Change-point tests and the bootstrap under long- and short-range dependence

Dissertation

Zur Erlangung des Doktorgrades der Naturwissenschaften an der Fakultät für Mathematik der Ruhr-Universität Bochum

July 2017
Preface

The thesis is based on my research from Oktober 2013 until July 2017. During this period, the following four papers originated: Change-point tests under local alternatives for long-range dependent processes (see also Tewes (2017)), Block bootstrap for the empirical process of long-range dependent data (see also Tewes (2016)), Sequential block bootstrap in a Hilbert space with application to change point analysis (see also Sharipov et al. (2016a)) and Convolved subsampling estimation with applications to block bootstrap (see also Tewes et al. (2017)).

Each of this research articles stands for itself. Hence they are presented in individual chapters (Chapters 3 to 6). However, they share following common topics: Change-point tests (in detail the Cramér-von Mises change-point test), block bootstrap and long-range dependence each play a central part in at least two of the four papers. Therefore, the articles are merged into one entity: The notation is adjusted, common features are highlighted and Chapter 2 serves as an introduction to all three parts.

The main aspect of this thesis is the theoretical investigation of the statistical methods. Then again, change-point analysis by itself is highly motivated by empirical examples. Hence we keep possible applications always in mind and in fact our methods are applied to network traffic and flood data. Moreover, all chapters include numerical results. Those serve to illustrate and motivate the statistical methods, and to show the validity for finite samples (as all our theorems are of asymptotic nature).

Finally, note that there are some small sections containing results that have not been published in the aforementioned papers: In Section 3.2.3 we state the asymptotic distribution of Kolmogorov-Smirnov statistic under weaker assumption. In Sections 3.3.1 and 3.3.2 additional change-point tests are considered, the CUSUM and the Wilcoxon test. Here we extend existing results. Finally, the asymptotic relative efficiency (a quantity that compares different tests asymptotically) is computed for several two-sample tests in Section 3.4.3.

At this point I like to thank several people, without whom this thesis would not exists in its actual form. First, I would like to express deep gratitude to my supervisor Herold Dehling, whose support I felt during the entire span of my studies. His mathematical knowledge and scientific advise was always of immense help.

Further, I am very grateful to Olimjon Sharipov and Martin Wendler. From the beginning of my doctoral research time to its end, they shared their expertise with me, making especially the start quite comfortable. Also I want to thank Dan Nordman. He made it possible for me
to visit Iowa State, where he was an incredibly nice host and a genuine advisor.
Financial support by the DFG through the SfB 823 “Statistical modelling of nonlinear dynamic processes” is thankful acknowledged.
I enjoyed working at the university a lot, which would have been not so without the great colleagues from our research group. All of them contributed to this thesis in different ways. In particular, I owe thanks to Jannis Buchsteiner for carefully checking large parts of this thesis.
I will miss our daily discussions about stochastics and countless other topics.
Final thanks belong to Jelena Milanovic and my parents for many and varied reasons.
## Contents

**List of Symbols**

1 **Introduction**
   1.1 Change-point analysis in time series ........................................... 1
   1.2 Main results of this thesis .......................................................... 3

2 **Probabilistic background** ................................................................. 7
   2.1 Long-range dependence ................................................................. 7
      2.1.1 Definition and motivation of long-range dependence ...................... 7
      2.1.2 Subordinated Gaussian processes and Hermite polynomials .......... 9
      2.1.3 Long memory linear processes ................................................ 12
      2.1.4 Fractional Brownian motion and Hermite processes ................. 14
   2.2 Empirical processes of LRD data ................................................. 16
   2.3 Short-range dependence and probability theory in Hilbert spaces ....... 20
      2.3.1 Mixing conditions in a Hilbert space \( H \) .................................. 20
      2.3.2 Brownian motion in \( H \) ........................................................ 21
      2.3.3 Convergence in \( D_{\mathcal H}[0,1] \) ............................................. 22
      2.3.4 Moment inequalities .............................................................. 24
   2.4 Bootstrap ......................................................................................... 24
      2.4.1 Two block bootstrap methods .................................................... 24
      2.4.2 Three types of convergence ...................................................... 26
      2.4.3 Bootstrap Consistency ............................................................... 28
      2.4.4 Subsampling .............................................................................. 30

3 **Distributional change under long-range dependence** ......................... 33
   3.1 Limit theorems for the empirical process and test statistics ............. 34
      3.1.1 The change-point setting .......................................................... 34
      3.1.2 Reduction principle for the empirical process of triangular arrays .... 36
      3.1.3 The empirical process under local alternatives ............................. 39
   3.2 Asymptotics for change-point statistics .......................................... 41
      3.2.1 Kolmogorov-Smirnov and Cramér-von Mises statistics .................. 41
      3.2.2 Examples ................................................................................. 45
      3.2.3 Weakening the conditions ......................................................... 49
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3</td>
<td>Application to further change-point tests</td>
<td>53</td>
</tr>
<tr>
<td>3.3.1</td>
<td>CUSUM test</td>
<td>53</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Wilcoxon test</td>
<td>58</td>
</tr>
<tr>
<td>3.4</td>
<td>Asymptotic relative efficiency</td>
<td>66</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Definition and motivation</td>
<td>66</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Change-point tests</td>
<td>67</td>
</tr>
<tr>
<td>3.4.3</td>
<td>Two-sample tests</td>
<td>72</td>
</tr>
<tr>
<td>3.5</td>
<td>Numerical results</td>
<td>79</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Fractional Gaussian Noise</td>
<td>79</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Unknown Hurst coefficient</td>
<td>80</td>
</tr>
<tr>
<td>3.5.3</td>
<td>$farima(0,d,0)$-processes</td>
<td>86</td>
</tr>
<tr>
<td>3.5.4</td>
<td>Short-range dependent effects</td>
<td>87</td>
</tr>
<tr>
<td>3.6</td>
<td>Proofs of the main results</td>
<td>88</td>
</tr>
<tr>
<td>3.6.1</td>
<td>Proof of Theorem 5</td>
<td>88</td>
</tr>
<tr>
<td>3.6.2</td>
<td>Proof of Theorem 6</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>Block bootstrap for the empirical process under long memory</td>
<td>99</td>
</tr>
<tr>
<td>4.1</td>
<td>Weak convergence of the bootstrapped empirical process</td>
<td>100</td>
</tr>
<tr>
<td>4.1.1</td>
<td>The case of subordinated Gaussian processes</td>
<td>100</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Estimating the Hermite coefficient function</td>
<td>105</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Smoothed empirical process</td>
<td>109</td>
</tr>
<tr>
<td>4.2</td>
<td>Applications to statistical testing</td>
<td>111</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Testing for monotonicity of transformations</td>
<td>111</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Goodness-of-fit tests</td>
<td>117</td>
</tr>
<tr>
<td>4.3</td>
<td>Proof of Lemma 4.1.3</td>
<td>120</td>
</tr>
<tr>
<td>5</td>
<td>Sequential block bootstrap in a Hilbert space with applications to change-point analysis</td>
<td>129</td>
</tr>
<tr>
<td>5.1</td>
<td>Limit theorems in a Hilbert space</td>
<td>130</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Functional central limit theorem</td>
<td>130</td>
</tr>
<tr>
<td>5.1.2</td>
<td>Bootstrapping the partial sum process</td>
<td>131</td>
</tr>
<tr>
<td>5.2</td>
<td>Application to change-point analysis</td>
<td>132</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Change in the mean of $H$-valued observations</td>
<td>132</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Local alternatives and multiple change-points</td>
<td>134</td>
</tr>
<tr>
<td>5.2.3</td>
<td>Cramer-von Mises test for change in the marginal distribution</td>
<td>136</td>
</tr>
<tr>
<td>5.3</td>
<td>Numerical results and real data analysis</td>
<td>139</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Application to flood data</td>
<td>139</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Finite sample performance of the CUSUSM test</td>
<td>141</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Finite sample performance of the Cramer-von Mises test</td>
<td>145</td>
</tr>
</tbody>
</table>
5.4 Proofs of the main results .................................................. 148

6 Convolved subsampling estimation with applications to block bootstrap 157
  6.1 Fundamental results for convolved subsampling .................... 158
  6.1.1 Convolved subsampling and connection to the bootstrap .......... 158
  6.1.2 Convolved subsampling and connections to block bootstrap ..... 159
  6.1.3 Limit theorems for the convolved distribution .................. 161
  6.1.4 Consistency of the subsampling variance ....................... 165
  6.2 Applications of convolved subsampling estimation ................. 166
  6.2.1 Convolved subsampling for general statistics under mixing .... 166
  6.2.2 Block bootstrap for mixing non-stationary time processes ...... 168
  6.2.3 Block bootstrap for linear time processes ...................... 170
  6.2.4 Block bootstrap under long-range dependence .................. 171
  6.3 Convolved subsampling in other contexts ......................... 173
  6.3.1 U-statistics ...................................................... 173
  6.3.2 Spectral estimators for non-stationary time series ............ 174
  6.4 Proofs ............................................................. 175
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1_A$</td>
<td>indicator function</td>
<td>2</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>nominal level of significance</td>
<td>28</td>
</tr>
<tr>
<td>$\alpha(k)$</td>
<td>alpha-mixing coefficient</td>
<td>20</td>
</tr>
<tr>
<td>$\bar{X}_n$</td>
<td>sample mean</td>
<td>5</td>
</tr>
<tr>
<td>$\beta$</td>
<td>power of a test</td>
<td>66</td>
</tr>
<tr>
<td>$\beta(k)$</td>
<td>absolute regularity coefficient</td>
<td>20</td>
</tr>
<tr>
<td>$B_H(t)$</td>
<td>fractional Brownian motion with Hurst parameter $H$</td>
<td>14</td>
</tr>
<tr>
<td>$C_H(B)$</td>
<td>space of continuous functions mapping from $B$ to $H$</td>
<td>22</td>
</tr>
<tr>
<td>$c_m$</td>
<td>$m$th Hermite coefficient</td>
<td>10</td>
</tr>
<tr>
<td>$C_{n,k}(x)$</td>
<td>$k$-fold convolved subsampling distribution</td>
<td>160</td>
</tr>
<tr>
<td>$\gamma(k)$</td>
<td>autocovariance function</td>
<td>8</td>
</tr>
<tr>
<td>$D$</td>
<td>long-range dependence parameter</td>
<td>8</td>
</tr>
<tr>
<td>$d_2(\cdot, \cdot)$</td>
<td>Mallow’s metric</td>
<td>175</td>
</tr>
<tr>
<td>$d_{BL}(\cdot, \cdot)$</td>
<td>metric, based on bounded Lipschitz functions</td>
<td>27</td>
</tr>
<tr>
<td>$D_H(B)$</td>
<td>space of càdlàg functions mapping from $B$ to $H$</td>
<td>22</td>
</tr>
<tr>
<td>$d_k$</td>
<td>near-epoch dependence coefficient</td>
<td>20</td>
</tr>
<tr>
<td>$d_{n,m}$</td>
<td>normalizing sequence under long memory (sometimes $d_n$)</td>
<td>16</td>
</tr>
<tr>
<td>$D_n(t, x)$</td>
<td>empirical bridge process</td>
<td>50</td>
</tr>
<tr>
<td>$E^*$</td>
<td>conditional expectation given the observation $X_1, \ldots, X_n$</td>
<td>25</td>
</tr>
<tr>
<td>$(\epsilon_i)_i$</td>
<td>innovations of a linear process</td>
<td>12</td>
</tr>
<tr>
<td>$\mathcal{F}^k_i$</td>
<td>$\sigma$-field generated by random variables $\xi_i, \ldots, \xi_k$</td>
<td>20</td>
</tr>
<tr>
<td>$F(x)$</td>
<td>cumulative distribution function</td>
<td>2</td>
</tr>
<tr>
<td>$F^{-1}(x)$</td>
<td>generalized inverse of $F$</td>
<td>28</td>
</tr>
<tr>
<td>$\hat{F}_n(x)$</td>
<td>empirical distribution function</td>
<td>4</td>
</tr>
<tr>
<td>$F_n(x)$</td>
<td>deterministic distribution function, $F_n(\omega) \to F(x)$</td>
<td>36</td>
</tr>
<tr>
<td>$\tilde{F}_{n,K}(x)$</td>
<td>smoothed empirical distribution function</td>
<td>109</td>
</tr>
<tr>
<td>$\tilde{F}_{n,i}(x)$</td>
<td>$E^*[n^{-1} \sum_{i=1}^n 1_{{X_i \leq x}}]$</td>
<td>104</td>
</tr>
<tr>
<td>$G(\cdot)$</td>
<td>(possibly nonlinear) transformation</td>
<td>9</td>
</tr>
<tr>
<td>$G_n(\cdot)$</td>
<td>transformations, such that $G_n(X) \sim F_n(\omega)(x)$</td>
<td>39</td>
</tr>
<tr>
<td>$H$</td>
<td>either Hurst coefficient or a Hilbert space</td>
<td>12, 4</td>
</tr>
<tr>
<td>$H_q(x)$</td>
<td>Hermite polynomial of order $q$</td>
<td>10</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>$J_m(x)$</td>
<td>$m$th Hermite coefficient of $1_{{G(\cdot) \leq x}}$</td>
<td>17</td>
</tr>
<tr>
<td>$\hat{J}_{m,n}(x)$</td>
<td>estimator for $J_m(x)$</td>
<td>106</td>
</tr>
<tr>
<td>$\mathcal{L}(X)$</td>
<td>law of $X$</td>
<td>27</td>
</tr>
<tr>
<td>$l$</td>
<td>block length for bootstrap and subsampling</td>
<td>25</td>
</tr>
<tr>
<td>$L_n(x)$</td>
<td>subsampling distribution function</td>
<td>5</td>
</tr>
<tr>
<td>$L(x)$</td>
<td>slowly varying function</td>
<td>8</td>
</tr>
<tr>
<td>$m$</td>
<td>Hermite rank</td>
<td>11</td>
</tr>
<tr>
<td>$N(\mu,\sigma^2)$</td>
<td>normal distribution with mean $\mu$ and variance $\sigma^2$</td>
<td>159</td>
</tr>
<tr>
<td>$O(\cdot)$</td>
<td>big $O$ (Landau notation)</td>
<td>18</td>
</tr>
<tr>
<td>$o(\cdot)$</td>
<td>small $o$ (Landau notation)</td>
<td>56</td>
</tr>
<tr>
<td>$o_p(\cdot)$</td>
<td>small $o$ in probability</td>
<td>44</td>
</tr>
<tr>
<td>$P^*$</td>
<td>conditional probability given $X_1,\ldots,X_n$</td>
<td>26</td>
</tr>
<tr>
<td>$P_k$</td>
<td>projection operator</td>
<td>23</td>
</tr>
<tr>
<td>$\phi$</td>
<td>standard normal density</td>
<td>10</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>standard normal distribution function</td>
<td>46</td>
</tr>
<tr>
<td>$\psi_\tau(t)$</td>
<td>change-point function</td>
<td>42</td>
</tr>
<tr>
<td>$q_1-\alpha$</td>
<td>upper $\alpha$-quantile</td>
<td>68</td>
</tr>
<tr>
<td>$S_{n,\text{SUB}}(x)$</td>
<td>subsampling distribution</td>
<td>159</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>variance of a real random variable</td>
<td>5</td>
</tr>
<tr>
<td>$\sigma^2_{n,\text{SUB}}$</td>
<td>variance of the subsampling distribution</td>
<td>161</td>
</tr>
<tr>
<td>$\theta$</td>
<td>parameter</td>
<td>2</td>
</tr>
<tr>
<td>$\hat{\theta}_{ck}$</td>
<td>estimator for $\theta$, based on observations $X_i,\ldots,X_k$</td>
<td>2</td>
</tr>
<tr>
<td>$T_n$</td>
<td>test statistic</td>
<td>3</td>
</tr>
<tr>
<td>$T_n^*$</td>
<td>bootstrapped statistic</td>
<td>3</td>
</tr>
<tr>
<td>$\tau$</td>
<td>location of the change-point is assumed to be $\lfloor n\tau \rfloor$</td>
<td>39</td>
</tr>
<tr>
<td>$\tau_b$</td>
<td>normalizing sequence (in the subsampling context)</td>
<td>30</td>
</tr>
<tr>
<td>$W(t)$</td>
<td>Brownian motion</td>
<td>5</td>
</tr>
<tr>
<td>$X_i^*$</td>
<td>bootstrapped random element</td>
<td>5</td>
</tr>
<tr>
<td>$Z$</td>
<td>random variable with normal distribution</td>
<td>4</td>
</tr>
<tr>
<td>$Z_{m,H}(t)$</td>
<td>Hermite process of order $m$</td>
<td>15</td>
</tr>
<tr>
<td>$\tilde{Z}_{m,H}(t)$</td>
<td>Hermite bridge</td>
<td>42</td>
</tr>
<tr>
<td>$|\cdot|$</td>
<td>unspecified norm</td>
<td>2</td>
</tr>
<tr>
<td>$|\cdot|_{TV}$</td>
<td>total variation norm</td>
<td>62</td>
</tr>
<tr>
<td>$\overset{D}{\rightarrow}$</td>
<td>convergence in distribution</td>
<td>4</td>
</tr>
<tr>
<td>$\overset{P}{\rightarrow}$</td>
<td>weak convergence with respect to $P^*$</td>
<td>27</td>
</tr>
<tr>
<td>$\sim$</td>
<td>distributed as or asymptotic equal to</td>
<td>25, 9</td>
</tr>
<tr>
<td>$\approx$</td>
<td>asymptotic proportional to</td>
<td>11</td>
</tr>
<tr>
<td>APC</td>
<td>almost periodically correlated</td>
<td>169</td>
</tr>
<tr>
<td>Symbol</td>
<td>Definition</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>ARE</td>
<td>asymptotic relative efficiency</td>
<td>66</td>
</tr>
<tr>
<td>LRD</td>
<td>long-range dependence</td>
<td>2</td>
</tr>
<tr>
<td>NED</td>
<td>near-epoch dependence</td>
<td>20</td>
</tr>
<tr>
<td>SRD</td>
<td>short-range dependence</td>
<td>11</td>
</tr>
</tbody>
</table>
1 Introduction

1.1 Change-point analysis in time series

Let us consider a time series $(X_i)_{i \geq 1}$, that is a sequence of observations, sampled at equally distanced time points. Examples of time series are vast, for instance climate data (such as rainfall, flood or temperature data), financial observations, such as stock returns, or network traffic, just to name a few. Statistical inference is often made under the assumption of stationarity of the sequence $(X_i)_{i \geq 1}$. That is, the distribution of $(X_{i+j}, \ldots, X_{i+d})$ is assumed to be independent of $j$. Reasonable deviations from stationarity are the existence of a trend or a seasonable component. Then again, these components still resemble some type of regular behavior of the time series. On the contrary, one might consider an abrupt change in a certain “characteristic” of the time series. This characteristic might be the mean, the variance, the dependence structure of the series or the marginal distribution function. The occurrence of such a change will be called change-point. Now assume that there is a given time series, possibly containing a change-point. When one speaks of change-point analysis, usually one of the following methods is applied to the series:

- **Offline-testing**: The data from the whole observation period are available. Based on those, we want to decide whether there is a change-point in the data or not. This decision is based on a statistical test.

- **Online-testing**: Only a small part of the data is given at the beginning of the procedure (the so-called training period). As new observations come in, one wishes to detect the occurrence of a change-point as soon as possible.

- **Estimation**: Assume that we already know that a change-point occurred. However, the time of this change is unknown and has to be estimated (using a data-dependent method).

- **Combined procedures**: Recently, there has been a growing interest in multiple change-point problems. Several new methods have been investigated, testing for changes and simultaneously estimating their location.

In this thesis we will solely consider offline-tests. However, the methods developed here might be used to construct online-tests and estimators as well. As mentioned before, detection of a
possible change will be done via statistical testing. In detail, we test the null-hypothesis of stationary data against the alternative of the existence of a change-point. Now, what type of statistic can be used for this test problem? The classical change-point test can be seen as a further development of a two-sample test. If the change occurs at the time point \( k_0 \), one might divide the series and apply a two-sample test to \( X_1, \ldots, X_{k_0} \) and \( X_{k_0+1}, \ldots, X_n \). Let \( \hat{\theta} \) be an estimator for the parameter of interest (the parameter in which the change occurs), say \( \theta \), then the mentioned two-sample test might be based on the difference \( \| \hat{\theta}_{1:k_0} - \hat{\theta}_{k_0+1:n} \| \), for a suitable norm \( \| \cdot \| \). However, the theory of change-point analysis has become so rich, as the time of the change is typically unknown. Thus a two-sample test would not suffice. Instead, all possible time points are taken into account, leading to the quantity \( \max_{1 \leq k < n} \| \hat{\theta}_{1:k} - \hat{\theta}_{k+1:n} \| \).

The change-points we are interested in this thesis are a mean-change in Hilbert space-valued data (\( \theta = \langle X_i \rangle \)) and a change in the marginal distribution of real-valued observations (\( \theta = F(\cdot) = P(X_i \leq \cdot) \)). The statistic developed above then becomes

\[
\max_{1 \leq k < n} \left\| \frac{1}{k} \sum_{i=1}^{k} X_i - \frac{1}{n-k} \sum_{i=k+1}^{n} X_i \right\|, \tag{1.1}
\]

for observations \( X_i \) taking values in a Hilbert space, or

\[
\max_{1 \leq k < n} \left\| \frac{1}{k} \sum_{i=1}^{k} 1\{X_i \leq x\} - \frac{1}{n-k} \sum_{i=k+1}^{n} 1\{X_i \leq x\} \right\|, \tag{1.2}
\]

with \( X_i \) being \( \mathbb{R} \)-valued, respectively. The quantity in (1.1) is called CUSUM statistic. Depending on the norm, the statistic in (1.2) is called Kolmogorov-Smirnov statistic (for \( \| \cdot \| = \sup_x |\cdot| \)) or Cramér-von Mises statistic (for \( \| \cdot \| = \int (\cdot)^2 dF \)). As a last step to a proper statistical test, critical values are needed. Ideally, they are taken as quantiles of the distribution of the test statistics. As we consider dependent observation, the exact distributions are unknown and we will get along with asymptotic distributions instead. In order to obtain the latter, we will make use of functional (non-)central limit theorems. The nature of these results is largely influenced by the dependence of the underlying random variables.

In this thesis we are dealing mainly with two types of time series. The first are long-range dependent (LRD) processes. The serial correlation of such sequences decays hyperbolically and consequently it is not absolutely summable. Analysis of such series was motivated by empirical evidence, see for instance Hurst (1951). In addition, we consider functional, short-range dependent time series. That is, each “data point” is a function, describing for instance a daily/monthly/yearly curve of observations. For such sequences, even a change in the mean (which is actually a change in the mean function) becomes a complex setting. In both situations we will derive the needed limit theorems. In the long memory case we will make further use of this results in order to compare different tests. This comparison is worked out via the so-called
asymptotic relative efficiency.

However, up to this point, this remains a purely theoretic procedure. The crux of this approach (that is using asymptotic critical values) is that the limiting distribution might depend on a nuisance parameter. Moreover, in the situations encountered in this thesis this parameter will be of infinite dimension. Thus a straightforward estimation is challenging.

Instead we will make use of the bootstrap. Developed by Efron (1979), its purpose is the approximation of the distribution of a statistic $T_n$ by reusing the original data. To this end one draws independent random variables from the empirical distribution of the original data. Eventually, these pseudo observations are used to construct a (bootstrap) statistic $T^*_n$, whose distribution approximates that of $T_n$.

The bootstrap was investigated for independent data, but many extensions to dependent observation have been proposed. Here we consider the block bootstrap, see Künsch (1989) for an important contribution. The data is reused block wise and thereby the dependence of the underlying sequence is maintained. Validity of the bootstrap has to be proved for each combination of statistic/stochastic process and dependency condition individually, and so will we do in this thesis. Eventually, critical values for our change-point tests can be deduced.

A different data resampling approach is due to Politis and Romano (1994a) - the so-called subsampling. It treats data blocks as small scale renditions of the original sample. Hence, the statistic of interest is computed on each block. Finally the empirical distribution of this set of statistics is used to approximate the distribution of the original statistic. In this thesis we will not study the subsampling distribution itself, but $k$-fold convolutions of it. In detail, we will prove convergence of this (empirical) distributions to a normal limit.

The motivation is two-fold: First, it turns out that, in the fundamental case of sample means, the block bootstrap is a $k$-fold convolution of the subsampling distribution (centered and normalized). Using this fact, we aim at proving results for the bootstrap under weaker conditions than previously considered.

Secondly, a general theory for convolved subsampling is of interest in its own right. It can be computationally less demanding than the bootstrap, while enhancing ordinary subsampling (see Lenart (2016) and Sharipov et al. (2016b) for numerical studies).

1.2 Main results of this thesis

The change-point procedures of this thesis are established (in principle) via the following three steps:

(i) A functional limit theorem is derived either for a partial sum process or an empirical process.

(ii) The change-point statistic is written as functional of the process from (i). Subsequently, its asymptotic distribution is investigated.
(iii) Bootstrap versions of the former results are established.

All steps are executed both under stationarity and under local alternatives. This general proceeding leads to the following findings:

Chapter 3 deals with the distributional change in long memory sequences and is mainly based on the article Change-point tests under local alternatives for long-range dependent processes, see Tewes (2017).

By the work of Dehling and Taqqu (1989) we know the limit behavior of the empirical process of stationary long memory data. One has

\[ \tau_n (F_n(x) - F(x)) \xrightarrow{D} J(x)Z, \]

with a normalizing sequence \((\tau_n)_n\), a deterministic function \(J(x)\) an a (possibly non-normal) random part \(Z\). Consequently, the limit is called semi-degenerate. In Theorems 5 and 6 we show that such a result still holds in presence of a change-point. Subsequently, we are able to deduce the asymptotic distributions of the Kolmogorov-Smirnov and Cramér-von Mises statistic (see (1.2)) in Theorem 7. To this end, general local alternatives are formulated for the first time under a large class of LRD sequences.

Two more change-point tests are investigated under long memory, the CUSUM test and the Wilcoxon test. To this end we extend the findings in Dehling et al. (2012, 2017)). The tests serve mainly as comparative value, but the asymptotic distributions are of interest in their own right too. Those are given in Theorems 8 and 9. Finally, we compute the asymptotic relative efficiency (ARE) of the mentioned tests in Examples 3.4.3 and 3.4.4 and Theorem 11.

The asymptotics of the four mentioned tests all depend on an infinite dimensional nuisance parameter, namely the function \(J(x)\). Our solution of this problem, presented in Chapter 4, is the bootstrap. The results are based on the article Block bootstrap for the empirical process of long-range dependent data, see Tewes (2016). We state limit theorems for the bootstrapped (smoothed) empirical process under long memory (Theorems 13 and 15), which has not been considered in the literature before. It will be shown that the limit process is of the form \(J(x)Z^*\). The function \(J(x)\) is the same as above, but the stochastic term \(Z^*\) is always normal. However, we propose an estimator for \(J(x)\) and show its consistency (Theorem 14), even if the bootstrap technically fails. This estimator might be used under stationarity as well as under change-point alternatives.

Chapter 5 treats a different change-point setting, namely the aforementioned mean-change in observations taking values in a Hilbert space \(H\). It is based on the paper Sequential block bootstrap in a Hilbert space with application to change point analysis, which is joint work with
Introduction

O. Sharipov and M. Wendler (see also Sharipov et al. (2016a)). We test for this mean-change using a Hilbert space analogue of the CUSUM test. To this end, a functional central limit theorem in \( H \) is established. In detail

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i) \overset{D}{\rightarrow} W(t),
\]

see Theorem 16. The process \((W(t))_t\) is a Hilbert-space analogue of Brownian motion with unknown covariance operator. This motivates a bootstrap version of the partial sum process, namely \( n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} (X_i^n - \bar{X}_n) \). Its convergence towards the same Brownian motion is stated in Theorem 17. Asymptotics of the change-point statistic and its bootstrap analogue are presented in Corollaries 5.2.1, 5.2.2 and 5.2.4.

Finally, Chapter 6 is based on the article *Convolved subsampling estimation with applications to block bootstrap*, see Tewes et al. (2017). It is joint work with D. Nordman and D. Politis. Let \( T_n \) be a statistic and \( L_n(x) \) its subsampling distribution. Consistency, that is \( \sup_x |L_n(x) - P(T_n \leq x)| \overset{P}{\rightarrow} 0 \), has been investigated for large classes of statistics and time series, see Politis et al. (1999) and Betken and Wendler (2017+). Here we consider the (centered and normalized) \( k \)-fold convolution of the subsampling distribution, that is

\[
C_{n,k}(x) = L_n \ast L_n \ast \cdots \ast L_n (x\sqrt{k} + km_n).
\]

Theorems 18 and 19 give conditions under whom consistency of \( C_{n,k}(x) \) follows from that of \( L_n(x) \). A central (and sometimes even necessary) requirement is the convergence of the subsampling variance, in detail

\[
\hat{\sigma}^2_{SUB,n} = \int x^2 dL_n(x) - \left( \int xdL_n(x) \right)^2 \overset{P}{\rightarrow} \sigma^2.
\]

Theorem 20 states several conditions to ensure this convergence, of whom some are easy to verify in praxis.
2 Probabilistic background

In this chapter we present some concepts from probability theory and mathematical statistics that are inevitable for the main results of this thesis. Thereby it lays not only the foundation of the theoretical work but also serves as a general introduction to the topics long-range dependence, empirical processes, probability in Hilbert spaces and block bootstrap.

Not all change-point tests that will be investigated in the following chapters are completely new. However, they are applied in situations never before considered. More precisely, we deal with real-valued, long-range dependent processes and short-range dependent sequences taking values in a Hilbert space. For these two classes of observations we state the basic definitions, some useful Lemmata and the most important results from literature. Special focus will be laid on the empirical process of long memory data. It has some unique asymptotic properties which carry over to the limit theorems of Chapters 3 and 4.

A second topic of this thesis is the bootstrap. In fact, we use the moving blocks bootstrap of Künsch (1989) and the nonoverlapping bootstrap of Carlstein (1986), which are both described below. In order to justify the bootstrap asymptotically, classical weak convergence does not suffice. Here we present two approaches used in the literature: Convergence of the conditional distribution, given the observations, and joint convergence of the bootstrap and the original statistic. Once such a convergence is established, the precise justification of the validity of the bootstrap is omitted in many articles. In this chapter we give arguments that ensure this validity.

2.1 Long-range dependence

2.1.1 Definition and motivation of long-range dependence

Several parts of this thesis are dedicated to time series with a very special dependence structure - so called long-range dependent (LRD) processes. The serial correlation of such sequences decays extremely slow, which has a dramatic effect on all kinds of statistical methods. During the 20th century the notion of long-range dependence has been developed, starting with empirical evidence through to a theoretical formulation. Some milestones are the work of Hurst (1951) (who studied the yearly discharge of the Nile river, showing features that cannot be explained via short-range dependence), the discovery of LRD in economics by Granger (1966) and the proof of a limit theorem, where non-normality of the limit is caused by the dependence of the
Long-range dependence observations, see Rosenblatt (1961).

Although this thesis is mainly focused on the theoretical aspects of LRD and its interplay with change-points, possible applications to real-life data are always kept in mind. Besides the mentioned hydrological data, LRD has been found in financial time series, network traffic, climate data in general and even in observations from biology. For comprehensive monographs on long-range dependence see Beran (1994) (with much emphasis on statistics), Beran et al. (2013) and the recently published book of Pipiras and Taqqu (2017).

In order to define LRD we make use of slowly varying functions, see Bingham et al. (1989) for a broad treatment.

**Definition 2.1.1.** A function $L$ is called *slowly varying at infinity*, if it is positive and for any $u > 0$

$$\frac{L(ux)}{L(x)} \to 1, \text{ as } x \to \infty.$$  

**Definition 2.1.2.** Let $(X_i)_{i \geq 1}$ be a second order stationary sequence of real-valued random variables with autocovariance function $\gamma(k) = EX_iX_{i+k} - EX_iEX_{i+k}$. We call $(X_i)_{i \geq 1}$ *long-range dependent* (LRD), if there is a $D \in (0, 1)$ and a slowly varying function $L(x)$, such that

$$\gamma(k) = k^{-D}L(k). \quad (2.1)$$

The terms *long memory* and *strong dependence* are often used synonymously for LRD, and we will do so in this thesis too. However, note that the latter term is a bit misleading, as the correlation between two observations $X_i$ and $X_j$ might be very weak. The crucial point is that it persists over large time periods. This has the effect that the autocovariance function is not absolutely summable.

To see this, we need some basic properties of slowly varying functions:

Let $L(x)$ be slowly varying at infinity. For any $\delta > 0$ there exists $C_1 > 0, C_2 > 0$ and $U \geq 0$, such that for all $u > U$

$$C_1u^{-\delta} \leq L(u) \leq C_2u^\delta. \quad (2.2)$$

(The result is an immediate consequence of *Potter’s Bound* (Theorem 1.5.6. in Bingham et al. (1989)).

Next let $c_k = L(k)k^p$ for arbitrary $p > -1$. Then, by Proposition 2.2.1 in Pipiras and Taqqu...
(2017), one has for $n \to \infty$

$$\sum_{k=1}^{n} c_k \sim \frac{L(n)n^{p+1}}{p+1}. \quad (2.3)$$

if now consider a covariance function satisfying (2.1). Then (2.3) implies

$$\sum_{k=-n}^{n} \gamma(k) = \gamma(0) + 2 \sum_{k=1}^{n} \gamma(k) \sim \frac{L(n)n^{-D+1}}{-D + 1},$$

which converges to infinity by (2.2). In fact, $\sum_{k=-\infty}^{\infty} \gamma(k) = \infty$ is sometimes found as another requirement for long-range dependence. Pipiras and Taqqu (2017) give further definitions, which are based on spectral theory or representations as linear processes. Alternatively, one might drop the use of covariance functions and therefore define long-range dependence for heavy-tailed data. Another approach, via so called long memory-stochastic volatility-models, considers uncorrelated time series, but with $(X_t^2)_{t \geq 1}$ showing features of LRD. However, for the models, situations and techniques investigated in this thesis, it suffices to work with Definition 2.1.2.

### 2.1.2 Subordinated Gaussian processes and Hermite polynomials

The model for LRD time series we are mainly interested in are subordinated Gaussian processes. Not only have many theoretic characteristics of long memory been first encountered for such processes, but they are still object of current research.

Let $(X_i)_{i \geq 1}$ be a stationary Gaussian process (that is $(X_{i_1}, \ldots, X_{i_d})$ has a multivariate normal distribution for all $d \geq 1$ and all $(i_1, \ldots, i_d)$), satisfying

$$EX_i = 0, \quad EX_i^2 = 1, \quad EX_iX_{i+k} = k^{-D}L(k), \quad (2.4)$$

for $0 < D < 1$ and a slowly varying function $L(x)$. Therefore, $(X_i)_{i \geq 1}$ mets our definition of long-range dependence. Examples for such processes will be given in the next two sections.

We will not only consider the $X_i$’s, but also transformations of these. Let $G: \mathbb{R} \to \mathbb{R}$ be a measurable function, then the sequence $(G(X_i))_{i \geq 1}$ is called subordinated Gaussian process.

The natural question arises wether such time series share the same dependence structure as the underlying Gaussian random variables, or if they are still long-range dependent at all? A concept that yields the answer is an $L^2$-representation of $G$ via polynomials. The Hermite polynomials $(H_q)_{q \geq 0}$ are given by

$$H_q(x) = (-1)^q e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}, \quad x \in \mathbb{R}.$$ 

---

1Here, $\sim$ means asymptotically equal to, in detail $a(n) \sim b(n)$ if $a(n)/b(n) \to 1$. 

---
Particularly,
\[ H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \quad H_3(x) = x^3 - 3x. \]

It can be shown (see Pipiras and Taqqu (2017), Propositions 5.1.1 and 5.1.3) that the collection \( (H_q)_{q \geq 0} \) forms an orthogonal basis of the space \( L^2(\phi) \), with \( EH_q(X)^2 = q! \). If the transformation \( G \) lies in this space (that is, if \( EG(X)^2 < \infty \)), one gets
\[
G(X) = \sum_{q=0}^{\infty} \frac{c_q}{q!} H_q(X),
\tag{2.5}
\]
with \( c_q = EG(X)H_q(X) \).

The convergence in (2.5) takes place in \( L^2(\phi) \), the right hand side of (2.5) is called Hermite expansion and the sequence \( (c_q)_{q \geq 0} \) Hermite coefficients. We will see that this concept matches very well with our notion of long-range dependence. The connection between long memory and the Hermite expansion is grounded in the following result, which is basically Proposition 5.1.4 in Pipiras and Taqqu (2017).

**Proposition 2.1.3.** Let \((X, Y)\) be a Gaussian vector with \( EX = EY = 0 \) and \( EX^2 = EY^2 = 1 \). Further, let \( G_1, G_2 \in L^2(\phi) \) with Hermite coefficients \((c_{1,q})_{q \geq 0}\) and \((c_{1,q})_{q \geq 0}\), respectively. Then
\[
EG_1(X)G_2(Y) = \sum_{q=0}^{\infty} c_{1,q}c_{2,q}(EXY)^q.
\]

Now we are able to examine the covariance structure of the subordinated process \((G(X_i))_{i \geq 1}\). Taking \( G = G_1 = G_2 \) in Proposition 2.1.3, one obtains
\[
Cov(G(X_i), G(X_i + k)) = \sum_{q=1}^{\infty} c_q^2q!(\gamma(k))^q
= \sum_{q=1}^{\infty} c_q^2q!(L(k)k^{-D})^q. \tag{2.6}
\]

The function \( L \) is slowly varying and \( D \in (0, 1) \), thus the first summand in (2.6) is asymptotically the dominating term. However, this term not necessarily corresponds to \( q = 1 \), as the Hermite coefficient \( c_1 \) might be zero. We therefore introduce the following definition: The number \( m \in \mathbb{N} \) with
\[
m = \min\{q \geq 1 \mid EG(X)H_q(X) \neq 0\}\]
is called *Hermite rank* of $G$. With this definition we get for (2.6)

$$
\sum_{q=1}^{\infty} c_q^2 q! (L(k)k^{-D})^q \sim c_m^2 m! L^m(k)k^{-mD}.
$$

(2.7)

It is easy to see that $L_m = c_m^2 m! L^m(k)$ is again slowly varying. Thus this covariance function satisfies condition (2.1), if $mD < 1$. To summarize, the transformation $G$ possibly affects the dependence of the time series in one of the following four ways:

- **$m = 1$**: The sequence $(G(X_i))_{i \geq 1}$ inherits the long-range dependence from $(X_i)_{i \geq 1}$.
- **$m > 1, mD < 1$**: The sequence $(G(X_i))_{i \geq 1}$ is still long-range dependent, but in a weaker sense than $(X_i)_{i \geq 1}$.
- **$mD > 1$**: The autocovariance function is absolutely summable, thus $(G(X_i))_{i \geq 1}$ is short-range dependent.
- **$mD = 1$**: Autocovariances are not absolutely summable, but condition (2.1) is not satisfied. This is sometimes referred to as *intermediate dependence*.

In this thesis we will exclusively consider the first two situations. For such sequences we obtain the following result.

**Lemma 2.1.4.** Let $(X_i)_{i \geq 1}$ be a stationary Gaussian process satisfying (2.4) and let $q \in \mathbb{N}$ with $qD < 1$. Then

$$
\text{Var} \left( \sum_{i=1}^{n} H_q(X_i) \right) \approx L_q(n) n^{2-qD},
$$

where the constant of proportionality\(^2\) is $2q!(1-qD)^{-1}(2-qD)^{-1}$.

Let also $G \in L^2(\phi)$ be a function with Hermite rank $m$, such that $mD < 1$. Then

$$
\text{Var} \left( \sum_{i=1}^{n} G(X_i) \right) \approx L_1(n) n^{2-mD},
$$

for a slowly varying function $L_1$.

The first part of this Lemma is Theorem 3.1 of Taqqu (1975) and the second follows from (2.6). This is the first result showing how LRD leads to clearly different effects than SRD (short-range dependence). In the latter case the variance of the partial sum is known to be of

\(^2\)Here, $\approx$ means *asymptotically proportional to*, in detail $a(n) \approx b(n)$ if there is a $C > 0$ with $a(n)/b(n) \to C$. 

order \( n \). This has also consequences for statistical applications and limit theorems. To obtain an asymptotic distribution, the right normalization for many statistics of LRD data is

\[
\left( \text{Var} \left( \sum_{i=1}^{n} H_q(X_i) \right) \right)^{1/2} \approx L^{m/2}(n)n^{1-mD/2} = L^{m/2}(n)n^H.
\]

The exponent \( H = 1 - mD/2 \) is called Hurst coefficient and the constant of proportionality is actually \( 2^m(1 - mD)^{-1}(2 - mD)^{-1} \), see Taqqu (1975). It is important to note that \( H > 1/2 \).

When we consider different long memory sequences in simulation studies (see Chapters 3 and 4), the different strengths of dependence will be described via the Hurst coefficient \( H \). It is usually unknown in praxis and the effect on statistical tests will be discussed in the mentioned chapters too.

### 2.1.3 Long memory linear processes

Besides subordinated Gaussian processes, an important class of LRD sequences are linear processes. A causal (one-sided) linear process is given through

\[X_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j},\]

with i.i.d. random variables \((\epsilon_j)_{j \geq 1}\), and \((a_k)_{k \in \mathbb{Z}}\) being some deterministic sequence. If one assumes \( E\epsilon_j = 0, E\epsilon_j^2 < \infty \) and \( \sum_{j=0}^{\infty} a_j^2 < \infty \), then \( X_n \) has finite second moments and the series is well defined (since it converges in \( L^2 \)). We establish long-range dependence via the coefficients \((a_j)_j\), requiring

\[a_j = L(j) j^{d-1},\]

for a slowly varying function \( L \) and \( d \in (0, 1/2) \). Lemma 2.1 in Beran et al. (2013) yields the covariance function under this assumption, in detail

\[EX_i X_{i+k} = \gamma(k) \sim L\gamma(k) k^{2d-1},\]  \hspace{1cm} (2.8)

with \( L\gamma(k) = E\epsilon_1^2 L^2(k) B(1-2d, d) \) and \( B(x, y) = \Gamma(x) \Gamma(b)/\Gamma(a + b) \) being the beta-function. Consequently, this process satisfies condition (2.1) with \( D = 1-2d \) and is long-range dependent.

**Example 2.1.5** (Gaussian linear process). Let \( \epsilon_i \) be i.i.d. standard normal and \((a_j)_j\) as above. Then the linear process \( X_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j} \) is a long memory, Gaussian process as considered in the previous section. In this case, the Hurst coefficient is given by \( H = 1 - D/2 = 1/2 + d \) (for a Hermite rank \( m = 1 \)).
On the other hand, any Gaussian process has a representation as a one-sided linear process if and only if its spectral density $\psi(\cdot)$ satisfies $\int_{-\pi}^{\pi} \log \psi(\lambda) d\lambda < \infty$, see Brockwell and Davis (1991).

**Example 2.1.6 (FARIMA(0, $d$, 0)-process).** Another special case are so-called FARIMA(0, $d$, 0)-processes, see Granger and Joyeux (1980). They are defined by

$$X_k = (I - B)^{-d}Z_k,$$

where $B$ is the backshift operator and $(Z_k)_{k \in \mathbb{Z}}$ a zero-mean, i.i.d. sequence with $EZ_0^2 < \infty$. Then $X_k$ has the following representation

$$X_k = \sum_{j=0}^{\infty} b_j Z_{k-j},$$

with $b_j$ being the coefficients of the Taylor expansion of $(1 - z)^d$. In detail, the coefficients are given by

$$b_j = \frac{\Gamma(j + d)}{\Gamma(j + 1) \Gamma(d)} \sim j^{d-1} \frac{\Gamma(d)}{\Gamma(1)}.$$

The asymptotic proportionality can be seen via Stirling’s formula $\Gamma(x) \sim \sqrt{2\pi e^{-x}}(x - 1)^{(x - 1/2)}$.

**Example 2.1.7 (FARIMA($p$, $d$, $q$)-process).** Finally we consider a model that combines FARIMA(0, $d$, 0)-process with the classical ARMA($p$, $q$)-series. Consider the polynomials

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \quad \text{and} \quad \psi(z) = 1 + \psi_1 z + \cdots + \psi_q z^q,$$

which have no common zeros and where $\phi$ has no unit roots. Then define $X_k$ by

$$X_k = \phi^{-1}(B)\psi(B)(I - B)^{-d}Z_k.$$

If $b_k$ are the coefficients of the Laurent expansion, in detail $\phi^{-1}(z)\psi(z)(1-z)^{-d} = \sum_{k=-\infty}^{\infty} b_k z^k$, one gets

$$X_k = \sum_{j=-\infty}^{\infty} b_j Z_{k-j} \quad \text{and} \quad b_k \sim \frac{\psi(1)}{\phi(1)} \frac{k^{d-1}}{\Gamma(d)}.$$

As (2.8) holds for two-sided linear processes too, $(X_i)_{i \geq 1}$ is indeed long-range dependent. Moreover, $\phi^{-1}(B)\psi(B)$ introduces short-range dependent effects to the model, making this concept quite flexible.

With the FARIMA($p$, $d$, $q$)-series we gave an example for long memory, two-sided linear pro-
cesses. They are usually more complicated to treat, but several results have been proved for this class, see for example Giraitis and Surgailis (2002).

Just as in the Gaussian case, one also might consider subordinated linear processes \( (G(X_i))_{i \geq 1} \). In view of empirical processes, special focus is laid on the transformation \( G_t(x) = 1_{x \leq t} \).

Two concepts have been investigated to analyze such sequences: An expansion in Appell polynomials (see Surgailis (1983) and Avram and Taqqu (1987)) and a martingale approach/Volterra expansion (see Ho and Hsing (1996)).

When we analyze linear processes in Chapter 4, our investigation will be based on results already existing. Therefore, further details on the techniques are omitted here.

### 2.1.4 Fractional Brownian motion and Hermite processes

We have seen that the variance of \( \sum_{i=1}^{n} X_i \) grows with a larger rate than \( n \), if \( (X_i)_{i \geq 1} \) is long-range dependent. Another striking difference from short-range dependence is the limit behavior of the partial sum process. For Gaussian random variables, Taqqu (1975) showed that \( \sum_{i=1}^{[nt]} X_i \) (properly normalized) does not converge towards Brownian motion but to fractional Brownian motion.

**Definition 2.1.8.** A zero-mean Gaussian process \( (B_H(t))_{t \in [0,1]} \) is called fractional Brownian motion (fBm) with Hurst parameter \( H \in (0,1) \) if its covariance function is given by

\[
\text{Cov}(B_H(t), B_H(s)) = \frac{\sigma^2}{2} \left( |t|^{2H} + |s|^{2H} - |t-s|^{2H} \right), \quad t, s \in \mathbb{R}.
\]

It is called standard fBm, if \( \sigma^2 = 1 \).

**Remark 2.1.9.** (i) For the special case \( H = 1/2 \) one obtains the classical Brownian motion.

(ii) An alternative definition of fractional Brownian motion might be given through a Wiener-Î­to-Dobrushin integral. In fact, define

\[
Z_H(t) = K_H \int_{\mathbb{R}} \frac{e^{i\lambda t} - 1}{i\lambda} \frac{1}{|\lambda|^{H-1/2}} dW(\lambda), \quad (2.9)
\]

with \( W \) being the complex, Gaussian random measure specified by \( W(A) = \overline{W(-A)} \) and \( EW(A)W(B) = |A \cap B| \) (for all Borel-sets \( A, B \)) and

\[
K_H = \left( \frac{H(H-1)}{2\Gamma(2-2H)\sin([H-1/2]\pi)} \right)^{1/2}.
\]

Then \( Z_H(t) \) is actually a standard fractional Brownian motion.

**Example 2.1.10** (Fractional Gaussian Noise). Using fBm we might define another Gaussian sequence that satisfies the conditions of section (2.1.2). To this end let \( (B_H(t))_{t \in \mathbb{R}} \) be standard
fractional Brownian motion and set
\[ X_i = B_H(i + 1) - B_H(i), \quad i \in \mathbb{Z}. \]

Then the autocovariance function of \((X_i)_{i \in \mathbb{Z}}\) reads as follows:
\[ EX_i X_{i+k} = \gamma(k) = \frac{1}{2} (|k + 1|^{2H} - 2|k|^{2H} + |k - 1|^{2H}). \]

One obtains for \(H > 1/2\)
\[ \gamma(k) \sim H(2H - 1)k^{-(2-2H)}, \]
see Samorodnitsky and Taqqu (1994). In other words, \((X_i)_{i \in \mathbb{Z}}\) fulfills condition (2.1) and is long-range dependent. The sequence is called fractional Gaussian noise, see Mandelbrot and van Ness (1968), and because of its easy derivation from fBm it has become quite popular for simulating long memory time series.

Next consider a subordinated Gaussian process \((G(X_i))_{i \geq 1}\). One might ask whether the limiting process of the partial sum process is still fractional Brownian motion or not. It turns out that the answer depends on the Hermite rank \(m\) of \(G\) and that we have to define further stochastic processes.

**Definition 2.1.11.** An \(m\)-th order Hermite process with self-similarity parameter \(H\) is defined by
\[
Z_{m,H}(t) = K_{m,H} \int_{\mathbb{R}^m} e^{it \sum_{j=1}^m \lambda_j} \frac{1}{i} \prod_{j=1}^m (\lambda_j^{-(1/2-(1-H)/m)}dW(\lambda_1) \cdots dW(\lambda_m)),
\]
with \(W\) being the same Gaussian measure as in Remark 2.1.9. It is called standard, if
\[ K_{m,H} = \left( \frac{H(2H - 1)}{m! \{2\Gamma((2 - 2H)/m)\sin([1/2 - (1 - H)/m]\pi)\}^m} \right)^{1/2}. \]
Moreover \(\int''\) denotes integration, with the hyper-diagonals \(\{\lambda_j = \pm \lambda_i \mid i \neq j\}\) being excluded.

For \(m = 1\) one gets fractional Brownian motion and for \(m = 2\) the so-called Rosenblatt process (see Taqqu (1975)). It is not Gaussian and the same is true for all higher-order Hermite processes. For a mathematical background on multiple Wiener-Itô integrals, see Major (1981). Now we might state the asymptotic behavior of the partial sum process. The following theorem was independently proved by Taqqu (1979) and Dobrushin and Major (1979).

**Theorem 1** (Taqqu (1975, 1979), Dobrushin and Major (1979)). Let \((X_i)_{i \geq 1}\) be a stationary Gaussian process satisfying (2.4). Let also \(G \in L^2(\phi)\) be a function with Hermite rank \(m\), such
that $mD < 1$. If we define $d_{n,m}$ by
\[
d_{n,m}^2 = \text{Var} \left( \sum_{i=1}^{n} H_m(X_i) \right),
\]
then, as $n \to \infty$,
\[
d_{n,m}^{-1} \sum_{i=1}^{\lfloor nt \rfloor} (G(X_i) - EG(X_i)) \xrightarrow{D} \frac{c_m}{m!} Z_{m,H}(t).
\]

Weak convergence takes place in $D[0,1]$, equipped with the uniform metric.

Remark 2.1.12. (i) In distinction to classical results, Theorem 1 is sometimes called non-central limit theorem.
(ii) The process $(Z_{m,H}(t))_{t \in [0,1]}$ will show up in several limit theorems of this thesis.

2.2 Empirical processes of LRD data

Consider a set of identically distributed random variables $Y_1, \ldots, Y_n$. A natural estimator for the distribution function $F(x) = P(Y_1 \leq x)$ is the empirical distribution function
\[
F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1\{Y_i \leq x\} \quad x \in \mathbb{R}.
\]
Under very general assumptions on $(Y_i)_{i \geq 1}$ (for instance ergodicity) one has for fixed $x \in \mathbb{R}$
\[
F_n(x) \to F(x) \quad n \to \infty.
\]
Moreover, by a Glivenko-Cantelli-type argument, it follows that this convergence is uniform. The difference between empirical and theoretic distribution function, in detail
\[
W_n(x) = F_n(x) - F(x) \quad x \in \mathbb{R},
\]
is called empirical process, and it has a vast number of application in statistical inference. Most obvious are maybe goodness-of-fit tests, considering $\|W_n(x)\|$ for different norms. To set critical values one might be interested in the asymptotic distribution. Donsker (1952) showed that $(\sqrt{n}W_n(x))_{x \in \mathbb{R}}$ converges weakly to Brownian Bridge, if the $Y_i$’s are independent. A further development, with applications to change-point tests, is the sequential empirical process $(W_{\lfloor nt \rfloor}(x))_{x \in \mathbb{R}, t \in [0,1]}$. It is a function in two arguments - $x$ and $t$. For independent observations, weak convergence has been proved by Kiefer (1972) and Müller (1970) and the limit process is called Kiefer-Müller process. It is a zero-mean Gaussian process with covariance
kernel

\[ E[K(x_1, t_1)K(x_2, t_2)] = \min(t_1, t_2)(F(\min(x_1, x_2)) - F(x_1)F(x_2)). \]

Many extensions to short-range dependent data have been studied, see for example Sen (1974) for mixing sequences. The limit is a two-parameter Gaussian process too, but with a much more complicated covariance kernel.

It was obtained in Dehling and Taqqu (1989) that the situation changes drastically when the observations are long-range dependent. Consider a subordinated Gaussian sequence as in section 2.1.2 and the associated sequential empirical process \( \sum_{i=1}^{\lfloor nt \rfloor} (1_{\{G(X_i) \leq x\}} - F(x)) \). Now for fixed \( x \in \mathbb{R} \) the indicators have an Hermite expansion (cf. (2.5))

\[ 1_{\{G(X_i) \leq x\}} - F(x) = \sum_{q=m_x}^{\infty} \frac{J_q(x)}{q!} H_q(X_i), \]

with Hermite coefficients (Hermite coefficient functions) \( J_q(x) = E[1_{\{G(X_1) \leq x\}}H_q(X_1)] \). Moreover, we define the Hermite rank of the class of functions \((1_{\{G(X_i) \leq x\}} - F(x))_{x \in \mathbb{R}}\) by

\[ m = \min \{ q > 0 \mid J_q(x) \neq 0 \text{ for some } x \}. \]

If we set \( d_{n,m}^2 = \text{Var}(\sum_{i=1}^{n} H_m(X_i)) \), then the mentioned limit theorem of Dehling and Taqqu (1989) reads as follows.

**Theorem 2** (Dehling, Taqqu). Let the class of functions \((1_{\{G(\cdot) \leq x\}} - F(x))_{x \in \mathbb{R}}\) have Hermite rank \( m \) and let \( 0 < D < 1/m \). Then

\[ \frac{1}{d_{n,m}^2} \sum_{i=1}^{\lfloor nt \rfloor} (1_{\{G(X_i) \leq x\}} - F(x)) \xrightarrow{D} \frac{J_m(x)}{m!} Z_{m,H}(t), \]

where the convergence takes place in \( D([0,1] \times [-\infty, \infty]) \), equipped with the uniform topology. \( (Z_{m,H}(t))_{t \in [0,1]} \) is an \( m \)-th order Hermite process.

The limit process is a product of a deterministic function \( J_m(x) \) and the Hermite process \( Z_{m,H}(t) \). Thus it is sometimes called semi-degenerate. The main part of the proof is a so-called reduction principle. Define

\[ S_n(x, t) = d_{n,m}^{-1} \sum_{i=1}^{\lfloor nt \rfloor} (1_{\{G(X_i) \leq x\}} - F(x)) - \frac{J_m(x)}{m!} H_m(X_i)). \]
Then it is shown in Dehling and Taqqu (1989) that for all $\epsilon > 0$

$$P \left( \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} \left| S_n(x, t) \right| \right) \to 0,$$

as $n \to \infty$. That is, the (sequential) empirical process gets approximated only by the first term of its Hermite expansion. This technique has become the main idea in the asymptotic analysis of empirical process for LRD. Ho and Hsing (1996) and Giraitis et al. (1996a) independently considered the empirical process of a linear processes $X_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}$. They showed that it can be approximated by $f(x) \sum_{i=1}^{n} X_i$, with $f(x)$ being the probability density of $X_i$. Later, this was extended to two-sided linear processes by Giraitis and Surgailis (2002).

The strongest result in this context was given in Wu (2003). Consider a one-sided linear process

$$X_i = \sum_{j=0}^{\infty} a_j \epsilon_{i-j}, \quad \text{with} \quad a_j \sim L(j)j^{-d},$$

for a slowly varying function $L$ and $d \in (1/2, 1)$. Moreover, let $f_k(x)$ and $F_k(x)$ (with $f_\infty(x) = f(x)$ and $F_\infty(x) = F(x)$) be probability density and distribution function of $X_{i,k} = \sum_{j=0}^{k} a_j \epsilon_{i-j}$, respectively. For $m^* \in \mathbb{N}$, define

$$S_n(t; m^*) = \sum_{i=1}^{n} \left( 1_{\{X_i \leq x\}} - \sum_{q=0}^{m^*} (-1)^q F^{(q)}(x) U_{q,i} \right), \quad U_{q,i} = \sum_{0 \leq j_1 < \ldots < j_q} \prod_{s=1}^{q} a_{j_s} \epsilon_{n-j_s}.$$

Here, the empirical process gets approximated by the first terms of what is called Volterra expansion. The next result is a slightly simplified version of Theorem 2 in Wu (2003).

**Theorem 3** (Wu (2003)). Assume $E|\epsilon_1|^{4+\gamma} < \infty$ for $\gamma \geq 0$, that $f_\kappa \in C^{p+1}$ for $\kappa > 0$ and $p \geq 0$, and that

$$\sum_{q=0}^{m^*+1} \int_{\mathbb{R}} |f^{(q)}(x)|^2 (1 + |x|)^\gamma dx < \infty.$$

Then, as $n \to \infty$,

$$E \left[ \sup_{x \in \mathbb{R}} (1 + |x|)^\gamma \left| S_n(x; m^*) \right|^2 \right] = O \left( n \log \log n + \theta_{n,m^*} \right),$$
with
\[
\theta_{n,m^*} = \begin{cases} 
O(n) & \text{if } (m^* + 1)(2d - 1) > 1, \\
O(n^{2-(m^*+1)(2d-1)}L^{2(m^*+1)}) & \text{if } (m^* + 1)(2d - 1) < 1, \\
O(n \left[ \sum_{i=1}^{n} L^{m^*+1}(i/i) \right]^2) & \text{if } (m^* + 1)(2d - 1) = 1.
\end{cases}
\]

Next consider transformations of linear processes, that is \((G(X_i))_{i \geq 1}\). Here too, the empirical process might be approximated by a linear combination of \(\sum_{i=1}^{n} U_{q,i} \), \(q = 1, \ldots, m^*\). Define
\[
S_n(t; G; m) = \sum_{i=1}^{n} \left( 1_{\{G(X_i) \leq x\}} - \sum_{q=0}^{m^*} (-1)^q K_q(x) U_{q,i} \right),
\]
with
\[
K_q(x) = k_x^{(q)}(0) \quad \text{and} \quad k_x(y) = P(G(X + x) \leq y).
\]
The smallest integer \(m \geq 1\) with \(K_m(x) \neq 0\) is called power rank. Note that the integer \(m^*\) in (2.5) might be larger than \(m\). In this case, the empirical process gets approximated by additional terms and this might improve the rate of convergence.

The function \(K_q(x)\) originates in a Volterra expansion (Ho and Hsing (1996), Wu (2003)) and plays a similar part as the Hermite coefficient function \(J_q(x)\). For a Gaussian linear process one gets \(K_q(x) = J_q(x)\). To see this, note
\[
k_x(y) = P(X + y \in G^{-1}((-\infty, x])) = \int_{G^{-1}((-\infty, x])} \phi(z - y)dz.
\]
Then, using the identity \(\phi^{(q)}(x) = (-1)^q H_q(x) \phi(x)\), one obtains
\[
k_x^{(q)}(y) = \int_{G^{-1}((-\infty, x])} H_q(z - y) \phi(z - y)dz \quad \text{and} \quad k_x^{(q)}(0) = E1_{\{G(X) \leq x\}} H_q(X).
\]
The last result in this section treats the scenario when a piecewise monotone transformation is applied to the linear process. One the one hand, the assumption of piecewise monotonicity simplifies the proof. On the other hand, it is satisfied by most transformations of interest.

**Corollary 2.2.1.** Assume that the conditions of Theorem 3 hold and let \(G\) be a measurable, piecewise monotone function. Then
\[
E\left[ \sup_{x \in \mathbb{R}} |S_n(x; G; m)|^2 \right] = O\left( n \log \log n + \theta_{n,m^*} \right),
\]
where \(\theta_{n,m^*}\) is of the same order as in Theorem 3.
Remark 2.2.2. (i) In contrast to Theorem 3, no weight function $(1 + |x|^{\gamma})$ is included in the result. For a weighted reduction principle under subordination see Buchsteiner (2015).

(ii) Interestingly, the empirical process of the original linear process has always power rank 1, while the process considered here might have arbitrary power rank.

2.3 Short-range dependence and probability theory in Hilbert spaces

2.3.1 Mixing conditions in a Hilbert space $H$

When the assumption of independence in classical result (central limit theorem, law of large numbers) is dropped, the dependence of random variables is often modeled via so-called mixing properties. Over the last five decades various notions of mixing have been investigated, ranging even to border of long memory. For a comprehensive overview we refer to the series of Bradley (2007). Here we introduce two of the classical mixing conditions. Denote by $F_{m}^\infty = \sigma(\xi_{-l}, \ldots, \xi_{m})$ the $\sigma$-field generated by $\xi_{-l}, \ldots, \xi_{m}$.

Definition 2.3.1. For a stationary sequence $(\xi_i)_{i \in \mathbb{Z}}$, taking values in a separable Banachspace, define

$$
\alpha(k) = \sup_{m \in \mathbb{Z}} \sup_{A \in F_{-m}^\infty, B \in F_{m+k}^\infty} |P(A \cap B) - P(A)P(B)|.
$$

Then $(\xi_i)_{i \in \mathbb{Z}}$ is called strongly mixing, if $\alpha(k) \to 0$ as $k \to \infty$.

Definition 2.3.2. For a stationary sequence $(\xi_i)_{i \in \mathbb{Z}}$, taking values in a separable Banachspace, define

$$
\beta(k) = \left| E \sup_{A \in F_k^\infty} [P(A|F_{-\infty}^0) - P(A)] \right|.
$$

Then $(\xi_i)_{i \in \mathbb{Z}}$ is called absolutely regular, if $\beta(k) \to 0$ as $k \to \infty$.

The notion of strong mixing is more general than absolute regularity, which will be assumed in Chapter 5. Anyway, the random variables we consider are functionals of absolutely regular processes. The following concept is due to Ibragimov (1968).

Definition 2.3.3. Let $(\xi_i)_{i \in \mathbb{Z}}$ be a stationary sequence of random variables, taking values in an arbitrary separable and measurable space. A stationary sequence $(X_n)_{n \in \mathbb{Z}}$ of $\mathbb{R}$-valued random variables is called $L_p$-near epoch dependent (NED($p$)) on $(\xi_i)_{i \in \mathbb{Z}}$, if there is a sequence $(d_k)_{k \in \mathbb{N}}$ with $d_k \to 0$ as $k \to \infty$ and

$$
E \left[ |X_0 - E[X_0|F_{-k}^k]|^p \right] \leq d_k.
$$
Hansen (1991) showed that GARCH-processes are near epoch dependent. In contrast, the strong mixing property fails to cover even some basic types of time series. For example, Andrews (1984) considered an AR-process that is not strongly mixing.

In what follows these notions are extended to Hilbert spaces. The motivation for studying dependent, Hilbert space-valued random variables is twofold. First, one might think of observations, actually taking values in such a space. Consider for instance daily temperature curves. It seems reasonable to model these as functional data and to assume some type of dependence between the curves.

Secondly, one might consider Hilbert spaces for technical reasons. For example, an empirical process can be treated as random element of a Hilbert space $H$. Then the dependence of the (real-valued) observations will carry over the random variables defined in $H$.

Now let $H$ be a separable (i.e. there exists a dense and countable subset) Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$.

The good news is that Definition 2.3.1 and 2.3.2 carry over directly to this space. Actually, one might consider random variables taking values in an arbitrary, separable, measurable space.

As for the NED condition, we have to define conditional expectation in Hilbert spaces. This can be done analogously to the definition for real-valued random variables, replacing the Lebesgue Integral by the Bochner Integral. For a treatment of the latter see the appendix of Chen and White (1996). Then $E[X|F]$ becomes an $H$-valued random variable and we say that $(X_n)_{n \in \mathbb{Z}}$ is $L^p$-near epoch dependent on $(\xi_i)_{i \in \mathbb{Z}}$ if

$$E \left[ \| X_0 - E[X_0 | \mathcal{F}_{-k}] \|^{p} \right] \leq d_k,$$

for some $d_k \to 0$.

### 2.3.2 Brownian motion in $H$

The mixing properties often correspond to the notion of asymptotic independence. Thus it is not surprising that limiting distributions of important statistics show a similar pattern as under independence. For instance, the partial sum process still converges to Brownian motion.

In order to define a Brownian motion in a separable Hilbert space $H$, we have to define Gaussianity in $H$ first. An $H$-valued random variable $N$ is said to be Gaussian, if for all $h \in H \setminus \{0\}$ the $\mathbb{R}$-valued variable $\langle N, h \rangle$ has a normal distribution.

We say that an $H$-valued random variable $X$ has mean $\mu \in H$, if $E\langle X, h \rangle = \langle \mu, h \rangle$ for all $h \in H$. We denote it by $EX$. Moreover, define the covariance operator $S: H \to H$ of $X$ (if it
exists) by
\[
\langle Sh_1, h_2 \rangle = E \left[ \langle X - EX, h_1 \rangle \langle X - EX, h_2 \rangle \right] \quad h_1, h_2 \in H.
\]
The distribution of a Gaussian, $H$-valued random variable is uniquely determined by its mean and its covariance operator. Next consider the space $C_H[0,1]$, defined as the set of all continuous $H$-valued functions on $[0,1]$. Under the supremum norm $\|f\|_\infty = \sup_{t \in [0,1]} \|f(t)\|_H$ it becomes a separable Banach space.

**Definition 2.3.4.** A $C_H[0,1]$-valued random element $W$ will be called Brownian motion in $H$ if

(i) $W(0) = 0$ almost surely;

(ii) The increments on disjoint intervals are independent;

(iii) For all $0 \leq s < t \leq 1$ the increment $W(t) - W(s)$ is a zero-mean, Gaussian random variable with covariance operator $(t - s)S$. $S : H \to H$ does not depend on $s$ or $t$.

Note that the distribution of a Brownian motion $W$ is uniquely determined by the covariance operator $S$ of $W(1)$.

**2.3.3 Convergence in $D_H[0,1]$**

In chapter 5 we will investigate the process $\sum_{i=1}^{\lfloor nt \rfloor} X_i$, that is the non-interpolated partial sum process of $H$-valued variables. Therefore, instead of $C_H[0,1]$ we have to study the space $D_H[0,1]$, the set of all càdlàg functions mapping from $[0,1]$ to $H$. An $H$-valued function on $[0,1]$ is said to be càdlàg, if it is right-continuous and the left limit exists for all $x \in [0,1]$. Analogously to the real valued case we define the Skorohod metric

\[
d(f,g) = \inf_{\lambda \in \Lambda} \left\{ \sup_{t \in [0,1]} \|f(t) - g \circ \lambda(t)\| + \|id - \lambda\|_\infty \right\} \quad f, g \in D_H[0,1],
\]

where $\Lambda$ is the class of strictly increasing, continuous mappings of $[0,1]$ onto itself and $\|\cdot\|$ is the Hilbert space norm. Moreover $id : [0,1] \to [0,1]$ is the identity function and $\circ$ denotes composition of functions. Most topological properties on $D[0,1] = D_R[0,1]$ carry over to the space $D_H[0,1]$ (for more details on $D_R[0,1]$ see the book of Billingsley (1968)). Equipped with the Skorohod metric, $D_H[0,1]$ becomes a separable Banach space.

Now let $(W_n)_{n \geq 1}$ be a sequence of $D_H[0,1]$-valued random functions. We will actually use two different techniques to verify weak convergence of $(W_n)_{n \geq 1}$, but the idea behind this techniques is basically the same:
• First, weak convergence of a finite dimensional approximation is proved;

• Secondly, we show that the entire process is approximated well by its finite distributions.

The difference of the two methods is now as follows: While the first technique approximates functions on \([0, 1]\) by \((f(t_1), \ldots, f(t_k))\), for a finite number of grid points, the second technique projects the function values (being elements of \(H\)) on a lower dimensional space.

Following the first method we have to prove weak convergence of

\[ W_{n,t_1,\ldots,t_k} = (W_n(t_1), \ldots, W_n(t_k)) \]

in \(H^k = H \times \cdots \times H\). Afterwards tightness might be established using the following criterion.

**Lemma 2.3.5.** Let \(\{W_n\}_{n \geq 1}\) be a sequence of \(D_H[0,1]\)-valued random functions with \(W_n(0) = 0\). Then \(\{W_n\}_{n \geq 1}\) is tight in \(D_H[0,1]\) if the following condition is satisfied:

\[
\lim_{\delta \to 0} \limsup_{n \to \infty} \frac{1}{\delta} P \left( \sup_{s \leq t \leq s + \delta} \|W_n(t)\| > \epsilon \right) = 0,
\]

for each positive \(\epsilon\) and each \(s \in [0,1]\).

Furthermore the weak limit of any convergent subsequence of \(\{W_n\}\) is in \(C_H[0,1]\), almost surely.

For real valued random variables this is Theorem 8.3 of Billingsley (1968), which carries over to \(D[0,1]\). It can be proved along the same lines, if \(H\)-space valued functions are treated. We will use this technique in the proof of Theorem 17.

For the second approach let \((e_i)_{i \geq 1}\) be an orthonormal basis of \(H\) and \(H_k\) be the closed linear span of \((e_i)_{1 \leq i \leq k}\). Then define \(P_k: H \to H_k\) as the projection operator and show convergence of the sequence \((P_kW_n)_{n \geq 1}\), given through

\[
P_kW_n(t) = \sum_{j=1}^{k} \langle W_n(t), e_j \rangle e_j. \tag{2.1}
\]

Finally, the next lemma formalizes how \((W_n)_{n \geq 1}\) should be approximated by \((P_kW_n)_{n \geq 1}\).

**Lemma 2.3.6.** Let \(\{W_n\}_{n \geq 1}\) be a sequence of \(D_H[0,1]\)-valued random functions. Let \(W^k\) be a Brownian motion in \(H_k\) with \(S^k\) being the covariance operator of \(W^k(1)\). Suppose the following conditions are satisfied:

(i) For each \(k \geq 1\), \(P_kW_n \xrightarrow{D} W^k\) in \(D_{H_k}[0,1]\) (as \(n \to \infty\)),

(ii) \(W^k \xrightarrow{D} W\) in \(D_H[0,1]\) (as \(k \to \infty\)),

(iii) \(\limsup_{n \to \infty} E \left( \sup_{t \in [0,1]} \|W_n(t) - P_kW_n(t)\|^2 \right) \to 0\) as \(k \to \infty\).
Then $W_n \xrightarrow{D} W$ in $D_H[0,1]$, where $W$ is a Brownian motion in $H$ with covariance operator $S$.

The result can be found in Chen and White (1998) (Lemma 4.1) with the slight modification that they use second moments in condition (iii). This approach ((2.1) and Lemma 2.3.6) will be used in the proof of Theorem 16.

### 2.3.4 Moment inequalities

A key tool in proofs of many limit theorems for mixing random variables are moment inequalities for partial sums. Here we give such an inequality for $H$-valued NED-processes.

**Lemma 2.3.7.** Let $(X_n)_{n \geq 1}$ be $H$-valued, stationary and $L_1$-near epoch dependent on an absolutely regular process with mixing coefficients $(\beta(m))_{m \geq 1}$ and approximation constants $(d_m)_{m \geq 1}$. If $EX_1 = 0$ and

(i) $E\|X_1\|^{4+\delta} < \infty$,

(ii) $\sum_{m=1}^{\infty} m^2 (d_m)^{(\delta+3)} < \infty$,

(iii) $\sum_{m=1}^{\infty} m^2 (\beta(m))^{(\delta+4)} < \infty$,

holds for some $\delta > 0$, then

$$E\|X_1 + X_2 + \cdots + X_n\|^4 \leq Cn^{2} \left( E\|X_1\|^{4+\delta} \right)^{\frac{1}{1+\delta}}.$$ 

This result follows from the proof of Lemma 2.24 of Borovkova et al. (2001), which is also valid for Hilbert spaces.

The next lemma is a special case of Theorem 1 of Móricz (1976). Here too, the proof carries over directly to Hilbert spaces.

**Lemma 2.3.8.** Let $(X_n)_{n \geq 1}$ be a stationary sequence of $H$-valued random variables such that $EX_1 = 0$, $E\|X_1\|^4 < \infty$ and for some $C > 0$

$$E\|X_1 + X_2 + \cdots + X_n\|^4 \leq Cn^{2}.$$ 

Then

$$E\max_{k \leq n}\|X_1 + X_2 + \cdots + X_k\|^4 \leq Cn^{2}.$$ 

### 2.4 Bootstrap

#### 2.4.1 Two block bootstrap methods

When the distribution of a test statistic $T_n(X_1, \ldots, X_n)$ is simulated by reusing the data $X_1, \ldots X_n$, one speaks of resampling. The resampling method that has drawn the most at-
tention over the last four decades is the bootstrap, which was developed by Efron (1979) for independent random variables. Assume that the observations $X_1, \ldots, X_n$ are i.i.d. and come from an unknown distribution $F(x)$. Define their empirical distribution function by $F_n(x)$. Efron (1979) then considered

$$X_i^* \sim F_n \quad i = 1, \ldots, n,$$

that is, random variables that are drawn independently from $F_n$. Eventually, the distribution of $T_n$ is approximated by the distribution of $T_n^* = T_n(X_1^*, \ldots, X_n^*)$. The latter might be calculated explicitly, however not without a huge computational burden. Thus it is often obtained by a Monte Carlo simulation, using the fact that arbitrary many samples $(X_{i,1}^*)_{i=1}^n, \ldots, (X_{i,J}^*)_{i=1}^n$ might be drawn.

Later on, an asymptotic justification of Efron’s method was given in Singh (1981) and Bickel and Freedman (1981). In detail, it is shown that the distribution of $T_n$ and the conditional distribution $T_n^*$ converge to the same weak limit. By conditional distribution we mean the distribution with respect to $P^* = P(\cdot|\sigma(X_1, \ldots, X_n))$. Ever since, validity of the bootstrap is usually established by this argument.

Now consider dependent observations. For example, let $(X_i)_{i \geq 1}$ be a stationary, mixing process with unknown distribution and suppose we are interested in the unknown mean $\mu \in \mathbb{R}$. Then, under some mild technical assumptions, $\sqrt{n}(\bar{X}_n - \mu) \rightarrow N(0, \sigma^2_{\infty})$, with

$$\sigma^2_{\infty} = \text{Var}(X_1) + 2 \sum_{k=2}^{\infty} \text{Cov}(X_1, X_k).$$

Next, let $(X_i^*)_{i \geq 1}$ be a bootstrap sample, independently drawn from $F_n$. Let further $E^*Y = E[Y|\sigma(X_1, \ldots, X_n)]$ and $\text{Var}^*Y = E^*Y^2 - (E^*Y)^2$. Then

$$\text{Var}^* \left( n^{-1/2} \sum_{i=1}^{n} (X_i^* - \mu) \right) = \text{Var}^*(X_1^*) = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 \rightarrow \text{Var}(X_1) \quad \text{a.s.,}$$

by the ergodic theorem. Thus Efron’s bootstrap fails here. Several modifications were suggested to overcome this issue, among them the block bootstrap. We will consider two different types of block bootstrap methods:

**Moving blocks bootstrap (MBB):** This procedure was proposed in the groundbreaking article of Künsch (1989). For a given sample $X_1, \ldots, X_n$ and a block length $l = l(n)$ define blocks of random variables by

$$I_j^M = (X_j, \ldots, X_{j+l-1}) \quad j = 1, \ldots, n - l + 1.$$
Then choose randomly with replacement \( k = k(n) \) blocks \( I_1^*, \ldots, I_k^* \) and define the bootstrap sample by
\[
X_1^*, \ldots, X_j^*, \ldots, X_{(k-1)l+1}^*, \ldots, X_{kl}^* = I_i^*
\]
Now this sample satisfies
\[
P \left( (X_{(j-1)l+1}^*, \ldots, X_{jl}^*) = I_i^M \right) = \frac{1}{n-l+1} \quad \text{for } j = 1, \ldots, k, \ i = 1, \ldots, n-l+1,
\]
a fact that will be used frequently in our proofs.

**Nonoverlapping block bootstrap (NBB):** A method that goes back to Carlstein (1986) uses blocks of the form
\[
I_j^N = (X_{(j-1)l+1}, \ldots, X_{jl}) \quad j = 1, \ldots, [n/l].
\]
Subsequently, bootstrap samples are generated analogously to the MBB. One obtains
\[
P \left( (X_{(j-1)l+1}^*, \ldots, X_{jl}^*) = I_i^N \right) = \frac{1}{[n/l]} \quad \text{for } j = 1, \ldots, k, \ i = 1, \ldots, [n/l].
\]
Finally, we briefly describe how the empirical process is bootstrapped. Let \( X_1, \ldots, X_n \) be a set of real-valued data with block bootstrap sample \( X_1^*, \ldots, X_{kl}^* \). Then the bootstrapped empirical process is is given by the empirical process of the bootstrap sample
\[
W_n^* = \sum_{i=1}^{kl} (1_{X_i^* \leq x} - E^*1_{X_i^* \leq x}).
\]
It is one of the central objects of this thesis.

## 2.4.2 Three types of convergence

We noted in the previous section that validity of the bootstrap is established via its asymptotic behavior. However, the usual weak convergence does to suffice. Here we describe three types of “bootstrap convergence” used in the literature. We will consider all three of them in this thesis.

**Almost sure weak convergence:** As above, define \( E^*Y = E[Y | \sigma(X_1, \ldots, X_n)] \), the conditional expectation given \( X_1, \ldots, X_n \). Analogously define \( Var^* \) and \( P^* \). Now let \( (Y_n^*)_{n \geq 1} \) be a sequence of random variables. We say that \( Y_n^* \) converges almost surely in distribution to \( Y \), if there is
a set \( A \subset \Omega \) with \( P(A) = 1 \) such that for all \( \omega \in A \)

\[
E^* f(Y_n^*) \to Ef(Y) \quad \text{for all bounded and continuous functions } f: B \to \mathbb{R},
\]

(2.1)
as \( n \to \infty \). Write

\[
Y_n^* \xrightarrow{D} Y \quad \text{a.s.}
\]

When proving this type of convergence one does not actually show (2.1). Assume that one has certain criteria for convergence at hand (such as finite-dimensional convergence, tightness etc.). Then it suffice to verify that these hold almost surely.

**Weak convergence in probability:** Let \( d(\cdot, \cdot) \) be a any metric on the space of distributions of \( B \)-valued random variables, metricizing weak convergence. Then \( Y_n^* \) converges weakly in probability towards \( Y \), if

\[
d(\mathcal{L}(Y_n^*|\sigma(X_1, \ldots, X_n)), \mathcal{L}(Y)) \xrightarrow{P} 0,
\]

(2.2)
as \( n \to \infty \). Write

\[
Y_n^* \xrightarrow{D} Y \quad \text{in probability.}
\]

The object \( d(\mathcal{L}(Y_n^*|\sigma(X_1, \ldots, X_n)), \mathcal{L}(Y)) \) might be not measurable. In this case (2.2) has to be formulated in *outer probability*. In detail, (2.2) has to be replaced by

\[
d(\mathcal{L}(Y_n^*|\sigma(X_1, \ldots, X_n)), \mathcal{L}(Y))^o \xrightarrow{P} 0,
\]

where \( f^o = \text{ess inf } f = \inf \{g \text{ measurable} | g \geq f\} \) is the essential infimum of \( f \).

For real-valued random variables the metric used most frequently is the *Kolmogorov distance*. Then (2.2) is equivalent to

\[
\sup_{x \in \mathbb{R}} |P^*(Y_n^* \leq x) - P(Y \leq x)| \xrightarrow{P} 0.
\]

As most test statistics are real-valued, this notion can be found in many results in the literature. Then again, those test statistics might be written as functional of more complex random variables - for example as functionals of the empirical process. A metric that suites this situation quite well is based on the bounded-Lipschitz norm. Define

\[
d_{BL}(\mu, \nu) = \sup_{f \in L(F)} \left| \int f \, d\mu - \int f \, d\nu \right|,
\]

(2.3)
where

\[ L(F) = \left\{ K : F \rightarrow \mathbb{R} : \sup_f |K(f)| \leq 1, |K(f_1) - K(f_2)| \leq \|f_1 - f_2\| \text{ for all } f_1, f_2 \right\} \]

is the set of functions whose bounded-Lipschitz norm is smaller than one. For more details on this metric see Araujo and Giné (1980). For an application to bootstrapping empirical processes see Giné and Zinn (1990) and Radulovic (1996a). We will use this notion in Chapter 4, then under long memory.

**Unconditional bootstrap convergence:** Let \( Y_n^* \) be a bootstrap version of \( Y_n \), both taking values in a space \( B \). Instead of “conditional convergence” one might prove \( Y_n^* \xrightarrow{D} Y \) in the common sense. However, if we are interested in proving asymptotic validity of confidence intervals (created from the bootstrap), this will not suffice. Therefore, consider \( J \) bootstrap replicas \( Y_{n,1}^*, \ldots, Y_{n,J}^* \). Then one might prove

\[ (Y_n, Y_{n,1}^*, \ldots, Y_{n,J}^*) \xrightarrow{D} (Y_0, Y_1, \ldots, Y_J), \]

where \( Y_0, Y_1, \ldots, Y_J \) are i.i.d. and weak convergence takes place in \( B^{J+1} \). As we will see below, this suffice to justify the bootstrap asymptotically.

### 2.4.3 Bootstrap Consistency

A central motivation for the bootstrap is the derivation of confidence intervals or, quite related, critical values for statistical testing. Let \( T_n \) be a statistic and \( T_{n}^* \) a bootstrapped version. Define

\[ K_n^* = P^*(T_n^* \leq x) \text{ and } K_n^{* - 1}(p) = \inf \{ x \in \mathbb{R} | K_n^*(x) \geq p \}. \]

We say that the bootstrap is **asymptotically consistent**, if

\[ P \left( T_n > K_n^{* - 1}(1 - \alpha) \right) \rightarrow \alpha, \]

as \( n \rightarrow \infty \). Lemma 23.3 in van der Vaart (1998) states that this actually holds, if \( T_n^* \) converges weakly in probability to the same continuous distribution as \( T_n \). The computation of \( K_n^{* - 1}(1 - \alpha) \) is possible, but computationally demanding. Therefore, it is common to approximate the quantile by a Monte Carlo simulation. Let \( T_{n,1}^*, \ldots, T_{n,J}^* \) be independent copies of \( T_n^* \) with empirical distribution function

\[ K_{n,j}^*(x) = \frac{1}{J} \sum_{j=1}^{J} 1\{ T_{n,j}^* \leq x \}. \]
and $K_{n,J}^{-1}(p) = \inf\{x \in \mathbb{R} \mid K_{n,J}^*(x) \geq p\}$. Then a critical value is defined by $K_{n,J}^{-1}(1 - \alpha)$. It might be justified via the next two results.

**Proposition 2.4.1.** Let $T_n$ be a real-valued statistic and $T_n^*$ its block bootstrap (non-overlapping or moving) counterpart. Assume further that, as $n \to \infty$,

$$
\sup_{x \in \mathbb{R}} |P(T_n \leq x) - K(x)| \to 0,
$$

and

$$
\sup_{x \in \mathbb{R}} |P^*(T_n^* \leq x) - K(x)| \to 0,
$$

for a continuous distribution function $K(x)$. Then, for any $\alpha \in (0, 1)$,

$$
\lim_{n \to \infty} \lim_{J \to \infty} P\left(T_n \geq K_{n,J}^{-1}(1 - \alpha)\right) = \alpha.
$$

**Proof.** Choose $\alpha \notin \mathbb{Q}$. Let us first show that, for any fixed sample size $n \in \mathbb{N}$, we have

$$
K_{n,J}^{-1}(1 - \alpha) \to K_n^{-1}(1 - \alpha) \quad \text{a.s.}(P^*),
$$

as $J \to \infty$. Note that the random variables $T_{n,1}^*, \ldots, T_{n,J}^*$ are i.i.d. with respect to the probability measure $P^*$, hence by the Glivenko-Cantelli theorem

$$
K_{n,J}^*(x) \to K_n^*(x) \quad \text{a.s.}(P^*).
$$

By Lemma 21.2 in van der Vaart (1998), this implies $K_{n,J}^{-1}(p) \to K_n^{-1}(p)$ for all points of continuity $p$ of $K_n^{-1}$. However, due to the block bootstrap structure, the distribution function of the bootstrapped statistic, $K_n^*(x)$, can be written as

$$
K_n^*(x) = P^*(T_n^* \leq x) = \frac{1}{N(n,l)} \sum_{i=1}^{N(n,l)} \{t_{i,n}(X_1,\ldots,X_n) \leq x\},
$$

for a certain $N(n,l) \in \mathbb{N}$ and functions $t_{i,n}: \mathbb{R}^n \to \mathbb{R}$ (whose precise structure is of no further importance). Thus $(1 - \alpha)$ is a point of continuity of $K_n^{-1}(x)$ for any $\alpha \notin \mathbb{Q}$ and this yields (2.7). Next, one has

$$
T_n - K_{n,J}^{-1}(1 - \alpha) \to T_n - K_n^{-1}(1 - \alpha) \quad \text{a.s.}(P^*),
$$

as $J \to \infty$. Note that almost sure convergence with respect to $P^*$ implies almost sure convergence with respect to $P$ and consequently

$$
\lim_{J \to \infty} P\left(T_n > K_{n,J}^{-1}(1 - \alpha)\right) \to P\left(T_n > K_n^{-1}(1 - \alpha)\right).
$$

By Lemma 23.3 in van der Vaart (1998) the right-hand side of (2.8) converges to $\alpha$, which
finishes the proof for \( \alpha \notin \mathbb{Q} \).

Now arguing as in the proof of Lemma 23.3 of van der Vaart (1998) this can be extended to all \( \alpha \in (0, 1) \). To see this, note that both sides of (2.6) are monotone functions of \( \alpha \) and that the right-hand side is continuous.

\[ \Box \]

**Proposition 2.4.2** (Bücher, Kojadinovic). For any integer \( J \geq 1 \) assume that there exists random variables \( T_n, T_{n,1}^*, \ldots, T_{n,J}^* \) such that

\[
(T_n, T_{n,1}^*, \ldots, T_{n,J}^*) \overset{D}{\to} (T_0, T_1, \ldots, T_J),
\]

where \( T \), the weak limit of \( T_n \) is a continuous random variable, and \( T_1, \ldots, T_J \) are independent copies of \( T \). Then, for any \( \alpha \in (0, 1) \),

\[
\lim_{J \to \infty} \lim_{n \to \infty} P \left( T_n \geq K_{n,j}^{-1}(1 - \alpha) \right) = \alpha.
\]

This result can be found in the supplement to Bücher and Kojadinovic (2016). It is interesting to note that compared to Proposition 2.4.1 the order of the limits has been changed.

**2.4.4 Subsampling**

Another resampling approach is due to Politis and Romano (1994a), the so-called subsampling. This method will play a central part in Chapter 6. Consider a real-valued statistic given by \( T_n = \tau_n(t_n(X_1, \ldots, X_n) - \theta) \), where \( \tau_n \) is a normalizing sequence and \( \theta \in \mathbb{R} \) some parameter of interest.

In order to approximate the distribution of \( T_N \), here too, blocks of neighboring observations are considered. In fact, one has \( I_1, I_2, \ldots, I_{N_n} \) with

\[
I_i = (X_{k(i)}, X_{k(i)+1}, \ldots, X_{k(i)+l-1}), \quad i = 1, \ldots N_n,
\]

for a block length \( l \in \mathbb{N} \). Depending on \( k(i) \), the blocks might be arranged circular, nonoverlapping or overlapping. Most common is the choice of moving blocks as in the bootstrap of Künsch (1987). In contrast to the bootstrap one does not draw random blocks. Instead, small scale versions of the statistic are computed for each block, in detail

\[
T_{l,i} = \tau_l(t_l(X_{k(i)}, X_{k(i)+1}, \ldots, X_{k(i)+l-1}) - t_n), \quad i = 1, \ldots N_n.
\]

Subsequently, the **subsampling distribution** is defined as the empirical distribution function of these statistics \( L_n(x) = \frac{1}{N_n} \sum_{i=1}^{N_n} 1_{\{T_{l,i} \leq x\}} \). Validity of the subsampling mechanism is usually
established via the following convergence:

$$\sup_{x \in \mathbb{R}} |L_n(x) - P(T_n \leq x)| \xrightarrow{P} 0.$$  \hspace{1cm} (2.9)

In fact, uniform convergence holds usually only if $T_n$ has a continuous limiting distribution. The motivation behind a subsampling estimator (just as it is for the bootstrap) is the derivation of critical values or confidence regions, respectively. Based on the convergence in (2.9), one might use empirical quantiles of the subsampling distribution. One obtains, as $n \to \infty$,

$$P \left( T_n > L_n^{-1}(1 - \alpha) \right) \to \alpha,$$  \hspace{1cm} (2.10)

where $\alpha \in (0,1)$ and $L_n^{-1}$ is the generalized inverse of $L_n(x)$. The next theorem is one of the classical results about the general validity of subsampling, see Politis et al. (1999). One assumes moving blocks, hence $k(i) = i$ and $N_n = n - l + 1$.

**Theorem 4** (Politis, Romano and Wolf (1999)). Let $(X_i)_{i \geq 1}$ be a strongly mixing sequence, such that $T_n$ converges weakly to some continuous limit law. Let further $\tau_l/\tau_n \to 0$, $l/n \to 0$ and $l \to \infty$ as $n \to \infty$. Then (2.9) and (2.10) hold.

Similar results exist for other classes of random variables, see for instance Betken and Wendler (2017+) for long memory, subordinated Gaussian series.
3 Distributional change under long-range dependence

Over the last two decades various authors have studied the change-point problem under long-range dependence and classical methods are often found to yield different results than under short-range dependence. The CUSUM test is studied in Csörgő and Horvath (1997) and compared to the Wilcoxon change-point test in Dehling et al. (2012). Ling (2007) investigates a Darling-Erdős-type result for a parametric change-point test, and estimators for the time of change are considered in Horvath and Kokoszka (1997) and Hariz et al. (2009). Moreover, the special features of long memory motivated new procedures. Beran and Terrin (1996) and Horvath and Shao (1999) are testing for a change in the linear dependence structure of the time series and Berkes et al. (2006) and Baek and Pipiras (2011) construct tests in order to discriminate between stationary long memory observations and short memory sequences with a structural change. For a general overview of change-point methods under LRD see Kokoszka and Leipus (2001) and the associated chapter in Beran et al. (2013).

The change-point problem which is the central aspect of this chapter, and of the whole thesis too, is the change of the marginal distributions of a time series \( \{Y_i\}_{i \geq 1} \). When testing for such a change-point, one often considers the empirical distribution function of the first \( k \) observations and that of the remaining observations. Taking a distance between the empirical distributions and the maximum over all \( k < n \) yields a natural statistic. Common distances are the supremum norm, which gives the Kolmogorov-Smirnov statistic or an \( L^2 \)-distance, which gives the Cramér-von Mises statistic. Both are widely used for goodness-of-fit tests, two-sample problems and change-point detection. In the LRD setting, only the Kolmogorov-Smirnov test and solely for linear processes has been investigated, see Giraitis et al. (1996b).

The first goal of this chapter is the derivation of the limit distributions of the aforementioned statistics under local alternatives. To this end we investigate subordinated Gaussian processes of the form \( G(X_1), \ldots, G(X_{k^*}), G_n(X_{k^*+1}), \ldots, G_n(X_n) \). Here \( G_n \) is a sequence of functions such that the distribution of \( G_n(X_1) \) converges to the distribution of \( G(X_1) \) in some suitable way.

Kolmogorov-Smirnov and Cramér-von Mises statistic are functionals of the sequential empirical process. Thus their asymptotic distributions rely on the limit of this process, which - under long-range dependence - has a very special structure (see Section 2.2). We will show that this unique asymptotic behavior still applies for the local alternatives described above.
The mathematically most challenging case is the situation when the Hermite rank \( m \) changes. The structure of the limiting process \( Z(t) \), e.g. the marginal distribution and the covariance structure, mainly depends on \( m \). However, a special feature of distributional changes in subordinated Gaussian processes is the fact that the Hermite rank may change, too. Hence the question arises which Hermite process will determine the limit distribution.

Our results differ in various ways from those obtained in Giraitis et al. (1996b), where changes in the coefficients of an LRD linear process were investigated. While the empirical process of LDR moving average sequences converges to fractional Brownian motion, we may encounter higher order Hermite processes. The possible change in the Hermite rank is therefore a novel feature in our investigation.

A second goal is the comparison of different change-point tests, which will be done via the notion of asymptotic relative efficiency (ARE) as introduced by Pitman (1948). To this end we will also consider the CUSUM test and the Wilcoxon test under quite general local alternatives and thereby extend the results of Dehling et al. (2017). However, the most outstanding result is about a simple mean-shift in Gaussian data. We will show that in this scenario the ARE of Kolmogorov-Smirnov, Cramér-von Mises, Wilcoxon and CUSUM test is 1. This is quite surprising as CUSUM and Wilcoxon test are designed to detect such change-points, whereas Kolmogorov-Smirnov and Cramér-von Mises test have power against all kinds of alternatives.

### 3.1 Limit theorems for the empirical process and test statistics

#### 3.1.1 The change-point setting

Let \((Y_i)_{i \geq 1}\) be a long-range dependent time series. The central aspect of this chapter is to detect a change in the marginal distribution of this sequence. That is we want to test the hypothesis

\[
\mathbf{H} : \quad P(Y_1 \leq x) = \cdots = P(Y_n \leq x) \quad \forall x \in \mathbb{R},
\]

against the alternative

\[
\mathbf{A} : \quad \exists x \in \mathbb{R}, \; k_0 < n : \quad P(Y_1 \leq x) = \cdots = P(Y_{k_0} \leq x) \neq P(Y_{k_0+1} \leq x) = \cdots = P(Y_n \leq x).
\]

The tests we want to investigate will be based on the difference of the empirical distribution of the data before the change \( F_{1:k_0}(x) = k_0^{-1} \sum_{i=1}^{k_0} 1_{\{Y_i \leq x\}} \) and that of the remaining observations \( F_{k_0+1:n}(x) = \sum_{i=k_0+1}^{n} 1_{\{Y_i \leq x\}} \). As the time of the (possible) change is assumed to be unknown, we have to take all \( k \in \{1, \ldots, n-1\} \) into account. Maximizing over all possible change-points
leads to the statistic
\[ \max_{1 \leq k < n} d(F_{1:k}(\cdot), F_{k_0+1:n}(\cdot)), \] (3.1)
for a suitable metric \(d(\cdot, \cdot)\). The goal is now two-fold: First, we want to deduce valid critical values for (3.1) and thereby establish a test procedure. Secondly, the performance of the test shall be investigated under the alternative. However, to do so we have to assume a certain model for the sequence \((Y_i)_{i \geq 1}\). In this chapter this will be a subordinated Gaussian process, as defined in section 2.1.2.

Therefore, let \((X_i)_{i \geq 1}\) be a stationary Gaussian process, with
\[ EX_i = 0, \quad EX_i^2 = 1 \quad \text{and} \quad \gamma(k) = EX_1X_{k+1} = k^{-D}L(k), \]
for \(0 < D < 1\) and a slowly varying function \(L\). The nonstationarity in terms of a change-point will be introduced via the transformation \(G\). In detail, consider the random variables
\[ G(X_1), \ldots, G(X_{k_0}), G^*(X_{k_0+1}), \ldots G^*(X_n). \]
As an example let \(F(x)\) and \(F^*(x)\) be two distribution functions which differ for some \(x \in \mathbb{R}\). Then define \(G(x) = F^{-1}(\Phi(x))\) and \(G^*(x) = F^{*-1}(\Phi(x))\), where \(\Phi\) is the distribution function of a standard normal random variable. Consequently one has \(G(X_i) \sim F\) and \(G^*(X_j) \sim F^*\). Thus we might model an arbitrary change from one marginal distribution to another, making subordinated processes very flexible in this context.

Our treatment of the test will be an asymptotic analysis. That is, critical values will be based on the asymptotic distribution of (3.1) and the power of the test will be computed for \(n \to \infty\). If we consider a fixed alternative, then the power of the test will converge to 1 as \(n\) goes to infinity. This is kind of a minimal requirement for any test, the procedure is then called consistent. However, when one wants to compare different test (in our case change-point test) a more elaborate analysis is needed. We therefore consider the following local alternatives:
\[ Y_{n,i} = \begin{cases} G(X_i), & \text{if } i \leq \lfloor n\tau \rfloor, \\ G_n(X_i), & \text{if } i \geq \lfloor n\tau \rfloor + 1, \end{cases} \] (3.2)
for an unknown \(\tau \in (0, 1)\). The transformations \(G\) and \((G_n)_{n \geq 1}\) are chosen in a way such that \(P(G_n(X_1) \leq x) \to P(G(X_1) \leq 1)\) as \(n \to \infty\).

Consider again (3.1), now with specific choices for the metric \(d(\cdot, \cdot)\). In detail, we treat the
supremum norm, leading to the Kolmogorov-Smirnov statistic

\[ T_n = \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} \left| \frac{|nt|}{n} \sum_{i=1}^{n} 1\{Y_{n,i} \leq x\} - \frac{|nt|}{n} \sum_{i=1}^{n} 1\{Y_{n,i} \leq x\} \right|, \quad (3.3) \]

and a (data-based) \(L^2\)-distance that yields the Cramér-von Mises statistic

\[ S_n = \sup_{t \in [0,1]} \int_{\mathbb{R}} \left| \frac{|nt|}{n} \sum_{i=1}^{n} 1\{Y_{n,i} \leq x\} - \frac{|nt|}{n} \sum_{i=1}^{n} 1\{Y_{n,i} \leq x\} \right|^2 d\hat{F}_n(x). \quad (3.4) \]

The asymptotic distributions of these statistics will be determined in three steps: First, we will treat an asymptotic expansion of the sequential empirical process of the triangular array \((G_n(X_i))_{i \leq n, n \in \mathbb{N}}\). Secondly, we will introduce weak convergence of the sequential empirical process of (3.7)

\[ \sum_{i=1}^{[nt]} \left(1\{Y_{n,i} \leq x\} - P(Y_{n,i} \leq x)\right) \quad t \in [0,1], \ x \in \mathbb{R}. \]

Finally, the limits of (3.3) and (3.4) will be investigated.

### 3.1.2 Reduction principle for the empirical process of triangular arrays

As described in section 2.2, reduction principles are the main tool in the analysis of empirical processes of long-range dependent data. More precisely, the empirical process gets approximated by only the first term (more general by the first \(q\) terms) of its Hermite expansion (Gaussian processes), Appell expansion or Volterra expansion (linear processes). However, all these approximations have in common that they are investigated for stationary observations.

Now consider \(G(X_1), \ldots, G(X_{[nt]}), G_n(X_{[nt]+1}), \ldots, G_n(X_n)\). For the random variables before the change-point we obtain the classical Hermite expansion (2.2). The empirical process then might be approximated just as in Dehling and Taqqu (1989). In contrast, the Hermite expansion of the indicator functions after the change is

\[ 1\{G_n(X_i) \leq x\} - F_{(n)}(x) \overset{L^2}{=} \sum_{q=m(n)}^{\infty} \frac{J_{q,n}(x)}{q!} H_q(X_i). \]

Two difficulties arise: First, \(m(n)\) might be smaller than \(m\), the Hermite rank of \(\{1\{G(\cdot) \leq x\}\}_{x \in \mathbb{R}}\). Secondly, the coefficients \(J_{q,n}(x)\) depend on \(n\) and might converge uniformly to 0. Thus, it is a priori not clear which term of the Hermite expansion is asymptotically dominant or if there are even more than one. The next result is a reduction principle that lays emphasis on this aspects. We will make use of it in the proof of Theorem 6 below, but it is also of interest in its
own right. Just as in Section 2.2, the normalization for the empirical process is given by \( d_{n,m} \),

\[
d_{n,m}^2 = \text{Var} \left( \sum_{i=1}^n H_m(X_i) \right) \sim L_m(n)n^{2H}.
\]

**Theorem 5.** Let \( \{G_n\}_n \) be a sequence of measurable functions and let \( m(n) \) be the sequence of Hermite ranks of \( \{1_{\{G_n(\cdot) \leq x\}}\}_{x \in \mathbb{R}} \). Then, for any \( m \in \mathbb{N} \) with \( m(n) \leq m < 1/D \) (for \( n \geq n_0 \)),

\[
P \left( \sup_{t \in (0,1)} \sup_{x \in \mathbb{R}} \frac{1}{d_{n,m}} \left| \sum_{i=1}^{[nt]} (1_{\{G_n(X_i) \leq x\}} - \sum_{q=0}^{m} \frac{J_{q,n}(x)}{q!} H_q(X_i)) \right| > \epsilon \right) \leq C n^{-\kappa (1 + \epsilon^{-3})},
\]

where \( C \) and \( \kappa \) do not depend on \( n \).

**Remark 3.1.1.** (i) Theorem 5 contains the reduction principle of Dehling and Taqqu (1989), see also (2.4), as a special case (\( G_n(x) = G(x) \) and \( m(n) = m \)).

(ii) Note that \( \{1_{\{G_n(\cdot) \leq x\}} - F(n)(x)\}_{x \in \mathbb{R}} \) might have a Hermite rank smaller than \( m \) (say \( m^* < m \)). Thus, one would expect \( d_{n,m}^{-1} \) as normalization. The weaker normalization \( d_{n,m}^{-1} \) is however possible since the empirical process is approximated by additional terms of the Hermite expansion, in detail those up to \( m \).

(iii) Theorem 2 in Wu (2003), see also Theorem 3 in Section 2.2 of this thesis, contains a stronger type of convergence. Where we state convergence in probability, Wu considers \( L^2 \)-convergence. Then again, he considers only the normal empirical process, while we also treat the sequential version. Moreover, we consider triangular arrays, which Wu (2003) does not.

**Corollary 3.1.2.** Let \( \{G_n\}_n \) be sequence of measurable functions and let \( m(n) \) be the sequence of Hermite ranks of \( \{1_{\{G_n(\cdot) \leq x\}}\}_{x \in \mathbb{R}} \). If further \( m^* \leq m(n) \leq m < 1/D \) for all \( n \geq n_0 \) and

\[
\frac{d_{n,q}}{d_{n,m}} \frac{J_{q,n}(x)}{q!} \to h_q(x) \quad \forall q \in \{m^*, \ldots, m\},
\]

uniformly in \( x \), then

\[
\frac{1}{d_{n,m}} \sum_{i=1}^{[nt]} (1_{\{G_n(X_i) \leq x\}} - F(n)(x)) \xrightarrow{D} \sum_{q=m^*}^{m} h_q(x) Z_q(t).
\]

\( (Z_q,t(t)_{t \in [0,1]} \) are uncorrelated, \( q \)th order Hermite processes.

**Proof.** Using the reduction principle, namely Theorem 5, it remains to show that

\[
d_{n,m}^{-1} \sum_{q=m^*}^{m} \frac{J_{q,n}(x)}{q!} \sum_{i=1}^{[nt]} H_q(X_i)
\]
converges to the desired limit processes. Define

\[ Z_{n,q}(t) = \frac{1}{d_{n,q}} \sum_{i=1}^{[nt]} H_q(X_i), \]

and note that because of \(1/m > D\) the sequences \((H_q(X_i))_{i \geq 1}\) are long-range dependent for \(q = m^*, \ldots, m\). Then we have by Theorem 4 in Bai and Taqqu (2013)

\[ (Z_{n,m^*}, \ldots, Z_{n,m}) \overset{D}{\to} (Z_{m^*}, \ldots, Z_m), \tag{3.6} \]

where convergence takes place in \((D[0,1])^{m-m^*+1}\), equipped with the uniform metric. Moreover, \((Z_q(t))_{t \in [0,1]}\) are uncorrelated Hermite processes of order \(q\). The functions \(h_q\) are elements of \(D[-\infty, \infty]\) and therefore they are also bounded, see Remark 3.1.4. Hence, we may apply the continuous mapping theorem and conclude that

\[ \left\{ \sum_{q=m^*}^{m} h_q(x) d_{n,q}^{-1} \sum_{i=1}^{[nt]} H_q(X_i) \right\}_{t,x} \]

converges in distribution to

\[ \left\{ \sum_{q=m^*}^{m} h_q(x) Z_q(t) \right\}_{t,x}, \]

where convergence takes place in \(D([0,1] \times [-\infty, \infty])\), equipped with the supremum norm. The result then follows by the uniform convergence of \(d_{n,m}/d_{n,q}J_{q,n}(x)\) towards \(q!h_q(x)\), the reduction principle and Slutsky’s theorem.

\[ \square \]

**Remark 3.1.3.** The Hermite processes occurring in the limit are dependent, see Proposition 1 in Bai and Taqqu (2013). To the best of our knowledge, such a limiting distribution has not been obtained for empirical processes in the literature.

**Remark 3.1.4.** In view of the proof of Corollary 3.1.2 it is important to note that the functions \(h_q\) are uniform limits of the càdlàg-functions \(J_{m,n}(x)\) and hence elements of \(D[-\infty, \infty]\). As a consequence they are also bounded (Pollard (1984)).

**Example 3.1.5.** There are indeed sequences of functions \(\{G_n\}_n\) that satisfy the conditions of Corollary 3.1.2. Consider the transformations

\[ G_n(x) = \begin{cases} a_n x^2, & \text{if } x \geq 0, \\ x^2, & \text{if } x < 0, \end{cases} \]
with $a_n \to 1$ and $a_n \neq 1$. Thus, we are in the situation of Theorem 5 with $m(n) = 1$ for all $n \in \mathbb{N}$. One obtains, as $a_n \to 1$,

$$\sup_{x \in \mathbb{R}} |J_{2,n}(x) - J_2(x)| \to 0,$$

with

$$J_2(x) = E[1_{\{X_1^2 \leq x\}}(X_1^2 - 1)] = -2\sqrt{x}\phi(\sqrt{x})1_{\{x \geq 0\}}.$$

If in addition $a_n \sim n^{-D/2}L^{1/2}(n) \sim d_{n,2}/d_{n,1}$, then

$$\sup_x \left| \frac{d_{n,1}}{d_{n,2}} J_{1,n}(x) - C_x \phi(\sqrt{x})1_{x \geq 0} \right| \to 0,$$

for some constant $C$ depending on $D$ only. Corollary 3.1.2 then holds with $m = 2$, $m^* = 1$, $h_1(x) = C_x \phi(\sqrt{x})1_{x \geq 0}$ and $h_2(x) = J_2(x)/2$.

### 3.1.3 The empirical process under local alternatives

Let us again consider the change-point model

$$Y_{n,i} = \begin{cases} G(X_i), & \text{if } i \leq \lfloor n\tau \rfloor, \\ G_n(X_i), & \text{if } i \geq \lfloor n\tau \rfloor + 1, \end{cases} \quad (3.7)$$

for measurable functions $G$ and $(G_n)_n$ and unknown $\tau \in (0,1)$. For $\tau = 0$ one gets a row-wise stationary triangular array, as considered in the previous section, and for $\tau = 1$ a stationary sequence, as in Dehling and Taqqu (1989). In what follows we will denote the distribution functions of $G(X_1)$ and $G_n(X_1)$ by $F$ and $F_n$, respectively.

To obtain weak convergence of the empirical process of (3.7) we have to make some assumptions on the structure of the change and the Hermite rank.

**Assumption A:**

A1. The class of functions $\{1_{\{G(\cdot) \leq x\}}\}_{x \in \mathbb{R}}$ has Hermite rank $m$ with $0 < D < 1/m$.

A2. Let $m(n)$ be the Hermite rank of $\{1_{\{G_n(\cdot) \leq x\}}\}_{x \in \mathbb{R}}$ and $m^* = \lim \inf_{n \to \infty} m(n)$. Then we assume

$$n^{(m-m^*)D(1+\delta)/2} \sup_{x \in \mathbb{R}} \left( P(\min\{G(X_1), G_n(X_1)\} \leq x) - P(\max\{G(X_1), G_n(X_1)\} \leq x) \right) \to 0,$$

for some $\delta > 0$. 

Theorem 6. If Assumption A holds, then
\[ \frac{1}{dn,m} \sum_{i=1}^{\lfloor nt \rfloor} \{ 1 \{ Y_n,i \leq x \} - P(Y_n,i \leq x) \} \overset{\mathcal{D}}{\to} \frac{J_m(x)}{m!} Z_{m,H}(t), \]
where \( J_m(x) \) is the Hermite coefficient function of \( 1_{\{ G(x) \leq x \}} \) and \( (Z_{m,H}(t))_t \) is an \( m \)-th order Hermite process. The convergence takes place in \( D([0,1] \times [-\infty, \infty]) \), equipped with the uniform topology.

Remark 3.1.6. (i) For given functions \( G(x) \) and \( G_n(x) \), Assumption A2 might easily being checked, see the examples below. It serves to ensure convergence of the Hermite coefficients \( J_{q,n}(x) = E[1_{\{ G_n(X_i) \leq x \}} H_q(X_i)] \). In detail,
\[ \sup_{x \in \mathbb{R}} \{ P(\min \{ G(X_1), G_n(X_1) \} \leq x) - P(\max \{ G(X_1), G_n(X_1) \} \leq x) \} \to 0 \]
implies, see the proof of Lemma 3.6.5,
\[ \sup_{x \in \mathbb{R}} |J_{q,n}(x) - J_q(x)| \to 0 \quad \forall q \in \mathbb{N}. \]  
(3.8)

By Assumption A1, \( J_1(x) = \ldots J_{m-1}(x) = 0 \) for all \( x \in \mathbb{R} \), yet \( J_m(x) \neq 0 \) for some \( x \). Together with (3.8) this implies \( m^* = \lim \inf_{n \to \infty} m(n) \leq m \).

(ii) Moreover, A2 implies convergence of the marginal distribution function. To see this, note
\[ |F_n(x) - F(x)| = \max \{ F_n(x), F(x) \} - \min \{ F_n(x), F(x) \} \]
\[ \leq P(\min \{ G(X), G_n(X) \} \leq x) - P(\max \{ G(X), G_n(X) \} \leq x) \]
and \( n^{(m-m^*)D(1+\delta)/2} = O(1) \). However, the converse is not always true. Consider for instance the functions \( G(x) = x \) and \( G_n(x) = G_1(x) = -x \) or the situation in Example 3.2.7. Then again, there are lots of natural choices of \( G \) and \( G_n \) for whom convergence of the marginal distribution functions (with a certain rate) implies Assumption A2. Among them \( G_n(x) = G(x) + \mu_n \) (mean-shift), \( G_n(x) = \sigma_n G(x) \) (change in variance) and
\[ G_n(x) = F_{(n)}^{-1} \circ \Phi(x) \quad \text{and} \quad G(x) = F^{-1} \circ \Phi(x). \]

(iii) Our assumptions explicitly allow for the Hermite rank to change together with the marginal distribution. However, the limit behavior seems to be untouched by this change. In fact, if compared to Corollary 3.1.2, only one Hermite process is involved in the limit of Theorem 6. Intuitively this corresponds to the idea that the change in distribution and the change in the Hermite coefficient, both caused by the difference of \( G \) and \( G_n \), are of the same order. For \( q < m \) this enforces the function \( J_{q,n}(x) \) to converge rather fast to 0. Technically this can be
explained through A2. If this assumption is dropped, we might actually encounter limits with multiple Hermite processes. Such a case will be considered in Example 3.2.7.

(iv) If A1 is violated, the sequence \( \{G(X_i)\}_{i \geq 1} \) is actually short-range dependent. For stationary observations, Csörgő and Mielniczuk (1996) showed convergence of the sequential empirical process to a two-parameter Gaussian process. Change-point alternatives have not been considered for such random variables yet, but would require fundamentally different proofs compared to our results.

### 3.2 Asymptotics for change-point statistics

#### 3.2.1 Kolmogorov-Smirnov and Cramér-von Mises statistics

We now apply the results concerning empirical processes to determine the asymptotic distribution of the Kolmogorov-Smirnov and Cramér-von Mises statistic. To reiterate, these are (now properly normalized)

\[
T_n = d_{n,m}^{-1} \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} \left| \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right|, \tag{3.1}
\]

and

\[
S_n = d_{n,m}^{-2} \sup_{t \in [0,1]} \int_{\mathbb{R}} \left| \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right|^2 d\hat{F}_n(x), \tag{3.2}
\]

respectively. To obtain a non degenerate limit under a sequence of local alternatives it is important to choose the right amount of change. For a mean-shift this is naturally the difference of the expectations before and after the change. For a general change we formulate the test problem as follows: We wish to test the hypothesis

\[ H : \text{ Assumption A1 holds and } G_n(x) = G(x) \text{ for all } x \in \mathbb{R} \text{ and } n \geq 1, \]

against the sequence of local alternatives

\[ A_n : \text{ Assumption A holds and, for } n \to \infty, \]

\[
\frac{n}{d_{n,m}} (F(x) - F_{(n)}(x)) \to g(x), \tag{3.3}
\]

uniformly in \( x \), where \( g(x) \) is a function that is measurable, of bounded total variation and whose support has positive Lebesgue measure.

**Remark 3.2.1.** Note that \( nd_{n,m}^{-1} \sim n^{mD/2}L^{-m/2}(n) \). Thus (3.3) implies

\[
n^{(m-m^*)D(1+\delta)/2} (F(x) - F_{(n)}(x)) \to 0,
\]
for \( \delta < m^*/(m - m^*) \) or \( m^* = m \). This again implies Assumption A2 for certain choices of functions \( G \) and \( G_n \), see Remark 3.1.6 (ii).

**Theorem 7.** (i) Under the hypothesis \( H \) of no change we have, as \( n \to \infty \),

\[
T_n \overset{D}{\to} \sup_{x \in \mathbb{R}} |J_m(x)/(m!)| \sup_{t \in [0,1]} \left| \tilde{Z}_{m,H}(t) \right|
\]

and

\[
S_n \overset{D}{\to} \int_{x \in \mathbb{R}} (J_m(x)/(m!))^2 \, dF(x) \sup_{t \in [0,1]} \left| \tilde{Z}_{m,H}(t) \right|^2,
\]

where \( \tilde{Z}_{m,H}(t) = Z_{m,H}(t) - tZ_{m,H}(1) \) and \( (Z_{m,H}(t))_{t \in [0,1]} \) is an \( m \)-th order Hermite process.

(ii) Under the sequence of local alternatives \( A_n \) we have, as \( n \to \infty \),

\[
T_n \overset{D}{\to} \sup_{x \in \mathbb{R}} \sup_{t \in [0,1]} \left| J_m(x)/(m!) \tilde{Z}_{m,H}(t) - g(x)\psi_r(t) \right|
\]

and

\[
S_n \overset{D}{\to} \sup_{t \in [0,1]} \int_{x \in \mathbb{R}} (J_m(x)/(m!) \tilde{Z}_{m,H}(t) - g(x)\psi_r(t))^2 \, dF(x),
\]

where \( \psi_r(t) \) is called change-point function and defined by

\[
\psi_r(t) = \begin{cases} 
  t(1 - \tau), & \text{if } t \leq \tau, \\
  \tau(1 - t), & \text{if } t > \tau.
\end{cases}
\]

**Proof.** We give the proof for a sequence of local alternatives. The asymptotic behavior under the hypothesis then is an immediate consequence. Obtain the following decomposition of the empirical bridge-process

\[
\frac{1}{d_{n,m}} \left( \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{ni} \leq x\}} - \frac{nt}{n} \sum_{i=1}^{n} 1_{\{Y_{ni} \leq x\}} \right)
\]

\[
= \frac{1}{d_{n,m}} \left( \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{ni} \leq x\}} - H_{n,i}(x) \right) - t \sum_{i=1}^{n} \left( 1_{\{Y_{ni} \leq x\}} - H_{n,i}(x) \right)
\]

\[
+ \left( t - \frac{nt}{n} \right) \frac{1}{d_{n,m}} \sum_{i=1}^{n} \left( 1_{\{Y_{ni} \leq x\}} - H_{n,i}(x) \right)
\]

\[
+ \frac{n}{d_{n,m}} \psi_{n,r}(t) \left( F(x) - F_{(n)}(x) \right),
\]

where

\[
\psi_{n,r}(t) = \begin{cases} 
  \frac{\lfloor nt \rfloor}{n} \left( 1 - \frac{\lfloor nt \rfloor}{n} \right), & \text{if } t \leq \tau, \\
  \frac{\lfloor nt \rfloor}{n} \left( 1 - \frac{\lfloor nt \rfloor}{n} \right), & \text{if } t > \tau.
\end{cases}
\]

By uniform convergence of \( n/d_{n,m}(F(x) - F_{(n)}(x)) \) and \( \psi_{n,r}(t) \) towards \( g(x) \) and \( \psi_r(t) \), respectively, Theorem 6 and the continuous mapping theorem, one gets that (3.4) converges weakly
towards
\[ J_m(x)/(m!) \left( Z_m(t) - tZ_m(t) \right) + \psi_r(t)g(x). \]

The convergence of the Kolmogorov-Smirnov type statistic then follows from continuity of the application of the supremum norm. The Cramér-von Mises statistic \( S_n \) can be written as \( S_n = \sup_{t \in [0,1]} M_n(t) \), where

\[
M_n(t) = d_{n,m}^{-2} \int_{\mathbb{R}} \left( \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} - \frac{nt}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 d\tilde{F}_n(x) \]
\[
= d_{n,m}^{-2} \int_{\mathbb{R}} \left( \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} - \frac{nt}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 dF(x) \quad (3.5) 
\]
\[
+ d_{n,m}^{-2} \int_{\mathbb{R}} \left( \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} - \frac{nt}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 d(\tilde{F}_n(x) - F(x)). \quad (3.6) 
\]

Due to the convergence of (3.4) and the continuous mapping theorem, (3.5) converges to the desired limit process. Thus, it remains to show that (3.6) is negligible. Therefore, obtain

\[
d_{n,m}^{-2} \int_{\mathbb{R}} \left( \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} - \frac{nt}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 d(\tilde{F}_n(x) - F(x)) 
\]
\[
= \int_{\mathbb{R}} \left( \frac{J_m(x)/(m!) \tilde{Z}_m(t) - \psi_r(t)g(x)}{n} \right)^2 d(\tilde{F}_n(x) - F(x)) \quad (3.7) 
\]
\[
+ \int_{\mathbb{R}} \left\{ d_{n,m}^{-2} \left( \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} - \frac{nt}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 
\quad - \left( \frac{J_m(x)/(m!) \tilde{Z}_m(t) - \psi_r(t)g(x)}{n} \right)^2 \right\} d(\tilde{F}_n(x) - F(x)). \quad (3.8) 
\]

Using the Skorohod-Dudley-Wichura representation theorem (whose conditions are satisfied because \( J_m(x)/(m!) \tilde{Z}_m(t) - \psi_r(t)g(x) \) lays almost surely in \( C([0,1] \times [-\infty, \infty]) \)), one can assume without loss of generality that

\[
d_{n,m}^{-2} \left( \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} - \frac{nt}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right)^2 
\quad - \left( \frac{J_m(x)/(m!) \tilde{Z}_m(t) - \psi_r(t)g(x)}{n} \right)^2 
\]
converges almost surely to 0, uniformly in \( x \) and \( t \). Thus, (3.8) converges to 0, uniformly in \( t \).
Next consider (3.7)

\[
\int_{\mathbb{R}} \left( J_m(x)/(m!) \hat{Z}_m(t) - \psi_+(t)g(x) \right)^2 d(\hat{F}_n(x) - F(x))
\]

\[
= (\hat{Z}_m(t))^2/(m!)^2 \int_{\mathbb{R}} J_m^2(x) \, d(\hat{F}_n(x) - F(x))
\]

\[
- 2\hat{Z}_m(t)\psi_+(t)/(m!) \int_{\mathbb{R}} J_m(x)g(x) \, d(\hat{F}_n(x) - F(x))
\]

\[
+ \psi_+^2(t) \int_{\mathbb{R}} g^2(x) \, d(\hat{F}_n(x) - F(x))
\]

\[
= I_n - II_n + III_n.
\]

As a consequence of Theorem 6 and \( F_n(x) \to F(x) \) one gets a weak Glivenko-Cantelli type convergence, in detail

\[
\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \leq \sup_{x \in \mathbb{R}} \left| \hat{F}_n(x) - \sum_{i=1}^{n} H_{n,i}(x) \right| + \frac{n - \lfloor n\tau \rfloor}{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{P} 0.
\]

Moreover, obtain that \( J_m(x) \) is of bounded variation (this was also noted in Dehling and Taqqu (1989)). To see this, let \([a, b]\) be an arbitrary interval and \( \{x_i\}_{i=0}^{n} \) a partition of this interval. Then

\[
\sum_{i=0}^{n-1} |J(x_{i+1}) - J(x_i)| = \sum_{i=0}^{n-1} |E[1_{\{x_i < G(X_1) \leq x_{i+1}\}} H_m(X_1)]|
\]

\[
\leq \sum_{i=0}^{n-1} E[1_{\{x_i < G(X_1) \leq x_{i+1}\}} |H_m(X_1)|]
\]

\[
= E \left[ \sum_{i=0}^{n-1} 1_{\{x_i < G(X_1) \leq x_{i+1}\}} |H_m(X_1)| \right]
\]

\[
= E \left[ 1_{\{G(X_1) \in [a,b]\}} |H_m(X_1)| \right]
\]

\[
\leq E|H_m(X_1)|.
\]

By the boundedness of \( J_m \), \( J_m^2 \) is also of bounded variation and thus integration by parts, together with the weak Glivenko-Cantelli-type result, yields

\[
I_n = -(\hat{Z}_m(t))^2/(m!)^2 \int_{\mathbb{R}} (\hat{F}_n(x) - F(x)) \, dJ_m^2(x) \xrightarrow{P} 0.
\]

By definition, the function \( g(x) \) is of bounded variation. Hence the same is true for \( g^2(x) \) and by the same arguments as above one gets \( III_n = o_P(1) \). Finally, \( II_n = o_P(1) \), which can be seen using Hölders’s inequality. This finishes the proof. \( \square \)
Remark 3.2.2. Note that our proof of the weak convergence of the Cramér-von Mises statistic would not work for short-range dependent time series. The reason is the completely different limit behavior of the sequential empirical process. Instead of the semi-degenerate process $J_m(x)Z_m(t)$ one gets a Gaussian process $K(t,x)$. While $J_m$ is of bounded variation, this is not the case for sample paths of $K$. Hence $\int_{\mathbb{R}} K(t,x) \, d(F_n(x) - F(x))$ cannot be treated simultaneously to $\int_{\mathbb{R}} J_m(x) \, d(F_n(x) - F(x))$.

Motivated by Theorem 7, we consider change-point tests based on the statistics $T_n$ and $S_n$. Critical values might be chosen as

$$\sup_{x \in \mathbb{R}} |J_m(x)/(m!)| q_{1-\alpha,m,H}$$

and

$$\int_{\mathbb{R}} (J_m(x)/(m!))^2 \, dF(x) q_{2-\alpha,m,H}^2$$

for the Kolmogorov-Smirnov test and the Cramér-von Mises test, respectively. Here $q_{1-\alpha,m,H}$ is the $(1-\alpha)$-quantile of $\sup_{t \in [0,1]} |\tilde{Z}_{m,H}(t)|$. Thereby the tests have asymptotically level $\alpha$ and nontrivial power against local alternatives.

The tests can be performed, if the right normalization for the empirical process, the supremum of $J_m(x)$ and the distribution of $\sup_{t \in [0,1]} |\tilde{Z}_{m,H}(t)|$ are known. In practical applications this might be not the case. Solutions are self-normalization (Shao (2011), Betken (2016)), estimating the the Hurst-coefficient (see for example Künsch (1987)) and bootstrap estimators for $J_m(x)$ (see Chapter 4 of this thesis).

3.2.2 Examples

**Example 3.2.3 (Mean-shift).** Let $G_n(x) = G(x) + \mu_n$ with $\mu_n \sim d_n/n$, then we get the typical change-in-mean problem. In the case of long-range dependent subordinated Gaussian processes this was considered in Dehling et al. (2012, 2017), Csörgő and Horvath (1997), Shao (2011) and Betken (2016). Let $f_G$ be the probability density of $G(X_1)$, and assume that it is continuous and of bounded variation. Then we obtain

$$\frac{n}{d_{n,m}} (F(x) - F_{(n)}(x)) = \frac{n}{d_{n,m}} (F(x) - F(x - \mu_n)) \to Cf_G(x),$$

where, due to continuity of $f_G$, the convergence holds uniformly.

**Example 3.2.4 (Change in the variance).** To describe the change-in-variance-problem define
\( G_n(x) = 1/(1 - \delta_n)G(x) \), with \( \delta_n \sim d_n/n \). For ease of notation let \( \delta_n = d_n/n \). Then we get

\[
\sup_{x \in \mathbb{R}} \left| \delta_n^{-1}(F(x) - F_n(x)) - xf_G(x) \right| \\
= \sup_{x \in \mathbb{R}} \left| \delta_n^{-1}(F(x) - F(x - \delta_n x)) - xf_G(x) \right| \\
= \sup_{x \in \mathbb{R}} \left| \frac{xF(x) - (x - \delta_n x)F(x - \delta_n x)}{\delta_n x} - F(x - \delta_n x) - xf_G(x) \right| \\
\leq \sup_{x \in \mathbb{R}} \left| \frac{xF(x) - (x - \delta_n x)F(x - \delta_n x)}{\delta_n x} - (xf_G(x) + F(x)) \right| + \sup_{x \in \mathbb{R}} |F(x - \delta_n x) - F(x)|. \tag{3.10}
\]

The derivative of \( xF(x) \) is \( xf_G(x) + F(x) \), hence (3.10) converges to 0. The convergence is uniform, if \( f_G \) and \( F \) are continuous. Since \( F \) is continuous, monotone and bounded, (3.11) converges to 0 too. Thus (3.3) holds with function \( g(x) = xf_G(x) \). Assume without loss of generality \( \sigma_n = 1/(1 - \delta_n) > 1 \), then

\[
P(\max\{G(X_1), G_n(X_1)\} \leq x) \\
= P(\sigma_n G(X_1) \leq x, G(x) \geq 0) + P(G(X_1) \leq x, G(X_1) \leq 0) \\
= \begin{cases} 
  F(x/\sigma_n), & \text{if } x \geq 0, \\
  F(x), & \text{if } x < 0.
\end{cases}
\]

The minimum can be treated analogously, hence Assumption A2 follows from convergence of the marginals.

Additionally, one might consider a combined change in mean and variance, given through \( G_n(x) = \sigma_n G(x) + \mu_n \). In this case, (3.3) holds with \( g(x) = f_G(x)(C_1 + C_2 x) \).

**Example 3.2.5** (Generalized inverse of a mixture distribution). By using the generalized inverse of a distribution function one could generate subordinated Gaussian processes with any given marginals, see for example Dehling et al. (2017). We use this for the change-point problem by setting

\[
G \equiv F^{-1} \circ \Phi \quad \text{and} \quad G_n \equiv F_{(n)}^{-1} \circ \Phi.
\]

For a continuous distribution function \( F^* \), define the mixture

\[
F_{(n)}(x) = (1 - \delta_n)F(x) + \delta_n F^*(x),
\]
with $\delta_n \sim d_n n^{-1}$. Then (3.3) holds with $g(x) = F^*(x) - F(x)$ and moreover

\[ P(\max\{G(X_1), G_n(X_1)\} \leq x) = P(\max\{F^{-1} \circ \Phi(X_1), F_{(n)}^{-1} \circ \Phi(X_1)\} \leq x) \]
\[ = P(\Phi(X_1) \leq \min\{F(x), F_{(n)}(x)\}) \]
\[ = \min\{F(x), F_{(n)}(x)\}. \]

Analogously, one has $P(\min\{G(X_1), G_n(X_1)\} \leq x) = \max\{F(x), F_{(n)}(x)\}$. Hence,

\[ P(\min\{G(X_1), G_n(X_1)\} \leq x) - P(\max\{G(X_1), G_n(X_1)\} \leq x) = |F_{(n)}(x) - F(x)|. \]

Thereby, Assumption A2 is satisfied too. For strongly mixing data similar local alternatives were considered by Inoue (2001).

**Example 3.2.6** ($\chi^2$-distribution). Consider a $\chi^2$-distribution given through $G(x) = x^2$ and note that the indicator functions have Hermite rank $m = 2$, see also Dehling and Taqqu (1989). Further let

\[ G_n(x) = \begin{cases} 
  a_n x^2, & \text{if } x \geq 0, \\
  x^2, & \text{if } x < 0,
\end{cases} \]

with Hermite ranks $m(n) = 1$ for all $n \in \mathbb{N}$. If $(a_n - 1) \sim d_n 2/n$, then one can show (similar to the case of a variance change in Example 3.2.4) that

\[ \frac{n}{d_{n,2}} \left( P(G(X_1) \leq x) - P(G_n(X_1) \leq x) \right) \rightarrow C\sqrt{x}\phi(\sqrt{x})1_{[0,\infty)}(x), \]

uniformly in $x$. As Assumption A2 is satisfied too, we may apply Corollary 7 (ii) with function $g(x) = C\sqrt{x}\phi(\sqrt{x})1_{[0,\infty)}(x)$ and $m = 2$.

**Example 3.2.7** (Multiple Hermite processes in the limit). In the previous example, both the marginal distribution and the Hermite rank changed. However, the limiting process seems to be untouched by this fact and one might ask whether this is intuitive or not.

It is caused by the fact that the change in the distribution and the change in the Hermite coefficients, both originating in the difference of the functions $G(x)$ and $G_n(x)$, are of the same order.

To get an additional Hermite process in the limit, one would need $(a_n - 1) \sim d_{n,2}/d_{n,1}$, see Corollary 3.1.2 and its proof. But then

\[ \frac{n}{d_{n,2}} \sup_x |F(x) - F_{(n)}(x)| = \frac{n}{d_{n,1}} \frac{d_{n,1}}{d_{n,2}} \sup_x |F(x) - F_{(n)}(x)| \rightarrow \infty, \]
and the test would have asymptotic power 1. To achieve nontrivial asymptotic power, one has to consider structural breaks that consist of two aspects and where only one is captured by the marginal distribution. To this end, define the transformations

\[ G(x) = \Phi^{-1}(F(|x|)) = \Phi^{-1}(2\Phi(|x|) - 1) \]

and

\[ G_n(x) = \Phi^{-1}(F_n^*(G_n^*(x))) + \mu_n, \]

where \( F_n^*(x) = P(G_n^*(X_i) \leq x) \) and \( G_n^*(x) = \begin{cases} a_n x^2, & \text{if } x \geq 0, \\ x^2, & \text{if } x < 0 \end{cases} \)

for some sequence \((a_n)_n\) with \( a_n \neq 1 \) and \( a_n \to 1 \). On the one hand, \( \{1_{G(\cdot) \leq x}\}_x \) has Hermite rank \( m = 2 \) and \( G(X_i) \sim N(0, 1) \). On the other hand, \( \{1_{G_n(\cdot) \leq x}\}_x \) has Hermite rank \( m(n) = 1 \) for all \( n \in \mathbb{N} \) and \( G_n(X_i) \sim N(\mu_n, 1) \). Now let \( \mu_n \sim d_n, 2/n \), then Example 3.2.3 applies and we obtain

\[ \frac{n}{d_n, 2} (F_n(x) - F(x)) = \frac{n}{d_n, 2} (\Phi(x - \mu_n) - \Phi(x)) \to C_{\phi}(x), \]

for any sequence \((a_n)_n\). In contrast, the convergence of the Hermite coefficients is highly influenced by \((a_n)_n\). If the sequence is chosen such that \((a_n - 1) \sim d_n, 2/d_n, 1 \) (therefore, it converges slower than \( \mu_n \)), then the sequential empirical process will converge towards

\[ K(x, t) = \begin{cases} J_2(x)/2Z_2(t), & \text{if } t \leq \tau, \\ \tilde{J}_1(x)Z_1(t) + J_2(x)/2Z_2(t), & \text{if } t > \tau. \end{cases} \]

Actually, this can be proved similar to Corollary 3.1.2. Moreover, the Kolmogorov-Smirnov statistic converges weakly to

\[ \sup_{t \in [0, 1]} \sup_{x \in \mathbb{R}} |K(x, t) - tK(x, 1) - \psi_r(t)C_{\phi}(x)|. \]

We find this example rather pathological, therefore such situations are excluded from the main results via Assumption A2.
3.2.3 Weakening the conditions

The normalized difference of the distribution functions before and after the change is given by

\[ g_n(x) := \frac{n}{d_{n,m}} (F(x) - F_n(x)). \]  

(3.12)

It determines the power of the Kolmogorov-Smirnov and Cramér-von Mises test. In the previous sections we required \( g_n(x) \) to converge uniformly towards some function \( g(x) \), see the formulation of local alternatives \( A_n \). This assumption is quite handy for proving a functional limit theorem in \( D([0,1] \times [-\infty, \infty]) \), but in some situations it is too restrictive.

For instance consider a mean-shift in a continuous distribution whose density is discontinuous. The function \( g_n(x) \) then converges pointwise to the density but of course this convergence cannot be uniform. Prominent examples are uniform, exponential and Pareto distribution. In Dehling et al. (2017) the power of CUSUM and Wilcoxon test against a mean-shift in Pareto distributed data is compared. In order to analyze Kolmogorov-Smirnov and Cramér-von Mises test in such situations too, we define a new sequence of local alternatives as follows:

\[ A_n^*: \quad F(x) \text{ is continuous, Assumption A and the following conditions holds: There are compact intervals} \]

\[ I_n = [a_n^{(1)}, a_n^{(2)}] \text{ and } I_n^* = [a_n^{(1)*}, a_n^{(2)*}] \text{ with } a_n^{(1)}, a_n^{(1)*} \in \mathbb{R} \cup \{-\infty\}, a_n^{(2)}, a_n^{(2)*} \in \mathbb{R} \cup \{+\infty\} \]

such that:

(i) \( |a_n^{(j)} - a_n^{(j)*}| \to 0 \) for \( j = 1, 2 \) (\( a_n^{(j)} \to \pm \infty \), if \( a_n^{(j)*} = \pm \infty \));

(ii) \( \text{supp}(f) \cup \text{supp}(f_n) \subset I_n^* \), where \( f \) and \( f_n \) are the probability densities of \( G(X_1) \) and \( G_n(X_1) \);

(iii) \( \sup_{x \in I_n}|g_n(x) - g(x)| \to 0 \) for a bounded function \( g(x) \);

(iv) \( \sup_{x \in I_n \setminus I_n^*}|g(x) - |g(a_n^{(j)})|||^+ \to 0 \) for \( j = 1, 2 \);

(v) \( \sup_{x \in I_n \setminus I_n^*}|g_n(x) - |g_n(a_n^{(j)})|||^+ \to 0 \) for \( j = 1, 2 \).

**Corollary 3.2.8.** Under the sequence of local alternatives \( A_n^* \)

\[ T_n \overset{D}{\to} \sup_{x \in \mathbb{R}} \sup_{t \in [0,1]} \left| J_m(x)/(m!) \tilde{Z}_{m,H}(t) - g(x)\psi_\tau(t) \right|. \]

**Proof.** By Theorem 2.1 (which holds under Assumption A) and the continuous mapping theorem, we have

\[ D_n(x,t) - g_n(x)\psi_{m,\tau}(t) \overset{D}{\to} \frac{J_m(x)}{m!} \tilde{Z}_m(t), \]  

(3.13)
with
\[ D_n(x, t) = \frac{1}{d_{n,m}} \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}}. \]  \hspace{1cm} (3.14)

being the empirical bridge process. By the Skorohod-representation theorem, we can assume without restriction that the convergence in (3.13) holds almost surely. Then one obtains further

\[
\left| T_{KS} - \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g_n(x)| \right|
\leq \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |D_n(x, t) - J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g_n(x)|
\leq \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} \left| (D_n(x, t) - \psi_{n,r}(t)g_n(x)) - J_m(x)/m! \bar{Z}_m(t) \right|
+ \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |\psi_{n,r}(t) - \psi_r(t)| \sup_{x \in \mathbb{R}} |g_n(x)|
= V_{n,1} + V_{n,2}.
\]

\( V_{n,1} \) converges almost surely to 0 by (3.13) and the Skorohod-representation theorem. Convergence of \( V_{n,2} \) follows from uniform convergence of \( \psi_{n,r}(t) \) towards \( \psi_r(t) \) and \( \sup_{x \in \mathbb{R}} |g_n(x)| < \infty \) (which is an immediate consequence of conditions (iii) - (iv)). Thus, in order to prove the corollary it suffices to show

\[
\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g_n(x)| \xrightarrow{P} \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g(x)|. \]  \hspace{1cm} (3.15)

To this end note

\[
\sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g_n(x)| - \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g(x)|
\leq \sup_{t \in [0,1]} \sup_{x \in I_n} |\psi_r(t)g_n(x) - \psi_r(t)g(x)|
+ \sup_{t \in [0,1]} \left\{ \sup_{x \in \mathbb{R}} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g_n(x)| - \sup_{x \in I_n} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g(x)| \right\}
+ \sup_{t \in [0,1]} \left\{ \sup_{x \in \mathbb{R}} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g(x)| - \sup_{x \in I_n} |J_m(x)/m! \bar{Z}_m(t) + \psi_r(t)g(x)| \right\}
= W_{n,1} + W_{n,2} + W_{n,3}.
\]

First note that \( F(x) \) is assumed to be continuous, hence it is also uniformly continuous and the same is true for \( J_m(x) \) (which can be seen easily through its definition). Moreover, by
monotone (or dominated) convergence and definition of $I_n$
\[
\sup_{x \notin I_n} |J_m(x)| \to 0,
\]
(3.16)
as $n \to \infty$. Finally let $x_0$ be such that $|J_m(x_0)| = \sup_x |J_m(x)|$ and note that for sufficiently large $n$
\[
x_0 \in I_n \quad \text{and} \quad |J_m(x_0)| > \sup_{x \notin I_n} |J_m(x)|.
\]
Now assume $|\tilde{Z}_m(t)| > C$ with
\[
C = \psi_\tau(\tau)m! \frac{\sup_x |g(x)| - |g(x_0)|}{|J_m(x_0)| - \sup_{x \notin I_{n_0}} |J_m(x)|}.
\]
Then by construction of $C$
\[
|J_m(x_0)\tilde{Z}_m(t)/m! + \psi_\tau(t)g(x_0)| \geq \sup_{x \notin I_n} |J_m(x)\tilde{Z}_m(t)/m! + \psi_\tau(t)g(x)|
\]
and consequently $W_{n,2}(t;\omega) = 0$ if $|\tilde{Z}_m(t;\omega)| > C$. Next let $|\tilde{Z}_m(t;\omega)| < C$. Then
\[
\sup_{x \in \mathbb{R}} |J_m(x)/m!\tilde{Z}_m(t) + \psi_\tau(t)g(x)| - \sup_{x \in I_n} |J_m(x)/m!\tilde{Z}_m(t) + \psi_\tau(t)g(x)|
\]
\[
\leq \sum_{j=1}^{2} \sup_{x \in I_{n}} \sup_{x \notin I_n} |J_m(x)/m!\tilde{Z}_m(t) + \psi_\tau(t)g(x)| - |J_m(a_n^{(j)})/m!\tilde{Z}_m(t) + \psi_\tau(t)g(a_n^{(j)})|
\]
\[
\leq 4 \sup_{x \notin I_n} |J_m(x)|C/m! + \sum_{j=1}^{2} \left( \sup_{x \notin I_n} |g(x)| - |g(a_n^{(j)})| \right) + .
\]
The first summand converges to 0 by condition (iv) of $A_n^\ast$ and the second summand by (3.16). Hence $\sup_t |W_{n,2}(t)| \to 0$. The same is true for $\sup_t |W_{n,3}(t)|$ if we replace condition (iv) by (v). This finishes the proof. \(\Box\)

Remark 3.2.9. If $g_n(x)$ converges uniformly to $g(x)$, the asymptotic distribution of the Kolmogorov-Smirnov statistic can be deduced directly from the limit of the empirical process (under long and short memory). However, for functions $g_n(x)$ and $g(x)$ as considered in $A_n^\ast$, our proof relies on the semi-degenerate structure of the limiting process. Therefore, it would be not valid under short memory.

Remark 3.2.10. The conditions of $A_n^\ast$ basically says that the convergence of $g_n(x)$ towards $g(x)$ has to be uniform on an interval $I_n$ and that this interval approximates the support of the involved random variables. Moreover, the functions $g_n(x)$ and $g(x)$ have to behave nicely on $\mathbb{R} \setminus I_n$. For given transformations $G$ and $G_n$, these conditions can be checked quite easily.
as the next example illustrates.

![Figure 3.1: Comparison of the functions $g_n(x)$ (solid line) and $g(x)$ (dashed line) for a mean-shift in uniform[0,1] (left) and Pareto(3) (right) distributed data. The interval $I_n$ is also drawn (thick line).](image)

**Example 3.2.11** (Mean-shift and discontinuous density). Let $G: \mathbb{R} \to \mathbb{R}$ be such that $G(X_i)$ has distribution function $F(x)$ and density $f(x)$ such that

- $F(x)$ is continuous on $[a, \infty)$,
- $F(x)$ is differentiable on $(a, \infty)$ but not in $a$,
- $F(x) = 0$ for $x < a$,
- $f(x)$ is bounded and continuous on $(a, \infty)$.

Moreover, let $G_n(x) = G(x) + \mu_n$ with $\mu_n > 0$ and $\mu_n \sim d_{n,m}/n$. Then define $I_n = [a + \mu_n, \infty)$ and $I^*_n = [a, \infty)$.

If $x \in I_n$, then $F$ is differentiable on $(x - \mu_n, x)$, hence by the mean value theorem

$$
\frac{F(x) - F(x - \mu_n)}{\mu_n} = f(\xi_n(x)), \quad \text{with } \xi_n(x) \in (x - \mu_n, x).
$$

By $|x - \xi_n(x)| \leq \mu_n \to 0$ and uniform continuity of $f: [a, \infty) \to \mathbb{R}$, it follows

$$
\sup_{x \in I_n} \left| \frac{(F(x) - F(x - \mu_n))}{\mu_n} - f(x) \right| \to 0.
$$
It remains to verify condition (iv). We get for \( x \in [a, a + \mu] \)
\[
|g_n(x) - g_n(a_n^{(1)})| = \frac{n}{d_{n,m}} (|F(x) - F(x - \mu)| - |F(a + \mu) - F(a)|)
\]
\[
= \frac{n}{d_{n,m}} (F(x) - F(a + \mu))
\]
\[
\leq 0.
\]
Moreover, by right-continuity of \( f(x) \),
\[
\sup_{x < a + \mu_n} |f(x) - f(a + \mu_n)| \to 0.
\]
Consequently, conditions (iv) and (v) are satisfied.

Part of our formulation of the local alternatives \( A_n^* \) is also Assumption A2, in detail
\[
n^{(m - m^*)D(1+\delta)/2} \sup_{x \in \mathbb{R}} (P(\min\{G(X_1), G_n(X_1)\} \leq x) - P(\max\{G(X_1), G_n(X_1)\} \leq x)) \to 0.
\]
However, for a simple mean-shift, this simplifies to
\[
\sup_{x \in \mathbb{R}} (F(x) - F(x - \mu_n)) \to 0,
\]
which holds due to boundedness of \( f(x) \). Therefore, Corollary 3.2.8 applies.

Similarly, one can verify the conditions for distributions whose density has jumps at both sides of the support.

### 3.3 Application to further change-point tests

#### 3.3.1 CUSUM test

In the previous sections the asymptotic behavior of Kolmogorov-Smirnov and Cramér-von Mises test has investigated. Now we want to compare those tests to other methods (in an asymptotic sense). Maybe the most widespread change-point test is the so called *CUSUM test*. It originates in the Gauss test, which is a two sample test based on the difference of sample means \( \bar{X} - \bar{Y} \). The extension to the change-point problem is given by
\[
\max_{1 \leq k < n} \left| \sum_{i=1}^{k} Y_i - \frac{k}{n} \sum_{i=1}^{n} Y_i \right|.
\]
A broad treatment, including weighted versions of (3.1), can be found in Csörgő and Horvath (1997). Under change-point alternatives, the CUSUM statistic has been considered in Dehling et al. (2017). They consider subordinated Gaussian processes \( Y_i = G(X_i) \) that exhibit long-
range dependence. For a mean-shift
\[ Y_{n,i} = \begin{cases} G(X_i), & \text{if } i \leq \lfloor n\tau \rfloor, \\ G(X_i) + \mu_n, & \text{if } i \geq \lfloor n\tau \rfloor + 1, \end{cases} \]
with \( \mu_n n/d_{n,m} \to \Delta \), they show
\[ C_n = \frac{1}{d_{n,m}} \max_{1 \leq k < n} \left| \sum_{i=1}^{k} Y_{n,i} - \frac{k}{n} \sum_{i=1}^{n} Y_{n,i} \right| \xrightarrow{p} \sup_{t \in [0,1]} \left| \frac{c_m}{m!} \tilde{Z}_m(t) + \psi_{\tau}(t)\Delta \right|. \]

However, other alternatives are thinkable. Take for instance a combined change in mean and variance, modeled via the transformation \( G_n(x) = \sigma_n x + \mu_n \). We have seen in Example 3.2.4 that the asymptotic distributions of Kolmogorov-Smirnov and Cramér-von Mises test are influenced by the additional change in the variance. On the contrary, one might expect that the CUSUM test is not sensitive to this part of the change. Consequently Kolmogorov-Smirnov and Cramér-von Mises test would become more efficient than the CUSUM test. To verify this conjecture we will investigate the CUSUM statistic \( C_n \) under similar local alternatives as in the previous sections.

Note that the CUSUM statistic might be written as a functional of the sequential empirical process. If \( EG(X_i) < \infty \), one has \( Y_t - EY_t = -\int (1_{\{Y_t \leq x\}} - F(x)) dx \) and thereby
\[ \frac{1}{d_{n,m}} \left( \sum_{i=1}^{k} Y_{n,i} - \frac{k}{n} \sum_{i=1}^{n} Y_{n,i} \right) = -\frac{1}{d_{n,m}} \int_{\mathbb{R}} \left( \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_{n,i} \leq x\}} - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} 1_{\{Y_{n,i} \leq x\}} \right) dx. \]

Unfortunately, integration (that is \( f \mapsto \int_{\mathbb{R}} f(x) dx \)) is not a continuous functional on \( D[-\infty, \infty] \), when this space is equipped with the uniform metric. One might circumvent this issue by using a weighted reduction principle of the empirical process as in Buchsteiner (2015). But neither an approximation by multiple terms of the Hermite expansion nor triangular arrays \( (G_n(X_i))_{i \leq n, \ n \geq 1} \) are covered by his results. Therefore, we leave the theory of empirical processes for the moment and consider the partial sum process directly. The next result gives a reduction principle for \( \sum_{i=1}^{\lfloor nt \rfloor} (G_n(X_i) - E G_n(X_i)) \), extending Theorem 4.1 of Taqqu (1975) to sequences of transformations \( G_n \).

**Proposition 3.3.1.** Consider the transformations \( (G_n)_{n \geq 1} \) with Hermite ranks \( m(n) \) and \( m = \lim \inf_{n \to \infty} m(n) < 1/D \). Assume further \( c_{q,n} = EG(X_1) H_q(X_1) \leq c^*_q, \ \forall q \geq 0 \) and
\[ \sum_{q=m}^{\infty} c_q^2 < \infty. \] Then, as \( n \geq n_0 \) and \( k \to \infty \)

\[
(i) \quad E \left[ \sum_{i=1}^{k} \left( G_n(X_i) - \sum_{q=0}^{m} \frac{c_{q,n}}{q!} H_q(X_i) \right) \right]^2 = o(d_{k,m}^2);
(ii) \quad E \left[ \sum_{i=1}^{k} (G_n(X_i) - EG_n(X_i)) \right]^2 - \sum_{q=1}^{m} \frac{c_{n,q}^2}{(q!)^2} E \left[ \sum_{i=1}^{k} H_q(X_i) \right] = o(d_{k,m}^2).
\]

Proof. Define the functions \( G_n^*(x) = G_n(x) - \sum_{q=0}^{m} \frac{c_{q,n}}{q!} H_q(x) \). One gets
\[
E \left[ \sum_{i=1}^{k} \left( G_n(X_i) - \sum_{q=0}^{m} \frac{c_{q,n}}{q!} H_q(X_i) \right) \right]^2 = E \left[ \sum_{i=1}^{k} G_n^*(X_i) \right]^2
= \sum_{j=0}^{k} (k - j) E [G_n^*(X_j)G_n^*(X_{j+1})].
\]

By Proposition 2.1.3 and for sufficiently larger \( n \) (such that \( m(n) \geq m \)), one has
\[
E [G_n^*(X_1)G_n^*(X_{j+1})] = \sum_{q=m+1}^{\infty} c_{q,n}^2 q! \rho^j(j) \leq \sum_{q=m+1}^{\infty} c_{q,n}^2 q! \rho^j(j),
\]
with \( \rho(j) = EX_1X_{j+1} \). As the coefficients of this new Hermite expansion no longer depend on \( n \) we can now proceed as in Taqqu (1975). That is
\[
E \left[ \sum_{i=1}^{k} G_n^*(X_i) \right]^2 \leq \sum_{j=0}^{k} (k - j) \sum_{q=m}^{\infty} c_{q,n}^2 q! \rho^j(j) \leq C \sum_{j=0}^{k} (k - j) \rho^j(j)^{m+1}
\]
with \( C = \sum_{q=m+1}^{\infty} c_{q,n}^2 < \infty \). Finally, it follows from (2.3) and \( \rho(j) \sim j^{-D} L(j) \) that
\[
k \sum_{j=0}^{k} \rho^j(j)^{m+1} \sim L(k)^{m+1} k^{2-(m+1)D} = o(d_{k,m}).
\]

This finishes the proof of (i). For the second part of the proposition obtain
\[
E \left[ \sum_{i=1}^{k} (G_n(X_i) - EG_n(X_i)) \right]^2 = E \left[ \sum_{i=1}^{k} G_n^*(X_i) \right]^2 + E \left[ \sum_{q=1}^{m} \frac{c_{n,q}^2}{(q!)^2} \sum_{i=1}^{k} H_q(X_i) \right]^2,
\]
as \( E[G_n^*(X_1)H_m(X_j)] = 0 \) (which follows from Proposition 2.1.3 and the fact that \( G_n^* \) has an Hermite rank larger than \( m \)). Thus (i) implies (ii). \( \square \)

Remark 3.3.2. Unlike in the reduction principle for the empirical process, we do not allow
sequences of transformations like \( G_n(x) = nG(x) \). In view of an application to local alternatives (where \( G_n \to G \) in some sense), this is not problematic.

**Assumption B:** Consider the triangular array

\[
Y_{n,i} = \begin{cases} 
G(X_i), & \text{if } i \leq \lfloor n\tau \rfloor, \\
G_n(X_i), & \text{if } i \geq \lfloor n\tau \rfloor + 1,
\end{cases}
\]

and let \( m \) be the Hermite rank of \( G \) and \( m^* = \lim \inf m^*(n) \) where \( m^*(n) \) is the Hermite rank of \( G_n \). Assume:

B1. There is a \( \Delta \in \mathbb{R} \), such that \( nd_{n,m}^{-1}(EG(X) - EG_n(X)) \to \Delta \).

B2. The Hermite rank \( m \) of \( G \) satisfies \( D < 1/m \).

B3. Let \( c_{q,n} = EG_n(X)H_q(X) \). Then \( |c_{q,n} - c_q| = o((d_{n,m}/d^*_{n,m})^2) \), for \( q = 0, \ldots, m \).

**Remark 3.3.3.** For the Hermite coefficients we need the rate of convergence \( o((d_{n,m}/d^*_{n,m})^2) \). Compared to Assumption A3 (where \( o(d_{n,m}/d^*_{n,m}) \) is needed), this is slightly stronger. However, if the Hermite rank does not change, this effect vanishes (as we have \( m = m^* \) in this case).

**Theorem 8.** Let Assumption B hold. Then, as \( n \to \infty \),

\[
C_n \overset{D}{\to} \sup_{t \in [0,1]} \left| \frac{c_m(x)}{(m!)^2} \sum_{i=1}^{\lfloor nt \rfloor} H_m(X_i) - \Delta \psi_\tau(t) \right|,
\]

with the change-point function \( \psi_\tau(t) \) defined in Theorem 7.

**Proof.** Define the normalized partial sum process \( S_{n,m}(t) = d_{n,m}^{-1} \sum_{i=1}^{\lfloor nt \rfloor} (Y_{n,i} - EY_{n,i}) \) for \( t \in [0,1] \). Fix a \( t < \tau \), then we have by virtue of Proposition 3.3.1 (i)

\[
E \left[ S_{n,m}(t) - \frac{1}{d_{n,m}(m!)} \sum_{i=1}^{\lfloor nt \rfloor} H_m(X_i) \right]^2 \to 0.
\]
Similarly, we obtain for $t > \tau$ and $c_{q,n,i} = EY_{n,i}H_q(X_i)$

$$E \left[ S_{n,m}(t) - \frac{1}{d_{n,m}} \sum_{q=m^*}^{m} \sum_{i=1}^{[nt]} \frac{c_{q,n,i}}{q!} H_q(X_i) \right]^2 \leq 2E \left[ S_{n,m}(\tau) - \frac{1}{d_{n,m}} \frac{c_m}{(m!)^2} \sum_{i=1}^{[nr]} H_m(X_i) \right]^2$$

$$+ 2E \left[ d_{n,m}^{-1} \sum_{i=[nt]}^{[nr]} \left( G_n(X_i) - EG_n(X_i) - \sum_{q=m^*}^{m} \frac{c_{q,n}}{q!} H_q(X_i) \right) \right]^2 \rightarrow 0,$$

again by part (i) of Proposition 3.3.1. Consequently the finite-dimensional laws of $S_{n,m}(t)$ will converge to the same as that of

$$\frac{1}{d_{n,m}} \sum_{q=m^*}^{m} \sum_{i=1}^{[nt]} \frac{c_{q,n,i}}{q!} H_q(X_i) = \frac{1}{d_{n,m}} \sum_{i=1}^{[nt]} \frac{c_m}{(m!)^2} H_m(X_i) + \frac{1}{d_{n,m}} \sum_{q=m^*}^{m} \frac{c_{q,n} - c_q}{q!} \sum_{i=[nt]}^{[nr]} H_q(X_i).$$

Here we have used that $c_q = 0$ for $q = 1, \ldots, m - 1$. The second summand is negligible, due to Assumption B3. Therefore Theorem 1 directly implies

$$S_{n,m}(t) \xrightarrow{fdd} \frac{c_m}{m!} Z_m(t),$$

for an $m$-th order Hermite process $(Z_m(t))_{t \in [0,1]}$. It remains to verify tightness. Let $0 \leq s < t \leq 1$ and assume $\tau \in [s,t]$ (actually, the next step is even simpler if $\tau < s$ or $\tau > t$). Then

$$E[S_{n,m}(t) - S_{n,m}(s)]^2 = \frac{1}{d_{n,m}^2} E \left[ \sum_{i=[ns]}^{[nt]} (Y_{n,i} - EY_{n,i}) \right]^2$$

$$\leq \frac{2}{d_{n,m}^2} E \left[ \sum_{i=[ns]}^{[nt]} (G(X_i) - EG(X_i)) \right]^2 + \frac{2}{d_{n,m}^2} E \left[ \sum_{i=[nt]}^{[nr]} (G_n(X_i) - EG_n(X_i)) \right]^2$$
By the second part of Proposition 3.3.1, it suffices to consider

\[
\frac{2}{d_{n,m}} \frac{c^2_m}{(m!)^2} E \left[ \sum_{i=\lfloor nt \rfloor}^{\lfloor ns \rfloor} H_m(X_i) \right]^2 + \frac{2}{d_{n,m}} \frac{c^2_q}{(q!)^2} \sum_{q=m^*}^{m} c_{q,n} E \left[ \sum_{i=\lfloor nt \rfloor}^{\lfloor ns \rfloor} H_q(X_i) \right]^2
\]

By Assumption B3, the quantity (3.3) is \(o_p(1)\). Moreover, Lemma 2.1.4 ensures that (3.2) is asymptotically proportional to \(|t-s|^{2H}\). As \(2H > 1\), this implies tightness by a well known criterion (see for example Theorem 15.6 in Billingsley (1968)).

We have thus shown that \(S_{n,m}(t)\) converges in \(D[0,1]\) weakly to \(c_m/(m!)^2 Z_m(t)\). The result then follows by the continuous mapping theorem, Slutsky’s theorem and Assumption B1.

### 3.3.2 Wilcoxon test

Another change-point test that has been considered under long-range dependence recently is the Wilcoxon-test. It is a robust alternative to the CUSUM test and is based on the statistic

\[
W_n = \sup_{t \in [0,1]} \left| \frac{1}{n d_{n,m}} \sum_{i=1}^{\lfloor nt \rfloor} \sum_{j=\lfloor nt \rfloor + 1}^{n} \left( 1\{Y_i \leq Y_j\} - \frac{1}{2} \right) \right|.
\]

Its asymptotic distribution has been investigated in Dehling et al. (2012, 2017) under stationarity and local alternatives, respectively. Moreover Betken (2016) considered a self-normalized version. In this section we generalize the results of Dehling et al. (2012, 2017) by using our findings about the sequential empirical process (see Section 2.2).

Let \(D_n(t, x)\) denote again denote the empirical bridge process of \((Y_i)_{i \geq 1}\), see (3.14). Assume that \(Y_1, \ldots, Y_n\) have a continuous distribution and let \(F_n(x)\) denote the empirical distribution
function of them. Then one obtains

\[
\int_{\mathbb{R}} D_n(t, x) dF_n(x) = \frac{1}{nd_{n,m}} \sum_{j=1}^{n} \left( \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_i \leq Y_j\}} - \frac{|nt|}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq Y_j\}} \right)
\]

\[
= \frac{1}{nd_{n,m}} \left( \sum_{j=\lfloor nt \rfloor+1}^{n} \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_i \leq Y_j\}} + \sum_{j=1}^{\lfloor nt \rfloor} \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_i \leq Y_j\}} - \frac{|nt|}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} 1_{\{Y_i \leq Y_j\}} \right)
\]

\[
= \frac{1}{nd_{n,m}} \left( \sum_{j=\lfloor nt \rfloor+1}^{n} \sum_{i=1}^{\lfloor nt \rfloor} 1_{\{Y_i \leq Y_j\}} + \frac{|nt|(|nt| + 1)}{2} - \frac{|nt| n(n + 1)}{2} \right)
\]

\[
= \frac{1}{nd_{n,m}} \sum_{j=\lfloor nt \rfloor+1}^{n} \sum_{i=1}^{\lfloor nt \rfloor} \left( 1_{\{Y_i \leq Y_j\}} - \frac{1}{2} \right). \tag{3.5}
\]

Note that the second last equality holds almost surely, since \(P(Y_i = Y_j) = 0\) for \(j \neq i\). Consequently, the Wilcoxon test statistic (3.4) has an easy expression as a functional of the empirical process. This matches the findings of Dehling et al. (2012, 2017). The proofs of their asymptotic results are based on the asymptotic theory of the empirical process. Therefore the Hermite rank of the indicator functions \((1_{\{G(x) \leq x\}})_x\) plays a central role. It determines the normalizing sequence \(d_{n,m}^{-1}\) and the long-run variance. In fact, as a direct consequence of Theorem 1 in Dehling et al. (2012), one has

\[
W_n \xrightarrow{D} \left| \int_{\mathbb{R}} \frac{J_m(x)}{m!} dF(x) \right| \sup_{t \in [0,1]} |\tilde{Z}_m(t)|,
\]

as \(n \to \infty\). However, the integral \(\int J_m(x) dF(x)\) might actually be zero, leading to a degenerate limit distribution which cannot be used for statistical inference. The following example illustrates that such transformations \(G\) actually exist.

**Example 3.3.4.** Let \(a > 0\) and define the transformation

\[
G(x) = \begin{cases} 
  x & \text{if } |x| > a, \\
  -x & \text{if } |x| \leq a.
\end{cases}
\]

Although the sequence \((G(X_i))_{i \geq 1}\) is non Gaussian, its marginals are standard normal. Moreover, we obtain for the Hermite coefficient functions

\[
J_m(x; a) = \begin{cases} 
  (-1)^m H_{m-1}(x) \phi(x) & \text{if } |x| > a, \\
  (-1)^m (H_{m-1}(a) + H_{m-1}(-a)) \phi(a) - (-1)^m H_{m-1}(-x) \phi(x) & \text{if } |x| \leq a.
\end{cases}
\]

On the one hand, \(J_1(x; a) \neq 0\), hence the Hermite rank of the indicators is \(m = 1\) for all \(a \in \mathbb{R}\).
On the other hand, define the function $\eta_1 : \mathbb{R} \to \mathbb{R}$ by

$$\eta_1(a) = \int_{\mathbb{R}} J_1(x) d\Phi(x) = -\int_{\mathbb{R}} \phi^2(x) dx + 2 \int_{-a}^{a} \phi^2(x) dx - 4a \phi(a).$$

(3.6)

This function is continuous and moreover $\eta_1(a) \to \int \phi^2(x) dx$ and $\eta_1(a) \to -\int \phi^2(x) dx$ as $a \to \infty$ and $a \to 0$, respectively. Thus there must be an $a \in \mathbb{R}$ with $\int_{\mathbb{R}} J_1(x; a) d\Phi(x) = 0$.

The reason for possibly degenerate limits in the results of Dehling et al. (2012) is the normalizing sequence $d_{n,m}$. It is based on the Hermite rank $m$ of the indicator functions. It turns out, that a non degenerate limiting distribution can be obtained if the Hermite rank of the transformation $F \circ G$ is used instead.

**Proposition 3.3.5.** Let $X_1 \sim N(0,1)$ and $G$ be a measurable transformation such that $P(G(X_1) \leq x) = F(x)$ for a continuous function $F$. Then

$$E[F(G(X_1))H_m(X_1)] = -\int_{\mathbb{R}} J_m(x) dF(x).$$

**Proof.** One has by continuity of $F$ and Fubini’s Theorem

$$\int_{\mathbb{R}} J_m(x) dF(x) = \int_{\mathbb{R}} \int_{\mathbb{R}} 1_{\{G(s) \leq x\}} \phi(s) H_m(s) ds \, dF(x)$$

$$= \int_{\mathbb{R}} (1 - F(G(s)) \phi(s) H_m(s) ds$$

$$= EH_m(X_1) - E[F(G(X_1))H_m(X_1)].$$

Motivated by Example 3.3.4 and Proposition 3.3.2, the next Theorem will be based on the Hermite rank of $F \circ G$. Moreover we will consider slightly more general local alternatives than Dehling et al. (2017). The reason is same as for the CUSUM test: We want to compare Kolmogorov-Smirnov and Cramér-von Mises test to the Wilcoxon test not only under a mean-shift but in more general situations.

**Assumption C:** Consider the triangular array

$$Y_{n,i} = \begin{cases} 
G(X_i), & \text{if } i \leq \lfloor n \tau \rfloor, \\
G_n(X_i), & \text{if } i \geq \lfloor n \tau \rfloor + 1,
\end{cases}$$

and let $m$ be the Hermite rank of $F \circ G$ and $m^* = \lim \inf m^*(n)$ where $m^*(n)$ is the Hermite rank of $(1_{\{G_n(\cdot) \leq x\}})_x$. Assume:

C1. There is a measurable function $g : \mathbb{R} \to \mathbb{R}$ such that $h_n(x) = n/d_{n,m}(F_n(x) - F(x)) - g(x)$
is of bounded total variation (uniformly in $n$) and
\[
\int_{\mathbb{R}} \left\{ \frac{n}{dn,m} (F_n(x) - F(x)) - g(x) \right\} dF(x) \to 0.
\]

C2. The Hermite rank $m$ of $F \circ G$ satisfies $D < 1/m$.

C3. Let $J_{q,n}(x) = E_1(G_n(X) \leq x)H_q(X)$. Then, for $q = 0, \ldots, m$,
\[
\sup_{x \in \mathbb{R}} |J_{q,n}(x) - J_q(x)| = o(d_{n,m}/d_{n,m^*}).
\]

**Theorem 9.** Let Assumption C hold. Then, as $n \to \infty$
\[
W_n \xrightarrow{D} \sup_{t \in [0,1]} \left| \int_{\mathbb{R}} \frac{J_m(x)}{(m!)^2} dF(x) \tilde{Z}_m(t) - \gamma \psi_\tau(t) \right|,
\]
where $\gamma = \int_{\mathbb{R}} g(x)dF(x)$ and the change-point function $\psi_\tau(t)$ is defined in Theorem 7.

**Remark 3.3.6.** (i) The second part of Assumption C1 is equivalent to: For $(X, X') \sim N(0, I_2)$ and $\gamma = \int g(x)dF(x)$ one has
\[
\frac{n}{dn,m} (P(G(X) \leq G_n(X')) - 1/2) \to \gamma.
\]
(ii) Alternatively to C1, one might require
\[
\sup_{x \in \mathbb{R}} \left| \frac{n}{dn,m} (F_n(x) - F(x)) - g(x) \right| \to 0,
\]
as $n \to \infty$. However, there are important examples which satisfy the conditions in Dehling et al. (2012), but fail to satisfy the uniform convergence above (take for instance a mean-shift in a distribution with non-continuous density, see Section 3.2.3). Therefore, we consider the (slightly complicated) condition C1, which is in fact satisfied in all reasonable situations (see for instance Corollary 3.3.7 below).
(iii) The null hypothesis is included in the Theorem via $\gamma = \int g(x)dF(x) = 0$.
(iv) A sufficient condition for C3 to hold is
\[
n^{(m-m^*)D(1+\delta)/2} \sup_{x \in \mathbb{R}} (P(\min\{G(X_1), G_n(X_1)\} \leq x) - P(\max\{G(X_1), G_n(X_1)\} \leq x)) \to 0,
\]
for some $\delta > 0$ which can be seen through the lines of the proof of Lemma 3.6.5. It is not hard to show that this condition is satisfied by most examples, see sections 3.2.2 and 3.2.3.

In the case of a mean-shift, Assumption C holds in a very general setting as the next Corollary indicates. This reproduces the findings in Dehling et al. (2017).
bounded total variation, while they only need a bounded density. Then again, we work with the Hermite rank of $F \circ G$ which Dehling et al. (2012) do not.

**Corollary 3.3.7.** Let $G$ be a measurable transformation with distribution $F$, such that $F \circ G$ has Hermite rank $m < 1/D$ and $G_n(x) = G(x) + \mu_n$ with $\mu_n n/d_{n,m} \to c$. Let further $F(x)$ be continuous and differentiable with a density $f$ of bounded total variation. Then the statement of Theorem 9 holds with $g(x) = cf(x)$ and $\gamma = c \int f^2(x) dx$.

**Proof.** We have to verify Assumption C1 and C3 for the triangular array

$$Y_{n,i} = \begin{cases} G(X_i), & \text{if } i \leq \lfloor n\tau \rfloor, \\ G(X_i) + \mu_n, & \text{if } i \geq \lfloor n\tau \rfloor + 1. \end{cases}$$

First note that for any partition $-\infty = x_0 < x_1 < \cdots < x_k = \infty$ we have by the mean-value theorem (with $\xi(x, n) \in [y - \mu_n, y]$)

$$\sum_{i=1}^{k} \left| \frac{n}{d_{n,m}} (F(x_i - \mu_n) - F(x_i)) - n d_{n,m} (F x_{i-1} - \mu_n) - F(x_{i-1})) \right|$$

$$= \frac{n}{d_{n,m}} \mu_n \sum_{i=1}^{k} |f(\xi(x, n)) - f(\xi(x_{i-1}, n))|$$

$$\leq \frac{n}{d_{n,m}} \mu_n \|f(x)\|_{TV}.$$ 

As $\mu_n n/d_{n,m} \to c$, we have $\|h_n\|_{TV} \leq c_1\|f\|_{TV}$. As $f$ is assumed to be of bounded total variation, the first part of C1 follows. For the second part note

$$\int \frac{n}{d_{n,m}} (F(x - \mu_n) - F(x))dF(x) \to c \int f^2(x)dx,$$

by boundedness of $f$. Thus it remains to verify C3. For a mean-shift one has (using again the mean-value theorem and $d_{n,m} \sim n^{H L^{m/2}(n)} = n^{1-mD/2L^{m/2}(n)}$)

$$n^{(m-m^*)D(1+\delta)/2} \sup_{x \in \mathbb{R}} \left( P(\min\{G(X_1), G_n(X_1)\} \leq x) - P(\max\{G(X_1), G_n(X_1)\} \leq x) \right)$$

$$= n^{(m-m^*)D(1+\delta)/2} \sup_{x \in \mathbb{R}} |F(x - \mu_n) - F(x)|$$

$$\leq n^{(m-m^*)D(1+\delta)/2} \mu_n \sup_{x \in \mathbb{R}} |f(x)|$$

$$= \frac{n^{5mD/2}} {n(1+\delta)^{m^*D/2}L^{m/2}(n) d_{n,m}} \mu_n \sup_{x \in \mathbb{R}} |f(x)|$$

$$\to 0,$$

for sufficiently small $\delta > 0$. By the proof of Lemma 3.6.5 below, this implies C3. \qed
Example 3.3.8. Consider once more the transformation from Example 3.3.4, now with \( a \in \mathbb{R} \) such that \( \int J_1(x; a) d\Phi(x) = 0 \). The Hermite coefficient function for \( q = 2 \) is \( J_2(x; a) = x \phi(a) \) (for any \( a \in \mathbb{R} \)). Consequently, \( \int J_2(x; a) d\Phi(x) = \int x \phi^2(x) = 0 \) and the second term of the Hermite expansion vanishes too. Finally, consider \( q = 3 \). We have

\[
\int \! J_3(x; a) d\Phi(x) = - \int \! (x^2 - 1) \phi^2(x) dx + 2 \int_{-a}^{a} (x^2 - 1) \phi^2(x) dx - 4a(a^2 - 1) \phi(a)
\]

Combining this with (3.7) yields

\[
\int \! J_3(x; a) d\Phi(x) = \begin{cases} \int \! (a^2 - x^2) \phi^2(x) dx - 2 \int_{-a}^{a} (a^2 - x^2) x^2 \phi^2(x) dx, & a < 0 \\ \int_{-\infty}^{a} (a^2 - x^2) \phi^2(x) dx + \int_{a}^{\infty} (a^2 - x^2) \phi^2(x) dx - \int_{-a}^{a} (a^2 - x^2) \phi^2(x) dx & a \geq 0 \end{cases} < 0.
\]

Consequently, Theorem 9 applies with Hermite rank \( m = 3 \).

Proof of Theorem 9. By the reduction principle for triangular arrays (Theorem 5) one has

\[
\frac{1}{d_{n,m}} \sum_{i=1}^{\lfloor nt \rfloor} \left( 1_{\{Y_{n,i} \leq x\}} - \sum_{q=0}^{m} \frac{J_{q,n,i}(x)}{q!} H_q(X_i) \right) \xrightarrow{P} 0,
\]

uniformly in \( t \in [0, 1] \) and \( x \in \mathbb{R} \). Here \( J_{q,n,i}(x) = E 1_{\{Y_{n,i} \leq x\}} H_q(X_i) \). Consequently, a similar reduction principle holds for the empirical bridge process

\[
D_n(t, x) = \frac{1}{d_{n,m}} \left( \sum_{i=1}^{\lfloor nt \rfloor} \sum_{q=0}^{m} \frac{J_{q,n,i}(x)}{q!} H_q(X_i) \right) \xrightarrow{P} 0,
\]

uniformly in \( t \) and \( x \). Combining this with (3.5), the result will follow if we can show

\[
\frac{1}{d_{n,m}} \sum_{q=0}^{m} \int_{\mathbb{R}} L_{n,q}(t, x) dF_n(x) \xrightarrow{D} \int_{\mathbb{R}} \frac{J_m(x)}{(m!)} dF(x) \tilde{Z}_m(t) - \gamma \psi(t),
\]

(3.8)
Application to further change-point tests

in $D[0, 1]$, with

$$L_{n,q}(t, x) = \sum_{i=1}^{\lfloor nt \rfloor} J_{q,i} H_q(X_i) - \frac{\lfloor nt \rfloor}{n} \sum_{i=1}^{n} J_{q,i} H_q(X_i), \quad q = 0, \ldots, m.$$  

To this end, we will treat each summand $L_{n,q}(t, x)$ separately. Let us start with $q = 0$, where randomness is only involved through the integrator. Define the function

$$\psi_{n,\tau}(t) = \begin{cases} \frac{\lfloor nt \rfloor}{n} \left(1 - \frac{|n\tau|}{n}\right), & \text{if } t \leq \tau, \\ \frac{|n\tau|}{n} \left(1 - \frac{nt}{n}\right), & \text{if } t > \tau. \end{cases}$$

Then one obtains

$$\frac{1}{d_{n,m}} \int_{\mathbb{R}} L_{n,0}(t, x) \, d\hat{F}_n(x) = \frac{n}{d_{n,m}} \psi_{n,\tau}(t) \int_{\mathbb{R}} \left(F(x) - F_n(x)\right) d\hat{F}_n(x)$$

$$= \psi_{n,\tau}(t) \int_{\mathbb{R}} \left(\frac{n}{d_{n,m}} (F(x) - F_n(x)) - g(x)\right) d(\hat{F}_n(x) - F(x))$$

$$+ \psi_{n,\tau}(t) \int_{\mathbb{R}} \frac{n}{d_{n,m}} (F(x) - F_n(x)) - g(x) \, dF(x)$$

$$+ \psi_{n,\tau}(t) \int_{\mathbb{R}} g(x) \, d\hat{F}_n(x) - F(x))$$

$$+ \psi_{n,\tau}(t) \int_{\mathbb{R}} g(x) \, dF(x).$$

Moreover, we have (by Theorem 6 and Assumption C3, respectively)

$$\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \leq \sup_{x \in \mathbb{R}} \left| n^{-1} \sum_{i=1}^{n} (1_{Y_{n,i} \leq x} - P(Y_{n,i} \leq x)) \right| + \frac{n - |n\tau|}{n} \sup_{x \in \mathbb{R}} |F(x) - F_n(x)|$$

$$\to 0,$$  

as $n \to \infty$. In combination with Assumption C1 and (uniform) convergence of $\psi_{n,\tau}(t)$ towards $\psi_{\tau}(t)$, this yields

$$\frac{1}{d_{n,m}} \int_{\mathbb{R}} L_{n,0}(t, x) \, d\hat{F}_n(x) \to \psi_{n,\tau}(t) \int_{\mathbb{R}} g(x) \, dF(x),$$

uniformly in $t$. 
Next consider \( q = 1, \ldots, m - 1 \) and write
\[
L_{n,q}(t,x) = \frac{J_q(x)}{q!} \left( \sum_{i=1}^{\lfloor nt \rfloor} H_q(X_i) - \frac{nt}{n} \sum_{i=1}^{n} H_q(X_i) \right)
+ \frac{J_{q,n}(x) - J_q(x)}{q!} \left( \sum_{i=\lfloor n\tau \rfloor + 1}^{\lfloor nt \rfloor} H_q(X_i) - \frac{nt}{n} \sum_{i=\lfloor n\tau \rfloor + 1}^{n} H_q(X_i) \right).
\]

However, by Assumption C2 and Proposition 3.3.5, \( \int J_q(x)dF(x) = 0 \) for \( q = 1, \ldots, m - 1 \). Thus
\[
\left| \frac{1}{d_{n,m}} \int_R L_{n,q}(t,x) d\hat{F}_n(x) \right| \leq \frac{d_{n,q}}{d_{n,m}} \int_R \left| J_{q,n}(x) - J_q(x) \right| d\hat{F}_n(x) \frac{1}{d_{n,q}} \left| \sum_{i=\lfloor n\tau \rfloor + 1}^{\lfloor nt \rfloor} H_q(X_i) - \frac{nt}{n} \sum_{i=\lfloor n\tau \rfloor + 1}^{n} H_q(X_i) \right|.
\]
By Theorem 1 the second factor is \( O_P(1) \) and by Assumption C3 the first one is \( o_P(1) \). Hence
\[
\frac{1}{d_{n,m}} \sum_{q=1}^{m-1} \int_R L_{n,q}(t,x) d\hat{F}_n(x) \overset{P}{\to} 0,
\]
as \( n \to \infty \). It remains to treat \( q = m \).
\[
\frac{1}{d_{n,m}} \int_R L_{n,m}(t,x) d\hat{F}_n(x) = \int_R \frac{J_m(x)}{m!} d\hat{F}_n(x) \frac{1}{d_{n,m}} \left( \sum_{i=1}^{\lfloor nt \rfloor} H_m(X_i) - \frac{nt}{n} \sum_{i=1}^{n} H_m(X_i) \right)
+ \int_R \frac{J_{m,n}(x) - J_m(x)}{(m!)^2} d\hat{F}_n(x) \frac{1}{d_{n,m}} \left( \sum_{i=\lfloor n\tau \rfloor + 1}^{\lfloor nt \rfloor} H_m(X_i) - \frac{nt}{n} \sum_{i=\lfloor n\tau \rfloor + 1}^{n} H_m(X_i) \right).
\]
The second summand of the right-hand side is negligible due to uniform convergence of the Hermite coefficient functions \( J_{m,n}(x) \). The first summand converges weakly to
\[
\int_R \frac{J_m(x)}{m!} dF(x) \tilde{Z}_m(t),
\]
which follows from the non-central limit Theorem of Taqqu (1975) and Dobrushin and Major (1979) (see also Theorem 1 in this thesis), the continuous mapping theorem and
\[
\int_R \frac{J_m(x)}{m!} d(\hat{F}_n(x) - F(x)) = -(m!)^{-1} \int_R (\hat{F}_n(x) - F(x)) dJ_m(x) \to 0.
\]
Here we have used (3.9) and the fact that $J_m(x)$ is of bounded variation, see (3.9).

3.4 Asymptotic relative efficiency

3.4.1 Definition and motivation

We have investigated four different change-point tests: Kolmogorov-Smirnov, Cramér-von Mises, CUSUM and Wilcoxon test. Consider a fixed alternative, in detail $Y_i = G_1(X_i)1_{\{i \leq \lfloor n \tau \rfloor\}} + G_2(X_i)1_{\{i > \lfloor n \tau \rfloor\}}$ with $EG_1(X_i) \neq EG_2(X_j)$. By the ergodic Theorem, all four test statistics will be of order $n/d_n m$ and the power of he tests will converge to 1. In other words, the procedures are consistent, which is kind of a minimal requirement for tests. However, if one wants to compare the efficiency of the tests this approach is to naive.

We have therefore considered local alternatives and derived the asymptotic distribution of the test statistics under these. Most important, the power of the tests against these alternatives does not converge to 1, but towards some probability $\beta \in (0, 1)$, called asymptotic power. We will now compare two different tests by the sample sizes, that are needed for each test to achieve the same asymptotic power. The (limit) ratio of the different sample sizes is then called (asymptotic) relative efficiency. This approach is due to Pitman (1948), for published articles see Noether (1950) and was formalized in Noether (1955).

In what follows we give a precise definition of the ARE in the very special context of our change-point setting. Of course it can be extended to all kinds of testing procedures.

**Definition 3.4.1.** Let $T_1$ and $T_2$ represent two change-point test procedures with a fixed level of significance $\alpha \in (0, 1)$. Consider the local alternatives

$$(G, G_{n_k}, \tau)$$ and a sample size $(n_k)_k$,

$$(G, \tilde{G}_{m_k}, \tau)$$ and a sample size $(m_k)_k$, such that $G_{n_k}(x) = \tilde{G}_{m_k}(x) = G_k(x)$ for all $k \geq 1$ and $x \in \mathbb{R}$. Let $\beta_1$ be the asymptotic power of the test $T_1$ against the local alternatives given by $(G, G_{n_k}, \tau, (n_k)_k)$ and $\beta_2$ be the asymptotic power of the test $T_2$ against the local alternatives given by $(G, \tilde{G}_{m_k}, \tau, (m_k)_k)$. If $\beta_1$ equals $\beta_2$, then the asymptotic relative efficiency (ARE) of the tests $T_1$ and $T_2$ is defined as

$$ARE(T_1, T_2) = \lim_{k \to \infty} \frac{m_k}{n_k}.$$

**Remark 3.4.2.** (i) There interpretation of this ratio is as follows: If $ARE(T_1, T_2) = b$, then $b$-times as much observations are needed for test $T_2$ than for test Test $T_1$ in order to achieve the same power $\beta$.
As there are other notions of relative efficiency, this quantity is sometimes called Pitman efficiency. Another type is the Bahadur efficiency, see van der Vaart (1998). Here one considers a fixed alternative \((G, G_2, \tau)\) and a level of significance \(\alpha_k \to 0\). If sample sizes \((n_k)_k\) and \((m_k)_k\) are such that the asymptotic powers of the tests coincide, the Bahadur efficiency is defined as \(\lim_{k \to \infty} m_k/n_k\). Wieand (1976) gives conditions under which the Pitman and the Bahadur efficiency coincide. However, he requires normal limits which is not the case for the Kolmogorov-Smirnov statistic.

(ii) Strictly speaking, the limit in Definition 3.4.1 may not always exist. In fact, it is not guaranteed that for every admissible power \(\beta\), a sequence of sample sizes \((n_k)_k\) exists such that the associated test has power \(\beta\). In what follows, we will obtain lower bounds for the ARE. We say that \(ARE(T_1, T_2) > b\), if there exists sequences \((m_k)_k\) and \((n_k)_k\) such that test \(T_1\) has asymptotic power larger than \(\beta\), \(T_2\) has asymptotic power below \(\beta\) and

\[
b = \lim_{k \to \infty} \frac{m_k}{n_k}.
\]

### 3.4.2 Change-point tests

It is the goal of this section to compare the Kolmogorov-Smirnov test and the Cramér-von Mises test to the CUSUM test in terms of asymptotic relative efficiency. Through the result in Dehling et al. (2017), which states the ARE of CUSUM and Wilcoxon test, we are able to compare the two newly investigated tests to the Wilcoxon test too. In detail, consider a mean-shift in a subordinated Gaussian process with \(EG^2(X_1) < \infty\) and Hermite rank \(m = 1\), then Theorem 4.2 in Dehling et al. (2012) states

\[
ARE(W, C) = \left(\frac{|c_1| \int f^2(x)dx}{|J_1(x)dF(x)|}\right)^{1/(1-H)}.
\]

By our Theorems 8 and 9 this can be extended to changes that are not necessarily mean-shifts. One obtains

\[
ARE(W, C) = \left(\frac{|c_1| \gamma}{|J_1(x)dF(x)| \Delta}\right)^{1/(1-H)},
\]

where \(\Delta = \lim_{n \to \infty} n/d_{n,1}(EG_n(X_1) - EG(X_1))\) and \(\gamma = \lim_{n \to \infty} n/d_{n,1}(P(G_n(X) \leq G(X')) - 1/2)\).

Such a sharp result will not be possible if we compare the Kolmogorv-Smirnov test to the CUSUM test, at least not for arbitrary local alternatives. The following difficulty arises. Consider a sample size \(n_k = b_1k\), then by Theorem 7 the asymptotic power of the KS test is

\[
\psi_1(b_1) = P\left(\sup_{x,t} \left| \frac{\tilde{Z}_m(t) + b_1^{-H} g(x) \psi_\tau(t)}{m!} \right| > q_{K_1,1-\alpha}\right).
\]
In contrast, consider the CUSUM test with a sample size \( m_k = b_2 k \). Its asymptotic power is

\[
\psi_2(b_2) = P \left( \sup_t \left| \frac{c_m(x)}{m!} \tilde{Z}_m(t) + b_2^{1-H} \Delta \psi_\tau(t) \right| > q_{C,1-\alpha} \right).
\]

Further, let \( \beta \in (0,1) \) be the power we want to achieve. The challenge is to find \( b_1, b_2 \), such that \( \psi_1(b_1) = \psi_2(b_2) = \beta \). In this case the ARE would equal \( b_1 / b_2 \). But little is known about the distribution of \( \sup_t |\tilde{B}_H(t) + f(t)| \), and even less if higher order Hermite processes are considered. Thus it seems hard to set the functions \( \psi_1(\cdot) \) and \( \psi_2(\cdot) \) in context to each other, which prevents a precise computation of the ARE in many cases.

Then again, there are special cases where this is possible and the ARE might be calculated or even bounded from below. Such is the case in the following two examples, which lead to interesting findings.

**Example 3.4.3** (Mean-shift in Gaussian data). Consider \( G(x) = x \) and \( G_n(x) = G(x) + \mu_n \), in other words a mean-shift in Gaussian data. As for the Hermite coefficient function, we get

\[
J_1(x) = -\phi(x),
\]

where \( \phi \) is the standard normal probability density. Thus, according to Corollary 7, the test statistic \( T_n \) converges towards

\[
\sup_{x \in \mathbb{R}} |\phi(x)| \sup_{t \in [0,1]} \left| \tilde{B}_H(t) + C \psi_\tau(t) \right| = (2\pi)^{-1/2} \sup_{t \in [0,1]} \left| \tilde{B}_H(t) + C \psi_\tau(t) \right|,
\]

whereas under the Null, that is we have a stationary standard Gaussian sequence, the limit distribution would be

\[
\sup_{x \in \mathbb{R}} |\phi(x)| \sup_{t \in [0,1]} \left| \tilde{B}_H(t) \right| = (2\pi)^{-1/2} \sup_{t \in [0,1]} \left| \tilde{B}_H(t) \right|.
\]

For the Cramér-von Mises statistic we obtain analogously the limit distributions

\[
\int \phi^3(x) dx \sup_{t \in [0,1]} \left| \tilde{B}_H(t) + C \psi_\tau(t) \right|^2 \quad \text{and} \quad \int \phi^3(x) dx \sup_{t \in [0,1]} \left| \tilde{B}_H(t) \right|^2
\]

under local alternative and hypothesis, respectively. Hence in this special case the CUSUM test, the Wilcoxon test (see Dehling et al. (2017) for each), the Kolmogorov-Smirnov test and the Cramér-von Mises test all have the same asymptotic power, namely

\[
P \left( \sup_{t \in [0,1]} |\tilde{B}_H(t) + C \psi_\tau(t)| > q_{1-\alpha,H} \right), \quad (3.1)
\]

where \( q_{1-\alpha,H} \) is the \((1-\alpha)\)-quantile of the maximum of a fractional Brownian bridge \( \sup_{t \in [0,1]} |\tilde{B}_H(t)| \). As a direct consequence, one gets that the ARE of the four tests is 1. This result is quite surprising, keeping in mind that CUSUM and Wilcoxon tests are designed to detect level-shifts, while our tests have power against all kinds of distributional changes.
Example 3.4.4 (Combined change in mean and variance). Let $G(x) = x$ and $G_k(x) = \sigma_k x + \mu_k$, that is a combined change of mean and variance in Gaussian data. If further $\mu_k k/d_k \rightarrow C_1 > 0$ and $(1 - 1/\sigma_k) k/d_k \rightarrow C_2 > 0$, then by example 3.2.4 the empirical bridge-type process converges to (for the sample size $k \rightarrow \infty$)

$$\phi(x)\tilde{B}_H(t) + \phi(x)(C_1 + C_2 x)\psi_\tau(t), \quad x \in \mathbb{R}, \; t \in [0, 1].$$

We now consider slightly modified Cramér-von Mises and CUSUM tests, in detail, instead of $[0, 1]$ the supremum is taken over $[\kappa_1, \kappa_2]$ for arbitrary $\kappa_1 \in (0, 1/2)$ and $\kappa_2 \in (1/2, 1)$.

The asymptotic distribution of the CUSUM test has been derived in Dehling et al. (2017), but only in the case of a mean-shift with constant variance. However, we might apply our Theorem 8 and conclude that the CUSUM statistic converges under this type of local alternatives to

$$\sup_{t \in [\kappa_1, \kappa_2]} \left| \int \phi(x) \left( \tilde{B}_H(t) + (C_1 + C_2 x)\psi_\tau(t) \right) dx \right| = \sup_{t \in [\kappa_1, \kappa_2]} \left| \tilde{B}_H(t) + C_1\psi_\tau(t) \right|.$$ 

Note that this is the same limit as under a mean-shift with constant variance and thus, too, the asymptotic power is the same as in Example 3.4.3.

The limiting distribution of the Cramér-von Mises statistic is given by

$$Z^2 = \sup_{t \in [\kappa_1, \kappa_2]} \int \phi^3(x) \left( \tilde{B}_H(t) + (C_1 + C_2 x)\psi_\tau(t) \right)^2 dx = \sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x) dx \left( \tilde{B}_H(t) + C_1\psi_\tau(t) \right)^2 + C_2^2 \psi^2_\tau(t) \int \phi^3(x) x^2 dx \right\},$$

and for its asymptotic power we obtain

$$P \left( Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) dx \right) = P \left( Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) dx , \sup_{t \in [\kappa_1, \kappa_2]} \{ \tilde{B}_H(t) \} > q_{1-\alpha,H} \right)$$

$$+ P \left( Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) dx , \sup_{t \in [\kappa_1, \kappa_2]} \{ \tilde{B}_H(t) \} \leq q_{1-\alpha,H} \right).$$
First assume \( \sup_t \{ \tilde{B}_H(t) \} \leq q = q_{1-\alpha,H} \) and consider \( C_1^* \), given by
\[
C_1^* = f^*(C_1, C_2, q, \tau, \kappa_1, \kappa_2) = \min_{t \in [\kappa_1, \kappa_2]} \left\{ \sqrt{q^2 + 2qC_1\psi_\tau(t)} + \left( C_1^2 + C_2^2 \left( \int \psi^3(x)x^2dx / \int \phi^3(x)dx \right) \right) \psi_\tau(t) - q \right\}.
\]

Now \( C_1^* \) is constructed in a way, such that (here we used the restriction to \([\kappa_1, \kappa_2]\) ) \( C_1^* > C_1 \) and for all \( \omega \in \Omega \) with \( \sup_t \tilde{B}_H(t; \omega) \leq q \)
\[
Z^2 = \sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x)dx \left( \tilde{B}_H(t) + C_1\psi_\tau(t) \right)^2 + C_2\psi_\tau^2(t) \int \phi^3(x)x^2dx \right\} > \sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x)dx \left( \tilde{B}_H(t) + C_1^*\psi_\tau(t) \right)^2 \right\}.
\]

If, on the other hand, \( \sup_t \{ \tilde{B}_H(t) \} > q_{1-\alpha,H} \), then (because \( C_1 > 0 \) ) automatically \( Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x)dx \). Combining these two findings with (3.2) we can bound the asymptotic power from below by
\[
P \left( Z^2 > q_{1-\alpha,H}^2 \int \phi^3(x) dx \right) = P \left( \sup_{t \in [\kappa_1, \kappa_2]} \int \phi^3(x) \left( \tilde{B}_H(t) + C_1^*\psi_\tau(t) \right)^2 dx \right) > q_{1-\alpha,H}^2 \int \phi^3(x)dx \right)
\geq P \left( \sup_{t \in [\kappa_1, \kappa_2]} \left| \tilde{B}_H(t) + C_1^*\psi_\tau(t) \right| > q_{1-\alpha,H} \right), \quad \text{(3.3)}
\]
for \( C_1^* > C_1 \).

Now we are ready to compute the ARE. To this end we chose different sample sizes for both tests. In detail, \( (n_k)_k \) for the Cramér-von Mises test and \( (m_k)_k \) for the CUSUM test. Moreover, the local alternatives are such that \( G_{n_k}^{CvM}(x) = G_{m_k}^{CUSUM}(x) = G_k(x) \) for all \( x \in \mathbb{R} \). Consequently,
\[
\mu_{n_k}^{(1)} = \mu_{m_k}^{(2)} = \mu_k \quad \text{and} \quad \sigma_{n_k}^{(1)} = \sigma_{m_k}^{(2)} = \sigma_k.
\]

For the CUSUM test, in order to achieve at least asymptotic power \( \beta \), its limit distribution
has to satisfy

\[ P \left( \sup_{t \in [0,1]} |\hat{B}_H(t) + C_1^* \psi_\tau(t)| > q_{\alpha,H} \right) \geq \beta. \]

In other words, \( C_1^* = \pi^{-1}(\beta) \), where

\[ \pi(C^*) = P \left( \sup_{t \in [0,1]} |\hat{B}_H(t) + C_1^* \psi_\tau(t)| > q_{\alpha,H} \right) \]

and \( \pi^{-1} \) is the generalized inverse. Therefore, the sample size of the CUSUM test has to be chosen such that

\[ C_1^* = \lim_{k \to \infty} \frac{m_k}{d_{m_k}} \mu^{(2)}_{m_k} = \lim_{k \to \infty} \frac{m_k}{d_{m_k}} \mu_k. \]  

We also obtain (as \( \mu_k \) and \( (1 - 1/\sigma_k) \) are of the same order)

\[ \lim_{k \to \infty} \frac{m_k}{d_{m_k}} \left( 1 - \frac{1}{\sigma_{m_k}^{(2)}} \right) = \lim_{k \to \infty} \frac{m_k}{d_{m_k}} \left( 1 - \frac{1}{\sigma_k} \right) = C_2^*, \]

for some \( C_2^* > 0 \).

Next we will select the sample size for the Cramér-von Mises test such that its asymptotic power might be bounded from below as in (3.3) (and therefore by \( \beta \)). To this end choose \( C_1 > 0 \), such that

\[ \tilde{f}(C_1) = f(C_1, C_1 \frac{C_2^*}{C_1}, \mu_k), \]

The function \( \tilde{f} : [0, \infty) \to [0, \infty) \) is monotone increasing, surjective and continuous (as the minimum is attained either in \( \kappa_1 \) or \( \kappa_2 \)). Therefore, such an \( C_1 \) always exists. By construction of the function \( f \) it follows that \( C_1 < C_1^* \). Now let the sample size of the Cramér-von Mises test satisfy

\[ C_1 = \lim_{k \to \infty} \frac{n_k}{d_{n_k}} \mu^{(1)}_{n_k} = \lim_{k \to \infty} \frac{n_k}{d_{n_k}} \mu_k. \]  

Moreover, we observe

\[ C_2 = \lim_{k \to \infty} n_k/d_{n_k} (1 - 1/\sigma_k) = \lim_{k \to \infty} (n_k/d_{n_k} \mu_k) (m_k/d_{m_k} \mu_k)^{-1} (m_k/d_{m_k} (1 - 1/\sigma_k)) = C_1 C_2^*/C_1^*. \]
Thus,
\[ f(C_1, C_2, q, \kappa_1, \kappa_2) = C_1^*, \]
and we observe for the asymptotic power of the Cramér-von Mises test
\[
P\left( \sup_{t \in [\kappa_1, \kappa_2]} \left\{ \int \phi^3(x) dx \left( \hat{B}_H(t) + C_1 \psi_\tau(t) \right)^2 + C_2 \psi_\tau^2(t) \int \phi^3(x)x^2 dx \right\} > q_{1-a,H} \int \phi^3(x)dx \right) 
\geq P\left( \sup_{t \in [\kappa_1, \kappa_2]} \left| \hat{B}_H(t) + C_1^* \psi_\tau(t) \right| > q_{1-a,H} \right) 
\geq \beta.
\]
So both tests have (at least) asymptotic power $\beta$ against the local alternatives $(G, G_k, \tau)$. Finally,
\[
\left( \frac{m_k}{n_k} \right)^{1-H} = \frac{m_k}{d_{mk}} \mu_k \frac{d_{mk}}{n_k} \mu_k L^{(1/2)}(m_k) \frac{L^{(1/2)}(n_k)}{C_1^*} \to C_1^* > 1,
\]
by construction of the sample sizes and the definition of slowly varying functions. Consequently
\[
ARE(CvM, CUSUM) = \left( \frac{C_1^*}{C_1} \right)^{1/(1-H)} > 1.
\]
In other words, the Cramér-von Mises test is asymptotically more efficient, no matter how small the additional variance-change is.

### 3.4.3 Two-sample tests

As illustrated in the previous section it is quite hard to calculate the ARE of different change-point tests. Therefore, we make two restrictions here: First assume that the parameter $\tau \in (0, 1)$ is known. This refers to a situation where a candidate for a possible change-point is at hand. For example, one might think of flood data before and after a reservoir dam is build. In this situation the tests become quite similar to classical two-sample problems. The second restriction is that we study only sequences with Hermite rank $m = 1$. Thereby the random variable in the limit is Gaussian, making the calculations much more simple.

The test statistics are then given as follows: The Kolmogorov-Smirnov statistic
\[
T_n = \frac{n}{d_n} \tau(1-\tau) \sup_{x \in \mathbb{R}} \left( \hat{F}_{1:|n\tau|}(x) - \hat{F}_{1+:|n\tau|:n}(x) \right),
\]
and the CUSUM statistic (which now becomes the test statistic of the Gauss test)

\[ G_n = \frac{n}{d_n^\tau} (1 - \tau) \left( \bar{Y}_{1:|n\tau|} - \bar{Y}_{1+|n\tau|:n} \right). \]

Note that both statistics refer to testing one-sided alternatives. This too, makes the calculations in the proof more straightforward.

Asymptotic efficiency of the Kolmogorov-Smirnov test has been investigated in a series of papers fifty years ago, see Ramachandramurty (1966), Capon (1965) and Yu (1971). However, all results are for independent data, while we deal with long-range dependence. Due to the semi-degenerate structure of the limiting process, our findings differ quite remarkably from existing results.

The next theorem gives the asymptotic distribution of \( T_n \) and \( G_n \), where the assumption are basically those of Sections 3.2 and 3.3.

**Theorem 10.** (i) Under the local Alternatives \( A^*_n \) with \( m = 1 \) we have

\[ T_n \xrightarrow{D} \sup_{x \in \mathbb{R}} \{ J_1(x)Z_{\tau,H} + \tau(1 - \tau)g^*(x) \}, \]

where \( Z_{\tau,H} \) is a normally distributed r.v. with zero-mean and variance \( \sigma^2_{\tau,H} = \tau^2H - \tau(1 + \tau^2 - (1 - \tau)^2H) + \tau^2. \)

(ii) Let Assumption B hold with \( m = 1 \). Then

\[ G_n \xrightarrow{D} c_1Z_{\tau,H} + \tau(1 - \tau)\Delta^*. \]

**Remark 3.4.5.** The normal random variable \( Z_{\tau,H} \) is given through

\[ Z_{\tau,H} = B_H(\tau) - \tau B_H(1), \]

for a fractional Brownian motion \( (B_H(t))_t \). The calculation of the variance then follows immediately from the structure of the covariance kernel of fractional Brownian motion, see Definition 2.1.8.

Parts (i) and (ii) of Theorem 10 might be proven along the same lines as Corollary 3.2.8 and Theorem 8, respectively. In fact, it is much simpler due to the restriction \( m = 1 \) and the fact that the two-sample statistics \( T_n \) and \( G_n \) are simplified versions of their change-point counterparts.

Note that the theorem only gives the asymptotic distribution under the alternative. Under the hypothesis, the two test statistics converge to \( \sup_x \{ J_1(x)Z_{\tau,H} \} \) and \( c_1Z_{\tau,H} \), respectively. If critical values are given as quantiles of these limits, we obtain the asymptotic power of the
Kolmogorov-Smirnov test

\[ \beta_1 = P \left( \sup_{x \in \mathbb{R}} \{ J_1(x)Z_{\tau,H} + \tau(1-\tau)g(x) \} > \sup_x |J_1(x)|Q_{\tau,H}(1-\alpha) \right), \]

where \( Q_{\tau,H}(1-\alpha) \) denotes the \((1-\alpha)\)-quantile of \( Z_{\tau,H} \). Moreover, the Gauss test has asymptotic power

\[ \beta_2 = P \left( c_1Z_{\tau,H} + \tau(1-\tau)b > |c_1|Q_{\tau,H}(1-\alpha) \right). \]

Next we give the main result of this section. It states the ARE for other distributions than in the previous section and more complex structural changes.

**Theorem 11.** Let the conditions of Theorem 10 hold. Further, let \( b > 0 \) be such that

\[ b^{1-H} = \sup_{x, J_1(x) \neq 0} \left\{ \frac{g^*(x)|c_1|}{|J_1(x)|\Delta^*} - \left( \frac{\sup_x |J_1(x)|}{|J_1(x)|} - 1 \right) \Phi^{-1}(1-\alpha) + 1 \right\}, \]

with \( \Phi^{-1} \) being the quantile function of a standard normal law and

\[ \frac{g^*(x)}{\Delta^*} = \lim_{k \to \infty} \frac{P(G(X_1) \leq x) - P(G_k(X_1) \leq x)}{E_G(X_1) - E_G(X_1)}, \]

uniformly in \( x \). Then

\[ \text{ARE}(KS,G) \geq b. \]

**Remark 3.4.6.** (i) Unlike in the case of a mean-shift in Gaussian data, the ARE depends on nominal size \( \alpha \) and the aimed asymptotic power \( \beta \). It increases with \( \beta \) and decreases with \( \alpha \), what was also noted for independent data. Moreover, for \( \beta \to 1 \) we obtain the lower bound

\[ b^{1-H} = \sup_{x, J_1(x) \neq 0} \frac{g^*(x)|c_1|}{\Delta^*|J_1(x)|}, \]

which does not depend on \( \alpha \) and might be infinity. For \( \alpha \to 0 \), one gets

\[ b^{1-H} = \frac{g^*(x_0)|c_1|}{\Delta^*|J_1(x_0)|}, \]

where \( x_0 \) is such that \( J_1(x_0) = \sup_{x \in \mathbb{R}} |J_1(x)| \).

(ii) One also notes that an increase of the serial dependence (an increase of \( H \)) supports the test that is already more efficient. Finally, unlike in the case of independent observations, our lower bound does not depend on the ratio of observations before and after the change (which is a function of \( \tau \)).
(iii) The function $\Psi: (0, \infty) \to (0, \infty)$ given by

$$
\Psi(y) = \sup_{x, J_1(x) \neq 0} \left\{ \frac{g^*(x)|c_1|}{|J_1(x)|\Delta^*} - \left( \frac{\sup_x |J_1(x)|}{|J_1(x)|} - 1 \right) \frac{\Phi^{-1}(1 - \alpha)}{\Phi^{-1}(\beta) + 1} y \right\}
$$

is monotone decreasing, hence there is at most one fix point. However, whether $\Psi$ is continuous or not depends on the interplay of the functions $g^*(x)$ and $J_1(x)$. Thus it is not clear if there is a fix point at all. For such cases we can state the following modified version of Theorem 11 (which can be proved along the same lines): If $b_0 > 0$ then

$$b_0^{1-H} \leq \Psi(b_0^{1-H})$$

implies

$$ARE(KS, G) \geq b_0.$$

**Example 3.4.7.** Consider a mean-shift in a long memory sequence with uniform $(U[0,1])$ marginals, modeled via

$$G(X_i) = \Phi(X_i) \quad \text{and} \quad G_n(X_i) = \Phi(X_i) + \mu_n,$$

with $\mu_n \sim d_n/n$. Then

$$J_1(x) = -\phi(\Phi^{-1}(x))1_{(0,1)}(x) \quad \text{and} \quad g^*(x)/\Delta^* = 1_{(0,1)}(x).$$

Consequently,

$$\Psi(y) = \sup_{x \in (0,1)} \left\{ \frac{1}{\phi(\Phi^{-1}(x))} \left( |c_1| - \frac{1}{\sqrt{2\pi}} \Phi^{-1}(1 - \alpha) \right) + \frac{\Phi^{-1}(1 - \alpha)}{\Phi^{-1}(\beta) + 1} y \right\}
$$

$$= \begin{cases} 
\infty & \text{for } y < |c_1|\sqrt{2\pi} \frac{\Phi^{-1}(\beta)+1}{\Phi^{-1}(1-\alpha)}, \\
|c_1|\sqrt{2\pi} & \text{for } y \geq |c_1|\sqrt{2\pi} \frac{\Phi^{-1}(\beta)+1}{\Phi^{-1}(1-\alpha)}.
\end{cases}
$$

Therefore, we have to distinguish two cases in order to compute the ARE. Let first $\Phi^{-1}(\beta)+1/\Phi^{-1}(1-\alpha) < 1$. Then

$$\Psi \left( |c_1|\sqrt{2\pi} \right) = |c_1|\sqrt{2\pi},$$

and hence

$$ARE(KS, G) \geq \left( |c_1|\sqrt{2\pi} \right)^{1/(1-H)}.$$
Now let $\Phi^{-1}(\beta)+1/\Phi^{-1}(1-\alpha) \geq 1$ and note that Remark 3.4.6 (iii) applies, that is $\Psi(y)$ has no fix point. However for all $y < |c_1|\sqrt{2\pi}(\Phi^{-1}(\beta)+1)/(\Phi^{-1}(1-\alpha))$ one has

$$y \leq \Psi(y) = \infty$$

and thus

$$ARE(KS, G) \geq \left( |c_1|\sqrt{2\pi}\Phi^{-1}(\beta)+1/\Phi^{-1}(1-\alpha) \right)^{1/(1-H)}.$$

Note that $|c_1|\sqrt{2\pi} \approx 0.7063982$, thus it depends on $\beta$ and $\alpha$ if the Kolmogorov-Smirnov test or the Gauss test is more efficient.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{ARE_graph.png}
\caption{Graphical evaluation of the ARE.}
\end{figure}

\textbf{Proof of Theorem 11.} Let us first consider the asymptotic power of the Kolmogorov-Smirnov test. Assume that the limit distribution of the test statistic is given by

$$\sup_{x \in \mathbb{R}} \{J_1(x)Z_{\tau,H} + \tau(1-\tau)g(x)\},$$

for some $g(x) \neq 0$. The asymptotic power then can be bounded from below in the following
However, this is just the power of the Gauss test when its limit is given by $c_1 Z_{\tau,H} + \tau(1-\tau)\Delta$ for all $x \in \mathbb{R}$ with $J_1(x) \neq 0$. We have also used the symmetry of $Z_{H,\tau}$. Thus

$$
P\left( \sup_{x \in \mathbb{R}} \{ J_1(x)Z_{\tau,H} + \tau(1-\tau)g(x) \} > \sup_{x \in \mathbb{R}}|J_1(x)|Q_{\tau,H}(1-\alpha) \right)
\geq P\left( \sup_{x \in \mathbb{R}} \{ J_1(x)Z_{\tau,H} + \tau(1-\tau)\frac{g(x)}{|J_1(x)|} \} > \frac{|c_1| \sup_{x \in \mathbb{R}}|J_1(x)|}{|J_1(x)|} Q_{\tau,H}(1-\alpha) \right)
= P\left( |c_1|Z_{\tau,H} + |c_1|\tau(1-\tau) \sup_{x,J(x)\neq 0} \left\{ \frac{g(x)}{|J_1(x)|} + \left(1 - \frac{\sup_{x \in \mathbb{R}}|J_1(x)|}{|J_1(x)|} \right) \frac{Q_{\tau,H}(1-\alpha)}{\tau(1-\tau)} \right\} > |c_1|Q_{\tau,H}(1-\alpha) \right),
$$

for all $x \in \mathbb{R}$ with $J_1(x) \neq 0$. However, this is just the power of the Gauss test when its limit is given by $c_1 Z_{\tau,H} + \tau(1-\tau)\Delta$ with

$$
\Delta = |c_1| \sup_{x,J(x)\neq 0} \left\{ \frac{g(x)}{|J_1(x)|} + \left(1 - \frac{\sup_{x \in \mathbb{R}}|J_1(x)|}{|J_1(x)|} \right) \frac{Q_{\tau,H}(1-\alpha)}{\tau(1-\tau)} \right\}. 
$$

Assume that both tests are applied for a sample of size $k$. Then, by Theorem 10,

$$
T_k \xrightarrow{D} \sup_{x \in \mathbb{R}} \{ J_1(x)Z_{\tau,H} + \tau(1-\tau)g^*(x) \},
$$

and

$$
G_k \xrightarrow{D} c_1 Z_{\tau,H} + \tau(1-\tau)\Delta^*,
$$

with

$$
g^*(x) = \lim_{k \to \infty} \frac{k}{d_k} (F(x) - F_{(k)}(x)) \quad \text{and} \quad \Delta^* = \lim_{k \to \infty} \frac{k}{d_k} (EG(X_1) - EG_k(X_1)).
$$

Next, consider different sample sizes $(n_k)_k$ for the Kolmogorov-Smirnov test and $(m_k)_k$ for the
Asymptotic relative efficiency

Gauss test, given by

\[ n_k = \lambda_1 k \quad \text{and} \quad m_k = \lambda_2 k, \]

with \( \lambda_1, \lambda_2 \in (0, \infty) \). However, both tests are still considered under the same structural change given by \((G, G_k, \tau)\). Note that Theorem 10 still holds under this circumstances with

\[ g(x) = \lim_{k \to \infty} \frac{n_k}{d_{n_k}} \left( F(x) - F_{(k)}(x) \right) \]

\[ = \lim_{k \to \infty} \frac{(\lambda_1 k)^{1-H} L^{1/2}(k)}{k^{1-H} L^{1/2}(\lambda_1 k)} \frac{k}{d_k} \left( F(x) - F_{(k)}(x) \right) \]

\[ = \lambda_1^{1-H} g^*(x), \]

uniformly in \( x \). In the last line we have used that \( L \) is slowly varying at infinity. By the same argument we obtain the limit distribution of the Gauss test with

\[ \Delta = \lambda_2^{1-H} \Delta^*. \tag{3.8} \]

Now fix \( \alpha, \tau \) and \( H \) and let \( \beta \in (\alpha, 1) \) be any admissible asymptotic power. To obtain such an asymptotic power with the Gauss test, the \( \Delta \) from (3.8) has to satisfy

\[ \beta = P \left( c_1 Z_{\tau,H} + \tau(1-\tau) \Delta > |c_1| Q_{\tau,H}(1-\alpha) \right). \tag{3.9} \]

In other words, \( \Delta \) has to be the \( \beta \)-quantile of a normally distributed random variable with mean \( Q_{\tau,H}(1-\alpha)|c_1|/(\tau(1-\tau)) \) and variance \( c_1^2 \sigma_{\tau,H}^2/(\tau(1-\tau))^2 \), which will be denoted by \( Q_{\alpha,\tau,H}(\beta) \).

To this end define

\[ m_k = k \left( \frac{\Delta}{\Delta^*} \right)^{1/(1-H)} = k \left( \frac{Q_{\alpha,\tau,H}(\beta)}{\Delta^*} \right)^{1/(1-H)}, \]

then (3.9) obviously holds.

Now let \( b \in (0, \infty) \) such that \( b^{1-H} \) solves the fix point problem (3.6) and chose

\[ n_k = \frac{m_k}{b} = \left( \frac{Q_{\alpha,\tau,H}(\beta)}{\Delta^*} \right)^{1/(1-H)} \frac{1}{b} k. \tag{3.10} \]

By definition of \((m_k)_k\) it follows that the Gauss test has power \( \beta \). Thus in order to finish the proof we have show that the Kolmogorov-Smirnov test has also at least power \( \beta \). First obtain
its limiting distribution, given by

\[ T_{n_k} \xrightarrow{D} \sup_{x \in \mathbb{R}} \left\{ J_1(x) Z_{\tau,H} + \tau(1 - \tau) \frac{Q_{\alpha,\tau,H}(\beta)}{\Delta^*} \frac{1}{b^{1-H}g^*(x)} \right\}. \]

Subsequently, we can bound the asymptotic power, using (3.7), by

\[ P \left( |c_1| Z_{\tau,H} + \tau(1 - \tau) \tilde{\Delta} > |c_1| Q_{\tau,H}(1 - \alpha) \right), \]

with

\[
\tilde{\Delta} = |c_1| \sup_{x, J(x) \neq 0} \left\{ \frac{g^*(x)}{|J_1(x)|} \frac{Q_{\alpha,\tau,H}(\beta)}{\Delta^*} \frac{1}{b^{1-H}} + \left( 1 - \frac{\sup_x |J_1(x)|}{|J_1(x)|} \right) \frac{Q_{\tau,H}(1 - \alpha)}{\tau(1 - \tau)} \right\}
\]

\[
= \frac{Q_{\alpha,\tau,H}(\beta)}{b^{1-H}} \sup_{x, J(x) \neq 0} \left\{ \frac{g^*(x)|c_1|}{\Delta^*|J_1(x)|} + \left( 1 - \frac{\sup_x |J_1(x)|}{|J_1(x)|} \right) \frac{Q_{\tau,H}(1 - \alpha)}{Q_{\alpha,\tau,H}(\beta)} \frac{|c_1|}{b^{1-H}} \right\}
\]

\[
= \frac{Q_{\alpha,\tau,H}(\beta)}{b^{1-H}} \sup_{x, J(x) \neq 0} \left\{ \frac{g^*(x)|c_1|}{\Delta^*|J_1(x)|} + \left( 1 - \frac{\sup_x |J_1(x)|}{|J_1(x)|} \right) \Phi^{-1}(1 - \alpha) \frac{\Phi^{-1}(\beta)}{\phi^{-1}(\beta) + 1} b^{1-H} \right\}
\]

Here we have used

\[
\frac{Q_{\tau,H}(1 - \alpha)}{Q_{\alpha,\tau,H}(\beta)} \frac{|c_1|}{\tau(1 - \tau)} = \frac{\Phi^{-1}(1 - \alpha)}{\Phi^{-1}(\beta) + 1}
\]

and the fact that \( b^{1-H} \) is a fix point. We have just shown that for a sample size \( n_k \) (given through (3.10)) the asymptotic power of the Kolmogorov-Smirnov test is larger or equal than \( \beta \). If \( (n_k^*)_k \) is a sequence such that the asymptotic power equals \( \beta \), this implies \( n_k^* \leq n_k \). Hence \( \lim_{k \to \infty} (m_k/n_k^*) \geq b \), and this finishes the proof.

distribution. Instead, we simulate \( J = 1000 \) Gaussian time series
\[
X_{j,1}, \ldots X_{j,n}, \quad j = 1, \ldots J,
\]
with Hurst coefficient \( H \). In the simulation study, we will use fractional Gaussian noise for this sequences. Subsequently, a Cramér-von Mises statistic is calculated for each of the \( J = 1000 \) Gaussian series, in detail
\[
S_{n,j} = \max_{1 \leq k < n} \int_{x \in \mathbb{R}} \left( \sum_{i=1}^{k} 1\{x_{j,i} \leq x\} - \frac{k}{n} \sum_{i=1}^{n} 1\{x_{j,i} \leq x\} \right)^2 \hat{F}_{n,j}(x) \quad j = 1, \ldots J.
\]
We then use the empirical quantiles of \( \{S_{n,j}\}_{j=1}^{J} \) as critical values. The Cramér-von Mises statistic is invariant under monotone transformations of the data (as is the Kolmogorov-Smirnov statistic). Hence the critical values are valid if our observations are monotone transformations of Gaussian data. We note that this is a strong assumption and that an accurate approximation of the empirical process for general long-range dependent data is an issue of future research. The CUSUM statistic is not invariant under monotone transformations. Therefore, the Wilcoxon change-point test is considered additionally.

In the first part of the simulation study, we treat realizations of a Gaussian process \( X_1, \ldots, X_n \), given by fractional Gaussian noise, see Example 2.1.10. For the implementation we have used the function \texttt{fgnSim} from the R-package \texttt{fArma}. Eventually, a change is added by \( Y_i = X_i + \mu 1\{i > \lfloor n\tau \rfloor\} \) and the three mentioned change-point tests are applied to \( Y_1, \ldots, Y_n \).

If the Hurst-coefficient is assumed to be known, the empirical size of the tests naturally equals the nominal one, due to the construction of the critical values. The empirical power of Cramér-von Mises (denoted by \( S_n \)), Wilcoxon (denoted by \( W_n \)) and CUSUM test (denoted by \( C_n \)) is displayed in Table 3.1. If the change occurs in the middle of the observation period, the three tests are showing almost exactly the same performance, which matches the theoretical results. For early changes (after 20% of the observations) the CUSUM test is slightly more accurate than the other tests. Depending on sample size and strength of dependence, either the Cramér-von Mises or the Wilcoxon test might be second best.

### 3.5.2 Unknown Hurst coefficient

In applications, true Hurst coefficient \( H \) is unknown, and in the following we will consider two different estimators. The first is the local Whittle estimator (denoted by \( \hat{H} \)) with bandwidth parameter \( m = [n^{2/3}] \), see Künsch (1987). However, if there is actually a change in the data, the local Whittle estimator is known to be biased. For the second estimator we therefore divide
Table 3.1: Empirical power, $H$ assumed to be known, size of level shift $\mu = 1$, relative change positions $\tau = 0.2$ and $\tau = 0.5$.

<table>
<thead>
<tr>
<th></th>
<th>Relative change position $\tau = 0.2$</th>
<th></th>
<th>Relative change position $\tau = 0.5$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$H = 0.6$</td>
<td>$H = 0.7$</td>
<td>$H = 0.8$</td>
</tr>
<tr>
<td>$n$</td>
<td>50 100 250 400</td>
<td>50 100 250 400</td>
<td>50 100 250 400</td>
</tr>
<tr>
<td>S</td>
<td>0.196 0.566 0.854 0.970</td>
<td>0.158 0.215 0.547 0.689</td>
<td>0.101 0.221 0.241 0.350</td>
</tr>
<tr>
<td>W</td>
<td>0.263 0.525 0.910 0.983</td>
<td>0.201 0.233 0.501 0.636</td>
<td>0.089 0.156 0.264 0.383</td>
</tr>
<tr>
<td>C</td>
<td>0.288 0.666 0.933 0.986</td>
<td>0.276 0.284 0.555 0.769</td>
<td>0.171 0.234 0.348 0.349</td>
</tr>
<tr>
<td></td>
<td>$H = 0.8$</td>
<td>$H = 0.9$</td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td>50 100 250 400</td>
<td>50 100 250 400</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>0.664 0.919 1.000 1.000</td>
<td>0.524 0.682 0.925 0.970</td>
<td>0.641 0.504 0.655 0.830</td>
</tr>
<tr>
<td>W</td>
<td>0.621 0.930 0.998 1.000</td>
<td>0.513 0.742 0.906 0.967</td>
<td>0.374 0.485 0.674 0.770</td>
</tr>
<tr>
<td>C</td>
<td>0.733 0.918 0.997 0.999</td>
<td>0.599 0.717 0.919 0.960</td>
<td>0.400 0.553 0.673 0.766</td>
</tr>
</tbody>
</table>
Numerical results

Table 3.2: Empirical size, estimated Hurst coefficient.

<table>
<thead>
<tr>
<th></th>
<th>$H = 0.6$</th>
<th></th>
<th>$H = 0.7$</th>
<th></th>
<th>$H = 0.8$</th>
<th></th>
<th>$H = 0.9$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>400</td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>400</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{\hat{H}}$</td>
<td>0.067</td>
<td>0.088</td>
<td>0.065</td>
<td>0.058</td>
<td>0.080</td>
<td>0.076</td>
<td>0.063</td>
<td>0.043</td>
</tr>
<tr>
<td>$S_{\hat{H}_k}$</td>
<td>0.082</td>
<td>0.105</td>
<td>0.085</td>
<td>0.083</td>
<td>0.122</td>
<td>0.143</td>
<td>0.107</td>
<td>0.081</td>
</tr>
<tr>
<td>$W_{\hat{H}}$</td>
<td>0.067</td>
<td>0.058</td>
<td>0.054</td>
<td>0.057</td>
<td>0.081</td>
<td>0.074</td>
<td>0.061</td>
<td>0.035</td>
</tr>
<tr>
<td>$W_{\hat{H}_k}$</td>
<td>0.070</td>
<td>0.100</td>
<td>0.102</td>
<td>0.078</td>
<td>0.122</td>
<td>0.118</td>
<td>0.100</td>
<td>0.081</td>
</tr>
<tr>
<td>$C_{\hat{H}}$</td>
<td>0.071</td>
<td>0.072</td>
<td>0.064</td>
<td>0.046</td>
<td>0.111</td>
<td>0.080</td>
<td>0.056</td>
<td>0.056</td>
</tr>
<tr>
<td>$C_{\hat{H}_k}$</td>
<td>0.085</td>
<td>0.090</td>
<td>0.103</td>
<td>0.081</td>
<td>0.116</td>
<td>0.127</td>
<td>0.095</td>
<td>0.078</td>
</tr>
</tbody>
</table>

the observations into two subsamples

$$X_1, \ldots, X_k \quad \text{and} \quad X_{k+1}, \ldots, X_n$$

and estimate $H$ on each set, using again the local Whittle estimator. Finally the new estimator is given by $\hat{H}_k = \hat{k}/n\hat{H}_1 + (n-\hat{k})/n\hat{H}_2$. Here $\hat{k}$ is the natural change-point estimator, associated with each test. For example, in case of the Cramér-von Mises test we use

$$\hat{k} = \min \left\{ 1 \leq k \leq n - 1 \mid U_{k,n} = \max_{1 \leq k \leq n-1} U_{k,n} \right\},$$

where

$$U_{k,n} = \int_{x \in \mathbb{R}} \left( \sum_{i=1}^{k} 1_{\{X_i \leq x\}} - \frac{k}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}} \right)^2 d\hat{F}_n(x).$$

Consistency of this estimator was shown in Hariz et al. (2009). Horvath and Kokoszka (1997) verified consistency for the analogous CUSUM-based estimator.

Empirical size and empirical power of the tests under unknown $H$ are displayed in Tables 3.2 and 3.3. Let us first compare the impact of the different estimators $\hat{H}$ and $\hat{H}_k$ on the finite sample performance of the Cramér-von Mises test. If we use the classical local Whittle estimator, the empirical size of the test is quite accurate and even matches the nominal size for
Table 3.3: Empirical Power, estimated Hurst coefficient, size of level shift $\mu = 1$, relative change positions $\tau = 0.2$ and $\tau = 0.5$.

<table>
<thead>
<tr>
<th>$H = 0.6$</th>
<th>$H = 0.7$</th>
<th>$H = 0.8$</th>
<th>$H = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td></td>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>250</td>
<td>400</td>
</tr>
<tr>
<td>$S_H$</td>
<td>0.178</td>
<td>0.263</td>
<td>0.597</td>
</tr>
<tr>
<td>$S_{H_k}$</td>
<td>0.248</td>
<td>0.491</td>
<td>0.834</td>
</tr>
<tr>
<td>$W_H$</td>
<td>0.190</td>
<td>0.298</td>
<td>0.625</td>
</tr>
<tr>
<td>$W_{H_k}$</td>
<td>0.291</td>
<td>0.532</td>
<td>0.850</td>
</tr>
<tr>
<td>$C_H$</td>
<td>0.289</td>
<td>0.413</td>
<td>0.706</td>
</tr>
<tr>
<td>$C_{H_k}$</td>
<td>0.378</td>
<td>0.596</td>
<td>0.874</td>
</tr>
<tr>
<td>$H = 0.8$</td>
<td></td>
<td>$H = 0.9$</td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>250</td>
<td>400</td>
</tr>
<tr>
<td>$S_H$</td>
<td>0.115</td>
<td>0.125</td>
<td>0.156</td>
</tr>
<tr>
<td>$S_{H_k}$</td>
<td>0.208</td>
<td>0.265</td>
<td>0.287</td>
</tr>
<tr>
<td>$W_H$</td>
<td>0.150</td>
<td>0.146</td>
<td>0.161</td>
</tr>
<tr>
<td>$W_{H_k}$</td>
<td>0.214</td>
<td>0.271</td>
<td>0.281</td>
</tr>
<tr>
<td>$C_H$</td>
<td>0.405</td>
<td>0.321</td>
<td>0.270</td>
</tr>
<tr>
<td>$C_{H_k}$</td>
<td>0.313</td>
<td>0.321</td>
<td>0.383</td>
</tr>
</tbody>
</table>

Relative change position $\tau = 0.5$

<table>
<thead>
<tr>
<th>$H = 0.6$</th>
<th>$H = 0.7$</th>
<th>$H = 0.8$</th>
<th>$H = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td></td>
<td>$n$</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>250</td>
<td>400</td>
</tr>
<tr>
<td>$S_H$</td>
<td>0.541</td>
<td>0.759</td>
<td>0.985</td>
</tr>
<tr>
<td>$S_{H_k}$</td>
<td>0.614</td>
<td>0.860</td>
<td>0.990</td>
</tr>
<tr>
<td>$W_H$</td>
<td>0.594</td>
<td>0.811</td>
<td>0.984</td>
</tr>
<tr>
<td>$W_{H_k}$</td>
<td>0.609</td>
<td>0.877</td>
<td>0.991</td>
</tr>
<tr>
<td>$C_H$</td>
<td>0.677</td>
<td>0.819</td>
<td>0.988</td>
</tr>
<tr>
<td>$C_{H_k}$</td>
<td>0.694</td>
<td>0.905</td>
<td>0.995</td>
</tr>
</tbody>
</table>

$H = 0.8$       |                | $H = 0.9$       |                |

| $n$             |                | $n$             |                |
| 50              | 100            | 250             | 400            | 50              | 100            | 250             | 400            |
| $S_H$           | 0.373          | 0.403           | 0.535           | 0.640           | 0.337          | 0.428           | 0.472           | 0.549           |
| $S_{H_k}$       | 0.454          | 0.563           | 0.649           | 0.711           | 0.412          | 0.480           | 0.582           | 0.604           |
| $W_H$           | 0.369          | 0.413           | 0.563           | 0.648           | 0.357          | 0.387           | 0.506           | 0.536           |
| $W_{H_k}$       | 0.443          | 0.566           | 0.662           | 0.710           | 0.433          | 0.528           | 0.544           | 0.599           |
| $C_H$           | 0.576          | 0.584           | 0.655           | 0.706           | 0.659          | 0.622           | 0.631           | 0.637           |
| $C_{H_k}$       | 0.535          | 0.599           | 0.716           | 0.760           | 0.604          | 0.574           | 0.637           | 0.638           |
Table 3.4: Empirical Power, estimated Hurst coefficient, relative change position $\tau = 0.5$, level shift of size $\mu = 1$ and change in variance from $\sigma^2 = 1$ to $\sigma_0^2 = 5/4$, nominal size $\alpha = 0.05$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$H = 0.6$</th>
<th>$H = 0.7$</th>
<th>$H = 0.8$</th>
<th>$H = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>400</td>
</tr>
<tr>
<td>$S_H$</td>
<td>0.701</td>
<td>0.931</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$S_{H_k}$</td>
<td>0.878</td>
<td>0.986</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$W_H$</td>
<td>0.609</td>
<td>0.812</td>
<td>0.989</td>
<td>1.000</td>
</tr>
<tr>
<td>$W_{H_k}$</td>
<td>0.734</td>
<td>0.973</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>$C_H$</td>
<td>0.660</td>
<td>0.899</td>
<td>0.999</td>
<td>1.000</td>
</tr>
<tr>
<td>$C_{H_k}$</td>
<td>0.588</td>
<td>0.916</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 3.5: Empirical Power, estimated Hurst coefficient, relative change position $\tau = 0.5, G_1(x) = x^2, G_2(x) = x^2 + x/2 + 1/2$, nominal size $\alpha = 0.05$, $H$ is the Hurst coefficient of the underlying Gaussian.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$H = 0.6$</th>
<th>$H = 0.7$</th>
<th>$H = 0.8$</th>
<th>$H = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>100</td>
<td>250</td>
<td>400</td>
</tr>
<tr>
<td>$S_H$</td>
<td>0.466</td>
<td>0.636</td>
<td>0.824</td>
<td>0.898</td>
</tr>
<tr>
<td>$S_{H_k}$</td>
<td>0.762</td>
<td>0.942</td>
<td>0.983</td>
<td>1.000</td>
</tr>
<tr>
<td>$W_H$</td>
<td>0.597</td>
<td>0.568</td>
<td>0.669</td>
<td>0.727</td>
</tr>
<tr>
<td>$W_{H_k}$</td>
<td>0.616</td>
<td>0.850</td>
<td>0.944</td>
<td>0.981</td>
</tr>
<tr>
<td>$C_H$</td>
<td>0.445</td>
<td>0.601</td>
<td>0.806</td>
<td>0.853</td>
</tr>
<tr>
<td>$C_{H_k}$</td>
<td>0.443</td>
<td>0.635</td>
<td>0.852</td>
<td>0.937</td>
</tr>
</tbody>
</table>
Distributional change under long-range dependence

\( n = 400 \) and \( H \leq 0.8 \). However, there is a loss in the empirical power. The power performance is much better, if the local Whittle estimator is modified. Actually, there is no loss in power if compared to the case where \( H \) was assumed to be known. Then again, the probability of a false rejection is higher than \( \alpha = 0.05 \), so the test is quite liberal.

Next we compare Cramér-von Mises, Wilcoxon and CUSUM test. The empirical size of the three tests is similar, no matter which estimator we choose and which situation we assume (sample size, Hurst coefficient), see Table 3.2.

In terms of empirical power the Cramér-von Mises and Wilcoxon test give similar results with the CUSUM test being slightly ahead for \( \tau = 1/2 \) and being clearly advantageous for early changes \( \tau = 1/5 \) (see Table 3.3).

We have to keep in mind that CUSUM and Wilcoxon test are designed to detect changes in the mean. On the contrary, the Cramér-von Mises test is a so called omnibus test and has power against arbitrary changes in the marginal distribution.

Therefore, we consider another situation, with the mean-shift being now accompanied by a small change in the variance. In detail,

\[
Y_i = \begin{cases} 
X_i & \text{for } i \leq \lfloor n\tau \rfloor, \\
\sigma X_i + \mu & \text{for } i > \lfloor n\tau \rfloor, 
\end{cases}
\]

for Gaussian \( \{X_i\}_{i \geq 1} \). The theoretic result in Example 3.4.4 indicates that in this scenario the Cramér-von Mises test should be advantageous. In fact, for all combinations of sample size \( n \) and Hurst coefficient \( H \) the empirical power against this change is always higher than the power against a mean-shift under constant variance. Moreover, the Cramér-von Mises test has clearly higher power than CUSUM and Wilcoxon test, see Table 3.4, which matches the theoretical findings of Example 3.4.4. Furthermore, we consider the change-point problem

\[
Y_i = \begin{cases} 
X_i^2 & \text{for } i \leq \lfloor n\tau \rfloor, \\
X_i^2 + aX_i + \mu & \text{for } i > \lfloor n\tau \rfloor, 
\end{cases}
\]

corresponding to a situation in which mean, variance, skewness and the Hermite rank change (see Example 3.2.7). Table 3.5 displays the empirical power of the three tests against this alternative and the picture is quite clear. The Cramér-von Mises test has the highest power for all combinations of \( H \) and \( n \), while the Wilcoxon test is second best. Also note that the Hermite rank of the pre-change transformation is \( m = 2 \). Consequently, these observations are short-range dependent for \( H < 0.75 \).
Table 3.6: Empirical size and power for farima\((0, 0.2, 0)\)-sequences, Hurst coefficient is estimated, nominal size \(\alpha = 0.05\).

<table>
<thead>
<tr>
<th></th>
<th>No change</th>
<th>Mean-shift (\mu = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>50 100 250 400</td>
<td>50 100 250 400</td>
</tr>
<tr>
<td></td>
<td>(\hat{S}_H)</td>
<td>(\hat{S}_{\hat{H}})</td>
</tr>
<tr>
<td></td>
<td>0.036 0.058 0.069 0.056</td>
<td>0.276 0.490 0.832 0.920</td>
</tr>
<tr>
<td></td>
<td>(\hat{S}_{\hat{H}})</td>
<td>(\hat{S}_{\hat{H}})</td>
</tr>
<tr>
<td></td>
<td>0.167 0.148 0.119 0.129</td>
<td>0.520 0.737 0.939 0.986</td>
</tr>
<tr>
<td></td>
<td>(\hat{W}_H)</td>
<td>(\hat{W}_{\hat{H}})</td>
</tr>
<tr>
<td></td>
<td>0.067 0.139 0.086 0.074</td>
<td>0.601 0.882 0.968 0.730</td>
</tr>
<tr>
<td></td>
<td>(\hat{W}_{\hat{H}})</td>
<td>(\hat{W}_{\hat{H}})</td>
</tr>
<tr>
<td></td>
<td>0.247 0.189 0.160 0.153</td>
<td>0.573 0.711 0.934 0.980</td>
</tr>
<tr>
<td></td>
<td>(\hat{C}_H)</td>
<td>(\hat{C}_{\hat{H}})</td>
</tr>
<tr>
<td></td>
<td>0.104 0.061 0.053 0.048</td>
<td>0.281 0.479 0.836 0.945</td>
</tr>
<tr>
<td></td>
<td>(\hat{C}_{\hat{H}})</td>
<td>(\hat{C}_{\hat{H}})</td>
</tr>
<tr>
<td></td>
<td>0.212 0.202 0.158 0.114</td>
<td>0.499 0.682 0.937 0.977</td>
</tr>
</tbody>
</table>

3.5.3 farima\((0, d, 0)\)-processes

For Gaussian long memory processes beyond fractional Gaussian noise, not only the Hurst coefficient determines the normalization. The theoretic normalization is given as \(Var(\sum_{i=1}^{n} X_i)\). We make use of its asymptotic expression and get

\[
d_{n,1} = n^H L^{1/2}(n)(H(2H - 1))^{-1/2},
\]

see Lemma 2.1.4. In this study we assume, as \(n \to \infty\), \(L(n) \to C\), which is quite common in the literature. For fractional Gaussian noise we have \(C = H(2H - 1)\), so the two factors just cancel out. In general the constant \(C\) is given through the limit

\[
\rho(k)k^{2-2H} \to C,
\]

as \(k \to \infty\). We suggest an estimator for \(C\) (which is quite heuristic) by:

\[
\hat{C} = \frac{1}{K} \sum_{k=1}^{K} \hat{\rho}(k)k^{2-2\hat{H}},
\]

with \(\hat{H}\) being one of the two estimators from above. Finally, we use the normalization

\[
\hat{d}_n = n^{\hat{H}} \hat{C}^{1/2}(\hat{H}(2\hat{H} - 1))^{1/2}.
\]

Usually estimators constructed via the spectral domain are favorable compared to such time-domain estimators. For a spectral domain approach see Abadir et al. (2009). However, using \(\hat{C}\), one might consider the normalization

\[
\hat{d}_n = n^{\hat{H}} \hat{C}^{1/2}(\hat{H}(2\hat{H} - 1))^{1/2}.
\]
Table 3.7: Empirical size and power for $\text{farima}(1, 0.2, 0)$-sequences with AR-coefficient $a_1 = 0.4$. Hurst coefficient is estimated, nominal size $\alpha = 0.05$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>400</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^h_H$</td>
<td>0.032</td>
<td>0.021</td>
<td>0.030</td>
<td>0.034</td>
<td>0.154</td>
<td>0.183</td>
<td>0.329</td>
<td>0.433</td>
</tr>
<tr>
<td>$S^h_{H_k}$</td>
<td>0.126</td>
<td>0.065</td>
<td>0.062</td>
<td>0.052</td>
<td>0.316</td>
<td>0.325</td>
<td>0.415</td>
<td>0.489</td>
</tr>
<tr>
<td>$W^h_H$</td>
<td>0.038</td>
<td>0.009</td>
<td>0.003</td>
<td>0.007</td>
<td>0.157</td>
<td>0.158</td>
<td>0.192</td>
<td>0.266</td>
</tr>
<tr>
<td>$W^h_{H_k}$</td>
<td>0.183</td>
<td>0.048</td>
<td>0.015</td>
<td>0.017</td>
<td>0.363</td>
<td>0.313</td>
<td>0.301</td>
<td>0.387</td>
</tr>
<tr>
<td>$C^h_H$</td>
<td>0.445</td>
<td>0.288</td>
<td>0.124</td>
<td>0.081</td>
<td>0.625</td>
<td>0.615</td>
<td>0.605</td>
<td>0.689</td>
</tr>
<tr>
<td>$C^h_{H_k}$</td>
<td>0.422</td>
<td>0.302</td>
<td>0.138</td>
<td>0.093</td>
<td>0.592</td>
<td>0.613</td>
<td>0.656</td>
<td>0.710</td>
</tr>
</tbody>
</table>

Table 3.8: Empirical size and power for $AR(1)$-sequences with AR-coefficient $a_1 = 0.6$. Hurst coefficient is estimated, nominal size $\alpha = 0.05$.

<table>
<thead>
<tr>
<th>$n$</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>400</th>
<th>50</th>
<th>100</th>
<th>250</th>
<th>400</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S^h_H$</td>
<td>0.016</td>
<td>0.008</td>
<td>0.004</td>
<td>0.002</td>
<td>0.135</td>
<td>0.150</td>
<td>0.324</td>
<td>0.576</td>
</tr>
<tr>
<td>$S^h_{H_k}$</td>
<td>0.097</td>
<td>0.030</td>
<td>0.012</td>
<td>0.004</td>
<td>0.282</td>
<td>0.260</td>
<td>0.374</td>
<td>0.556</td>
</tr>
<tr>
<td>$W^h_H$</td>
<td>0.036</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
<td>0.149</td>
<td>0.106</td>
<td>0.106</td>
<td>0.213</td>
</tr>
<tr>
<td>$W^h_{H_k}$</td>
<td>0.141</td>
<td>0.018</td>
<td>0.000</td>
<td>0.000</td>
<td>0.336</td>
<td>0.249</td>
<td>0.235</td>
<td>0.343</td>
</tr>
<tr>
<td>$C^h_H$</td>
<td>0.467</td>
<td>0.279</td>
<td>0.017</td>
<td>0.006</td>
<td>0.641</td>
<td>0.592</td>
<td>0.606</td>
<td>0.721</td>
</tr>
<tr>
<td>$C^h_{H_k}$</td>
<td>0.412</td>
<td>0.219</td>
<td>0.026</td>
<td>0.006</td>
<td>0.654</td>
<td>0.631</td>
<td>0.678</td>
<td>0.794</td>
</tr>
</tbody>
</table>

The estimator in (3.1) is only defined under long memory, that is $H > 0.5$ (or in this situation $\hat{H} > 0.5$). Therefore, we modify both estimators by considering $\max(\hat{H}, 0.501)$ instead of $\hat{H}$. The effect of this modification on short memory processes will be seen in the next section. However, for $\text{farima}(0, d, 0)$-sequences it seems to work quite well, see Table 3.6. Note that critical values are still deduced from fractional Gaussian noise. The finite sample performance (under the hypothesis as well as under a mean-shift) is very similar to the case where the data comes from fractional Gaussian noise. Meaning, the Cramér-von Mises test has good properties and the different tests yield very similar results, again matching the theoretic findings.

### 3.5.4 Short-range dependent effects

Finally, we have considered deviations from purely LRD sequences by simulating $\text{farima}(1, d, 0)$-time series and short memory $AR(1)$-processes. First, we have applied the tests to $\text{farima}(0, d, 1)$-sequences, which are still long-range dependent. Table 3.7 indicates that the empirical power of the Cramér-von Mises test is less than in the case of $\text{farima}(0, d, 0)$-processes. However, the test works principally well, meaning that
the power increases with the number of observations while the empirical size stays close to the nominal size. For CUSUM and Wilcoxon test, this seems to be not the case.

For the (purely short-range dependent) AR(1)-processes we make two observations: First, due to the assumption of LRD ($H > 0.5$) the normalization is too strong and the statistics converge to 0, at least under stationarity. If the structural change is big enough, the tests might still detect the change (see Table 3.8). However, there is a certain loss in power. Secondly, Cramér-von Mises test and CUSUM test are showing a quite different finite sample performance. While under LRD (in concordance with the theory) their empirical size and power is always very similar, we now observe situations where the Cramér-von Mises test has empirical power 0.374 and the CUSUM test 0.678, see the results in Table 3.8. Again, this matches the theoretical fact that under short memory the tests show a different asymptotic behavior.

3.6 Proofs of the main results

3.6.1 Proof of Theorem 5

It is the goal to approximate the sequential empirical process by a linear combination of multiple partial sum processes. The indicator function $1_{\{G_n(x_j) \leq x\}}$ has the Hermite expansion

$$1_{\{G_n(x_j) \leq x\}} = \sum_{q=0}^{\infty} \frac{J_{q,n}(x)}{q!} H_q(X_j).$$

Remind that $J_{q,n}(x) = E[1_{\{G_n(X_j) \leq x\}}H_q(X_j)]$ and especially $J_{0,n}(x) = P(G_n(X_j) \leq x) = F_n(x)$. Now let $L_{m,n,j}(x)$ be the Hermite expansion up to $m$, in detail

$$L_{m,n,j}(x) = \sum_{q=0}^{m} \frac{J_{q,n}(x)}{q!} H_q(X_j).$$

Let $m(n)$ be the Hermite rank of $(1_{\{G_n(x_j) \leq x\}})_x$. Then we have by the conditions of Theorem 5 that $m^* \leq m(n) \leq m$ for some $m^* \leq m < 1/D$. Thus

$$L_{m,n,j}(x) = F_n(x) + \sum_{q=m^*}^{m} \frac{J_{q,n}(x)}{q!} H_q(X_j).$$

Moreover, define

$$S_n(l; x) = \frac{1}{d_{n,m}} \sum_{j=1}^{l} (1_{\{G_n(x_j) \leq x\}} - L_{m,n,j}(x)).$$
Finally, let $S_n(k; x, y) = S_n(k; y) - S_n(k; x)$, $L_{m,n,j}(x, y) = L_{m,n,j}(y) - L_{m,n,j}(x)$ and $J_{n,q}(x, y) = J_{n,q}(y) - J_{n,q}(x)$.

We will make use of a chaining technique similar to those of Dehling and Taqqu (1989). To this end, define

$$
\Lambda_n(x) := \int_{\{G_n(s) \leq x\}} \left( \sum_{q=0}^{m} \frac{|H_q(s)|}{q!} \right) \phi(s) \, ds
$$

and observe that $J_{q,n}(x, y)/q!$ is bounded by $\Lambda_n(x, y) = \Lambda_n(y) - \Lambda_n(x)$, for all $n \in \mathbb{N}$ and all $q = 0, \ldots, m$. Furthermore, $\Lambda_n$ is monotone, $\Lambda_n(-\infty) = 0$ and $\Lambda_n(\infty) = \int_{\mathbb{R}} \left( \sum_{q=0}^{m} \frac{|H_q(s)|}{q!} \right) \phi(s) \, ds = C < \infty$, for all $n \in \mathbb{N}$.

Define partitions, similarly to Dehling and Taqqu (1989), but now depending on $n$, by

$$
x_i(k) = x_i^{(n)}(k) = \inf \{x | \Lambda_n(x) \geq \Lambda_n(+\infty)2^{-k}\} \quad i = 0, \ldots, 2^k - 1
$$

for $k = 0, \ldots, K$, with the integer $K$ chosen below. Then we have

$$
\Lambda_n(x_i(k) -) - \Lambda_n(x_{i-1}(k)) \leq \Lambda_n(+\infty)2^{-k}.
$$

Note that the right hand side of (3.2) does not depend on $n$. Based on these partitions we can define chaining points $i_k(x)$ by

$$
x_{i_k(x)}(k) \leq x < x_{i_k(x)+1}(k),
$$

for each $x$ and each $k \in \{0, 1, \ldots, K\}$, see Dehling and Taqqu (1989).

**Lemma 3.6.1.** Define the chaining points as above. Suppose the following two conditions hold:

(i) There are constants $\gamma > 0$ and $C > 0$, not depending on $n$, such that for all $k \leq n$

$$
E|S_n(k; x, y)|^2 \leq C \left( \frac{k}{n} \right)^{n-\gamma} F_{(n)}(x, y).
$$

(ii) For all $\epsilon > 0$ and all $n \in \mathbb{N}$ there is a real number $K = K(n, \epsilon)$, such that for all $\lambda > 0$

$$
P \left( \sup_{x \in \mathbb{R}} \left| \frac{1}{dn,m} \sum_{j=1}^{l} L_{m,n,j}(x_{i_k(x)}(K), x_{i_k(x)+1}(K)-) \right| > \epsilon \right) \leq C \left( \frac{l}{n} \right)^{2-m^*} n^{\lambda-m^*}.
$$
Then there is a constant $\rho > 0$, such that for all $n \in \mathbb{N}$ and all $\epsilon > 0$ the following holds:

$$P \left( \sup_x |S_n(l; x)| > \epsilon \right) \leq C \left( \frac{l}{n} \right)^{-\gamma} n^{-2(K(n, \epsilon) + 3)} + C \left( \frac{l}{n} \right)^{2-m^*D} n^{\lambda-m^*D}.$$ 

Proof. Due to definition of the chaining points each point $x$ is linked to $-\infty$ in detail

$$-\infty = x_{i_0(x)}(0) \leq x_{i_1(x)}(1) \leq \cdots \leq x_{i_K(x)}(K) \leq x < x_{i_K(x)+1}(K)$$

We have

$$S_n(l; x) = \sum_{k=1}^{K} S_n(l; x_{i_k-1}(x)(k-1), x_{i_k}(x)(k)) + S_n(l; x_{i_K(x)}(K), x). \quad (3.3)$$

The last summand of the right hand side of (3.3) can be treated as follows

$$|S_n(l; x_{i_K(x)}(K), x)| = \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} \sum_{k=1}^{K} \left( 1_{x_{i_K(x)}(K) < G_n(x_j) \leq x_{i_K(x)}(K)} - L_{m,n,j}(x_{i_K(x)}(K), x) \right) \right|$$

$$\leq \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} \sum_{k=1}^{K} \left( 1_{x_{i_K(x)}(K) < G_n(x_j) < x_{i_K(x)}(K)+1(K)} - L_{m,n,j}(x_{i_K(x)}(K), x) \right) \right|$$

$$+ 2 \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x) \right|$$

$$= \left| S_n(l; x_{i_K(x)}(K), x) \right|$$

$$+ 2 \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x) \right|. \quad (3.4)$$

By (3.3) and (3.4) we get, using $\sum_{k=1}^{\infty} (k + 2)^{-2} < 1/2$,

$$P \left( \sup_x |S_n(l; x)| > \epsilon \right) \leq P \left( \sup_x |S_n(l; x)| > \epsilon \sum_{k=1}^{K+1} (k + 2)^{-2} + \epsilon/2 \right)$$

$$\leq \sum_{k=1}^{K} P \left( \max_x |S_n(l; x_{i_k-1}(x)(k-1), x_{i_k}(x)(k))| > \epsilon/(k + 2)^2 \right) \quad (3.5)$$

$$+ P \left( \max_x |S_n(l; x_{i_K(x)}(K), x_{i_K(x)+1}(K)-)| > \epsilon/(K + 3)^2 \right) \quad (3.6)$$

$$+ P \left( 2d_{n,m}^{-1} \left| \sum_{j=1}^{l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_K(x)+1}(K)-) \right| > \epsilon/2 \right). \quad (3.7)$$
Further, by condition (i) of Lemma 3.6.1 and the Markov inequality we get
\[
P \left( \max_{x} |S_n(l; x_{i_k(x)}(k), x_{i_{k+1}(x)}(k+1))| > \epsilon/(k+2)^2 \right) \\
\leq \sum_{i=0}^{2^{k+1}-1} P \left( S_n(l; x_i(k+1), x_{i+1}(k+1)) > \epsilon/(k+2)^2 \right) \\
\leq C \sum_{i=0}^{2^{k+1}-1} \left( \frac{l}{n} \right) n^{-\gamma} \frac{(k+2)^4}{\epsilon^2} F_n(x_i(k+1), x_{i+1}(k+1)) \\
\leq C \left( \frac{l}{n} \right) n^{-\gamma} \frac{(k+2)^4}{\epsilon^2}. \tag{3.8}
\]

The constant \( C \) in (3.8) is the constant of condition (i) in Lemma 3.6.1 and thus independent of \( n \). In the next line this \( C \) gets multiplied with \( \Lambda_n(\lambda, +\infty) \), which is a constant by itself. Thus the \( C \) in the inequality above is a universal constant, not depending on \( n \). The same is true for \( \gamma \).

Using the same arguments we get
\[
P \left( \max_{x} |S_n(l; x_{i_K(x)}(K), x_{i_{K+1}(x)}(K)-)| > \epsilon/(K+3)^2 \right) \leq C \left( \frac{l}{n} \right) n^{-\gamma} \frac{(K+3)^4}{\epsilon^2}.
\]

Finally, we have by condition (ii) of Lemma 3.6.1
\[
P \left( 2d_{n,m}^{-1} \left| \sum_{j \leq l} L_{m,n,j}(x_{i_K(x)}(K), x_{i_{K+1}(x)}(K)-) \right| > (\epsilon/2) \right) \leq C \left( \frac{l}{n} \right) 2^{-m^*D} n^{\lambda-m^*D},
\]
for all \( \lambda > 0 \). Combining the estimates for (3.5), (3.6) and (3.7) we arrive at
\[
P \left( \sup_{x} |S_n(l; x)| > \epsilon \right) \leq C \left( \frac{l}{n} \right) n^{-\gamma} \epsilon^{-2} \sum_{k=1}^{K+1} (k+2)^4 + C \left( \frac{l}{n} \right) 2^{-m^*D} n^{\lambda-m^*D} \\
\leq C \left( \frac{l}{n} \right) n^{-\gamma} \epsilon^{-2} (K+3)^5 + C \left( \frac{l}{n} \right) 2^{-m^*D} n^{\lambda-m^*D}.
\]

which finishes the proof. \qed

**Lemma 3.6.2.** There exist constants \( \gamma \) and \( C \), not depending on \( n \), such that for all \( k \leq n \)
\[
E |S_n(k; x, y)|^2 \leq C \left( \frac{k}{n} \right) n^{-\gamma} F_n(x, y).
\]

The proof is very close to the proof of Lemma 3.1 in Dehling and Taqqu (1989). However, for further results it is crucial that \( C \) and \( \gamma \) only depend indirectly on the function \( G_n \), namely through the Hermite rank. Thus we give a detailed proof to highlight this fact.
Proof. First, obtain the Hermite expansion

\[ 1_{\{x < G_n(X_i) \leq y\}} - F_n(x, y) = \sum_{q=m^*}^{\infty} \frac{J_{q,n}(x, y)}{q!} H_q(X_i). \]

Secondly, we have by orthogonality of the \( H_q(X_i) \) and \( EH_q^2(X_i) = q! \)

\[
\sum_{q=m^*}^{\infty} \frac{J_{q,n}^2(x, y)}{q!} = \sum_{q=m^*}^{\infty} E\left( \frac{J_{q,n}(x, y)}{q!} H_q(X_i) \right)^2 \\
= E\left( \sum_{q=m^*}^{\infty} \frac{J_{q,n}(x, y)}{q!} H_q(X_i) \right)^2 \\
= E\left( 1_{\{x < G_n(X_i) \leq y\}} - F_n(x, y) \right)^2 \\
= F_n(x, y)(1 - F_n(x, y)) \\
\leq F_n(x, y).
\]

This yields

\[
E(d_{n,m}S_n(k; x, y))^2 = \sum_{q=m+1}^{\infty} \frac{J_{q,n}^2(x)}{q!} \frac{1}{q!} \sum_{i,j \leq k} EH_q(X_i)H_q(X_j) \\
\leq F_n(x, y) \sum_{i,j \leq k} |r(i - j)|^{m+1}.
\]

Note that the second factor of the product in the last line may depend indirectly on the function \( G_n \), because \( G_n \) determines \( m \), however this is the only influence. For different combinations of \( m \) and \( D \) the term \( \sum_{i,j \leq k} |r(i - j)|^{m+1} \) might have a different asymptotic order. However, in all cases we get (see page 1777 in Dehling and Taqqu (1989))

\[
\frac{1}{d_{n,m}} \sum_{i,j \leq k} |r(i - j)|^{m+1} \leq C n^{mD - 2} L^{-m(n)} k^{1/(2-(m+1)/D)} L_1(k) \\
\leq C \left( \frac{k}{n} \right)^{1/(2-(m+1)/D)} n^{mD-1-(-D)} L_1(k) L^{-m(n)}.
\]

The result then follows because \( L \) and \( L_1 \) are slowly varying. \( \square \)

Lemma 3.6.3. Let \( n \in \mathbb{N} \) and \( \epsilon > 0 \). Define the chaining points and \( L_{m,n,j}(x) \) as in (3.1). Set

\[
K = K(n, \epsilon) = \left\lceil \log_2 \left( \frac{(m - m^* + 2)A_n(+\infty)}{\epsilon} \frac{nd_{n,m}}{n} \right) \right\rceil + 1.
\]
Then there is a constant $C > 0$, such that for all $\lambda > 0$

$$P \left( \sup_{x \in \mathbb{R}} \left| \frac{1}{d_{n,m}} \sum_{j=1}^{l} L_{m,n,j}(x_{iK}(x), x_{iK(x)+1}(K)) \right| > \epsilon \right) \leq C \left( \frac{l}{n} \right)^{2-m^*D} n^{\lambda-m^*D}.$$  

**Proof.** By construction of the chaining points we have for $q = 0, \ldots, m$

$$\sup_{x \in \mathbb{R}} |J_{q,n}(x_{iK}(x), x_{iK(x)+1}(K))|/q! \leq \Lambda_n(+\infty)2^{-K}.$$  

Thus for all $x \in \mathbb{R}$

$$\frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} L_{m,n,j}(x_{iK(x), x_{iK(x)+1}(K)}) \right| \leq \sum_{q=0}^{m} |J_{q,n}(x_{iK(x), x_{iK(x)+1}(K)})|/q! \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} H_q(X_j) \right| \leq \Lambda_n(+\infty)2^{-K} \sum_{q=0}^{m} \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} H_q(X_j) \right|.$$  

By definition of $K$

$$\frac{2^K \epsilon}{(m - m^* + 2) \Lambda_n(+\infty)} \geq \frac{n}{d_{n,m}}.$$  

Therefore, we get by Markov’s inequality for $q = m^*, \ldots, m$

$$P \left( \Lambda_n(+\infty)2^{-K} \frac{1}{d_{n,m}} \left| \sum_{j=1}^{l} H_q(X_j) \right| > \epsilon/(m - m^* + 2) \right) \leq P \left( \left| \sum_{j=1}^{l} H_q(X_j) \right| > n \right) \leq C \frac{l^2}{n^2} \leq C \frac{l^{2-qD}}{n^2} \leq C \left( \frac{l}{n} \right)^{2-m^*D} n^{\lambda-m^*D}.$$  

For $q = 0$ the term is deterministic, thus the probability is 0.  

**Proof of Theorem 5.** The two conditions of Lemma 3.6.1 are satisfied (see Lemma 3.6.2 and Lemma 3.6.3 ) with

$$K = \left\lfloor \log_2 \left( \frac{(m - m^* - 2) \Lambda_n(+\infty)}{\epsilon} d_{n,m}^{-1} \right) \right\rfloor + 1.$$
Note that \((K + 3)^{5} \leq C\varepsilon^{-1}n^{\delta}\) for any \(\delta > 0\), see Dehling and Taqqu (1989), page 1781. By this fact and by virtue of Lemma 3.6.1

\[
P\left(\sup_{x}\left|S_{n}(l; x)\right| > \epsilon\right) \leq C\left(\frac{l}{n}\right)^{\delta} + C\left(\frac{l}{n}\right)^{2-m^*D}n^{\lambda-m^*D}
\]

\[
\leq Cn^{-\rho}\left\{\left(\frac{l}{n}\right)\epsilon^{-3} + \left(\frac{l}{n}\right)^{2-m^*D}\right\},
\]

with \(\rho = \min(\gamma - \delta, m^*D - \lambda)\). Now choose \(\delta < \gamma\), then \(\rho > 0\) and we have thus proven a reduction principle in \(x\). It remains to verify uniformity in \(l\). For \(n = 2^{r}\) one gets by the same arguments as in the proof of Theorem 3.1 in Dehling and Taqqu (1989)

\[
P\left(\max_{l \leq n}\sup_{x}\left|S_{n}(l; x)\right| > \epsilon\right) \leq Cn^{-k}(1 + \epsilon^{-3})
\]

for any \(0 < \epsilon \leq 1\) and universal constants \(C\) and \(\epsilon\). Next consider arbitrary \(n\) and define for \(r\) such that \(2^{r-1} < n \leq 2^{r}\)

\[
S_{n}^{\ast}(l, x) = \frac{1}{d_{2^{r}, m}} \sum_{j=1}^{l}(1_{\{G_{n}(X_{j}) \leq x\}} - L_{m,n,j}(x)) \quad \text{for} \quad l \leq 2^{r},
\]

where \(\{G_{n}(X_{j})\}_{n \in \mathbb{N}, j \leq 2^{r}}\) is a (slightly modified) array. One obtains

\[
P\left(\max_{l \leq n}\sup_{x}\left|S_{n}^{\ast}(l; x)\right| > \epsilon\right) \leq C(2^{r})^{-k}(1 + \epsilon^{-3}).
\]

Hence

\[
P\left(\max_{l \leq n}\sup_{x}\left|S_{n}(l; x)\right| > \epsilon\right) \leq P\left(\max_{l \leq n}\sup_{x}\left|S_{n}^{\ast}(l; x)\right| > \epsilon d_{n,m}/d_{2^{r}, m}\right)
\]

\[
\leq \leq C(2^{r})^{-k}\left(1 + \epsilon^{-3} \left(\frac{d_{2^{r}, m}}{d_{n,m}}\right)^{3}\right)
\]

\[
\leq Cn^{-k}(1 + \epsilon^{-3}).
\]

The last line holds since \(d_{2^{r}, m}/d_{n,m}\) is uniformly bounded away from 0 and \(\infty\). □
3.6.2 Proof of Theorem 6

We start by proving a reduction principle for the empirical process in presence of a change point. Consider the array \( \{Y_{n,i}\}_{n \in \mathbb{N}, i \leq n} \), defined in (3.7), and let \( H_{n,i}(x) = P(Y_{n,i} \leq x) \). Define

\[
S_n^{(\tau)}(t, x) = \frac{1}{d_{n,m}} \sum_{i=1}^{\lfloor nt \rfloor} \left( 1_{\{Y_{n,i} \leq x\}} - H_{n,i}(x) - \sum_{q=m}^{m^*} \frac{J_{q,n,i}(x)}{q!} H_q(X_i) \right),
\]

where \( J_{q,n,i}(x) = E[1_{\{Y_{n,i} \leq x\}} H_q(X_i)] \). Note that \( J_{q,n,i}(x) = 0 \) if \( i \leq \lfloor n\tau \rfloor \) and \( q \leq m \).

**Lemma 3.6.4.** Under the conditions of Theorem 6 we can find constants \( C \geq 0 \) and \( \kappa > 0 \), such that for all \( \epsilon > 0 \)

\[
P \left( \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_n^{(\tau)}(t, x)| > \epsilon \right) \leq Cn^{-\kappa}(1 + \epsilon^{-3}).
\]

**Proof.** Define

\[
S_{n,1}(t, x) = \frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor nt \rfloor} \left( 1_{\{G(X_j) \leq x\}} - F(x) - \frac{J_m(x)}{m!} H_m(X_j) \right)
\]

and

\[
S_{n,2}(t, x) = \frac{1}{d_{n,m}} \sum_{j=1}^{\lfloor nt \rfloor} \left( 1_{\{G_n(X_j) \leq x\}} - F_n(x) - \sum_{q=m}^{m^*} \frac{J_{q,n}(x)}{q!} H_q(X_j) \right).
\]

By Theorem 5 we have

\[
P \left( \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_{n,i}(t, x)| > \epsilon \right) \leq Cn^{-\kappa}(1 + \epsilon^{-3}) \quad i = 1, 2. \quad (3.9)
\]

Next obtain

\[
S_n^{(\tau)}(t, x) = \begin{cases} 
S_{n,1}(t, x), & \text{if } t \leq \tau, \\
S_{n,2}(t, x) + S_{n,1}(\tau, x) - S_{n,2}(\tau, x), & \text{if } t > \tau.
\end{cases}
\]

Therefore, we get for all \( n \in \mathbb{N} \) and all \( \epsilon > 0 \), using (3.9) several times,

\[
P \left( \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_n^{(\tau)}(t, x)| > \epsilon \right) \leq 2P \left( \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_{n,1}(t, x)| > \epsilon/4 \right)
\]

\[
+ 2P \left( \sup_{t \in [0,1]} \sup_{x \in \mathbb{R}} |S_{n,2}(t, x)| > \epsilon/4 \right)
\]

\[
\leq 4Cn^{-\kappa}(1 + \epsilon^{-3}).
\]

\[\square\]
Lemma 3.6.5. Let Assumption A hold. Then for all $q \leq m$

$$
\sup_{x \in \mathbb{R}} d_{n,m^*}/d_{n,m} |J_{q,n}(x) - J_q(x)| \to 0, \quad (3.10)
$$
as $n \to \infty$.

**Proof.** Using Hölder’s inequality, one has for any $p \in \mathbb{N}$

$$
|J_{q,n}(x) - J_q(x)| = |E \left( \left( 1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}} \right) H_q(X_i) \right)|
\leq \left( E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|^{|p+1)/p| \right) \|H_q(X_i)\|_{L^{p+1}}
\leq C (E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|^{p/(p+1)}).
$$

Now obtain

$$
E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|
= P (\{G_n(X_1) \leq x, G(X_1) > x \} \cup \{G_n(X_1) > x, G(X_1) \leq x \})
= 1 - P (\{G_n(X_1) \leq x, G(X_1) \leq x \}) - P (\{G_n(X_1) > x, G(X_1) > x \})
= P(\min\{G_n(X_1), G(X_1)\} \leq x) - P(\max\{G_n(X_1), G(X_1)\} \leq x)
= o(n^{(m^* - m)D(1+\delta)/2}),
$$

for some $\delta > 0$. The last line holds uniformly due to Assumptions A2. Finally,

$$
d_{n,m^*}/d_{n,m} |J_{q,n}(x) - J_q(x)|
\leq C n^{(m-m^*)D/2} L^{(m^*-m)/2} (n) \left( E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|^{p/(p+1)} \right)
\leq C \left( n^{(m-m^*)D(p+1)/p} E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|^{p/(p+1)} \right)
\leq C \left( n^{(m-m^*)D(1+1/p)/p} E|1_{\{G_n(X_i) \leq x\}} - 1_{\{G(X_i) \leq x\}}|^{p/(p+1)} \right).
$$

Finally, choosing $p > 1/\delta$ this implies (3.10).
Proof of Theorem 6. By definition of the functions $J_{q,n,i}$ we get

$$\frac{1}{d_{n,m}} \sum_{i=1}^{\lfloor nt \rfloor} \sum_{q=m^*}^{m} \frac{J_{q,n,i}(x)}{q!} H_q(X_i)$$

$$= \frac{1}{d_{n,m}} \frac{J_m(x)}{m!} \sum_{i=1}^{\lfloor nt \rfloor} H_m(X_i)$$

$$+ 1\{t>\tau\} \sum_{q=m^*}^{m-1} \frac{1}{d_{n,m}} \frac{J_{q,n}(x)}{q!} \frac{1}{d_{n,q}} \sum_{i=\lfloor n\tau \rfloor+1}^{\lfloor nt \rfloor} H_q(X_i)$$

$$+ 1\{t>\tau\} \frac{J_{m,n}(x) - J_m(x)}{m!} \frac{1}{d_{n,m}} \sum_{i=\lfloor n\tau \rfloor+1}^{\lfloor nt \rfloor} H_m(X_i).$$

The second and the third summands are negligible due to the uniform convergence of the functions $J_{q,n}$ (see Lemma 3.6.5). The first summand converges in distribution towards

$$\frac{J_m(x)}{m!} Z_m(t),$$

see Dehling and Taqqu (1989). Together with Lemma 3.6.4 this finishes the proof. \qed
4 Block bootstrap for the empirical process under long memory

Especially for empirical processes the blockwise bootstrap of Künsch (1989) provides a strong nonparametric tool. Consider first a short-range dependent time series. Then under some technical assumptions the normalized empirical process of this series converges to a zero-mean Gaussian process $K(x)$ with covariance kernel $h(x,y) = E[K(x)K(y)]$. This covariance kernel is unknown in practice and hard to estimate, due to its infinite dimension. Several authors (e.g. Bühlmann (1994), Naik-Nimbalkar and Rajarshi (1994) and Peligrad (1998)) treated this problem by successfully applying the block bootstrap to the empirical process.

In the case of long-range dependence the situation is different. Here the empirical process has been analyzed for various classes of random variables, see Section 2.2. In all situations the empirical process converges weakly to a so-called semi-degenerate process $g(x)Z$. Here $g(x)$ is a deterministic function and $Z$ a possibly non Gaussian, real-valued random variable. Hence the limiting distribution is not as hard to treat as in the short memory case. For linear processes $X_i = \sum_{j=1}^{\infty} b_j \xi_{i-j}$, the function $g(x)$ is just the probability density of the $X_i$ and might be estimated. However, in the case of nonlinear transformations the function $g(x)$ is unknown and hence a resampling method might be of interest. Lahiri (1993) considered the block bootstrap for the sample mean of long memory processes and showed that it is valid if and only if the non-bootstrapped sample mean (properly normalized) converges to a normal limit. The sampling window method does not suffer from this issue (see Hall et al. (1998)) and has become more popular for statistical inference on long memory time series (see Nordman and Lahiri (2005), Zhang et al. (2013), Bai et al. (2016) and Betken and Wendler (2017+)).

In this chapter we investigate the block bootstrap for the empirical process. To the best of our knowledge such a result does not exist in the literature, for any type of long memory. If the Hermite rank is larger than one, the bootstrap fails to estimate the true distribution of the empirical process. However, we show that even in this case it can be used to consistently estimate the unknown function $g(x)$.

Subsequently, we test the hypothesis that the process is a monotone transformation of a Gaussian time series. If the data come from such a process, $g(x)$ can be estimated in a more direct manner, namely by $\hat{g}_n(x) = -\phi(\Phi^{-1}(\hat{F}_n(x)))$, where $\phi$ and $\Phi$ are probability density and distribution function of a standard normal, respectively. But this estimator fails to be consistent in a general setting. We use this fact to develop a new testing procedure for monotonicity of
Further applications are tests which are based on empirical distribution functions, such as Wilcoxon, Kolmogorov-Smirnov and Cramèr-von Mises test. Under long-range dependence the function \( g(x) \) shows up in the limit of the associated test statistics, see Chapter 3.

4.1 Weak convergence of the bootstrapped empirical process

4.1.1 The case of subordinated Gaussian processes

Consider a stationary Gaussian process \((X_i)_{i \geq 1}\) with standard normal marginals and autocovariance function \( \rho(k) = E[X_i X_{i+k}] = k^{-D}L(k) \). Further assume that our data come from transformations of these random variables, just as described in Section 2.1.2. In detail, let \( Y_i = G(X_i) \) with \( G \) being measurable.

In the previous section several change-point statistics were investigated for this type of sequences. As a special feature of long-range dependence, their limits are of a very coherent type. That is \( \left( a_m/m! \right) \sup_{t \in [0,1]} |\tilde{Z}_m(t)| \), for a Hermite bridge \((\tilde{Z}_m(t))_{t \in [0,1]}\) and a deterministic factor \( a_m \). The latter depends on the unknown transformation \( G \) and is therefore unknown by itself. It is given through:

<table>
<thead>
<tr>
<th>Kolmogorov-Smirnov</th>
<th>Cramér-von Mises</th>
<th>Wilcoxon</th>
<th>CUSUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_m )</td>
<td>( \sup_{x \in \mathbb{R}}</td>
<td>J_m(x)</td>
<td>)</td>
</tr>
</tbody>
</table>

Apparently, the function \( J_m(x) = E 1\{G(X_i) \leq x\} H_m(X_1) \) plays a central role. To the best of our knowledge there exists no procedure in the literature to estimate it. To reiterate, it is the coefficient of the first non-zero term in the Hermite expansion of \( 1\{G(X_1) \leq x\} - F(x) \). Moreover, it appears in the limit of the empirical process, in detail

\[
\frac{1}{d_n} \sum_{i=1}^{n} (1\{Y_i \leq x\} - F(x)) \overset{D}{\to} \frac{J_m(x)}{m!} Z_m, \tag{4.1}
\]

which is an immediate consequence of Theorem 1 in Dehling and Taqqu (1989) (see also Theorem 2). Here

\[
d_n^2 = d_{n,m}^2 = \text{Var} \left( \sum_{i=1}^{n} H_m(X_i) \right).
\]

We drop the subscript \( m \), as the Hermite rank is assumed to be fixed in this chapter. By considering a bootstrap version of (4.1), we hope to gain information about \( J_m(x) \).

By bootstrap we mean block bootstrap and more precisely the MBB of Künsch (1989), see Section 2.4.1 for an introduction. For a realization of a subordinated Gaussian process \( Y_1, \ldots, Y_n \),
define \( n - l + 1 \) blocks of length \( l = l(n) \). Then choose randomly with replacement \( p = p(n) \) blocks, such that the bootstrap sample \( Y^*_1, \ldots, Y^*_p \) satisfies

\[
P\left( (Y^*_{(j-1)l+1}, \ldots, Y^*_jl) = I_i \right) = \frac{1}{n-l+1} \quad \text{for} \quad j = 1, \ldots, p, \quad i = 1, \ldots, n - l + 1.
\]

The common choice for the number of blocks is \( p = \lfloor n/l \rfloor \), however, our proof works for all \( p \to \infty \). Further denote the blocks of indices by

\[
B_i = (i, \ldots, i + l - 1) \quad i = 1, \ldots, n - l + 1.
\]

A central issue in the analysis of subordinated processes is that the underlying sequence, here the Gaussian \((X_i)_{i \geq 1}\) is unobservable. However, on might (theoretically) define a block bootstrap version of this process by \( X^*_1, \ldots, X^*_pl \), with

\[
P\left( (X^*_{(j-1)l+1}, \ldots, X^*_jl) = I^*_i \right) = \frac{1}{n-l+1} \quad \text{for} \quad j = 1, \ldots, p, \quad i = 1, \ldots, n - l + 1,
\]

and \( I^*_i = (X_{i}, \ldots, X_{i+l-1}) \). Obviously, the observable bootstrap sample can be written as

\[
Y^*_i = G(X^*_i) \quad \text{for} \quad i = 1, \ldots, pl.
\]

This representation will be a central point of the proofs below. In the case of long-range dependence the classical block bootstrap has been applied to subordinated Gaussian processes by Lahiri (1993) and to linear sequences by Kim and Nordman (2011). Both consider the bootstrap of the sample mean.

In what follows \( E^* \) will denote conditional expectation given the sample \( X_1, \ldots, X_n \). Analogously \( P^* \) denotes conditional probability and \( \mathcal{D} \) weak convergence with respect to \( P^* \).

The mentioned result of Lahiri (1993) then reads as follows. It is important to note that conditioning on \( Y_1, \ldots, Y_n \) does not suffice, which gets clear when checking the proofs of our results.

**Theorem 12** (Lahiri). Let \( l = O(n^{1-\epsilon}) \) for some \( 0 < \epsilon < 1 \) and \( p^{-1} + l^{-1} = o(1) \). If \( EG(X_1)^2 < \infty \) then

\[
\frac{1}{pl^2d_l} \sum_{i=1}^{pl} (Y^*_i - E^*Y^*_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_m) \quad \text{in probability},
\]

where \( \sigma_m = E[G(X_1)H_m(X_1)]/m! \).

Two things are remarkable. The first is that the bootstrap weakens the dependence of the
random variables, thus the different normalization is needed, satisfying
\[ d_{pl}/(p^{1/2}d_i) \to 0. \]
The second is that the limit is always normal. However, for Hermite ranks larger than 1 the partial sum of long-range dependent data converges towards a nonnormal limit, see Taqqu (1975), Taqqu (1979) and Dobrushin and Major (1979). Hence, the bootstrap technically fails in this case.

Next consider not only one but \( J \) independently drawn bootstrap samples denoted by
\[ Y_{1}^{*\,(j)}, \ldots, Y_{pl}^{*\,(j)} \quad j \in \{1, \ldots, J\}. \]
Define further
\[
S_n = \frac{1}{d_n} \sum_{i=1}^{n} (Y_i - EY_1) \quad \text{and} \quad S_n^{*\,(j)} = \frac{1}{p^{1/2}d_i} \sum_{i=1}^{pl} (Y_{i}^{*\,(j)} - E^{*\,1}Y_{i}^{*\,(j)}),
\]
the normalized partial sums of the original and the bootstrap samples, respectively. The next result is an unconditional version of Theorem B, which can be deduced from the conditional version (at least for real-valued random variables).

**Corollary 4.1.1.** Let the conditions of Theorem 12 hold. Then
\[
\left( S_n, S_n^{*\,(1)}, \ldots, S_n^{*\,(J)} \right) \overset{D}{\to} \sigma_m \left( Z_m, Z^{(1)}, \ldots, Z^{(J)} \right),
\]
where \( Z_m \) is defined as above and \( Z^{(j)} \) are independent standard normal random variables, independent of \( Z_m \).

**Proof.** For simplicity assume \( J = 2 \). Let \( x, y_1, y_2 \in \mathbb{R} \), then
\[
P \left( S_n \leq x, S_n^{*\,(1)} \leq y_1, S_n^{*\,(2)} \leq y_2 \right)
\]
\[
= E \left[ 1_{\{S_n \leq x\}} 1_{\{S_n^{*\,(1)} \leq y_1\}} 1_{\{S_n^{*\,(2)} \leq y_2\}} \right]
\]
\[
= E \left[ 1_{\{S_n \leq x\}} E^{*\,1} 1_{\{S_n^{*\,(1)} \leq y_1\}} 1_{\{S_n^{*\,(2)} \leq y_2\}} \right]. \tag{4.2}
\]
By Theorem 1, \( 1_{\{S_n \leq x\}} \) converges in distribution to \( 1_{\{Z_m \leq x\}} \). By Theorem 12, the random variables \( E^{*\,1} 1_{\{S_n^{*\,(1)} \leq y_1\}} \) and \( E^{*\,1} 1_{\{S_n^{*\,(2)} \leq y_2\}} \) converge in probability towards \( P(Z^{(1)} \leq y_1) \) and \( P(Z^{(2)} \leq y_2) \), respectively. As all three random variables are bounded, this implies convergence of (4.2) to \( P(Z_m \leq x) P(Z^{(1)} \leq y_1) P(Z^{(2)} \leq y_2) \).
Now define for each bootstrap sample the (normalized) empirical process by

$$W_n^*(j)(x) = \frac{1}{p^{1/2}d_l} \sum_{i=1}^{pl} \{Y_i^{*(j)} \leq x\} - E^*[1_{\{Y_i^{*(j)} \leq x\}}]. \quad (4.3)$$

The main theorem of this chapter then reads as follows.

**Theorem 13.** Let the class of functions $\{1_{G(t) \leq x} - F(t), -\infty < x < \infty\}$ have Hermite rank $m$ and let $0 < D < 1/m$. Let further the block length satisfy $l = O(n^{1-\epsilon})$ for some $0 < \epsilon < 1$ and $p^{-1} + l^{-1} = o(1)$. Then:

$$W_n^*(x) \xrightarrow{D} \frac{J_m(x)}{m!} Z \quad \text{in probability.}$$

$Z$ is standard normal and weak convergence takes place in $D([-\infty, \infty])$, equipped with the uniform topology. Moreover:

$$\left( W_n(x), W_n^{(1)}(x), \ldots, W_n^{*(j)}(x) \right) \xrightarrow{D} \frac{J_m(x)}{m!} \left( Z_m, Z^{(1)}, \ldots, Z^{(j)} \right),$$

where convergence takes place in $(D([-\infty, \infty]))^{j+1}$, equipped with the uniform topology. $Z_m$ is defined as above and $Z^{(j)}$ are independent standard normal random variables, independent of $Z_m$.

**Remark 4.1.2.** (i) Similar to the empirical process of LRD data (see (4.1)) the bootstrapped version has a semi-degenerate limit. However, the normalization in (4.3) is weaker than in (4.1) and the random part of the limit is always Gaussian, just as for the bootstrapped sample mean.

(ii) We decided to formulate Theorem 13 in two different ways. The first one is a conditional result and by weak convergence in probability we mean

$$d_{BL}(\mathcal{L}(W_n^*(\cdot)|X_1, \ldots, X_n), \mathcal{L}(J_m(\cdot)/m!Z)) \xrightarrow{P} 0.$$ 

Here $d_{BL}$ is based on the bounded-Lipschitz norm, see (2.4.1). It defines a metric on the distributions of $D[-\infty, \infty]$-valued random variables and was used in Radulovic (1996a) in the short memory case. It turns out that this metric and the semi-degenerate limit of the empirical process match very well.

(iii) The second statement is an unconditional result. It is argued in Bücher and Kojadinovic (2016) that this type of convergence might simplify possible applications, such as the functional delta method.

The main part of the proof of weak convergence of the non-bootstrapped empirical process is a reduction principle and this technique has become popular for empirical processes of LRD.
data ever since. Define
\[
R_n(x) = \frac{1}{d_n} \sum_{i=1}^{n} \left( 1_{\{Y_i \leq x\}} - F(x) - J_m/m!(x)H_m(X_i) \right).
\] (4.4)

Dehling and Taqqu (1989) have shown that \( R_n(x) \) converges uniformly and in probability towards 0. It is our aim to prove Theorem 13 using a reduction principle too. Consider the bootstrapped version of (4.4)
\[
R_{n,l}^*(x) = \frac{1}{d_{l}^{1/2}} \sum_{i=1}^{l} \left( 1_{\{Y_i^* \leq x\}} - \tilde{F}_{n,l}(x) - J_m(x)/m!(H_m(X_i^*) - \tilde{\mu}_{n,l}(H_m)) \right),
\] (4.5)
where
\[
\tilde{\mu}_{n,l}(H_m) = l^{-1}E^* \left[ \sum_{j \in B_1} H_m(X_i^*) \right] \quad \text{and} \quad \tilde{F}_{n,l}(x) = l^{-1}E^* \left[ \sum_{j \in B_1} 1_{\{Y_i^* \leq x\}} \right].
\] (4.6)

Lemma 4.1.3 (Bootstrap uniform weak reduction principle). Let the conditions of Theorem 13 hold. Then there are constants \( C \) and \( \rho \) such that
\[
P \left( \sup_{x \in R} |R_{n,l}^*(x)| > \epsilon \right) \leq C l^{-\rho} (\epsilon^{-3} + 1),
\]
for all \( \epsilon > 0 \).

Remark 4.1.4. One might go through the proof step by step and obtain
\[
\rho = \min\{2H - 1, (2 - 2H)/m, (2 - 2H)\epsilon\} - \delta,
\]
for any \( \delta > 0 \). Here \( \epsilon > 0 \) is such that \( l = O(n^{1-\epsilon}) \). We do not believe this result to be sharp.

The proof of the lemma is quite lengthy and will be presented below. Once the reduction principle is established, we might prove convergence of the bootstrapped empirical process.

Proof of Theorem 13. The unconditional statement follows directly from the reduction principle (Lemma 4.1.3) and Corollary 4.1.1.

For the conditional statement we will make use of the metric defined by
\[
d_{BL}(\mu, \nu) = \sup_{f \in L(F)} \left| \int f d\mu - \int f d\nu \right|.
\]
where
\[ L(\mathcal{F}) = \left\{ K : \mathcal{F} \to \mathbb{R} : \sup_f |K(f)| \leq 1, |K(f_1) - K(f_2)| \leq \|f_1 - f_2\| \forall f_1, f_2 \right\} \]
is the set of functions whose bounded-Lipschitz norm is smaller than one, see also Section 2.4.2. Then we obtain

\[ d_{BL}(L^*_{n}(W^*_n(\cdot)), L(J_m(\cdot)/m!Z)) \leq d_{BL}(L^*_{n}(W^*_n(\cdot)), L_n^*(J_m(\cdot)Z^*_n/m!)) + d_{BL}(L_n^*(J_m(\cdot)Z^*_n/m!), L(J_m(\cdot)/m!Z)). \tag{4.7} \]

For \( K \in L(D[-\infty, \infty]) \) we get

\[ |E^*[K(W^*_n(\cdot)) - E^*[K(J_m(\cdot)Z^*_n/m!)]| \leq E^* \left[ \min\{1, \sup_x |W^*_n(x) - J_m(x)Z^*_n/m!| \right], \]

which converges to zero by Lemma 4.1.3. Thus, the first summand of (4.7) goes to zero too. As for the second summand, let again \( K(\cdot) \) be an element of \( L(D[-\infty, \infty]) \) and define the function \( f_K : \mathbb{R} \to \mathbb{R} \) by \( f_K(z) = K(J(\cdot)z/m!) \). Then \( f_K \in L(\mathbb{R}) \) (due to boundedness of \( J_m(x) \)) and one obtains

\[ |E^*[K(J_m(\cdot)Z^*_n/m!)] - E[K(J_m(\cdot)Z/m!)]| = |E^*f_K(Z^*_n) - Ef_K(Z)| \leq \sup_{f \in L(\mathbb{R})} |E^*f(Z^*_n) - Ef(Z)| = d_{BL}(L^*_n(Z^*_n), L(Z)). \]

However, by Theorem 12 (Theorem 2.2 of Lahiri (1993)) \( Z^*_n \) converges weakly towards \( Z \). As \( d_{BL} \) metricizes weak convergence, we get \( d_{BL}(L^*_n(Z^*_n), L(Z)) \to 0 \) in probability.

### 4.1.2 Estimating the Hermite coefficient function

Comparing the asymptotic distributions in (4.1) and Theorem 13, one might conclude that the bootstrap fails, if \( m > 1 \). However, even in this case we can make use of Theorem 13 to estimate the function \( J_m(x) \) (up to its sign). As stated above, this function \( J_m(x) \) plays a central role in statistical inference for subordinated Gaussian processes.

For \( J \) bootstrap iterations denote the empirical process of the \( j \)-th sample by \( W^*_n(j)(x) \) and
define the estimator

\[ \hat{J}_{n,J,m}(x) = m! \left( \frac{1}{J} \sum_{j=1}^{J} (W_n^*(j)(x))^2 \right)^{1/2}. \tag{4.8} \]

Then Theorem 13 (in detail its unconditional version) immediately yields

\[ \lim_{J \to \infty} \left[ \lim_{n \to \infty} P \left( \sup_{x \in \mathbb{R}} \left| \hat{J}_{n,J,m}(x) - J_m(x) \right| > \epsilon \right) \right] = 0, \]

for all \( \epsilon > 0 \). However, this type of convergence is a bit unnatural and moreover the estimator (4.8) is just a Monte-Carlo approximation of \( m! \text{Var}^*(W_n^*(x)) \). This (conditional) variance can be explicitly computed, in fact

\[ \text{Var}^*(W_n^*(x)) = \frac{1}{n-l+1} \sum_{i=1}^{n-l+1} W_{n,l,i}(x)^2, \tag{4.9} \]

with \( W_{n,l,i}(x) = d_i^{-1} \sum_{j=i}^{i+l-1} (1_{\{Y_j \leq x\}} - \hat{F}_{n,l}(x)) \). For fixed \( x \in \mathbb{R} \) its consistency follows from Lemma 3.1 in Lahiri (1993). Moreover, we conjecture the following uniform consistency.

**Conjecture 4.1.5.** Let the conditions of Theorem 13 hold. Then

\[ \sup_{x \in \mathbb{R}} \left| (m!)^2 \text{Var}^*(W_n^*(x)) - J_m^2(x) \right| \xrightarrow{P} 0. \]

We will show that this conjecture holds in the special case of linear Gaussian processes.

**Theorem 14.** Consider a linear process given through \( X_i = \sum_{k=1}^{\infty} a_k \epsilon_{i-k} \), where the \( \epsilon_i \)'s are i.i.d. standard normal and \( (a_k)_k \) is regularly varying with exponent \( H - 3/2 \). Let moreover \( G: \mathbb{R} \to \mathbb{R} \) be a measurable, piecewise monotone transformation.

If further the block length satisfies \( l = O(n^{1-\epsilon}) \) for some \( (1-H)/(2-H) < \epsilon < 1 \) and \( p^{-1} + l^{-1} = o(1) \), the statement of Conjecture 4.1.5 holds.

In what follows we will consider the estimator

\[ \hat{J}_{n,m}(x) = m! (\text{Var}^*(W_n^*(x)))^{1/2}. \]

**Remark 4.1.6.** (i) For this kind of long memory sequences we can make use of the reduction principle of Wu (2003) (and thereby Corollary 2.2.1), which is the strongest result of this type. (ii) Any Gaussian process has a representation as a one-sided linear process if and only if its spectral density \( \psi() \) satisfies \( \int_{-\pi}^{\pi} \log \psi(\lambda) d\lambda < \infty \), see Brockwell and Davis (1991). The two most prominent models of long memory Gaussian sequences, fractional Gaussian noise and farima-processes, satisfy this condition, see Examples 2.1.6 and 2.1.10.
The second assumption, piecewise monotonicity of the transformation \( G \), is also very weak and it is satisfied by nearly all examples that have been considered in the literature.

(iii) Finally, the assumption made on the block length is satisfied by any reasonable choice of \( l \) too.

**Proof of Theorem 14.** First consider the slightly modified estimator

\[
\hat{J}^F_n(x) = \frac{1}{n-l+1} \sum_{i=1}^{n-l+1} \frac{1}{d_i^2} \left( \sum_{j=i}^{i+l-1} \left\{ Y_j \leq x \right\} - F(x) \right)^2,
\]

where \( \tilde{F}_{n,l}(x) \) is replaced by \( F(x) \). Then

\[
\hat{J}^F_n(x) = \frac{J^2_n(x)}{(m!)^2} \frac{1}{n-l+1} \sum_{i=1}^{n-l+1} \frac{1}{d_i^2} \left( \sum_{j=i}^{i+l-1} H_m(X_j) \right)^2
\]

\[
+ \frac{1}{n-l+1} \sum_{i=1}^{n-l+1} \frac{1}{d_i^2} \left( \sum_{j=i}^{i+l-1} \left\{ Y_j \leq x \right\} - F(x) - J_m(x)/m!H_m(X_j) \right)^2
\]

\[
+ \frac{2}{m!} \frac{1}{n-l+1} \sum_{i=1}^{n-l+1} \frac{1}{d_i^2} \left( \sum_{j=i}^{i+l-1} \left\{ Y_j \leq x \right\} - F(x) - J_m(x)/m!H_m(X_j) \right) \left( \sum_{j=i}^{i+l-1} H_m(X_j) \right)
\]

\[= I_n(x) + II_n(x) + III_n(x).\]

By Lemma 3.2 of Lahiri (1993)

\[I_n(x) \xrightarrow{P} (m!)^{-2} J^2_m(x),\]

and this convergence is uniform due to boundedness of \( J_m(x) \). Moreover,

\[E[\sup_{x \in \mathbb{R}} II_n(x)] \leq E \left[ \sup_{x \in \mathbb{R}} \left( \frac{1}{d_l} \sum_{j=1}^{l} \left\{ Y_j \leq x \right\} - F(x) - J_m(x)/m!H_m(X_j) \right) \right]^2 = o(1),\]

by Corollary 2.2.1. Consequently \( II_n(x) = o_P(1) \), uniformly in \( x \), and using Cauchy-Schwarz inequality one obtains the same for \( III_n(x) \). Thus

\[\sup_{x \in \mathbb{R}} \left| \frac{1}{m!} J^2_n(x) - J_m(x)^2 \right| \xrightarrow{P} 0,\]

and it remains to verify

\[\sup_{x \in \mathbb{R}} \left| \hat{J}^F_n(x) - \text{Var}^*(W^*_n(x)) \right| \xrightarrow{P} 0. \tag{4.10}\]
To this end, obtain

\[
\left( \frac{l}{d_l} \right)^2 \left( \tilde{F}_{n,l}(x) - F(x) \right)^2 \leq 2 \left( \frac{l}{d_l} \right)^2 \left( F_n(x) - F(x) \right)^2 + 2 \left( \frac{l}{d_l} \right)^2 \left( \tilde{F}_{n,l}(x) - F_n(x) \right)^2.
\]

By Theorem 2 (Theorem 1.1 in Dehling and Taqqu (1989)) the first summand of the right hand side is \(O_P((ld_n)^2/(nd_l)^2)\) and consequently \(o_p(1)\). As for the second summand, note that

\[
\tilde{F}_{n,l}(x) - F_n(x) = \sum_{i=1}^n b_{n,i} \mathbb{1}_{\{Y_i \leq x\}},
\]

with

\[
b_{n,i} = \begin{cases} 
\frac{1}{n-i+1} - \frac{1}{n} & \text{for } i < l, \\
\frac{1}{n-l+1} - \frac{1}{n} & \text{for } l \leq i \leq n - l + 1, \\
\frac{1}{n-l+1} - \frac{1}{n} & \text{for } i > n - l + 1.
\end{cases}
\]

By the assumption made on the block length \((l = O(n^{1-\epsilon})\) with \(1 - \epsilon < 1/(2 - H)\)), one then obtains

\[
\frac{l}{d_l} \sup_{x \in \mathbb{R}} |\tilde{F}_{n,l}(x) - F_n(x)| \leq \frac{l}{d_l} \sum_{i=1}^n |b_{n,i}| \leq \frac{l}{d_l} \frac{l - 1}{n - l + 1} = o_p(1).
\]

Thus (4.10) is \(o_p(1)\) too, and this finishes the proof.

\[\square\]

Apparently \(\hat{J}_{n,m}(x)\) estimates \(|J_m(x)|\) instead of \(J_m(x)\). However, note that estimating \(J_m(x)\) is an ill-posed problem, at least if \(m\) is odd. For example, consider the two samples

\[G(X_1), \ldots, G(X_n) \quad \text{and} \quad G(-X_1), \ldots, G(-X_n),\]

where \((X_i)_{i=1}^n\) is a stationary Gaussian LRD process and \(G : \mathbb{R} \to \mathbb{R}\) is such that \(\{1_{\{G(\cdot) \leq x\}} - F(t), -\infty < x < \infty\}\) has odd Hermite rank \(m\). On the one hand, Gaussianity of the \(X_i\) implies

\[\{G(X_i)\}_{i=1}^n =^D \{G(-X_i)\}_{i=1}^n.\]

On the other hand, their Hermite coefficient functions differ, in detail

\[J_m(x) = E[1_{\{G(-X_i) \leq x\}} H_m(X_i)] = -E[1_{\{G(-X_i) \leq x\}} H_m(-X_i)] = -J_m(x).\]

Moreover, note that for Kolmogorov Smirnov and Cramér-von Mises test it suffices to estimate \(|J_m(x)|\).
4.1.3 Smoothed empirical process

For continuous distributions the true \( J_m(x) \) is often a smooth function. For example let \( G(x) = x \), then \( J_1(x) = -\phi(x) \), where \( \phi \) is the standard normal probability density. On the contrary, the estimator \( \hat{J}_{n,m}(x) \) is not even a continuous function. In what follows we will therefore consider the smoothed empirical process and its bootstrapped version.

Let \( H : \mathbb{R} \to \mathbb{R} \) be any continuous distribution function and \((h_n)_{n \in \mathbb{N}}\) a positive, possibly random sequence converging to 0 as \( n \) goes to infinity. The smoothed empirical distribution function is then defined by

\[
\tilde{F}_{n,H}(x) = \frac{1}{n} \sum_{i=1}^{n} H \left( \frac{x - Y_i}{h_n} \right) = \int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dF_n(y).
\]

Analogously the smoothed empirical process is defined by

\[
\tilde{W}_{n,H}(x) = \int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dW_n(y).
\] (4.11)

For independent data, weak convergence of (4.11) was obtained by van der Vaart (1994) and Yukich (1992). They considered empirical processes indexed by Donsker-classes \( F \). Radulovic and Wegkamp (2000) extended these results beyond Donsker-classes and also considered a bootstrapped version. To the best of our knowledge, neither the smoothed empirical process, nor its bootstrap has been treated in the long memory scenario.

Let \( Y_1^{*(j)}, \ldots, Y_p^{*(j)} \), be the \( j \)-th block-bootstrap sample. Then define

\[
\tilde{W}_{n,H}^{*(j)}(x) = \frac{1}{dpl^{1/2}} \left( \sum_{j=1}^{pl} H \left( \frac{x - Y_i^{*(j)}}{h_n} \right) - \sum_{i=1}^{p} \sum_{j \in B_i} H \left( \frac{x - Y_i}{h_n} \right) \right)
\]

\[
= \int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dW_{n}^{*(j)}(y).
\]

The next theorem states weak convergence of the (bootstrapped) smoothed empirical process.

**Theorem 15.** Let the conditions of Theorem 13 hold and let \( F \) be continuous on \( \mathbb{R} \). If \( H \) is a continuous distribution function and \( h_n > 0 \) with \( h_n = o_P(1) \), then

\[
\tilde{W}_{n,H}^{*}(x) \overset{D}{\to} \frac{J_m(x)}{m!} Z \quad \text{in probability.}
\]

\( Z \) is standard normal and weak convergence takes place in \( D([\infty, \infty]) \), equipped with the
weak convergence of the bootstrapped empirical process uniform topology. Moreover:
\[
\left( \tilde{W}_n(x), \tilde{W}_n^{*}(1)(x), \ldots, \tilde{W}_n^{*}(J)(x) \right) \xrightarrow{D} \frac{J_m(x)}{m!} \left( Z_m, Z^{(1)}, \ldots, Z^{(J)} \right),
\]
where the convergence takes place in \((D([-\infty, \infty]))^{J+1}\), equipped with the uniform topology. \(Z_m\) is defined as above and \(Z^{(j)}\) are independent standard normal random variables, independent of \(Z_m\).

**Remark 4.1.7.**

(i) There are no restrictions on the bandwidth \(h_n\) besides being positive and \(o_P(1)\). Therefore, it might be data-driven.

(ii) Again, by weak convergence in probability we mean that the metric \(d_{BL}(\cdot, \cdot)\) of the two distributions converges in probability to zero.

**Proof of Theorem 15.**

Once more we will make use of a reduction principle. Therefore consider
\[
\tilde{W}_n^{*}(x) = \int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dW_n^{*}(y)
= S_n^*(1/m!) \int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dJ_m(y) + \int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dR_n^*(y),
\]
where \(S_n^* = \frac{1}{d^{p/2}} \sum_{i=1}^{pl} (H_m(X_i^*) - E^*[H_m(X_i^*)])\) and
\[
R_n^{*(j)}(x) = \frac{1}{d^{p/2}} \sum_{i=1}^{pl} \left( 1 \{ Y_i^{*(j)} \leq x \} - \frac{1}{m} \sum_{i=1}^{pl} H_m(X_i^*) - \frac{1}{m!} \left( H_m(X_i^{*(j)}) - E^*[H_m(X_i^{*(j)})] \right) \right),
\]
which does not depend on \(h_n\). Note that for fixed \(x\) and \(h_n\) the function \(H_{h_n,x}(y) = H((x - y)/h_n)\) is monotone decreasing. Integration by parts then yields
\[
\int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dR_n^{*(j)}(y) = - \int_{\mathbb{R}} R_n^{*(j)}(y) dH_{h_n,x}(y)
\leq \sup_{y \in \mathbb{R}} |R_n^{*(j)}(y)| \xrightarrow{P} 0,
\]
where the convergence in probability is due to Lemma 4.1.3 (the bootstrap reduction principle). Moreover, one gets for \(n \to \infty\)
\[
\int_{\mathbb{R}} H \left( \frac{x - y}{h_n} \right) dJ_m(y) \xrightarrow{P} J_m(x),
\]
uniformly in \(x\), which can be seen as follows. Denote by \(\xi\) a random variable, such that \(-\xi\)
has distribution function $H(y)$. Moreover, define the function $f: (0, \infty) \to \mathbb{R}$ by
\[
f(h) = \sup_{x \in \mathbb{R}} \left| \int_{\mathbb{R}} H \left( \frac{x - y}{h} \right) dJ_m(y) - J_m(x) \right|.
\]
Then, as $h \searrow 0$, we get from uniform continuity and boundedness of $J_m(x)$
\[
f(h) = \sup_x \left| - \int J_m(y) dH_{h,x}(y) - J_m(x) \right| = \sup_x |E[J_m(x - h_n \xi) - J_m(x)]| \to 0.
\]
Consequently, $f(h) = o_P(1)$ and (4.13) holds. Combining (4.12) and (4.13) then yields
\[
\hat{W}_{n,H}^*(x) = \frac{J_m(x)}{m!} S_n^* + o_P(1),
\]
uniformly in $x$. Now the conditional as well as the unconditional result can be obtained analogously to the proof of Theorem 13.

Motivated by Theorem 15 we define the estimator
\[
\tilde{J}_{n,m}(x) = \left( Var^*(\hat{W}_n^*) \right)^{1/2} = \frac{1}{n - l + 1} \sum_{i=1}^{n-l+1} \hat{W}_{n.l,i}(x)^2.
\]
Consistency of this estimator might be proven along the same lines as consistency of the non-smoothed estimator $\hat{J}_{m,n}(x)$.

**Corollary 4.1.8.** Let the conditions of Theorem 14 hold. If $F$ and $H$ are continuous and $h_n$ is positive and $o_P(1)$, then
\[
\sup_{x \in \mathbb{R}} |\tilde{J}_{n,m}(x) - m! |J_m(x)| \xrightarrow{P} 0.
\]

In Figure 4.1 we compare $\tilde{J}_{n,m}(x)$ to the true function $J_m(x)$ for different transformations $G(x)$. For the smoothing we have used the R-package **kerdiest**, see also Quintela-del-Rio and Estevez-Perez (2012). There are lots of possible combinations of choices for the smoothing bandwidth $h_n$, the block length $l$ and the number of drawn blocks $p$. For the bandwidth selection we have used the data driven method of Quintela-del-Rio and Estevez-Perez (2012), moreover $l = \lfloor n^{1/2} \rfloor$ and $p = \lfloor n/l \rfloor$.

### 4.2 Applications to statistical testing

#### 4.2.1 Testing for monotonicity of transformations

The function $\hat{J}_{m,n}(x)$ (which we denote here by $\hat{J}_{m,n}^{(1)}(x)$) consistently estimates $|J_m(x)|$ for all choices of transformations $G$. Now assume that the data come from a monotone transformation
of a Gaussian process. This assumption makes statistical inference for LRD processes much more straightforward, see for example the estimation of the Hurst coefficient below.

In order to test for monotonicity we will make use of the results from Section 4.1.2 and the following proposition. Let $X \sim N(0, 1)$, $F(x) = P(G(X) \leq x)$ and $\phi$ and $\Phi$ denote probability density and distribution function of $X$, respectively.

**Proposition 4.2.1.** (i) If $G$ is a monotone increasing (monotone decreasing) function, then

$$G(x) = F^{-1}(\Phi(x)) \quad (G(x) = F^{-1}(\Phi(-x)))$$

for almost all $x \in \mathbb{R}$. $F^{-1}(x)$ is the generalized inverse of $F(x)$. 

---

Figure 4.1: Comparison of $\tilde{J}_{n,m}(x)$ and the true function $J_m(x)$ (solid line) for $G(x) = x$ (first row), $G(x) = \exp(x)$ (second row) and $G(x) = x^2$ (third row). The Hurst parameter is set to $H = 0.6$ (left), $H = 0.7$ (center) and $H = 0.8$ (right) and the sample size is $n = 500$. The Hurst coefficient is assumed to be known (dotted line) or estimated using the local Whittle estimator (dashed line).
(ii) Let $G$ be such that $G(X)$ has a continuous distribution function. Then

$$J_1(x) = -\phi(\Phi^{-1}(F(x))) \quad (J_1(x) = \phi(\Phi^{-1}(F(x)))),$$

if and only if $G$ is almost everywhere monotone increasing (monotone decreasing).

**Proof of Proposition 4.2.1.** We will only prove the parts concerning monotone increasing functions. The results for monotone decreasing functions will follow immediately by considering $G^*(x) = G(-x)$.

For part (i) let $x \in \mathbb{R}$ and $X \sim N(0, 1)$ and recall the definition of the generalized inverse

$$F^{-1}(z) = \inf\{y \in \mathbb{R} \mid F(y) \geq z\}.$$

By the monotonicity of $G$

$$F(G(x)) = P(G(X) \leq G(x)) > P(X \leq x) = \Phi(x).$$

But this implies $G(x) \geq F^{-1}(\Phi(x))$, due to the definition of the generalized inverse.

Now let $x$ be a point of continuity of $G$ and choose $y \in \mathbb{R}$ such that $F(y) \geq \Phi(x)$. We will show that this implies $y \geq G(x)$.

Assume the contrary, namely $G(x) > y$. Then, by continuity of $G$ in $x$,

$$F(y) = P(G(X) \leq y) = P(X \leq \sup\{s \in \mathbb{R} \mid G(s) \leq y\}) < P(X \leq x) = \Phi(x),$$

which contradicts the assumption made on $y$. We have just shown that for all $x$

$$F(y) \geq \Phi(x) \Rightarrow y > G(x).$$

But this implies $G(x) \leq F^{-1}(\Phi(x))$, again by the definition of the generalized inverse. Due to monotonicity, the discontinuity points of $G$ have Lebesgue measure 0, and this finishes the proof of (i).

The “if”-part of (ii) follows directly from (i). For the “only-if”-part define

$$\tilde{G}(x) = \Phi^{-1}(F(G(x))).$$

Then, by the assumption made on $J_1(x)$ and the continuity of $F$, we have

$$\tilde{J}_1(x) = E1_{\{G(x) \leq x\}}X = J_1(F^{-1}(\Phi(x))) = -\phi(x).$$
Thus $\tilde{G}(X) \sim N(0, 1)$ and its Hermite coefficient function is $\tilde{J}_1(x) = -\phi(x)$. The latter implies

$$E\tilde{G}(X)X = -\int \tilde{J}_1(x) \, dx = 1.$$ 

Together with $\tilde{G}(X) \sim N(0, 1)$ this yields

$$\tilde{G}(x) = x,$$

almost everywhere. By definition of $\tilde{G}$ we get

$$F(G(x)) = \Phi(x).$$

Then $G$ must be monotone increasing almost everywhere. □

Part (ii) of Proposition 4.2.1 motivates another estimator for $J_1(x)$, namely

$$\hat{J}^{(2)}_{n,1} = \phi(\Phi^{-1}(\hat{F}_n(x))),$$

where $\hat{F}_n(x)$ is the empirical distribution function. However, this estimator is only consistent (again up to its sign) if the observations come from a monotone transformation of Gaussian data. This is illustrated in Figure 4.2, where both estimators are applied to $\{G_0(X_i)\}_{i=1}^n$ with

$$G_0(x) = \begin{cases} x & \text{for } |x| \geq 1, \\ -x & \text{for } |x| < 1. \end{cases}$$

Figure 4.2: Comparison of the different estimators $\hat{J}^{(1)}_{n,1}(x)$ (solid line) and $\hat{J}^{(2)}_{n,1}(x)$ (dashed line) together with the true function $J_1(x)$ (dotted line). Both are applied to $G_0(X_i)$, where $(X_i)_i$ is fractional Gaussian noise with Hurst-coefficient $H = 0.7$ and sample size $n = 500, 1000, 2500$ (from left to right).

The sequence $\{G_0(X_i)\}_{i\geq1}$ is non-Gaussian, but its marginals are standard normal, meaning
\(P(G_0(X_i) \leq x) = \Phi(x)\). We will now use Proposition 4.2.1 to investigate the test problem:

\[H: G \text{ is either monotone increasing or monotone decreasing;}
\]

vs. \(A: G \text{ is neither monotone increasing nor monotone decreasing.}\)

As test statistic we will use an \(L^2\) difference between the two estimators, in detail

\[M_n = \int \left( \hat{j}^{(1)}_{n,1}(x) - \hat{j}^{(2)}_{n,1}(x) \right)^2 d\hat{F}_n(x).\]

A direct consequence of Proposition 4.2.1 and the Glivenko-Cantelli theorem is \(M_n \xrightarrow{P} 0\), if the hypothesis holds. Under the alternative

\[M_n \xrightarrow{P} \int \left( J_1(x) - \phi(\Phi^{-1}(F(x))) \right)^2 dF(x),\]

with \(J_1(x) \neq \phi(\Phi^{-1}(F(x)))\). In order to develop an ordinary test one needs a different normalization of \(M_n\) and an asymptotic distribution. However, this seems quite challenging and it is not clear if such a distribution exists. We therefore try to simulate the law of \(M_n\) by simulating Gaussian time series \((X^*_i)_{i \geq 1}\) (fractional Gaussian noise in the simulation study). For each series, a statistic \(M^*_n\) is computed and eventually critical values are deduced from empirical quantiles.

A crucial point is the right strength of dependence of the simulated processes. Under the hypothesis \((G(x) \text{ is monotone})\) one might estimate the original Gaussian sequence \((X_i)_{i \geq 1}\) by

\[\hat{X}_i = \Phi^{-1}(\hat{F}_n(G(X_i))) \quad i = 1, \ldots, n.\]  

(4.1)

We then estimate the Hurst coefficient of \((\hat{X}_i)_{i \geq 1}\) as in the previous chapter (local Whittle estimator) and simulate \(X^*_i\) as fractional Gaussian noise with parameter \(\hat{H}\).

There might be more sophisticated types of parametric bootstrap (see for example Andrews et al. (2006)), but our method seems to be quite satisfying for finite samples.

We considered several transformations of a farima process and a comparison of empirical size (Table 4.1) and power (Table 4.2) shows that the test actually might discriminate between hypothesis and alternative. Then again, the test is highly sensitive to the true Hurst coefficient, as it fails for small values of \(H\), here \(H = 0.6\). The choice of the block length \(l\) seems to be of minor importance.

A great advantage of a monotone transformation is the possibility of estimating the underlying \((X_i)_{i \geq 1}\) via (4.1). Eventually, statistical inference might be made based upon \((\hat{X}_i)_{i \geq 1}\) instead of \((Y_i)_{i \geq 1}\). For example, in what follows we modify the classical local Whittle estimator by adding a test for monotonicity. If the hypothesis is not rejected, \(H\) is estimated using \((\hat{X}_i)_{i \geq 1}\). Figure (4.3) shows that this modification clearly reduces the bias of the estimator when the
Table 4.1: Empirical size of the test for monotonicity of $G$, with $Y_i = X_i$ being a Gaussian farima process. The nominal size is $\alpha = 0.05$ and critical values are deduced from $999$ iterations of $M^*_n$.

<table>
<thead>
<tr>
<th></th>
<th>$H = 0.6$</th>
<th>$H = 0.7$</th>
<th>$H = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 500$</td>
<td>$l = 22$</td>
<td>0.447</td>
<td>0.121</td>
</tr>
<tr>
<td>$l = 25$</td>
<td>0.460</td>
<td>0.115</td>
<td>0.010</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>$l = 31$</td>
<td>0.378</td>
<td>0.056</td>
</tr>
<tr>
<td>$l = 35$</td>
<td>0.368</td>
<td>0.054</td>
<td>0.004</td>
</tr>
<tr>
<td>$n = 2500$</td>
<td>$l = 50$</td>
<td>0.295</td>
<td>0.029</td>
</tr>
<tr>
<td>$l = 55$</td>
<td>0.303</td>
<td>0.027</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 4.2: Empirical power of the test for monotonicity of $G$, with several transformations being applied to Gaussian farima processes. The nominal size is $\alpha = 0.05$ and critical values are deduced from $999$ iterations of $M^*_n$.

<table>
<thead>
<tr>
<th></th>
<th>$H = 0.6$</th>
<th>$H = 0.7$</th>
<th>$H = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 500$</td>
<td>$l = 22$</td>
<td>0.432</td>
<td>0.474</td>
</tr>
<tr>
<td>$l = 25$</td>
<td>0.451</td>
<td>0.499</td>
<td>0.443</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>$l = 31$</td>
<td>0.420</td>
<td>0.449</td>
</tr>
<tr>
<td>$l = 35$</td>
<td>0.436</td>
<td>0.430</td>
<td>0.394</td>
</tr>
<tr>
<td>$n = 2500$</td>
<td>$l = 50$</td>
<td>0.470</td>
<td>0.381</td>
</tr>
<tr>
<td>$l = 55$</td>
<td>0.470</td>
<td>0.385</td>
<td>0.307</td>
</tr>
</tbody>
</table>

data come from a monotone transformation, while for non-monotone $G$ there is no negative impact.

Figure 4.3: Boxplots of the two different local Whittle estimators for $1000$ time series of length $n = 1000$. The true Hurst coefficient is $H = 0.7$.

Finally, we would like to mention the test of Beran et al. (2016). The authors suggest a
procedure to test for the Hermite rank of $G$ being equal to 1, which is a more general hypothesis than the transformation $G$ being monotone. Moreover their test is also based on the bootstrap, but in a completely opposite way. They are using the fact that under long memory the bootstrap fails for all Hermite ranks larger than 1 and therefore under their alternative. On the contrary our test profits from the consistency of the bootstrap-based estimator even under the alternative.

![Figure 4.4: Ethernet traffic data in bytes per second at a LAN at Bellcore (see the R-package longmemo). The comparison of the estimators $\hat{J}_{n,1}^{(1)}(x)$ (solid line) and $\hat{J}_{n,1}^{(2)}(x)$ (dashed line) suggests that only the upper tail of the data might be modeled as monotone transformation of a Gaussian. However, our test does not reject the hypothesis.](image)

### 4.2.2 Goodness-of-fit tests

A typical application of bootstrapped empirical processes are goodness-of-fit tests. When testing the hypothesis $P(Y_i \leq x) = F_0(x)$ against $P(Y_i \leq x) \neq F_0(x)$, two classical statistics are the Kolmogorov-Smirnov and Cramér-von Mises statistic. They are given by

$$T_{n,1} = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

and

$$T_{n,2} = \int_{\mathbb{R}} (F_n(x) - F(x))^2 dF(x),$$

respectively. The Kolmogorov-Smirnov statistic has been investigated in Beran and Ghosh (1990, 1991) and Koul and Surgailis (2010). However, a bootstrap approximation has not been applied, yet. The asymptotic distributions can be obtained directly from (4.1):

$$\frac{n}{d_n} T_{n,1} \xrightarrow{D} (ml)^{-1} \sup_{x \in \mathbb{R}} |J_m(x)||Z_m|$$

and

$$\frac{n}{d_n^2} T_{n,2} \xrightarrow{D} (ml)^{-2} \int_{\mathbb{R}} (J_m(x))^2 dF(x)||Z_m|^2.$$

Both limiting distributions depend on the function $|J_m(x)|$. Hence for all Hermite ranks $m \geq 1$ one might replace $J_m(x)$ by $\hat{J}_{n,m}(x)$ (or the smoothed version $\tilde{J}_{n,m}(x)$). Then (asymptotically) valid critical values are obtained by the appropriate quantiles of $|Z_m|$. 
In the important case $m = 1$ there is a second way of deriving critical values, namely by directly bootstrapping the statistics. In detail, consider

$$T_{n,1}^* = \sup_{x \in \mathbb{R}} |W_n^*(x)| \quad \text{and} \quad T_{n,2}^* = \int_{\mathbb{R}} (W_n^*(x))^2 dF(x).$$

Next, one might produce $J$ bootstrap iterations and denote by $T_{n,i}^{*(j)}$ the associated goodness-of-fit statistics and by $Q_{n,J,i}(1 - \alpha)$ their empirical quantiles. The following result is then an immediate consequence of Theorem 13. It states that the bootstrap consistently estimates the distribution of the two statistics and that critical values are asymptotically valid.

**Corollary 4.2.2.** (i) Let the conditions of Theorem 13 hold and $m = 1$. For $i = 1, 2$ one has

$$\sup_{x \in \mathbb{R}} |P(T_{n,i} \leq x) - P^*(T_{n,i}^* \leq x)| \overset{P}{\rightarrow} 0.$$

(ii) Moreover, for $\alpha \in (0, 1)$ and for $i = 1, 2$ one has

$$\lim_{J \to \infty} \lim_{n \to \infty} P(T_{n,i} > Q_{n,J,i}(1 - \alpha)) = \alpha.$$

**Proof of Corollary 4.2.2.** Part (i) is a direct consequence of the conditional result in Theorem 13. Part (ii) follows from the unconditional result in Theorem 13 and Proposition 2.4.2 (see also Proposition F.1 in the supplement to Bücher and Kojadinovic (2016)).

In a small simulation study we investigate the empirical properties of the two resampling methods when combined with the Kolmogorov-Smirnov test. Note that convergence of the empirical process towards $J_m(x)/m!Z_m$ is quite slow. Therefore, deducing critical values from the quantile of $\sup_x |\hat{J}_{m,n}(x)/m!Z_m|$ leads to inaccurate results even for $n = 10000$. Hence this method is omitted in the following.

On the contrary, using quantiles of the bootstrap distribution seems to be promising, see Table 4.3. Here the empirical size of the Kolmogorov-Smirnov (KS) test is displayed for different sample sizes and Hurst coefficients. The latter is assumed to be unknown and estimated using the local Whittle estimator of Künsch (1987). The correct normalization of the empirical process (and therefore of the KS statistic) is given through $C_{m,H}L^{m/2}(n)n^H$, where $C_{m,H}$ is the constant of proportionality from Lemma 2.1.4. Here we assume $L(n) \to C$ and the combined constant $C_0 = C_{m,H}C$ is estimated using $\hat{H}$ and the empirical autocorrelation function (for details see the numerical results in Section 3.5.3). Eventually, we normalize the KS statistic by $\hat{C}_0n^{\hat{H}}$.

For small $n$ the test is a bit oversized, however for $n = 500$ empirical and nominal size already match. The problem is with the large samples. As Figure 4.5 and Table 4.3 indicate, the empirical size decreases with $n$ and falls clearly beneath the nominal size, here $\alpha = 0.05$. 

Table 4.3: Empirical size of the Kolmogorov-Smirnov test for $H: F(x) = \Phi(x)$. The used resampling methods are combined bootstrap / block bootstrap / sampling window with $J = 499$ bootstrap iterations. The nominal size is $\alpha = 0.05$

<table>
<thead>
<tr>
<th>Count</th>
<th>$l$</th>
<th>Block bootstrap</th>
<th>$H = 0.6$</th>
<th>$H = 0.7$</th>
<th>$H = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 50$</td>
<td>$l = 7$</td>
<td>$0.134 / 0.119 / 0.191$</td>
<td>$0.172 / 0.139 / 0.201$</td>
<td>$0.209 / 0.152 / 0.184$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$l = 8$</td>
<td>$0.120 / 0.105 / 0.161$</td>
<td>$0.166 / 0.122 / 0.184$</td>
<td>$0.231 / 0.154 / 0.201$</td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>$l = 10$</td>
<td>$0.124 / 0.110 / 0.149$</td>
<td>$0.158 / 0.105 / 0.154$</td>
<td>$0.158 / 0.095 / 0.131$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$l = 12$</td>
<td>$0.103 / 0.088 / 0.142$</td>
<td>$0.130 / 0.088 / 0.133$</td>
<td>$0.156 / 0.086 / 0.141$</td>
<td></td>
</tr>
<tr>
<td>$n = 500$</td>
<td>$l = 22$</td>
<td>$0.068 / 0.048 / 0.051$</td>
<td>$0.073 / 0.037 / 0.057$</td>
<td>$0.098 / 0.055 / 0.047$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$l = 25$</td>
<td>$0.086 / 0.059 / 0.049$</td>
<td>$0.069 / 0.043 / 0.055$</td>
<td>$0.097 / 0.056 / 0.055$</td>
<td></td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>$l = 31$</td>
<td>$0.061 / 0.038 / 0.041$</td>
<td>$0.066 / 0.035 / 0.044$</td>
<td>$0.086 / 0.047 / 0.034$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$l = 35$</td>
<td>$0.049 / 0.032 / 0.036$</td>
<td>$0.069 / 0.035 / 0.037$</td>
<td>$0.084 / 0.046 / 0.044$</td>
<td></td>
</tr>
<tr>
<td>$n = 5000$</td>
<td>$l = 70$</td>
<td>$0.043 / 0.021 / 0.028$</td>
<td>$0.044 / 0.021 / 0.031$</td>
<td>$0.051 / 0.028 / 0.019$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$l = 75$</td>
<td>$0.040 / 0.020 / 0.022$</td>
<td>$0.044 / 0.022 / 0.020$</td>
<td>$0.046 / 0.027 / 0.034$</td>
<td></td>
</tr>
</tbody>
</table>

Not only stands this in clear contrast to the short range dependent case. Doukhan et al. (2015) consider different bootstrap methods and due to underestimating covariances the empirical size is always larger than the nominal. But also, our findings indicate that the asymptotic calculus of Corollary 4.2.2 does not hold even for large sample sizes as $n = 10000$. This should lead to a painful loss of power.

Figure 4.5: Empirical size of the KS test as function of $n$ using block bootstrap (dashed line), the sampling window method (dotted line) and the combined bootstrap (solid line) for $H = 0.6$ (left) and $H = 0.8$ (right).

We therefore suggest a new resampling method, which is in fact a combination of the former two. Consider

$$\tilde{T}_{n,1} = \sup_{x \in \mathbb{R}} |\delta W_n^*(x) + \sqrt{1 - \delta^2 \tilde{J}_{1,n}(x)} \tilde{Z}|,$$
where $W^*_n$ is the bootstrapped empirical process, $\hat{J}_{1,n}$ the estimator from (4.14), $\delta \in [0, 1]$ and $\tilde{Z} \sim N(0, 1)$, independent of $(X_i)_{i \geq 1}$ and independent of the resampling mechanism. Here we call it combined bootstrap. Critical values are deduced from Monte Carlo simulation, just as in classical bootstrap procedures. Due to the independence of $Z$, Theorem 14 and Corollary 4.2.2, $\tilde{T}_{n,1}^*$ converges weakly towards $\sup_x |J_1(x)Z|$, hence the critical values are asymptotically valid.

Moreover, here the asymptotic calculus holds for much smaller sample sizes as it was the case for the classical block bootstrap, see Figure 4.5 and Table 4.3. For $\delta$ we have chosen $\delta = \delta_n = (l/n)^{H-1/2}$. The reason is quite heuristic, however, we think that future work on a second order approximation of the bootstrapped empirical process might yield a theoretic justification.

Finally, we have compared our approach to the sampling window method. For long-range dependent processes and general statistics its validity has been proved by Betken and Wendler (2017+). For small $n$ it is less accurate than the block bootstrap and for large $n$ it shows the same flaws. Thus for the KS test our methods seem advantageous, as long as the Hermite rank equals 1.

### 4.3 Proof of Lemma 4.1.3

Throughout the proofs we use the notation:

\[
R_n(x, y) = R_n(y) - R_n(x), \quad F(x, y) = F(y) - F(x)
\]
\[
\tilde{F}_{n,l}(x, y) = \tilde{F}_{n,l}(y) - \tilde{F}_{n,l}(x), \quad J_m(x, y) = J_m(y) - J_m(x).
\]

**Lemma 4.3.1** (Dehling, Taqqu). There exists constants $\gamma > 0$ and $C > 0$ such that for all $n \in \mathbb{N}$

\[
E|R_n(x, y)|^2 \leq Cn^{-\gamma}(F(y) - F(x)).
\]

The next result is Lemma 3.1. of Lahiri (1993).

**Lemma 4.3.2** (Lahiri). Define $\tilde{\mu}_{n,l}(H_m)$ as in (4.6). If the conditions of Theorem 13 hold,

(i) $\tilde{\mu}_{n,l}(H_m) = o_P(d_l/l)$ and

(ii) $E[|\mu_{n,l}(H_m)|^2] \leq Cn^2/\sigma^2$.

The next lemma extends the previous one to indicator functions.

**Lemma 4.3.3.** Define $\tilde{F}_{n,l}(x)$ as in (4.6). If the conditions of Theorem 13 hold,

\[
E \left( F(x, y) - \tilde{F}_{n,l}(x, y) \right)^2 \leq Cn^2/\sigma^2 F(x, y).
\]
Proof. Since the Hermite rank equals \( m \) we obtain the following expansion:

\[
1_{x < Y_j \leq y} - F(x, y) \overset{L^2}{=} \sum_{q=m}^{\infty} J_q(x, y)/q! H_q(X_i).
\]

By definition of \( \tilde{F}_{n,l}(x) \) we have

\[
F(x) - \tilde{F}_{n,l}(x) = F(x) - \frac{1}{l} \frac{1}{(n - l + 1)} \sum_{j=1}^{n} a_{n,j} 1_{Y_j \leq x}
= \frac{1}{l} \frac{1}{(n - l + 1)} \sum_{j=1}^{n} a_{n,j} (F(x) - 1_{Y_j \leq x}),
\]

where

\[
a_{n,j} = \begin{cases} 
    j, & \text{if } j < l, \\
    l, & \text{if } l \leq j \leq n - l + 1, \\
    n - j + 1 & \text{if } j > n - l + 1.
\end{cases}
\]

Note that \( a_{n,j} \leq l \) for all \( j \). By orthogonality of the \( H_q(X_i) \),

\[
\sum_{q=m}^{\infty} J_q^2(x, y)/q! \leq F(x, y)
\]

and moreover

\[
E \left( F(x, y) - \tilde{F}_{n,l}(x, y) \right)^2 = \frac{1}{l^2} \frac{1}{(n - l + 1)^2} \sum_{q=m}^{\infty} J_q^2(x, y) \frac{1}{q!} \sum_{i,j \leq n} a_{n,i} a_{n,j} E[H_q(X_i) H_q(X_j)]
\leq \frac{1}{(n - l + 1)^2} F(x, y) \sum_{i,j \leq n} |r(i - j)|^m.
\]

The conclusion then follows because of \( d_n^2 \sim \sum_{i,j \leq n} |r(i - j)|^m \).

\[ \square \]

**Lemma 4.3.4.** Using the moving blocks bootstrap mechanism described above, one gets

\[
E^* [R_{n,l}^*(x, y)]^2 \leq \frac{1}{(n - l + 1)^2} C \sum_{i=1}^{n-1+l} R_{i,l}^2(x, y)
+ \frac{1}{d_l^2} C l^2 \left( F(x, y) - \tilde{F}_{n,l}(x, y) \right)^2
+ \frac{1}{d_l^2} C J_m^2(x, y)/(m!)^2 l^2 \left( \tilde{\mu}_{n,l}(H_m) \right)^2,
\]
where $\mu_{n,l}(H_m)$ and $\tilde{F}_{n,l}(x,y)$ are defined as in (4.6) and

$$R_{i,l}(x) = \frac{1}{d_l} \sum_{j \in B_i} (1_{Y_j \leq x} - F(x) - J_m(x)/m! H_m(X_j)).$$

**Proof.** The results follows directly from the construction of the bootstrap sample, in detail

$$E^*|R_{n,l}^*(x)|^2 = \frac{1}{d_l^2} \left( \sum_{j=1}^{kl} (1_{1 \{Y_j \leq x\} - \tilde{F}_{n,l}(x) - J_m(x)/m! H_m(X_j) - \tilde{\mu}_{n,l}(H_m)) \right)^2$$

$$= \frac{1}{d_l^2} \left( \sum_{j \in B_i} (1_{1 \{Y_j \leq x\} - \tilde{F}_{n,l}(x) - J_m(x)/m! H_m(X_j) - \tilde{\mu}_{n,l}(H_m)) \right)^2$$

$$= \frac{1}{d_l^2} \left( \sum_{i=1}^{n-l+1} \left( \sum_{j \in B_i} (1_{1 \{Y_j \leq x\} - \tilde{F}_{n,l}(x) - J_m(x)/m! H_m(X_j) - \tilde{\mu}_{n,l}(H_m)) \right)^2$$

$$\leq \frac{1}{d_l^2} \left( \sum_{i=1}^{n-l+1} \left( \sum_{j \in B_i} (F(x) - \tilde{F}_{n,l}(x)) \right)^2$$

$$+ \frac{\ell^2}{d_l^2} \left( \sum_{j \in B_i} (F(x) - \tilde{F}_{n,l}(x)) \right)^2$$

$$+ \frac{\ell^2}{d_l^2} \left( \sum_{j \in B_i} (F(x) - \tilde{F}_{n,l}(x)) \right)^2$$

$$\leq \frac{1}{d_l^2} \left( \sum_{i=1}^{n-l+1} \left( \sum_{j \in B_i} (F(x) - \tilde{F}_{n,l}(x)) \right)^2$$

$$+ \frac{\ell^2}{d_l^2} \left( \sum_{j \in B_i} (F(x) - \tilde{F}_{n,l}(x)) \right)^2$$

$$+ \frac{\ell^2}{d_l^2} \left( \sum_{j \in B_i} (F(x) - \tilde{F}_{n,l}(x)) \right)^2.$$

\[\square\]

**Proof of Lemma 4.1.3.** We will prove the result by using exactly the same chaining points as in Dehling and Taqqu (1989). Define

$$\Lambda(x) := F(x) + \int_{\{G(s) \leq x\}} \frac{|H_m(s)|}{m!} \phi(s) \, ds.$$

The function $\Lambda$ is monotone, $\Lambda(-\infty) = 0$, $\Lambda(+\infty) < \infty$ and $\max\{F(x,y), J_m(x,y)/m!\} \leq \Lambda(y) - \Lambda(x)$.

Define for $k = 0, 1, \ldots, K$ refining partitions of $\mathbb{R}$,

$$-\infty = x_0(k) \leq x_1(k) \leq \cdots \leq x_{2^k}(k) = \infty,$$

by

$$x_i(k) = \inf\{x \in \mathbb{R} \mid \Lambda(x) \geq \Lambda(+\infty) i 2^{-k}\}, \quad i = 0, 1, \ldots, 2^k - 1,$$
with $K$ to be chosen later. We obtain the important estimate

$$\Lambda(x_i(k)) - \Lambda(x_{i-1}(k)) \leq \Lambda(+\infty)2^{-k}.$$ 

Based on these partitions we can define chaining points $i_k(x)$ by

$$x_{i_k(x)}(k) \leq x < x_{i_k(x)+1}(k),$$

for each $x$ and each $k \in \{0, 1, \ldots, K\}$, see Dehling and Taqqu (1989). By doing so every point $x$ is linked to $-\infty$, in detail

$$-\infty = x_{i_0(x)}(0) \leq x_{i_1(x)}(1) \leq \cdots \leq x_{i_K(x)}(K) \leq x.$$ 

Writing $R_{n,l}^*(x, y) = R_{n,l}^*(y) - R_{n,l}^*(x)$ we obtain the decomposition

$$R_{n,l}^*(x) = \sum_{i=0}^{K-1} R_{n,l}^*(x_{i_k(x)}(k), x_{i_k+1(x)}(k+1)) + R_{n,l}^*(x_{i_K(x)}(K), x). \quad (4.1)$$ 

Let us first consider the last term of $(4.1)$.

$$|R_{n,l}^*(x_{i_K(x)}(K), x)|$$

$$= \left| d_l^{-1} p^{-1/2} \sum_{j=1}^{pl} \left( 1_{\{x_{i_K(x)}(K) < Y^*_j \leq x\}} - \tilde{F}_{n,l}(x_{i_K(x)}(K), x) \right. \right.$$ 

$$\left. - \frac{1}{m!} J_m(x_{i_K(x)}(K), x) (H_m(X^*_j) - \tilde{\mu}_{n,l}(H_m)) \right) \right|$$

$$\leq d_l^{-1} p^{-1/2} \sum_{j=1}^{pl} \left( 1_{\{x_{i_K(x)}(K) < Y^*_j \leq x\}} + \tilde{F}_{n,l}(x_{i_K(x)}(K), x) \right)$$

$$+ \left| \frac{1}{(m)!} J_m(x_{i_K(x)}(K), x) d_l^{-1} p^{-1/2} \sum_{j=1}^{pl} (H_m(X^*_j) - \tilde{\mu}_{n,l}(H_m)) \right|$$

$$\leq |R_{n,l}^*(x_{i_k(x)}(K), x_{i_k(x)+1}(K))|$$

$$+ 2pl d_l^{-1} p^{-1/2} \tilde{F}_{n,l}(x_{i_K(x)}(K), x_{i_K(x)+1}(K))$$

$$+ 2\Lambda(\infty)2^{-K} d_l^{-1} p^{-1/2} \sum_{j=1}^{pl} (H_m(X^*_j) - \tilde{\mu}_{n,l}(H_m))$$
Proof of Lemma 4.1.3

\[ P^* \left( \sup_x |R_{n,l}^*(x)| > \epsilon \right) \]
\[ \leq P^* \left( \sup_x |R_{n,l}^*(x)| > \epsilon \sum_{k=0}^{K-1} (k + 3)^{-2} + \epsilon/2 \right) \]
\[ \leq \sum_{k=0}^{K-1} P^* \left( \max_x |R_{n,l}^*(x_k(k), x_{i_k+1}(k+1))| > \epsilon/(k + 3)^2 \right) \]
\[ + P^* \left( \max_x |R_{n,l}^*(x_k(K), x_{i_k+1}(K))| > \epsilon/(K + 3)^2 \right) \]
\[ + P^* \left( \max_x 2pld_i^{-1}p^{-1/2} |\tilde{F}_{n,l}(x_k(K), x_{i_k+1}(K)) - F(x_{i_k}(K), x_{i_k}(K))| > \epsilon/(K + 4)^2 \right) \]
\[ + P^* \left( 2\Lambda(+\infty)2^{-K}d_i^{-1}p^{-1/2} \sum_{j=1}^{pl} (H_m(X_j^*) - E^*[H_m(X_j^*)]) \right) > (\epsilon/2) - 2\Lambda(+\infty)pld_i^{-1}p^{-1/2}2^{-K} \right). \]

Further we get, using Markov’s inequality,
\[ P^* \left( \max_x |R_{n,l}^*(x_k(k), x_{i_k+1}(k+1))| > \epsilon/(k + 3)^2 \right) \]
\[ \leq \sum_{i=0}^{2^{k+1}-1} P^* \left( R_{n,l}^*(x_i(k+1), x_{i+1}(k+1)) > \epsilon/(k + 3)^2 \right) \]
\[ \leq \sum_{i=0}^{2^{k+1}-1} E^* \left[ R_{n,l}^*(x_i(k+1), x_{i+1}(k+1)) \right]^2 \left( (k + 3)^4 / \epsilon^2 \right). \]

It is our goal to show that \( E[P^*(\sup_{x \in \mathbb{R}}|R_{n,l}^*(x)| > \epsilon)] \to 0 \) as \( n \to \infty \). To this end we take expectations of every summand of the right-hand side of (4.2).
Making successive use of (4.3) and Lemma 4.3.4 we obtain

\[
E \left[ P^* \left( \text{max}_x |R^*_{n,l}(x_{i_k(x)}(k), x_{i_{k+1}(x)}(k+1))| > \epsilon/(k+3)^2 \right) \right] 
\]

\[
= C \sum_{i=0}^{2^{k+1}-1} E[R^2_{i}(x_i(k+1), x_{i+1}(k+1)) \frac{(k+3)^4}{\epsilon^2}
\]

\[
+ C \sum_{i=0}^{2^{k+1}-1} \frac{l^2}{d_i^2} E \left( F(x_i(k+1), x_{i+1}(k+1)) - \tilde{F}_{n,l}(x_i(k+1), x_{i+1}(k+1)) \right)^2 \frac{(k+3)^4}{\epsilon^2}
\]

\[
+ C \sum_{i=0}^{2^{k+1}-1} \frac{J_m^2(x_i(k+1), x_{i+1}(k+1))}{(m!)^2} \frac{1}{d_i^l} E \left( \tilde{\mu}_{n,l}(H_m) \right)^2 \frac{(k+3)^4}{\epsilon^2}
\]

\[
\leq C \sum_{i=0}^{2^{k+1}-1} l^{-\gamma} F(x_i(k+1), x_{i+1}(k+1)) \frac{(k+3)^4}{\epsilon^2}
\]

\[
+ C \sum_{i=0}^{2^{k+1}-1} \frac{l^2 d_n^2}{d_i^2 n^2} F(x_i(k+1), x_{i+1}(k+1)) \frac{(k+3)^4}{\epsilon^2}
\]

\[
+ C \sum_{i=0}^{2^{k+1}-1} \Lambda(x_i(k+1), x_{i+1}(k+1)) \frac{1}{d_i^l} E \left( \tilde{\mu}_{n,l}(H_m) \right)^2 \frac{(k+3)^4}{\epsilon^2}.
\]

We have also used Lemma 4.3.3 and

\[
E|R_{t,i}(y) - R_{t,i}(x)|^2 \leq C l^{-\gamma}(F(y) - F(x)),
\]

which is implied by Lemma 4.3.1.

Note that \((l/n)^2(d_n/d_i)^2 \leq C l^\lambda\) for some \(\lambda > 0\) and \(\Lambda(x_i(k+1), x_{i+1}(k+1))^2 \leq C 2^{-(k+1)}\).

Thus setting \(\eta = \min\{\gamma, \lambda\}\) yields

\[
E \left[ P^* \left( \text{max}_x |R^*_{n,l}(x_{i_k(x)}(k), x_{i_{k+1}(x)}(k+1))| > \epsilon/(k+3)^2 \right) \right] 
\]

\[
= C \left( l^{-\eta}(k+3)^4 \epsilon^{-2} + 2^{-(k+1)} l^2 / d_i^2 E[\tilde{\mu}_{n,l}(H_m)]^2 \right).
\]

In the same way we get

\[
E \left[ P^* \left( \text{max}_x |R^*_{n,l}(x_{i_k(x)}(K), x_{i_{k+1}(x)}(K+1))| > \epsilon/(K+3)^2 \right) \right] 
\]

\[
\leq C l^{-\eta}(K+3)^4 \epsilon^{-2} + C 2^{-K} l^2 / d_i^2 E[\tilde{\mu}_{n,l}(H_m)]^2
\]
take the expectation in (4.4) therefore yields

\[ E[P^* \left( \max_x 2pd_t^{-1}p^{-1/2} |\tilde{F}_{n,l}(x_{i_{K}(x)}(K), x_{i_{K}(x)+1}(K)) - F(x_{i_{K}(x)}(K), x_{i_{K}(x)+1}(K))| > \epsilon/(K+4)^2 \right) \]

\[ \leq \sum_{i=0}^{2K-1} 2^{pi} \frac{(K+4)^4}{e^2} E \left( F(x_i(K), x_{i+1}(K)) - \tilde{F}_{n,l}(x_i(K), x_{i+1}(K)) \right)^2 \]

\[ \leq Cl^{-\eta}(K+4)^4. \]

Choose now

\[ K = \left[ \log_2 \left( \frac{8\Lambda(\infty)}{\epsilon} l d_t^{-1}p^{1/2} \right) \right] + 1, \]

hence \( 2\Lambda(\infty)pld_t^{-1}p^{-1/2}e^{-K} \leq \epsilon/4 \). It remains to treat the last probability in (4.2). By our choice of \( K \) it can be bounded by

\[
P^* \left( d_t^{-1}p^{-1/2} \left| \sum_{j=1}^{pl} (H_m(X_j^*) - E^*[H_m(X_j^*)]) \right| > \epsilon \frac{2K-1}{4\Lambda(\infty)} \right) \leq d_t^{-2}p^{-1}E^* \left[ \sum_{j=1}^{pl} (H_m(X_j^*) - E^*[H_m(X_j^*)]) \right] \frac{16}{\epsilon^2} \Lambda(\infty)^2 2^{-2K+2}. \quad (4.4)
\]

By the proof of Theorem B (see the proof of Lemma 3.2 in Lahiri (1993)) we get

\[ d_t^{-2}p^{-1}E^* \left( \sum_{j=1}^{pl} (H_m(X_j^*) - E^*[H_m(X_j^*)]) \right)^2 \leq C. \]

Taking expectation in (4.4) therefore yields

\[
E \left[ P^* \left( d_t^{-1}p^{-1/2} \left| \sum_{j=1}^{pl} (H_m(X_j^*) - E^*[H_m(X_j^*)]) \right| > \epsilon \frac{2K-1}{4\Lambda(\infty)} \right) \right] \leq C \frac{16}{\epsilon^2} \Lambda(\infty)^2 2^{-2K+2} \leq Cl^{-2}p^{-1}d_t^2.
\]

We have now found estimates for expectations of all summands of (4.2). Combining these estimates we find

\[
E \left[ P^*(\sup_x R_{n,l}^*(x) > \epsilon) \right] \leq Cl^{-\eta} \epsilon^{-2} \sum_{k=0}^{K+1} (k+3)^4 + l^2d_t^{-2}E[\tilde{\mu}_{n,l}(H_m)]^2 \sum_{k=0}^{K} 2^{-(k+1)} + Cl^{-2}p^{-1}d_t^2 \leq Cl^{-\eta} \epsilon^{-2} (K+4)^5 + Cl^2d_t^{-2}E[\tilde{\mu}_{n,l}(H_m)]^2 + Cl^{-2}p^{-1}d_t^2 \leq Cl^{-\eta} \epsilon^{-2} (K+4)^5 + Cl^{-\eta} + Cl^{2\eta^{-2}}.
\]
In the last line we have used $l^2 d_l^{-2} E[\tilde{\mu}_{n,l}(H_m)]^2 \leq Cl^{-\lambda} \leq Cl^{-\eta}$ (see Lemma 4.3.2 (ii)) and $l^{-2p-1}d_l^2 \leq l^{2H-2}p^{-1}L^{m/2}(l) \leq l^{-\alpha}$ for some $\alpha > 0$.

The definition of $K$ yields

$$(K + 4)^5 \leq C (|\log(\epsilon^{-1})|^5 + |\log(pl)|^5) \leq C\epsilon^{-1}l^\delta,$$

for any $\delta > 0$ and a constant $C$, depending on $\delta$. Choose $\delta = \eta/2$ and $\rho = \min\{\eta - \delta, \alpha\}$, then

$$E\left[ P^*(\sup_x |R_{n,l}^*(x)| > \epsilon) \right] \leq Cl^{-\rho}(\epsilon^{-3} + 1).$$

$\square$
Sequential block bootstrap in a Hilbert space with applications to change-point analysis

5 Sequential block bootstrap in a Hilbert space with applications to change-point analysis

An important example of Hilbert space-valued observations are functional data. In the last decade statistical methods for such data have received great attention, among them environmental data analysis, see Hörmann and Kokoszka (2010). Due to a strong seasonal effect, for example in temperature or hydrological data, such time series are non-stationary and thus change-point analysis is a complex topic. A possible solution is to look at annual curves instead of the whole time series. In this case, observations become functions. The method of functional principal components was used by Kokoszka et al. (2008) in testing for independence in the functional linear model and by Benko et al. (2009) in two sample tests for $L^2[0,1]$-valued random variables, a method that was extended to change point analysis by Berkes et al. (2009). Another approach is due to Fraiman et al. (2014) who used record functions to detect trends in functional data. In contrast to all former approaches, we will take the fully functional observations into account. In fact, we will investigate the CUSUM statistic based on random functions, or more general random elements taking values in a Hilbert space.

Another change-point problem that will be considered is the abrupt change of the marginal distribution of random variables, now taking values in $\mathbb{R}^d$. This setting has been intensively studied in Chapter 3 (under long memory) and the Kolmogorov-Smirnov and Cramér-von Mises test were found to be suitable methods. Both are based on the empirical process which has been treated as random element of the space $D[-\infty, \infty]$ of cadlag functions. In this chapter we will pursue a different approach. For real-valued observations $X_1, \ldots, X_n$ define $Y_i$ by $Y_i(t) = 1_{\{X_i \leq t\}}$. The objects $Y_i$ are no longer real valued random variables, but take values in a function space. Here we will no longer make use of the cadlag space $D[-\infty, \infty]$, but the Hilbert space $L^2$. It turns out that the CUSUM statistic of the Hilbert space-valued sequence $(Y_i)_{i \geq 1}$ becomes the Cramér-von Mises statistic of the finite dimensional observation $(X_i)_{i \geq 1}$. Thereby, properties of the Cramér-von Mises test might be deduced from Hilbert space theory. Critical values for change point tests are often deduced from asymptotics. The CUSUM statistic can be expressed as a functional of the partial sum process $\sum_{i=1}^{\lfloor nt \rfloor} X_i$ for $t \in [0,1]$, whose asymptotic behaviour for $H$-valued data was investigated by Chen and White (1998) for mixing-gales and near epoch dependent processes. For statistical inference, one needs control over the
asymptotic distribution. Due to dependence and the infinite dimension of the \( \{X_i\}_{i \geq 1} \), the asymptotic distribution depends on an unknown infinite dimensional parameter - the covariance operator. Our solution is the bootstrap, which has been successfully applied to many statistics in the case of real or \( \mathbb{R}^d \)-valued data. For a combination with the change-point problem, see for instance Gombay and Horvath (1999) and Inoue (2001). For Hilbert spaces, only Politis and Romano (1994b) and recently Dehling et al. (2015) established the asymptotic validity of the bootstrap. The results of Politis and Romano (1994b) can only handle bounded random variables. Thus, indicator functions and the Cramér-von Mises statistic can be bootstrapped by their method, but general functional data cannot.

In this chapter, we extend the result of Dehling et al. (2015) by a sequential component, i.e. we are bootstrapping the partial sum process instead of bootstrapping the sample mean. This is inevitable for change-point problems, where the location of possible change-points is typically unknown.

5.1 Limit theorems in a Hilbert space

5.1.1 Functional central limit theorem

Let \( H \) be a separable (i.e. there exists a dense and countable subset) Hilbert space with inner product \( \langle \cdot , \cdot \rangle \) and norm \( \| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle} \). We say that an \( H \)-valued random variable \( X \) has mean \( \mu \in H \) if \( E \langle X, h \rangle = \langle \mu, h \rangle \) for all \( h \in H \). We denote it by \( EX \). Moreover, define the covariance operator \( S : H \rightarrow H \) of \( X \) (if it exists) by

\[
\langle Sh_1, h_2 \rangle = E [ \langle X - EX, h_1 \rangle \langle X - EX, h_2 \rangle ] \quad h_1, h_2 \in H,
\]

see also Section 2.3.2. For more details and a generalization to Banach spaces, see the book of Ledoux and Talagrand (1991).

Let \( (X_i)_{i \geq 1} \) be a sequence of (possibly dependent) \( H \)-valued random variables. For more information on the different dependency conditions in \( H \), see Section 2.3.1. It is our aim to prove a functional central limit theorem for such sequences, in other words we want to prove weak convergence of the partial sum process \( n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \mu) \), \( t \in [0,1] \). It is an element of the space \( D_H[0,1] \) and we will equip this space with the Skorohod metric, see Section 2.3.3.

Theorem 16. Let \( (X_n)_{n \in \mathbb{Z}} \) be \( L_1 \)-near epoch dependent on a stationary, absolutely regular sequence \( (\xi_n)_{n \in \mathbb{Z}} \) with \( EX_1 = \mu \in H \) and assume that the following conditions hold for some \( \delta > 0 \)

1. \( E \| X_1 \|^{4+\delta} < \infty \),
2. \( \sum_{m=1}^{\infty} m^2 \langle a_n \rangle^{\delta/(\delta+3)} < \infty \),
3. \( \sum_{m=1}^{\infty} m^2 \langle \beta(m) \rangle^{\delta/(\delta+4)} < \infty \).
Then, as \( n \to \infty \),

\[
\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \mu) \right)_{t \in [0,1]} \overset{D}{\to} (W(t))_{t \in [0,1]},
\]

where weak convergence takes place in the space \( D_H[0,1] \) and where \( (W(t))_{t \in [0,1]} \) is a Brownian motion in \( H \) and \( W(1) \) has the covariance operator \( S : H \to H \), defined by

\[(Sx,y) = \sum_{i=-\infty}^{\infty} E[\langle X_0 - \mu, x \rangle \langle X_i - \mu, y \rangle], \quad \text{for } x,y \in H. \tag{5.1}\]

Furthermore, the series in (5.1) converges absolutely.

**Remark 5.1.1.**

(i) For a fixed \( t \in [0,1] \) one gets the partial sum of \( H \)-valued random variables. Weak convergence has been established for uniform and strong mixing processes by Dehling (1983), for martingale difference sequences by Jakubowski (1980), for \( L^2 \)-NED processes by Chen and White (1998) and for \( L^1 \)-NED processes by Dehling et al. (2015).

(ii) A functional central limit theorem was given by Walk (1977) for martingale difference sequences and by Chen and White (1998) in the near epoch dependent case. They assume strong mixing, which is more general than absolute regularity. Then again, we require \( L_1 \)-near epoch dependence, while they use \( L_2 \)-near epoch dependence, which implies our conditions and is therefore more restrictive.

### 5.1.2 Bootstrapping the partial sum process

Functional central limit theorems have several classical applications. In this thesis they mainly serve to prove weak convergence of change-point statistics. However, the problem arises that the limiting distribution may be unknown in praxis, or even if it is known, it might depend on an infinite dimensional parameter, in our case the covariance operator \( S \).

To circumvent this problem, we will use the nonoverlapping block bootstrap of Carlstein (1986) to construct a process with the same limiting distribution as \( \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} (X_i - \mu) \). As described in Section 2.4.1, one constructs \( k = \lfloor n/p \rfloor \) nonoverlapping blocks \( I_i \) of length \( p \). Drawing independently and with replacement from these blocks yields the bootstrap sample \( X_{1}^{\ast}, \ldots, X_{pl}^{\ast} \) with

\[
P\left( (X_{(j-1)p+1}^{\ast}, \ldots, X_{jp}^{\ast}) = I_i \right) = \frac{1}{k} \quad \text{for } i,j = 1, \ldots, k.
\]

The fact that the random variables \( (X_i)_{i \geq 1} \) take values in Hilbert space \( H \) has no impact on the bootstrap mechanism. Obviously, the bootstrap sample is \( H \)-valued too. Now we can
define a bootstrapped version of the partial sum process by

$$W_{n,p}^*(t) = \frac{1}{\sqrt{kp}} \sum_{i=1}^{\lfloor kpt \rfloor} (X_i^* - E^* X_i^*),$$  \hspace{1cm} (5.2)

which is again an element of $D_{H[0,1]}$. As usual, $E^*$ and $P^*$ denote conditional expectation and probability, respectively, given $\sigma(X_1, \ldots, X_n)$. Further, $\xrightarrow{D^*}$ denotes weak convergence with respect to $P^*$. The next result establishes the asymptotic distribution of (5.2).

**Theorem 17.** Let $(X_n)_{n \in \mathbb{Z}}$ be $L_1$-near epoch dependent on a stationary, absolutely regular sequence $(\xi_n)_{n \in \mathbb{Z}}$ with $EX_1 = \mu$ and assume that the following conditions hold for some $\delta > 0$

1. $E\|X_1\|^{4+\delta} < \infty$,

2. $\sum_{m=1}^{\infty} m^2 (a_m)^{\delta/(\delta+3)} < \infty$,

3. $\sum_{m=1}^{\infty} m^2 (\beta(m))^{\delta/(\delta+4)} < \infty$.

Further, let the block length be nondecreasing, $p(n) = O(n^{1-\epsilon})$ for some $\epsilon$ and $p(n) = p(2^l)$ for $n = 2^{l-1} + 1, \ldots, 2^l$, for all $l \in \mathbb{N}$. Then

$$\left( W_{n,p}^*(t) \right)_{t \in [0,1]} \xrightarrow{D^*} \left( W(t) \right)_{t \in [0,1]} \text{ a.s.},$$

where $(W(t))_{t \in [0,1]}$ is a Brownian motion in $H$ and $W(1)$ has the covariance operator $S : H \rightarrow H$, defined in Theorem 16.

**Remark 5.1.2.** (i) The result is stated in terms of almost sure weak convergence, which is strongest of the three notions described in Section 2.4.2.

(ii) Weak convergence of the object (5.2) has not been considered before. For fixed $t \in [0,1]$, Politis and Romano (1994b) (for bounded random variables) and recently Dehling et al. (2015) established the asymptotic validity of the bootstrap. For real-valued observations, Calhoun (2015) considered the bootstrapped partial sum process and proved consistency under $L^2$-NED.

### 5.2 Application to change-point analysis

#### 5.2.1 Change in the mean of $H$-valued observations

Let us consider the following change point problem. Given $X_1, \ldots, X_n$, we want to test the null hypothesis

$$H_0: \quad EX_1 = \cdots = EX_n$$
against the alternative

$H_A: \quad EX_1 = \cdots = EX_k \neq EX_{k+1} = \cdots = EX_n,$

for some $k \in \{1, \ldots, n-1\}$.

For real-valued variables, asymptotics of CUSUM-type tests have been extensively studied by Csörgő and Horvath (1997). They investigated tests for i.i.d. data, weakly dependent data, and for long range dependent processes. The third case was extended by Dehling et al. (2012).

For functional data, Berkes et al. (2009) have developed estimators and tests for a change point in the mean, which is extended by Hörmann and Kokoszka (2010), Aston and Kirch (2012), Horvath et al. (2014) and Torgovitski (2016) to weakly dependent data. They reduce the dimension of the observations with the help of functional principal components, while motivated by Theorems 1 and 2 - we consider the test statistic

$$T_n = \max_{1 \leq m < n} \frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{m} X_i - \frac{m}{n} \sum_{i=1}^{n} X_i \right\|$$

based on full functional data. To obtain critical values without dimension reduction, one would have to know the long run covariance operator, which is an infinite dimensional parameter and thus hard to estimate. Therefore, we propose to use the bootstrap analogue $T^*_n$ of $T_n$ with

$$T^*_n = \max_{1 \leq m < kp} \frac{1}{\sqrt{kp}} \left\| \sum_{i=1}^{m} X^*_i - \frac{m}{kp} \sum_{i=1}^{kp} X^*_i \right\|.$$

The next result states that $T_n$ and $T^*_n$ have the same limiting distribution, which is a direct consequence of Theorems 1 and 2 and the continuity of both the maximum function and the Hilbert space norm.

**Corollary 5.2.1.** (i) Under the conditions of Theorem 16,

$$T_n \overset{D}{\to} \max_{t \in [0,1]} \|W(t) - tW(1)\|,$$

where $(W(t))_{t \in [0,1]}$ is the Brownian motion defined in Theorem 16.

(ii) Under the conditions of Theorem 17

$$T^*_n \Rightarrow_{*} \max_{t \in [0,1]} \|W(t) - tW(1)\| \quad \text{a.s.}$$

The Corollary motivates the following test procedure, which is typical for bootstrap tests:

(i) Compute $T_n$.

(ii) Simulate $T^*_j,n$ for $j = 1, \ldots, J$. 
(iii) Based on the independent (conditional on $X_1, \ldots, X_n$) random variables $T^*_{n,1}, \ldots, T^*_{n,J}$, compute the empirical $(1 - \alpha)$-quantile $q_{n,J}(\alpha)$.

(iv) If $T_n > q_{n,J}(\alpha)$ reject the null hypothesis.

By Corollary 5.2.1 and Proposition 2.4.1, the proposed test has an asymptotically significance level of $\alpha$.

5.2.2 Local alternatives and multiple change-points

Next, we derive the asymptotic distribution of the (bootstrapped) change-point statistic under a sequence of converging alternatives. Define the triangular array of $H$-valued random variables

$$Y_{n,i} = \begin{cases} X_i & \text{if } i \leq \lceil n\tau \rceil, \\ X_i + \Delta_n & \text{if } i > \lceil n\tau \rceil, \end{cases}$$

for $n \in \mathbb{N}$ and $i \leq n$. Here, $\lceil n\tau \rceil$ is the unknown change-point for some $\tau \in (0, 1)$ and $(\Delta_n)_n$ is an $H$-valued deterministic sequence with

$$\|\sqrt{n}\Delta_n - \Delta\| \to 0,$$

for $n \to \infty$ and some $\Delta \in H$.

Now we want to test the Hypothesis $\Delta_n = 0$ against the sequence of Alternatives where $\Delta, \Delta_n \in H \setminus \{0\}$.

Note that a bootstrap sample $(Y^*_{n,i})_{i \leq kp, n \geq 1}$ can be created analogously to $(X^*_i)_{i \leq kp}$. Then we can define the statistics $T_n$ and $T^*_n$, now based on $Y_{n,i}$ and $Y^*_{n,i}$, respectively.

**Corollary 5.2.2.** (i) Consider an array $(Y_{n,i})_{n \in \mathbb{N}, i \leq n}$, defined as above. If the conditions of Theorem 16 hold for $(X_i)_{i \geq 1}$, then under the sequence of local alternatives

$$T_n \Rightarrow \max_{t \in [0,1]} \|W(t) - tW(1) + \psi_\tau(t)\Delta\|, \quad (5.1)$$

where $(W(t))_{t \in [0,1]}$ is the Brownian motion defined in Theorem 16 and the function $\psi_\tau : [0, 1] \to \mathbb{R}$ is defined by

$$\psi_\tau(t) = \begin{cases} t(1 - \tau) & \text{if } t \leq \tau \\ (1 - t)\tau & \text{if } t > \tau. \end{cases}$$

(ii) If the conditions of Theorem 17 are satisfied, then under the sequence of local alternatives

$$T^*_n \Rightarrow_* \max_{t \in [0,1]} \|W(t) - tW(1)\| \text{ a.s..} \quad (5.2)$$
Proof. Part (i) can be obtained by arguments similar to the case of real-valued random variables, see Theorem 2.1 in Dehling et al. (2017) or Theorem 8 in this thesis.

To verify part (ii) define random variables $U_1, \ldots, U_k$, where $U_i$ is the number of the $i$th drawn block. Clearly the random variables $U_1, \ldots, U_k$ are independent and uniformly distributed on $\{1, \ldots, k\}$.

Note that the random variables in the blocks $B_1, \ldots, B_{\lfloor k\tau \rfloor}$ are of the form $X_i$ and the variables of the blocks $B_{\lfloor k\tau \rfloor + 2}, \ldots, B_k$ are of the form $X_i + \Delta_n$. The change point occurs in the block $B_{\lfloor k\tau \rfloor + 1}$, so this block contains shifted and non-shifted variables.

This subdivision in different types of blocks leads to the following decomposition of the process

$$\frac{1}{\sqrt{kp}} \left( \sum_{i=1}^{\lfloor kpt \rfloor} Y_{n,i}^* - \frac{\lfloor kpt \rfloor}{kp} \sum_{i=1}^{kp} Y_{n,i}^* \right) = \frac{1}{\sqrt{kp}} \left( \sum_{i=1}^{\lfloor kpt \rfloor} X_i^* - \frac{\lfloor kpt \rfloor}{kp} \sum_{i=1}^{kp} X_i^* \right)$$

$$+ \sqrt{kp\Delta_n} R_{n,k,p}(t),$$

where

$$R_{n,k,p}(t) = \frac{1}{kp} \left( \sum_{i=1}^{\lfloor kt \rfloor} 1\{U_i > \lfloor k\tau \rfloor + 1\} \right)^p - \frac{1}{kp} \frac{\lfloor kpt \rfloor}{kp} \sum_{i=1}^{k} 1\{U_i > \lfloor k\tau \rfloor + 1\}$$

$$\quad + \frac{1}{kp} (\lfloor k\tau \rfloor + 1)p - \lfloor n\tau \rfloor \sum_{i=1}^{\lfloor kt \rfloor} 1\{U_i = \lfloor k\tau \rfloor + 1\}$$

$$- \frac{1}{kp} (\lfloor k\tau \rfloor + 1)p - \lfloor n\tau \rfloor \frac{\lfloor kpt \rfloor}{kp} \sum_{i=1}^{k} 1\{U_i = \lfloor k\tau \rfloor + 1\}$$

$$+ 1\{U_{\lfloor kt \rfloor + 1} > \lfloor k\tau \rfloor + 1\} \frac{1}{kp} (\lfloor kpt \rfloor - \lfloor kt \rfloor)p$$

$$+ 1\{U_{\lfloor k\tau \rfloor + 1} = \lfloor k\tau \rfloor + 1\} \frac{1}{kp} \max\{\lfloor kpt \rfloor - \lfloor n\tau \rfloor, 0\}. \quad (5.3) \quad (5.4) \quad (5.5)$$

By part (ii) of Corollary 5.2.1 and $\sqrt{kp\Delta_n} \to \Delta$ it remains to show that

$$P^*\left( \sup_{t \in [0,1]} |R_{n,k,p}(t)| > \epsilon \right) \to 0 \quad \forall \epsilon > 0, \text{ a.s.}$$

as $n \to \infty$. But this holds because $R_{n,k,p}$ is independent of the $X_i$ and: (5.3) + (5.4) and (5.5)
+ (5.6) are each $o_P(1)$. To see this observe

$$\frac{1}{k} \sum_{i=1}^{\lfloor kt \rfloor} 1 \{ U_i > \lfloor k\tau \rfloor + 1 \} \xrightarrow{P} t(1 - \tau),$$

uniformly in $t$.

The quantity in (5.7) is $o_P(1)$ because $(\lfloor kpt \rfloor - \lfloor kt \rfloor p)/(kp) \to 0$. Finally, (5.8) is $o_P(1)$ because $P(U_{\lfloor kl \rfloor + 1} = [k\tau] + 1) = k^{-1}$.

Remark 5.2.3. (i) The deterministic element $\Delta \in H$ describes the amount of the change, while $\psi_\tau$ describes its location. Together they discriminate the limits of (5.1) and (5.2) and hence they are responsible for the asymptotic power. Note that the maximum of $\psi_\tau$ is $\tau(1 - \tau)$. Thus the power decreases, if the change occurs near the beginning or near the end of the observation period. We will illustrate this issue in the simulation study.

(ii) The above test problem is that of at most one change point (AMOC). However, especially in functional time series multiple changes are thinkable. Our statistic can be extended to allow such alternatives in the same way as the classical CUSUM statistic, see Erasmus and Lombard (1988). The statistic in Erasmus and Lombard (1988) can be written as a continuous functional of the partial sum process. Hence, by our Theorems 16 and 17 and the continuous mapping Theorem, one obtains weak convergence of this statistic and its bootstrapped version.

Another possibility of testing for multiple changes is the binary segmentation method, see Vostrikova (1981) and Venkatraman (1992). We will apply this method to real-life functional data.

5.2.3 Cramér-von Mises test for change in the marginal distribution

We will now apply the results to random variables, whose realizations are not truly functional. Consider, for example, the real-valued random variables $X_1, \ldots, X_n$ and the problem of testing for changes in the underlying marginal distributions:

$$H_0: \quad P(X_1 \leq t) = \cdots = P(X_n \leq t) \quad \forall t \in \mathbb{R}$$

against

$$H_A: \quad P(X_1 \leq t) = \cdots = P(X_k \leq t) \neq P(X_{k+1} \leq t) = \cdots = P(X_n \leq t),$$

for some $k \in \{1, \ldots, n-1\}$ and $t \in \mathbb{R}$. Asymptotic tests have been investigated by Csörgő and Horvath (1997), Horvath and Shao (1996) and Szyszkowicz (1994) in the independent case, by Inoue (2001) for strong mixing data, and by Giraitis et al. (1996b) for long-memory linear processes. Moreover, a test under long-range dependent, subordinated Gaussian processes has been investigated in Chapter 3 of this thesis.
As test statistic we have considered the difference of the empirical distribution functions before and after possible change-points. Its asymptotic distribution is derived via weak convergence of the sequential empirical process, with convergence taking place in \((D[0,1] \times [-\infty, \infty])\), equipped with the uniform metric. The other results mentioned, proceed in a similar way with the uniform metric sometimes replace by the Skorohod metric.

Here we pursue a different approach. First, note that the (sequential) empirical process is just the partial sum (process) of the indicator functions

\[ 1_{\{X_i \leq t\}}, \quad t \in \mathbb{R}. \]  

(5.9)

The idea of this section is to interpret (5.9) as random elements in a Hilbert space \(H\). In fact, the connection of empirical process and Hilbert space theory has been utilized many times, see for example Politis and Romano (1994b), Tsudaka and Nishiyama (2014) and Dedecker et al. (2015).

There are several possible choices for the space \(H\). Here we consider \(H = L^2(\mathbb{R}, \mu)\), with the measure given by \(\mu(A) = \int_A w(t) dt\) for some positive, bounded function \(\int_\mathbb{R} w(t) dt < \infty\). An important special case is \(w(t) = F(t)\) with \(F(t)\) being the distribution function of \(X_1\). Equipped with the inner product

\[ \langle f, g \rangle_w = \int_\mathbb{R} f(t)g(t)w(t) \, dt, \]

\(H\) becomes a separable Hilbert space. By Fubini’s Theorem, we have

\[ E\left[ \int_\mathbb{R} 1_{\{X_i \leq t\}} h(t) w(t) \, dt \right] = \int_\mathbb{R} F(t) h(t) w(t) \, dt \quad \text{for all } h \in H. \]

Hence, by the definition it follows that the mean of (5.9) (as an \(H\)-valued element) is just the distribution function of \(X\). So the change-in-the-mean-problem (in \(H\)) becomes a change-in-distribution-problem (in \(\mathbb{R}\)). Furthermore, the arithmetic mean becomes the empirical distribution function. Note that this still holds when we consider \(\mathbb{R}^d\)-valued data, then considering the space \(H = L^2(\mathbb{R}^d, \mu^d)\).

The CUSUM statistic in \(H\) now becomes a Cramér-von Mises-type change-point statistic

\[ T_{n,w} = \max_{1 \leq m \leq n-1} \frac{1}{n} \int_{\mathbb{R}^d} \left( \sum_{i=1}^{m} 1_{\{X_i \leq t\}} - \frac{m}{n} \sum_{i=1}^{n} 1_{\{X_i \leq t\}} \right)^2 w(t) \, dt. \]  

(5.10)

Note that the another change-point statistic, the Kolmogorov Smirnov statistic (3.1), cannot be treated via this approach. The reason is the application of the norm \(\| \cdot \|_\infty = \sup_{x \in \mathbb{R}} |\cdot|\), which is not continuous with respect to the \(L^2\)-norm of the Hilbert space.

This \(L^2\) approach to change-point analysis was also considered by Tsudaka and Nishiyama.
(2014), albeit for independent observations. Due to the dependence, the limit distribution obtained here is quite complicated, which is why we consider a resampling method.

The empirical process has been bootstrapped by several authors, among them Bühlmann (1994), Naik-Nimbalkar and Rajarshi (1994), Peligrad (1998) (all block bootstrap), Doukhan et al. (2015) (wild bootstrap) and Kojadinovic and Yan (2012) (weighted bootstrap). All the results are for weakly dependent data, while in Chapter 4 of this thesis the block bootstrap is considered for LRD data. Furthermore, convergence is always proved in the space of càdlàg-functions. The only exception is due to Politis and Romano (1994b). They use Hilbert space theory in combination with the stationary bootstrap.

We do not spend much time on the empirical process here and consider directly the bootstrapped Cramér-von Mises statistic. It is given by

$T_{n,w}^* = \max_{1 \leq m \leq kp-1} \frac{1}{kp} \int_{\mathbb{R}^d} \left( \sum_{i=1}^{m} 1\{X_i^* \leq t\} - \frac{m}{kp} \sum_{i=1}^{kp} 1\{X_i^* \leq t\} \right)^2 w(t) \, dt,$

(5.11)

where the sample $X_1^*, \ldots, X_{kp}^*$ is produced by the nonoverlapping block bootstrap for $\mathbb{R}^d$-valued data. We will now state conditions, under which the bootstrap is valid.

**Corollary 5.2.4.** Let $(X_n)_{n \in \mathbb{N}}$ be $\mathbb{R}^d$ valued random variables, $L_1$-near epoch dependent on a stationary, absolutely regular sequence $(\xi_n)_{n \in \mathbb{Z}}$, such that for some $\delta > 0$

1. $\sum_{m=1}^{\infty} m^2 (a_m)^{\delta/(\delta+3)} < \infty$,

2. $\sum_{m=1}^{\infty} m^2 (\beta_m)^{\delta/(\delta+4)} < \infty$.

Let the block length $p$ be nondecreasing with $p(n) = O(n^{1-\epsilon})$ for some $\epsilon > 0$ and $p(n) = p(2^l)$ for $n = 2^{l-1} + 1, \ldots, 2^l$.

Then, almost surely, the conditional distribution of $T_{n,w}^*$, given $X_1, \ldots, X_n$, converges to the same limit as the distribution of $T_{n,w}$, as $n \to \infty$.

**Proof.** Note that producing a bootstrap sample $X_1^*, \ldots, X_{kp}^*$ first, and then analyzing the indicators

$1\{X_1^* \leq \cdot\}, \ldots, 1\{X_{kp}^* \leq \cdot\},$

is the same as if we first look upon the indicators as $H$-valued random variables $Y_1, \ldots, Y_n$ and then generate $Y_1^*, \ldots, Y_{kp}^*$. Thus it suffices to verify the conditions of Theorems 16 and 17, respectively. In this case, Corollary 5.2.1 would hold, leading to the desired result.

The moment condition in Theorem 17 is automatically satisfied, due to the definition of $w(t)$. Moreover, the dependence conditions are satisfied because of Lemma 2.2 in Dehling et al. (2015) and the Lipschitz-continuity of the mapping $x \mapsto 1\{x \leq \cdot\}$. \qed
5.3 Numerical results and real data analysis

5.3.1 Application to flood data

![Image of flood data analysis](image)

Figure 5.1: Left: Process $\frac{1}{\sqrt{n}} \| \bar{X}_k - k/n \bar{X}_n \|$ (black line) computed from 103 annual flow curves of the river Chemnitz and 0.95 level of significance (dashed line) computed from 999 bootstrap iterations. Right: Average annual flow curves of the time period 1910 - 1964 (grey line) and the time period 1965 - 2012 (black line).

![Image of flow curves](image)

Figure 5.2: Processes $n^{-1/2} \| \bar{X}_k - k/n \bar{X}_{n_1} \|$, $k = 1, \ldots, n_1$ (left) and $(n - n_1)^{-1/2} \| \bar{X}_{n_1+1:k} - k/(n-n_1) \bar{X}_{n_1+1:n} \|$, $k = n_1 + 1, \ldots, n$ (right) and 0.95 level of significance (dashed lines) computed from 999 bootstrap iterations each. Right: Average annual flow curves of the time period 1910 - 1964 (grey line) and the time period 1965 - 2012 (black line).

To illustrate our methods we apply the tests, described in the previous subsections, to hydrological observations.

The first data set contains average daily flows of the river Chemnitz at Goeritzhain for the time period 1910 - 2012. Thus one gets 103 annual flow curves which can be interpreted as realizations of $\mathbb{R}^{365}$-valued random variables. Alternatively, one could smoothen the curves and hence get functional data. Moreover, it seems reasonable to assume dependence between the curves.
Let \( X_i \) be the \( i \)th annual curve, taking its value in \( \mathbb{R}^{365} \). Figure 5.1 shows the process

\[
\frac{1}{\sqrt{n}} \left\| \sum_{i=1}^{k} X_i - \frac{k}{n} \sum_{i=1}^{n} X_i \right\| \quad k = 1, \ldots, n - 1.
\]

The value of the test statistic is the maximum of this process, which is attained in 1964. Because it is larger than the bootstrapped 5% level of significance, the test indicates that there has been a change in structure of the annual flow curves.

Figure 5.1 (right) illustrates the character of this change by comparing the average flow curves based on the data before and after 1964. It is a common method to use the time point in which the maximum is attained as an estimator for the time of the change. This procedure was investigated by Aue et al. (2009) for independent functional observations and by Aue et al. (2015) for functional time series.

In order to search for further change points we made use of the binary-segmentation procedure, see Vostrikova (1981) and Venkatraman (1992). In detail, we have applied our change-point test to the subsamples \( X_1, \ldots, X_{55} \) and \( X_{56}, \ldots, X_{103} \). Figure 5.2 shows the associated CUSUM processes and the bootstrapped 5% levels of significance. In neither case an additional change is detected, hence we do not have to worry about a Bonferroni-correction for multiple testing.

Of course, there are other methods to deal with this data set. One might adopt the methodology of Robbins et al. (2011), used to detect changes in storm frequency and strengths. Here one might jointly test for changes in the yearly flood counts and the corresponding river heights.

As a second example, we look at annual maximum flows (the flows are annual maximums over daily observations) of the river Elbe at Dresden for the time period 1850 - 2012, see Figure 5.3.

![Figure 5.3: Annual maximum flows of the river Elbe at Dresden from 1850 to 2012.](image)
In the statistical analysis of floods annual maxima are typically modeled as independent. However, such time series often display some correlation in truth. Classical methods of extreme value theory sometimes fail if observations are dependent, a problem that is bypassed by our method. Moreover, the data seem to have heavy tails. But Corollary 3 does not require any moment conditions and hence we may apply the test for distributional change to these $\mathbb{R}$-valued observations. For this purpose, the Cramér-von Mises statistic (5.10) has been computed, together with 999 iterations of its bootstrap version (5.11). Figure 5.4 (left) shows the process

$$\frac{1}{n} \int \left( \sum_{i=1}^{k} 1\{X_i \leq x\} - \frac{k}{n} \sum_{i=1}^{n} 1\{X_i \leq x\} \right)^2 \phi(x) \, dx \quad k = 1, \ldots, n - 1,$$

where we have used the probability density of the $N(2000, 2000^2)$ distribution as weight function $\phi(\cdot)$. The value of the test statistic equals the maximum of this process, which is larger than the bootstrapped level of significance and therefore a change is detected.

Finally Figure 5.4 (right) compares the empirical distribution functions based on the data before and after 1900, which is where the maximum is attained. The comparison indicates that moderately severe floods have become less frequent.

### 5.3.2 Finite sample performance of the CUSUSM test

In this simulation study we will apply our CUSUM test to realizations of functional time series, given by

$$Y_i(t) = X_i(t) + \mu(t)1\{t > k\}, \quad t \in [0, 1]. \quad (5.1)$$

The function $\mu: [0, 1] \to \mathbb{R}$ describes the change, $k$ is the time of this change and $(X_i(t))_{i \geq 1}$ is a sequence of independent, identically distributed functional data. In detail, we consider
Table 5.1: Empirical size for bootstrapped test \( T_{n,p}^* \) and the fpca-based test \( S_{n,d} \); independent, identically distributed functional data; nominal size \( \alpha = 0.1 \), sample size \( n \), bootstrap block length \( p \), number of principal components \( d \).

<table>
<thead>
<tr>
<th></th>
<th>( T_{n,p}^* )</th>
<th>( S_{n,d} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BB</td>
<td>BM</td>
</tr>
<tr>
<td>( n = 50 )</td>
<td>( p = 1 )</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td>( p = 3 )</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>( p = 5 )</td>
</tr>
<tr>
<td></td>
<td>( p = 7 )</td>
<td>0.108</td>
</tr>
<tr>
<td></td>
<td>( n = 50 )</td>
<td>( p = 1 )</td>
</tr>
<tr>
<td></td>
<td>( p = 4 )</td>
<td>0.088</td>
</tr>
<tr>
<td></td>
<td>( n = 100 )</td>
<td>( p = 7 )</td>
</tr>
<tr>
<td></td>
<td>( p = 10 )</td>
<td>0.093</td>
</tr>
</tbody>
</table>

trajectories of the Brownian Motion (BM) and the Brownian Bridge (BB). Both are transformed to functional objects by the R-function `Data2fd`, using a basis of B-spline functions. The CUSUM test is then applied to these sequences, where critical values are obtained from \( J = 499 \) bootstrap iterations. Moreover, empirical size (empirical probabilities that the hypothesis is falsely rejected) and empirical power are deduced from 1000 simulation runs. They are displayed in tables 5.1 and 5.2, together with the original results of Berkes et al. (2009). The performances of the two tests are quite similar, under the hypothesis as well as under the alternative. The test \( (S_{n,d}) \) by Berkes et al. (2009) is based on functional principal components analysis (fpca) (where \( d \) components are used) and is designed for independence. Thus, critical values are not valid for dependent observations. Our test has good properties even if one does not assume independence and chooses a block length greater than one.

Next consider weakly dependent functional data. As model for the sequence \( (X_i(t))_{i \geq 1} \) in (5.1) we now use functional autoregressive processes of order 1 (FAR(1)), formally

\[
X_i(t) = \int_0^1 \psi(t, s) X_{i-1}(s) \, ds + \epsilon_i(t),
\]

(5.2)

see Bosq (2000). The \( (\epsilon_i(t))_{i \geq 1} \) are independent and Gaussian and \( \psi(s, t) \) is a kernel function, satisfying

\[
\|\psi\|^2_{L^2([0,1]^2)} = \int_0^1 \int_0^1 \psi^2(s, t) \, ds \, dt < 1.
\]
Table 5.2: Empirical power for bootstrapped test ($T^*_{n,p}$) and the fpca-based test ($S_{n,d}$); independent, identically distributed functional data; $\mu(t) = \sin(t)$, change after 50% of the observations; nominal size $\alpha = 0.1$, sample size $n$, bootstrap block length $p$, number of principal components $d$.

<table>
<thead>
<tr>
<th></th>
<th>$T^*_{n,p}$</th>
<th></th>
<th>$S_{n,d}$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BB</td>
<td>BM</td>
<td>BB</td>
<td>BM</td>
</tr>
<tr>
<td>$p = 1$</td>
<td>1.000</td>
<td>0.823</td>
<td>1.000</td>
<td>0.998</td>
</tr>
<tr>
<td>$p = 3$</td>
<td>1.000</td>
<td>0.820</td>
<td>1.000</td>
<td>0.726</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>$p = 5$</td>
<td>1.000</td>
<td>0.771</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p = 7$</td>
<td>1.000</td>
<td>0.752</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n = 100$</td>
<td>$p = 7$</td>
<td>1.000</td>
<td>0.965</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p = 10$</td>
<td>1.000</td>
<td>0.941</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As kernel functions we use

$$\psi_G(s,t) = C_1 \exp((s^2 + t^2)/2) \quad \text{or} \quad \psi_W(s,t) = C_2 \min(s,t),$$

the so-called Gaussian- or Wiener kernels, respectively. One obtains

$$\|\psi_G\|_{L^2([0,1]^2)} \approx C_1 (0.6832)^{-1} \quad \text{and} \quad \|\psi_W\|_{L^2([0,1]^2)} = C_2 6^{-1/2}.$$

Note that the $L^2$-norms of the kernel functions indicate the strength of the dependence in the sequences.

In the simulation study we have reproduced the implementation mode of Torgovitski (2016), using the R-package *fda*. The $\epsilon_i(t)$ are created from Brownian bridges, which are then transformed to functional data objects by the R-function *Data2fd*, using 25 B-spline functions. We set $X_{-99}(t) = \epsilon_{-99}(t)$ and $X_i$ as in (5.2) for $i \geq -98$. Using a burn-in period of length 100, we discard $X_{-99}, \ldots, X_0$. Afterwards a function $\mu(t)$ is eventually added to $X_k(t), \ldots, X_n(t)$, describing the structural change. Finally, the CUSUM test is applied.

Table 5.3 shows the empirical size of the test. For almost all combinations of dependencies and block lengths it is higher then the nominal size. However, as long as the dependence is not too strong ($\|\psi\| \leq 0.2$ for $n = 50$, $\|\psi\| \leq 0.4$ for $n = 100$) this happens to an acceptable degree. For $\|\psi\| = 0.6$ the probability of a type I error becomes too high. This is hardly surprising, as bootstrapped tests suffer from this issue even if the observations are real valued, see table 5.5 below. Finally one might compare the outcome of the test for the different FAR(1)-models.

The test performs better if functional observations are generated using the Gaussian kernel, but only for the right choice of block length.
Table 5.3: Empirical size for bootstrapped test ($T_{n,p}^*$) and the fpca-based test ($S_{n,d}$): FAR(1)-model with Gaussian-/Wiener-kernel; nominal size $\alpha = 0.1$, sample size $n$, bootstrap block length $p$, number of principal components $d$

<table>
<thead>
<tr>
<th>$n=50$</th>
<th>$|\psi|_{L^2}$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{n,p}^*$</td>
<td>$p=4$</td>
<td>0.118/0.122</td>
<td>0.138/0.154</td>
<td>0.171/0.189</td>
<td>0.274/0.263</td>
</tr>
<tr>
<td></td>
<td>$p=5$</td>
<td>0.106/0.124</td>
<td>0.131/0.099</td>
<td>0.159/0.149</td>
<td>0.214/0.235</td>
</tr>
<tr>
<td></td>
<td>$p=6$</td>
<td>0.120/0.133</td>
<td>0.143/0.120</td>
<td>0.151/0.168</td>
<td>0.193/0.216</td>
</tr>
<tr>
<td></td>
<td>$p=7$</td>
<td>0.117/0.110</td>
<td>0.149/0.116</td>
<td>0.153/0.142</td>
<td>0.209/0.203</td>
</tr>
<tr>
<td></td>
<td>$p=8$</td>
<td>0.139/0.125</td>
<td>0.157/0.144</td>
<td>0.164/0.146</td>
<td>0.195/0.211</td>
</tr>
<tr>
<td>$S_{n,d,h}$</td>
<td>$d=1$, $h=1$</td>
<td>0.072/0.096</td>
<td>0.087/0.098</td>
<td>0.087/0.099</td>
<td>0.089/0.112</td>
</tr>
<tr>
<td></td>
<td>$d=3$, $h=1$</td>
<td>0.051/0.046</td>
<td>0.043/0.046</td>
<td>0.034/0.038</td>
<td>0.032/0.027</td>
</tr>
<tr>
<td></td>
<td>$d=1$, $h=2$</td>
<td>0.099/0.101</td>
<td>0.082/0.080</td>
<td>0.070/0.052</td>
<td>0.047/0.048</td>
</tr>
<tr>
<td></td>
<td>$d=3$, $h=2$</td>
<td>0.049/0.068</td>
<td>0.059/0.040</td>
<td>0.053/0.028</td>
<td>0.031/0.025</td>
</tr>
<tr>
<td></td>
<td>$d=1$, $h=3$</td>
<td>0.100/0.105</td>
<td>0.091/0.092</td>
<td>0.065/0.055</td>
<td>0.048/0.040</td>
</tr>
<tr>
<td></td>
<td>$d=3$, $h=3$</td>
<td>0.072/0.082</td>
<td>0.071/0.064</td>
<td>0.051/0.039</td>
<td>0.031/0.035</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$n=100$</th>
<th>$|\psi|_{L^2}$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_{n,p}^*$</td>
<td>$p=6$</td>
<td>0.102/0.123</td>
<td>0.139/0.131</td>
<td>0.157/0.177</td>
<td>0.236/0.222</td>
</tr>
<tr>
<td></td>
<td>$p=8$</td>
<td>0.118/0.119</td>
<td>0.117/0.133</td>
<td>0.162/0.136</td>
<td>0.199/0.187</td>
</tr>
<tr>
<td></td>
<td>$p=10$</td>
<td>0.107/0.113</td>
<td>0.114/0.116</td>
<td>0.123/0.128</td>
<td>0.151/0.170</td>
</tr>
<tr>
<td></td>
<td>$p=11$</td>
<td>0.128/0.114</td>
<td>0.104/0.131</td>
<td>0.135/0.120</td>
<td>0.170/0.169</td>
</tr>
<tr>
<td></td>
<td>$p=13$</td>
<td>0.138/0.132</td>
<td>0.146/0.133</td>
<td>0.178/0.183</td>
<td>0.204/0.201</td>
</tr>
<tr>
<td>$S_{n,d,h}$</td>
<td>$d=1$, $h=1$</td>
<td>0.115/0.086</td>
<td>0.078/0.088</td>
<td>0.093/0.106</td>
<td>0.146/0.139</td>
</tr>
<tr>
<td></td>
<td>$d=3$, $h=1$</td>
<td>0.097/0.065</td>
<td>0.055/0.079</td>
<td>0.056/0.061</td>
<td>0.088/0.074</td>
</tr>
<tr>
<td></td>
<td>$d=1$, $h=2$</td>
<td>0.098/0.090</td>
<td>0.090/0.085</td>
<td>0.067/0.057</td>
<td>0.067/0.060</td>
</tr>
<tr>
<td></td>
<td>$d=3$, $h=2$</td>
<td>0.079/0.066</td>
<td>0.060/0.061</td>
<td>0.051/0.045</td>
<td>0.039/0.032</td>
</tr>
<tr>
<td></td>
<td>$d=1$, $h=3$</td>
<td>0.098/0.093</td>
<td>0.087/0.090</td>
<td>0.082/0.073</td>
<td>0.055/0.055</td>
</tr>
<tr>
<td></td>
<td>$d=3$, $h=3$</td>
<td>0.063/0.077</td>
<td>0.074/0.054</td>
<td>0.050/0.040</td>
<td>0.041/0.041</td>
</tr>
</tbody>
</table>
Table 5.4: Empirical power for bootstrapped test \( T_{n,p}^* \) and the fpca-based test \( S_{n,d} \); FAR(1)-model with Gaussian-/ Wiener-kernel; \( \mu(t) = \sin(t) \), change after 50% of the observations; nominal size \( \alpha = 0.1 \), sample size \( n \), bootstrap block length \( p \), number \( d \) of principal components \( d \)

<table>
<thead>
<tr>
<th>( n=50 )</th>
<th>( | \psi |_{L^2} )</th>
<th>0.1</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p = 4 )</td>
<td>1.000/1.000</td>
<td>0.997/1.000</td>
<td>0.988/0.986</td>
<td>0.927/0.894</td>
<td></td>
</tr>
<tr>
<td>( p = 5 )</td>
<td>0.999/0.999</td>
<td>0.998/0.999</td>
<td>0.975/0.972</td>
<td>0.861/0.850</td>
<td></td>
</tr>
<tr>
<td>( T_{n,p}^* )</td>
<td>( p = 6 )</td>
<td>1.000/0.999</td>
<td>0.996/0.998</td>
<td>0.968/0.971</td>
<td>0.859/0.830</td>
</tr>
<tr>
<td></td>
<td>( p = 7 )</td>
<td>1.000/1.000</td>
<td>0.996/0.999</td>
<td>0.969/0.957</td>
<td>0.858/0.828</td>
</tr>
<tr>
<td></td>
<td>( p = 8 )</td>
<td>0.998/0.998</td>
<td>0.989/0.990</td>
<td>0.961/0.949</td>
<td>0.808/0.808</td>
</tr>
<tr>
<td>( d = 1, h = 1 )</td>
<td>0.994/0.990</td>
<td>0.993/0.979</td>
<td>0.948/0.930</td>
<td>0.893/0.811</td>
<td></td>
</tr>
<tr>
<td>( d = 3, h = 1 )</td>
<td>0.999/0.997</td>
<td>0.998/0.993</td>
<td>0.991/0.990</td>
<td>0.989/0.947</td>
<td></td>
</tr>
<tr>
<td>( d = 1, h = 2 )</td>
<td>0.952/0.933</td>
<td>0.918/0.913</td>
<td>0.830/0.730</td>
<td>0.685/0.571</td>
<td></td>
</tr>
<tr>
<td>( S_{n,d} )</td>
<td>( d = 3, h = 2 )</td>
<td>0.033/0.022</td>
<td>0.020/0.023</td>
<td>0.015/0.020</td>
<td>0.018/0.014</td>
</tr>
<tr>
<td></td>
<td>( d = 1, h = 3 )</td>
<td>0.453/0.401</td>
<td>0.374/0.316</td>
<td>0.282/0.210</td>
<td>0.158/0.105</td>
</tr>
<tr>
<td></td>
<td>( d = 3, h = 3 )</td>
<td>0.028/0.026</td>
<td>0.027/0.022</td>
<td>0.028/0.024</td>
<td>0.020/0.022</td>
</tr>
</tbody>
</table>

Table 5.4 shows the empirical power of the test. We consider the same alternative as Torgovitski (2016), that is \( \mu(t) = \sin(t) \) and \( k = \lfloor n/2 \rfloor \). The power is very good and decreases only slightly as the dependence grows.

An alternative to our test is an adaption of the fpca-based method of Berkes et al. (2009) to weakly dependent data. The behavior of the test \( S_{n,d} \) in the case of a FAR(1)-model was investigated in Torgovitski (2016) using \( d \) principal components and a Bartlett-type estimator of the long run covariance matrix with bandwidth \( h \). We have copied the results of the simulation study of Torgovitski (2016) and added them to tables 5.3 and 5.4. Contrary to our results, the empirical size is clearly beneath the nominal size. Depending on the choice of projection dimension \( (d) \) and the selection of the bandwidth for variance estimation \( (h) \), the empirical power might vanish. In contrast our test has good power properties for all block lengths \( (p) \).

### 5.3.3 Finite sample performance of the Cramér-von Mises test

In a second simulation study we investigate the finite sample performance of the Cramér-von Mises-type change-point test. We are considering different block lengths \( p \) and three kinds of dependencies. The data generating process is an AR(1)-process satisfying

\[
X_t = a_1 X_{t-1} + \epsilon_t,
\]

where \( a_1 \) is a parameter and \( \epsilon_t \) is a white noise process.
Table 5.5: Empirical size for Cramér-von Mises/CUSUM test for a real-valued AR(1)-process; nominal size $\alpha = 0.05$, sample size $n$, bootstrap block length $p$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$p$</th>
<th>$a_1 = 0.2$</th>
<th>$a_1 = 0.5$</th>
<th>$a_1 = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>4</td>
<td>0.050/0.064</td>
<td>0.127/0.140</td>
<td>0.230/0.318</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.044/0.063</td>
<td>0.085/0.097</td>
<td>0.212/0.226</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.046/0.051</td>
<td>0.076/0.075</td>
<td>0.155/0.145</td>
</tr>
<tr>
<td>100</td>
<td>6</td>
<td>0.035/0.078</td>
<td>0.082/0.111</td>
<td>0.254/0.225</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.056/0.062</td>
<td>0.059/0.079</td>
<td>0.171/0.173</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.047/0.048</td>
<td>0.072/0.054</td>
<td>0.131/0.123</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.056/0.064</td>
<td>0.074/0.085</td>
<td>0.122/0.126</td>
</tr>
<tr>
<td>200</td>
<td>8</td>
<td>0.061/0.078</td>
<td>0.091/0.061</td>
<td>0.201/0.208</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.040/0.061</td>
<td>0.064/0.085</td>
<td>0.149/0.156</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.055/0.061</td>
<td>0.067/0.057</td>
<td>0.137/0.140</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.042/0.057</td>
<td>0.066/0.063</td>
<td>0.100/0.104</td>
</tr>
</tbody>
</table>

Table 5.6: Empirical power for Cramér-von Mises/CUSUM test for a real-valued AR(1)-process; change of height $\mu$, relative change-position $\tau$; nominal size $\alpha = 0.05$, sample size $n$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\mu = 0.5, \tau = 0.5$</th>
<th>$a_1 = 0.2$</th>
<th>$a_1 = 0.5$</th>
<th>$a_1 = 0.8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.230/0.233</td>
<td>0.161/0.156</td>
<td>0.207/0.194</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.302/0.315</td>
<td>0.280/0.262</td>
<td>0.206/0.206</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>0.669/0.700</td>
<td>0.456/0.462</td>
<td>0.258/0.295</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu = 1, \tau = 0.5$</th>
<th>$n$ = 50</th>
<th>0.686/0.678</th>
<th>0.313/0.431</th>
<th>0.351/0.335</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 100</td>
<td>0.847/0.851</td>
<td>0.695/0.708</td>
<td>0.419/0.373</td>
</tr>
<tr>
<td></td>
<td>n = 200</td>
<td>0.995/0.998</td>
<td>0.937/0.945</td>
<td>0.640/0.630</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\mu = 1, \tau = 0.2$</th>
<th>$n$ = 50</th>
<th>0.225/0.234</th>
<th>0.108/0.114</th>
<th>0.157/0.159</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 100</td>
<td>0.171/0.175</td>
<td>0.249/0.269</td>
<td>0.137/0.159</td>
</tr>
<tr>
<td></td>
<td>n = 200</td>
<td>0.734/0.773</td>
<td>0.530/0.513</td>
<td>0.268/0.295</td>
</tr>
</tbody>
</table>
with $a_1 \in \{0.2, 0.5, 0.8\}$ and $(\epsilon_i)_{i \geq 1}$ iid with $\epsilon_i \sim N(0, 1 - a_1^2)$. In all situations we have calculated critical values from $J = 999$ bootstrap-iterations and empirical size and power from $m = 1000$ iterations of the test. In addition, we have applied the classical CUSUM test to the data, which compares sample means. For the execution of this test see Section 5.3.2 and consider the special case $H = R$. The number of bootstrap-iterations is set to 999 too.

Table 5.5 reports empirical sizes under the hypothesis of no change. For the low correlation case ($a_1 = 0.2$) the performance is quite good, even for small sample sizes like $n = 50$. When $a_1 = 0.8$ the empirical size is drastically larger than the nominal one. This is typical for bootstrap tests due to an underestimation of covariances, see for example Doukhan et al. (2015). Altogether there are only marginal differences between Cramèr-von Mises and CUSUM test. Note that for the different tests, different choices of block length are advantageous.

Regarding the power of our test we choose for each sample size and AR-coefficient the block length that provides the best empirical size under this circumstances. We start with the following change-in-mean model:

$$ Y_i = \begin{cases} 
X_i & \text{for } i \leq \lfloor n \tau \rfloor \\
X_i + \mu & \text{for } i > \lfloor n \tau \rfloor 
\end{cases} $$

Table 5.6 gives an overview of the empirical power against this alternative for $\mu \in \{0.5, 1\}$ and $\tau \in \{0.2, 0.5\}$. We see that a level shift of height $\mu = 0.5$ in an AR-process with $a_1 = 0.8$ is to small to be detected. However, for larger shifts ($\mu = 1$) in the center of the observation period ($\tau = 0.5$) the power of our test is notably good. Finally, note that in order to detect early changes, a sufficiently large sample size is required and the dependence must not be too strong.

The CUSUM test is designed to detect changes in the mean. If critical values can be deduced from a known asymptotic distribution, the CUSUM test is supposed to have greater power then our test. However, if critical values are investigated by the bootstrap, table 5.6 indicates that both tests have similar power properties.

To illustrate the power of our test against several alternatives, consider a change in the skewness of a process. Therefore we need a second data generating process $X'_i = a_1 X'_{i-1} + \epsilon'_i$, independent
of the first one, and define

\[ Y_i = \begin{cases} 
X_i^2 + X_i'^2 & \text{for } i \leq \lfloor n\tau \rfloor \\
4 - (X_i^2 + X_i'^2) & \text{for } i > \lfloor n\tau \rfloor.
\end{cases} \]

Table 5.7 shows that against this alternative the power of the Cramér-von Mises test is excellent for \( n = 200 \) and coefficients \( a_1 \leq 0.5 \). The same table illustrates the power of the CUSUM test. Apparently, this test does not see changes in the skewness when the mean is unmodified.

To summarize, the Cramér-von Mises test can be used as an omnibus test for change in the marginal distribution without prespecifying the type of a change and the time series model. In the case of a change in mean, the power is not much lower compared to the classical CUSUM test. All in all, the test that is based on the Cramér-von Mises statistic seems advantageous.

### 5.4 Proofs of the main results

**Proof of Theorem 16.** Let \( W_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} (X_i - \mu) \). By Lemma 2.3.6 we have to verify the following three conditions:

(i) For each \( k \geq 1 \), \( P_k W_n \xrightarrow{D} W^k \) in \( D_{H_2}[0,1] \) (as \( n \to \infty \)),

(ii) \( W^k \xrightarrow{D} W \) in \( D_{H_2}[0,1] \) (as \( k \to \infty \)),

(iii) \( \lim \sup_{n \to \infty} E \left( \sup_{t \in [0,1]} \| W_n(t) - P_k W_n(t) \|^4 \right) \to 0 \) as \( k \to \infty \).

To show (i) we start with the special case \( H = \mathbb{R} \). Let \( EX_1 = 0 \). Then by Lemma 2.23 of Borovkova et al. (2001) we have

\[ \frac{1}{n} E \left( \sum_{i=1}^{n} X_i \right)^2 \to \sigma^2, \]

where \( \sigma^2 = \sum_{i=-\infty}^{\infty} E(X_0 X_i) \) and this series converges absolutely.

Furthermore, by Lemma 2.4. of Dehling et al. (2015) we have

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \xrightarrow{D} N(0,\sigma^2). \] (5.1)

In order to show convergence of the finite dimensional distributions of \( (W_n(t))_{t \in [0,1]} \), we will show

\[ (W_n(t), W_n(1) - W_n(t)) \xrightarrow{D} (\sigma W(t), \sigma(W(1) - W(t))), \] (5.2)

where \( W \) is standard Brownian motion in \( \mathbb{R} \). This can be easily adopted to dimensions higher than 2. Remember that \( (X_i)_{i \geq 1} \) is \( L_1 \)-near epoch dependent on an absolutely regular process
$(\epsilon_i)_{i \in \mathbb{Z}}$ and $\mathcal{F}^m_{-l} = \sigma(\epsilon_{-l}, \cdots, \epsilon_m)$. We proceed as in the proof of Theorem 21.1 in Billingsley (1968). Define

$$U_{n,|nt|} = \frac{1}{\sqrt{n}} E\left[\sum_{i=1}^{[nt]-2jn} X_i \mid \mathcal{F}_{-\infty}^{[nt]-jn}\right]$$

and

$$V_{n,|nt|} = \frac{1}{\sqrt{n}} E\left[\sum_{i=1+|nt|+2jn}^{n} X_i \mid \mathcal{F}^{\infty}_{|nt|+jn}\right],$$

for positive integers $j_n \to \infty$. Billingsley (1968) shows

$$|U_{n,|nt|} - W_n(t)| \xrightarrow{P} 0 \quad \text{and} \quad |V_{n,|nt|} - (W_n(1) - W_n(t))| \xrightarrow{P} 0,$$

and thus by (5.1) and Slutsky’s theorem we obtain for fixed $t$

$$U_{n,|nt|} \xrightarrow{D} \sigma W(t) \quad \text{and} \quad V_{n,|nt|} \xrightarrow{D} \sigma(1-W(t)). \tag{5.4}$$

Further, for all Borel sets we get by definition of $U_{n,|nt|}$ and $V_{n,|nt|}$

$$|P(U_{n,|nt|} \in H_1, V_{n,|nt|} \in H_2) - P(U_{n,|nt|} \in H_1) P(V_{n,|nt|} \in H_2)| \leq \alpha(F_{-\infty}^{[nt]-jn}, \mathcal{F}^{\infty}_{|nt|+jn}) \leq \alpha(j_n) \to 0, \tag{5.5}$$

as $n \to \infty$, where $\alpha(\cdot)$ is the strong mixing coefficient. $\alpha(j_n)$ converges to 0 because the $(\epsilon_i)_{i \in \mathbb{Z}}$ are absolute regular and this implies strong mixing. Combining (5.4) with (5.5) we arrive at

$$(U_{n,|nt|}, V_{n,|nt|}) \xrightarrow{D} (\sigma W(t), \sigma(1-W(t))),$$

where weak convergence takes place in $\mathbb{R}^2$. However, because of (5.3) this implies (5.2).

If we can show that the set

$$\left\{ \max_{s \leq t \leq s+\delta} \frac{1}{\delta} (W_n(t) - W_n(s))^2 \mid 0 \leq s \leq 1, 0 \leq \delta \leq 1, n \leq N(s, \delta) \right\} \tag{5.6}$$

is uniformly integrable, then according to Lemma 2.2 in Wooldridge and White (1988) $W_n$ is tight in $D[0,1]$, equipped with the Skorohod topology. Furthermore the weak limit is almost surely in $C[0,1]$.

So fix $s \in [0,1]$ and $\delta \in [0,1]$. By the proof of Lemma 2.24 in Borovkova et al. (2001) we obtain

$$E \left( \sum_{i=\lceil ns \rceil +1}^{n(\delta+s)} X_i \right)^4 \leq C(\lceil n(\delta+s) \rceil - \lceil ns \rceil)^2.$$
Next, Theorem 1 of Móricz (1976) together with the moment inequality stated above implies

\[
E \left( \max_{s \leq t \leq s + \delta} \left| \sum_{i=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} X_i \right| \right)^4 \leq C (\lfloor n(\delta + s) \rfloor - \lfloor ns \rfloor)^2. \tag{5.7}
\]

Now we will show uniform integrability of (5.6). Using first Hölder’s- and Markov’s inequality and then (5.7), one obtains

\[
E \left( \max_{s \leq t \leq s + \delta} \frac{1}{\delta} (W_n(t) - W_n(s))^2 \mathbf{1}_{\max \frac{1}{\delta} (W_n(t) - W_n(s))^2 \geq K} \right) \leq \frac{1}{K} \frac{1}{\delta^2} E \left( \max_{s \leq t \leq s + \delta} |W_n(t) - W_n(s)| \right)^4
\leq \frac{1}{K} \frac{1}{n^2 \delta^2} E \left( \max_{s \leq t \leq s + \delta} \left| \sum_{i=\lfloor ns \rfloor + 1}^{\lfloor nt \rfloor} X_i \right| \right)^4
\leq C \frac{1}{K} \frac{1}{n^2 \delta^2} (\lfloor n(\delta + s) \rfloor - \lfloor ns \rfloor)^2.
\]

Because the last term tends to 0 as \( K \to \infty \), (5.6) is uniformly integrable and the partial sum process converges in tight in \( D[0,1] \). Together with (5.2), this implies weak convergence towards a Brownian motion \( W \) with

\[
W(1) =_D N(0, \sigma^2).
\]

Now consider an arbitrary separable Hilbert space \( H \). For fixed \( h \in H \setminus \{0\} \), the sequence \( (\langle X_i, h \rangle)_{i \in \mathbb{N}} \) is a sequence of real valued random variables. The mapping \( x \mapsto \langle x, h \rangle \) is Lipschitz-continuous with constant \( \|h\| \) and therefore by Lemma 2.2 of Dehling et al. (2015), \( (\langle X_i, h \rangle)_{i \in \mathbb{N}} \) is \( L_1 \)-near epoch dependent on an absolute regular process with approximating constants \( (\|h\| a_m)_{m \in \mathbb{N}} \). Moreover, it has finite \( (4 + \delta) \)-moments, because

\[
E|\langle X_1, h \rangle|^{4+\delta} \leq \|h\|^{4+\delta} E\|X_1\|^{4+\delta} < \infty.
\]

Thus we can apply the functional central limit theorem in \( D[0,1] \) (proved in the lines above) and get

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \langle X_i, h \rangle \overset{D}{\to} W_h(t), \tag{5.8}
\]
where $W_h$ is a Brownian motion with $EW_h(1)^2 = \sigma^2(h)$ and

$$\sigma^2(h) = \sum_{i=-\infty}^{\infty} E\langle (X_0, h)(X_i, h) \rangle.$$

Define the covariance operator $S: H \to H$ by

$$\langle Sh_1, h_2 \rangle = \sum_{i=-\infty}^{\infty} E\langle (X_0, h_1)(X_i, h_2) \rangle.$$

Then $\langle Sh, h \rangle = \sigma^2(h)$ holds for all $h \in H \setminus \{0\}$.

Now we are able to verify condition (i) of Lemma 2.3.6. By the isometry and isomorphism between $H_k$ and $\mathbb{R}^k$ it suffices to show for all $k \geq 1$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} Y_i^{(k)} \overset{D}{\to} Y^{(k)}(t), \quad (5.9)$$

where $Y_i^{(k)} = (\langle X_i, e_1 \rangle, \ldots, \langle X_i, e_k \rangle)^t$ and $Y^{(k)}$ is Brownian motion in $\mathbb{R}^k$, whose covariance matrix corresponds to $S^k = P_kSP_k$. By (5.8), we obtain for all $k \geq 1$ and all $\lambda_1, \ldots, \lambda_k$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \langle X_i, \sum_{j=1}^{k} \lambda_j e_j \rangle \overset{D}{\to} W\sum_{j=1}^{k} \lambda_j e_j(t).$$

But this implies (5.9), because of the Cramér-Wold device, the arguments used for verifying (5.2) and the fact that univariate tightness in $D[0,1]$ implies tightness in $D_{\mathbb{R}^k}[0,1]$. Thus condition (i) is satisfied. For condition (ii) we need that $W^k \overset{D}{\to} W$ as $k$ goes to infinity. But this holds, because

$$W^k =_D P_k W$$

and

$$\sup_{t \in [0,1]} \|P_k W - W\| \to 0 \quad \text{a.s.,} \quad (5.10)$$

for $k \to \infty$. (5.10) holds pointwise due to Parseval’s identity. Uniform convergence then follows from the almost sure continuity of $W$.

Thus it remains to prove the validity of condition (iii). Define the operator $A_k: H \to H$ by $A_k = I - P_k$, where $I$ is the identity operator on $H$, and note that the mapping $h \mapsto A_k(h)$ is Lipschitz-continuous with Lipschitz-constant 1. Consequently, $(A_k(X_i))_{i \in \mathbb{N}}$ is a 1-approximating functional with the same constants as $(X_i)_{i \in \mathbb{N}}$. From Lemma 2.3.7 it follows

$$E\|A_k(X_1) + \cdots + A_k(X_n)\|^4 \leq Cn^2 \left(E\|A_k(X_1)\|^{4+\delta}\right)^{\frac{1}{1+\delta}}. \quad (5.11)$$
Observe that
\[
E \left( \sup_{t \in [0,1]} \| W_n(t) - P_k W_n(t) \|^4 \right) = \frac{1}{n^2} E \left( \max_{1 \leq m \leq n} \left\| \sum_{i=1}^m A_k(X_i) \right\|^4 \right)
\]
and note that the term on the right hand side is bounded by \( C (E \| A_k(X_1) \|^{4+\delta})^{\frac{1}{1+\delta}} \), due to (5.11) and Lemma 2.3.8. The constant \( C \) does not depend on \( k \) so it suffices to show
\[
E \| A_k(X_1) \|^{4+\delta} \xrightarrow{k \to \infty} 0.
\] (5.12)

By Parseval’s identity and the orthonormality of the \( e_i \) one obtains
\[
\| A_k(X_1) \|^2 = \sum_{i=k+1}^{\infty} \langle X_1, e_i \rangle^2 = \sum_{i=k+1}^{\infty} \langle X_1, e_i \rangle^2 \xrightarrow{k \to \infty} 0 \text{ a.s.}
\]

Further \( \| A_k(X_1) \|^{4+\delta} \leq \| X_1 \|^{4+\delta} < \infty \) almost surely and thus, by dominated convergence, (5.12) holds. But this implies condition (iii) and therefore finishes the proof.

**Proof of Theorem 17.** Assume \( E X_1 = 0 \) and define
\[
S^*_{n,i} := \frac{1}{\sqrt{p}} \sum_{j=(i-1)p+1}^{ip} (X^*_j - E^* X^*_j)
\]
and
\[
R^*_{n,kp}(t) := \frac{1}{\sqrt{kp}} \sum_{j=[kt]}^{\lfloor kpt \rfloor} (X^*_j - E^* X^*_j).
\]

In what follows, we decompose the \( W_{n,kp} \) into the partial sum process of the independent blocks and the remainder. In detail,
\[
W^*_{n,kp}(t) = \frac{1}{\sqrt{k}} \sum_{i=1}^{\lfloor kt \rfloor} S^*_{n,i} + R^*_{n,kp}(t).
\]

We start by proving that \( R^*_{n,kp} \) is negligible, i.e.
\[
R^*_{n,kp}(\cdot) \xrightarrow{P^*} 0 \text{ a.s.}
\] (5.13)
uniformly as \( n \to \infty \). Note, that \( R^*_{n,kp}(t) \) is the sum over the first \( l \) variables of a randomly generated block, where \( l = l(k,p,t) = \lfloor kpt \rfloor - \lfloor kt \rfloor p \). Thus, for fixed \( t \) we have
\[
\| R^*_{n,kp}(t) \| \leq \frac{1}{\sqrt{kp}} \max_{1 \leq j \leq k} \left\| \sum_{i=j(p-1)+1}^{j(p-1)+l} (X_i - E^* X^*_i) \right\|.
\]
Sequential block bootstrap in a Hilbert space with applications to change-point analysis

Taking the supremum over \( t \), we get

\[
\sup_{t \in [0,1]} \|R_{n,kp}^*(t)\| \leq \frac{1}{\sqrt{kp}} \max_{1 \leq j \leq k} \max_{1 \leq l \leq p} \left\| \sum_{i=j(p-1)+1}^{j(p-1)+l} (X_i - E^*X_i^*) \right\|
\]

\[=: Y_n.\]

We will show that \( Y_n \) converges to 0 almost surely.

For \( n \in \{2^{l-1} + 1, \ldots, 2^l\} \) observe that

\[
Y_n \leq \frac{2}{\sqrt{2^l}} \max_{j \leq k(2^{l-1})} \max_{1 \leq m \leq p(2^{l-1})} \left\| \sum_{i=j(p-1)+1}^{j(p-1)+m} (X_i - E^*X_i^*) \right\|
\]

\[=: Y'_l.\]

Taking the sum instead of the maximum, we can begin to bound the fourth moments of \( Y'_l \):

\[
E|Y'_l|^4 = \frac{16}{2^{2l}} E \left( \max_{j \leq k(2^{l-1})} \max_{m \leq p(2^{l-1})} \left\| \sum_{i=j(p-1)+1}^{j(p-1)+m} (X_i - E^*X_i^*) \right\|^4 \right)
\]

\[\leq \frac{16}{2^{2l}} \sum_{j=1}^{k(2^{l-1})} E \left( \max_{m \leq p(2^{l-1})} \left\| \sum_{i=j(p-1)+1}^{j(p-1)+m} (X_i - E^*X_i^*) \right\|^4 \right)
\]

\[= \frac{16k(2^{l-1})}{2^{2l}} E \left( \max_{m \leq p(2^{l-1})} \left\| \sum_{i=1}^{m} (X_i - E^*X_i^*) \right\|^4 \right).\]

The last line holds since \((X_i)_{i \in \mathbb{N}}\) and \(E^*X_i^*\) does not depend on the block in which \(X_i^*\) is, but only on the position of \(X_i^*\) in this block. We want to make use of Lemma A2. For \( p = p(2^{l}) \) and \( k = k(2^{l}) \) we obtain using the Minkowski inequality

\[
E \left\| \sum_{i=1}^{p} (X_i - E^*X_i^*) \right\|^4 = E \left\| \sum_{i=1}^{p} X_i - \frac{1}{k} \sum_{i=1}^{kp} X_i \right\|^4
\]

\[\leq \left\{ \left( E \left\| \sum_{i=1}^{p} X_i \right\|^4 \right)^{1/4} + \left( E \left\| \frac{1}{k} \sum_{i=1}^{kp} X_i \right\|^4 \right)^{1/4} \right\}^4
\]

\[= O(p^2).\]

In the last line we have used Lemma 2.3.7 and the fact that the first summand is the dominating term.
Next by virtue of Lemma 2.3.8 we obtain
\[
E \left( \max_{m \leq p(2^l)} \left\| \sum_{i=1}^{m} (X_i - E^* X_i^*) \right\|^4 \right) = O(p^2).
\]

Thus \( E|Y'_l|^4 = O\left(\frac{p^{(2^l)}}{2^l}\right) = O((2^{-\epsilon})^l) \), because of \( p(n) = O(n^{1-\epsilon}) \), see the definition of the block length. An application of the Markov inequality and the Borel-Cantelli Lemma implies that
\[
Y'_l \xrightarrow{l \to \infty} 0 \text{ a.s.}.
\]

Now \( Y_n \leq Y'_l \) for \( n \in \{2^{l-1}, \ldots, 2^l\} \) and thus \( Y_n \) converges almost surely to 0 as \( n \) tends to infinity. Finally, this leads to
\[
E^* \left( \sup_{t \in [0,1]} \| R_n^*(t) \| \right) \leq E^* Y_n = Y_n \to 0 \text{ a.s.}
\]

and we have just proved (5.13).

In order to verify convergence of the bootstrap process in \( D_H[0,1] \), it suffices to show that
\[
V^*_{n,kp}(t) = \frac{1}{\sqrt{k}} \sum_{i=1}^{[kt]} S^*_{n,i}
\]

converges to the desired Gaussian process.

We first establish the finite dimensional convergence. For \( 0 \leq t_1 < \cdots < t_l \leq 1 \) and \( l \geq 1 \) consider the increments
\[
(V^*_{n,kp}(t_1), V^*_{n,kp}(t_2) - V^*_{n,kp}(t_1), \cdots, V^*_{n,kp}(t_l) - V^*_{n,kp}(t_{l-1})).
\]

Note that the random variables \( S^*_{n,i} \) are independent, conditional on \( (X_i)_{i \in \mathbb{Z}} \), so it is enough to treat \( V^*_{n,kp}(t_i) - V^*_{n,kp}(t_{i-1}) \) for some \( i \leq l \). By the consistency of the bootstrapped sample mean of \( H \)-valued data (see Dehling et al. (2015)), there is a subset \( A \) of the underlying probability space with \( P(A) = 1 \), so that for all \( \omega \in A \) the central limit theorem holds:
\[
\frac{1}{\sqrt{k}} \sum_{i=1}^{k} S^*_{n,i} \overset{D}{\to} N, \quad (5.14)
\]

where \( N \) is a Gaussian \( H \)-valued random variable with mean zero and covariance operator
Sequential block bootstrap in a Hilbert space with applications to change-point analysis

$S: H \to H$ defined by

$$\langle Sx, y \rangle = \sum_{i=\infty}^\infty E[\langle X_0, x \rangle \langle X_i, y \rangle], \text{ for } x, y \in H.$$ 

For $\omega \in A$ and arbitrary $t_i > t_{i-1}$ it follows by (5.14) that

$$V_{n,kp}^*(t_i) - V_{n,kp}^*(t_{i-1}) = \frac{1}{\sqrt{k}} \sum_{i=\lfloor kt_{i-1} \rfloor + 1}^{\lfloor kt_i \rfloor} S_{n,i}^*$$

$$= \frac{\sqrt{\lfloor kt_1 \rfloor - \lfloor kt_{i-1} \rfloor}}{\sqrt{k}} \frac{1}{[kt_i] - [kt_{i-1}]} \sum_{i=\lfloor kt_{i-1} \rfloor + 1}^{\lfloor kt_i \rfloor} S_{n,i}^*$$

where $N$ is is the Gaussian random element described above. Thus the one-dimensional distributions converge almost surely. But because of the conditional independence this implies the finite dimensional convergence.

By Lemma 2.3.5, tightness will follow if we can show that

$$\lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{\delta} P^*\left( \sup_{0 \leq t \leq \delta} \|V_{n,kp}(t)\| > \epsilon \right) = 0 \ a.s. \quad (5.15)$$

for all $\epsilon > 0$.

Using first Chebychev's inequality and then Rosenthal's inequality (see Rosenthal (1970) and Ledoux and Talagrand (1991) for validity in Hilbert spaces) we obtain

$$\frac{1}{\delta} P^*\left( \sup_{0 \leq t \leq \delta} \frac{1}{\sqrt{k}} \left\| \sum_{i=1}^{\lfloor kt \rfloor} S_{n,i}^* \right\| > \epsilon \right)$$

$$\leq \frac{1}{\delta} \frac{1}{k^2 \epsilon^4} E^* \left( \max_{1 \leq j \leq [k \delta]} \left\| \sum_{i=1}^{j} S_{n,i}^* \right\| \right)$$

$$\leq \frac{1}{\delta} \frac{1}{k^2 \epsilon^4} C \left\{ [k \delta] E^* \|S_{n,1}^*\|^4 + \left( [k \delta] E^* \|S_{n,1}^*\|^2 \right)^2 \right\}$$

$$\leq C \frac{1}{\delta} \frac{k \delta}{k^2 \epsilon^4} E^* \|S_{n,1}^*\|^4 + C \frac{1}{\delta} \frac{k^2 \delta^2}{k^2 \epsilon^4} (E^* \|S_{n,1}^*\|^2)^2$$

$$= I_n + II_n,$$

where $I_n$ and $II_n$ are the respective summands. By the construction of the bootstrap sample
and Minkowski’s inequality we get
\[
I_n = \frac{1}{k^4} \frac{1}{k} \sum_{i=1}^{k} \left( \frac{1}{\sqrt{p}} \left\| \sum_{j \in B_i} (X_j - \bar{X}_{n,kp}) \right\|^4 \right) = C \frac{1}{e^4} \frac{1}{k^2} \sum_{i=1}^{k} \left( \frac{1}{\sqrt{p}} \left\| \sum_{j \in B_i} X_j \right\|^4 \right) + C \frac{1}{e^4} \frac{1}{k} \left\| \bar{X}_{n,kp} \right\|^4
\]
\[
= \tilde{I}_{n,1} + \tilde{I}_{n,2}.
\]

A suitable strong Law of Large numbers (see Lemma 2.7 in Dehling et al. (2015)) yields
\[
\frac{p^{1/2}}{k^{1/4}} \bar{X}_{n,kp} = \frac{1}{(kp)^{1/2}k^{1/2+1/4}} \sum_{i=1}^{k} X_i \to 0 \ \text{a.s.,}
\]
as \(n \to \infty\). Hence, \(\tilde{I}_{n,2}\) converges almost surely to 0. Regarding \(\tilde{I}_{n,1}\), note that for \(n \in \{2^{l-1}, \ldots, 2^l\}\)
\[
\tilde{I}_{n,1} \leq 16C \frac{1}{e^4} k^{(2^l)} \sum_{i=1}^{k(2^l)} \left( \frac{1}{\sqrt{p(2^l)}} \left\| \sum_{j \in B_i} X_j \right\|^4 \right) := I_{l,1}.
\]

We get by a fourth moment bound (see Lemma 2.3.7)
\[
E(I_{l,1}) = O \left( 1/k(2^l) \right) = O \left( 2^{-l/4} \right),
\]
because \(k = \lfloor n/p(n) \rfloor\) and \(p(n) = O(n^{1-\epsilon})\). Hence, by Markov’s inequality and the Borel-Cantelli Lemma \(I_{l,1} \to 0\) almost surely for \(l \to \infty\). Consequently, \(\tilde{I}_{n,1} \to 0\) almost surely for \(n \to \infty\) and thus \(I_1 \to 0\).

In Dehling et al. (2015) it is shown that \(E^* \| S_{n,1}^* \|^2\) converges almost surely to \(E \| N \|^2\), where \(N\) is Gaussian with the covariance operator defined above. Therefore, \(E \| N \|^2\) is almost surely bounded and we obtain
\[
II_n = \delta \frac{\epsilon}{e^4} (E^* \| S_{n,1}^* \|^2)^2 \overset{n \to \infty}{\longrightarrow} \delta \frac{\epsilon}{e^4} (E \| N \|^2)^2 \overset{\delta \to \infty}{\longrightarrow} 0 \ \text{a.s.}
\]
which implies (5.15) and therefore finishes the proof. \(\square\)
6 Convolved subsampling estimation with applications to block bootstrap

Subsampling and block bootstrap are two common nonparametric tools for statistical inference under dependence; see Politis et al. (1999) and Lahiri (2003), respectively, for monographs on these.

As noted in Politis et al. (1999) (cf. sec. 3.9), subsampling is often valid under weak assumptions about the dependent process, basically requiring that a non-degenerate (possibly non-normal) limit exist for the sampling distribution being approximated. In contrast, the block bootstrap applies to mean-like statistics with normal limits and typically requires comparatively much stronger assumptions for its validity. Case-by-case treatments are commonly needed to validate the bootstrap across differing dependence conditions. However, while perhaps not widely recognized, subsampling can in fact be used to verify the block bootstrap in some cases, which is a theme of this work.

We investigate estimators defined by the $k$-fold self-convolution of a subsampling distribution, and establish a new and general theory for their consistency to normal limits. There are two basic motivations for considering such convolved subsampling. The first is that, in the fundamental case of sample means, the block bootstrap estimator is a $k$-fold self-convolution of a subsampling distribution (centered and normalized), where the level $k$ of convolution corresponds to the number of resampled blocks. This observation was originally noted by Politis et al. (1999), who suggested this aspect as a potential technique for showing the validity of the bootstrap. Specifically, they conjectured that convolved subsampling might provide a route for establishing the block bootstrap under minimal conditions for non-stationary, strongly mixing processes, in analogy to bootstrap results existing for stationary, mixing series due to Radulovic (1996b, 2012). For the bootstrap under dependence, the findings of Radulovic (1996b, 2012) for the sample mean have stood out as an exception, verifying the method under the same weak assumptions as subsampling (i.e., conditions essentially needed for a limit law to exist).

By investigating the convolved subsampling approach here, we can answer the above conjecture affirmatively. Moreover, we show convolved subsampling leads to a simple and unified procedure for establishing the block bootstrap for sample means for further types of processes under much weaker conditions than previously considered, such as linear time processes, long-memory sequences, (non-stationary) almost periodic time series, and spatial fields. Hence, convolved subsampling estimation allows for bootstrap consistency under dependence to be
generally extended under the same weak assumptions used by subsampling, containing the conclusions of Radulovic (1996b, 2012) for stationary series as a special case.

While connections to the bootstrap are useful, our study of convolved subsampling estimation is intended to be broad, applying also to general statistics with normal limits and with arbitrary levels of convolution. Consistency results often do not require particular assumptions about the underlying dependent process, but are rather formulated in terms of mild convergence properties of the original subsampling distribution and its variance. Furthermore, we show that a consistent subsampling variance is not only sufficient, but essentially necessary, for the consistency of convolved subsampling (and the block bootstrap in some cases). Due to its importance, we also provide tools for verifying the consistency of subsampling variance estimators.

For general statistics beyond the sample mean, the convolved subsampling distribution may differ from the block bootstrap, which relates to a second motivation for our development. That is, a general theory for convolved subsampling is of interest in its own right, as the approach can be computationally less demanding than the block bootstrap while also potentially enhancing ordinary subsampling for approximating sampling distributions with normal limits. In fact, there has been recent interest in establishing generalized types of subsampling estimation for complicated statistics under various dependence structures, where numerical studies suggest such methods exhibit better finite sample performance than standard subsampling; for example, see Lenart (2016) and Sharipov et al. (2016b) for spectral estimates and U-statistics, respectively, with time series. While not formally recognized as such, however, these proposed methods are exactly convolved subsampling estimators. By exploiting this realization, our results can facilitate future work and allow such previous findings with generalized subsampling to be demonstrated in an alternative, simpler manner with weaker assumptions.

6.1 Fundamental results for convolved subsampling

6.1.1 Convolved subsampling and connection to the bootstrap

Consider data $X_1, \ldots, X_n$ from a real-valued process governed by a probability structure $P$. For concreteness, we may envision such observations arising from a time series process $\{X_t\}$, though spatial and other data schemes may be treated as well. Based on $X_1, \ldots, X_n$, consider the problem of approximating the distribution of

$$T_n \equiv \tau_n(t_n(X_1, \ldots, X_n) - t(P)),$$

involving an estimator $t_n \equiv t_n(X_1, \ldots, X_n)$ of a parameter $t(P)$ and a sequence of positive scaling factors $\tau_n$ yielding a distributional limit for $T_n$. For example, if $t_n(X_1, \ldots, X_n) \equiv \bar{X}_n = \sum_{i=1}^n X_i/n$ is the sample mean, then $t(P)$ may correspond to a common process mean $\mu$ and
$T_n$ may be defined with usual scaling $\tau_n = \sqrt{n}$ under weak time dependence. Denote the sampling distribution function of $T_n$ as $F_n(x) = P(T_n \leq x)$, $x \in \mathbb{R}$.

We next define the subsampling estimator of $F_n$; see Politis and Romano (1994a). For a positive integer $b \equiv b_n < n$, let \{$(X_i, \ldots, X_{i+b-1}) : i = 1, \ldots, N_n$\} denote the set of $N_n \equiv n - b + 1$ overlapping data blocks, or subsamples, of length $b$. To keep blocks relatively small, the block size is often assumed to satisfy 

\[
\frac{b-1}{n} + \frac{b}{n} + \frac{\tau_b}{\tau_n} \to 0 \quad \text{as} \quad n \to \infty.
\]

For each subsample, we compute the statistic as $t_{n,b,i} = t_b(X_i, \ldots, X_{i+b-1})$ and define a “scale $b$” version of $T_n \equiv \tau_n(t_n(X_1, \ldots, X_n) - t(P))$ as $\tau_b[t_{n,b,i} - t_n]$ for $i = 1, \ldots, N_n$. Letting $I(\cdot)$ denote the indicator function, the subsampling estimator of $F_n$ is given by

\[
S_{n,\text{SUB}}(x) = \frac{1}{N_n} \frac{1}{N_n} \sum_{i=1}^{N_n} I\left(\tau_b[t_{n,b,i} - t_n] \leq x\right), \quad x \in \mathbb{R},
\]  

(6.1)

or the empirical distribution of subsample analogs \{$\tau_b[t_{n,b,i} - t_n]$\}$_{i=1}^{N_n}$ (cf. Politis et al. (1999)). Suppose that $S_{n,\text{SUB}}$ is consistent for the distribution of $T_n$, where the latter has an asymptotically normal $N(0, \sigma^2)$ limit for some $\sigma > 0$, that is, as $n \to \infty$,

\[
T_n \xrightarrow{d} N(0, \sigma^2),
\]  

(6.2)

\[
\sup_{x \in \mathbb{R}} |S_{n,\text{SUB}}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0,
\]  

(6.3)

where $\Phi(\cdot)$ is the standard normal distribution function. We wish to consider estimators of the distribution $F_n$ of $T_n$ formed by self-convolutions of the subsampling estimator $S_{n,\text{SUB}}$. This provides a general class of block resampling estimators in its own right, but also has explicit connections to block bootstrap estimators in the important case that the statistic of interest $t_n(X_1, \ldots, X_n) = \bar{X}_n$ is a sample mean, as described next.

### 6.1.2 Convolved subsampling and connections to block bootstrap

Let $k_n \in \mathbb{N}$ be a sequence of positive integers and define a triangular array \{${Y_{n,1}^*, \ldots, Y_{n,k_n}^*}$\}$_{n \geq 1}$, where, for each $n$, \{${Y_{n,i}^*}$\}_{i=1}^{k_n} are iid variables following the subsampling distribution $S_{n,\text{SUB}}$, as determined by (6.1) from data $X_1, \ldots, X_n$. For $n \geq 1$, define a centered and scaled sum

\[
Z_n^* = \frac{1}{\sqrt{k_n}} \sum_{j=1}^{k_n} (Y_{n,j}^* - m_{n,\text{SUB}})
\]  

(6.4)

where $m_{n,\text{SUB}} \equiv \int xdS_{n,\text{SUB}}(x) = N_n^{-1} \sum_{i=1}^{N_n} \tau_b[t_{n,b,i} - t_n]$ is the mean of the subsampling distribution $S_{n,\text{SUB}}$, and let

\[
C_{n,k_n}(x) \equiv P_s(Z_n^* \leq x), \quad x \in \mathbb{R},
\]
denote the induced resampling distribution $P_*$ of $Z_n^*$. Then, $C_{n,k_n}$ represents the $k_n$-fold self-convolution of the subsampling distribution $S_{n,SUB}$, with appropriate centering/scaling adjustments. That is,

$$C_{n,k_n}(x) = S_{n,SUB} * S_{n,SUB} * \cdots * S_{n,SUB}(x \sqrt{k_n} + k_n m_n), \quad x \in \mathbb{R}.$$  

We consider $C_{n,k_n}$ as an estimator of the distribution $F_n$ of $T_n$ and formulate general conditions under which this convolved subsampling distribution is also consistent.

As suggested earlier, such results have direct implications for block bootstrap estimation as well, because the convolved subsampling estimator $C_{n,k_n}$ exactly matches a block bootstrap estimator in the basic sample mean case $t_n(X_1, \ldots, X_n) = \bar{X}_n$. To illustrate, consider approximating the distribution of $T_n = \sqrt{n}(\bar{X}_n - \mu)$ where $t(P) \equiv \mu = E\bar{X}_n$ and $\tau_n = \sqrt{n}$. In this setting, the block bootstrap uses an analog

$$T^*_n = \sqrt{n_1}(\bar{X}^*_{n_1} - E_{*} \bar{X}^*_{n_1})$$

based on the average $\bar{X}^*_{n_1} \equiv n_1^{-1} \sum_{i=1}^{n_1} X^*_i$ from a block bootstrap sample $X^*_1, \ldots, X^*_{n_1}$ of size $n_1 \equiv k_n b$, which is defined by drawing $k_n$ blocks of length $b$, independently and with replacement, from the subsample collection $\{(X_i, \ldots, X_{i+b-1}) : i = 1, \ldots, N_n\}$ and pasting these together (where above $E_{*} \bar{X}^*_{n_1} = N_n^{-1} \sum_{i=1}^{N_n} b^{-1} \sum_{j=i}^{i+b-1} X_j$ denotes the bootstrap expectation of $\bar{X}^*_{n_1}$); see ch. 2, Lahiri (2003). Most typically, the number of blocks resampled is taken as $k_n = \lfloor n/b \rfloor \to \infty$ so that the bootstrap sample re-creates the approximate length $\lfloor n/b \rfloor b \approx n$ of the original sample. The bootstrap distribution of $T^*_n$ here is then equivalent to the convolved subsampling distribution $C_{n,k_n}$. This is because $T^*_n$ has the same resampling distribution as $Z_n^*$ in (6.4) as a sum of $k_n$ iid block averages $(Y^*_{n,i} - m_{n,SUB})/\sqrt{k_n}$, with each $Y^*_{n,i}$ drawn from $S_{n,SUB}$ in (6.1) where $t_n = \bar{X}_n$ and $\tau_n[t_{n,b,i} - t_n] = \sqrt{b}[b^{-1} \sum_{j=i}^{i+b-1} X_j - X_n]$, $1 \leq i \leq N_n$, for the sample mean case. Consequently, if convolved subsampling estimators $C_{n,k_n}$ are shown to be valid under weak conditions, such results entail that block bootstrap estimation is as well. In the following, we make comprehensive use of the fact that $C_{n,k_n}$ is always and exactly a block bootstrap estimator whenever the underlying statistic $t_n(X_1, \ldots, X_n) = \bar{X}_n$ is a sample mean; this holds true across all the varied dependent data structures considered here, including settings where the block bootstrap formulation (6.5) itself requires some modification (cf. long-range dependence in Sections 6.2.3 and 6.2.4).
6.1.3 Limit theorems for the convolved distribution

From (6.1) and the subsampling mean $m_{n,\text{SUB}} \equiv \int xdS_{n,\text{SUB}}(x) = N_n^{-1} \sum_{j=1}^{N_n} \tau_b[t_{n,b,i} - t_n]$, we have the variance of the original subsampling distribution $S_{n,\text{SUB}}$ as

$$\hat{\sigma}^2_{n,\text{SUB}} \equiv \int (x - m_{n,\text{SUB}})^2 dS_{n,\text{SUB}}(x) = \frac{1}{N_n} \sum_{j=1}^{N_n} (\tau_b[t_{n,b,i} - t_n] - m_{n,\text{SUB}})^2,$$

which estimates the asymptotic variance $\sigma^2$ of $T_n$ as in (6.2) (cf. Politis et al. (1999)). Note that $\hat{\sigma}^2_{n,\text{SUB}}$ is also the variance of the convolved subsampling distribution $C_{n,k_n}$ (i.e., the variance of the iid sum from (6.4)). Correspondingly, $\hat{\sigma}^2_{n,\text{SUB}}$ is then a block bootstrap variance estimator when applied to sample means.

This section provides basic distributional results for convolved subsampling estimators, describing when and how these have normal limits. These findings do not involve particular assumptions about the process $\{X_t\}$, but are instead expressed through properties of the original subsampling distribution $S_{n,\text{SUB}}$ and, specifically, convergence of the subsampling variance $\hat{\sigma}^2_{n,\text{SUB}}$. Such subsampling properties can often be verified under weak assumptions about a process, allowing the limit behavior of convolved estimators $C_{n,k_n}$ and the block bootstrap, to be established under minimal conditions. Theorem 18 and Proposition 6.1.1 address the important case where the original subsampling distribution $S_{n,\text{SUB}}$ has a normal limit (6.3), as is often natural when the statistic $T_n \overset{d}{\to} N(0,\sigma^2)$ is asymptotically normal. These findings are expected to be the most practical for establishing convolved subsampling $C_{n,k_n}$ estimation with normal targets (6.2). Dropping the condition that $S_{n,\text{SUB}}$ converges to a normal law but assuming convolved estimators $C_{n,k_n}$ are based on increasing convolution $k_n \to \infty$ of $S_{n,\text{SUB}}$, Theorem 19 and Corollary 6.1.2 characterize the convergence of $C_{n,k_n}$ to normal limits through the subsampling variance $\hat{\sigma}^2_{n,\text{SUB}}$. In many problems involving the block bootstrap for sample means (cf. Section 6.2), where $T_n$ has a normal limit (6.2), these results provide both necessary and sufficient conditions for the validity of the block bootstrap as well as convolved subsampling generally. Finally, because convergence $\hat{\sigma}^2_{n,\text{SUB}} \overset{p}{\to} \sigma^2$ of the subsampling variance emerges as central to the behavior of convolved estimators $C_{n,k_n}$, Section 6.1.4 develops basic results for establishing this feature.

Theorem 18 provides a sufficient condition for the general validity of the convolved estimator $C_{n,k_n}$ via fundamental subsampling quantities, $S_{n,\text{SUB}}$ and $\hat{\sigma}^2_{n,\text{SUB}}$.

**Theorem 18.** Suppose (6.3) holds (i.e., $\sup_{x \in \mathbb{R}} |S_{n,\text{SUB}}(x) - \Phi(x/\sigma)| \overset{p}{\to} 0$) and $\hat{\sigma}^2_{n,\text{SUB}} \overset{p}{\to} \sigma^2 > 0$ as $n \to \infty$. Then,

$$\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \overset{p}{\to} 0 \quad \text{as } n \to \infty$$

for any positive integer sequence $k_n$.

Furthermore, when (6.2) holds additionally (i.e., $T_n \overset{d}{\to} N(0,\sigma^2)$), then $C_{n,k_n}$ is consistent for
the distribution $F_n$ of $T_n$,

$$\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - F_n(x)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$ 

To re-iterate, the integer sequence $k_n$, $n \geq 1$, need not even be convergent in Theorem 18. The consistency of the subsampling variance estimator $\hat{\sigma}^2_{n,\text{SUB}}$ automatically guarantees that, for any amount $k_n$ of convolution of $S_{n,\text{SUB}}$, the convolved subsampling estimator $C_{n,k_n}$ will have a normal limit if the subsampling distribution $S_{n,\text{SUB}}$ does. In other words, if (6.2)-(6.3) hold so that $S_{n,\text{SUB}}$ is consistent, then $C_{n,k_n}$ will be as well provided $\hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{P} \sigma^2$. When the statistic $t_n(X_1, \ldots, X_n) = \bar{X}_n$ is a sample mean, then $C_{n,k_n}$ again denotes a block bootstrap estimator based on $k_n$ resampled blocks, which is thereby consistent under Theorem 18 for any sequence $k_n$, including the common choice $k_n = \lceil n/b \rceil \to \infty$.

Proposition 6.1.1 next characterizes the convolved subsampling estimator $C_{n,k_n}$ under bounded levels $k_n$ of convolution. In this case, a normal limit for the subsampling estimator $S_{n,\text{SUB}}$ entails the same for the convolved estimator $C_{n,k_n}$, provided the mean $m_{n,\text{SUB}} \equiv \int x dS_{n,\text{SUB}}(x)$ of the subsampling distribution converges to zero. But, if the subsampling mean $m_{n,\text{SUB}}$ converges in this fashion, a normal limit for $C_{n,k_n}$ with bounded $\{k_n\}$ is equivalent to a normal limit for the original subsampling distribution $S_{n,\text{SUB}}$.

**Proposition 6.1.1.** Suppose $\sup_n k_n < \infty$.

(i) If (6.3) holds (i.e., $\sup_{x \in \mathbb{R}} |S_{n,\text{SUB}}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0$), then

$$\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0 \quad \text{as } n \to \infty$$

if and only if $m_{n,\text{SUB}} \equiv \int x dS_{n,\text{SUB}}(x) \xrightarrow{P} 0$.

(ii) If $m_{n,\text{SUB}} \xrightarrow{P} 0$ as $n \to \infty$, then (6.3) holds if and only if

$$\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0 \quad \text{as } n \to \infty.$$ 

When the original subsampling estimator $S_{n,\text{SUB}}$ is consistent for a distribution with a normal limit (i.e., (6.2)-(6.3)), both Theorem 18 and Proposition 6.1.1 show that the convolved subsampling estimator $C_{n,k_n}$ is consistent under an additional subsampling moment condition. With bounded levels $k_n$ of convolution, the additional condition under Proposition 6.1.1 is that the subsampling mean converge $m_{n,\text{SUB}} \xrightarrow{P} 0$. But, for general and potentially unbounded $k_n$, the additional condition from Theorem 18 for consistency of $C_{n,k_n}$ is a convergent subsampling variance $\hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{P} \sigma^2$. With diverging amounts $k_n \to \infty$ of convolution, which is often encountered in practice and in connection to the block bootstrap, it turns out that convergence $\hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{P} \sigma^2$ is also necessary for consistency of the convolved estimator $C_{n,k_n}$, as treated in the next section.
We next consider the behavior of convolved subsampling estimators for unbounded convolution $k_n \to \infty$ as $n \to \infty$, which arises, for example, with the block bootstrap $C_{n,k_n}$ for sample means with $k_n = \lfloor n/b \rfloor$ resampled blocks. Results here do not explicitly require convergence of the original subsampling estimator $S_{n, SUB}$ to a normal limit (6.3). While a reasonable condition in problems where the target quantity $T_n \overset{d}{\to} N(0, \sigma^2)$ is asymptotically normal, limits for $S_{n, SUB}$ are not directly necessary for convolved estimators $C_{n,k_n}$ to yield valid normal limits from increasing convolution $k_n$ of $S_{n, SUB}$. However, convergence of the subsampling variance $\hat{\sigma}^2_{n, SUB}$ is crucial, as shown in Theorem 19.

**Theorem 19.** Suppose $k_n \to \infty$ and $\int_{|x| \geq \sqrt{k_n} \epsilon} x^2 dS_{n, SUB}(x) \overset{P}{\to} 0$ for each $\epsilon > 0$ as $n \to \infty$.

(i) Then,

$$\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \overset{P}{\to} 0$$

if and only if $\hat{\sigma}^2_{n, SUB} \overset{P}{\to} \sigma^2 > 0$ as $n \to \infty$.

(ii) When $\hat{\sigma}^2_{n, SUB} \overset{P}{\to} \sigma^2 > 0$ as $n \to \infty$, then

$$\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - F_n(x)| \overset{P}{\to} 0$$

if and only if (6.2) holds (i.e., $\sup_{x \in \mathbb{R}} |F_n(x) - \Phi(x/\sigma)| \to 0$ for the distribution $F_n$ of $T_n$).

For an unbounded sequence $k_n \to \infty$ of convolution (e.g., block bootstrap with $k_n = \lfloor n/b \rfloor$ concatenated blocks), Theorem 19 imposes no direct assumption on the convergence of the original subsampling distribution, but rather that $S_{n, SUB}$ fulfills a mild truncated second moment property. From this, the consistency of the convolved subsampling estimator $C_{n,k_n}$ to a normal limit is completely determined by the behavior of the subsampling variance $\hat{\sigma}^2_{n, SUB}$ under Theorem 19. Furthermore, when $\hat{\sigma}^2_{n, SUB}$ converges, the convolved estimator $C_{n,k_n}$ is again valid for estimating the distribution $F_n$ of the target quantity $T_n$ with a normal limit (Theorem 19(ii)).

The following corollary of Theorem 19 shows that a convolved estimator $C_{n,k_n}$ will quite generally have a normal limit, provided that the subsampling variance converges $\hat{\sigma}^2_{n, SUB} \overset{P}{\to} \sigma^2 > 0$ and that some other basic feature exists for the subsampling distribution $S_{n, SUB}$ or for composite statistics $\{\tau_b[t_{n,b,i} - t_n] \equiv \tau_b[t_b(X_1, \ldots, X_{i-b+1}) - t_n(X_1, \ldots, X_n)]\}_{i=1}^{N_n} \equiv n^{-b+1}$ defining $S_{n, SUB}$ in (6.1). Essentially, Corollary 6.1.2 entails that the truncated second moment assumption in Theorem 19 is mild in conjunction with $\hat{\sigma}^2_{n, SUB} \overset{P}{\to} \sigma^2$.

**Corollary 6.1.2.** Suppose one of the following conditions (C.1)-(C.4) holds:

(C.1) for some distribution $J_0$ with variance $\sigma^2 > 0$, $S_{n, SUB}(x) \overset{P}{\to} J_0(x)$ as $n \to \infty$ for any continuity point $x \in \mathbb{R}$ of $J_0$;

(C.2) for some $\epsilon_0 > 0$, $N_n^{-1} \sum_{i=1}^{N_n} [\tau_b(t_{n,b,i} - t_n)]^{2+\epsilon_0} = O_p(1)$.
(C.3) the subsample-based sequence \( \{ T^2_{b,i} \equiv \tau_b^2(t_{n,b,i} - t(P))^2 : i = 1, \ldots, N_n \} \) is uniformly integrable and \( T_n \equiv \tau_n(t_n - t(P)) = O_p(\tau_n/\tau_b) \)

(C.4) \( \{ X_t \} \) is stationary, \( \{ T^2_n : n \geq 1 \} \) is uniformly integrable, and \( \tau_b/\tau_n = O(1) \).

Then, as \( n \to \infty \),

\[
\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{p} 0
\]

for any sequence \( k_n \) with \( \lim_{n \to \infty} k_n = \infty \) if and only if \( \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 > 0 \).

**Remark 6.1.3.** For reference, note \( \tau_b/\tau_n \to 0 \) often holds with subsample scaling as \( n \to \infty \) so that, for example, the conditions \( \tau_b/\tau_n = O(1) \) and \( T_n = O_p(\tau_n/\tau_b) \) are mild.

Hence, if \( k_n \to \infty \) and \( \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 \), then the convolved estimator \( C_{n,k_n} \) will converge to a normal limit if the subsampling distribution \( S_{n,\text{SUB}} \) is convergent (C.1) or has an appropriate stochastically bounded moment (C.2), or if the subsampling statistics related to computing \( S_{n,\text{SUB}} \) have uniformly integrable second moments (C.3)-(C.4). Condition (C.4) is a special case of (C.3) under stationarity, and corresponds to an underlying assumption of Radulovic (1996b, 2012) for examining the block bootstrap estimator \( C_{n,k_n} \) of a sample mean with stationary, mixing processes; see also Remark 2 to follow. When restricted to Condition (C.1), the “\( \Leftarrow \)” part of Corollary 6.1.2 corresponds to an initial convolved subsampling result due to Politis et al. (1999) (Proposition 4.4.1) for unbounded convolution \( k_n \to \infty \), which was developed for establishing the block bootstrap estimator \( C_{n,k_n} \) for the sample mean of non-stationary data, as re-considered here in Section 6.2.2. Note that, for inference with \( T_n \) having a normal \( N(0, \sigma^2) \) limit (6.2), Condition C.1 in Corollary 6.1.2 is perhaps most natural and approached by verifying convergence \( S_{n,\text{SUB}} \) to a normal (6.3). In which case, the implication of Corollary 6.1.2 (involving \( k_n \to \infty \)) for guaranteeing that convolved subsampling and block bootstrap estimators replicate normal limits when \( \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 \) also becomes a special case of Theorem 18 (involving any \( k_n \)).

**Remark 6.1.4.** For block bootstrap estimation of the sample mean \( T_n = \sqrt{n}(\bar{X}_n - E X_1) \) with strongly mixing, stationary processes, Radulovic (1996b, 2012) provides necessary and sufficient conditions for convergence of \( C_{n,k_n} \) (with \( k_n = [n/b] \to \infty \)) to a normal limit, assuming \( \{ T^2_n : n \geq 1 \} \) is uniformly integrable. Under such assumptions, the main result there is that normal limits for both \( C_{n,k_n} \) and \( T_n \) are equivalent. In comparison, the necessary and sufficient conditions for normality of the block bootstrap estimator \( C_{n,k_n} \) for a mean in Theorem 19 are perhaps more basic in that the conclusions of Radulovic (1996b, 2012), under the additional assumptions made there, follow from Theorem 19 (cf. Corollary 6.1.2). In this sense, Theorem 19 broadly re-frames the findings in Radulovic (1996b, 2012), by not involving particular process assumptions (i.e., stationarity or mixing) and applying to convolved subsampling estimators \( C_{n,k_n} \) with general statistics and arbitrarily increasing convolution levels \( k_n \to \infty \).
Further connections to, and extensions of, the results of Radulovic (1996b, 2012) are made in Section 6.2.1 for strongly mixing processes.

### 6.1.4 Consistency of the subsampling variance

Theorems 18-19 demonstrate that the subsampling variance $\hat{\sigma}^2_{n,\text{SUB}}$ plays a key role in the convergence of the convolved subsampling estimator $C_{n,k,n}$ generally, and of the block bootstrap for the sample mean in particular. However, convergence of the subsampling distribution $S_{n,\text{SUB}}$ itself is often much easier to directly establish under weak assumptions about the process $\{X_t\}$; see Politis et al. (1999) and Section 6.2 to follow. This raises a further question considered next: if one knows that subsampling estimator $S_{n,\text{SUB}}$ is consistent (6.3) for a normal limit, then when will the subsampling variance $\hat{\sigma}^2_{n,\text{SUB}}$ be convergent as well, thereby guaranteeing (from Theorem 18) that the convolved estimator $C_{n,k,n}$ is also consistent? As shown in Theorem 20, a general characterization is possible as well as simple sufficient conditions based on moment properties of subsample statistics (e.g., $T^2_b$).

For $n \geq 1$, recall $T_n \equiv \tau_n(t_n(X_1, \ldots, X_n) - t(P))$ and additionally define $T_{n,i} \equiv \tau_n(t_n(X_{i}, \ldots, X_{i+n-1}) - t(P))$ for $i \geq 1$ from the statistic applied to $(X_{i}, \ldots, X_{i+n-1})$. Based on $N_n \equiv n - b + 1$ subsample observations of length $1 \leq b \equiv b_n < n$, define a distribution function

$$D_{n,b}(x) \equiv \frac{1}{N_n} \sum_{i=1}^{N_n} P(T_{b,i} \leq x), \quad x \in \mathbb{R},$$

as an average of subsample-based probabilities.

**Theorem 20.** Suppose (6.3) and $T_n = o_p(\tau_n \tau_b)$ as $n \to \infty$.

(i) Then, $\hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 > 0$ as $n \to \infty$ if and only if, for each $\epsilon > 0$,

$$\lim_{m \to \infty} \sup_{n \geq m} P \left( \frac{1}{N_n} \sum_{i=1}^{N_n} T^2_{b,i} I(|T_{b,i}| > m) > \epsilon \right) = 0. \quad (6.7)$$

(ii) Additionally, (6.7) holds whenever $\{Y^2_b : b \geq 1\}$ is uniformly integrable, where $Y_b$ denotes a random variable with distribution $D_{n,b}$, $n \geq 1$, from (6.6) (i.e., $P(Y_b \leq x) = D_{n,b}(x)$, $x \in \mathbb{R}$). If (6.3) and $T_n = o_p(\tau_n \tau_b)$ hold, uniform integrability of $\{Y^2_b : b \geq 1\}$ is equivalent to $\int x^2 dD_{n,b}(x) = N_n^{-1} \sum_{i=1}^{N_n} \tau^2_b \to \sigma^2$ as $n \to \infty$.

(iii) (6.7) also holds whenever $\{X_t\}$ is stationary and $\{T^2_b : b \geq 1\}$ is uniformly integrable.

**Remark 6.1.5.** As $T_n$ is typically tight, the assumption $T_n = o_p(\tau_n \tau_b)$ is often satisfied by a standard condition on block length: $b \to \infty$ with $b/n + \tau_b/\tau_n \to 0$. Block conditions are not, in fact, used or required in statements of Theorems 18-20 above. However, block assumptions are usually needed to show the original subsampling estimator $S_{n,\text{SUB}}$ is convergent as in (6.3), and examples of Section 6.2 shall impose block length conditions for this purpose.
Theorem 20 connects convergence (6.3) of subsampling distributions $S_{n,\text{SUB}}$ to the convergence of subsampling variances $\hat{\sigma}^2_{n,\text{SUB}}$ in a way involving no further conditions on the process or statistic beyond mild types of uniform integrability. For example, with non-stationary processes $\{X_t\}$, Theorem 20(ii) converts the problem of probabilistic convergence $\hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2$ into a more approachable one of subsample-moment convergence $N_n^{-1} \sum_{i=1}^{N_n} ET_{b,i}^2 \rightarrow \sigma^2$. To frame another implication of Theorem 20, note that many inference problems with time series involve a stationary process $\{X_t\}$ and a statistic $T_n$ with a normal limit (6.2) such that $\{T^2_n : n \geq 1\}$, and consequently $\{T^2_b : b \geq 1\}$, is uniformly integrable; see Remark 2. In such problems, it suffices to simply establish the consistency of the subsampling estimator $S_{n,\text{SUB}}$ (6.3) and then the consistency of subsampling variance $\hat{\sigma}^2_{n,\text{SUB}}$ follows with no further effort (by Theorem 20(iii)) along with the consistency of the convolved subsampling estimator $C_{n,k_n}$ (by Theorem 18). Again, with sample means, $C_{n,k_n}$ is a block bootstrap distribution and $\hat{\sigma}^2_{n,\text{SUB}}$ is a block bootstrap variance estimator, so both will be consistent in this setting by showing that $S_{n,\text{SUB}}$ is consistent. This strategy has two advantages with the block bootstrap: showing the consistency of $S_{n,\text{SUB}}$ is often an easier prospect than considering either $C_{n,k_n}$ or $\hat{\sigma}^2_{n,\text{SUB}}$ directly, and the consistency of $S_{n,\text{SUB}}$ (and thereby the bootstrap) can typically be established under weak process assumptions.

To illustrate, Section 6.2 applies the basic results here for establishing the convolved subsampling estimator $C_{n,k_n}$, as well as the block bootstrap for sample means, under differing dependence structures.

### 6.2 Applications of convolved subsampling estimation

#### 6.2.1 Convolved subsampling for general statistics under mixing

For mixing stationary time series, Radulovic (1996b) proved consistency of block bootstrap estimation for $T_n = \tau_n(t_n(X_1, \ldots, X_n) - t(P))$ based on the sample mean $t_n(X_1, \ldots, X_n) = \bar{X}_n$ with $t(P) = E X_1$ and $\tau_n = \sqrt{n}$. The assumptions made were quite weak, requiring only

(a1) a stationary, $\alpha$-mixing process fulfilling (6.2) (i.e., $T_n \overset{d}{\rightarrow} N(0, \sigma^2)$) and block lengths $b^{-1} + b/n \rightarrow 0$ as $n \rightarrow \infty$;

(a2) uniformly integrable $\{T^2_n : n \geq 1\}$.

From results in Section 6.1.3 and the equivalence between the block bootstrap and the convolved subsampling estimator $C_{n,k_n}$ for the sample mean, a different perspective is possible for the bootstrap findings in Radulovic (1996b). Under only assumption (a1) above, the subsampling estimator $S_{n,\text{SUB}}$ is consistent (i.e., (6.3) holds) for the asymptotically normal distribution of $T_n = \sqrt{n}(\bar{X}_n - EX_1)$ (cf. Theorem 3.2.1, Politis et al. (1999)), implying, by Theorem 18 here, that the block bootstrap estimator $C_{n,k_n}$ would be consistent if the subsampling variance converges $\hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2$. But, if $S_{n,\text{SUB}}$ is consistent for a normal limit by (a1), assumption (a2)
then guarantees that \( \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 \) holds by Theorem 20. Furthermore, under (a2) and with \( k_n = \lfloor n/b \rfloor \to \infty \) resampled blocks as in Radulovic (1996b, 2012), convergence \( \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 \) becomes even necessary here by Theorem 19. Hence, \( \alpha \)-mixing serves to show that the original subsampling estimator \( S_{n,\text{SUB}} \) is consistent; after which, uniform integrability and stationary assure both \( \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 \) and consistency of the block bootstrap estimator \( C_{n,k_n} \) by Theorems 19-20.

Under analogously weak assumptions as those of Radulovic (1996b), Theorem 21 next provides the general consistency of convolved subsampling estimation for general statistics arising from mixing, and possibly non-stationary, time processes. When applied to a sample mean \( t_n(X_1, \ldots, X_n) = \bar{X}_n \), so that \( C_{n,k_n} \) is a block bootstrap estimator, this result extends those of Radulovic (1996b) in two ways: by allowing potential non-stationarity series and by permitting arbitrary levels \( k_n \) of convolution/block resampling (rather than the single choice \( k_n = \lfloor n/b \rfloor \)). When the statistic \( t_n(X_1, \ldots, X_n) \) is not a sample mean, \( C_{n,k_n} \) may again no longer correspond to the block bootstrap but has interest as a block resampling estimator representing a composite of subsampling and bootstrap; see also Section 6.3.

**Theorem 21.** Let \( \{X_t\} \) be a (possibly non-stationary) strongly mixing sequence. Suppose \( b^{-1} + b/n + \tau_b/n \to 0 \) as \( n \to \infty \); \( T_n = o_p(\tau_n/n) \); (6.7) holds; and that \( Y_b \xrightarrow{d} N(0, \sigma^2) \) as \( n \to \infty \), for some \( \sigma^2 > 0 \), where each random variable \( Y_b, b \equiv b_n \geq 1 \), has distribution function \( D_{n,b} \) from (6.6). Then, as \( n \to \infty \),

\[
\sup_{x \in \mathbb{R}} |S_{n,\text{SUB}}(x) - \Phi(x/\sigma)| \xrightarrow{p} 0 \quad \text{and} \quad \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2
\]

and, for any positive integer sequence \( k_n \),

\[
\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{p} 0.
\]

Furthermore, if (6.2) additionally holds (i.e., \( T_n \xrightarrow{d} N(0, \sigma^2) \)), then \( S_{n,\text{SUB}} \) and \( C_{n,k_n} \) (with any \( k_n \)) are consistent for the distribution \( F_n \) of \( T_n \):

\[
\sup_{x \in \mathbb{R}} |S_{n,\text{SUB}}(x) - F_n(x)| \xrightarrow{p} 0 \quad \text{and} \quad \sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - F_n(x)| \xrightarrow{p} 0.
\]

While providing a broad result on the validity of convolved subsampling estimation for mixing processes, Theorem 21 also expands the general subsampling results of Politis et al. (1999) (ch. 4.2), which focused on \( S_{n,\text{SUB}} \) for mixing series, to further include consistency of the subsampling variance \( \hat{\sigma}^2_{n,\text{SUB}} \). That is, when dropping (6.7), the remaining Theorem 21 assumptions match those of Theorem 3.2.1-4.2.1 of Politis et al. (1999) for the consistency of \( S_{n,\text{SUB}} \) to a normal limit; including (6.7) in Theorem 21 is then necessary for \( \hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 \) by Theorem 20 and assures convergence of \( C_{n,k_n} \) by Theorem 18.
If the process \( \{X_t\} \) is actually stationary, we immediately obtain the following result.

**Corollary 6.2.1.** Let \( \{X_t\} \) be a stationary, strongly mixing sequence. Suppose also \( b^{-1} + b/n + \tau_b/\tau_n \rightarrow 0 \) as \( n \rightarrow \infty \); that (6.2) holds; and that (6.7) holds (e.g., uniform integrability of \( \{T_n^2 : n \geq 1\} \) suffices). Then, as \( n \rightarrow \infty \), the convergence results of Theorem 21 hold.

Section 6.2.2 next provides some further refinements with mixing processes in the sample mean case, where \( C_{n,k} \) matches the block bootstrap.

### 6.2.2 Block bootstrap for mixing non-stationary time processes

Consider a strongly mixing, potentially non-stationary sequence \( \{X_t\} \) having a common mean parameter \( \mathbb{E}X_t = \mu \in \mathbb{R} \), which is estimated by the sample mean \( \bar{X}_n \). In this setting and under conditions where \( T_n \equiv \sqrt{n}(\bar{X}_n - \mu) \) has a normal limit (6.2), Fitzenberger (1998) established the consistency of the block bootstrap for estimating the distribution of \( T_n \). The result, however, required the existence of a \((4 + \delta)\)-moment (i.e., \( \sup_t \mathbb{E}|X_t|^{4+\delta} < \infty \) for some \( \delta > 0 \)) along with stringent mixing conditions and restrictions on the block length \( b = o(n^{1/2}) \). Politis et al. (1999) (example 4.4.1) showed that the subsampling estimator \( S_{n,\text{SUB}} \) is consistent under weaker conditions, including only a \((2 + \delta)\)-moment. For the block bootstrap with \( k_n = \lfloor n/b \rfloor \) resampled blocks, Politis et al. (1999) also proved bootstrap consistency by applying convolved subsampling in this problem, using a weaker block assumption \( b = o(n) \) than Fitzenberger (1998) but otherwise with same remaining strong assumptions about the process. However, Politis et al. (1999) (remark 4.4.4) conjectured that the block bootstrap might be established under non-stationary using the same weak moment/mixing conditions as the subsampling estimator \( S_{n,\text{SUB}} \), just as in the case of stationary mixing processes (cf. Radulovic (1996b)). We confirm this by the following Theorem 22.

**Theorem 22.** Let \( \{X_t\} \) be a sequence of (not necessarily stationary) strongly mixing random variables with common mean \( \mu \). For some \( \delta > 0 \), suppose that \( \sup_t \mathbb{E}|X_t|^{2+\delta} < \infty \) and \( \sum_{k=1}^{\infty} \alpha(k)^{\delta/(2+\delta)} < \infty \). Assume also that, for some \( \sigma^2 > 0 \),

\[
\lim_{n \to \infty} \sup_{i \geq 1} \left| \text{Var} \left( n^{-1/2} \sum_{t=i}^{i+n-1} X_t \right) - \sigma^2 \right| = 0.
\]

Then, as \( n \to \infty \), \( T_n = \sqrt{n}(\bar{X}_n - \mu) \overset{d}{\to} N(0, \sigma^2) \) (i.e., (6.2) holds). Additionally, if \( b^{-1} + b/n \to 0 \) as \( n \to \infty \), then

\[
\sup_{x \in \mathbb{R}} |S_{n,\text{SUB}}(x) - \Phi(x/\sigma)| \overset{p}{\to} 0 \quad \text{and} \quad \delta_{n,\text{SUB}}^2 \overset{p}{\to} \sigma^2
\]
and, for any positive integer sequence \( k_n \),

\[
\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0.
\]

Hence, with any number \( k_n \) of concatenated blocks, the block bootstrap estimator \( C_{n,k_n} \) is valid for the distribution of the sample mean under mild assumptions for mixing, and possibly non-stationary, processes. Note that the assumptions of Theorem 22 resemble those essentially needed for a central limit theorem (CLT) itself (cf. Theorem 16.3.5, Athreya and Lahiri (2006)). In particular, the assumptions also match those commonly used in the stationary case for establishing the block bootstrap; see Section 3.2 of Lahiri (2003). With the same moment condition as Politis et al. (1999) (Theorem 4.4.1), Theorem 22 additionally shows that the original subsampling estimator \( S_{n,\text{SUB}} \) is consistent under non-stationarity with even weaker mixing assumptions than considered previously.

\[
\sum_{k=1}^{\infty} \frac{(k+1)^2}{\alpha(k)\delta/(8+\delta)} < \infty.
\]

The central message of Theorem 22, however, is that the convolved subsampling approach allows the block bootstrap estimator \( C_{n,k_n} \) for the sample mean to be established under weak conditions similarly to \( S_{n,\text{SUB}} \).

Next consider the block bootstrap in another important example of non-stationarity, involving certain periodically correlated time series. Here the mean function \( \mu(t) \equiv \mathbb{E}X_t \) is not constant, as in Theorem 22, but rather an almost periodic function. A real-valued function \( f \) is almost periodic if, for every \( \epsilon > 0 \), there is an \( n(\epsilon) \in \mathbb{N} \) such that in every interval \( I_{n(\epsilon)} \) of length \( n(\epsilon) \) or greater, there is an integer \( p \in I_{n(\epsilon)} \) such that

\[
\sup_{t \in \mathbb{Z}} |f(t+p) - f(t)| < \epsilon.
\]

see Corduneanu (1989). For such functions the limit \( M(f) \equiv \lim_{n \to \infty} n^{-1} \sum_{i=s}^{s+n-1} f(i) \) exists and does not depend on \( s \). Moreover, if the set \( \Lambda = \{ \lambda \in [0,2\pi) : M(g_{\lambda}) \neq 0 \} \) is finite for \( g_{\lambda}(t) \equiv f(t)e^{-i\lambda t}, t \in \mathbb{Z} (i = \sqrt{-1}) \), then

\[
\left| \frac{1}{n} \sum_{i=s}^{s+n-1} (f(i) - M(f)) \right| \leq \frac{C}{n}
\]

holds for some \( C > 0 \) not depending on \( n \) or \( s \) by Cambanis et al. (1994). Hence, \( M(f) \) represents the mean value of an almost periodic function \( f \). A time series is called almost periodically correlated (APC) if its mean and autocovariance functions are almost periodic, that is, for every fixed \( \tau \in \mathbb{Z} \),

\[
\mu(t) = \mathbb{E}X_t \quad \text{and} \quad \rho_{\tau}(t) = \mathbb{E}X_tX_{t+\tau}
\]

are almost periodic as functions of \( t \); see Hurst (1951). For an APC series \( \{X_t\} \), a parameter
of interest is then \( t(P) = \lim_{n \to \infty} n^{-1} \sum_{i=s}^{s+n-1} \mu(i) \) as a summary of the process mean structure, which is estimated by \( \bar{X}_n \). Synowiecki (2007) showed that the block bootstrap consistently estimates the sampling distribution of \( T_n = n^{1/2}(\bar{X}_n - \mu) \) under appropriate conditions. By applying the convolved subsampling technique, we may extend the bootstrap results of Synowiecki (2007) (Corollary 3.2) by substantially weakening the assumptions made there about \((4 + \delta)-moments and \( \sum_{k=1}^{\infty} k\alpha(k)\delta/(4+\delta) < \infty \).

**Corollary 6.2.2.** Let \( \{X_t\} \) be an APC sequence of strongly mixing random variables such that \( \sup_t \mathbb{E}|X_t|^{2+\delta} < \infty \) and \( \sum_{k=1}^{\infty} \alpha(k)^{\delta/(\delta+2)} < \infty \) for some \( \delta > 0 \), and suppose the set \( \Lambda = \{ \lambda \in [0,2\pi) : M(g_{\lambda}) \neq 0 \} \) is finite for \( g_{\lambda}(t) \equiv \mu(t)e^{-it\lambda}, \ t \in \mathbb{Z}, \) with \( \mu(t) = \mathbb{E}X_t \). Then, all conclusions of Theorem 22 hold for \( T_n = n^{1/2}(\bar{X}_n - \mu) \) as \( n \to \infty \).

### 6.2.3 Block bootstrap for linear time processes

Based on a sample \( X_1, \ldots, X_n \), next consider inference about the mean \( \mathbb{E}X_t = \mu \in \mathbb{R} \) of a stationary time process \( \{X_t\} \) prescribed as

\[
X_t = \mu + \sum_{j \in \mathbb{Z}} a_j \varepsilon_{t-j}, \quad t \in \mathbb{Z},
\]

(6.2)

in terms of iid variables \( \{\varepsilon_t\} \) with mean zero and finite variance \( \mathbb{E}\varepsilon_t^2 \in (0, \infty) \) and a real-valued sequence \( \{a_j\} \) of constants where \( \sum_{j \in \mathbb{Z}} a_j^2 < \infty \). The linear series \( \{X_t\} \) need not be mixing and, depending on constants \( \{a_j\} \), can potentially exhibit either weak or strong forms of time dependence. See for instance Section 2.1.3 on long memory, linear processes. Considering the sample mean \( \bar{X}_n \) as an estimator of the process mean \( \mu \), suppose that

\[
\lim_{n \to \infty} n^\alpha \text{Var}(\bar{X}_n) = \sigma^2
\]

(6.3)

for some \( \sigma^2 > 0 \) and exponent \( \alpha \in (0,1] \) depending on the process \( \{X_t\} \). When \( \alpha = 1 \), the sample mean’s variance decays at a rate \( O(n^{-1}) \) with the sample size, as typical for weakly, or short-range, dependent processes. However, when \( \alpha \in (0,1) \), the sample mean has a variance with comparatively slower decay \( O(n^{-\alpha}) \), which may be associated with processes exhibiting strong, or long-range, forms of dependence (cf. Lemma 2.1.4). If one considers Gaussian, long memory, linear processes (see Example 2.1.5), then \( \alpha \) corresponds to the Hurst-coefficient \( H \).

Based on (6.3), define \( T_n \equiv n^{\alpha/2}(\bar{X}_n - \mu) \) in terms of scaling \( \tau_n \equiv n^{\alpha/2} \). In this setting, the convolved subsampling \( C_{n,k_n} \) once again corresponds to the block bootstrap estimator based on \( k_n \) resampled blocks, but there is a wrinkle to note. Recalling from (6.5) that the bootstrap sample mean \( \bar{X}_{n_1}^{*} \) is created from a bootstrap sample of length \( n_1 = k_n b \), the bootstrap analog of \( T_n \) here is given by

\[
T_n^* = b^{(1-\alpha)/2}(n_1)^{\alpha/2}(X_{n_1}^{*} - \mathbb{E}_v \bar{X}_{n_1}^{*})
\]

(6.4)
rather than $T_n^* = (n_1)^{\alpha/2}(\bar{X}_{n_1}^* - E_\pi \bar{X}_{n_1}^*)$. While intuitive, the latter definition is incorrect and produces a degenerate bootstrap limit, shown by Lahiri (1993). Instead, scalar adjustment $b(1-\alpha)/2$ is required in the bootstrap version (6.4) of $T_n$. Note that the same modified normalizing sequence was used for the bootstrapped empirical process in Chapter 4. The adjustment disappears under weak dependence $\alpha = 1$ whereby bootstrap versions in (6.5) and (6.4) then match.

Interestingly, the convolved subsampling estimator $C_{n,k_n}$ automatically corresponds to the correct bootstrap rendition $T_n^*$ in (6.4) under both weak $\alpha = 1$ and strong $\alpha \in (0, 1)$ dependence. Considering the sample mean from stationary linear processes (6.2) ranging over short- or long-range dependence, Kim and Nordman (2011) showed the consistency of the block bootstrap distribution $C_{n,k_n}$ (when $k_n = \lfloor n/b \rfloor$) and bootstrap variance $\hat{\sigma}_n^2$. Via convolved subsampling, we may generalize their results. For linear processes $\{X_t\}$ satisfying (6.2)-(6.3), the sample mean $T_n \equiv n^{\alpha/2}(\bar{X}_n - \mu)$ has a normal limit (6.2) (cf. Davydov (1970)), and Nordman and Lahiri (2005) showed the consistency of the subsampling estimator $S_{n,\text{SUB}}$ (i.e., (6.3) holds) under mild assumptions. Hence, by primitively assuming (6.2)-(6.3) to hold, Corollary 6.2.3 next extends the block bootstrap to general stationary processes with sample means satisfying a variance condition (6.3), which includes results of Kim and Nordman (2011) for linear processes as a special case.

**Corollary 6.2.3.** Let $\{X_t\}$ be a stationary process with mean $\mu \in \mathbb{R}$ satisfying (6.3) for some $\alpha \in (0, 1]$, and suppose that (6.2)-(6.3) hold for $T_n \equiv n^{\alpha/2}(\bar{X}_n - \mu)$. Then, as $n \to \infty$,

$$\hat{\sigma}_n^2 \overset{p}{\to} \sigma^2 \quad \text{and} \quad \sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \overset{p}{\to} 0$$

for any positive integer sequence $k_n$.

Corollary 6.2.3 is an application of Theorems 18 and 20 for stationary processes which may not be mixing. Our exposition has assumed the exponent $\alpha \in (0, 1]$ to be known. Upon replacing $\alpha$ with an estimator $\hat{\alpha} \equiv \hat{\alpha}(X_1, \ldots, X_n)$ where $|\hat{\alpha} - \alpha| \log n \overset{P}{\to} 0$, the conclusions of Corollary 6.2.3 still hold; see Remark 3 of Kim and Nordman (2011) for further details.

**6.2.4 Block bootstrap under long-range dependence**

This section briefly mentions the block bootstrap with additional types of long-memory sequences. Beyond linear processes, the sample mean of a long-range dependent sequence may converge to a non-normal limit, such as the case for certain subordinated Gaussian processes considered by Taqqu (1975, 1979) and Dobrushin and Major (1979) (cf. Section 2.1.2). For such time series, Lahiri (1993) proved that the block bootstrap sample mean always has a normal limit, so that the block bootstrap fails if the original sample mean is asymptotically non-normal (and the same holds true for the bootstrapped empirical process, see Theorem 13). This result is in concordance with our Theorem 19(ii).
Zhang et al. (2013) considered subsampling estimation for a wider class of long-range dependence series that includes both subordinated Gaussian processes as well types of linear processes (6.2). Namely, sequences $X_t = K(Z_t)$, $t \in \mathbb{Z}$, formed by a measurable transformation $K$ of a long-range dependent linear process

$$Z_t = \varepsilon_t + \sum_{j=1}^{\infty} j^{-\beta} L(j) \varepsilon_{t-j}, \quad t \in \mathbb{Z},$$

defined with iid mean zero, finite variance innovations $\{\varepsilon_t\}$, an index parameter $1/2 < \beta < 1$, and slowly varying function $L(\cdot)$. They distinguish two cases, depending on $\beta$ and the power $p \geq 1$ of $K(\cdot)$. In the first case (i.e., $p(2\beta - 1) > 1$), the transformation $K(\cdot)$ diminishes long-range dependence, and the sample mean converges to a normal limit. In the second case (i.e., $p(2\beta - 1) < 1$), the transformed process $X_t = K(Z_t)$ remains strongly dependent and the sample mean has a normal limit only when $p = 1$.

Assuming a constant function $L(\cdot) = C$ in the above formulation, the variance of a sample mean satisfies (6.3) (i.e., $\lim_{n \to \infty} n^a \text{Var}(\bar{X}_n) = \sigma^2 > 0$) with a long-memory exponent $\alpha \equiv \min\{1, p(2\beta - 1)\} \in (0, 1]$ that changes between cases of weak $\alpha = 1$ or strong $\alpha = p(2\beta - 1) \in (0, 1)$ dependence (cf. Lemma 1, Zhang et al. (2013)). For the sample mean, Zhang et al. (2013) established consistency of several subsampling estimators as well as convergence of $\hat{\sigma}^2_{n,\text{SUB}}$.

Thus, by slightly re-casting results of Zhang et al. (2013) and applying our Corollary 6.2.3, we may show the validity of the block bootstrap $C_{n,k}$ for estimating the distribution of $T_n \equiv n^{\alpha/2}(\bar{X}_n - \mu)$, $\mu = E X_t$, for transformed linear processes exhibiting either short- or long-range dependence. To the best of our knowledge, the bootstrap has not yet been investigated for such processes.

**Corollary 6.2.4.** For $X_t = K(Z_t)$, $t \in \mathbb{Z}$, as above, suppose (6.3) holds for $\alpha = \min\{1, p(2\beta - 1)\} \in (0, 1]$ along with conditions of Theorem 1 in Zhang et al. (2013) (involving a block $b \propto n^a$ for some $a \in (0, 1)$) with either $p(2\beta - 1) > 1$, or $p = 1$ and $(2\beta - 1) < 1$. Then, for $T_n = n^{\alpha/2}(\bar{X}_n - \mu)$ as $n \to \infty$, both (6.2)-(6.3) hold and

$$\hat{\sigma}^2_{n,\text{SUB}} \xrightarrow{p} \sigma^2 \quad \text{and} \quad \sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{p} 0$$

for any positive integer sequence $k_n$.

As with Gaussian subordinated processes (see Lahiri (1993)), consistency of the block bootstrap or convolved estimator $C_{n,k_n}$ holds only in cases where the sample mean follows a CLT. For statistics other than the sample mean, Betken and Wendler (2017+) proved the consistency of the subsampling estimator $S_{n,\text{SUB}}$ with transformations of Gaussian processes, provided that the original statistic converges under mild assumptions. When this limit is normal, consistency of the convolved estimator will follow by our Theorem 19 by showing convergence of
\[ \hat{\sigma}^2_{n, SUB} \] (which, as Betken and Wendler (2017+) consider only stationary processes, can hold by Theorem 20 and uniform integrability of the statistic).

6.3 Convolved subsampling in other contexts

6.3.1 U-statistics

U-statistics are a class of nonlinear functionals for prescribing statistics, such as the sample variance. Suppose that \( X_1, \ldots, X_n \) arise from a stationary process and, based on a symmetric kernel \( h: \mathbb{R}^2 \to \mathbb{R} \), define a (bivariate) U-statistic as

\[ t_n \equiv t_n(X_1, \ldots, X_n) = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} h(X_i, X_j), \]

which estimates a target parameter \( t(P) \equiv \int h(x, y) dG(x) dG(y) \), where \( G \) denotes the marginal distribution of \( X_t \). Consider the problem of estimating the distribution of \( T_n \equiv \sqrt{n}(t_n - t(P)) \), with scaling \( \tau_n = \sqrt{n} \), under weak time dependence. The subsampling distribution \( S_{n, SUB} \) is defined by computing the U-statistic \( t_{n, \bar{b}, i} = [b(b-1)]^{-1} \sum_{i \leq j < j \leq i+\bar{b}-1} h(X_{i,j}, X_{j,j}) \) on each length \( b \) subsample \( \{(X_i, \ldots, X_i+b-1)\}_{i=1}^{N_{n,b}} \) in (6.1). In contrast, block bootstrap versions of U-statistics have a formulation similar to (6.5); see Dehling and Wendler (2010), Sharipov and Wendler (2012) and Leucht (2012). That is, a bootstrap sample \( X^*_1, \ldots, X^*_{n_1}, n_1 = k_n b \) is generated by resampling \( k_n \) blocks of length \( b \) (typically \( k_n = \lfloor n/b \rfloor \)) and then the U-statistic \( t^*_{n_1} \equiv t_{n_1}(X^*_1, \ldots, X^*_{n_1}) \) is calculated from the complete bootstrap sample to create a bootstrap rendition \( T^*_n = \sqrt{n}(t^*_{n_1} - E_t t^*_{n_1}) \) of \( T_n \). In this setting, the bootstrap distribution \( T^*_n \) would not generally correspond that of a \( t_{n,b} \)-fold convolution \( C_{n,k_n} \) of the subsampling distribution \( S_{n, SUB} \), as occurred in the sample mean case (Section 6.1.2).

However, Sharipov et al. (2016b) recently considered an alternative block resampling estimator for U-statistics, which matches the convolved subsampling estimator \( C_{n,k_n} \) here based on the subsampling estimator \( S_{n, SUB} \) for \( T_n \) described above. Note that, for stationary mixing data, Dehling and Wendler (2010) (Theorem 1.8-Lemma 3.6) provide a CLT for the relevant U-statistic: \( T_n \xrightarrow{d} N(0, \sigma^2) \) and \( E T_n^2 \to \sigma^2 \) as \( n \to \infty \) where \( \sigma^2 \equiv 4 \sum_{k=-\infty}^{\infty} \text{Cov}(h_1(X_0), h_1(X_k)) \) for \( h_1(x) = \int h(x, y) dG(y) \). Under mixing conditions and with \( k_n = \lfloor n/b \rfloor \to \infty \), Sharipov et al. (2016b) established that \( C_{n,k_n} \) captures this limiting normal distribution of \( T_n \) and also showed the consistency of the variance \( \hat{\sigma}^2_{n, SUB} \) of \( C_{n,k_n} \). The argument there involved decomposing the bootstrap U-statistic \( T^*_n \) into a linear part, coinciding with a sample mean from the usual block bootstrap, and degenerate part shown to be negligible. However, the general convolution result in Theorem 21 for mixing processes provides an alternative, and much simpler, approach. From \( T_n \xrightarrow{d} N(0, \sigma^2) \) and \( E T_n^2 \to \sigma^2 \), all of the conditions of Theorem 21 automatically hold, proving that \( C_{n,k_n} \) is consistent for the distribution of \( T_n \) for any convolution level \( k_n \) and also that \( \hat{\sigma}^2_{n, SUB} \xrightarrow{d} \sigma^2 \). This approach also weakens the block assumptions used by
Sharipov et al. (2016b) \( (i.e., b = O(n^\epsilon) \) for some \( \epsilon \in (0, 1) \) to \( b^{-1} + b/n \to 0 \) under Theorem 21. As a side note, simulations in Sharipov et al. (2016b) also indicate that convolved subsampling \( C_{n,k_n} \) exhibits better coverage accuracy than traditional subsampling \( S_{n,\text{SUB}} \) for U-statistics, and has performance comparable to the standard block bootstrap. This finding suggests consideration of convolved subsampling estimators even when these do not match the block bootstrap.

**Remark 6.3.1.** Convolved subsampling can reduce skewness compared to distributional estimates from subsampling; see Politis et al. (1999) (sec. 10.2). This aspect may partially explain the better U-statistic coverage mentioned above. For example, in approximating sample means, the distribution of \( T_n = \sqrt{n}(\bar{X}_n - \mu) \) often has approximate skewness \( \gamma/\sqrt{n} \) for some constant \( \gamma \). In this case, the corresponding subsampling estimator \( S_{n,\text{SUB}} \) has a larger approximate skewness \( \gamma/\sqrt{b} \), while a more fully convolved estimator (bootstrap) \( C_{n,k_n} \) with \( k_n \approx n/b \) has skewness approximately \( \gamma/\sqrt{n} \), where a better matching skewness can impact higher-order accuracy.

### 6.3.2 Spectral estimators for non-stationary time series

As described in Section 6.2.2, almost periodically correlated (APC) time series \( \{X_t\} \) are an important example of non-stationary sequences. Beyond the mean function, inference about the correlation structure is also of interest. Based on a sample \( X_1, \ldots, X_n \), a symmetric kernel \( w(\cdot) \) and a bandwidth choice \( L_n \), Lenart (2011, 2016) considered kernel estimators

\[
t_n(X_1, \ldots, X_n) = \frac{1}{2\pi n} \sum_{t=1}^{n} \sum_{s=1}^{n} \frac{1}{L_n} w \left( \frac{t-s}{L_n} \right) X_t X_s e^{-i\upsilon t} e^{i\upsilon s}
\]

for an extended spectral density \( t(P) \equiv t(P)(\upsilon, \omega), (\upsilon, \omega) \in (0, 2\pi]^2 \), used to represent the almost periodic covariance function \( c_\tau(t) \equiv \text{Cov}(X_t, X_{t+\tau}), t \in \mathbb{Z} \), for a given \( \tau \in \mathbb{Z} \); see Lenart (2011, 2016) for details.

For \( T_n \equiv \tau_n(t_n - t(P)) \) with scaling \( \tau_n = \sqrt{n/L_n} \), Lenart (2011) (Theorems 3.1-3.2) proved a CLT \( T_n \overset{d}{\to} N(0, \sigma^2) \) and moment convergence \( E T_n^2 \to \sigma^2 \) with mixing APC series, which was extended in Lenart (2016) to multivariate data. Due to the complicated form of \( \sigma^2 \), a subsampling estimator \( S_{n,\text{SUB}} \) for the distribution of \( T_n \) may be computed as in (6.1) with analog statistics \( t_{n,b,i} \) and scaling \( \tau_b = \sqrt{b/L_b} \) defined from subsamples \( \{(X_i, \ldots, X_{i+b-1})\}_{i=1}^{N_n} \). Lenart (2011) proved the consistency of the estimator \( S_{n,\text{SUB}} \), while Lenart (2016) proposed a generalized resampling method which essentially corresponds a convolved subsampling estimator \( C_{n,k_n} \) induced from \( S_{n,\text{SUB}} \) (though Lenart (2016) also considered \( S_{n,\text{SUB}} \) defined with possibly non-uniform weights on \( \{t_{n,b,i}\}_{i=1}^{N_n} \)). In particular, Lenart (2016) established the consistency of \( C_{n,k_n} \) through bootstrap arguments requiring much stronger mixing and moment assumptions than needed for the convergence \( T_n \overset{d}{\to} N(0, \sigma^2) \) and \( E T_n^2 \to \sigma^2 \). However, the gen-
eral convolved subsampling result in Theorem 21 may alternatively be used here with mixing non-stationary ACP series.

To apply Theorem 21 with blocks where $b^{-1} + b/n + \tau_b/\tau_n \to 0$ as $n \to \infty$, one requires that $Y_b \xrightarrow{d} N(0, \sigma^2)$ and that (6.7) holds, where $Y_b, b \equiv b_n \geq 1$, denotes a sequence of variables with distribution $D_n,b(\cdot)$ from (6.6). But, the same conditions needed for $T_n \xrightarrow{d} N(0, \sigma^2)$ and $EY^2_\sigma \to \sigma^2$ also yield $Y_b \xrightarrow{d} N(0, \sigma^2)$ and $EY^2_\sigma \to \sigma^2$ (cf. Theorems 3.1-3.2 and 4.1, Lenart (2011)). Furthermore, mixing and $Y_b \xrightarrow{d} N(0, \sigma^2)$, along with $T_n = O_p(1)$ and $\tau_n/\tau_b \to \infty$, guarantee that (6.3) holds (i.e., $\sup_{x \in \mathbb{R}} |S_{n,SUB}(x) - \Phi(x/\sigma)| \xrightarrow{p} 0$) and that consequently (6.7) follows from Theorem 20(ii) by $EY^2_\sigma \to \sigma^2$. That is, the same minimal conditions for a CLT with APC series suffice for the consistency of convolved subsampling $C_{n,k_n}$ by the general result of Theorem 21.

### 6.4 Proofs

**Lemma 6.4.1.** Let $\{X_n\}$ be a sequence of real-valued random variables such that $X_n \xrightarrow{d} X$ and $\text{Var}(X_n) \to \text{Var}(X) < \infty$ as $n \to \infty$. Then, $EX_n \to EX \in \mathbb{R}$ and $EX^2_n \to EX^2 < \infty$.

*Proof.* Write $m_n = EX_n$ and $Y_n = X_n - m_n$. As $EY^2_n$ is bounded, $\{Y_n\}$ is uniformly integrable and hence tight. Because $\{X_n\}$ is also tight by $X_n \xrightarrow{d} X$, it holds that $m_n = X_n - Y_n$ is tight and so must be a bounded sequence. Since $\{Y_n\}$ and $\{m_n\}$ are uniformly integrable, the sum $\{X_n = Y_n + m_n\}$ is also uniformly integrable and $EX_n \to EX$ follows. $\Box$

*Proof of Theorem 18.* We show distributions converge in probability through Mallow’s metric $d_2(\cdot, \cdot)$; see Remark 5 and Bickel and Freedman (1981) (sec. 8). For random variables $X,Y$ with $X \sim F$ and $Y \sim G$, we also denote $d_2(X,Y) \equiv d_2(F,G)$. For $n \geq 1$, recall the random variable $Z^*_n \equiv k_n^{-1/2} \sum_{i=1}^{k_n}(Y^*_{n,i} - m_{n,SUB})$ from (6.4) has the convolved distribution $C_n,k_n$ based on iid variables $\{Y^*_{n,i}\}_{i=1}^{k_n}$ with distribution $S_{n,SUB}$ and mean $m_{n,SUB} \equiv \int x dS_{n,SUB}(x)$. Let $Z_1, \ldots, Z_{k_n}$ be iid $N(0, \sigma^2)$ variables. Then,

$$ [d_2(C_n,k_n, \Phi(\cdot/\sigma))]^2 = \left[ d_2 \left( Z^*_n, \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} Z_i \right) \right]^2 \leq \sum_{i=1}^{k_n} \left[ d_2 \left( Y^*_{n,i} - m_{n,SUB}, Z_i \right) \right]^2 = \left[ d_2 \left( Y^*_{n,1}, Z_1 \right) \right]^2 - [m_{n,SUB}]^2 $$

holds by Lemmas 8.7-8.8 of Bickel and Freedman (1981) (for the inequality and the last equality, respectively). By (6.2)-(6.3) and $\sigma^2_{n,SUB} \xrightarrow{p} \sigma^2$, for any arbitrary subsequence $\{n_j\} \subset \{n\}$, one may extract a further subsequence $\{\ell \equiv n_k\} \subset n_j$ such that $Y^*_{\ell,1} \xrightarrow{d} Z_1$ and $\sigma^2_{\ell,SUB} \to \sigma^2$.
as $\ell \to \infty$ (a.s.$(P)$). By Lemma 1, this implies $m_{\ell, \text{SUB}} \to 0$ and $\int x^2 dS_{\ell, \text{SUB}}(x) \to \sigma^2$ as $\ell \to \infty$ (a.s.$(P)$). Together, $Y_{\ell,1}^* \overset{d}{\to} Z_1$ and $\int x^2 dS_{\ell, \text{SUB}}(x) \to \sigma^2$ (where $Y_{\ell,1}^* \sim S_{\ell, \text{SUB}}$) are equivalent to $[d_2(Y_{\ell,1}^*, Z_1)]^2 \to 0$ as $\ell \to \infty$ (cf. Lemma 8.3, Bickel and Freedman (1981)). Thus, $|d_2(C_{\ell,k_1}, \Phi(\cdot \sigma))|^2 \to 0$, implying $\sup_{x \in \mathbb{R}} |C_{\ell,k_1}(x) - \Phi(x/\sigma)| \to 0$ as $\ell \to \infty$ (a.s.$(P)$).

As the subsequence $\{n_j\}$ was arbitrary, Theorem 18 follows.

Proof of Proposition 6.1.1. Recall that $C_{n,k_n}(x) \equiv P_{\ast}(Z_n^* \leq x)$, $x \in \mathbb{R}$, is defined by the resampling distribution of $Z_n^*$ from (6.4), and define $\tilde{C}_{n,k_n}(x) \equiv P_{\ast}(Z_n^* + m_{n, \text{SUB}} \sqrt{k_n} \leq x)$, $x \in \mathbb{R}$, where $Z_n^* + m_{n, \text{SUB}} \sqrt{k_n} = \sum_{i=1}^{k_n} Y_{n,i}/\sqrt{k_n}$ is the distribution of $S_{n, \text{SUB}}$. To show Proposition 6.1.1, note that distributions $C_{n,k_n}$ and $\tilde{C}_{n,k_n}$ can only match asymptotically if and only if $m_{n, \text{SUB}} \sqrt{k_n} \xrightarrow{x} 0$, which is equivalent to $m_{n, \text{SUB}} \equiv \int xdS_{n, \text{SUB}}(x) \xrightarrow{P} 0$ as $\{k_n\}$ is a bounded positive integer sequence. Consequently, Proposition 6.1.1 follows by showing that, for bounded $\{k_n\}$, (6.3) holds if and only if $\sup_{x \in \mathbb{R}} |\tilde{C}_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0$.

Let $\phi_n(t)$ and $\tilde{\phi}_{n,k_n}(t) = [\phi_n(t/\sqrt{k_n})]^{k_n}$, $t \in \mathbb{R}$, denote the characteristic functions of $S_{n, \text{SUB}}$ and $\tilde{C}_{n,k_n}$. Suppose (6.3) holds so that, for any subsequence $\{n_j\} \subset \{n\}$, extract a further subsequence $\{\ell \equiv n_k\} \subset \{n_j\}$ such that $\sup_{x \in \mathbb{R}} |S_{\ell, \text{SUB}}(x) - \Phi(x/\sigma)| \to 0$ as $\ell \to \infty$ (a.s.$(P)$).

Then, for any given $T > 0$,

$$\Delta_\ell(T) \equiv \max_{|t| \leq T} |\phi_{\ell}(t) - e^{-t^2 \sigma^2/2}| \to 0$$

as $\ell \to \infty$ (a.s.$(P)$) by the Levy continuity theorem (cf. Athreya and Lahiri (2006), Theorem 10.3.1). Fix $t_0 \in \mathbb{R}$ and set $T_0 = |t_0|$. Then, using that $|\prod_{i=1}^{n} w_i - \prod_{i=1}^{n} z_i| \leq \sum_{i=1}^{n} |w_i - z_i|$ for complex numbers $\{w_i, z_i\}_1^n$ with $|w_i|, |z_i| \leq 1$, we have

$$|\tilde{\phi}_{\ell,k_1}(t_0) - e^{-t_0^2 \sigma^2/2}| \leq k_1|\phi_{\ell}(t_0/\sqrt{k_n}) - e^{-t_0^2 \sigma^2/(2k_1)}| \leq k_1 \Delta_\ell(T_0) \to 0$$

as $\ell \to \infty$ (a.s.$(P)$) by $\sup_{n} k_n < \infty$. Hence, for all $t_0 \in \mathbb{R}$, $\tilde{\phi}_{\ell,k_1}(t_0) \to e^{-t_0^2 \sigma^2/2}$ as $\ell \to \infty$, implying $\sup_{x \in \mathbb{R}} |\tilde{C}_{\ell,k_1}(x) - \Phi(x/\sigma)| \to 0$ (a.s.$(P)$). As the subsequence $\{n_j\}$ was arbitrary, we have $\sup_{x \in \mathbb{R}} |\tilde{C}_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0$ by the equivalence of convergence in probability to almost sure convergence among subsequences.

Next suppose $\sup_{x \in \mathbb{R}} |\tilde{C}_{n,k_n}(x) - \Phi(x/\sigma)| \xrightarrow{P} 0$ so that, for any subsequence $\{n_j\} \subset \{n\}$, extract a further subsequence $\{\ell \equiv n_k\} \subset \{n_j\}$ such that $\sup_{x \in \mathbb{R}} |\tilde{C}_{\ell,k_1}(x) - \Phi(x/\sigma)| \to 0$ as $\ell \to \infty$ (a.s.$(P)$). Fix $t_0$ and define $T = |t_0| \sup_{n} k_n$. By the Levy continuity theorem,

$$\sup_{|t| \leq T} |\tilde{\phi}_{\ell,k_1}(t) - e^{-t^2 \sigma^2/2}| = \sup_{|t| \leq T} |\phi_{\ell}(t/\sqrt{k_1})|^{k_1} - [e^{-t^2 \sigma^2/(2k_1)}]^{k_1} \to 0$$

as $\ell \to \infty$, implying $[\phi_{\ell}(t_0)/e^{-t_0^2 \sigma^2/2}]^{k_1} \to 1$ (a.s.$(P)$) from bounded $\{k_1\}$. As $1 \leq k_1 \leq \sup_{n} k_n < \infty$, we then have $\phi_{\ell}(t_0) \to e^{-t_0^2 \sigma^2/2}$ as $\ell \to \infty$ for any $t_0 \in \mathbb{R}$, so that $\sup_{x \in \mathbb{R}} |S_{\ell, \text{SUB}}(x) - \Phi(x/\sigma)| \to 0$ (a.s.$(P)$). As the subsequence $\{n_j\}$ was arbitrary, $\sup_{x \in \mathbb{R}} |S_{n, \text{SUB}}(x) - \Phi(x/\sigma)| \xrightarrow{P}$
Proof of Theorem 19. For \( n \geq 1 \), again let \( \{Y_{n,i}\}_{i=1}^{k_n} \) be iid with distribution \( S_{n,\text{SUB}} \) and mean \( m_{n,\text{SUB}} \equiv \int xdS_{n,\text{SUB}}(x) \), so that \( Z_n^* \equiv k_n^{-1/2} \sum_{i=1}^{k_n}(Y_{n,i}^* - m_{n,\text{SUB}}) \) from (6.4) follows the convolved distribution \( C_{n,k_n} \). For \( \epsilon > 0 \), define quantities \( \Delta_1n(\epsilon) \equiv k_nP_*(Y_{n,1}^* \geq \sqrt{k_n}\epsilon) = k_n\int_{|x| \geq \sqrt{k_n}\epsilon} 1dS_{n,\text{SUB}}(x) \),

\[
\Delta_2n(\epsilon) \equiv \sqrt{k_n} \int_{|x| \geq \sqrt{k_n}\epsilon} |x|dS_{n,\text{SUB}}(x),
\]

\[
\Delta_3n(\epsilon) \equiv \int_{|x| \geq \sqrt{k_n}\epsilon} x^2dS_{n,\text{SUB}}(x),
\]

taking into account

\[
\Delta_1n(\epsilon) \leq \epsilon \Delta_2n(\epsilon) \leq \epsilon^2 \Delta_3n(\epsilon)
\]

and, by assumption, \( \Delta_3n(\epsilon) \overset{\mathbb{P}}{\to} 0 \) for any \( \epsilon > 0 \). For any \( \{n_j\} \subset \{n\} \), extract a further subsequence \( \{\ell \equiv n_k\} \subset \{n_j\} \) such that \( \Delta_3\ell(1/m) \to 0 \) for any integer \( m \geq 1 \) as \( \ell \to \infty \) (as \( P \)), implying that \( \lim_{\ell \to \infty} \Delta_3\ell(\epsilon) = 0 \) for any \( \epsilon > 0 \) and \( j = 1, 2, 3 \) (as \( P \)). In particular, as \( |\tilde{\Delta}_2\ell(1)| \leq \Delta_3\ell(1) \to 0 \) as \( \ell \to \infty \) for \( \tilde{\Delta}_3\ell(1) \equiv \sqrt{k_\ell} \int_{|x| \geq \sqrt{k_\ell}\epsilon} x^2dS_{\ell,\text{SUB}}(x) \), note that \( Z_\ell^* \) can have a normal \( N(0, \sigma^2) \) limit law if and only if \( Z_\ell^* - \tilde{\Delta}_2\ell(1) \to \infty \) (as \( P \)). Also, from \( \lim_{\ell \to \infty} \Delta_1\ell(\epsilon) = 0 \) for any \( \epsilon > 0 \), the array \( \{Y_{\ell,i}^*/\sqrt{k_\ell}\}_{i=1}^{k_\ell} \) is infinitesimal. Hence, by classical CLT results with independent, infinitesimal random variables (cf. Chow and Teicher (1988), Theorem 3(ii), ch. 12.2),

\[
Z_\ell^* - \tilde{\Delta}_2\ell(1) = \frac{1}{\sqrt{k_\ell}} \sum_{i=1}^{k_\ell} \left[ Y_{\ell,i}^* - \int_{|x| < \sqrt{k_\ell}\epsilon} x^2dS_{\ell,\text{SUB}}(x) \right]
\]

will have a normal \( N(0, \sigma^2) \) limit if and only if

\[
\Gamma_\ell(\epsilon) \equiv \int_{|x| < \sqrt{k_\ell}\epsilon} x^2dS_{\ell,\text{SUB}}(x) - \left( \int_{|x| < \sqrt{k_\ell}\epsilon} x^2dS_{\ell,\text{SUB}}(x) \right)^2 \to \sigma^2
\]

holds for any \( \epsilon > 0 \) as \( \ell \to \infty \) (as \( P \)). But, for any \( \epsilon > 0 \),

\[
|\Gamma_\ell(\epsilon) - \sigma_{\ell,\text{SUB}}^2|
\]

\[
= \left| \int_{|x| \geq \sqrt{k_\ell}\epsilon} x^2dS_{\ell,\text{SUB}}(x) + \left( \int x^2dS_{\ell,\text{SUB}}(x) \right)^2 - \left( \int_{|x| < \sqrt{k_\ell}\epsilon} x^2dS_{\ell,\text{SUB}}(x) \right)^2 \right|
\]

\[
\leq \Delta_3\ell(\epsilon) + 2\Delta_2\ell(\epsilon) \int |x|dS_{\ell,\text{SUB}}(x) \to 0
\]

by the almost surely boundedness of \( \int |x|dS_{\ell,\text{SUB}}(x) \) (as a consequence of \( \Delta_2(\epsilon) \to 0 \)). Thus, \( \Gamma_\ell(\epsilon) \to \sigma^2 \) for any \( \epsilon > 0 \) if and only if \( \sigma_{\ell,\text{SUB}}^2 \to \sigma^2 \) as \( \ell \to \infty \) (as \( P \)). As the subsequence
\{n_j\} was arbitrary, the result now follows. \hfill \Box

Proof of Corollary 6.1.2. Suppose first \(\hat{\sigma}_{n,\text{SUB}}^2 \overset{p}{\to} \sigma^2 > 0\) and \(k_n \to \infty\). Fix \(\epsilon > 0\). Then, the normal convergence for \(C_{n,k_n}\) in Corollary 6.1.2 will follow from Theorem 19 by showing \(Y_n(\epsilon) \equiv \int_{|x| \geq \sqrt{k_n}} x^2 dS_{n,\text{SUB}}(x) \overset{p}{\to} 0\) under any one of Conditions (C.1)-(C.4). If Condition (C.1) holds, then for any subsequence \(\{n_j\} \subset \{n\}\), extract a further subsequence \(\{\ell \equiv n_k\} \subset \{n_j\}\) such that \(\hat{\sigma}_{\ell,\text{SUB}}^2 \to \sigma^2\) and \(Y_\ell \overset{d}{\to} Y_0\) as \(\ell \to \infty\) (a.s.\((P)\)), where \(Y_\ell\) denotes a random variable with distribution \(S_{\ell,\text{SUB}}\) and \(Y_0\) is a random variable with distribution \(J_0\). As \(\hat{\sigma}_{\ell,\text{SUB}}^2\) and \(\sigma^2\) are the variances of \(Y_\ell\) and \(Y_0\), Lemma 6.4.1 yields \(\int x^2 dS_{\ell,\text{SUB}}(x) \to EY_0^2 < \infty\) from which \(Y_\ell(\epsilon) \to 0\) holds by \(k_\ell \to \infty\) as \(\ell \to \infty\) (a.s.\((P)\)) (i.e., \(\{Y_\ell^2 : \ell \geq 1\}\) is uniformly integrable from \(\int x^2 dS_{\ell,\text{SUB}}(x) \to EY_0^2 < \infty\) and \(Y_\ell \overset{d}{\to} Y_0\)). As the subsequence \(\{n_j\}\) was arbitrary, \(Y_n(\epsilon) \overset{p}{\to} 0\) follows. If Condition (C.2) holds, then, for any \(C > 0\), Markov’s inequality gives \(P(Y_n(\epsilon) > C) \leq P(\int x^2 + \epsilon_0 dS_{n,\text{SUB}}(x) > (\epsilon\sqrt{k_n}/C) \to 0\) as \(n \to \infty\) because \(\int x^2 + \epsilon_0 dS_{n,\text{SUB}}(x) = N_n^{-1} \sum_{i=1}^{N_n} [\tau_b(t_{n,b,i} - t_n)]^2 + \epsilon_0 = O_p(1)\) and \(k_n \to \infty\). If Condition (C.3) holds, then we use the inequality

\[
Y_n(\epsilon) \leq 2 \left(\frac{\tau_n T_n}{\tau_n}\right)^2 I(|\tau_n T_n/\tau_n| \geq 2^{-1} \epsilon\sqrt{k_n}) + 2 \sum_{i=1}^{N_n} T_{b,i}^2 I(|T_{b,i}| \geq 2^{-1} \epsilon\sqrt{k_n})
\]

to bound, for any \(C > 0\),

\[
P(Y_n(\epsilon) > C) \leq \frac{4}{C \sqrt{k_n}} \sum_{i=1}^{N_n} E T_{b,i}^2 I(|T_{b,i}| \geq 2^{-1} \epsilon\sqrt{k_n}) \to 0
\]
as \(n \to \infty\) by \(\sup_{b \geq 1} \sup_{1 \leq i \leq N_n} ET_{b,i}^2 I(|T_{b,i}| \geq 2^{-1} \epsilon\sqrt{k_n}) \to 0\) and \(|\tau_b T_n/\tau_n| = O_p(1)\). Under Condition (C.4), the above becomes

\[
P(Y_n(\epsilon) > C) \leq C^{-1} 4 ET_b^2 I(|T_b| \geq 2^{-1} \epsilon\sqrt{k_n}) + P(|\tau_b T_n/\tau_n| \geq 2^{-1} \epsilon\sqrt{k_n}) \to 0
\]

with \(|\tau_b T_n/\tau_n| = O_p(1)\) by \(\sup_{n \geq 1} ET_b^2 < \infty\) and \(\tau_b/\tau_n = O(1)\). Next consider the converse of Corollary 6.1.2. For \(A_n \equiv \max_{1 \leq i \leq N_n} |\tau_b(t_{n,b,i} - t_n)|, n \geq 1\), pick an increasing integer sequence \(k_{n+1} > k_n \geq 1\) such that \(P(A_n \geq \sqrt{k_n}) < 2^{-n}\) for each \(n \geq 1\), where \(\{\tau_b(t_{n,b,i} - t_n)\}_{i=1}^{N_n}\) are the subsample statistics defining \(S_{n,\text{SUB}}\) in (6.1). By the Borel-Cantelli lemma, \(A_n < \sqrt{k_n}\) holds eventually for large \(n\) (a.s.\((P)\)). Next recall that \(C_{n,k_n}\) corresponds to the distribution of \(Z_{n,n}^\epsilon = k_n^{-1/2} \sum_{i=1}^{k_n} (Y_{n,i}^\epsilon - m_{n,\text{SUB}})\) from (6.4), where \(\{Y_{n,i}^\epsilon\}_{i=1}^{k_n}\) are iid variables following \(S_{n,\text{SUB}}\). Fix an arbitrary integer \(m \geq 1\). As \(k_n \to \infty\), it holds that \(P^*(|Y_{n,i}^\epsilon| > m^{-1} \sqrt{k_n}) \leq S_{n,\text{SUB}}(-m^{-1} \sqrt{k_n}) + 1 - S_{n,\text{SUB}}(m^{-1} \sqrt{k_n}) \overset{p}{\to} 0\) under any one of the Conditions (C.1)-(C.4), which can be established similarly to arguments above. As \(C_{n,k_n}\) converges by assumption for \(\{k_n\}\), then for any subsequence \(\{n_j\} \subset \{n\}\), we may
extract a further subsequence \( \{ \ell \equiv n_k \} \subset \{ n_j \} \) such that \( \sup_{x \in \mathbb{R}} |C_{\ell,ki}(x) - \Phi(x/\sigma)| \to 0 \) holds, along with \( P^*(Y_i^* \mid k_i \leq \ell) > m^{-1} \sqrt{k_i} \to 0 \) for any \( m \geq 1 \), as \( \ell \to \infty \) (a.s.(P)). As the row-wise independent array \( \{ Y_{\tau,i}/\sqrt{k_i} \}_{i=1}^\ell \), \( \ell \geq 1 \), is infinitesimal and \( Z_i^* \) has a normal \( N(0, \sigma^2) \) limit, it follows that

\[
\lim_{\ell \to \infty} \left[ \int_{|x| < \sqrt{k_\ell}} x^2 dS_{\ell,SB}(x) - \left( \int_{|x| < \sqrt{k_\ell}} x dS_{\ell,SB}(x) \right)^2 \right] = \sigma^2
\]

(a.s.(P)) by classical convergence results to normal laws (Chow and Teicher (1988), ch. 12.2, Theorems 2-3). However, \( \int_{|x| < \sqrt{m}} x^2 dS_{\ell,SB}(x) = \int x^2 dS_{\ell,SB}(x) \) and \( \int_{|x| < \sqrt{m}} x dS_{\ell,SB}(x) = \int x dS_{\ell,SB}(x) = m_{\ell,SB} \) eventually for large \( \ell \) (a.s.(P)) as \( A_\ell < \sqrt{k_\ell} \) eventually. Hence,

\[
\lim_{\ell \to \infty} \tilde{\sigma}^2_{\ell,SB} = \lim_{\ell \to \infty} \left[ \int x^2 dS_{\ell,SB}(x) - (m_{\ell,SB})^2 \right] = \sigma^2.
\]

Now \( \tilde{\sigma}_{n,SB}^2 \overset{p}{\to} \sigma^2 \) follows in Corollary 6.1.2 as \( \{ n_j \} \) was arbitrary.

\[\Box\]

**Proof of Theorem 20.** We first establish \( \tilde{\sigma}_{n,SB}^2 \overset{p}{\to} \sigma^2 \) assuming (6.7). Let \( \epsilon > 0 \) and \( \delta > 0 \), set \( Z \sim N(0, \sigma^2) \), and define \( \tilde{S}_n(x) \equiv n^{-1} \sum_{i=1}^n I[T_{b,i} \leq x] = S_{n,SB}(x + \tau_b(t_n - t(P))) \), \( x \in \mathbb{R} \), as the empirical distribution of the subsample copies \( T_{b,i} = \tau_b(X_i, \ldots, X_{i+b-1}) - t(P) \), \( i = 1, \ldots, N_n \equiv n - b + 1 \) of length \( b = b_n \in [1, n] \). By (6.7) and noting \( \int_{|x| > m} x^2 d\tilde{S}_n(x) = N_n^{-1} \sum_{i=1}^n \int_{|T_{b,i}| > m} x^2 \) for \( m > 0 \), choose and fix integer \( m \geq 1 \) such that \( EZ^2I[|Z| > m] \leq \epsilon/3 \) and \( P(\int_{|x| > m} x^2 d\tilde{S}_n(x) > \epsilon/3) < \delta \) for all \( n \geq m \). For any \( \{ n_j \} \subset \{ n \} \), extract a further subsequence \( \{ \ell \equiv n_k \} \subset \{ n_j \} \) such that \( \sup_{x \in \mathbb{R}} |\tilde{S}_{\ell}(x) - \Phi(x/\sigma)| \to 0 \) as \( \ell \to \infty \) (a.s.(P)) by (6.3) and \( T_n \equiv \tau_n(t_n - t(P)) = o_p(\tau_n/\tau_b) \). Hence, as \( \ell \to \infty \) (a.s.(P)), these two limits imply \( \sup_{x \in \mathbb{R}} |\tilde{S}_{\ell}(x) - \Phi(x/\sigma)| \to 0 \) holds as well as, by the Dominated Convergence Theorem (DCT), \( \int_{|x| \leq m} x^2 d\tilde{S}_\ell(x) \to EZ^2I[|Z| \leq m] \). As the subsequence \( \{ n_j \} \) was arbitrary, \( \sup_{x \in \mathbb{R}} |\tilde{S}_n(x) - \Phi(x/\sigma)| \overset{p}{\to} 0 \) and \( \int_{|x| \leq m} x^2 d\tilde{S}_n(x) \overset{p}{\to} EZ^2I[|Z| \leq m] \) hold as \( n \to \infty \). Then, we may bound

\[
\lim_{n \to \infty} P \left( \left| \int x^2 d\tilde{S}_n(x) - \sigma^2 \right| > \epsilon \right) \\
\leq \lim_{n \to \infty} P \left( \left| \int x^2 d\tilde{S}_n(x) \right| > \epsilon/3 \right) \\
+ \lim_{n \to \infty} P \left( \left| \int x^2 d\tilde{S}_n(x) - EZ^2I[|Z| \leq m] \right| > \epsilon/3 \right) \\
\leq \delta
\]

so that, as \( \epsilon, \delta > 0 \) were arbitrary, \( \int x^2 d\tilde{S}_n(x) \overset{p}{\to} \sigma^2 \) follows directly, implying also \( \int x d\tilde{S}_n(x) \overset{p}{\to} \sigma \).
Lemma 1, the sequence of random variables \( \sigma \{ \bar{Y}_j \} \) for all \( j \) is uniformly integrable. In part(ii) of Theorem 20, note that (6.3) and \( T_n = o_p(\tau_n/n) \) entail that \( \tilde{S}_n(x) \) has distribution \( \tilde{S}_\ell(x) - \Phi(x/\sigma) \) for each \( x \). Hence, by the DCT applied to \( |\tilde{S}_n(x)| \leq 1 \), then \( D_{n,b}(x) = E\tilde{S}_n(x) \to \Phi(x/\sigma) \) holds for any \( x \in \mathbb{R} \) so that \( Y_b \overset{d}{\to} N(0,\sigma^2) \) follows. Consequently, \( EY_b^2 \to \sigma^2 \) is equivalent to \( Y_b^2 \), \( b \equiv b_n \geq 1 \), being uniformly integrable.

Proof of Theorem 21. By the assumptions, (6.3) follows (i.e., \( S_{n,SUB} \) converges to a normal
limit) by Theorem 4.2.1 of Politis et al. (1999). Then, assumption (6.7) with (6.3) gives \( \hat{\sigma}^2_{n,B} \overset{p}{\rightarrow} \sigma^2 \) by Theorem 20 and the convergence of \( C_{n,k,n} \) follows from Theorem 18.  

**Proof of Theorem 22.** We first establish a CLT for \( T_n = \sqrt{n}(X_n - \mu) = n^{-1/2} \sum_{t=1}^n (X_t - EX_t) \) using results from Athreya and Lahiri (2006) (ch. 16). For integers \( n \geq 1, i \geq 1, \) and real \( M > 1, \) define sum quantities \( T_{n,i} \equiv n^{-1/2} \sum_{t=i}^{n+i-1} (X_t - EX_t) \) as well as truncated versions \( T_{n,i}^{(1)}(M) \) and \( T_{n,i}^{(2)}(M) \) for any \( i, n, M. \) By assumption,

\[
\sup_{i \geq 1} |\text{Var}(T_{n,i}) - \sigma^2| \to 0 \quad \text{as } n \to \infty \tag{6.3}
\]

holds and Athreya and Lahiri (2006) (p. 526) show that, under the mixing and moment assumptions, there exists some \( C > 0 \) (not depending on \( M \) or \( n \)) such that

\[
\sup_{i \geq 1} \text{E}[T_{n,i}^{(2)}(M)]^2 \leq C \left( M^{-3\delta/4} + \sum_{k=\lfloor M^{1/4} \rfloor}^{\infty} \alpha(k)^{5/(2+\delta)} \right) \equiv \Lambda(M) \tag{6.4}
\]

for all \( M > 1 \) and \( n \geq 1; \) note \( \lim_{M \to \infty} \Lambda(M) = 0 \) also holds by the mixing assumptions. The proof of Athreya and Lahiri (2006) (Theorem 16.3.2) provides a CLT (6.2) for \( T_n \equiv T_{n,1} \) assuming bounded random variables \( \{X_t\}, \) but the same arguments hold immediately for \( T_{n,1}^{(1)}(\log n) \) (i.e., variables truncated at \( \log n \)) provided that \( \lim_{n \to \infty} \sup_{i \geq 1} |\text{Var}(T_{n,i}^{(1)}(\log n)) - \sigma^2| = 0. \) The latter follows by (6.3)-(6.4) here, so that we have \( T_{n,1}^{(1)}(\log n) \overset{d}{\to} N(0, \sigma^2) \) as \( n \to \infty. \) Also, for \( i = \sqrt{-1} t \) and \( t \in \mathbb{R}, \) if \( \phi_{b,i}(t) \equiv \text{E}e^{itT_{b,i}^{(2)}(\log b)} \) denotes the characteristic function of \( T_{b,i}^{(1)}(\log b) \) for \( i = 1, \ldots, n, \) \( n \equiv n - b + 1, \) the same proof of Athreya and Lahiri (2006) (Theorem 16.3.2) shows

\[
\max_{1 \leq i \leq N_n} \left| \phi_{b,i}(t) - e^{-t^2\sigma^2/2} \right| \to 0, \tag{6.5}
\]

for each \( t \in \mathbb{R} \) as \( n \to \infty \) using (6.3)-(6.4) with \( b^{-1} + b/n \to 0. \)

Now from \( T_{n,1}^{(1)}(\log n) \overset{d}{\to} N(0, \sigma^2) \) and \( T_{n,1}^{(2)}(\log n) \overset{p}{\to} 0 \) as \( n \to \infty, \) where the latter follows from \( \text{E}[T_{n,1}^{(2)}(\log n)]^2 \leq \Lambda(\log n) \to 0 \) under (6.4), we obtain \( T_n = T_{n,1}^{(1)}(\log n) + T_{n,1}^{(2)}(\log n) \overset{d}{\to} N(0, \sigma^2) \) in Theorem 22 by Slutsky’s theorem.

Furthermore, if \( \phi_{b,i}(t) \equiv \text{E}e^{itT_{b,i}} \) denotes the characteristic function of \( T_{b,i} = T_{b,i}^{(1)}(\log b) + \)
$T_{b,i}^{(2)}(\log b), 1 \leq i \leq N_n$, then

$$\max_{1 \leq i \leq N_n} \left| \phi_{b,i}(t) - e^{-t^2\sigma^2/2} \right| \leq \max_{1 \leq i \leq N_n} \left| \phi_{b,i}^{(1)}(t) - e^{-t^2\sigma^2/2} \right| + \max_{1 \leq i \leq N_n} \left| \phi_{b,i}(t) - \phi_{b,i}^{(1)}(t) \right| \to 0$$

for each $t \in \mathbb{R}$ as $n \to 1$ by (6.5) and

$$\max_{1 \leq i \leq N_n} \left| \phi_{b,i}(t) - \phi_{b,i}^{(1)}(t) \right| \leq \max_{1 \leq i \leq N_n} E \left| e^{\overrightarrow{T_{b,i}^{(2)}}(\log b)} - 1 \right| \leq |t| [\Lambda(\log b)]^{1/2} \to 0$$

from (6.4) along with $|e^{i(x+y)} - e^{iy}| = |e^{ix} - 1| \leq |x|$ for $x, y \in \mathbb{R}$. From this, we obtain that if $Y_b$ denotes a random variable with distribution function $D_{n,b}$ from (6.6), then the characteristic function of $Y_b$ satisfies $Ee^{itY_b} = N_n^{-1} \sum_{i=1}^{N_n} \phi_{b,i}(t) \to e^{-t^2\sigma^2/2}$ for each $t \in \mathbb{R}$ as $n \to \infty$. Hence, $Y_b \overset{d}{\to} N(0, \sigma^2)$ as $n \to \infty$ which further implies $\{Y_b^2 : b \geq 1\}$ is uniformly integrable (cf. Lemma 1) as the second moment of $Y_b$ here is $EY_b^2 = N_n^{-1} \sum_{i=1}^{N_n} Var(T_{b,i}) \to \sigma^2$ by (6.3). Now Theorem 22 follows from Theorem 21 as $T_n = O_p(1) = o_p((n/b)^{1/2})$ and $Y_b \overset{d}{\to} N(0, \sigma^2)$ hold and, by Theorem 20(ii), (6.7) does as well.

**Proof of Corollary 6.2.2.** Recall here $T_n \equiv \sqrt{n}(\bar{X}_n - t(P))$ for $t(P) = M(\mu)$ based on $\mu(t) = E\tilde{X}_t, t \in \mathbb{Z}$, and the subsample estimator $S_{n,\text{SUB}}$ is the empirical distribution (6.1) of $\sqrt{b}(\bar{X}_{b,i} - \bar{X}_n)$ for $\bar{X}_{b,i} \equiv \sum_{t=i}^{i+b-1} X_t/b, i = 1, \ldots, N_n \equiv n - b + 1$.

Consider the time series $\{Y_t\}$, defined by $Y_t = X_t - \mu(t)$, having mean zero $\mu_Y = 0$. From a sample $Y_1, \ldots, Y_n$, write $T_{Y,n} \equiv \sqrt{n}(\bar{Y}_n - \mu_Y)$ based on the sample mean $\bar{Y}_n$, and let $\hat{\sigma}_{Y,n,\text{SUB}}^2$ and $S_{Y,n,\text{SUB}}$ denote the subsample variance and distribution estimators for $T_{Y,n}$ as derived from the subsample quantities $b^{1/2}(\bar{Y}_{b,i} - \bar{Y}_n)$ for $\bar{Y}_{b,i} \equiv \sum_{t=i}^{i+b-1} Y_t/b, 1 \leq i \leq N_n$. As $\{Y_t\}$ is an APC strongly mixing time series, Lemma A.6 of Synowiecki (2007) yields

$$\sup_{i \geq 1} \left| \text{Var} \left( \frac{i+n-1}{n} \sum_{t=i}^{i+n-1} Y_t \right) - \sigma^2 \right| \to 0$$

as $n \to \infty$ for some $\sigma > 0$. The assumptions of Theorem 22 then hold for $\{Y_t\}$ so that

$$T_{Y,n} \overset{d}{\to} N(0, \sigma^2), \quad \sup_{x \in \mathbb{R}} |S_{Y,n,\text{SUB}}(x) - \Phi(x/\sigma)| \overset{P}{\to} 0, \quad \hat{\sigma}_{Y,n,\text{SUB}}^2 \overset{P}{\to} \sigma^2, \quad (6.6)$$

as $n \to \infty$ with $b^{-1} + b/n \to 0$. Using (6.1), it holds that

$$\sup_{i \geq 1} \left| \frac{1}{n} \sum_{t=i}^{i+n-1} [\mu(t) - M(\mu)] \right| \leq \frac{C}{n}, \quad \sup_{i \geq 1} |\bar{X}_{b,i} - \bar{Y}_{b,i} - M(\mu)| \leq \frac{C}{b} \quad (6.7)$$

for all $b, n \geq 1$ with some $C > 0$ (not depending on $n, b$). By (6.6)-(6.7), the limit distribution
of $T_n$ follows as $T_n \overset{d}{\to} N(0, \sigma^2)$ by $T_n - T_{Y,n} = n^{-1/2} \sum_{t=1}^n [\mu(t) - M(\mu)] = O(n^{-1/2})$. Likewise, by (6.7), the difference of subsample statistics

$$d_n \equiv \max_{1 \leq i \leq N_n} |\sqrt{b}(\bar{X}_{b,i} - \bar{X}_n) - \sqrt{b}(\bar{Y}_{b,i} - \bar{Y}_n)| \leq C b^{1/2} (b^{-1} + n^{-1}) \to 0$$

as $n \to \infty$. Consequently, using that $S_{Y,n,SUB}(x - d_n) \leq S_{n,SUB}(x) \leq S_{Y,n,SUB}(x + d_n)$ holds for all $x \in \mathbb{R}$, we find that

$$\sup_{x \in \mathbb{R}} |S_{n,SUB}(x) - \Phi(x/\sigma)| \leq \sup_{x \in \mathbb{R}} |S_{Y,n,SUB}(x) - \Phi(x/\sigma)| + \sup_{x \in \mathbb{R}} |\Phi((x + d_n)/\sigma) - \Phi(x/\sigma)| \overset{P}{\to} 0$$

by (6.6) and the continuity of $\Phi(\cdot)$. Hence, (6.3) holds or $S_{n,SUB}$ is consistent. Finally, by (6.6), Theorem 20(i) yields

$$\lim_{m \to \infty} \Delta_m(\epsilon) = 0, \quad \Delta_m(\epsilon) \equiv \sup_{n \geq m} P \left( N_n^{-1} \sum_{i=1}^{N_n} b[\bar{Y}_{b,i}]^2 I(\sqrt{b}|\bar{Y}_{b,i}| > m) > \epsilon \right) \quad (6.8)$$

for each $\epsilon > 0$. Fixing $\epsilon > 0$ and using (6.7), we have

$$\sup_{n \geq m} P \left( N_n^{-1} \sum_{i=1}^{N_n} b[\bar{X}_{b,i} - M(\mu)]^2 I(\sqrt{b}|\bar{X}_{b,i} - M(\mu)| > m) > \epsilon \right) \leq \sup_{n \geq m} P \left( b^{-1} C^2 + N_n^{-1} \sum_{i=1}^{N_n} b[\bar{Y}_{b,i}]^2 I(\sqrt{b}|\bar{Y}_{b,i}| > m - b^{-1/2} C) > \epsilon/2 \right) \leq \Delta_{m-1}(\epsilon/4)$$

where the last inequality follows for any large $m$ such that $b^{-1/2} C < 1$ and $b^{-1} C^2 < \epsilon/4$ hold for $n \geq m - 1$ based on $C > 0$ in (6.7). By this, (6.8) and (6.3) (i.e., $S_{n,SUB}$ consistency), $\hat{\sigma}^2_{n,SUB} \overset{P}{\to} \sigma^2$ holds by Theorem 20(i) and then $\sup_{x \in \mathbb{R}} |C_{n,k_n}(x) - \Phi(x/\sigma)| \overset{P}{\to} 0$ follows by Theorem 18.

**Proof of Corollary 6.2.3.** By (6.2) and (6.3), $\{T_n^2 : n \geq 1\}$ is uniformly integrable where $T_n \equiv n^{\alpha/2}(\bar{X}_n - \mu)$. This implies (6.7) by Theorem 20(iii). Now Corollary 6.2.3 follows from Corollary 1 under (6.3).

**Proof of Corollary 6.2.4.** This follows as a special case of Corollary 6.2.3.
List of Figures

3.1 Difference of mean-shifted distributions. ................................. 52
3.2 Graphical evaluation of the ARE. ............................................. 76

4.1 Comparison of Hermite coefficient function $J_m(x)$ and estimator $\hat{J}_{m,n}(x)$. . . . 112
4.2 Comparison of two different estimators for the Hermite coefficient function. . . . 114
4.3 Boxplots for modified local Whittle estimation. ............................... 116
4.4 Ethernet traffic data and estimated Hermite coefficient function. .............. 117
4.5 Empirical size of two-sample Kolmogorov-Smirnov test for LRD data. ............ 119

5.1 CUSUM process for the river Chemnitz together with bootstrapped level of significance. ................................................................. 139
5.2 Searching for additional change-points in the Chemnitz data set. ................. 139
5.3 Annual maximum flows of the river Elbe at Dresden from 1850 to 2012. ........ 140
5.4 Cramèr-von Mises change-point test for annual maximum flows of the river Elbe. 141
### List of Tables

3.1 Empirical power of change-point tests when the Hurst coefficient is known. 81
3.2 Empirical size of change-point tests when the Hurst coefficient is estimated. 82
3.3 Empirical power of change-point tests against mean-shift. The Hurst coefficient is estimated. 83
3.4 Empirical power of change-point tests against changes in mean and variance. The Hurst coefficient is estimated. 84
3.5 Empirical power of change-point tests against distributional change. The Hurst coefficient is estimated. 84
3.6 Empirical power of change-point tests for $\text{farima}(0,0.2,0)$-sequences. 86
3.7 Empirical power of change-point tests for $\text{farima}(1,0.2,0)$-sequences. 87
3.8 Empirical power of change-point tests for short memory $\text{AR}(1)$-sequences. 87
4.1 Empirical size of the test for monotonicity of transformations. 116
4.2 Empirical power of the test for monotonicity of transformations. 116
4.3 Empirical size of two-sample Kolmogorov-Smirnov test for LRD data. 119
5.1 Empirical size for bootstrapped and fpca-based tests for independent functional data. 142
5.2 Empirical power for bootstrapped and fpca-based tests for independent functional data. 143
5.3 Empirical power for bootstrapped and fpca-based tests for FAR(1) series. 144
5.4 Empirical power for bootstrapped and fpca-based tests for FAR(1) series. 145
5.5 Empirical size for Cramér-von Mises/CUSUM test for a real-valued AR(1)-process. 146
5.6 Empirical power for Cramér-von Mises/CUSUM test against a mean-shift in real-valued AR(1)-processes. 146
5.7 Empirical power for Cramér-von Mises/CUSUM test against a change in skewness. 147
Bibliography


Bibliography 191


