$ for all $1\le i\le n$. Let*

$$g(\mathbf{x};\mathbf{y})=\log f(\mathbf{x};\mathbf{y}),$$

$$g_i(\mathbf{x};\mathbf{y})=\frac{\partial}{\partial y_i}g(\mathbf{x};\mathbf{y})$$

and

$$g_{i_1,\ldots,i_r}(\mathbf{x};\mathbf{y})=\frac{\partial^r g(\mathbf{x};\mathbf{y})}{\partial y_{i_1}\ldots\partial y_{i_r}}$$

for all $\mathbf{x}\in\mathbb{R}^m$ and $\mathbf{y}=(y_1,\ldots,y_d)\in\Phi$. The true values of the parameters under H_0 are denoted by $\varphi_{\mathbf{0}}$.

We assume that

C.1 $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ *have probability densities* $f(\mathbf{x};\varphi_1), \ldots, f(\mathbf{x};\varphi_n)$ *with respect to* ν, *where* ν *is a* σ*-finite measure on* $\mathbb{R}^m$

C.2 $F(\mathbf{x};\varphi)$ *generates distinct measures for* $\varphi \in \Phi$

C.3 *For each* k, $1 \leq k \leq n$, *we can find unique estimators* $\widehat{\varphi}_k, \widehat{\varphi}_k^*$ *such that*

$$\sum_{j=1}^{k} g_i(\mathbf{X}_j;\widehat{\varphi}_k) = 0, \quad 1 \leq i \leq d,$$

$$\sum_{j=k+1}^{n} g_i(\mathbf{X}_j;\widehat{\varphi}_k^*) = 0, \quad 1 \leq i \leq d$$

C.4 *There is an open interval*
$Phi_0 \subseteq \Phi \subseteq \mathbb{R}^d$ *containing* φ_0 *such that* $g_i(\mathbf{x};\mathbf{y}), g_{i,j}(\mathbf{x};\mathbf{y})$ *and* $g_{i,j,k}(\mathbf{x};\mathbf{y}), 1 \leq i,j,k \leq d$, *exist and are continuous in* $\mathbf{y}$, *for all* $\mathbf{x} \in \mathbb{R}^m$ *and* $\mathbf{y} \in \Phi_0$

C.5 *There are functions* $M_1(\mathbf{x})$ *and* $M_2(\mathbf{x})$ *such that* $|g_i(\mathbf{x};\mathbf{y})| \leq M_1(\mathbf{x}), |g_{i,j}(\mathbf{x};\mathbf{y})| \leq M_2(\mathbf{x})$ *and* $|g_{i,j,k}(\mathbf{x};\mathbf{y})| \leq M_2(\mathbf{x})$ *for all* $\mathbf{x} \in \mathbb{R}^m$ *and* $\mathbf{y} \in \Phi_0$ *and* $1 \leq i,j,k \leq d$. *The functions* M_1 *and* M_2 *satisfy*

$$\int_{\mathbb{R}^m} M_1(\mathbf{x})\nu(dx) < \infty$$

and

$$E_{\varphi_0} M_2(\mathbf{X}_1) < \infty$$

C.6 $E_{\mathbf{y}} g_i(\mathbf{X}_1;\mathbf{y}) = 0$ *for all* $1 \leq i \leq d$ *and* $\mathbf{y} \in \Phi_0$

C.7 $J_{i,j}(\mathbf{y}) = E_{\mathbf{y}} g_i(\mathbf{X}_1;\mathbf{y}) g_j(\mathbf{X}_1;\mathbf{y}) = -E_{\mathbf{y}} g_{i,j}(\mathbf{X}_1;\mathbf{y}), 1 \leq i,j \leq d$, *and* $J^{-1}(\mathbf{y})$ *exist and are continuous for all* $\mathbf{y} \in \Phi_0$, *where* $J(\mathbf{y}) = \{J_{i,j}(\mathbf{y}), 1 \leq i,j \leq d\}$ *is the information matrix*

C.8 $var_{(\varphi_0)} g_{i,j}(\mathbf{X}_1;\varphi_0) < \infty$ *for all* $1 \leq i,j \leq d$

C.9 $E_{(\varphi_0)}|g_{i,j}(\mathbf{X}_1;\varphi_0)|^{\mu} < \infty$ *for all* $1 \leq i,j \leq d$ *with some* $\mu > 2$

Then if H_0 *and all the necessary regularity conditions C.1 - C.9 hold, we have*

$$\lim_{n\to\infty} P\left(A(\log n)\left(\max_{0\leq k\leq n-1} 2\log(\Lambda_k)\right)^{1/2} \leq t + D_d \log(n)\right) = \exp(-2e^{-t})$$

for all t, *where*

$$A(x) = \sqrt{2\,\log x}$$

and

$$D_d(x) = 2\log x + (d/2)\log\log x - \log\Gamma(d/2),$$

where $\Gamma(t)$ *is the gamma function defined*

$$\Gamma(t) = \int_0^{\infty} y^{t-1}\exp(-y)dy.$$

Remark A.1.2. *If we know the parameters $\varphi_1 = \ldots = \varphi_k = \varphi_0$ before the change and we do not know the parameters $\varphi_{k+1} = \ldots = \varphi_n = \varphi$ after the change, the null hypothesis and the alternative have the form:*

$$\begin{aligned} &H_0 : \varphi_1 = \varphi_2 = \ldots = \varphi_n = \varphi_0 \\ &H_A : \textit{there exists} \quad k \in \{0, \ldots, n-1\} \quad \textit{such that} \\ &\qquad \varphi_1 = \ldots = \varphi_k = \varphi_0, \\ &\qquad \varphi_{k+1} = \ldots = \varphi_n = \varphi \quad , \textit{where} \quad \varphi \neq \varphi_0. \end{aligned}$$

Then the twice log–likelihood ratio has a form

$$\max_{0 \le k \le n-1} \left(2 \log(\Lambda_k^{(0)}) \right) = \max_{0 \le k \le n-1} 2 \left[L_k^*(\widehat{\varphi}_k^*) - L_k^*(\varphi_0) \right]. \tag{A.2}$$

The asymptotic distribution under H_0 and the assumptions of Theorem A.1.1 of the statistic $\max_{0 \le k \le n-1} \left(2 \log(\Lambda_k^{(0)}) \right)$ *is given by*

$$\lim_{n \to \infty} P\left(A(\log n) \left(\max_{0 \le k \le n-1} \left(2 \log(\Lambda_k^{(0)}) \right)^{1/2} \right) \le t + D_d \log(n) \right) = \exp(-e^{-t}).$$

A.2 The extreme value distributions

The extreme value distributions formally arise as limiting distributions for maxima or minima of a sequence of random variables. We concentrate on maxima, as the results for minima can be obtained by replacing random variables $X_1, X_2, \ldots,$ by their negatives $-X_1, -X_2, \ldots$

Suppose $X_1, X_2, \ldots$ are independent random variables with a common distribution function F. The distribution of the maximum $M_n = \max\{X_1, \ldots, X_n\}$ is

$$P\{M_n \leq x\} = P\{X_1 \leq x, \ldots, X_n \leq x\} = F^n(x).$$

We denote by

$$x_F = \sup\{x \in \mathbb{R}; F(x) < 1\}$$

the right endpoint of F. We obtain for all $x < x_F$,

$$P\{M_n \leq x\} = F^n(x) \to 0 \quad \text{as} \quad n \to \infty,$$

and, in the case $x_F < \infty$, we have for $x \geq x_F$ that

$$P\{M_n \leq x\} = F^n(x) = 1.$$

Thus $M_n \xrightarrow{P} x_F$ as $n \to \infty$. This fact does not provide a lot of information.

It turns out that we can get interesting results if we renormalize: define scaling constants $a_n > 0$ and b_n so that

$$\begin{aligned} P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} &= P\left\{M_n \leq a_n x + b_n\right\} \\ &= F^n(a_n x + b_n) \to H(x) \quad \text{as} \quad n \to \infty, \end{aligned}$$

where H is nondegenerate distribution function.

There are only three types of limiting distribution and these are given by following theorem, which is the basis of classical extreme value theory.

Theorem A.2.1. (Fisher−Tippet theorem) *Let $X_1, X_2, \ldots$ be i.i.d. random variables. If there exist norming constants $a_n > 0$ and $b_n \in \mathbb{R}$ and some non-degenerate distribution function H such that*

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} \to H,$$

then H belongs to the type of one following three distribution functions:

$$\begin{aligned}
\textit{Fréchet}: \quad & \Phi_\alpha(x) = 0, && \textit{for} \quad x \le 0 \\
& = \exp\{-x^{-\alpha}\}, && \textit{for} \quad x > 0, \quad \alpha > 0 \\
\textit{Weibull}: \quad & \Psi_\alpha(x) = \exp\{-(-x)^{\alpha}\}, && \textit{for} \quad x \le 0 \quad \alpha > 0 \\
& = 1, && \textit{for} \quad x > 0, \\
\textit{Gumbel}: \quad & \Lambda(x) \;\; = \exp\{-e^{-x}\}, && \textit{for} \quad x \in \mathbb{R}.
\end{aligned}$$

The three types of extreme value distributions may be combined into a single family known as the generalized extreme value distribution (abbreviated to GEV) given by

$$H(x;\mu,\psi,\xi) = \exp\left\{-\left(1+\xi\,\frac{x-\mu}{\psi}\right)^{-\frac{1}{\xi}}\right\} \quad \psi > 0, \mu \in \mathbb{R}, \xi \in \mathbb{R} \tag{A.3}$$

defined on the region for which $1 + \xi(x-\mu)/\psi > 0$. In (A.3), μ is a location parameter, ψ is a scale parameter and ξ is a shape parameter. The case $\xi > 0$ is the Fréchet type with $\alpha = 1/\xi$, the case $\xi < 0$ is the Weibull type with $\alpha = -1/\xi$, while the case $\xi = 0$ is the Gumbel distribution as a result from following limit

$$\lim_{\xi\to 0} H(x;\mu,\psi,\xi) = \exp\left\{-\exp\left(-\frac{x-\mu}{\psi}\right)\right\}.$$

For the GEV, the density $h(x;\mu,\psi,\xi)$ is obtained by differentiating (A.3) with a respect to x

$$h(x;\mu,\psi,\xi) = \frac{1}{\psi}\left(1+\xi\,\frac{x-\mu}{\psi}\right)^{-\frac{1}{\xi}-1} \exp\left\{-\left(1+\xi\frac{x-\mu}{\psi}\right)^{-\frac{1}{\xi}}\right\}, \tag{A.4}$$

provided $1 + \xi(x-\mu)/\psi > 0$. Here are a few basic properties of the GEV distribution. The mean exists if $\xi < 1$ and the variance if $\xi < \frac{1}{2}$; more generally the k'th moment exists if $\xi < \frac{1}{k}$. The mean and the variance are given by

$$\begin{aligned}
\mu_1 &= E(X) = \mu + \frac{\psi}{\xi}\Gamma(1-\xi) - 1, && \xi < 1, \\
\mu_2 &= E(X-\mu_1)^2 = \frac{\psi^2}{\xi^2}\Gamma(1-2\xi) - \Gamma^2(1-\xi), && \xi < \frac{1}{2},
\end{aligned}$$

where $\Gamma(\cdot)$ is the Gamma function. In the limiting case $\xi \to 0$, these reduce to

$$\mu_1 = \mu + \psi\gamma, \quad \mu_2 = \frac{\psi^2\pi^2}{6},$$

where γ is Euler's constant.

A.3 Limit theorems

Theorem A.3.1. (Smith) *In general we assume θ is real-valued and $(\alpha, \beta) \in \Phi \subseteq \mathbb{R}^2$. Let $(\theta_0, \alpha_0, \beta_0)$ denote the true values of (θ, α, β). The maximum likelihood estimator, when it exists, will be denoted by $(\widehat{\theta}, \widehat{\alpha}, \widehat{\beta})$. The probability density is of the form*

$$f(x; \theta, \alpha, \beta) = (x - \theta)^{\alpha - 1} g(x - \theta; \alpha, \beta) \quad (\theta < x < \infty)$$

Assume conditions

$$E\Big(\frac{\partial}{\partial\theta}\big(\log f(X_i; \varphi_{\mathbf{0}})\big)\Big) = 0,$$
$$E\Big(\frac{\partial}{\partial\alpha}\big(\log f(X_i; \varphi_{\mathbf{0}})\big)\Big) = 0,$$
$$E\Big(\frac{\partial}{\partial\beta}\big(\log f(X_i; \varphi_{\mathbf{0}})\big)\Big) = 0$$

and

$$\begin{aligned}
m_{\theta\theta} &= E\{\frac{\partial}{\partial\theta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\frac{\partial}{\partial\theta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\theta^2}\log(f(X_i;\, \varphi_{\mathbf{0}}))\}, \\
m_{\alpha\alpha} &= E\{\frac{\partial}{\partial\alpha}\log(f(X_i;\, \varphi_{\mathbf{0}}))\frac{\partial}{\partial\alpha}\log(f(X_i;\, \varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\alpha^2}\log(f(X_i;\, \varphi_{\mathbf{0}}))\}, \\
m_{\beta\beta} &= E\{\frac{\partial}{\partial\beta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\beta^2}\log(f(X_i;\, \varphi_{\mathbf{0}}))\}, \\
m_{\theta\alpha} = m_{\alpha\theta} &= E\{\frac{\partial}{\partial\theta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\frac{\partial}{\partial\alpha}\log(f(X_i;\, \varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\theta\partial\alpha}\log(f(X_i;\, \varphi_{\mathbf{0}}))\}, \\
m_{\theta\beta} = m_{\beta\theta} &= E\{\frac{\partial}{\partial\theta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\theta\partial\beta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\},
\end{aligned}$$

$$\begin{aligned}
m_{\alpha\beta} = m_{\beta\alpha} &= E\{\frac{\partial}{\partial\alpha}\log(f(X_i;\, \varphi_{\mathbf{0}}))\frac{\partial}{\partial\beta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\} \\
&= -E\{\frac{\partial^2}{\partial\alpha\partial\beta}\log(f(X_i;\, \varphi_{\mathbf{0}}))\}.
\end{aligned}$$

are satisfied, and moreover assume

- *All second order partial derivatives of $g(x;\theta,\alpha,\beta)$ exist and are continuous in $0 < x < \infty$, $(\alpha,\beta) \in \Phi$. Moreover $c(\alpha,\beta) = \alpha^{-1}\lim_{x\to 0} g(x;\alpha,\beta)$ exists, is positive and finite for each (α,β), and is twice continuously differentiable as a function of (α,β).*
- *If $h(x;\alpha,\beta)$ is any of $\frac{\partial^2}{\partial x\partial\alpha}\log g(x;\alpha,\beta)$, $\frac{\partial^2}{\partial x\partial\beta}\log g(x;\alpha,\beta)$, $\frac{\partial^2}{\partial\alpha^2}\log g(x;\alpha,\beta)$, $\frac{\partial^2}{\partial\alpha\partial\beta}\log g(x;\alpha,\beta)$, $\frac{\partial^2}{\partial\beta^2}\log g(x;\alpha,\beta)$. Then as $\theta\to\theta_0$, $\alpha\to\alpha_0$, $\beta\to\beta_0$,*

$$E_0|h(X-\theta;\alpha,\beta) - h(X-\theta_0;\alpha_0,\beta_0)| \to 0,$$

  where E_0 is expectation with respect to $f(.;\theta_0,\alpha_0,\beta_0)$. If $\alpha(\alpha_0,\beta_0) > 2$, we require the same of $x(x,\alpha,\beta) = \frac{\partial^2}{\partial x^2}\log g(x;\alpha,\beta)$.
- *For each $\epsilon > 0$, $\delta > 0$, there exists a function $h_{\epsilon,\delta}$ such that*

$$|\frac{\partial^2}{\partial x^2}\log g(x;\alpha,\beta)| < \frac{\epsilon}{x^2} + h_{\epsilon,\delta}(y,\alpha_0,\beta_0)$$

  whenever $|\alpha-\alpha_0| < \delta$, $|\beta-\beta_0| < \delta$, $|x-y| < \delta$, and $h_{\epsilon,\delta}$ satisfies

$$\int_0^\infty h_{\epsilon,\delta}(x,\alpha_0,\beta_0) f(x;0,\alpha_0,\beta_0) < \infty.$$

Suppose $\alpha > 1$, $\boldsymbol{M}$ is strictly positive-definite. Then there exist a sequence $(\widehat{\theta}_k,\widehat{\alpha}_k,\widehat{\beta}_k)$ of solutions to the likelihood equations such that

$$\widehat{\theta}_k - \theta_0 = O_p(k^{\frac{1}{2}}) \quad \widehat{\alpha}_k - \alpha_0 = O_p(k^{\frac{1}{2}}) \quad \widehat{\beta}_k - \beta_0 = O_p(k^{\frac{1}{2}}).$$

Theorem A.3.2. (Marcinkiewicz−Zygmund law) *Let $X_1, X_2, \ldots$ be i.i.d. Let $0 < r < 2$. Establish the equivalence*

$$E|X|^r < \infty \quad \textit{if and only if} \quad \frac{1}{n^{\frac{1}{r}}}\sum_{k=1}^{n}(X_k - c) \to 0 \quad \textit{a.s. for some } c.$$

If so, then $c = EX$ when $1 \le r < 2$, while c is arbitrary when $0 < r < 1$.

Theorem A.3.3. (Law of the Iterated Logarithm) *Let $X_1, X_2, \ldots$ be i.i.d. Consider the partial sums $S_n = X_1 + \cdots + X_n$.*
If $EX = 0$ and $\sigma^2 \equiv Var|X| < \infty$, then

$$\limsup_{n\to\infty}\frac{S_n}{\sigma\sqrt{2\,n\log\log n}} = 1 \quad \textit{a.s.} \quad \textit{while} \quad \liminf_{n\to\infty}\frac{S_n}{\sigma\sqrt{2\,n\log n\log n}} = -1 \quad \textit{a.s.}$$

Theorem A.3.4. *Let $X, X_1, X_2, \ldots$ be i.i.d. Then*

$$\mathsf{E}\,|X| < \infty \quad \textit{if and only if} \quad \frac{1}{n}\max_{1\le k\le n}|X_k| \to 0 \quad \textit{a.s.}$$

Lemma A.3.5. *Let h be a continuously differentiable real-valued function of $p+1$ variables and let H denote the gradient vector of h. Suppose that the scalar product of x and $H(x)$ is negative whenever $|x| = 1$. Then h has a local maximum, at which $H = 0$, for some $|x| < 1$.*

A.4 Strong-mixing sequences

Definition A.4.1. *Let*

$$\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots$$

be a strictly stationary sequence of random variables defined on a probability space $(\Omega, \mathcal{B}, P)$. *For* $a \le b$, *define* $\mathcal{M}_a^b$ *as the* σ*-field generated by the random variables* $\xi_a, \ldots, \xi_b$; *define* $\mathcal{M}_{-\infty}^a$ *as the* σ*-field generated by the random variables* $\ldots, \xi_{a-1}, \xi_a$; *and define* $\mathcal{M}_a^\infty$ *as the* σ*-field generated by the random variables* $\xi_a, \xi_{a+1}, \ldots$. *The sequence* $\{\xi_n, n \in \mathbb{Z}\}$ *is said to satisfy strong-mixing condition if there exists a sequence of real numbers* $\alpha(n)$ *satisfying*

$$\lim_{n\to\infty} \alpha(n) = 0$$

such that

$$|P(A \cap B) - P(A)P(B)| \le \alpha(n)$$

for all $A \in \mathcal{M}_{-\infty}^k$ *and* $B \in \mathcal{M}_{k+n}^\infty$ *and all* $k, n \ge 1$.

Before we start with listing properties of strong-mixing sequences, we introduce conditions for linear processes to be strong-mixing, see Withers [27].

Lemma A.4.2. *Let* Z_j *be independent random variables on* $\mathbb{R}$ *with characteristic functions* ϕ_j *such that*

$$K = (2\pi)^{-1} \max_j \int |\phi_j(t)| dt < \infty \tag{A.5}$$

and

$$\gamma = \max_j \mathsf{E}|Z_j|^\delta < \infty \quad \textit{for some} \quad \delta > 0. \tag{A.6}$$

Let g_j *be complex numbers such that*

$$G_t = S_t(\min(1,\delta))^{\max(1,\delta)} \longrightarrow 0 \quad \textit{as} \quad t \longrightarrow \infty, \tag{A.7}$$

where

$$S_t(\delta) = \sum_{\nu=t}^{\infty} |g_\nu|^\delta. \tag{A.8}$$

Then for all t, $X_{nt} = \sum_{j=0}^n g_j Z_{t-j}$ *converges in probability to a random variable* X_t *as* $n \to \infty$. *Suppose*

$$M_0 = \sup_{m,s,k\ge 1} \sup_{\alpha,\beta,\nu} \max_t \left| \frac{\partial}{\partial \nu_t} P\left(W + \nu \in \bigcup_1^s Dj \right) \right| < \infty, \tag{A.9}$$

where

$$\begin{aligned} D_j &= X_{t=k}^{k+m-1}(\alpha_{jt}, \beta_{jt}), \quad \nu = (\nu_k, \ldots, \nu_{k+m-1}), \\ W &= (W_k, \ldots, W_{k+m-1}), \quad W_t = X_{t-1,t}. \end{aligned}$$

Then for X_t,

$$\alpha(k) \le 2(4M_0 + \gamma)\alpha_0(k), \quad \textit{where} \quad \alpha_0(k) = \sum_{t=k}^{\infty} G_t. \tag{A.10}$$

As an application, consider the general A.R.M.A. process, written as

$$\prod_{j=1}^{p}(1-\rho_j\,U)\,X_t = f_q(U)\,Z_t,$$

where $U\,x_t = x_{t-1}$, $f_q(z)=\sum_{l=0}^{q} b_l\,z_l$.

Corollary A.4.3. *If the A.R.M.A. process X_t satisfies* (A.5), (A.6), (A.9) *and condition*

$$r = \max_{j=1}^{p} |\rho_j| < 1 \tag{A.11}$$

then for all $r_0 > r$, $\alpha(k) = O(r_0^{\lambda\,k})$ where $\lambda = \delta(1+\delta)^{-1}$.

Following lemmas summarize moment inequalities for strong-mixing processes.

Lemma A.4.4. *Let ξ and η be two random variables measurable $\mathcal{F}$ and $\mathcal{G}$ respectively. Let $r,s,t \geq 1$ with $r^{-1}+s^{-1}+t^{-1}=1$. If $||\xi||_s < \infty$ and $||\eta||_t < \infty$ then*

$$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \leq 10(\rho(\mathcal{F}\mathcal{G}))^{1/r}||\xi||_s||\eta||_t.$$

Moreover, if $||\xi||_\infty < \infty$ and $||\eta||_\infty < \infty$ then

$$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \leq 4\rho(\mathcal{F}\mathcal{G})||\xi||_\infty||\eta||_\infty.$$

Here

$$\rho(\mathcal{F},\mathcal{G}) = \sup|P(AB)-P(A)P(B)|$$

the supremum being extended over all $A \in \mathcal{F}$ and $B \in \mathcal{G}$.

Corollary A.4.5. *If ξ is measurable $M_{-\infty}^{n}$ and $||\xi||_\infty < \infty$, and if η is measurable M_{n+k}^{∞} $(k \geq 0)$ and $||\eta||_\infty < \infty$ there exists a constant M, such as*

$$|\mathsf{E}\{\xi\eta\} - \mathsf{E}\{\xi\}\mathsf{E}\{\eta\}| \leq M\,\alpha(k).$$

As a consequence of a previous corollary is also a following result dealing with the sums of strong-mixing sequences.
Let ξ_n is stationary, bounded, strong-mixing sequence of random variables. Let

$$S_n = \xi_1 + \xi_2 + \ldots + \xi_n$$

and $S_0 = 0$.

Corollary A.4.6. *If ξ_0 is bounded by C and $\mathsf{E}\xi_0 = 0$, and if $\sum \alpha(i) < \infty$, then*

$$\mathsf{E}S_n^4 \leq K_\alpha C^4 n^2,$$

where K_α depends on α alone.

Proof. Can be found in [24]. □

The following theorem gives moment inequalities for the maximum of partial sums as well as a convergency rate in the strong law of large numbers, see Kirch [20] and Serfling [25].

Theorem A.4.7. *Let $\{Y(i), i \in \mathbb{Z}\}$ be a strictly stationary sequence with $\mathsf{E}Y(i) = 0$, $i \in \mathbb{Z}$. Assume there are $\delta, \Delta > 0$ with*

$$\mathsf{E}|Y(i)|^{2+\delta+\Delta} \leq D_1 \quad \textit{for all} \quad i \in \mathbb{Z}$$

and there is a sequence $\alpha(k)$ with $\alpha_\gamma(k) \leq \alpha(k), k \in \mathbb{N}$, and

$$\sum_{k=0}^{\infty} (k+1)^{\delta/2} \alpha(k)^{\Delta/(2+\delta+\Delta)} \leq D_2(\delta, \Delta),$$

where α_γ is the corresponding mixing coefficient. Then it holds

a)

$$\mathsf{E}\left(\max_{l=1,\ldots,n} \left| \sum_{j=1}^{l} Y(j) \right|^{2+\delta} \right) \leq D\, n^{(2+\delta)/2},$$

where D only depends on δ and the joint distribution of the $Y(i)$.

b)

$$\frac{1}{n} \left| \sum_{j=1}^{n} Y(j) \right| = O\left(\frac{(\log n)^{1/(2+\delta)} (\log \log n)^{2/(2+\delta)}}{n^{1/2}} \right) \quad a.s.$$

A main tool in the change-point analysis is to make use of an invariance principle. We present an almost sure invariance principle for sums of $d-$dimensional random vectors satisfying strong-mixing condition (for details we refer to Kuelbs and Philipp [21].)

Theorem A.4.8. *Let $\{\xi_n, n \leq 1\}$ be in a weak sense stationary sequence of $\mathbb{R}^d$-valued random vectors, centered at expectations and having $(2+\delta)$th moments with $0 < \delta \leq 1$, uniformly bounded by 1. Suppose that $\{\xi_n, n \leq 1\}$ satisfies a strong-mixing condition with*

$$\alpha(n) \ll n^{-(1+\epsilon)(1+2/\delta)} \quad \epsilon > 0.$$

Write

$$\xi_n = (\xi_{n1}, \ldots, \xi_{nd}).$$

Then the two series in

$$\gamma_{ij} = \mathsf{E}\xi_{1i}\xi_{1j} + \sum_{k \geq 2} \mathsf{E}\xi_{1i}\xi_{kj} + \sum_{k \geq 2} \mathsf{E}\xi_{ki}\xi_{1j} \tag{A.12}$$

converge absolutely. Denote the matrix $((\gamma_{ij}))(1 \leq i, j \leq d)$ by Γ. Then we can redefine the sequence $\{\xi_n, n \leq 1\}$ on a new probability space together with Brownian motion $X(t)$ with covariance matrix Γ such that

$$\sum_{n \leq t} \xi_n - X(t) \ll t^{1/2-\lambda} \quad a.s.$$

for some $\lambda > 0$ depending on ϵ, δ and d only.

The symbol $\ll$ denotes that the left-hand side is bounded by an unspecified constant times the right-hand side; in the other words, the $\ll$ symbol is used instead of the O notation.

Remark A.4.9. *Specially for $d = 1$ and a sequence of random variables $\{X_n, n \leq 1\}$ satisfying conditions of Theorem A.4.8 with $\mathsf{E}X_n = 0, \quad \sigma_X^2 = EX_n^2$ we obtain in relation* (A.12)

$$\sigma^2 = \sigma_X^2 \left(1 + 2 \sum_{i=1}^{\infty} \rho(i) \right).$$

A.5 Rank statistics

The asymptotics for the rank statistic is the main tool to derive the validity of the block permutation method. For more details and proofs we refer to Hušková [15], Antoch and Hušková [2] and Kirch [20].

Let $(X_1, \dots, X_n)$ be i.i.d. random variables with common continuous distribution function F and let $(R_1, \dots, R_n)$ be the corresponding ranks. Consider the simple linear rank statistic

$$S_k(\mathbf{a}) = \sum_{i=1}^{k} \left(a_n(R_i) - \overline{a}_n\right), \qquad k = 1, \dots, n,$$

where $a_n(1), \dots, a_n(n)$ are scores satisfying:

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |(a_n(R_i) - \overline{a}_n)|^{\nu} \le D_1 \tag{A.13}$$

for some $\nu > 2$ and

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} |(a_n(R_i) - \overline{a}_n)|^{2} \ge D_2, \tag{A.14}$$

where $\overline{a}_n = \frac{1}{n}\sum_{i=1}^{n} a_n(i)$ and $D_1, D_2 > 0$ are some constants.

The main theorem for ranks statistics states:
Theorem A.5.1. *Let $(X_1, \dots, X_n)$ be i.i.d. random variables with common distribution functions F. Let assumptions* (A.13), (A.14) *are satisfied. Then, as $n \to \infty$, it holds for all $t \in \mathbb{R}$*

$$\lim_{n\to\infty} P\left(A(\log n) \max_{1\le k<n} \left\{ \sqrt{\frac{n}{k(n-k)}} \frac{1}{\widehat{\sigma}_n(\mathbf{a})} |S_k(\mathbf{a})| \right\} - D(\log n) \le t \right) = \exp(-2e^{-t}) \quad a.s.,$$

where

$$A(x) = \sqrt{2\,\log x},$$

$$D(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log \pi,$$

$$\sigma_n^2(\mathbf{a}) = \frac{1}{n-1} \sum_{i=1}^{n} \left(a_n(i) - \overline{a}_n\right)^2.$$

Bibliography

[1] Antoch J., Hušková M., Prášková Z., Effect of dependency on statistics for determination of change. *J. statist. Plan. Inf.* **60**: 291 – 310, 1997.

[2] Antoch J. and Hušková M., Permutation tests for a change point analysis. *Statist. Probab. Letters* **53**: 37 – 46, 2001.

[3] Antoch J. and Jarušková D., Testing a homogeneity of stochastics processes. *Kybernetika* **43**: 415 – 430, 2007.

[4] Bilingsley, P., *Convergence of Probability Measures*, Wiley, New York, 1968.

[5] Brockwell P. J., Davis R. A., Time series: Theory and Methods. Springer–Verlag, 1991

[6] Camuffo, D., Jones, P. (Eds.), Improved understanding of past climatic variability from early daily european instrumental sources. *Climatic Change* **53**: 1 – 3, 2002.

[7] Csörgő, M. and Horváth, L., *Limit theorems in change – point analysis.* John Wiley, Chichester, 1997.

[8] Darling D.A., Erdös P., A llimit theorem for the maximum of normalized sums of independent random variables. *Duke Math J.*, vol.**23**: 143 – 155, 1956.

[9] Dehling H., Philipp W., Almost sure invariance principles for weakly dependent vector-valued random variables. *The annals of Probability*, vol.**10**, No.**3**: 689 – 701, 1982.

[10] Embrechts, P., Klüppelberg, C., Mikosch, T., *Modelling extremal events.* Springer, Berlin, 1997.

[11] Gombay, E. and Horváth, L., An application of the maximum likelihood test to the change – point problem. *Stoch. Proc. Appl.* **50**: 161 – 171, 1994.

[12] Feller W., An Introduction to Probability Theory and Its Applications, Vol. 1, 3rd ed. New York: Wiley, 1968.

[13] Horváth, L., Horváth, Z., Hušková M., Ratio tests for change point detection. *IMS COLLECTIONS. Beyond Parametrics in Interdisciplinary Research*, Vol.**1**: 293 – 304, 2007.

[14] Horváth, L., Kokoszka, P., Steinebach, J., Testing for changes in multivariate dependent observations with an application to temperature changes., *J. Multivariate Analysis* **68**: 96 – 119, 1999.

[15] Hušková, M., Limit theorems for rank statistics. *Statist. Probab. Lett.* **32**: 42 – 55, 1997. Jandhyala, V.K., Fotopoulos, S.B., Evaggelopoulos, N., Change-point methods for Weibull models with applications to detection of trends in extreme temperatures. *Environmetrics* **10**: 547 – 564, 1999..

[16] Jarušková, D., Some problems with application of change-point detection methods to environmental data. *Environmetrics* **8**: 469 – 483, 1997.

[17] Jarušková, D., Testing appearance of linear trend. *J. Statist. Plann. Infer.* **70**: 263 – 276, 1998.

[18] Jarušková, D., Maximum log-likelihood test for a change in three parameter Weibull distribution. *J. Statist. Plann. Inference*, **137**, 1805 – 1815, 2007.

[19] Jarušková, D. and Rencová, M., Analysis of annual maximal and minimal temperatures for some European cities by change-point methods. *Environmetrics* **19**: 221 – 233, 2007.

[20] Kirch, C., Block permutation principles for the change analysis of dependent data. *J. Statist. Plann. Inference*, **137**:2453 – 2474, 2007.

[21] Kuelbs, J. and Philipp, W., Almost sure invariance principles for partial sums of mixing B-valued random variables. *The Annals of Probability* **57**: 477-480, 1980.

[22] Leadbetter M. R., Lindgren G. and Rootzén H., *Extremes and related properties of random sequences and processes.* Springer Verlag, Heidelberg, 1983.

[23] Rencová, M., Extremes of temperature series (in Czech). *In: Antoch, J., Dohnal, G. (Eds.), Robust'04*, JČMF, Praha, 2004.

[24] Rio, E., Covariance inequalities for strongly mixing processes. *Annales de l'Institut Henri Poincar*, **4**: 587 – 597, 1993.

[25] Serfling R. J., Moment inequalities for the maximum cumulative sum. *The Annals of Mathematical Statistics* **41**: 1227 – 1234, 1970.

[26] Smith R. L., Maximum likelihood estimation in a class of nonregular cases. *Biometrika* **72**: 67 – 90, 1985.

[27] Withers, C.S., Conditions for linear processes to be strong-mixing. *Z. Wahrsch. verw. Geb.*,57:477-480,1981

[28] Yao Y. C. and Davis R. A., The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhya* **48**: 339 – 353, 1984.