Topics in Real-Valued and Circular Time Series Models Under Long-Range Dependence

Doctoral thesis
for obtaining the academic degree
Doctor of Natural Sciences
(Dr. rer. nat.)

submitted by

Britta Steffens

at the

Faculty of Sciences
Department of Mathematics and Statistics

Konstanz, 2022

Tag der mündlichen Prüfung: 07.02.2023

1. Referent: Prof. Dr. Jan Beran
2. Referent: Prof. Dr. Yuanhua Feng

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor Prof. Dr. Jan Beran for his continuous support, guidance and advice, and his tremendous patience during my studies. I am very grateful that he gave me the chance to learn from and work with him, and that he supported my interests in applying the theory to “real-world“ data examples.

I also want to thank Dr. Sucharita Ghosh for valuable collaboration.

Furthermore, I thank Prof. Dr. Yuanhua Feng for his willingness to review this thesis.

Special thanks go to Dr. Volker Bürkel for his valuable feedback and support throughout my studies.

For interesting discussions (also beyond the topics of our projects) and their advice, and for never boring coffee breaks or meetings, I would like to thank Prof. Dr. Michael Kupper, and my former colleagues and cubicle neighbors at the University of Konstanz, in particular, Veronika Weißer, Dieter Schell, Timo Dimitriadis, Klaus Telkmann, Gilbert Koch, Haiyan Liu, Julian Vill, Freya Bachmann, Markus Schlipf and Jeremy Näscher.

Last but not least, I am very thankful to my family and friends for the immense support, for suffering my moods, and for their enormous patience and encouragement.

Summary

This thesis deals with two special nonparametric regression models in the context of long-range dependent processes. Both models and settings are motivated by examples from meteorology: temperatures and wind directions, respectively.

Part I focuses on a real-valued *seasonal fixed-design model* with a linear long-memory residual process. We introduce *periodic processes* and propose a hypothesis test for a change in the expected number of *seasonal exceedances*. An appropriate test statistic in this context is of *Kolmogorov-Smirnov-type*. We define it based on so-called *periodic empirical processes*. The proof of the asymptotic distribution of our test statistic is based on the *uniform weak reduction principle* from Dehling and Taqqu (1989), which we extend to periodic processes involving *periodic sample means*. We show that the limiting behavior of long-range dependent periodic empirical processes is essentially the same as of long-range dependent linear empirical processes.

Part II deals with a *circular-circular regression model* under the assumption of long-range dependence. This is a random-design model where both, the response and the explanatory variables are random and have values in $[0, 2\pi)$. We define the circular long-memory process by Gaussian subordination and estimate the mean direction using a *circular Nadaraya-Watson kernel estimator*. In the context of such long-range dependent models one observes dichotomous asymptotic behavior of the kernel estimator depending on the size of the bandwidth. Here, optimal convergence of our estimator is meant in the sense that *the leading error term does not depend on the choice of the bandwidth*. We determine a condition on the sequence of bandwidths which ensures optimal convergence. Based on this, we prove uniform consistency of our circular kernel estimator and derive asymptotic confidence bands. Thereby, we extend the results by Beran and Ghosh (2020) to nonparametric circular regression and the results by Mielniczuk and Wu (2004) to circular random-design models.

In either part, our theoretical results are illustrated by applying them to the above mentioned meteorological data sets.

Zusammenfassung

Diese Arbeit betrachtet zwei spezielle nicht-parametrische Regressionsmodelle unter der Annahme eines *long-memory* Fehlerprozesses. Beide Modelle und Situationen sind durch Beispiele aus der Meteorologie motiviert: Temperaturen bzw. Windrichtungen.

In Teil I befassen wir uns mit einem reellwertigen saisonalen Modell mit fixem Design und linearem *long-memory* Fehlerprozess. Wir führen periodische Prozesse ein und definieren einen Hypothesentest für eine Änderung in der erwarteten Anzahl an saisonalen Überschreitungen. Eine geeignete Teststatistik in diesem Kontext ist vom *Kolmogorov-Smirnov-Typ*, welche wir für sogenannte periodische empirische Prozesse definieren. Der Beweis für die asymptotische Verteilung unserer Teststatistik beruht auf dem *uniform weak reduction principle* von Dehling und Taqqu (1989), welches wir auf periodische Prozesse unter Einbeziehung des periodischen Stichprobenmittels erweitern. Wir zeigen, dass unter der Annahme von *long memory* das asymptotische Verhalten von periodischen empirischen Prozessen im Wesentlichen das gleiche ist wie das von linearen empirischen Prozessen.

Teil II behandelt ein zirkuläres *long-memory* Regressionsmodell mit zufälligem Design. Bei diesem Modell haben die zufälligen unabhängigen und abhängigen Variablen jeweils Werte in $[0, 2\pi)$. Wir definieren den zirkulären *long-memory* Prozess durch Gaußsche Subordination und schätzen die *mean direction* unter Verwendung eines zirkulären Nadaraya-Watson Kernschätzers. Im Zusammenhang mit solchen long-memory Modellen können wir ein dichotomes, von der Größe der Bandweite abhängiges, asymptotisches Verhalten des Kernschätzers beobachten. Optimale Konvergenz unseres Schätzers ist hier gemeint im Sinne der *Unabhängigkeit des führenden Terms von der Wahl der Bandweite*. Um Optimalität zu gewährleisten geben wir eine Bedingung an die Sequenz der Bandweiten an. Basierend darauf weisen wir gleichmäßige Konsistenz unseres Schätzers nach und leiten asymptotische Konfidenzintervalle her. Dabei erweitern wir sowohl die Ergebnisse von Beran und Ghosh (2020) auf nicht-parametrische zirkuläre Regressionmodelle als auch die Ergebnisse von Mielniczuk und Wu (2002) auf zirkuläre *random-design* Modelle.

In beiden Teilen wenden wir abschließend unsere theoretischen Ergebnisse auf die oben genannten meteorologischen Datenbeispiele an.

Contents

List of Tables

List of Figures

Introduction

This thesis contains results on statistical inference for long-memory processes in the context of different models with real-valued and circular ($[0, 2\pi)$-valued) state spaces. In the real-valued setting, we consider seasonal exceedances under long-memory and thereby extend the uniform reduction principle for empirical processes from Dehling and Taqqu (1989) [106]. In the circular setting, we derive asymptotic confidence bands for the mean direction and thereby generalize the results in parametric circular regression by Beran and Ghosh (2020) [34] and the results on real-valued random-design models by Mielniczuk and Wu (2004) [298]. We illustrate our technical results by applying them to data examples from meteorology: daily temperatures and daily wind directions, respectively.

The general set-up of the real-valued situation is the following: A discrete-time additive time series model is given by

$$Y_t = m(X_t) + Z_t, \qquad t \in \mathbb{Z}, \tag{1}$$

where $(Y_t, X_t)_{t\in\mathbb{Z}}$ is a (real-valued) bivariate process, m is the mean function and $(Z_t)_{t\in\mathbb{Z}}$ is a stationary zero-mean error process. The mean function can be decomposed into a long-term trend and a seasonal component. For the analysis of real-valued time series see standard textbooks, such as Heiler (1970) [200], Priestley (1981) [328], Cleveland et al. (1990) [80], Ghysels and Osborn (2001) [154], Brockwell and Davis (2016) [65], Beran (2017) [25], Kitagawa (2020) [245], and references therein.

Rewriting model (1) in a slightly different manner with circular response variables Θ_t, circular explanatory variables Ψ_t, and mean direction ν, i.e.,

$$\Theta_t = [\nu(\Psi_t) + Z_t] \ mod\, 2\pi, \qquad t \in \mathbb{Z}, \tag{2}$$

we obtain a general regression model for circular time series, which is our framework in the circular situation. Circular or directional data arise in many scientific areas such as meteorology and biology. Typical examples are, e.g., wind directions or animal navigation and movement. First approaches to the analysis of circular time series can be found in the publications of

Rayleigh (1888) [336] and of Pearson (1905) [314]. Since the 1970s various methods for the analysis of circular time series have been developed and lateron circular methods have been extended to circular time series analysis, see among others, Wehrly and Johnson (1980) [413], Breckling (1989) [62], and Fisher and Lee (1994) [140].

Either setting captures various types of specific time series models: Firstly, models with deterministic or random explanatory variables leading to a fixed- or a random-design model, respectively, and secondly, models with error processes that exhibit different dependence structures such as independence, weak or short-range dependence, and strong or long-range dependence. In literature, models of both set-ups have mostly been studied under the assumption of independent and identically distributed (iid) or weakly dependent errors. However, in many applications such as oceanography, meteorology, geophysics, biology and finance, time series with serial or long-range dependent errors are observed.

This thesis focuses on long-range dependent processes, i.e., processes that exhibit long-memory behavior. Such processes are characterized by their hyperbolically decaying correlations and their spectral density having a pole at the origin. This means, for a long-memory process (Z_t), if γ denotes the autocovariance function of (Z_t) and f the spectral density of (Z_t), then, for the long-memory parameter $d \in \left(0, \frac{1}{2}\right)$,

$$\gamma(k) = \operatorname{cov}(Z_t, Z_{t+k}) \underset{k\to\infty}{\sim} c_\gamma |k|^{2d-1}$$

and

$$f(\lambda) \underset{\lambda\to 0}{\sim} c_f |\lambda|^{-2d},$$

respectively, where $c_\gamma, c_f > 0$ are constants.

Statistical inference for real-valued long-memory processes in general is considered, e.g., in textbooks of Beran (1994) [22], Giraitis et al. (2012) [158], Beran et al. (2013) [33], and references therein. In the circular environment, long-range dependence has gained more interest during the last 10 years and is, e.g., considered in Di Marzio et al. (2012) [114], Beran and Ghosh (2020) [34] and Beran, Steffens and Ghosh (2021a) [38] (as well as summarized in Beran, Steffens and Ghosh (2022) [40]) in different settings.

Generally, the investigation in time series analysis is based on the estimation of time series components and of unknown memory parameters. Depending on the regression model and the dependence structure of the error process, different estimation methods are in use and have been widely studied. In Part I of this thesis, the focus is on a *real-valued fixed-design model*, and in Part II on a *circular random-design model*, in either part under the assumption of long-range dependent residuals. The model components are estimated using nonparametric regression,

namely *kernel density estimation* and *kernel regression*. Nonparametric estimation under long-range dependence is, e.g., discussed in Beran and Feng (2002a-c) [28], [29], [30] for estimation of a (non-periodic) trend component, and in Beran, Steffens and Ghosh (2018) [37] for estimation of a seasonal component.

Due to the fact that results are commonly based on asymptotic distributions, the convergence rate of the estimators is of basic interest. For the trend component, the convergence rate strongly depends on the long-memory parameter, whereas for the seasonal component the rate is of the same order as in the iid case. This is shown in Beran, Steffens and Ghosh (2018) [37] for real-valued processes and in Beran and Ghosh (2020) [34] for circular processes. One main issue in this context is the choice of an optimal kernel bandwidth, which basically amounts to a trade-off between minimizing bias and variance of the estimator. In the long-range dependent random-design setting, we observe an interesting phenomenon, the so-called *smoothing dichotomy*, which occurs if the bandwidth of the estimator is too small to recover the long-memory behavior in the process. This phenomenon is widely discussed in the real-valued setting, see, e.g., Wu and Mielniczuk (2002) [420], Mielniczuk and Wu (2004) [298], and references therein. Such a dichotomous behavior is observed in the circular situation as well.

In detail: **Part I** deals with the real-valued situation. We consider the general nonparametric (real-valued) time series model $(Y_t)_{t\in\mathbb{Z}}$ given in (1) as a fixed-design model of the form

$$Y_t = m_t + Z_t = \mu_t + S_t + Z_t, \qquad t \in \mathbb{Z}, \tag{3}$$

where m_t is the unknown mean function, decomposed into a long-term trend μ_t and a seasonal component S_t with even period $T_0 \geq 2$, and $(Z_t)_{t\in\mathbb{Z}}$ is a stationary long-range dependent error process with zero mean and finite variance.

In our situation (3), we are interested in a change in the expected number of seasonal exceedances above a certain threshold within a given time interval. The topic came up when analyzing temperature data where the question arose how to detect a significant change in the number of days (of a specific month) over the years at which the temperature exceeds a certain threshold. Here, we develop a hypothesis test for such a question, investigate the asymptotic distribution of the associated test statistic, and derive rejection regions. This means, for some $0 < c < \infty$, seasonal time points t_j, $j = 1, \ldots, k \leq T_0$, and N periods, we define

$$\chi_i = \sum_{j=1}^{k} \mathbb{1}\Big\{Y_{t_j+(i-1)T_0} \leq c\Big\}, \qquad i = 1, \ldots, N,$$

as the *number of seasonal exceedances* over some level c in period number i at seasonal time

points $t_1, \ldots, t_k$, and we test the hypothesis

$$H_0 : \; E[\chi_1] = \ldots = E[\chi_N]$$

against the alternative

H_1: there is at least one i $(1 \leq i \leq N-1)$ such that either for all $l \leq i$ and for all $m \geq i+1$, $E[\chi_l]$ is smaller than $E[\chi_m]$ or for all l, m, $E[\chi_l]$ is larger than $E[\chi_m]$.

Our test statistic is based on the idea to compare the average number of exceedances over different periods (of years). Precisely, we define the test statistic in terms of so-called *seasonal empirical processes*. This means the test statistic M is given by

$$M = \sup_{\tau \in (0,1)} \sum_{j=1}^{k} D_j(\tau), \qquad k \leq T_0,$$

which is a *Kolmogorov-Smirnov-type test statistic* adapted to our situation of several periods. Here,

$$D_j(\tau) = c_{N,\tau} \cdot \left(F^*_{Z;N-[N\tau],t_j}(c - Y_{t_j}) - F_{Z;[N\tau],t_j}(c - Y_{t_j}) \right), \qquad j = 1, \ldots, k \leq T_0,$$

where $F_{Z;[N\tau],t_j}$ is the empirical distribution function of the first $[N\tau]$ observations, $F^*_{Z;N-[N\tau],t_j}$ is the empirical distribution of the last $N - [N\tau]$ observations, and $c_{N,\tau}$ is some suitable normalizing constant. In order to derive asymptotic rejection regions, we determine the asymptotic distribution of the test statistic. This is based on the extension of the *uniform weak reduction principle* for empirical processes from Dehling and Taqqu (1989) [106] to seasonal empirical processes involving seasonal sample means.

This first part concludes with an application to daily temperature data. Before applying our test procedure, we have to estimate the deterministic time series components and unknown long-memory parameters. Estimation of time series components is performed by trigonometric regression and the unknown long-memory parameters are obtained through Whittle quasi-likelihood estimation together with fitting a fractional ARIMA(p,d,0) process of appropriate order p to the data by use of the BIC selection criterion (see, e.g., Giraitis et al., 2012 [158] and Beran et al., 2013 [33]).

Part II of the thesis deals with circular time series models and directional data, i.e., data that can be represented by points on the unit sphere. Due to the different topologies on the real line and the unit sphere, statistical analysis of directional data differs from the real-valued case. Consequently, in particular, an appropriate notion of „*variation*“ has to be established. Hence,

modeling directional data requires both, a class of models on the unit sphere and a theory of statistical inference for these models.

We focus on a circular time series model for a bivariate process $(\Theta_t, \Psi_t)_{t\in\mathbb{Z}}$, where both, (Θ_t) and (Ψ_t) represent angles, i.e., we deal with a *circular-circular regression model* of the type (2)

$$\Theta_t = [\nu(\Psi_t) + Z_t] \ mod\ 2\pi, \qquad t \in \mathbb{Z}.$$

Here, $\nu : [0, 2\pi) \to \mathbb{R}$ is the unknown (conditional) mean direction and (Z_t) again is a stationary zero-mean long-memory error process. The circular process (Ψ_t) is assumed to be defined by Gaussian subordination, i.e.,

$$\Psi_t := G^{-1}(\Phi(X_t)), \qquad t \in \mathbb{Z},$$

where (X_t) is a stationary long-memory process with zero mean and unit variance, Φ denotes the standard normal distribution function and $G : [0, 2\pi) \to [0, 1]$ is a continuous circular distribution function.

This approach was introduced into the context of circular long-memory processes by Beran and Ghosh (2020) [34]. Based on the fact that such Gaussian subordinated processes have a unique Hermite polynomial representation, Beran and Ghosh show that long-range dependence in (X_t) leads to long-range dependence in (Ψ_t), where the long-memory parameter depends on the Hermite rank m of the particular Hermite polynomial representation. An earlier approach is given by Di Marzio et al. (2012) [114], where the long-memory property of the circular process is defined via a specific covariance matrix. Both publications consider estimation of the mean direction in a fixed-design regression model (under the assumption of long-range dependence).

However, our focus is on a circular random-design regression model. We are interested in nonparametric kernel estimation of the (conditional) mean direction ν using circular kernels under the assumption of long-range dependence for both, the error process (Z_t) and the Gaussian long-memory process (X_t). A famous example in this context are daily wind directions at different locations. They, on the one hand, exhibit long-memory behavior and, on the other hand, are presumably related to each other such that they require a circular-circular regression model.

For any $\psi^0 \in [0, 2\pi)$ our sequence of kernel estimators - a circular version of the *Nadaraya-Watson estimator* - is given by

$$\hat{\nu}(\psi^0) = \frac{n^{-1} \sum_{t=1}^{n} K_b(\psi_t - \psi^0)\Theta_t}{\hat{f}_n(\psi^0)},$$

where K_b is a circular kernel function of even order $2k$, $k \geq 1$, and $(\hat{f}_n)_n$ is a sequence of circular kernel density estimators of the marginal density f of Ψ. As mentioned above, the crucial issue in this context is the smoothing dichotomy. Instead of calculating an optimal bandwidth by minimizing a suitable criterion, we define a sequence of bandwidths as *asymptotically optimal* provided it no longer has any impact on the convergence rate of the estimator. This consequently results in an "*asymptotically optimal convergence rate*". We show that the existence of such an optimal sequence of bandwidths depends on the relationship of the long-memory parameters of the two processes (Z_t) and (X_t), denoted by d_Z and d_X, and the order of the kernel function. Based on this condition, we prove asymptotic normality and uniform consistency of the sequence of circular kernel estimators and derive asymptotic confidence band for our kernel estimator. Moreover, we develop a test procedure for testing for the existence of such an optimal sequence of bandwidths. The technical statement of the hypothesis reads

$$H_0 : \; d_Z \leq \max\left\{\frac{1}{5}d_X, \; \frac{1}{2(4k+1)}\right\}$$

against the alternative

$$H_1 : \; d_Z > \max\left\{\frac{1}{5}d_X, \; \frac{1}{2(4k+1)}\right\},$$

where the specific form of the right-hand side essentially results from the representation of the convergence rate of our estimator. Our test statistic now is based on the estimators of the two long-memory parameters d_Z and d_X and their asymptotic normality.

Summarizing the above: On the one hand, we generalize the results by Beran and Ghosh (2020) [34] to a nonparametric circular-circular regression model and on the other hand, we extend the results by Mielniczuk and Wu (2004) [298] for nonparametric estimation in the context of real-valued random-design models to the circular case under the assumption of independence of (Z_t) and (X_t). In this specific situation, we obtain an explicit condition for the existence of an optimal sequence of bandwidths, which entails formulation of an hypothesis test.

We complete this second part with an application to daily wind directions. Due to the seasonal pattern of such data, we perform a deseasonalization method and then estimate the mean direction using the circular Nadaraya-Watson estimator. The unknown long-memory parameters are obtained through variants of Whittle estimation together with fitting a fractional ARIMA(p,d,0)-process of appropriate order p by use of the BIC selection criterion.

Although the two parts deal with different issues, they have a few things in common: a nonparametric regression model, the dependence structure of the residual process, and estimation of time series components and of long-memory parameters. It can be shown that the theory in the real-valued state space has similar counterparts in the circular state space involving a

new distance measure. Additionally, in both cases we apply our results to data examples from meteorology where long-range dependence is a reasonable assumption.

The thesis is structured as follows:

Part I - Inference for real-valued stationary long-memory processes: The introductory Chapter 1 recalls some basic definitions and theorems in the theory of long-range dependent processes and introduces examples of this type of processes. The focus is, in particular, on processes which occur as limit processes and which will be addressed in the subsequent Chapter 2. In this second chapter, limit theorems, especially the uniform weak reduction principle both, by Dehling and Taqqu (1989) [106] for Gaussian subordinated processes, and by Ho and Hsing (1996) [206] and Giraitis and Surgailis (1999) [165] for linear processes, are presented. Chapter 3 reviews results in nonparametric estimation of model components with a focus on kernel density estimation, and kernel and locally weighted regression. Furthermore, methods for model selection combined with estimation of unknown model parameters, mainly the Whittle estimation and modified versions, are addressed. This part's last chapter (Chapter 4) covers our recently published theoretical results (Beran, Steffens and Ghosh, 2021b [39]) on testing for a change in the expected number of seasonal exceedances and illustrates these theoretical results on temperature data at different locations in Germany.

Part II - Inference for bivariate circular long-memory processes: This part's first chapter (Chapter 5) gives a short overview of general definitions and notions in circular statistics, and of circular distributions and processes. In particular, a class of circular long-memory processes that are generated via Gaussian subordination is introduced. Chapter 6 contains definitions, notations and main results in the context of parametric and nonparametric circular regression including circular kernels and circular kernel estimators. In this part's last chapter (Chapter 7), our recently published theoretical results (Beran, Steffens and Ghosh, 2021a [38]) on the asymptotics of our circular Nadaraya-Watson estimator are presented and these theoretical results are applied to daily wind directions at three locations in the United States.

Part I

Inference for real-valued long-memory processes

Chapter 1

Long-memory processes and their asymptotic properties

Throughout this Part I, our focus is on long-range dependent and stationary processes. We consider discrete-time $\mathbb{R}$-valued stochastic processes $(X_t)_{t\in\mathbb{Z}}$, $(X_t)_{t\in\mathbb{N}}$ or $(X_t)_{t=1,\dots,n}$, and write (X_t) for short.

In this first chapter, whenever not otherwise specified, notions and results follow Beran et al. (2013) [33] and Pipiras and Taqqu (2017) [325].

1.1 Basic definitions and notations

We introduce *stationary long-range dependent processes* by use of two main characteristics: the *autocovariance function* and the *spectral density*.

Definition 1.1.1. *For a (real-valued) stochastic process (X_t) with $E\left[X_t^2\right] < \infty$ and $\mu_t = E[X_t]$ the **autocovariance function** is defined as*

$$\begin{aligned} \gamma:\ \mathbb{Z}\times\mathbb{Z} &\to \mathbb{R} \\ (s,t) &\mapsto \gamma(s,t) := cov(X_s, X_t) = E\left[(X_s-\mu_s)(X_t-\mu_t)\right]. \end{aligned} \tag{1.1}$$

We focus on stationary processes for which the autocovariance function has a special form.

Definition 1.1.2. *Let (X_t) be a stochastic process.*

- *(X_t) is said to be **stationary** if*

  (i) (X_t) has finite second moments, i.e., $E\left[|X_t|^2\right] < \infty$, $t \in \mathbb{Z}$,

  (ii) (X_t) has a constant mean, i.e., there exists $\mu \in \mathbb{R}$ such that $E[X_t] = \mu$ for all $t \in \mathbb{Z}$,

(iii) the autocovariance function γ defined in (1.1) is time-shift invariant, i.e., for all $t, k \in \mathbb{Z}$:

$$cov(X_t, X_{t+k}) = cov(X_k, X_0) = \gamma(k, 0).$$

*Other notions are **weakly stationary**, **second-order stationary**, **covariance stationary** or **widesense stationary**.*

- *(X_t) is said to be **strictly** or **strongly stationary** if for all $k, l \in \mathbb{Z}$ and $t_1, \ldots, t_l \in \mathbb{Z}$:*

$$(X_{t_1}, \ldots, X_{t_l}) \stackrel{d}{=} (X_{t_1+k}, \ldots, X_{t_l+k}),$$

where $\stackrel{d}{=}$ denotes equality in distribution.

For a stationary process (X_t), (iii) means that the autocovariance function only depends on the time-lag k. Therefore, we redefine the *autocovariance function of a stationary process* as

$$\begin{aligned} \gamma : \mathbb{Z} &\to \mathbb{R}, \\ k &\mapsto \gamma(k) = \operatorname{cov}(X_t, X_{t+k}). \end{aligned}$$

Thus, under stationarity, since $\gamma(0) = var(X_t)$, $t \in \mathbb{Z}$, the **autocorrelation function** of a process (X_t) is given by

$$\varrho(k) = \frac{\gamma(k)}{\gamma(0)}, \qquad k \in \mathbb{Z}.$$

Remark. *If a strictly stationary process (X_t) has finite second moments, (X_t) is (second-order) stationary. This does not hold, for instance, for an iid process with standard Cauchy distribution. In particular, for Gaussian processes (X_t) (see Definition 1.2.17) strict stationarity and weak stationarity are equivalent.*

A more general notion of stationarity is the so-called **difference stationarity**, which is given if a non-stationary process (X_t) can be transformed into a stationary process (Y_t) by taking differences, i.e.,

$$Y_t = (1 - B)X_t = X_t - X_{t-1},$$

and iterations of these. Here B denotes the **backshift operator** defined by

$$BX_t = X_{t-1}. \tag{1.2}$$

Such difference-stationary processes are also called **integrated processes**. An important example are *integrated autoregressive moving average processes* (*ARIMA processes*). These processes are described in detail in Chapter 1.2.1.

Standard examples of stationary and non-stationary processes are the following:

Examples 1.1.3. *(i) A (second-order) stationary process is the so-called* ***white noise process*** *defined as an uncorrelated process (ε_t) with zero mean and finite variance σ_ε^2. If the random variables are iid, this process is moreover strictly stationary. If, in addition, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, this process is called* ***Gaussian white noise****.*

(ii) An important example of a non-stationary (stochastic) process built from a white noise process is the so-called ***random walk*** *defined as a process (X_t) with*

$$X_t = X_{t-1} + \varepsilon_t = X_0 + \sum_{j=1}^{t} \varepsilon_j,$$

where $X_0 = 0$, and (ε_t) is a white noise process as defined in (i). However, this process is difference-stationary, since taking one difference leads to a stationary white noise process.

(iii) Any ***moving average process*** *(X_t)* ***of order*** *q (**MA(q) process**), $q \geq 0$, which is a moving average over the white noise process (ε_t) defined in (i), i.e.,*

$$X_t = \sum_{j=0}^{q} \psi_j \varepsilon_{t-j} = \psi(B)\varepsilon_t,$$

where ψ is a polynomial of the form

$$\psi(z) = 1 + \sum_{j=1}^{q} \psi_j z^j, \qquad q \geq 1, \tag{1.3}$$

and B denotes the backshift-operator defined in (1.2), is stationary.

(iv) Let (X_t) be an ***autoregressive process of order*** *p (**AR(p) process**), $p \geq 0$, i.e., (X_t) is the solution of the equation*

$$\sum_{j=0}^{p} \varphi_j X_{t-j} = \varphi(B)X_t = \varepsilon_t, \qquad t \in \mathbb{Z}, \tag{1.4}$$

with white noise process (ε_t), polynomial ϕ of the form

$$\varphi(z) = 1 - \sum_{j=1}^{p} \varphi_j z^j, \qquad p \geq 1, \tag{1.5}$$

and backshift-operator B defined in (1.2).

This process is stationary if and only if the polynomial $\varphi(z)$ has no roots on the unit circle,

i.e., if $\varphi(z) \neq 0$ *for* $|z| = 1$.

Hence, e.g., the AR(1) process $\varepsilon_t = X_t - \varphi_1 X_{t-1}$ *is stationary if and only if* $|\varphi_1| \neq 1$. *This shows, why the random walk process defined in (ii), which is an AR(1) process with* $\varphi_1 = 1$, *is a non-stationary process.*

While, in time domain, the dependence structure of a process is characterized by the autocovariance function, in frequency domain (after *Fourier transformation*) it is characterized by the *spectral density function.*

Definition 1.1.4. *For a stationary process* (X_t) *with autocovariance function* γ, *the* ***spectral density*** *is (formally) defined as the Fourier transform of the autocovariance function*

$$f(\lambda) := \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \gamma(k) e^{-ik\lambda}, \qquad \lambda \in [-\pi, \pi]. \tag{1.6}$$

Conversely, the (formally defined) Fourier transform of the spectral density f is the autocovariance function γ given by

$$\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\nu) e^{ik\nu} d\nu, \qquad k \in \mathbb{Z}.$$

For an overview of spectral analysis of time series see, e.g., Priestley (1981) [328].

In the following, we introduce the notion of *long-range dependent stationary process* in terms of the autocovariance function and the spectral density.

The two standard definitions by Karamata and Zygmund (Karamata, 1930a,b, 1933 [235], [236], [237] and Zygmund, 1968 [427]) involve a parameter $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$ and so-called *slowly varying functions.* We follow the outline in Beran et al. (2013) [33].

Definition 1.1.5. *A function* $L : [c, \infty)$ *with* $c \geq 0$ *is said to be*

(i) ***slowly varying at infinity (in Karamata's sense)*** *if* L *is positive (and measurable) for* x *large enough and, for any* $u \in \mathbb{R}^+$

$$L(ux) \underset{x \to \infty}{\sim} L(x).$$

(ii) ***slowly varying at infinity (in Zygmund's sense)*** *if* L *is positive for* x *large enough, and for any* $\delta > 0$ *there exists some finite* $x_0(\delta) > 0$ *such that for* $x > x_0(\delta)$, *the functions* $x^{\delta} L(x)$ *and* $x^{-\delta} L(x)$ *are monotone.*

In each case, a function L *is called* ***slowly varying at the origin*** *if* $\tilde{L}(x) = L(x^{-1})$ *is slowly varying at infinity.*

Here and throughout the thesis, in this context, for both, sequences and functions, the notation $a_n \sim b_n$ means that the ratio a_n/b_n converges to one as n tends to infinity.

Lemma 1.1.6. *Let $L(k)$ be such that $L(k) \to c$, $c > 0$ constant, as $k \to \infty$. Then L is slowly varying at infinity.*

This is the simplest example of a slowly varying function.

Remark. *It is shown, for instance, in Beran et al. (2013) [33] that the class of slowly varying functions in Zygmund's sense is a proper subset of the class of slowly varying functions in Karamata's sense.*

For a detailed introduction into slowly varying functions and the more general notion of *regularly varying functions* see, e.g., Bingham et al. (1989) [56], and for more results under the assumption of long-range dependence and for comparison of different characterizations, see, e.g., Beran et al. (2013) [33] and Pipiras and Taqqu (2017) [325].

Based on any of these two notions, we can formally define *(linear) dependence.*

Definition 1.1.7. *Let (X_t) be a stationary process with autocovariance function $\gamma(k)$, $k \in \mathbb{Z}$, and spectral density f of the form (1.6), and let L_f be symmetric and slowly varying at zero. If for $\lambda \in [-\pi,\ \pi]$ and $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$,*

$$f(\lambda) = L_f(\lambda)|\lambda|^{-2d},$$

then (X_t) is said to be

(i) ***long-range dependent*** *if $d \in \left(0, \frac{1}{2}\right)$,*

(ii) ***intermediate dependent*** *if $d = 0$ and $\lim_{\lambda \to 0} L_f(\lambda) = \infty$,*

(iii) ***short-range dependent*** *if $d = 0$ and $\lim_{\lambda \to 0} L_f(\lambda) = c_f \in (0, \infty)$,*

(iv) ***antipersistent*** *if $d \in \left(-\frac{1}{2}, 0\right)$.*

This parameter d is the so-called ***memory parameter****.*

Alternative notions of long-range dependence that are equivalently used in literature are **long memory**, **strong dependence** or **persistence**.

In general, stationary processes (X_t) are distinguished by the decay rate of their autocovariance function γ and/or by the asymptotic behavior of their spectral density f (at the origin). The following theorem is given in Beran et al. (2013) [33], Theorem 1.3:

Theorem 1.1.8. *Let (X_t) be a stationary process with autocovariance function γ and spectral density f, and let L_γ and L_f be slowly varying functions in Zygmunds' sense.*

(i) If L_γ is slowly varying at infinity such that

$$\gamma(k) = L_\gamma(k)|k|^{2d-1}, \qquad k \in \mathbb{Z},$$

either $d \in \left(0, \frac{1}{2}\right)$ or $d \in \left(-\frac{1}{2}, 0\right)$, and $\sum_{k\in\mathbb{Z}} \gamma(k) = 0$, then

$$f(\lambda) \underset{\lambda\to 0}{\sim} L_f(\lambda)|\lambda|^{-2d}, \qquad \lambda \in [-\pi, \pi],$$

with

$$L_f(\lambda) = L_\gamma(\lambda^{-1})\pi^{-1}\Gamma(2d)\sin\left(\frac{\pi}{2} - \pi d\right). \tag{1.7}$$

(ii) If L_f is slowly varying at the origin and of bounded variation on (a, π) for all $a > 0$ such that for $0 < \lambda < \pi$,

$$f(\lambda) = L_f(\lambda)|\lambda|^{-2d},$$

where $d \in \left(-\frac{1}{2}, 0\right) \cup \left(0, \frac{1}{2}\right)$, then

$$\gamma(k) \underset{k\to\infty}{\sim} L_\gamma(k)|k|^{2d-1}, \tag{1.8}$$

with

$$L_\gamma(k) = 2L_f(k^{-1})\Gamma(1-2d)\sin(\pi d). \tag{1.9}$$

In the context of a long-range dependent process, one also speaks of *slowly (hyperbolically) decaying correlations* or of a *spectral density having a pole at the origin.*

Taking into account the relationship between the behavior of the spectral density at the origin and the summability of the autocovariances, we obtain a further statement (see Beran et al., 2013 [33], Corollary 1.1):

Corollary 1.1.9. *Under the same assumptions as in Theorem 1.1.8 (ii), we have*

(i) long-range dependence: for $d \in \left(0, \frac{1}{2}\right)$,

$$\sum_{k\in\mathbb{Z}} \gamma(k) = 2\pi \lim_{\lambda\to 0} f(\lambda) = \infty,$$

(ii) intermediate dependence: for $d = 0$ and $\lim_{\lambda\to 0} L_f(\lambda) = \infty$,

$$\sum_{k\in\mathbb{Z}} \gamma(k) = 2\pi f(0) = \lim_{\lambda\to 0} f(\lambda) = \infty,$$

(iii) short-range dependence: for $d = 0$ *and* $\lim_{\lambda \to 0} L_f(\lambda) = c_f \in (0, \infty)$,

$$\sum_{k \in \mathbb{Z}} \gamma(k) = 2\pi f(0) = 2\pi c_f \in (0, \infty),$$

(iv) antipersistence: for $d \in \left(-\frac{1}{2}, 0\right)$,

$$\sum_{k \in \mathbb{Z}} \gamma(k) = 2\pi f(0) = 0.$$

Remark. *Note that Theorem 1.1.8 only includes long-range dependence and antipersistence, since the case of short-range dependence with* $d = 0$ *in (1.8) would imply non-summable autocovariances, which then would (under additional regularity conditions) correspond to intermediate dependence with (at the origin) divergent spectral density.*

Based on this corollary, one can shortly summarize the conditions on the memory parameter d for the three main dependence types (i), (iii) and (iv). For more details on different models exhibiting this special dependence type of *intermediate dependence*, see, e.g., Granger and Ding (1996) [170].

Remarks. *(i)* (X_t) *exhibits long-range dependence if* $d \in \left(0, \frac{1}{2}\right)$, *which means that the spectral density* f *has a hyperbolic pole at the origin or that* (X_t) *has slowly (hyperbolically) decaying correlations.* (X_t) *is short-range dependent if* $d = 0$ *and the correlations are summable, and antipersistent if* $d \in \left(-\frac{1}{2}, 0\right)$.

(ii) The different parameter ranges can be interpreted as follows: A value $d \in \left(0, \frac{1}{2}\right)$ *means that high values of the process will be followed by high values for a long time; a value* $d \in \left(-\frac{1}{2}, 0\right)$ *indicates a tendency of eventually switching between high and low values; and for* $d = 0$ *we have a random process with neither a tendency to switching nor a tendency of maintaining the same behavior.*

For the particular case where the slowly varying function tends to a constant (see Lemma 1.1.6), the asymptotic behavior is expressed as follows:

Corollary 1.1.10. *A stationary process* (X_t) *exhibits long-range dependence if there exists some* $d \in \left(0, \frac{1}{2}\right)$ *and constants* $0 < c_f, c_\gamma < \infty$ *such that*

$$\gamma(k) \underset{k \to \infty}{\sim} c_\gamma |k|^{2d-1}, \tag{1.10}$$

or

$$f(\lambda) \underset{\lambda \to 0}{\sim} c_f |\lambda|^{-2d}, \tag{1.11}$$

where the memory parameter d quantifies the decay rate.

Instead of the memory parameter d, the **Hurst parameter** or **Hurst exponent** $H = d + \frac{1}{2}$ is frequently used. This parameter is named after Harold Edward Hurst (1880-1978) who observed fluctuations of the Nile River water levels over a long-term period (see, e.g., Hurst (1951) [216]). Using this Hurst parameter, the asymptotic behavior of the autocovariance function γ and the spectral density f (from Theorem 1.1.8) becomes

$$\gamma(k) \underset{k\to\infty}{\sim} L_\gamma(k)|k|^{2H-2}, \tag{1.12}$$

and

$$f(\lambda) \underset{\lambda\to 0}{\sim} L_f(\lambda)|\lambda|^{1-2H},$$

respectively, where L_f and L_γ are slowly varying functions. Consequently, for $H \in \left(\frac{1}{2}, 1\right)$, this process exhibits long-range dependence.

We complete this short introduction with a few notes on slowly varying functions that are, e.g., useful to investigate the asymptotic behavior of different long-range dependent quantities such as the sample mean. The following results can be found in Pipiras and Taqqu (2017) [325].

A standard result in the theory of slowly varying functions is *Karamata's Theorem*:

Proposition 1.1.11. ***(Karamata)***

Let L be slowly varying at infinity (in Karamata's sense) and let $p > -1$. Then, for

$$c_j = L(j)j^p, \quad j \geq 1, \tag{1.13}$$

we have

$$\sum_{j=1}^{n} c_j \underset{n\to\infty}{\sim} \frac{L(n)n^{p+1}}{p+1}.$$

This proposition can be directly used to prove the asymptotic behavior of the *sample mean variance*:

Corollary 1.1.12. *Let (X_t) be a stationary process with autocovariance function γ of the form (1.8) with slowly varying function L_γ, then we obtain for the sample mean $\bar{X_n} = \frac{1}{n}\sum_{t=1}^{n} X_t$:*

$$var\left(\bar{X_n}\right) \underset{n\to\infty}{\sim} \frac{L_\gamma(n)n^{2d-1}}{d(2d+1)}.$$

This follows from Karamata's Theorem (Proposition 1.1.11) with $p = 2d - 1 \in (-1, 0)$ and $L = L_\gamma$.

If we restict p in Karamata's Theorem to the interval $\left(-1, -\frac{1}{2}\right)$ we obtain another useful asymptotic result:

Proposition 1.1.13. *Let L be a slowly varying function and let $p \in \left(-1, -\frac{1}{2}\right)$. Then (1.13) implies that*

$$\sum_{j=1}^{n} c_j c_{j+k} \underset{k\to\infty}{\sim} (L(k))^2 k^{2p+1} \frac{\Gamma(p+1)\Gamma(-2p-1)}{\Gamma(-p)}$$

with Gamma function

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

As we will see in Chapter 1.2.1, this proposition can be applied to the coefficients of a *linear process* to prove the long-memory property.

1.2 Various examples of long-memory processes

There are various examples of processes that exhibit long-range dependence. All of them are characterized by the decay rate of their correlations or the asymptotic behavior of their spectral density. Since in this thesis, we focus in Part I on *linear long-memory processes* and in Part II on *Gaussian subordinated long-memory processes*, we restrict our considerations to these two classes of long-memory processes. Important representatives of the class of *linear long-memory processes* are *fractional ARIMA- (fARIMA)* processes, which we invoke in our data examples as suitable model selection candidates. Furthermore, we introduce *self-similar long-memory processes* that occur as limiting distributions and are of interest in our asymptotic results of Part I.

For definitions and results in this chapter, we mainly refer to Beran (1994) [22], Beran et al. (2013) [33], and Pipiras and Taqqu (2017) [325].

1.2.1 Linear long-memory processes

In this first section, we consider the simplest long-range dependent processes, the *linear long-memory processes*, and introduce their main representatives, the *fractional ARIMA processes*. These processes are constructed out of linear processes by specifying the decay rate of the coefficients.

One important theorem in the context of stationary processes is the *Wold-decomposition theorem*, named after Herman Wold (Wold, 1954 [418]). This theorem states that any (second-order) stationary process has a linear representation:

Theorem 1.2.1. ***(Wold decomposition)***

Any zero-mean non-deterministic stationary process (X_t) can be uniquely represented as an infinite weighted sum of an elapsed white noise process (ε_t) (which is a MA(∞) process), and a deterministic term $\mu_t \in \mathbb{R}$, i.e.,

$$X_t \stackrel{L^2}{=} \mu_t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \qquad t \in \mathbb{Z}, \tag{1.14}$$

with coefficients $a_j \in \mathbb{R}$, $j = 0, 1, \ldots$, such that

$$\sum_{j=0}^{\infty} a_j^2 < \infty. \tag{1.15}$$

Using the backshift operator B defined in (1.2), we can rewrite (1.14) as

$$X_t = \mu_t + \left(\sum_{j=0}^{\infty} a_j B^j \right) \varepsilon_t = \mu_t + A(B)\varepsilon_t, \qquad t \in \mathbb{Z},$$

with linear filter $A(B) = \sum_{j=0}^{\infty} a_j B^j$.
If $\mu_t = 0$ the process is called ***purely non-deterministic****.*

This *Wold representation* is strongly related to the general notion of *linear processes*.

Definition 1.2.2. *(i) A process*

$$X_t = \mu + \sum_{j=-\infty}^{\infty} a_j \varepsilon_{t-j}, \qquad t \in \mathbb{Z},$$

with (ε_t) iid, is called a ***linear process****.*

(ii) By taking into account only past values, i.e., $a_j = 0$ for $j < 0$, we obtain a ***causal linear process***

$$X_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \qquad t \in \mathbb{Z}.$$

Remarks. *(i) The class of causal purely non-deterministic linear processes with finite second moments, i.e., of processes of the form*

$$X_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \qquad t \in \mathbb{Z}, \tag{1.16}$$

coincides with the class of ***weakly stationary linear processes****.*

(ii) Note that sometimes, the notion of linear process refer to white noise innovations (ε_t) instead of iid innovations. One can conclude that any purely non-deterministic stationary process is either a linear process with iid innovations or can be represented as linear combination with white noise innovations according to Wold's decomposition theorem.

(iii) Equation (1.15) together with $\sigma_\varepsilon^2 = var(\varepsilon_t) < \infty$ ensures that (X_t) has finite second moments.

(iv) A linear process (X_t) with finite second moments converges in the L^2-sense.

(v) A causal linear process is a process that is independent of the future.

We introduce the long-range dependence property for *weakly stationary linear processes* by determining the asymptotic behavior of the coefficients a_j, $j = 0, 1, \ldots$.

Corollary 1.2.3. *Let (X_t) be a weakly stationary linear process of the form (1.16) and let L_a be a function slowly varying at infinity (in Zygmund's sense). Then (X_t) is*

- *long-range dependent if, for $d \in \left(0, \frac{1}{2}\right)$, the coefficients a_j are of the form*

$$a_j = L_a(j) j^{d-1} \tag{1.17}$$

- *short-range dependent if*

$$\sum_{j=0}^{\infty} |a_j| < \infty \quad \text{and} \quad \sum_{j=0}^{\infty} a_j \neq 0,$$

- *antipersistent if, for $d \in \left(-\frac{1}{2}, 0\right)$, the coefficients a_j are of the form*

$$a_j = L_a(j) j^{d-1} \quad \text{and} \quad \sum_{j=0}^{\infty} a_j = 0.$$

Note that under long-range dependence, condition (1.17) implies that the coefficients are non-summable, i.e., $\sum_{j=0}^{\infty} a_j = \infty$.
As a consequence, we obtain the following corollary:

Corollary 1.2.4. *Let (X_t) be a weakly stationary linear process with autocovariance function γ and spectral density f, and let L_a, L_γ and L_f be slowly varying functions in Zygmund's sense. If, for $d \in \left(0, \frac{1}{2}\right)$, the coefficients satisfy (1.17), then*

(i) the autocovariances are of the form

$$\gamma(k) = L_\gamma(k) |k|^{2d-1}, \qquad \text{as } k \to \infty,$$

with

$$L_\gamma(k) \underset{k\to\infty}{\sim} \sigma_\varepsilon^2 \left(L_a(k)\right)^2 \frac{\Gamma(d)\Gamma(1-2d)}{\Gamma(1-d)},$$

(ii) and the spectral density is given by

$$f(\lambda) = L_f(\lambda)|\lambda|^{-2d}, \qquad as\ \lambda \to 0,$$

with

$$L_f(\lambda) \underset{\lambda\to 0}{\sim} \frac{\sigma_\varepsilon^2}{2\pi} \left(L_a(\lambda^{-1})\right)^2 \Gamma(d)^2.$$

Let's look at an example:

Example 1.2.5. *Let (X_t) be a weakly stationary linear process of the form (1.16) and let $\sigma_\varepsilon^2 = var(\varepsilon_t)$.*
We assume that, for $d \in \left(0, \frac{1}{2}\right)$, the coefficients a_j, $j = 0, 1, \ldots$, coincide with the coefficients in the power series representation

$$A(z) = (1-z)^{-d} = \sum_{j=0}^{\infty} a_j z^j.$$

Then, due to the properties of the Gamma function (and using Stirling's formula), the coefficients are of the form

$$\begin{aligned} a_j = (-1)^j \binom{-d}{j} &= (-1)^j \frac{\Gamma(1-d)}{\Gamma(j+1)\Gamma(1-d-j)} \\ &\underset{j\to\infty}{\sim} \frac{1}{\Gamma(d)} j^{d-1}. \end{aligned}$$

This means, the coefficients satisfy condition (1.17), and thus, according to Corollary 1.2.3, the process (X_t) exhibits long-memory.
Moreover, by applying Proposition 1.1.13 with $p = d-1$ and, since $\Gamma(d)\cdot\Gamma(d+1) = \frac{\pi}{\sin(\pi d)}$, we obtain statement (i) from Corollary 1.2.4, i.e., the autocovariances are asymptotically of the form

$$\gamma(k) = \sigma_\varepsilon^2 \sum_{j=0}^{\infty} a_j a_{j+k} \underset{k\to\infty}{\sim} L_\gamma(k)|k|^{2d-1}$$

with

$$\begin{aligned} L_\gamma(k) &= \sigma_\varepsilon^2 \frac{1}{\Gamma(d)^2} \frac{\Gamma(d)\Gamma(1-2d)}{\Gamma(1-d)} \\ &= \frac{\sigma_\varepsilon^2}{\pi} \Gamma(1-2d)\sin(\pi d). \end{aligned}$$

From the same Corollary 1.2.4, we further obtain the asymptotic behavior of the spectral density as

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi}\left|\sum_{j=0}^{\infty} a_j e^{-ij\lambda}\right|^2 \underset{\lambda\to 0}{\sim} L_f(\lambda)|\lambda|^{-2d}$$

with

$$\begin{aligned} L_f(\lambda) &= \frac{\sigma_\varepsilon^2}{2\pi}\frac{1}{(\Gamma(d))^2}\Gamma(d)^2 \\ &= \frac{\sigma_\varepsilon^2}{\pi}. \end{aligned}$$

Note that this is the spectral density of a stationary fractional ARIMA(0,d,0) process (see Definition 1.2.6).

The so-called *fractional ARIMA processes* are among the best known representatives of processes that exhibit long-memory behavior. These processes belong to a useful class of linear processes that may exhibit either of the three above mentioned dependence structures. They were first introduced by Granger and Joyeux (1980) [173] and Hosking (1981) [214] as an extension of both the *autoregressive moving average (ARMA) processes* and the *integrated ARMA (ARIMA) processes.*

Definition 1.2.6. *Let (X_t) be a stationary solution of*

$$\varphi(B)(1-B)^d X_t = \psi(B)\varepsilon_t, \tag{1.18}$$

where B is the backshift operator. φ, ψ are polynomials of order p, q corresponding to the AR-polynomial and the MA-polynomial defined in (1.5) and (1.3), respectively, such that they have no common roots.

If $d \in \mathbb{R}$ is the minimal value such that (X_t) satisfies (1.18), then (X_t) is called

(i) a ***ARMA(p,q) process*** *if $d = 0$,*

(ii) a ***ARIMA(p,d,q) process*** *if $d \in \mathbb{N}_0$,*

(iii) a ***fractional ARIMA(p,d,q) process*** *if $d \in \left(-\frac{1}{2}, \frac{1}{2}\right)$.*

By restricting d to $\left(-\frac{1}{2}, \frac{1}{2}\right)$ in (iii), we obtain a class of stationary processes that is extensively discussed in literature due to its special properties of the spectral density.

Note that in case of fractional d, the binomial series expansion $(1-B)^d$ uses the generalized

binomial coefficient, which can be represented by the Gamma function, i.e.,

$$\binom{d}{k} = \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)},$$

resulting in

$$(1-B)^d = \sum_{k=0}^{\infty} (-1)^k B^k \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}.$$

Remarks. *(i) The parameter d is accountable for the long-term behavior and the parameters p, q account for the short-range properties of the process.*

(ii) An ARIMA(p,0,q) process is a ARMA(p,q) process, which always is short-range dependent.

(iii) An ARIMA(0,0,0) process corresponds to Gaussian white noise defined in Example 1.1.3 (i).

(iv) The random walk, given in Example 1.1.3 (ii) as a well-known non-stationary process, is simply an ARIMA(0,1,0) process.

(v) A general dth-difference stationary process is a fractional $ARIMA(0, d, 0)$ process; for $d = 1$ we have the random walk.

Remark. *To understand why the restriction on d to the interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$ is reasonable in studying stationary processes, note first that for any real value $\tilde{d} \geq \frac{1}{2}$, the process (X_t) is non-stationary. Secondly, note that for any $|\tilde{d}| \geq \frac{1}{2}$, one reduces the situation to $-\frac{1}{2} \leq d < \frac{1}{2}$ by splitting off the integer part of $\tilde{d}$ and taking appropriate differences.*

1.2.2 Gaussian subordinated long-memory processes

Another class of long-memory processes can be generated via *Gaussian subordination.* This class of so-called *Gaussian subordinated long-memory processes* is considered in various papers, see, e.g., Rosenblatt (1961, 1979) [344], [345], Taqqu (1975, 1979), [382], [385], Dobrushin and Major (1979) [119], and Dobrushin (1980) [118]. As we will see in Part II, this class is later extended to the circular state space by Beran and Ghosh (2020) [34].

First, we review main notions and properties in the context of Gaussian subordination such as the *Hermite polynomial representation.* More information on this are summarized in Beran et al. (2013) [33].

Definition 1.2.7. *Let $Z \sim N(0,1)$ and let $G : \mathbb{R} \to \mathbb{R}$ be an arbitrary measurable function, then the random variable $G(Z)$ is called* ***Gaussian subordinated*** *if it satisfies the following*

conditions:

$$E\left[G(Z)\right] = 0 \quad \textit{and} \quad E\left[G^2(Z)\right] < \infty.$$

Then G is in

$$L^2(\mathbb{R}, \varphi) := \left\{ G : \mathbb{R} \to \mathbb{R},\ ||G||^2 = \int G^2(x)\varphi(x)dx < \infty \right\}$$

equipped with the scalar product

$$\langle G, H \rangle = \int_{-\infty}^{\infty} G(z)H(z)\varphi(z)dz,$$

where φ denotes the standard normal density function and H are the so-called **Hermite polynomials**.

Accordingly, a **Gaussian subordinated process** is a process (X_t) such that any X_t is a Gaussian subordinated random variable.

Moreover, the set of *Hermite polynomials*, which satisfy the so-called *Hermite differential equation*

$$f''(x) - xf'(x) + jf(x) = 0, \qquad j = 0, 1, \ldots,$$

form a complete orthogonal basis of the Hilbert space $L^2(\mathbb{R}, \varphi)$, and for the scalar product holds:

$$\langle H_m, H_n \rangle = \int_{\mathbb{R}} H_m(z)H_n(z)\varphi(z)dz = \delta_{mn} \cdot m! \tag{1.19}$$

with *Kronecker delta* δ_{mn}.

A representation of the Hermite polynomials is given as follows:

Definition 1.2.8. *The **jth Hermite polynomial** $H_j(x)$, $j \in \mathbb{N}$, can be expressed as*

$$H_j(x) = (-1)^j \exp\left(\frac{x^2}{2}\right) \frac{d^j}{dx^j} \exp\left(\frac{-x^2}{2}\right),$$

i.e.,

$$H_0(x) = 1, \quad H_1(x) = x, \quad H_2(x) = x^2 - 1, \ldots$$

Using these polynomials, we obtain the unique and orthogonal L^2-representation, the so-called **Hermite polynomial expansion**, of a function $G \in L^2(\mathbb{R}, \varphi)$.

Lemma 1.2.9. *A Gaussian subordinated random variable $G(Z)$ can be represented as*

$$\begin{aligned} G(Z) &= \sum_{k=0}^{\infty} \frac{\langle G, H_k \rangle}{||H_k||^2} \cdot H_k(Z) \\ &= E\left[G(Z)\right] + \sum_{k=m}^{\infty} \frac{J(k)}{k!} H_k(Z) \end{aligned}$$

with ***Hermite coefficients***

$$\frac{\langle G, H_k \rangle}{||H_k||^2} = \frac{J(k)}{||H_k||^2} = \frac{J(k)}{\langle H_k, H_k \rangle} =: \frac{J(k)}{k!}, \quad k = 1, 2, \dots \ .$$

Here, $m \geq 1$ is the so-called ***Hermite rank*** *of G defined as the smallest integer in the sum such that $J(k) \neq 0$, i.e., $m := \min\{k : J(k) \neq 0\}$.*

Note that in some references the notion of *Hermite coefficients* refers to just the scalar product

$$J(k) = \langle G, H_k \rangle = E\left[G(Z) H_k(Z)\right].$$

In the following, we recall some important results concerning the covariance of Gaussian subordinated random variables:

Lemma 1.2.10. *Let $Z_1, Z_2 \sim N(0,1)$ be two jointly distributed random variables with $\gamma_Z = cov(Z_1, Z_2)$. Then*

$$cov(H_k(Z_1), H_l(Z_2)) = \delta_{kl} \cdot k! \cdot \gamma_Z^k$$

with Kronecker delta δ_{kl}.

In the special case of iid standard normal random variables Z_1, Z_2, we obtain the scalar product given in (1.19).

Moreover, if $G(Z_1), G(Z_2)$ are the corresponding Gaussian subordinated random variables. Then

$$cov(G(Z_1), G(Z_2)) = \sum_{k=1}^{\infty} \frac{J^2(k)}{k!} \cdot \gamma_Z^k. \tag{1.20}$$

The second part of this lemma follows directly from the orthogonality of the Hermite polynomials. Hence, the covariance of Gaussian subordinated random variables is decomposed into (orthogonal) contributions of the coefficients $J(k)$. Applying this result to a Gaussian subordinated long-memory process, we obtain an (asymptotic) representation of its autocovariance function.

Lemma 1.2.11. *Let (Z_t) be a stationary long-range dependent Gaussian process with zero mean, unit variance, autocovariance function γ_Z and spectral density f_Z, and let G be such that $(G(Z_t))$*

is a Gaussian subordinated process. Then the autocovariances of $G(Z_t)$ asymptotically behave as

$$\begin{aligned}\gamma_{G(Z)}(k) &= \sum_{l=m}^{\infty} \frac{J^2(l)}{l!}\gamma_Z^l(k) \\ &\underset{k\to\infty}{\sim} \frac{J^2(m)}{m!}L_{\gamma_Z}^m(k)|k|^{m(2d-1)},\end{aligned} \tag{1.21}$$

where L_{γ_Z} is slowly varying (in Zygmund's sense) at infinity associated with γ_Z.

This means that, as k tends to infinity, the asymptotic behavior of $\gamma_{G(Z)}$ is determined by the leading first non-zero term of the sum.

This directly follows from Lemma 1.2.10 and the asymptotic behavior of the autocovariances of the Gaussian long-memory process (Z_t).

As a consequence, we obtain the following results:

Corollary 1.2.12. *Let $m \geq 1$ and $d \in \left(0, \frac{1}{2}\right)$, and let*

$$d_m := \frac{1}{2}\left(1 - m(2d-1)\right). \tag{1.22}$$

If $m(1-2d) < 1$ or $d > \frac{1}{2}\left(1 - \frac{1}{m}\right)$, respectively, then the subordinated process (X_t) exhibits long memory with parameter d_m.

Moreover, if (X_t) has Hermite rank $m = 1$, the memory parameter d_m coincides with the memory parameter d of the Gaussian long-memory process (Z_t).

Remark. *The statements hold, since the asymptotic behavior of the subordinated process $(G(Z_t))$ in general reads*

$$\gamma_{G(Z)}(k) \underset{k\to\infty}{\sim} \frac{J^2(m)}{m!}L_{\gamma_Z}^m(k)|k|^{2d_m-1},$$

and consequently, with (1.21), m and d satisfy on the one hand $m(2d-1) > -1$, and on the other hand $2d_m - 1 = m(2d-1)$.

Remark 1.2.13. *In statistical inference, the question arises how to estimate the Hermite rank m. For Gaussian subordinated processes, Hermite rank $m = 1$ is a reasonable choice, since any monotonously increasing function F has Hermite rank 1 and any marginal distribution function G can be obtained by a monotone increasing transformation of the Gaussian distribution, e.g., by $G(z) := F^{-1}(\Phi(z))$, where Φ is the standard normal distribution function. In general, higher Hermite ranks lead to limiting distributions that are no longer normal. Hermite-Rosenblatt processes are one example of such non-normal limiting distributions.*

Further explanations are, for instance, given in Menéndez et al. (2013) [297].

Furthermore, although, to the best of our knowledge, no method on estimation of m is discussed in literature, there exist papers, e.g., Beran et al. (2016) [35] and Tewes (2018) [390], proposing hypotheses tests for the Hermite rank to be one.

We complete this short introduction with a remark, which we will need in Section 2.2.2.

Remark 1.2.14. *For dealing with processes subordinated to not necessarily Gaussian linear processes (Z_t) other approaches are needed. One approach requires expansions based on* ***Appell polynomials*** *$A_k(z)$, $k \geq 0$, which are defined by their generating function*

$$\sum_{k=0}^{\infty} \frac{x^k A_k(z)}{k!} = \frac{e^{xz}}{E\left[e^{xZ_t}\right]}.$$

Appell polynomials are distribution specific and consequently not necessarily orthogonal with respect to the associated L^2-norm. They are a generalization of Hermite polynomials (where the associated distribution is a standard normal distribution).

For the definition of this class of polynomials, see, e.g., Appell (1880) [8] and for more details, see e.g. Beran et al. (2013) [33]. Statistical inference of such classes of processes and their relation to linear processes is, e.g., discussed in Surgailis (1981, 1982, 2000) [373] [374] [375], Giraitis (1985) [155], Giraitis and Surgailis (1986, 1989) [162] [163], Avram and Taqqu (1987) [11], Surgailis and Vaičiulis (1999) [377] and Schützner (2009) [351] . For a review of Appell polynomials see Surgailis (2003) [376] and Schützner (2006) [350].

Another approach for this is the martingale difference decomposition, which is considered in Ho and Hsing (1996, 1997) [206] [207], Giraitis and Surgailis (1999) [165], Wu (2003) [419] in the context of limit theorems of empirical processes that originate from linear processes. See also Hsing (2000) [215] for a review on this approach.

1.2.3 Self-similar processes

Another class of stochastic processes that exhibit a certain type of long-range dependence are the so-called *self-similar processes*. They were first introduced by Kolmogorov (1941) [248] in a theoretical context, and later Mandelbrot (1965) [269] and Mandelbrot and Wallis (1968, 1969c-e) [279], [282], [283], [284] introduced them into statistics.

Although, we will define our long-memory processes as either linear long-memory processes (in Part I) or Gaussian subordinated long-memory processes (in Part II), we will review this special class of processes as they play an important role as limiting processes or limiting distributions.

For definitions and properties we refer to Beran (1994) [22] and Beran et al. (2013) [33].

Definition 1.2.15. *Let $(X_t)_{t\in\mathbb{R}}$ be a continuous-time stochastic process. (X_t) is called **self-similar** with **self-similarity parameter H** , if for any positive stretching factor c the rescaled process $(c^{-H}X_{ct})$ with time scale ct is equal in distribution to the original process (X_t), i.e., if for any finite sample $t_1, \ldots, t_k$ and any positive constant c,*

$$c^{-H}(X_{ct_1}, \ldots, X_{ct_k}) \stackrel{d}{=} (X_{t_1}, \ldots, X_{t_k}),$$

where $\stackrel{d}{=}$ denotes equality in distribution.

This process is invariant in distribution under suitable scaling of time and space, and therefore useful for modeling space-time scaling of random phenomena.

We restrict our considerations to self-similar processes whose increments are stationary. These special self-similar processes link the long-memory property to the self-similarity parameter via the autocovariance function in the following way:

Lemma 1.2.16. *Let (X_t) be a self-similar process with zero mean, variance σ_X^2, and with stationary increments $\xi_t = X_t - X_{t-1}$. Then the autocovariances of (X_t) are of the form*

$$\gamma_X(t,s) = \frac{\sigma_X^2}{2}\left(|t|^{2H} - |t-s|^{2H} + |s|^{2H}\right)$$

with self-similar parameter $H = d + \frac{1}{2}$.

The analogue applies for the stationary increment process (ξ_t):

$$\gamma_\xi(k) = \frac{\sigma_\xi^2}{2}\left(|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right).$$

Moreover, for $H \in \left(\frac{1}{2}, 1\right)$,

$$\gamma(k) \underset{k\to\infty}{\sim} c_\gamma |k|^{2H-2},$$

where $c_\gamma = \sigma_\xi^2 H(2H-1)$. According to (1.12), this process exhibits long-range dependence.

Thus, in this context, the self-similarity parameter H coincides with the Hurst parameter introduced in Chapter 1.1.

A well-known representative of self-similar processes is the *fractional Brownian motion* which is a generalization of the *Brownian Motion* and occurs in many limit theorems as a candidate for limiting distributions. For a detailed introduction to these processes, see, e.g., Pipiras and Taqqu (2000b, 2003) [323], [324], Embrechts and Maejima (2002) [128], Bender et al. (2007) [19], Mishura (2008) [299] and Mandelbrot (2013b) [275] . A bibliographical guide in this context is given by Taqqu (1986) [386].

In the following, we recall some definitions and properties of processes that are related to (fractional) Brownian Motion.

Definition 1.2.17. *A stochastic process (ξ_t) is called a* ***Gaussian process*** *if its finite-dimensional distributions are Gaussian, i.e., if for any $n \geq 1$ and any $t_1, \ldots, t_n$ the random vector $(\xi_{t_1}, \ldots, \xi_{t_n})$ has a multivariate Gaussian distribution.*

A special Gaussian process, the *Brownian Motion* (or *Wiener process*), named after Robert Brown (1773-1858) who used it to model the random motion of pollen particles in a fluid in 1827, was first described independently by Albert Einstein in 1905 [127] and Marian von Smoluchowski in 1906 [402]. For ease of notation, in this section, we denote processes (X_t) as $(X(t))$.

Definition 1.2.18. *Let $(B(t))$ be a Gaussian process that satisfies the following properties:*

(i) $B(0) = 0$ *almost surely,*

(ii) $B(t)$ *has independent Gaussian increments with zero mean such that*

$$B(t) - B(s) \sim N(0, \sigma^2|t - s|),$$

(iii) $B(t)$ *is almost surely continuous in* t.

Then $(B(t))$ is called ***Brownian Motion*** *$(B(t))$.*

An expository paper on Brownian Motion is given by Csörgő (1979) [83]. We summarize some properties of a Brownian Motion:

Lemma 1.2.19. *(i) A Brownian Motion $(B(t))_{t\geq 0}$ has the following properties:*

(a) $E[B(t)] = 0$ *for all* $t \in \mathbb{R}$

(b) $var(B(t)) = t$ *and* $cov(B(s), B(t)) = \min\{s, t\}$

(c) $B(t) = B(t) - B(0) \sim N(0, t)$

(ii) If $(B(t))_{t\geq 0}$ is a Brownian Motion, then for any $t_1 < \ldots < t_n$ the random vector $(B(t_1), \ldots, B(t_n))$ has a multivariate normal distribution with zero mean and $cov(B(s), B(t)) = \min\{s, t\}$.

(iii) If $(B(t))_{t\geq 0}$ is a Brownian Motion, then $(-B(t))_{t\geq 0}$ is also a Brownian Motion.

A generalization of the Brownian Motion is the *fractional Brownian Motion*, introduced by Mandelbrot and Van Ness in 1968 [278]. As the simplest example for a self-similar process that exhibits long-range dependence, it is characterized by a special form of the autocovariance function.

Definition 1.2.20. *For any Hurst parameter $H \in (0,1)$, the **fractional Brownian Motion** $(B_H(t))_{t\in[0,1]}$ is a continuous-time zero-mean Gaussian process restricted to $[0,1]$ and starting at zero, i.e., $B_H(0) = 0$, with autocovariance function*

$$cov(B_H(t), B_H(s)) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right).$$

Here, the Hurst parameter H describes the smoothness of the resultant motion: the larger H the smoother the motion.

Fractional Brownian Motion captures all three dependence structures:

Lemma 1.2.21. *For $H > \frac{1}{2}$ the fractional Brownian Motion exhibits long-range dependence; for $H < \frac{1}{2}$ the process is antipersistent; and for $H = \frac{1}{2}$ we obtain the Brownian Motion, a short-range dependent process.*

Remark. *Note that, in discrete time, the stationary increment process of discrete-time fractional Brownian Motion is the so-called discrete-time fractional Gaussian noise. According to Lemma 1.2.21, this holds, in particular, for discrete-time Brownian Motion and discrete-time Gaussian noise.*

Another process that is related to Brownian Motion is a *Brownian Bridge*.

Definition 1.2.22. *A **Brownian Bridge** $(B^0(t))_{t\in[0,1]}$ is a Gaussian process with the following properties:*

(i) $B^0(0) = B^0(1) = 0$,

(ii) $E[B^0(t)] = 0$ *for all* $t \in [0,1]$,

(iii) $cov\left(B^0(t), B^0(s)\right) = \min\{t,s\} - ts$ *for* $t, s \in [0,1]$,

(iv) $B^0(t)$ *is almost surely continuous in* $t \in [0,1]$.

In other words: A *Brownian Bridge* is a standard Brownian Motion $(B(t))_{t\in[0,1]}$ pinned to zero at *both ends* of the interval, i.e., $B(0) = 0 = B(1)$.

The simplest way to construct a Brownian Bridge from a Brownian Motion, and vice versa, is the following:

Lemma 1.2.23. *Let $(B(t))_{t\in[0,1]}$ be a Brownian Motion, and let*

$$B^0(t) = B(t) - tB(1), \quad t \in [0,1].$$

Then $(B^0(t))_{t\in[0,1]}$ is a Brownian Bridge.

Conversely, a standard Brownian Motion $(B(t))_{t\in[0,1]}$ *can be constructed from a Brownian Bridge* $(B^0(t))_{t\in[0,1]}$ *by*

$$B(t) = B^0(t) - tZ, \quad t \in [0,1],$$

where $Z \sim N(0,1)$ *is independent of* $(B^0(t))$.

Analogously as in Definition 1.2.22, we define the **fractional Brownian Bridge** $(B_H^0(t))_{t\in[0,1]}$ with Hurst parameter $H \in (0,1)$ as a fractional Brownian Motion $(B_H(t))_{t\in[0,1]}$ pinned to zero at both ends. Moreover, we can construct a *fractional Brownian Bridge* by replacing $B(t)$ in Lemma 1.2.23 by a fractional Brownian Motion $B_H(t)$.

Finally, for asymptotic considerations, we need to introduce another representative of the class of self-similar processes with long-memory: the *Hermite-Rosenblatt processes*. These processes are defined using *multiple Wiener-Itô integrals* (see, e.g., Ito (1951) [220] and Major (1981) [268]) and extend the class of Gaussian self-similar processes to the class of non-Gaussian self-similar processes. A representation of self-similar processes in terms of *multiple Wiener-Itô integral* is, e.g., given in Taqqu (1978) [384]. As we will see in Chapter 2, these processes play an important role as limiting processes.

Definition 1.2.24. *Let* $m \geq 1$ *and* $H \in (0,1)$. *Then*

$$Z_{m,H}(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} Q_{m,t}(x_1, \dots, x_m; H) dB(x_1) \dots dB(x_m) \tag{1.23}$$

with kernel $Q_{m,t}$ *of the form*

$$Q_{m,t}(x_1, \dots, x_m; H) = \int_0^t \left(\prod_{i=1}^m (s - x_i)_+^{-\left(\frac{1}{2} + (1-H)/m\right)} \right) ds,$$

where $(a)_+$ *denotes the positive part of* a, *is called* ***Hermite process*** *or* ***Hermite-Rosenblatt process*** *on* $\mathbb{R}$. *The notion* ***Rosenblatt process*** *originally refers to the case* $m = 2$ *only (see, e.g., Taqqu 1975 [382]).*

Here, the kernels Q guarantee that the process is self-similar with stationary increments.

Remark 1.2.25. *Note that*

(i) for $m = 1$ *in (1.23), we obtain standard fractional Brownian Motion.*

(ii) for $m > 1$ *in (1.23), this process is no longer Gaussian.*

Chapter 2

Limit theorems

One main topic in statistical inference addresses large sample and asymptotic behavior of different statistics. Our main statements in this thesis concern the asymptotic distributions of the test statistic in Part I and of the kernel estimator of the mean direction in Part II. In the iid case, *central limit theorems (CLTs)* are of interest. In the context of long-memory processes, limit theorems involve *limits of multiple Wiener-Itô integrals*, in particular Hermite-Rosenblatt processes, which are defined in Definition 1.2.24. These limiting processes play an important role in *weak reduction principles* from Taqqu (1975) [382] and Dehling and Taqqu (1989) [106] on which our main statements are based. These principles deal with weak convergence of empirical processes that originate either from Gaussian subordinated or from linear processes and use knowledge about weak limits of partial sums of those processes. The case of Gaussian subordinated processes is based on Hermite polynomial expansion (see Chapter 1.2.2) and has been studied by Dehling and Taqqu (1989) [106]. For linear processes, Appell polynomials (see Remark 1.2.14) or martingale difference expansions are the crucial tools. Weak convergence of the respective empirical processes is proven by Ho and Hsing (1996) [206], Giraitis et al. (1996a) [157], and Giraitis and Surgailis (1999) [165].

In the context of empirical processes, we study weak convergence in the so-called *Skorohod space*. Hence, the first chapter recalls the general definition of weak convergence and introduces the Skorohod spaces. In the second chapter, we state the *uniform weak reduction principle* and *functional limit theorems (FCLTs)* of empirical processes that either originate from Gaussian subordinated or from linear processes. These results can be found summarized in Beran et al. (2013) [33], and in more detail in the references mentioned above.

2.1 Weak convergence

We are interested in weak convergence of random elements on a *polish space* $\mathcal{S}$, i.e., on a separable and complete metric space, equipped with the Borel-σ-algebra $\mathcal{B}$. Regarding the notion of weak convergence, we refer to Billingsley (1999, 2013) [53], [54].

Definition 2.1.1. *(i) A sequence (μ_n) of probability measures on $(\mathcal{S}, \mathcal{B})$* ***converges weakly*** *to a probability measure μ on $(\mathcal{S}, \mathcal{B})$, denoted by*

$$\mu_n \xrightarrow{w,\mathcal{S}} \mu,$$

if and only if

$$\lim_{n\to\infty} \int f d\mu_n \to \int f d\mu \qquad \text{for all } f \in \mathcal{C}_b(\mathcal{S}),$$

where $\mathcal{C}_b(\mathcal{S})$ is the space of bounded continuous functions on $\mathcal{S}$.

(ii) A sequence (X_n) of random variables with distributions (P_n) on $\mathcal{S}$ ***converges weakly*** *to a random variable X with distribution P, denoted by*

$$X_n \xrightarrow{w,\mathcal{S}} X,$$

or equivalently (X_n) ***converges in distribution*** *to X, denoted by*

$$X_n \xrightarrow{d} X,$$

if and only if $P_n \xrightarrow{w,\mathcal{S}} P$.

In this case, weak convergence can be expressed as

$$\lim_{n\to\infty} E\left[f(X_n)\right] = E\left[f(X)\right], \qquad \text{for all } f \in \mathcal{C}_b(\mathcal{S}). \tag{2.1}$$

For an overview of weak convergence of empirical processes see, e.g., the textbook of Wellner and van der Vaart (1996) [396].

In order to study weak convergence of empirical processes, we need to define the space $D(I)$, which is the **space of cadlag (continue à droîte, limite à gauche) functions** on some closed interval $I \subseteq \bar{\mathbb{R}}$, i.e., of functions on I that are right-continuous with left-hand limits. In case of $I = \bar{\mathbb{R}}$ the functions are right-continuous on $\{-\infty\} \cup \mathbb{R}$ and have left-hand limits on $\mathbb{R} \cup \{+\infty\}$. In other words, $D(I)$ is the space of real-valued functions on I that have *discontinuities of the first kind*. Hence, it serves to model processes with jumps, for instance, empirical distribution functions, martingales, Poisson or Lévy processes. The space $D(I)$ equipped with the so-called

Skorohod J_1 topology, see, e.g., Skorokhod (1956) [369] and Billingsley (1999) [53], is also called **Skorohod space**. And with the **Skorohod metric**, it becomes a separable and metric space. When restricted to the space of bounded continuous functions, this topology is equivalent to the *uniform topology*. Introducing an equivalent metric makes $D(I)$ even a polish space (see Billingsley, 1968 [51]).

Remark. *Note that proving weak convergence on $D(I)$ involves proving (i) weak convergence of the respective finite-dimensional distributions and (ii) tightness of the set of probability measures. Further, on polish spaces, one makes use of the fact that tightness is equivalent to relative compactness (Theorem of Prokhorov [330]). See, e.g., Giraitis and Surgailis (1999) [165] for comments on this.*

In this thesis, we consider weak convergence in three different Skorohod spaces, in $D([0,1])$, and in the generalizations $D([0,1]\times\mathbb{R})$ and $D([0,1]\times\mathbb{R}^k)$, $k \geq 1$, respectively. The latter two spaces will be needed in the context of uniform weak convergence of partial empirical processes as the main issue in functional limit theorems.

For generalizations of the Skorohod space $D([0,1])$ see, e.g., Bickel and Wichura (1971) [49], Billingsley (1999, 2013) [53], [54], and references therein.

2.2 The uniform weak reduction principle

This second chapter focuses on the *uniform weak reduction principle* from Dehling and Taqqu (1989) [106] for Gaussian subordinated processes and the extension to linear processes (independently) proven by Ho and Hsing (1996) [206] and Giraitis et al. (1996a) [157]. Since the reduction principle *reduces* the limiting behavior of sums of Gaussian subordinated processes to the limiting behavior of sums of Hermite polynomials, we also recall weak limits of various partial sums.

Weak limits of partial sums of general Gaussian subordinated processes were first studied by Rosenblatt (1961) [344] and later independently developed by Taqqu (1975, 1977, 1979) [382], [383], [385], Dobrushin and Major (1979) [119], and Dobrushin (1980) [118]. Further results are given in Breuer and Major (1983) [64], Giraitis and Surgailis (1985) [161], Ho and Sun (1987, 1990) [208], [209], Dehling and Taqqu (1988, 1989) [105], [106], and Arcones (1994) [9]. Weak convergence of partial sums of general long-memory processes is considered, e.g., in Davydov (1970) [102], Gorodetskii (1978) [168], Lang and Soulier (2000) [253], and Wang et al. (2003) [409].

First, we will recall two important limit theorems in the context of empirical distribution functions and empirical processes. Let (Z_t) be a stationary Gaussian process with marginal distri-

bution function $F_Z = P(Z_t \leq z)$ and let, for $Z_1, \ldots, Z_n$,

$$F_{n,Z}(z) = \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{Z_t \leq z\}$$

be the **empirical distribution function**. The *Glivenko-Cantelli Theorem* (Glivenko, 1933 [167], Cantelli, 1933 [73]) states that in case of iid random variables Z_t the empirical distribution function $F_{n,Z}$ is a *uniformly consistent estimator* of F_Z, i.e.,

$$\sup_{z \in \mathbb{R}} |F_{n,Z}(z) - F_Z(z)| \xrightarrow{a.s.} 0,$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence.

We define

$$\zeta_{n,Z}(z) := F_{n,Z}(z) - F_Z(z) \tag{2.2}$$

as the **empirical process** and

$$\zeta_{[n\tau],Z}(z) := F_{[n\tau],Z}(z) - F_Z(z), \qquad \tau \in [0,1], \tag{2.3}$$

as the **partial empirical process**. *Donsker's invariance principle* (Donsker, 1951 [120]) asserts that for iid random variables (Z_t) with distribution function F_Z the empirical process $\zeta_{n,Z}$ scaled by $\sqrt{n}$ converges weakly in $D(\mathbb{R})$ (equipped with the sup-topology) to a Gaussian process W with stationary increments, zero mean and

$$\operatorname{cov}(W(s), W(t)) = \min\{F_Z(s), F_Z(t)\} - F_Z(s)F_Z(t).$$

Remark. *For uniformly distributed Z_t, $t \in [0,1]$, where $F_Z(t) = t$, the limiting Gaussian process W becomes the Brownian Bridge defined in Definition 1.2.22.*

This result has been extended to weakly dependent random variables under certain mixing conditions yielding the same convergence rate. For a discussion in the case of Gaussian subordinated and/or linear processes, see, e.g., Billingsley (1969) [52], Berkes and Philipp (1977) [44], Doukhan et al. (1995) [122], Csörgő and Mielniczuk (1996) [92], Shao and Yu (1996) [357], and Doukhan and Surgailis (1998) [121].

In the case of Gaussian long-memory processes, the limiting behavior of empirical processes changes considerably.

Lemma 2.2.1. *Let (Z_t) be a Gaussian long-memory process with marginal density function p_Z and let $\zeta_{n,Z}$ be the respective empirical process. Then*

$$n^{\frac{1}{2}-d}\zeta_{n,Z}(z) \xrightarrow{w,D(\mathbb{R})} W(z), \qquad d \in \left(0, \frac{1}{2}\right),$$

where $W(z) := p_Z(z) \cdot c \cdot W$, $W \sim N(0,1)$, and $c > 0$ is a constant.

This means, compared to the short-memory case, the convergence rate is much slower ($n^{\frac{1}{2}-d}$ with $0 < d < \frac{1}{2}$ instead of $n^{-\frac{1}{2}}$) and the limiting process $W(z)$ is Gaussian but **degenerate** (or **perfectly correlated/totally dependent**).

Such limit theorems are, e.g., studied by Rosenblatt (1961), Dobrushin and Major (1979) and Taqqu (1979). This convergence still holds for linear long-memory processes. However, for Gaussian subordinated sequences, the limiting process is no longer necessarily Gaussian. Limit theorems of empirical processes generated by either Gaussian subordinated or linear long-memory processes are discussed by various authors, for instance by Dehling and Taqqu (1989) [106], Giraitis et al. (1996) [157], Ho and Hsing (1996) [206], Giraitis and Surgailis (1999) [165], Wu (2003) [419], Csörgő et al. (2006) [96], Csörgő and Kulik (2008a, 2008b) [87], [88], and Beran and Telkmann (2018) [41].

In the subsequent two sections, we review important limit theorems in the context of long-memory processes such as the *weak reduction principle*, the *uniform weak reduction principle* and *functional central limit theorems (FCLTs)* for both, Gaussian subordinated and linear long-memory processes, and their respective partial empirical processes.

From this section on, we use the following terminology: *Gaussian subordinated (partial) empirical process* is supposed to mean (partial) empirical process that originate from a Gaussian subordinated (long-memory) process (see Chapter 1.2.2). Analogously for linear (long-memory) processes (see Chapter 1.2.1), and periodic (long-memory) processes (see Chapter 4.2 ff). Furthermore, for any such process (X_t) with marginal distribution function F, we define the process

$$\zeta_{X_t}(x) := \mathbb{1}\{X_t \leq x\} - F(x),$$

which refers to the single contributions of the (partial) empirical process $\zeta_{[n\tau],X}$, i.e.,

$$\frac{1}{[n\tau]} \sum_{t=1}^{[n\tau]} \zeta_{X_t}(x) = \zeta_{[n\tau],X}(x), \qquad \tau \in [0,1].$$

2.2.1 Uniform weak reduction principle for Gaussian subordinated processes

This section reviews the "*original version*" of the *uniform weak reduction principle* proven by Dehling and Taqqu (1989) [106] for Gaussian subordinated empirical processes.

Let (Z_t) be a stationary long-memory process, with autocovariance function of the form

$$\gamma_Z(k) \underset{k\to\infty}{\sim} L_\gamma(k)|k|^{2d-1}, \qquad d \in \left(0, \frac{1}{2}\right),$$

where L_γ is slowly varying at infinity, and let $G(Z_t)$ be a Gaussian subordinated (long-memory) process as defined in Definition 1.2.7. Recall from Chapter 1.2.2 that any Gaussian subordinated process has a Hermite polynomial expansion with Hermite rank $m \geq 1$, i.e.,

$$G(Z_t) = \sum_{k=m}^{\infty} \frac{J(k)}{k!} H_k(Z_t), \tag{2.4}$$

and consequently (see Lemma 1.2.11),

$$\gamma_G(k) \underset{k\to\infty}{\sim} \frac{J^2(m)}{m!} L_\gamma^m(k)|k|^{m(2d-1)}.$$

Furthermore, recall from Corollary 1.2.12 that such a Gaussian subordinated process exhibits long-range dependence if m is such that $0 < m(1-2d) < 1$. This restriction on the Hermite rank m is necessary to pass the long-memory property from (Z_t) on to the Gaussian subordinated process $G(Z_t)$.

For the contributions $\zeta_{G(Z_t)}$ of the (partial) empirical process $\zeta_{[n\tau],G(Z_t)}$, we obtain a similar Hermite polynomial representation:

$$\zeta_{G(Z_t)}(z) = \sum_{k=m}^{\infty} \frac{J(k,z)}{k!} H_k(Z_t), \tag{2.5}$$

where $H_k(Z_t)$, $k = 1, 2, \ldots$, are the Hermite polynomials and $m = m(z)$ is the Hermite rank of $\zeta_{G(Z_t)}$. Here, the Hermite coefficients $J(k,z)$ are given by

$$J(k,z) = E\left[\mathbb{1}\left\{G(Z_t) \leq z\right\} H_k(Z_t)\right]. \tag{2.6}$$

This suggests that weak convergence of Gaussian subordinated empirical processes $\zeta_{n,G(Z_t)}$ and of Gaussian subordinated partial empirical processes $\zeta_{[n\tau],G(Z_t)}$ can be reduced to weak convergence of sums of Hermite polynomials. This is stated and proven by Taqqu (1975) [382] in the *weak reduction principle*.

Proposition 2.2.2. *(**Weak reduction principle**)*

Let $G(Z_t)$ be a Gaussian subordinated long-memory process having a Hermite polynomial expansion of the form (2.4) with Hermite rank m such that $0 < m(1-2d) < 1$, $d \in \left(0, \frac{1}{2}\right)$. Let $d_n^2 \sim var\left(\sum_{t=1}^{n} G(Z_t)\right) = n^{2-(m(1-2d))} L_\gamma^m(n)$ and let $\tau \in [0,1]$.

If, as $n \to \infty$, the limit of the finite-dimensional distributions of $d_n^{-1} \sum_{t=1}^{[n\tau]} H_m(Z_t)$, denoted by $\bar{S}_{H_m}$, exists, then

$$d_n^{-1} \sum_{t=1}^{[n\tau]} G(Z_t) \xrightarrow{w, D([0,1])} \frac{J(m)}{m!} \bar{S}_{H_m}.$$

This principle asserts that the limiting distribution of normalized partial sums of Gaussian subordinated long-memory processes is determined by the first non-zero term of the Hermite expansion of $G(Z_t)$ only provided the limiting distribution $\bar{S}_{H_m}$ exists.

Taqqu (1975, 1979) [382], [385], and Dobrushin and Major (1979) [119] applied this principle to obtain weak convergence in $D([0,1])$ of partial empirical processes $\zeta_{[n\tau],G(Z_t)}$:

Proposition 2.2.3. *Let $G(Z_t)$ and d_n^2 be as in Proposition 2.2.2, and let $\zeta_{[n\tau],G(Z_t)}$, $\tau \in [0,1]$, have a Hermite polynomial expansion of the form (2.5). Then we obtain the following statements on weak convergence of normalized partial sums of Hermite polynomials and of partial empirical processes:*

(i)

$$d_n^{-1} \sum_{t=1}^{[n\tau]} H_m(Z_t) \xrightarrow{w, D([0,1])} Z_{m,H}(\tau),$$

where $Z_{m,H}(\tau)$, $\tau \in [0,1]$, is a Hermite-Rosenblatt process with Hurst parameter $H = d + \frac{1}{2}$ (see Definition 1.2.24).

(ii) For fixed $z \in \mathbb{R}$:

$$d_n^{-1}[n\tau]\zeta_{[n\tau],G(Z_t)} \xrightarrow{w, D([0,1])} \frac{J(m,z)}{m!} Z_{m,H}(\tau), \tag{2.7}$$

where $J(m,z)$ is as in (2.6).

The *uniform weak reduction principle* proposed by Dehling and Taqqu (1989) [106], Theorem 3.1, now extends the *weak reduction principle* for partial empirical processes $\zeta_{[n\tau]}$ uniformly in both variables $0 \le \tau \le 1$ and $-\infty \le z \le \infty$. In this case, the Hermite rank m of the class of Hermite polynomial expansions (2.5 is defined as

$$m = \inf_{z \in \mathbb{R}} \{m(z)\}. \tag{2.8}$$

Theorem 2.2.4. *(**Uniform weak reduction principle**)*

Let $G(Z_t)$ be a stationary Gaussian subordinated long-memory process. Let $\zeta_{[n\tau],G(Z_t)}$ be the corresponding empirical process with Hermite rank m as in (2.8) such that $0 < m(1-2d) < 1$, $d \in \left(0, \frac{1}{2}\right)$, and let d_n^2 be as in Proposition 2.2.2.

Then, for all $0 < \varepsilon \leq 1$,

$$P\left(\sup_{z\in\mathbb{R},\, 0\leq\tau\leq 1} d_n^{-1}\left|[n\tau]\zeta_{[n\tau],G(Z_t)} - \frac{J(m,z)}{m!}\sum_{t=1}^{[nt]} H_m(Z_t)\right| > \varepsilon\right) \longrightarrow 0.$$

Applying Proposition 2.2.3 (2.7), this principle states that the normalized partial empirical process $\zeta_{[n\tau],G(Z_t)}$ converges weakly in $D([0,1]\times\mathbb{R})$ towards a degenerate m-th order Hermite-Rosenblatt process $Z_{m,H}(\tau)$ with Hurst parameter $H = d + \frac{1}{2}$. This is formulated as a *FCLT* for the partial empirical process in Dehling and Taqqu (1989) [106], Theorem 1.1:

Theorem 2.2.5. *(**FCLT for Gaussian subordinated empirical processes**)*

Let $G(Z_t)$ be a stationary Gaussian subordinated long-memory process. Let $\zeta_{[n\tau],G(Z_t)}$, $\tau \in [0,1]$, be the associated empirical process with Hermite rank m as in (2.8) such that $0 < m(1-2d) < 1$, $d \in \left(0, \frac{1}{2}\right)$ and let d_n^2 be as in Proposition 2.2.2. Then

$$\left\{d_n^{-1}[n\tau]\zeta_{[n\tau],G(Z_t)}(z) \;:\; \tau\in[0,1],\; z\in\mathbb{R}\right\} \xrightarrow{w,D([0,1]\times\mathbb{R})} \left\{\frac{J(m,z)}{m!}Z_{m,H}(\tau) \;:\; \tau\in[0,1],\; z\in\mathbb{R}\right\}.$$

Remark. *Note that the terminology "functional" corresponds to convergence on function spaces such as $D([0,1])$ or $D([0,1]\times\mathbb{R})$. Hence, all theorems in this section that are stated for partial empirical processes are functional limit theorems.*

Finally, let

$$\tilde{\zeta}_{n,G(Z_t)}(\tau, z) := [n\tau]\zeta_{[n\tau],G(Z_t)}(z)$$

define the **two-parameter empirical process** and let

$$\tilde{Z}_{m,H}(\tau, z) := \frac{J(m,z)}{m!}Z_{m,H}(\tau),$$

then Theorem 2.2.5 states the following:

Corollary 2.2.6. *The uniform weak reduction principle also holds in the sense of **weak convergence of two-parameter processes**, i.e.,*

$$d_n^{-1}\tilde{\zeta}_{n,G(Z_t)} \xrightarrow{w,D([0,1]\times\mathbb{R})} \tilde{Z}_{m,H}.$$

As mentioned in the beginning, the uniform weak reduction principle (Theorem 2.2.5) reveals three main aspects:

(i) in contrast to the statement in *Donsker's invariance principle* for iid sequences, in case of long-memory processes, the limiting process is not necessarily Gaussian, except for Hermite rank $m = 1$;

(ii) the convergence rate is always lower than in the iid or short-memory case;

(iii) unlike for iid sequences, the limiting process is degenerate (or totally dependent/perfectly correlated).

Let's look at two examples:

Example 2.2.7. *We consider the simplest Gaussian subordinated long-memory process* $G(Z_t) = Z_t$*, with autocovariance function* $\gamma_Z(k) \sim c_\gamma |k|^{2d-1}$ *and spectral density* $f(\lambda) \sim c_f |\lambda|^{-2d}$*,* $d \in \left(0, \frac{1}{2}\right)$*. Since*

$$\begin{aligned} J(1,z) &= E\left[\mathbb{1}\left\{Z_t \leq z\right\} H_1(Z_t)\right] \\ &= E\left[\mathbb{1}\left\{Z_t \leq z\right\} \cdot Z_t\right] \\ &= -\varphi(z) \neq 0, \end{aligned} \tag{2.9}$$

where $\varphi(z) = \Phi(z)'$ *is the standard normal probability density function, the corresponding Hermite expansion always has Hermite rank* $m = 1$*.*

(i) Proposition 2.2.3 (i) yields weak convergence of the normalized partial sums of Hermite polynomials to $Z_{1,H}(\tau)$*,* $\tau \in [0,1]$*, which is fractional Brownian Motion with Hurst parameter* $H = d + \frac{1}{2}$*, i.e.,*

$$d_n^{-1} \sum_{t=1}^{[n\tau]} H_1(Z_t) = d_n^{-1} \sum_{t=1}^{[n\tau]} Z_t \xrightarrow{w, D([0,1])} Z_{1,H}(\tau) = B_H(\tau),$$

where $d_n^2 \sim c_f \cdot v(d) n^{2d+1}$ *and* $v(d) = 2\Gamma(1-2d)\sin(\pi d)$*.*

Consequently,

$$n^{-\frac{1}{2}-d} \sum_{t=1}^{[n\tau]} Z_t \xrightarrow{w, D([0,1])} \sqrt{c_f \cdot v(d)} B_H(\tau).$$

Applying the FCLT for empirical processes (Theorem 2.2.5), we obtain with (2.9),

$$\begin{aligned} n^{-\frac{1}{2}-d}[n\tau]\zeta_{[n\tau],Z_t} &\xrightarrow{w,D([0,1]\times\mathbb{R})} \sqrt{c_f \cdot v(d)} \cdot \frac{J(1,z)}{1!} Z_{1,H}(\tau) \\ &= \sqrt{c_f \cdot v(d)} \cdot J(1,z) B_H(\tau) \\ &= -\sqrt{c_f \cdot v(d)} \cdot \varphi(z) B_H(\tau). \end{aligned}$$

(ii) In the special case of $\tau = 1$, we obtain weak convergence of the sample mean $\bar{Z}_n = \frac{1}{n}\sum_{t=1}^{n} Z_t$ to a degenerate standard normal random variable:

$$\begin{aligned} n^{-\frac{1}{2}-d}\sum_{t=1}^{n} Z_t &= n^{-\frac{1}{2}-d} n \bar{Z}_n \\ &= n^{\frac{1}{2}-d}\bar{Z}_n \\ &\xrightarrow{w,D([0,1])} \sqrt{c_f \cdot v(d)} B_H(1) \\ &= \sqrt{c_f \cdot v(d)} W, \end{aligned} \tag{2.10}$$

where $W \sim N(0,1)$, and c_f and $v(d)$ as given in (i).

2.2.2 Uniform weak reduction principle for linear processes

We now pass over to C-0-linear processes $Z_t = \sum_{i=1}^{n} a_j \varepsilon_{t-j}$ defined in Definition 1.2.2 with (ε_t) iid and coefficients a_j, $j = 1, 2, \ldots$, that are square-summable and are such that (Z_t) exhibit long-range dependent (see (1.17)). Results on limit theorems for linear long-memory processes and the corresponding linear empirical processes are given in Giraitis et al. (1996) [157], Ho and Hsing (1996) [206], Giraitis and Surgailis (1999) [165], and Wu (2003) [419].

Given a Gaussian C-0-linear processes (Z_t), we obtain a special representation in terms of Hermite polynomials for the associated empirical process ζ_{Z_t}, namely

$$\zeta_{Z_t}(z) = \sum_{k=1}^{\infty} \frac{(-1)^k p^{(k)}(z)}{k!} H_k(Z_t) = -p(z)Z_t + \ldots,$$

where $p^{(k)}$ is the k-th derivative of the Gaussian probability density function, and $H_k(Z_t)$, $k = 1, 2, \ldots$, are the Hermite polynomials. Based on the *uniform reduction principle* from Dehling and Taqqu (1989) (Theorem 2.2.5) we obtain weak convergence to a degenerate Gaussian random variable which corresponds to the limiting distribution of the sample mean $\bar{Z}_n$. This is the same result as in Example 2.2.7 (ii).

For general linear processes (that are not necessarily Gaussian), Hermite representation (2.5) no

longer holds. Therefore, in literature, there exist two further approaches for proving *(functional) CLTs* in the context of linear empirical processes.

One approach is based on an expansion using *Appell polynomials* A_k, which are already mentioned in Remark 1.2.14, i.e.,

$$\zeta_{Z_t}(z) = \sum_{k=1}^{\infty} \frac{(-1)^k p_Z^{(k)}(z)}{k!} A_k(Z_t),$$

where $p_Z^{(k)}$ denotes the k-th derivative of the marginal distribution function p_Z of (Z_t). This leads to a representation which is similar to (2.5), but Appell polynomials are unlike Hermite polynomials not orthogonal in $L^2(\mathbb{R})$. And although, for Gaussian processes, the limiting behavior of normalized sums of Appell polynomials is the same as for normalized sums of Hermite polynomials, the proof of a FCLT is not directly transferable.

Ho and Hsing (1996) [206] and Giratis and Surgailis (1999) [165] use another approach based on the *martingale difference decomposition*, i.e.,

$$\zeta_{Z_t}(z) = \sum_{k=0}^{\infty} \left(P\left(Z_t \leq z \mid \mathcal{F}_{t-k}\right) - P\left(Z_t \leq z \mid \mathcal{F}_{t-k-1}\right)\right),$$

where $\mathcal{F}_t = \sigma(Z_j : j \leq t)$ is the σ-algebra generated by the past variables Z_j, $j \leq t$. This expansion again is orthogonal and replaces the orthogonal Hermite expansion in (2.5). The authors showed that, under suitable moment conditions on (ε_t) and on the corresponding probability density p_ε, the empirical process has a similar asymptotic behavior as in the Gaussian subordinated case. Precisely, the result reads as follows:

Theorem 2.2.8. ***(Weak reduction principle for linear processes)***
Let (Z_t) be a linear long-memory process with marginal density function p_Z and let $\zeta_{n,Z}(z)$ be the corresponding linear empirical process. Assume that for the innovations (ε_t) the following holds:

(i) $E\left[|\varepsilon_1|^{4+\beta}\right] < \infty$ *for some* $\beta > 0$ *and*

(ii) the probability density p_ε is such that

$$\sup_{z\in\mathbb{R}} |p_\varepsilon^{(r)}(z)| + \int |p_\varepsilon^{(r)}|^2 dz < \infty, \qquad r = 0, 1, 2.$$

Then the normalized linear empirical process converges uniformly in probability, moreover

$$n^{\frac{1}{2}-d} \sup_{z\in\mathbb{R}} \left|\zeta_{n,Z}(z) + p_Z(z)\sqrt{c_f \cdot v(d)}W\right| \xrightarrow{P} 0,$$

and consequently,

$$n^{\frac{1}{2}-d}\zeta_{n,Z} \xrightarrow{w,D(\mathbb{R})} -p_Z(z)\sqrt{c_f \cdot v(d)}W,$$

where $W \sim N(0,1)$, c_f *corresponds to the spectral density of* (Z_t), *and* $v(d) = 2\Gamma(1-2d)\sin(\pi d)$.

Extending this result to weak convergence in $D([0,1] \times \mathbb{R})$ of linear partial empirical processes uniformly in $\tau \in [0,1]$ and in $z \in \mathbb{R}$, we obtain weak convergence to degenerate fractional Brownian Motion:

Theorem 2.2.9. ***(FCLT for linear empirical processes)***

Let $\zeta_{n,Z_t}(\tau, z) := [n\tau]\zeta_{[n\tau],Z_t}(z)$ *be the two-parameter process associated with the linear partial empirical process* $\zeta_{[n\tau],Z_t}$.

Then,

$$n^{-\frac{1}{2}-d}\zeta_{n,Z_t}(\tau, z) \xrightarrow{w,D([0,1]\times\mathbb{R})} -p_Z(z)\sqrt{c_f \cdot v(d)} \cdot B_H(\tau),$$

where p_Z *is the probability density function of* (Z_t), $B_H(\tau)$ *denotes fractional Brownian Motion with Hurst parameter* $H = d + \frac{1}{2}$, *and* c_f *and* $v(d)$ *are as in Theorem 2.2.8.*

Note that for a Gaussian linear long-memory process, we obtain the same results as given in Example 2.2.7 (i).

Moreover, a *(multivariate) FCLT* for weak convergence in $D\left([0,1] \times \mathbb{R}^k\right)$, $k \geq 1$, of linear partial empirical processes, is proposed by Giraitis et al. (1996a) [157] and Giraitis and Surgailis (1999) [165]. As we will see in Chapter 4, this multivariate FCLT can be directly applied in the field of hypothesis testing to derive asymptotic distributions of *Kolmogorov-Smirnov-type test statistics* (see, e.g., Giraitis et al. (1996b) [160]). Since our test statistic is of Kolmogorov-Smirnov-type adapted to seasonal time series, we will extend both, the uniform weak reduction principle and the FCLT to so-called *periodic empirical processes*.

Chapter 3

On model selection and estimation for long-memory processes

In this chapter, we turn to the estimation of time series components and unknown model parameters.

A *standard (real-valued) time series model* is of the form

$$Y_t = m(X_t) + \epsilon_t, \qquad t \in \mathbb{Z}, \tag{3.1}$$

where $X_t \in \mathbb{R}$, $t \in \mathbb{Z}$, are identically distributed explanatory variables (random or fixed), m is the unknown mean function and (ϵ_t) is a stationary zero-mean error process with finite variance. There are various textbooks on the analysis of such time series, see, e.g., Heiler (1970) [200], Priestley (1981) [328], Cleveland et al. (1990) [80], Ghysels and Osborn (2001) [154], Brockwell and Davis (2016) [65], Kitagawa (2020) [245], and references therein. For a summary of main mathematical foundations in time series analysis, see Beran (2017) [25].

In application, in a first step, time series components have to be estimated and subtracted. In a second step, the dependence structure of the residual series has to be specified and, based on this, a suitable model out of a class of appropriate model candidates, e.g., out of the class of fractional ARIMA(p,d,0)-processes, has to be selected using some model selection criterion. Thus, in the following, we will focus on *estimating the unknown mean function* m under the assumption of long-range dependence as well as on model selection. In particular, we review commonly used *parametric and nonparametric estimation methods*, mainly *kernel estimation*, of the marginal density of (X_t), and of model components, such as trend and seasonal component. Furthermore, we discuss parametric estimation of long-memory parameters, the *Whittle estimation and modified versions*, and the model selection procedure using various *model selection*

criteria.

These methods are applied to the data examples of both parts (see Chapters 4.5 and 7.6), where we consider a time series model with seasonal component and a long-range dependent residual process. Each seasonal component is estimated by *trigonometric regression* (Section 3.1.2), the unknown long-memory parameters are estimated using *Whittle* or *local Whittle estimation* (Section 3.2.1), and a fractional ARIMA(p,d,0)-process is fitted to the residual series using the Bayesian information criterion (BIC) as model selection criterion (Section 3.2.2). In addition, this chapter also serves as an introductory chapter for defining kernel estimators in the circular setting in Part II (see Chapters 6.2 and 6.3).

3.1 Nonparametric regression and kernel density estimation of real-valued processes

In this chapter, we consider the standard (real-valued) time series model of the form (3.1) with unknown (deterministic) mean function m, observable explanatory variables X_t, $t \in \mathbb{Z}$, and a second-order stationary error process (ϵ_t) with zero mean and finite variance σ_ϵ^2. Such nonparametric regression models include two types of model structure: *random-design models*, where X_t are random design variables, and *fixed-design models*, where $X_t (= x_t)$ are, in general, equidistant design variables.

An overview on nonparametric estimation methods in general can be found in Fan and Gijbels (1996) [131], Härdle et al. (1997, 2004) [192], [193], and Fan and Yao (2008) [132]. Smoothing methods are, e.g., discussed in the monographs of Silverman (1986) [368], Eubank (1988) [129], Müller (1988) [305], Härdle (1990) [191], Wand and Jones (1994) [403], and Ghosh (2018) [153].

In the following, we review some basic definitions and properties of kernel functions and two commonly used kernel estimation methods: the *kernel density estimation* and the *kernel estimation of the mean function.* For estimation of the mean function, we focus on two main techniques in nonparametric regression (due to their statistical properties and their advantages regarding boundary effects), namely *kernel regression* and *locally weighted regression.* We further discuss the impact of the dependence structure on the convergence rate of those estimators and define two types of estimators whose convergence rates are affected by the strength of dependence. In addition, in the context of long-memory regression models, we observe the so-called *phenomenon of smoothing dichotomy*, which is shortly taken up in Chapter 3.1 and discussed in more detail in Chapter 7.1.

3.1.1 Kernel density estimation

The *kernel density estimation (KDE)* is a nonparametric method for estimating the probability density function of a random variable. It serves as an improvement of the histogram method leading to smoother continuous curves by use of suitable *kernel functions* as weighting functions. The basic idea of this point-wise estimation is to weigh the observations according to their distance with respect to each particular point. The smoothness of the curve now depends on the choice of the weights and the distances taken into account.

KDE in general is considered in textbooks such as Silverman (1986) [368], Wand and Jones (1994) [403], and in Scott (2015) [353] in the multivariate setting. For short-range dependent processes, KDE is, e.g., discussed in Altman (1990) [6], Hall and Hart (1990) [181], Hermann et al. (1992) [203], and Wu and Mielniczuk (2002) [420]. KDE in the context of long-range dependent processes is, e.g., addressed in Hall and Hart (1990) [181], Cheng and Robinson (1991) [75], Csörgő and Mielniczuk (1995a) [89], Beran et al. (1999) [31], Honda (2000a) [210], Wu and Mielniczuk (2002) [420], and Beran and Schumm (2017) [36]. And for KDE in a long-range dependent multivariate setting, see, e.g., Beran and Telkmann (2018) [41].

At first, we review definitions and properties of kernel functions in general.

Definition 3.1.1. *A non-negative, unimodal and symmetric probability density function K with*

$$\int_{\mathbb{R}} K(u)u^j du = \begin{cases} 1, & j = 0, \\ 0, & j = 1, \dots, k-1, \\ (-1)^k (k!)\beta_k, & j = k, \end{cases}$$

where $\beta_k \neq 0$ is some constant, is called ***kernel function of order k****.*

We restrict the considerations to kernels that satisfy the following **kernel conditions**:

(K1) K is a kernel of order 2,

(K2) K is symmetric with respect to zero,

(K3) K is Lipschitz continuous on its compact support $[-1, 1]$.

Second-order kernels that are mostly in use are members of the so-called *Beta family* and are of the form

$$K^\gamma(u) = C_\gamma (1 - u^2)^\gamma \mathbb{1}_{\{|u| \leq 1\}}(u), \qquad \gamma \in \mathbb{N},$$

with some non-negative integer γ and a normalizing constant

$$C_\gamma = \frac{\Gamma(\gamma + 3/2)}{\sqrt{\pi}\Gamma(\gamma + 1)}$$

with gamma function Γ (see, e.g., Müller, 1984 [303]). For the so-called **degree of smoothness** $\gamma = 0, 1, 2, 3$, the kernel K^γ represents the *uniform*, the *Epanechnikov*, the *biweight* and the *triweight kernel*, respectively. In this context, the degree of smoothness indicates that for $\gamma \geq 1$ the $(\gamma - 1)-$th derivative of the kernel is Lipschitz continuous. But there are also *kernels of higher order* or *of non-compact support* in use. A well-known example for a kernel of order ∞ is the *Gaussian kernel* given by

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), \qquad u \in \mathbb{R}.$$

See, e.g., Hall and Marron (1988) [183] for a discussion on the choice of the order of the kernel.

Regarding a kernel function K, its scaled version is defined by

$$K_b(u) := b^{-1} K\left(\frac{u}{b}\right), \qquad u \in \mathbb{R}. \tag{3.2}$$

Thus, if the kernel K has support $[-1, 1]$ the **scaled kernel** K_b has support $[-b, b]$. The *smoothing parameter* $b > 0$ is called **kernel bandwidth**, which, for the estimation procedure, intuitively has to be „*as small as possible and as large as necessary*“.

Based on these scaled kernels, we define a general kernel estimator for the marginal density f_X of (X_t), which was first introduced by Rosenblatt (1956) [343] and Parzen (1962) [312].

Definition 3.1.2. *Let K satisfy the kernel conditions (K1) to (K3). Then, given a sample $X_1, \ldots, X_n$ of a time series (X_t), the **Parzen-Rosenblatt estimator (PR-estimator)** is defined as*

$$\hat{f}_X(x) \;=\; \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{X_t - x}{b}\right) = \frac{1}{n} \sum_{t=1}^{n} K_b\left(X_t - x\right), \tag{3.3}$$

which is simply a weighted average.

For asymptotic considerations, a **consistency condition** on the sequence of bandwidths (b_n) is necessary:

(BW1) (b_n) is a sequence of bandwidths that satisfies:

$$b_n \to 0 \quad \text{and} \quad n b_n \to \infty \qquad \text{as } n \to \infty. \tag{3.4}$$

(BW2) A slightly stronger condition requires (in addition to (i)):

$$b_n \to 0 \quad \text{and} \quad n b_n^3 \to \infty \qquad \text{as } n \to \infty. \tag{3.5}$$

Either condition ensures that the sequence of bandwidths does not converge to zero too fast (see also the phenomenon of smoothing dichotomy addressed in Part II, Chapter 7.1).

Assuming that the sequence of bandwidths satisfies the consistency condition (BW1), it is shown by Parzen (1962) [312] in the iid case that the PR-estimator is consistent and asymptotically normal. The same holds in the case of short-range and long-range dependence (see, e.g., Hall and Hart, 1990 [181] and Robinson, 1997 [342]). But, unless the consistency condition holds, the asymptotic behavior of long-range dependent sequences might be completely different (see, e.g., Csörgő and Mielniczuk, 1995a [89]). As mentioned in the beginning of this chapter, under long-memory, the kernel density estimator shows dichotomous behavior depending on the size of the bandwidth. For small bandwidths, the asymptotic behavior is the same as in the iid case, irrespective of long-range dependence. Large bandwidths, however, (depending on the strength of persistence and the convergence rate of the bandwidth) yield either asymptotic normality of the estimator with different normalizing sequences, or convergence to a suitably normalized Hermite-Rosenblatt process (see Definition 1.2.24). This smoothing dichotomy was proven by Ho (1996) [205] and was discussed in the special context of linear processes by Wu and Mielniczuk (2002) [420].

In order to obtain good statistical properties of the kernel density estimator, two variables have to be appropriately chosen: the kernel bandwidth b controlling the amount of smoothing and the kernel function K accounting for the shape of the estimator.

Choice of kernel bandwidth

A crucial problem in kernel estimation is the choice of an "*optimal*" bandwidth, since it amounts to a trade-off between minimizing bias and variance of the estimator.

In order to define *optimality*, the *mean squared error (MSE)* serves as a local criterion: Let $\hat{f}$ be the kernel estimator given in (3.3). In the iid and short-memory case, at some fixed point $x_0 \in \mathbb{R}$,

$$\begin{aligned} MSE(\hat{f}(x_0)) &= E\left[\left(\hat{f}(x_0) - f(x_0)\right)^2\right] \\ &= \frac{b_n^4}{4}(\mu_2(K))^2|f''(x_0)|^2 + \frac{R(K)}{nb_n} + o\left(\max\left\{b_n^4, (nb_n)^{-1}\right\}\right) \\ &= \mathcal{O}\left(\max\left\{b_n^4, (nb_n)^{-1}\right\}\right), \end{aligned}$$

where $\mu_2(K) = \int x^2 K(x)dx$ is the second moment of the kernel and $R(K)^2 = \int K^2(x)dx$ is the curvature of the kernel function. Minimizing this *asymptotic MSE (AMSE)* (with respect to b_n) then leads to a *locally optimal bandwidth* of order $\mathcal{O}\left(n^{-\frac{1}{5}}\right)$, and the order of the optimal MSE becomes $\mathcal{O}\left(n^{-\frac{4}{5}}\right)$.

In general, one is interested in an *overall optimal bandwidth*. A common procedure in bandwidth selection is the minimization of the *asymptotic mean integrated squared error (asymptotic MISE)*, where the MISE, as a global criterion, takes the whole curve into account. However, this procedure does not correct possible *boundary effects* that occur due to symmetry of the kernel. Accordingly, in the iid and short-memory case, the MISE of the kernel estimator is given by

$$\begin{aligned} MISE(\hat{f}; b_n) &= E\left[\int \left(\hat{f}(x) - f(x)\right)^2 dx\right] \\ &= \frac{b_n^4}{4}(\mu_2(K))^2 R(f'') + \frac{R(K)}{nb_n} + o\left(\max\left\{b_n^4, (nb_n)^{-1}\right\}\right) \\ &= \mathcal{O}\left(\max\left\{b_n^4, (nb_n)^{-1}\right\}\right), \end{aligned} \tag{3.6}$$

where $\mu_2(K)$ and $R(K) = \int K^2(x)dx$ as above and $R(f'')$ is the overall curvature of the density f. For both, iid and short-memory sequences, minimizing the asymptotic MISE again results in an optimal bandwidth of order $\mathcal{O}\left(n^{-\frac{1}{5}}\right)$.

In the long-memory case, the asymptotic behavior of the estimator strongly depends on the size of the kernel bandwidth. For small kernel bandwidths, the convergence is of the same order as in the iid case, and for large bandwidths, the convergence strongly depends on the strength of the memory parameter. This phenomenon is called *smoothing dichotomy* (see Hall and Hart, 1990 [181]). Since, for large bandwidths, in addition, the limiting behavior varies, the notion *smoothing trichotomy* was introduced by Wu and Mielniczuk (2002) [420]. In Beran et al. (2013) [33], Theorem 5.17 ff., this trade-off is discussed in detail and an overview of convergences rates depending on the size of the kernel bandwidth and of the memory parameter is given.

Note that the notion of the MISE and consequently of the optimal bandwidth requires knowledge about the second derivatives of the unknown density f. Thus, in practice, *(iterative) bandwidth selection procedures* are needed for "*optimal estimation*" of the density. In literature, the most frequently used (practical) methods are *Rules of Thumb*, *Cross-validation*, *Plug-in methods*, and *smoothed Bootstrap techniques*. See, e.g., Müller (1984) [303], Silverman (1986) [368], Hall and Marron (1987) [182], Gasser et al. (1991) [148], Jones et al. (1991) [230], Hall et al. (1992) [184], Browne (2000) [67], and Arlot and Celisse (2010) [10] for introductions to these methods. A review and a comparison of different methods on bandwidth selection (mostly dealing with iid samples) is given by Wand and Jones (1994) [403], Jones et al. (1996) [231], Heidenreich et al. (2013) [199], Köhler et al. (2014) [246], and references therein. Robinson (1989) [339] for instance developed a *cross-validation method*. *Plug-in approaches* and *data-driven procedures* in general are discussed in Chiu (1989) [78], Hermann et al. (1992) [203], Hermann and Gasser (1994) [202], and under the assumption of long-range dependence in Hart and Vieu (1990) [195], Ray and Tsay (1997) [334], Beran and Feng (2002b) [29], and Beran, Steffens and Ghosh (2018) [37].

A discussion on the bandwidth selection from a theoretical perspective is given in Hall et al. (1995a,b) [185], [186].

Choice of kernel function

With respect to asymptotic considerations, the shape of the kernel function does not play a key role in KDE. In fact, if the bandwidth is chosen optimally, all symmetric and unimodal kernels perform almost the same way, and the loss in efficiency when using a non-optimal kernel is quite small (see, e.g., Silverman (1986) [368] for a discussion on the optimal choice of the kernel). Moreover, since the asymptotic convergence of the MISE does not depend on the choice (or rather on the shape) of the kernel (see equation (3.6)), it does not much affect the efficiency of the estimate.

Remark 3.1.3. *So far, we considered second-order kernels, but the order k of the kernel has an important effect on the convergence rate of the optimal estimator: The higher the order of the kernel the faster the convergence, although, as the order of the bias reduces to* $\mathcal{O}\left(b^{2k}\right)$*, the finite sample variance is usually larger than in the case of a second-order kernel. However, the use of higher-order kernels, which were first introduced by Bartlett (1963) [16], additionally needs rather strong assumptions, whereas a second-order kernel needs minimal assumptions only. For a discussion on higher-order kernels see, e.g., Parzen (1962) [312], Gasser and Müller (1979, 1984) [145], [146], Gasser et al. (1985) [147], and Jones (1995) [229].*

In order to avoid boundary effects that occur when using symmetric kernels, one commonly uses *asymptotically equivalent kernels*. This issue is considered for instance in Lejeune (1985) [260], Müller (1987) [304], Lejeune and Sarda (1992) [261], Ruppert and Wand (1994) [346], Feng (1999) [133], Beran and Feng (2002a) [28], and Beran, Steffens and Ghosh (2018) [37]. A summary can be found in Beran et al. (2013) [33].

3.1.2 Nonparametric regression

Back to the regression model given in (3.1). The relationship between the variables X and Y in can be formalized using the conditional expectation of Y given $X = x$, as

$$m(x) := E\left[Y|X = x\right] = \frac{\int y f_{X,Y}(x, y)dy}{f_X(x)} \tag{3.7}$$

with joint probability density function $f_{X,Y}$ and marginal density function f_X.

Kernel regression estimators for the mean function, also denoted as *kernel regressors*, are weighted local averages of the response variables Y_t. In general, one considers *linear smoothers*

of the form

$$\hat{m}(x) = \sum_{t=1}^{n} w_t(x) Y_t$$

with weights $w_t(x)$, $t = 1, \ldots, n$, for which $\sum_{t=1}^{n} w_t(x) = 1$.

The references given in the introductory section mainly consider nonparametric regression under iid errors. Under long-memory, nonparametric regression is as well widely discussed in literature, see, e.g., Altmann (1990) [6], Hall and Hart (1990) [181], Hart (1991) [194], Csörgő and Mielniczuk (1995b) [90], Wang (1996) [405], Hidalgo (1997) [204], Ray and Tsay (1997) [334], Robinson (1997) [342], Ghosh (2001) [151], Beran and Feng (2002a-c) [28], [29], [30], and Beran et al. (2002) [32], see also Beran et al. (2013) [33] and references therein.

All these references consider a fixed-design model of the form (3.1), where X_t are real-valued deterministic explanatory variables. The random-design model is addressed in Györfi et al. (1989) [180], Cheng and Robinson (1991, 1994) [75], [76], Csörgő and Mielniczuk (1995b, 1999, 2000) [90], [93], [94], Mielniczuk and Wu (2004) [298], Zhao and Wu (2008) [426], and Kulik and Lorek (2011) [250]. Random-design regression is discussed in Künsch et al. (1993) [252], Ho (1996) [205], Koul et al. (2004) [249], Guo and Koul (2008) [179], and Kulik and Wichelhaus (2012) [251].

As mentioned earlier, in the long-range dependence setting, one observes dichotomous asymptotic behavior depending on the type of model (fixed or random) and the size of the bandwidth. In this context, *dichotomy* is meant in the sense that long-range dependence has a non-negligible impact on the convergence rate of the estimator or it does not have an impact. Csörgő and Mielniczuk (1999) [93] give a comparison between the fixed- and the random-design model. They show that the asymptotic variance of the estimator in the random-design model is smaller (has a larger order) than in the fixed-design model while the bias of the kernel estimator is of the same order. This *design type-dichotomy* is for instance also addressed in Csörgő and Mielniczuk (1995b, 2000) [90], [94], Ho (1996) [205], Csörgő (2002) [84], and Mielniczuk and Wu (2004) [298]. In addition, Bryk and Mielniczuk (2007, 2008) [69], [70] and Bryk (2014) [68] introduced a randomization scheme of the grid points in a fixed-design regression model in order to decrease the dependence between the observations and thus to diminish the impact of strong dependence. The *bandwidth-dichotomy* is discussed in Csörgő and Mielniczuk (1995) [90] for Gaussian subordinated processes, and in Csörgő (2002) [84], Wu and Mielniczuk (2002) [420], and Mielniczuk and Wu (2004) [298] for linear processes. This dichotomy introduces a new aspect to the question of bandwidth selection when long-range dependence is expected. Since we observe the same dichotomous phenomenon in the circular setting in Part II, we will refer to this smoothing dichotomy and review the results by Mielniczuk and Wu (2004) [298] in detail

later (see Chapter 7.1).

Kernel regression

In the *random-design situation* of (3.1), where X_t, $t = 1, \ldots, n$, are random explanatory variables, the joint density $f_{X,Y}$ and the marginal density f_X in (3.7) are unknown and have to be estimated using appropriate kernel estimators. If we use the simplest estimator for the joint density, namely the *product kernel*, the numerator in (3.7) becomes

$$\int y f(x,y) dy = \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{X_t - x}{b}\right) Y_t,$$

where K is a kernel function with kernel bandwidth b satisfying the kernel conditions (K1) to (K3) and the consistency condition (BW1), respectively.

Obviously, we can replace the denominator f_X in (3.7) by the PR-estimator $\hat{f}_X$ in (3.3). The resulting **Nadaraya-Watson estimator (NW-estimator)** estimator was first introduced by Nadaraya and Watson in 1964 (Nadaraya, 1964 [308], and Watson, 1964 [410]), and is defined by

$$\hat{m}_{NW}(x) = \frac{\sum_{t=1}^{n} K_b\left(X_t - x\right) Y_t}{\sum_{t=1}^{n} K_b\left(X_t - x\right)} \tag{3.8}$$

which is a weighted average of variables Y_t in a neighborhood of x.

The numerator relates to the so-called **Priestley-Chao estimator (PC-estimator)** (Priestley and Chao, 1972 [329]), i.e.,

$$\hat{m}_{PC}(x) = \frac{1}{n} \sum_{t=1}^{n} K_b\left(X_t - x\right) Y_t, \tag{3.9}$$

which is an estimator for the mean function m for equidistant deterministic variables $x_1, \ldots, x_n$. Hence, together with (3.3), we can rewrite the NW-estimator (3.8) as

$$\hat{m}_{NW}(x) = \frac{\hat{m}_{PC}(x)}{\hat{f}_X(x)}. \tag{3.10}$$

This estimator naturally extends kernel estimation in the fixed-design situation to the problem of *estimating an unknown conditional expectation function.*

For asymptotic considerations, the sequence of bandwidths has to satisfy either consistency condition (BW1) or (BW2), the kernel function K has to satisfy the kernel conditions (K1) to (K3), and $\mu_2(K) := \int_{-1}^{1} x^2 K(x) dx$ and $R(K) := \int_{-1}^{1} K^2(x) dx$ are finite. Moreover, we assume that m is twice continuously differentiable.

We obtain the following asymptotic properties:

(i)
$$E\left[\hat{m}_{PC}(x)\right] = f_X(x)m(x) + \mathcal{O}\left(b^2\right),$$
which means that the bias is not negligible unless the density f_X is the uniform distribution on $[0,1]$.

(ii) Provided the kernel density estimator $\hat{f}_X$ is a consistent estimator for the marginal density f_X, we obtain convergence in probability of the NW-estimator given in (3.10).

In this case, $\hat{m}_{NW}$ is a consistent estimator for the mean function m and we can conclude that the NW-estimator for a random-design model is a modification of the PC-estimator (3.9) divided by a consistent estimate of the density f_X.

In the *fixed-design situation* of (3.1), one usually assumes that $x_1, \ldots, x_n$ are equidistant and fixed design variables. Thus, the marginal density f_X is the uniform distribution in x_i, $i = 1, \ldots, n$, which simplifies the estimation of the mean function. In addition, for modeling slight changes in the mean function, one commonly uses *rescaled time points*, i.e., an equally spaced sequence $x_t = x_{t,n} = \frac{t}{n} \in [0,1]$, $t = 0, \ldots, n$. This takes into account that the noise varies much faster compared to the mean function. As a consequence, for $n \to \infty$, the number of variables x_t becomes more and more dense in $[0,1]$. In the framework of locally stationary time series, rescaling time points to the interval $[0,1]$ allows for consistent estimation of the mean function. Using these rescaled time points, the *fixed-design regression model* can be rewritten as

$$Y_t = Y_{t,n} = m(x_t) + \epsilon_t = m\left(\frac{t}{n}\right) + \epsilon_t, \tag{3.11}$$

and the PC-estimator defined in (3.9) at some point $t_0 \in \{0, \ldots, n\}$ then reduces to

$$\hat{m}_{PC}(t_0) = \frac{1}{nb} \sum_{t=1}^{n} K\left(\frac{t - t_0}{nb}\right) Y_t,$$

which is a weighted average of the observations in some neighborhood of $\frac{t_0}{n} \in [0,1]$. Analogously as in the random-design case, one can show that the kernel estimator is a consistent estimator. Under the assumption of iid residuals, the asymptotically optimal bandwidth is of order $\mathcal{O}\left(n^{-1/5}\right)$ yielding an optimal MISE of order $\mathcal{O}\left(n^{-4/5}\right)$ (analogously as in (3.6)). This is the same result as for the PR-estimator.

For the *fixed-design regression model* (3.11), e.g., Hall and Hart (1990) [181] considered kernel regression for μ under short-range dependence and showed that the optimal MISE is of the same order as in the iid case. They also showed that, under the assumption of long-memory

the MISE is of order $\mathcal{O}\left(\max\left\{b4, (nb)^{2d-1}\right\}\right)$ resulting in an asymptotically optimal bandwidth of order $\mathcal{O}\left(n^{\frac{1-2d}{5-2d}}\right)$. Consequently, the convergence of the estimator is slower compared to the iid and short-range case, see, e.g., Beran et al. (2013) [33]. Note that this is the same rate of convergence as for the general KDE. And again, from a practical perspective, the expression of the optimal bandwidth is not directly applicable since it involves unknown quantities similar to the KDE that have to be estimated by use of special procedures. Those bandwidth selection procedures are generally speaking the same as in KDE.

The *random-design model* (3.1) for a bivariate process (Y_t, X_t), where, in addition, the residual process depends on both random variables Y_t and X_t, is considered in Mielniczuk and Wu (2002) [420] and Wu and Mielniczuk (2004) [298] for both, short-range and long-range dependent variables. They derive asymptotic formulas for the NW-estimator under various conditions and investigate functional limit laws for the estimator. It is shown that the asymptotic distribution of the estimator strongly depends on the amount of smoothing and the memory parameter of both random variables. In the random-design setting under long-range dependence, again the smoothing dichotomy/trichotomy phenomenon is observed. This means, the convergence rate strongly depends on the size of the bandwidth yielding similar results as for kernel density estimation under long-range dependence (see Chapter 3.1.1) and which is in contrast to the fixed-design kernel regression. Beran et al. (2013) [33], Theorem 7.28 ff, gives a detailed overview of the limiting behavior according to the size of the bandwidth and the strength of the memory parameter.

One widely discussed problem in kernel estimation are the so-called *boundary effects*, which occur when estimating the mean function using symmetric kernels at boundary points $t \in [0, b) \cup (1-b, 1]$. To remove these effects, many different methods are proposed in literature. One method uses so-called *boundary kernels* that are, e.g., defined based on *truncated asymptotically equivalent kernels*. This issue is, for instance, considered in Gasser and Müller (1979) [145], Müller (1984, 1991) [303], [306], Gasser et al. (1985) [147], and Müller and Wang (1994) [307]. The differences in estimation at interior and boundary points are pointed out in Lejeune (1985) [260], Lejeune and Sarda (1992) [261], Ruppert and Wand (1994) [346], and Feng (1999) [133], and methods for boundary correction are, e.g., discussed in Jones (1993) [228] for KDE under iid residuals, and in Beran and Feng (2001, 2002a) [27], [28] for estimation of the mean function under short- and long-memory.

Another approach to overcome the problem of boundary effects is the **locally weighted regression** which include an automatic adaptation at the boundary. Typical examples are the *local linear*, the *local polynomial*, and the *local trigonometric regression*.

Locally weighted regression

The *locally weighted regression (LWR)* is a kernel-based method where the kernel smoothing procedure is applied to traditional regression by using *locally weighted least squares*. These LWR-estimators are asymptotically equivalent to certain kernel estimators and, in particular, for finite samples, they are more attractive due to the absence of boundary effects. In fact, the bias of the LWR-estimator at the boundary is substantially reduced compared to that of the general kernel estimators. Further practical advantages are that the LWR-estimator can be adapted to the particular model design (fixed or random) and that the estimation procedure applies as well to the estimation of the derivatives. On the contrary, the general kernel estimator is more useful for asymptotic considerations (due to the simple kernel representation). For more details concerning asymptotic equivalence between kernel estimators and local regressors, see Hastie and Loader (1993) [196] besides the above mentioned references.

The LWR method is frequently used when the deterministic component in the nonparametric model (3.1) shows both, non-periodic and periodic features.

Accordingly, the general (fixed-design) time series model (3.11) can be decomposed into three different terms:

$$Y_t = m(x_t) + \epsilon_t = \mu(x_t) + S(x_t) + \epsilon_t, \qquad t \in \mathbb{N}. \tag{3.12}$$

This means the mean function m is decomposed into a twice-continuously differentiable *non-periodic component* μ and a *seasonal component* S, and (ϵ_t) is (as before) a stationary zero-mean error process with finite variance σ_ϵ^2. The seasonal component is (for identifiability reasons) usually assumed to be **exactly periodic** with period $T_0 \geq 2$, i.e.,

$$S(t + T_0) = S(t) \quad \text{and} \quad \sum_{t=1}^{T_0} S(t) = 0. \tag{3.13}$$

Nonparametric estimation of μ and S in the iid case, mainly local polynomial fitting and local trigonometric regression, is considered for instance in Feng (1999, 2013) [133], [134], Heiler and Feng (2000) [201], and Feng and McNeil (2008) [135]. Local polynomial fitting in the context of dependent residuals is discussed in Stone (1977) [372], Cleveland (1979) [79], Müller (1987) [304], Hastie and Loader (1993) [196], Ruppert and Wand (1994) [346], Fan and Gijbels (1996) [131], and Beran and Feng (2002c) [30]. Estimation of the seasonal component S of a detrended time series with long-range dependent residuals is, e.g., considered in Yajima (1988, 1991) [423], [424] for parametric methods, and in Beran, Steffens and Ghosh (2018) [37] for nonparametric methods. These references mainly refer to fixed-design models. For an overview on local polynomial fitting in the random-design case see, e.g., Fan and Gijbels (1996) [131] and references therein.

Since, in both parts of the thesis (see Chapter 4 and Chapter 7), we consider a time series model with periodic component, we shortly review the overall principle of both, fitting trigonometric series and the LWR for such additive time series models.

We will only focus on the fixed-design setting with rescaled time points $x_t = x_{t,n} = \frac{t}{n} \in [0,1]$ as introduced in (3.11), and the model (3.12) accordingly reduces to

$$Y_{t,n} = m\left(\frac{t}{n}\right) + \epsilon_t = \mu\left(\frac{t}{n}\right) + S\left(\frac{t}{n}\right) + \epsilon_t, \qquad t = 1, \ldots, n. \tag{3.14}$$

For any nonparametric regression model $Y_{t,n}$ of the form (3.14) and estimator $\hat{Y}_{t,n}$, LWR leads to the *weighted least squares problem*

$$\min_b \left\{ \sum_{t=1}^{n} \left(Y_{t,n} - \hat{Y}_{t,n}\right)^2 K\left(\frac{t-t_0}{nb}\right) \right\},$$

where K is any kernel function with bandwidth b satisfying the conditions (K1) to (K3) on the kernel, and either the consistency condition (BW1) or (BW2) on the sequence of bandwidths.

For any mean function m the LWR-approach is the following: We model m at some point t_0 as

$$m(t_0) = \mathbf{X}^T \beta(t_0),$$

where $\mathbf{X}$ is the *matrix of regressors*, superscript ^{T} denotes the transpose of $\mathbf{X}$ and $\beta(t_0)$ is the vector of *local regression coefficients*. Minimizing this with respect to the bandwidth b yields the generally known *weighted least squares estimator*

$$\hat{\beta}(t_0) = (\mathbf{X}^T \mathbf{K}(t_0) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{K}(t_0) \mathbf{y},$$

where $\mathbf{K}(t_0) = diag\left(K_b\left(\frac{t-t_0}{n}\right)\right)$, $t = 1, \ldots, n$, is a diagonal matrix with the kernel function K_b evaluated at points $\frac{t-t_0}{n}$ on the diagonal, and $\mathbf{y}$ is the *vector of observations*.

For the long-term trend function μ, *local polynomial fitting* is commonly used. Provided that $\mu(x)$ is $(p+1)$-times continuously differentiable in a small neighborhood of x, it is shown, e.g. by Fan and Gijbels 1996 [131], that in the iid case, the asymptotically optimal bandwidth is of order $\mathcal{O}\left(n^{-\frac{1}{2p+3}}\right)$.

Remarks. *(i) For $p = 1$, one obtains the usual convergence rates for kernel regressors.*

(ii) For $p = 0$, the NW-estimator (3.8) is a special case of this class of local polynomial estimators, i.e., the NW-estimator is the constant local least squares estimator

$$\hat{m}_{NW}(t_0) = \hat{\beta}_0(t_0).$$

The long-range dependent and the short-range dependent case is considered in Beran and Feng (2002a) [28]. In either case, the MISE is of order $\mathcal{O}\left(\max\left\{b^{2(p+1)}, (nb)^{2d-1}\right\}\right)$ yielding an optimal bandwidth of order $\mathcal{O}\left(n^{\frac{2d-1}{2p+3-2d}}\right)$. This means that the order of both, the MISE and the optimal bandwidth, strongly depend on the memory parameter d. Thus, in the long-memory case, i.e., for $d \in \left(0, \frac{1}{2}\right)$, the convergence rate is slower than in the short-memory or the iid case.

For the seasonal component S, *local trigonometric regression* is frequently in use. Feng (1999) [133] introduced this method with kernels of order 2 as weighting functions in the iid case, and Beran, Steffens and Ghosh (2018) [37] extended his results to the long-range dependent case and to kernels of even order $2k$, $k > 1$. Either paper assume that the seasonal component is exactly periodic (see (3.13)), and can be modeled as a *Fourier series* of the form

$$S(t;n) = \sum_{j=1}^{\left[\frac{T_0}{2}\right]} \beta_j\left(\frac{t}{n}\right)\cos\left(\lambda_j t\right) + \sum_{j=1}^{\left[\frac{T_0}{2}\right]-1} \beta_{j+\left[\frac{T_0}{2}\right]}\left(\frac{t}{n}\right)\sin\left(\lambda_j t\right), \tag{3.15}$$

where $T_0 \geq 2$ is the *period*, $\lambda_j = \frac{2\pi j}{T_0}$ are the *Fourier frequencies*, and the *regression coefficients* $\beta_j : [0,1] \to \mathbb{R}$, $j = 1, \ldots, T_0$, are functions of rescaled time $\frac{t}{n}$, $t = 1, \ldots, n$.

For asymptotic considerations, one assumes that the coefficient functions β_j, $j = 1, \ldots, T_0$, are twice continuously differentiable and the assumptions on the kernel function (K1) to (K3) and the consistency condition (BW2) on the sequence of bandwidths are satisfied. Under these assumptions, Feng (1999) showed that (in the iid case) the optimal sequence of bandwidths is of order $\mathcal{O}\left(n^{-\frac{1}{5}}\right)$ leading to an optimal MISE of order $\mathcal{O}\left(n^{-\frac{4}{5}}\right)$, which is the same order as for the local polynomial estimator with $p = 1$.

Beran, Steffens and Ghosh (2018) [37] considered local trigonometric regression of the seasonal component S under long-range dependence. It is proven that the optimal sequence of bandwidths is of order $\mathcal{O}\left(n^{-\frac{1}{5}}\right)$. Interestingly, this is the same order as in the iid case, i.e., the convergence rate is independent of the strength of persistence characterized by the long-memory parameter d. This is in contrast to the situation of estimating the mean function where the order of the optimal bandwidth is larger than in the uncorrelated case. In addition, using the fact that the estimators of trend and seasonal component are asymptotically independent of each other, these components can be estimated separately and with different kernel functions yielding different asymptotically optimal bandwidths. For the proof of independence of μ and S and a data-driven algorithm for simultaneous estimation of seasonal and trend component, see Beran, Steffens and Ghosh (2018) [37].

3.2 Estimation of model parameters and model selection

This chapter is divided into two sections. In the first section, we assume to have knowledge about the underlying model and have a look at different parameter estimation methods, mainly the *Whittle estimation* and its variants, and discuss properties of these estimators. In application, such knowledge is not necessarily present. Moreover, although we have chosen the class of parametric models, both, the dimension of the parameter vector and, in this context, the choice of the "*best model*" are still not determined. Hence, one has to combine estimation of unknown parameters with fitting an appropriate model to the sample. This procedure and the commonly used model selection criteria are discussed in the second section.

3.2.1 The Whittle estimation and its local variants

Since the discovery of long-range dependent behavior in the area of hydrology by Hurst (1951) [216], estimation of unknown long-memory parameters has gained great interest. An overview of different estimation methods including heuristic approaches and approaches both, in time and in frequency domain, is given in Beran et al. (2013) [33] and references therein. In this section, we review the commonly used methods.

A first heuristic approach is the so-called *rescaled range (R/S) statistic* which was introduced by Hurst in 1951 [216] and further investigated in the 70s by Mandelbrot and Wallis (1969a-c) [280], [281], [282], Mandelbrot (1975) [271], and Mandelbrot and Taqqu (1979) [277]. This *R/S statistic*, described in detail in Hurst (1965) [217], measures the variability of a time series by relating the range of the partial sums of deviation from mean over a specific period of time to the standard deviation. This and other heuristic methods are useful to gain a first idea of the Hurst parameter H but they are less suitable for statistical inference. For more details, see, e.g., Beran et al. (2013) [33].

Here, we will focus on the so-called **Whittle estimation**. Let $(x_1, \ldots, x_n)$ be a random sample from a stationary zero-mean Gaussian time series (X_t). We are interested in fitting a parametric long-memory model, e.g., a fractional ARIMA(p,d,0)-model, *optimal* to the sample, and hence want to estimate all parameters of the presumed long-memory process. In this context, *optimal* is meant in the sense of *minimizing a certain quality-of-fit function.*

In general, apart from the Hurst parameter, there are additional unknown parameters that have to be estimated: the scale parameters σ_ϵ^2 and/or c_f, and the parameters characterizing the short-memory behavior. Possible approaches are parametric estimation methods in time or in frequency domain that are based on *maximum likelihood estimation (MLE).* We will shortly review the *exact Gaussian MLE* as an approach in the time domain, and *Whittle's approximate MLE* and the *local Whittle estimation* as periodogram-based approaches, i.e., approaches in the

frequency domain. For more details see, e.g., Beran et al. (2013) [33], Beran (2017) [25], and references therein.

In case of a stationary zero-mean Gaussian process (X_t), for a random vector $(X_1, \ldots, X_n)$ and the unknown parameter vector η, the *approximate likelihood function* which, except for constant terms, coincides with the *log-likelihood function*, is given by

$$\mathcal{L}_n(\eta) := n^{-1} \log\left(|\Sigma_n(\eta)|\right) + n^{-1} x^T \Sigma_n^{-1}(\eta) x, \tag{3.16}$$

where $|\Sigma_n(\eta)|$ denotes the determinant of the matrix $\Sigma_n(\eta)$. Minimizing this log-likelihood function leads to the **exact Gaussian MLE**

$$\hat{\eta}_{n,\mathrm{MLE}} = argmin\left\{\mathcal{L}_{n,\mathrm{exact}}(\eta)\right\}.$$

Consistency, *asymptotic normality*, and *asymptotic efficiency* of the exact Gaussian MLE for stationary Gaussian processes with short-memory dependence were derived by Hannan (1973) [188]. Consistency and asymptotic normality for fractional Gaussian ARIMA(0,d,0)-processes have been proven by Yajima (1985) [422], and for general Gaussian long-memory processes with a possibly unknown mean by Dahlhaus (1989, 2006) [98], [100]. Moreover, Dahlhaus (1989) [98] showed that for a Gaussian long-memory process, although its spectral density has a pole at the origin, the exact MLE is asymptotically efficient. In case of non-Gaussian processes (X_t), the respective estimator is called *pseudo- or quasi-MLE* and (under fairly general regularity conditions) is still consistent and asymptotically normal.

The exact Gaussian MLE requires calculation of both, the determinant of the covariance matrix Σ_n and the inverse of the covariance matrix Σ_n^{-1}. Here, Σ_n is a *Toeplitz matrix* (see, e.g. Grenander and Szegö, 1958 [175]). For approximation of the inverse of such a matrix, there are methods both in time and in frequency domain.

Time domain approximations of the Gaussian likelihood can, e.g., be found in Shaman (1975, 1976) [355], [356], Bhansali (1982) [48], and Coursol and Dacunha-Castelle (1982) [82]. One useful approximation of the Gaussian likelihood involves *invertible linear processes*, i.e., MA(∞)-processes, which can be generalized to fractional ARIMA processes. This is discussed in detail, e.g., in Beran (1995) [23] and Beran et al. (1998, 2013) [26], [33].

In the frequency domain, a first approximation of the Gaussian MLE, the so-called **Whittle's estimation**, was introduced by Whittle (1951, 1953) [415], [416] for short-memory time series. The overall idea is to decompose Σ_n^{-1} using the discrete Fourier transform of the time series and, based on this, formulate a frequency-domain representation of the Gaussian likelihood, the **Whittle likelihood**. See, e.g., Rao and Yang (2021) [333] (and references therein) for

a discussion on the connection between the Gaussian and the Whittle likelihood. In the last decades, the Whittle likelihood and applications of this estimation method have gained increasing interest due to computational advantages and the asymptotic properties of the corresponding Whittle estimator. For Whittle estimation in the context of long-memory time series, see, e.g., Beran (1995), Robinson (1995), Dahlhaus (2000), Hurvich and Chen (2000), Giraitis and Robinson (2011), Choudhuri et al. (2004), Abadir et al. (2007), Shao and Wu (2007) and Giraitis et al. (2012), and see Cheung and Hassler (2020) [77] for Whittle-type estimation under long-memory and non-stationarity.

The *Whittle estimate* is a MLE approximate for Gaussian processes in the frequency domain and is based on rewriting the likelihood function in terms of the periodogram of the underlying process. Let $I_{n,X}(\cdot)$ denote the periodogram of the process (X_t), i.e.,

$$I_{n,X}(\lambda) = (2\pi n)^{-1}\Big|\sum_{t=1}^{N} X_t e^{it\lambda}\Big|^2.$$

Note further that $\log(|\Sigma_n(\eta)|)$ in (3.16) can be approximated by

$$\lim_{n\to\infty} \log(|\Sigma_n(\eta)|) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log(f_X(\lambda;\eta))d\lambda = \log(\sigma_\epsilon^2) - \log(2\pi),$$

where f_X is the spectral density of (X_t), see, e.g., Grenander and Szegö (1958) [175].

Then the **approximate Whittle log-likelihood** in terms of the periodogram is of the form

$$\mathcal{L}_{n,\mathrm{W}}(\eta) := \log(\sigma_\epsilon^2) + (2\pi)^{-1}\int_{-\pi}^{\pi} \frac{I_{n,X}(\lambda)}{f_X(\lambda;\eta)}d\lambda, \tag{3.17}$$

and can be further approximated by replacing the integral by a Riemann sum, i.e.,

$$\mathcal{L}_{n,\mathrm{W,R}}(\eta) := \log(\sigma_\epsilon^2) + \frac{2}{n}\sum_{j=1}^{N} \frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j;\eta)} \approx \mathcal{L}_{n,\mathrm{W}}(\eta), \tag{3.18}$$

where $N = \left[\frac{n-1}{2}\right]$, and $\lambda_j = \frac{2\pi j}{n}$, $j = 1,\ldots,N$, are the *Fourier frequencies*.

This Whittle estimator is consistent and has the same asymptotic distribution as the exact MLE, thus is asymptotically normal. Beran (1986) [20], Fox and Taqqu (1986) [143], and Dahlhaus (1989) [98] showed this for a Gaussian process (X_t). Hence, in this case, the Whittle estimator is asymptotically efficient. Later, Giraitis and Surgailis (1990) [164] and Horvàth and Shao (1999) [213] generalized this (under regularity conditions) to linear processes (X_t). However, proven by Giraitis and Taqqu (1999) [166], for Gaussian subordinated processes, asymptotic

efficiency (in general) is no longer valid.

For such non-Gaussian long-memory processes, another *approximate Whittle likelihood* is of interest. Since the long-memory behavior is determined by the asymptotic behavior of the spectral density at the origin, the Whittle estimator introduces a large bias at higher frequencies. Thus, the idea is to use local estimators, i.e., to restrict the estimation to an arbitrarily small neighborhood of the origin. This is done by applying the so-called **local Whittle estimation**, which is a semiparametric approach, since assumptions on the spectral density are only needed at small frequencies. By using only the lowest $m \leq N$ frequencies in (3.18), we obtain the **approximate local Whittle likelihood function**

$$\mathcal{L}_{n,\text{local W}}(\eta) \approx \frac{2}{m} \sum_{j=1}^{m} \left(\log \left(f_X(\lambda_j; \eta) + \frac{I_{n,X}(\lambda_j)}{f_X(\lambda_j; \eta)} \right) \right).$$

Robinson (1995a,b) [340], [341] showed that (under mild regularity conditions) this local Whittle estimator is consistent and asymptotically normal. This means, for the local Whittle estimator $\hat{\eta}$, as $m \to \infty$ and $\frac{m}{n} \to 0$,

$$\sqrt{m} \left(\hat{\eta} - \eta \right) \xrightarrow{d} \frac{1}{2} W \tag{3.19}$$

where W is a standard normal random variable.

Further periodogram-based methods are, e.g., the *log-periodogram regression* (Geweke and Porter-Hudak, 1983 [150]), and *variants of the Whittle and local Whittle estimator*, see, e.g., Dahlhaus (1988, 1990) [97], [99], Velasco (1999) [397], Shimotsu and Phillips (2005) [363], Abadir et al. (2007) [1], and Shao (2010) [358]. For a review on semiparametric aproaches, see, e.g., Moulines and Soulier (2003) [302]. Furthermore, Sykulski et al. (2019) [378] recently proposed the so-called *debiased Whittle likelihood* accounting for the finite sample bias in the periodogram.

3.2.2 Model selection

This second paragraph addresses the choice of the "*best model*" among a finite set of certain model candidates using a certain model selection criterion. In general, such model selection criteria measure the quality of each model relative to all other models in terms of the relative amount of lost information. The commonly used ones are the *Akaike's information criterion (AIC)* and the *Bayesian information criterion (BIC)*, and as an alternative to these two, the *Hannan-Quinn information criterion (HIC)*. For a discussion on these three criteria in the context of iid or short-memory time series models see, e.g., Akaike (1969a, 1973) [4] [5], Parzen (1974) [313], Shibata (1976, 1980) [360], [361], Schwarz (1978) [352], Hannan and Quinn (1979) [190], Hannan (1980) [189], and Quinn (1980) [332]. For model selection under the assumption of long-range dependence, see, e.g., Beran (1989) [21] and Beran et al. (1998, 1999) [26], [31].

Besides, there are many more model selection criteria depending on the objective, such as model selection for inference vs. for prediction, and on the framework, such as parametric vs. nonparametric. For an overview of different model selection techniques see, e.g., Ding et al. (2018) [111] and references therein.

In the following, we shortly review the model selection procedure and the three model selection criteria mentioned above in the context of fitting a fractional ARIMA(p,d,0)-model, along the same lines as in Beran et al. (2013) [33], since we use this procedure in our data examples (see Chapters 4.5 and 7.6).

Let (X_t) be a fractional ARIMA(p^0,d,0) process with unknown autoregressive order p^0 and unknown AR-coefficients $(\phi_1^0, \ldots, \phi_p^0) \geq 0$, and let $\eta^0 = (\sigma_{\epsilon,0}^2, d^0, p^0)$ denote the $(p+2)$-dimensional *true parameter vector*, where $\sigma_{\epsilon,0}$ is the unknown variance of the underlying white noise process (ϵ_t). We consider another process (Y_t), independent of (X_t), with $Y_t \stackrel{d}{=} X_t$. Based on a sample $X = (X_1, \ldots, X_n)$, for some order $p \geq p^0$, let $\hat{\eta}(X, p)$ denote the estimator for $\eta(p)$ for a fractional ARIMA(p,d,0)-process. Then, given random samples X and $Y = (Y_1, \ldots, Y_n)$, the quality of the fit is measured by minimizing the risk function

$$R(p) = E\left[L\left(p; \hat{\eta}(X, p)\right)\right],$$

where the loss function L is given by

$$L\left(p; \hat{\eta}(X, p)\right) = -2E\left[\mathcal{L}(Y, \hat{\eta}(X, p))\right],$$

and $\mathcal{L}$ is some approximate log-likelihood function, e.g., the approximate Whittle log-likelihood given in (3.17).

This minimization leads to model selection criteria (with respect to some $\alpha > 0$) of the form

$$AIC_\alpha(p) = n \ln\left(\hat{\sigma}_\epsilon^2\right) + \alpha \cdot (p + 2),$$

and the selected order is obtained as

$$\hat{p} = \underset{0 \leq p \leq p_{\max}}{argmin} \left\{AIC_\alpha(p)\right\},$$

where $p_{\max}$ is a predefined maximal order of the models that are used for comparison. To avoid overfitting, the parameter α accounts for *penalizing the number of estimated parameters*. This means, the notion $AIC_\alpha(p)$ generalizes the three commonly used criteria: $\alpha = 2$ yields the AIC, $\alpha = \ln(n)$ corresponds to the BIC, and for $\alpha = 2c \ln(\ln(n))$ with $c > 1$, we obtain the HIC.

Remarks. *(i) The AIC is asymptotically efficient for the nonparametric framework, and the BIC is consistent and asymptotically efficient for the parametric framework. The HIC is asymptotically equivalent to BIC, but although it ensures strong consistency for the parametric framework, it is only rarely used in practice (see, e.g., Burnham and Anderson (2002, 2004) [71], [72]).*

(ii) The AIC is suitable when one is interested in model prediction and the BIC is more appropriate when the true model is included in the set of model candidates.

(iii) Recently, a new model selection criterion, the Bridge criterion (BC), has been proposed by Ding et al. (2016, 2017) [109], [110] aiming to bridge the asymptotic properties of the AIC and the BIC.

Chapter 4

Testing the number of exceedances

In this part's last chapter we consider a seasonal time series under the assumption of long-range dependence and address the question whether there exists a *change in the seasonal exceedance probability*. All these technical results are published in Beran, Steffens and Ghosh (2021b) [39]. Figures, simulations, and the test procedure were conducted in R 4.1.0, R Core Team.

As a possible application we look at daily average temperatures at Konstanz over a time period of 10 years (provided by DWD, `https://opendata.dwd.de/`) displayed in Figure 4.1.

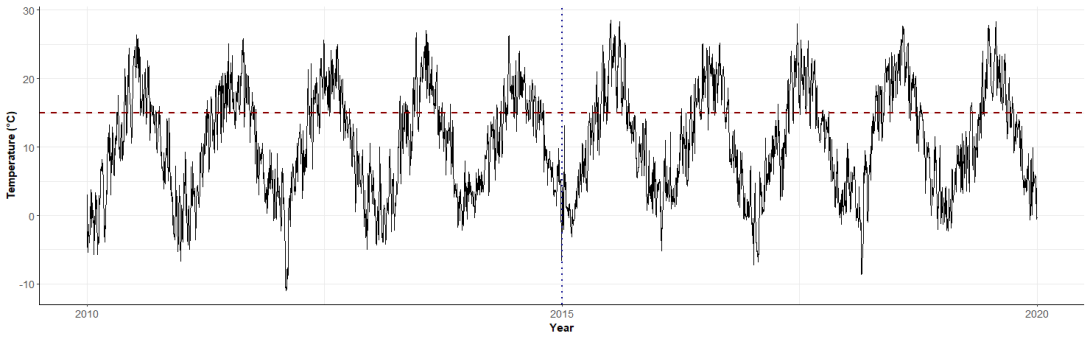

Figure 4.1: Daily average temperature series for Konstanz between 2010 and 2020 together with an arbitrary temperature threshold of 15 degrees in dashed red line. The dotted blue line illustrates the day (January 1, 2015) at which, for comparison, we split the whole time period of 10 years into two smaller periods of 5 years each.

We are interested in a change in the expected number of days at which the daily temperature exceeds the 15 degrees (dashed red line) comparing, for instance, the first 5 years to the last 5 years, illustrated by the dotted blue line. In addition, for classifying this in the context of climate change, one may only focus on the temperatures within a certain range in time, e.g., the

month of April as the beginning of spring. In the applied literature, there are several recently published works discussing this or similar questions in order to investigate, e.g., whether there is a shift in the seasons, whether there is a change in the duration of the season or even whether some of the seasons seem to disappear, see, for instance, Wang et al. (2021) [408] and references therein.

Back to the opening question: As we are testing for a change in the expected number of seasonal exceedances, we define our test statistic in terms of *periodic partial empirical distribution functions*. We derive the asymptotic distribution of this statistic by applying the *uniform weak reduction principle* which we prove for so-called *periodic processes*, and, based on this, establish rejection regions for our hypothesis test. A data example with temperatures at different locations in Germany that differ in geographical area and/or altitude illustrates the technical results. Here, for estimating the seasonal component, we apply trigonometric regression (as introduced in Chapter 3.1.2 in the context of locally weighted regression), and to account for long-memory behavior, we fit a fractional ARIMA(p,d,0)-process of appropriate order p and estimate the unknown parameters using Whittle estimation combined with the BIC selection criterion, which is addressed in Chapter 3.2.

4.1 Short literature review

Before we introduce our framework in detail, we will have a short literature review on the topics that were not discussed in the previous chapters.

The *theory of exceedances over thresholds* can be viewed as a part of the extreme value theory, which has a long history, see, e.g., Gumbel (1958, 2004) [176], [177] for statistics of extremes, and Gumbel and von Schelling (1950) [178] and Sarkadi (1957) [347] for the distribution of the number of exceedances. Estimation and prediction of the exceedance probability are considered in different fields of application. These methods are, for instance, applied to examine possible factors that have an impact on climate change. Extreme value theory in the field of stochastic processes is considered, for instance, in Leadbetter and Rootzén (1988) [256] and Leadbetter et al. (2012) [255], and Davison and Smith (1990) [101], e.g., derived models for exceedances beyond high thresholds. A further related topic is the number of crossings, especially the *number of upcrossings*, which also is of great interest in the applications (see, e.g., Nordin and Rosbjerg (1970) [309] for upcrossings in hydrology). In the case of stochastic processes with continuous time index set, one considers *sojourns* and *excursions above high levels*, which is, e.g., discussed in the book of Berman (2017) [46] and references therein.

Testing for a change in the number of exceedances is a special topic in *change-point theory*. Change-point testing in general is considered in various papers. For instance, Kokoszka and

Leipus (1998) [247], Horvath et al. (1999) [212], and Vogelsang (1998, 1999) [398], [399] developed tests for a change in the mean. Approaches for detecting a change in the marginal distribution function of iid random variables is discussed in Csörgő and Horvath (1988) [85] and Brodsky and Darkhovsky (2013) [66]. Under the assumption of long-range dependence, several standard methods have been studied leading to main differences in asymptotic behavior of test statistics and/or estimators. Csörgő and Horvath (1997) [86], e.g., considered the CUSUM test and Dehling et al. (2013) [107] the Wilcoxon change-point test as examples for nonparametric tests. Giraitis et al. (1996) [160] studied the *Kolmogorov-Smirnov-type test statistic*, and Tewes (2017) [389], e.g., considered the same test statistic and the Cramer-von Mises-type statistic in the context of subordinated Gaussian long-memory processes. Moreover, there are papers on testing for a change in the long-memory parameter, see, e.g., Beran and Terrin (1996) [42] and Horvath and Shao (1999) [213]. For an overview and more references see, e.g., Giraitis et al. (2003) [156], Perron (2006) [315], and Beran et al. (2013) [33].

Last but not least, one important problem in the context of long-memory models is to distinguish between long-memory and structural breaks in short-memory models such as mean-in-time shifts. This is, e.g., discussed in Granger and Hyung (1999, 2004) [171], [172], Sibbertsen (2004) [365] and Shimotsu (2006) [362], and recently, several approaches on testing for single breaks in a long-memory model are published, see, e.g., Berkes et al. (2006) [45], Wang (2008) [406], Baek and Pipiras (2012) [14], Shao (2011) [359], Dehling et al. (2013) [107], and Betken (2016) [47]. Testing for multiple breaks is, e.g., considered in Lavielle and Moulines (2000) [254], and approaches in a multivariate setting are discussed in Sibbertsen et al. (2020) [366]. A survey is provided by Wenger et al. (2019) [414].

A common statistic in the context of change-point issues of the marginal distribution function is the **Kolmogorov-Smirnov-type statistic (K-S-type statistic)** which is defined by

$$T_N = \sup_{(\tau,x)\in[0,1]\times\mathbb{R}} \{V_N(\tau,x)\},$$

with the two-parameter process

$$V_N(\tau,x) := \frac{[N\tau]\,(N-[N\tau])}{N}\left(F_{[N\tau]}(x) - F^*_{N-[N\tau]}(x)\right).$$

Here, $F_{[N\tau]}$ is the partial empirical distribution function of the first $[N\tau]$ variables and $F^*_{N-[N\tau]}$ is the partial empirical distribution function of the last $N-[N\tau]$ variables. This statistic is, e.g., discussed in former works of Picard (1985) [321], Deshayes and Picard (1986) [108], Hawkins (1988) [198], Leipus (1988) [258], Szyszkowicz (1994) [379], and Csörgő and Szyszkowicz (1994) [95]. It is shown, for instance, in Giraitis et al. (1996) [160] that the two-parameter

process $V_N(\tau, x)$ (suitably normalized) converges to a random field $W(\tau, x)$, which coincides with the *Kiefer process* in the iid situation and with a degenerate Gaussian field under the assumption of long-memory. Hence, a general asymptotic approach is based on the asymptotic behavior of two-parameter empirical processes and makes use of the respective uniform weak reduction principle. Besides the above mentioned references, general limit theorems are established in Yao (1987) [425], Dumbgen (1991) [126], Giraitis and Leipus (1992) [159], and, particularly for long-range dependent variables, in Ray and Tsay (2002) [335], Wang (2007) [404], and Dehling et al. (2013) [107].

As we will see in the subsequent chapter, for testing for a change in the number of seasonal exceedances, we define such a K-S-type test statistic in the framework of seasonal time series models with stationary long-range dependent residuals.

4.2 Framework and notations

We consider the seasonal time series introduced in (3.12) of the form

$$Y_t = \mu_t + S_t + Z_t, \qquad t \in \mathbb{N}, \tag{4.1}$$

where (Z_t) is a stationary zero-mean error process with finite variance, μ_t is a non-periodic long-term trend and S_t is an exactly periodic seasonal component with period $T_0 \geq 2$, i.e., for identifiability (see (3.13)),

$$\sum_{t=1}^{T_0} S_t = 0 \qquad \text{and} \qquad S_{t+T_0} = S_t.$$

(Z_t) is assumed to be a weakly stationary linear process for the form (1.16) with long-range dependence, i.e.,

$$Z_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \qquad t \in \mathbb{N}, \tag{4.2}$$

where (ε_t) is a zero mean white-noise process with finite variance $\sigma_\varepsilon^2 = var(\varepsilon_t)$, and the coefficients a_j, $j = 1, 2, \ldots$, are such that

$$a_j \underset{j\to\infty}{\sim} c_a j^{d-1}, \qquad d \in \left(0, \frac{1}{2}\right) \quad \text{and} \quad 0 < c_a < \infty. \tag{4.3}$$

This special form of the coefficients a_j has been prementioned in (1.17). Due to this asymptotic representation, we know from Corollary 1.2.3 that the spectral density of (Z_t) has a pole at the

origin of the form

$$f_Z(\lambda) \underset{\lambda \to 0}{\sim} c_f |\lambda|^{-2d}, \qquad 0 < c_f < \infty, \tag{4.4}$$

and the autocovariances are slowly decaying according to

$$\gamma_Z(k) \underset{k \to \infty}{\sim} c_\gamma |k|^{2d-1}, \qquad 0 < c_\gamma < \infty. \tag{4.5}$$

Our assumptions can be shortly summarized as follows:

(A1) (Y_t) is a seasonal time series of the form (4.1) with a deterministic and exactly periodic seasonal component S_t with period $T_0 \geq 2$, and under H_0, the long-term trend μ_t is constant, i.e., $\mu_t = \mu$.

(A2) (Z_t) in (4.1) is a weakly stationary linear long-memory process of the form (4.2) with coefficients of the form (4.3) satisfying the properties (4.4) and (4.5).

In order to formulate a test for a *change in the expected number of seasonal exceedances*, we introduce so-called **periodic processes** and associated definitions and notations.

Let (Z_t) be a linear process and let $n \in \mathbb{N}$ denote the time range. For period $T_0 \geq 2$, we define $N := \left[\frac{n}{T_0}\right]$ as the **number of periods**. Now, the idea is, after dividing the whole time range into N sub-periods, to look at *fixed periodically repeated time points* only.

Definition 4.2.1. *(i) We call a time point $t_j \in \{1, \ldots, T_0\}$, $j = 1, \ldots, k$, $k \leq T_0$, a **seasonal time point**.*

*(ii) For a fixed seasonal time point t_j and $i = 1, \ldots, N$, we define the **periodic process** $Z_{t_j,i}$ as*

$$Z_{t_j,i} := Z_{t_j+(i-1)T_0}. \tag{4.6}$$

(iii) Let $(Z_{t_j,i})$ be a periodic process and let $t_1 < \ldots < t_k$ be seasonal time points. Then, for some $0 < c < \infty$, we define

$$\chi_i = \sum_{j=1}^{k} \mathbb{1}\left\{Z_{t_j,i} > c\right\}$$

*as the **number of seasonal exceedances** of some level c in period number i at periodically repeated time points t_j.*

Definition 4.2.2. *Let $\tau \in [0, 1]$. For a time series (X_t) and the corresponding periodic process $(X_{t_j,i})$ associated with seasonal time points $t_1 < \ldots < t_k$, we further define*

(i) the ***periodic sample mean*** *as*

$$\bar{x}_{N,t_j} \quad := \quad \frac{1}{N}\sum_{i=1}^{N} X_{t_j,i}$$

(ii) the ***periodic partial empirical distribution function*** *of* (X_t) *as*

$$F_{X;N,t_j}(\tau, x) \quad := \quad \frac{1}{[N\tau]}\sum_{i=1}^{[N\tau]} \mathbb{1}\left\{X_{t_j,i} \leq x\right\}$$

(iii) and the ***complementary periodic partial empirical distribution function*** *of* (X_t) *as*

$$F^*_{X;N,t_j}(\tau, x) \quad := \quad \frac{1}{N-[N\tau]}\sum_{i=[N\tau]+1}^{N} \mathbb{1}\left\{X_{t_j,i} \leq x\right\}.$$

(iv) For the right tail distributions, i.e., for

$$\frac{1}{N}\sum_{i=1}^{N} \mathbb{1}\left\{X_{t_j,i} \geq x\right\} \qquad \textit{and} \qquad \frac{1}{N-[N\tau]}\sum_{i=[N\tau]+1}^{N} \mathbb{1}\left\{X_{t_j,i} \geq x\right\},$$

we always write

$$\bar{F}_{X;N,t_j} = 1 - F_{X;N,t_j} \qquad \textit{and} \qquad \bar{F}^*_{X;N,t_j} = 1 - F^*_{X;N,t_j},$$

respectively.

(v) For $\tau = 1$*, we obtain the* ***periodic empirical distribution function****, and instead of* $F_{X;N,t_j}(1, x)$*, we write* $F_{X;N,t_j}(x)$ *for short, since it no longer depends on* τ*.*

This means, calculating the periodic sample mean $\bar{x}_{N,t_j}$ corresponds to taking the mean over all periods N for fixed seasonal time points $t_j \in \{1, \ldots, T_0\}$, and the periodic empirical distribution function $F_{X;N,t_j}(x)$ counts the periods for which, at a fixed time point $t_j \in \{1, \ldots, T_0\}$, the periodic process is below some value x.

Remarks. *(i) We made a difference in the terminologies "seasonal" and "periodic" in the last two Definitions to emphasize the two ways of building these terms. This means, e.g., the seasonal exceedances (Definition 4.2.1 (iii)) are defined as the number of exceedances within one period or rather within one "season", and the periodic partial empirical distributions functions (Definition 4.2.2 (ii)) are defined at fixed seasonal time points summed up over all periods.*

(ii) The notion „complementary“ in Definition 4.2.2 (iii) refers to the empirical distribution of just the last $N - [N\tau]$ observations.

We are interested in testing for a *change in the expected number of seasonal exceedances*. This means, our goal is to test the null hypothesis

$$H_0 : \; E[\chi_1] = E[\chi_2] = \ldots = E[\chi_N]$$

against the alternative

H_1: there is at least one i $(1 \leq i \leq N-1)$ such that either for all $l \leq i$ and for all $m \geq i+1$, $E[\chi_l]$ is smaller than $E[\chi_m]$, or for all l, m, $E[\chi_l]$ is larger than $E[\chi_m]$.

As mentioned above, a natural statistic for this is the *K-S-type test statistic* (4.1). Since we are interested in the number of exceedances at several fixed seasonal time points, we use this statistic in a slightly different version. Our resulting test statistic is given by

$$M := \sup_{\tau \in (0,1)} V(\tau), \tag{4.7}$$

where

$$V(\tau) := \sum_{j=1}^{k} D_{j,N}(\tau) \tag{4.8}$$

and

$$\begin{aligned} D_{j,N}(\tau) &:= \frac{[N\tau](N-[N\tau])}{N} \left[\bar{F}^*_{Y;N,t_j}(\tau, c) - \bar{F}_{Y;N,t_j}(\tau, c)\right] \\ &= \frac{[N\tau](N-[N\tau])}{N} \left[F_{Z;N,t_j}\left(\tau, c-\mu-S_{t_j}\right) - F^*_{Z;N,t_j}\left(\tau, c-\mu-S_{t_j}\right)\right]. \end{aligned}$$

The idea behind this test statistic is the following: We look at a portion $\tau \in (0,1)$ of all N periods. In order to detect a change in the number χ_l of seasonal exceedances at one fixed seasonal time point t, we compare the value of χ_l before and after $[N\tau]$ periods by summing up the differences of the upper and the lower average number of exceedances (suitably normalized). This is where the periodic partial empirical distribution functions come in. If, in addition, we have several seasonal time points t_j at which we wish to study the occurrence of exceedances, we have to sum up all these differences with respect to j. Finally, we choose our test statistic as the supremum of the latter sum with respect to portions τ. In other words, we take that portion τ that yields the maximum average differences in the number of exceedances.

Consequently, in order to derive asymptotic rejection regions of our test statistic, we study the asymptotic distributions of $V(\tau)$ and M, and therefore investigate the asymptotic behavior of the periodic empirical distribution functions.

4.3 Uniform weak reduction principle for periodic processes

The investigation of the asymptotic behavior of periodic empirical distribution functions is based on the *uniform weak reduction principle* from Dehling and Taqqu (1988) [105] for Gaussian subordinated processes (see Proposition 2.2.4), which we extend to periodic processes. Recall the results for Gaussian subordinated (Chapter 2.2.1) and linear processes (Chapter 2.2.2). We will show that the limiting distribution is in principle the same as in the case of linear processes, where the convergence rate is adapted to the periodic nature of the process.

A first important result concerns the asymptotic behavior of the autocovariance function of a periodic process.

Proposition 4.3.1. *Under Assumption (A2), any periodic process $(Z_{t,i})$ of the form (4.6) is a stationary linear long-memory process with covariances*

$$cov(Z_{t,i}, Z_{t,i+k}) \underset{k\to\infty}{\sim} c_\gamma |kT_0|^{2d-1}.$$

Proof. This follows directly from the definition of periodic processes and the asymptotic behavior of the autocovariance function of the underlying linear process (Z_s), since

$$\begin{aligned} \operatorname{cov}(Z_{t,i}, Z_{t,i+k}) &= \operatorname{cov}\Big(Z_{t+(i-1)T_0}, Z_{t+((i+k)-1)T_0}\Big) \\ &= \gamma(kT_0) \\ &\underset{k\to\infty}{\sim} c_\gamma |kT_0|^{2d-1}. \end{aligned}$$

■

This theorem hypothesizes that establishing a *uniform weak reduction principle for periodic processes* (and a *FCLT*) can be ascribed to the uniform weak reduction principle for linear processes given in Chapter 2.2.2, Theorem 2.2.9.

In order to link the periodic empirical process with the periodic sample mean, we first study the asymptotic behavior of periodic sample means under Assumptions (A1) and (A2).

Let (Z_t) be as in (A2). Recall that, under the assumption of long-range dependence, the common sample mean $\bar{z} = \frac{1}{n}\sum_{s=1}^{n} Z_s$ converges weakly towards a degenerate Gaussian random variable (see (2.10) in Example 2.2.7 (ii)), i.e.,

$$n^{\frac{1}{2}-d}\bar{z} \xrightarrow{d} \sqrt{c_f v(d)}W, \tag{4.9}$$

where $W \sim N(0,1)$, c_f corresponds to the spectral density of (Z_t) and $v(d) = 2\Gamma(1-2d)\sin(\pi d)$.

For long-range dependent periodic processes, we obtain a very similar result:

Proposition 4.3.2. *Let $(Z_{t,i})$ be a periodic process of the form (4.6). Then, under the given assumptions on the underlying process (Z_s), we obtain the following convergence of the periodic sample mean $\bar{z}_{N,t}$:*

$$N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\bar{z}_{N,t} \xrightarrow{d} \sqrt{c_f v(d)}W. \tag{4.10}$$

Moreover,

$$N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\left(\bar{z}_{N,1},\ldots,\bar{z}_{N,T_0}\right) \xrightarrow{d} \sqrt{c_f v(d)}W\cdot(1,\ldots,1)^T, \tag{4.11}$$

where $W\sim N(0,1)$, c_f corresponds to the spectral density of (Z_t), and $v(d)=2\Gamma(1-2d)\sin(\pi d)$.

Proof. The first part of the theorem, (4.10), directly follows from (4.9) together with the asymptotic behavior of the autocovariance function of a periodic long-memory process given in Proposition 4.3.1.

For the second part, (4.11), we need to show convergence in distribution for each vector component. This convergence holds, in particular, if we show it for any linear combination of these components. Therefore, let $\alpha=(\alpha_1,\ldots,\alpha_{T_0})^T\in\mathbb{R}^{T_0}$. Then, for any α_t and $\bar{z}_{N,t}$, $t=1,\ldots,T_0$, it is sufficient to show that

$$N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\sum_{t=1}^{T_0}\alpha_t\bar{z}_{N,t} \xrightarrow{d} \sqrt{c_f v(d)}\sum_{t=1}^{T_0}\alpha_t\cdot W.$$

Note that $\sum_{t=1}^{T_0}\alpha_t\bar{z}_{N,t}$ can be decomposed into

$$\alpha_1(\bar{z}_{N,1}-\bar{z}_{N,2})+\ldots+(\alpha_1+\ldots+\alpha_{T_0-1})(\bar{z}_{N,T_0-1}-\bar{z}_{N,T_0})+\alpha_{T_0}\bar{z}_{N,T_0}.$$

Since the convergence of the last term $\alpha_{T_0}\bar{z}_{N,T_0}$ is given by the first part of the theorem, it remains to show that each of the differences converges towards 0 in distribution.

For simplicity, we present the proof only for the first term, i.e., $\alpha_t=0$ for $t\geq 3$, $\alpha_1=1$ and $\alpha_2=-1$, and show

$$N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\left(\bar{z}_{N,1}-\bar{z}_{N,2}\right) \xrightarrow{d} 0.$$

The same arguments then apply to the other terms.

For the variance, we obtain

$$\begin{aligned}\operatorname{var}(\bar{z}_{N,1}-\bar{z}_{N,2}) &= N^{-2}\operatorname{var}\left(\sum_{i=1}^{N}\left(Z_{1+(i-1)T_0}-Z_{2+(i-1)T_0}\right)\right)\\ &= 2N^{-2}\sum_{k=-(N-1)}^{N-1}(N-|k|)\left(\gamma_Z(kT_0)-\gamma_Z(kT_0+1)\right).\end{aligned}$$

Consequently, with the asymptotic behavior of the autocovariances and using the Riemann

sum, we obtain

$$\begin{aligned} N^{-2} \sum_{k=-(N-1)}^{N-1} (N-|k|)\gamma_Z(kT_0) &= T_0^{2d-1} N^{-2} c_\gamma \sum_{\substack{k=-(N-1) \\ k\neq 0}}^{N-1} (N-|k|)|k|^{2d-1} + o(N^{2d-1}) \\ &= T_0^{2d-1} N^{2d-1} c_f v(d) + o(N^{2d-1}) \end{aligned}$$

and

$$\begin{aligned} N^{-2} \sum_{k=-(N-1)}^{N-1} (N-|k|)\gamma_Z(kT_0+1) &= T_0^{2d-1} N^{-2} c_\gamma \sum_{\substack{k=-(N-1) \\ k\neq 0}}^{N-1} (N-|k|)|k+T_0^{-1}|^{2d-1} + o(N^{2d-1}) \\ &= T_0^{2d-1} N^{-2} c_\gamma \sum_{\substack{k=-(N-1) \\ k\neq 0}}^{N-1} (N-|k|)|k|^{2d-1} + o(N^{2d-1}) \\ &= T_0^{2d-1} N^{2d-1} c_f v(d) + o(N^{2d-1}). \end{aligned}$$

Hence

$$N^{1-2d} T_0^{1-2d} \mathrm{var}(\bar{z}_{N,1} - \bar{z}_{N,2}) \xrightarrow{P} 0$$

and therefore

$$N^{\frac{1}{2}-d} T_0^{\frac{1}{2}-d} (\bar{z}_{N,1} - \bar{z}_{N,2}) \xrightarrow{d} 0.$$

Finally, applying (4.10), the result follows.

■

This theorem states that periodic sample means converge weakly towards a degenerate or perfectly correlated Gaussian random vector with marginal distributions independent of the seasonal time point $t \in \{1, \ldots, T_0\}$. This is a special behavior in the context of long-memory processes and the usual asymptotic behavior of long-range dependent sample means. Note that this is in contrast to the iid and the short-range dependent case, where the sample means are *asymptotically independent* or *asymptotically uncorrelated* random variables. See also the discussion in Chapter 2.

Extending the results of the *uniform weak reduction principle* for linear processes given in Theorem 2.2.8 to periodic processes yields the following two theorems:

Theorem 4.3.3. ***(Uniform weak reduction principle for periodic processes)***

Suppose that the assumptions of Proposition 4.3.2 hold. Let $\delta > 0$ and denote by $p_Z = F_Z'$ the marginal density function of the linear process (Z_s). Then

$$\lim_{n\to\infty} P\left(N^{\frac{1}{2}-d} \max_{t=1,\ldots,T_0} \sup_{z\in\mathbb{R}} |F_{Z;N,t}(z) - F_Z(z) + p_Z(z)\bar{z}_{N,t}| > \delta \right) = 0$$

and

$$\lim_{n\to\infty} P\left(N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\max_{t=1,\ldots,T_0}\sup_{z\in\mathbb{R}}\left|F_{Z;N,t}(z)-F_Z(z)+p_Z(z)\sqrt{c_f v(d)}W\right|>\delta\right)=0,$$

where $W\sim N(0,1)$, and c_f and $v(d)$ are as in Proposition 4.3.2.

Proof. Recall from Proposition 4.3.1 that for $t\in\{1,\ldots,T_0\}$ and $N=\left[\frac{n}{T_0}\right]$, $Z_{t+(i-1)T_0}$ is a stationary linear long-memory process with autocovariance function

$$\mathrm{cov}(Z_i,Z_{i+k})=\gamma_Z(kT_0)\underset{k\to\infty}{\sim}c_\gamma|T_0^{2d-1}k^{2d-1}|.$$

The uniform weak reduction principle for linear processes (Theorem 2.2.8) implies for any $0<\delta\leq 1$:

$$\lim_{n\to\infty} P\left(N^{\frac{1}{2}-d}\sup_{z\in\mathbb{R}}|F_{Z;N,t}(z)-F_Z(z)+p_Z(z)\bar{z}_{N,t}|>\delta\right)=0.$$

Since T_0 is finite, we obtain

$$\lim_{n\to\infty} P\left(N^{\frac{1}{2}-d}\max_{t=1,\ldots,T_0}\sup_{z\in\mathbb{R}}|F_{Z;N,t}(z)-F_Z(z)+p_Z(z)\bar{z}_{N,t}|>\delta\right)=0.$$

Finally, applying Proposition 4.3.2 results in

$$\lim_{n\to\infty} P\left(N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\max_{t=1,\ldots,T_0}\sup_{z\in\mathbb{R}}|F_{Z;N,t}(z)-F_Z(z)+p_Z(z)\sqrt{c_f v(d)}W|>\delta\right)=0.$$

■

So far, we have proven *uniform weak convergence for periodic empirical processes*. For establishing a *(multivariate) FCLT*, we introduce the following notations:

For $\underline{\mathbf{t}}=(t_1,\ldots,t_k)^T$, $\underline{\mathbf{z}}=(z_1,\ldots,z_k)^T$ and $\tau\in[0,1]$, let

$$\begin{aligned}
F_{Z;N,\underline{\mathbf{t}}}(\tau,\underline{\mathbf{z}}) &:= \left(F_{Z;N,t_1}(\tau,z_1),\ldots,F_{Z;N,t_k}(\tau,z_k)\right),\\
F_Z(\underline{\mathbf{z}}) &:= \left(F_Z(z_1),\ldots,F_Z(z_k)\right),\\
p_Z(\underline{\mathbf{z}}) &:= \left(p_Z(z_1),\ldots,p_Z(z_k)\right)=\left(F_Z'(z_1),\ldots,F_Z'(z_k)\right),
\end{aligned}$$

and $I_k=[0,1]\times\mathbb{R}^k$.

Based on these notations, we can state a FCLT for a sequence of periodic empirical processes $F_{Z;N,\underline{\mathbf{t}}}(\tau,\underline{\mathbf{z}})-F_Z(\underline{\mathbf{z}})$:

Theorem 4.3.4. ***(FCLT for periodic empirical processes)***

Let

$$S_N(\tau,\underline{\boldsymbol{z}}):=N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\left[F_{Z;N,\underline{\mathbf{t}}}(\tau,\underline{\boldsymbol{z}})-F_Z(\underline{\boldsymbol{z}})\right],\qquad(\tau,\underline{\boldsymbol{z}})\in I_k.$$

Then, under the same assumptions as in Proposition 4.3.2, the following convergences hold:

$$S_N(\tau, z) \xrightarrow{w, D(I_1)} p_Z(z)\sqrt{c_f v(d)}B_H(\tau), \qquad (\tau, z) \in I_1, \tag{4.12}$$

and moreover,

$$S_N(\tau, \underline{z}) \xrightarrow{w, D(I_k)} p_Z(\underline{z})\sqrt{c_f v(d)}B_H(\tau), \qquad (\tau, \underline{z}) \in I_k, \tag{4.13}$$

where $B_H(\tau)$, $\tau in [0,1]$, denotes fractional Brownian Motion with $H = d + \frac{1}{2}$, and c_f and $v(d)$ are as in Proposition 4.3.2.

Proof. Let $t_j \in \{1, \ldots, T_0\}$ and $\tau \in [0,1]$. As $Z_{t+(i-1)T_0}$ is a linear process with autocovariance function

$$\mathrm{cov}(Z_i, Z_{i+k}) = \gamma_Z(kT_0) \underset{k\to\infty}{\sim} c_\gamma |T_0^{2d-1} k^{2d-1}|,$$

the uniform weak reduction principle for periodic empirical processes given in Theorem 4.3.3 (together with the respective FCLT for linear processes given in Theorem 2.2.9) implies

$$N^{\frac{1}{2}-d} T_0^{\frac{1}{2}-d} \left[F_{Z;N,t}(\tau, z) - F_Z(z)\right] \xrightarrow{w, D(I_1)} -p_Z(z)\sqrt{c_f v(d)}B_H(\tau),$$

where $I_1 = [0,1] \times \mathbb{R}$. And with Lemma 1.2.19 (iii), i.e., since $-B_H(\tau)$ again is a fractional Brownian Motion, we obtain weak convergence in $D(I_1)$ (4.12).

For $\underline{\mathbf{t}}$ and $F_{Z;N,\underline{\mathbf{t}}}(\tau, \underline{\mathbf{z}})$, due to degeneracy, the multivariate version of weak convergence of periodic sample means (see Proposition 4.3.2) together with the uniform weak reduction principle for periodic processes (see Theorem 4.3.3) and following the same arguments as in Giraitis et al. (1996b) [160] and Giraitis and Surgailis (1999) [165], signify that the fractional Brownian Motion with Hurst parameter $H = d + \frac{1}{2}$ is the same for all $j = 1, \ldots, k$, and consequently, (4.13) holds.

■

Summarizing the above, we have shown that the limiting behavior of long-range dependent periodic empirical processes is in principle the same as of long-range dependent linear empirical processes, since $N = \left[\frac{n}{T_0}\right]$ and therefore $[NT_0]^{\frac{1}{2}-d} = n^{\frac{1}{2}-d}$.

4.4 Testing for the expected number of exceedances

These results are applied to derive the asymptotic distribution of our K-S-type test statistic defined by (4.7) and (4.8).

Theorem 4.4.1. *Let $p_Z = F_Z'$ be the marginal density function of the linear process (Z_t). Then, under (A1) and (A2), we obtain weak convergence of the form*

$$N^{-(d+\frac{1}{2})}T_0^{\frac{1}{2}-d}V(\tau) \xrightarrow{w,D(I_k)} \left(\sqrt{c_f v(d)}\sum_{j=1}^{k} p_Z\left(c-\mu-S_{t_j}\right)\right)\cdot B_H^0(\tau)$$

and

$$N^{-(d+\frac{1}{2})}T_0^{\frac{1}{2}-d}M \xrightarrow{d} \left(\sqrt{c_f v(d)}\sum_{j=1}^{k} p_Z\left(c-\mu-S_{t_j}\right)\right)\cdot \sup_{\tau\in(0,1)} B_H^0(\tau),$$

where c_f and $v(d)$ are as in Proposition 4.3.2, and $(B_H^0(\tau))$ is a fractional Brownian Bridge with Hurst parameter $H = d + \frac{1}{2}$ defined in Definition 1.2.22.

Proof. Setting $c_j = c-\mu-S_{t_j}$ and $\underline{\mathbf{c}} = (c_1,\ldots,c_k)$, we can decompose $(D_{j,N}(\tau))$ into a sum of two empirical processes as follows (for each t_j):

$$\begin{aligned}
D_{j,N}(\tau) &= \frac{[N\tau](N-[N\tau])}{N}\left[F_{Z;N,t_j}(\tau,c_j) - F^*_{Z;N,t_j}(\tau,c_j)\right] \\
&= \frac{[N\tau](N-[N\tau])}{N}\left(\frac{1}{[N\tau]}\sum_{i=1}^{[N\tau]}\mathbb{1}\left\{Z_{t_j+(i-1)T_0}\le c_j\right\} - \frac{1}{N-[N\tau]}\sum_{i=[N\tau]+1}^{N}\mathbb{1}\left\{Z_{t_j+(i-1)T_0}\le c_j\right\}\right) \\
&= \left[\left(1-\frac{[N\tau]}{N}\right)\sum_{i=1}^{[N\tau]}\mathbb{1}\left\{Z_{t_j+(i-1)T_0}\le c_j\right\} - \frac{[N\tau]}{N}\sum_{i=[N\tau]+1}^{N}\mathbb{1}\left\{Z_{t_j+(i-1)T_0}\le c_j\right\}\right] \\
&= \left[\sum_{i=1}^{[N\tau]}\mathbb{1}\left\{Z_{t_j+(i-1)T_0}\le c_j\right\} - \frac{[N\tau]}{N}\sum_{i=1}^{N}\mathbb{1}\left\{Z_{t_j+(i-1)T_0}\le c_j\right\}\right] \\
&= N\left[F_{Z;N,t_j}(\tau,c_j) - \frac{[N\tau]}{N}F_{Z;N,t_j}(c_j)\right]
\end{aligned}$$

and further expand into

$$\frac{D_{j,N}(\tau)}{N} = \left(F_{Z;N,t_j}(\tau,c_j) - F_Z(c_j)\right) + \left(F_Z(c_j) - \frac{[N\tau]}{N}F_{Z;N,t_j}(c_j)\right).$$

Hence, we get

$$N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\left(\frac{1}{N}\sum_{j=1}^{k}D_{j,N}(\tau)\right) = N^{\frac{1}{2}-d}T_0^{\frac{1}{2}-d}\sum_{j=1}^{k}\left[\underbrace{\left(F_{Z;N,t_j}(\tau,c_j)-F_Z(c_j)\right)}_{(*)} + \underbrace{\left(F_Z(c_j)-\frac{[N\tau]}{N}F_{Z;N,t_j}(c_j)\right)}_{(**)}\right],$$

which is the sum of a periodic partial empirical process and a periodic empirical process.

Applying the FCLT, Theorem 4.3.4, to the two empirical processes now yields convergence to degenerate fractional Brownian Motion for (*) and convergence to a degenerate standard normal random variable for (**).

Recall that

$$V(\tau) = \sum_{j=1}^{k} D_{j,N}(\tau).$$

Finally, based on the construction of a fractional Brownian Bridge out of a fractional Brownian Motion given in Lemma 1.2.23, we obtain the following limiting distribution:

$$\begin{aligned} N^{-(d+\frac{1}{2})} T_0^{\frac{1}{2}-d} V(\tau) &= N^{\frac{1}{2}-d} T_0^{\frac{1}{2}-d} \left(\frac{1}{N} \sum_{j=1}^{k} D_{j,N} D_{j,N}(\tau) \right) \\ &\xrightarrow{w, D(I_k)} \left(\sqrt{c_f v(d)} \sum_{j=1}^{k} p_Z(c_j) \right) \cdot (B_H(\tau) - \tau \cdot B_H(1)) \\ &= \left(\sqrt{c_f v(d)} \sum_{j=1}^{k} p_Z(c_j) \right) \cdot B_H^0(\tau). \end{aligned}$$

■

Based on this result, we derive asymptotic rejection regions for our test statistic M:

For $\alpha \in (0,1)$, let $q_{1-\alpha, B_H^0}$ denote the $(1-\alpha)$-quantile of $\sup_{\tau \in (0,1)} B_H^0(\tau)$. Then H_0 is rejected at the level of significance α, if

$$M > N^{d+\frac{1}{2}} T_0^{d-\frac{1}{2}} \cdot \left(\sqrt{c_f v(d)} \sum_{j=1}^{k} p_Z(c_j) \right) \cdot q_{1-\alpha, B_H^0}. \tag{4.14}$$

In practice, for applying this (theoretical) result, we need to estimate several unknown parameters: Trigonometric regression can be applied for deseasonalization (see Chapter 3.1.2), kernel density estimation (discussed in Chapter 3.1.1) can be applied for estimating the marginal density p_Z, the long-memory parameters d and c_f can be obtained by applying variants of Whittle estimation together with fitting a fractional ARIMA-model by use of the BIC model selection criterion (see Chapter 3.2), and simulating a fractional Brownian Motion with a consistent estimate of the Hurst parameter $\hat{H} = \hat{d} + \frac{1}{2}$ yields the asymptotic $(1-\alpha)$-quantile of $B_{\hat{H}}^0(\tau)$, $\tau \in (0,1)$.

4.5 Data example: Daily temperature data

In this last chapter, we apply our test procedure and theoretical results to temperature data from different locations in Germany over a period of more than 45 years (provided by DWD, `https://opendata.dwd.de/`).

To address possible different long-memory and seasonal effects depending on the altitudes and the geographical area, we have chosen two mountainous locations, Brocken (Sachsen-Anhalt,

altitude 1141 m) and Zugspitze (Bayern, altitude 2962 m), one island in the North Sea, Norderney (Niedersachsen, altitude 12 m), one coastal location at the Baltic Sea, Rostock-Warnemünde (Mecklenburg-Vorpommern, 4 m) and two cities, Konstanz (Baden-Württemberg, altitude 428 m) and Düsseldorf (Nordrhein-Westfalen, altitude 37 m).

Due to seasonality, we assume that the seasonal component with period T_0 can by modeled by a Fourier series expansion of the form (3.15). The unknown coefficients could be estimated, e.g., by nonparametric regression, such as locally weighted regression, or by a simple parametric regression, as described in Chapter 3.1.2. We use the latter and fit a trigonometric series of order 3. Furthermore, for the sake of simplicity, we assume that $T_0 = 365$, which means, we omit the 29th of February in leap years.

The first figure, Figure 4.2, shows (as an example) daily average temperatures for Konstanz (upper two panels) and Zugspitze (lower two panels) between January 1st, 1973, and December 31st, 2021, together with the estimated seasonal component (red line) in (A) and (C), and the deseasonalized residual series in (B) and (D).

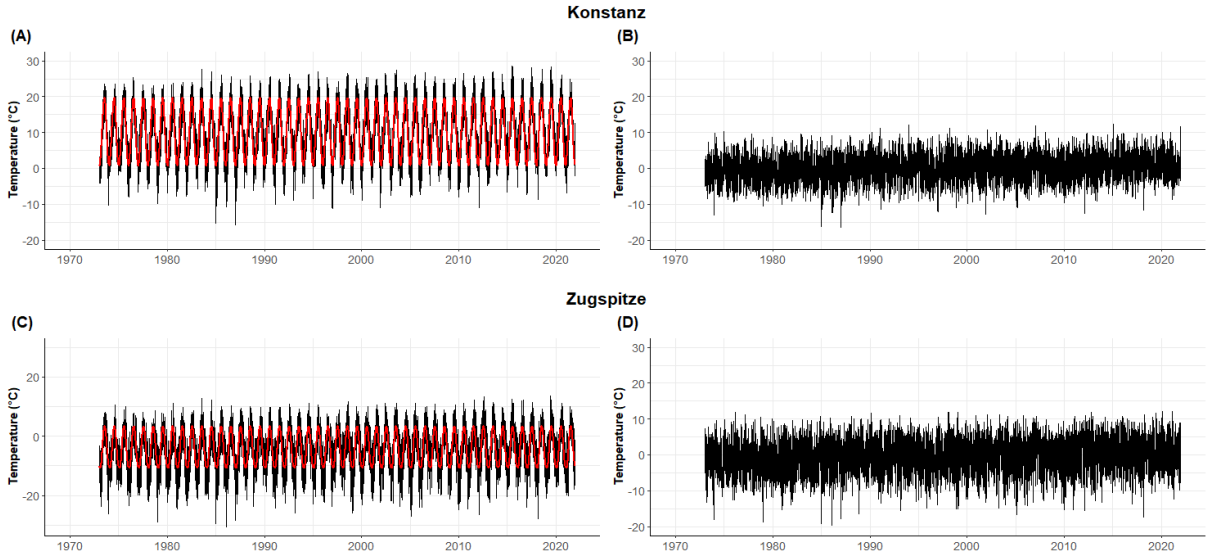

Figure 4.2: Temperature series for Konstanz and Zugspitze between 1973 and 2021. (A) & (C) display daily average temperatures together with estimated seasonal series in solid red line; (B) & (D) depict estimated deseasonalized residual series.

For testing for the number of seasonal exceedances, we consider both, daily April temperatures characterizing the beginning of spring and daily December temperatures characterizing the beginning of winter. Since we regard different climate zones, we expect differences not only between months but also between locations. The corresponding temperature boxplots in Figure 4.3 of

April temperature data and in Figure 4.4 of December temperature data, respectively, each split into time periods of about 10 years, 1973 to 1980, ..., 2011 to 2021, give a first impression of the change in temperatures over the years.

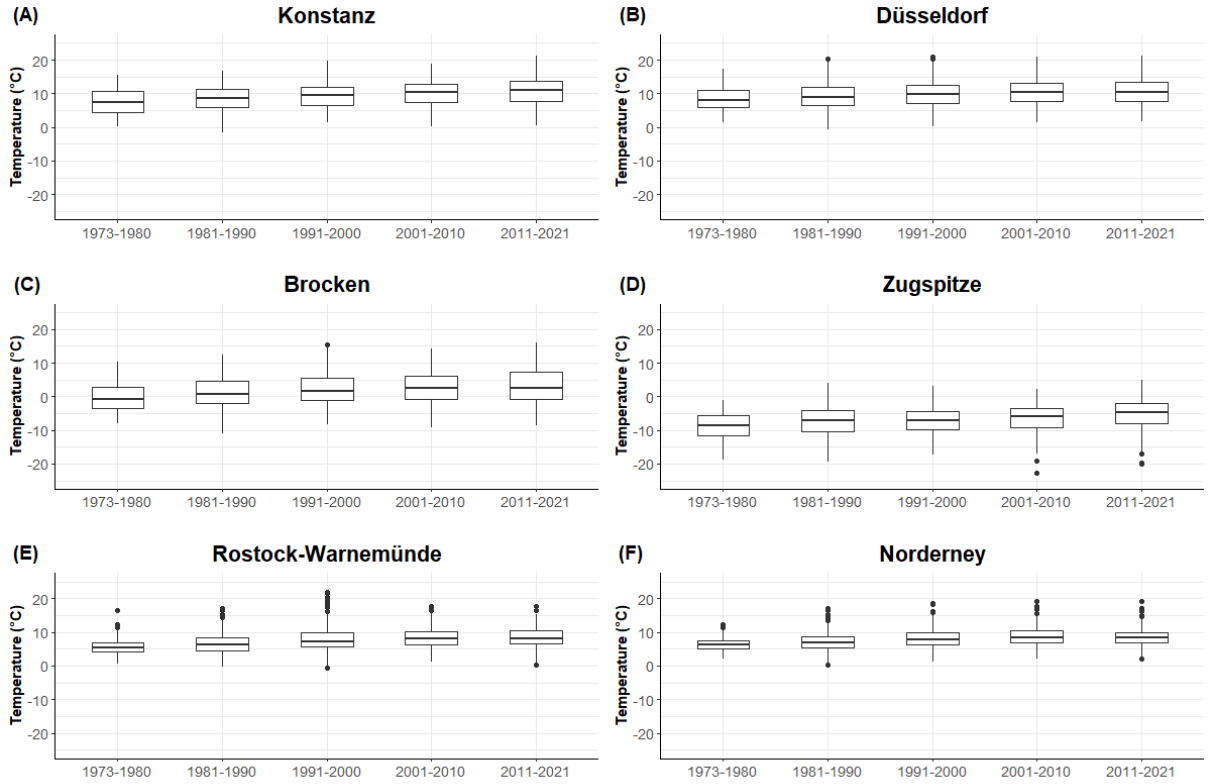

Figure 4.3: Boxplots of April temperature data at the six locations over time periods of about 10 years.

Particularly for April, median temperatures constantly increase, which is not the case for December temperatures.

In order to make results on our test procedure comparable among the different geographic positions and altitudes, we provide location- and month-specific temperature thresholds. To this end, we set c equal to the overall location- and month-specific 10%-quantile of temperatures, and obtain the following thresholds, displayed in Table 4.1.

	Konstanz	Düsseldorf	Brocken	Zugspitze	Rostock	Norderney
c_{Apr}	4	5	-3.9	-13.2	3.8	4.7
c_{Dec}	-2.62	-0.82	-7.7	-16.3	-2.22	-0.9

Table 4.1: Location- and month-specific thresholds c for April and December temperatures determined as the overall 10%-quantiles of April and of December temperatures, respectively.

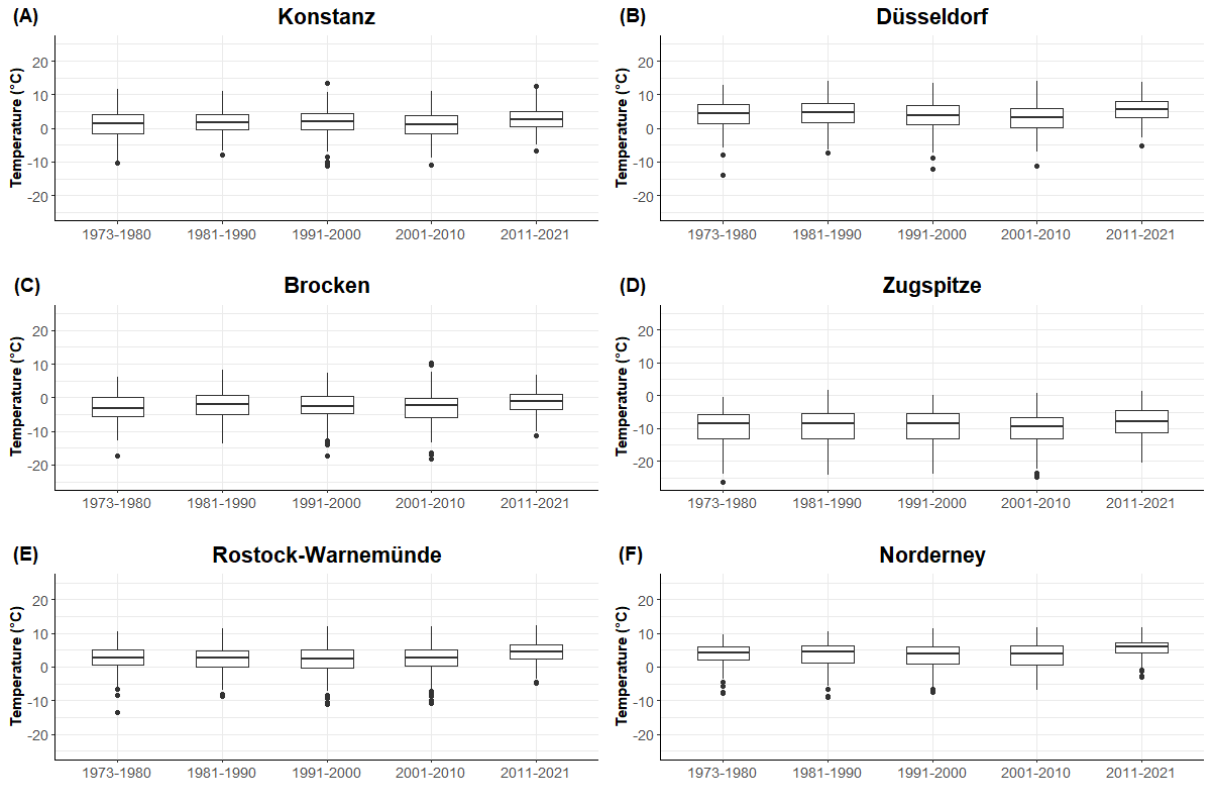

Figure 4.4: Boxplots of December temperature data at the six locations over time periods of about 10 years.

We are interested in a change in the expected number of seasonal exceedances. Hence, by testing the hypothesis

$$H_0 : \; E[\chi_1] = E[\chi_2] = \ldots = E[\chi_N]$$

against the alternative

$$H_1\text{: there is at least one } i \ (1 \leq i \leq N-1) \text{ such that either for all } l \leq i \text{ and for all } m \geq i+1,$$
$$E[\chi_l] \text{ is smaller than } E[\chi_m], \text{ or for all } l, m, \; E[\chi_l] \text{ is larger than } E[\chi_m]$$

at level of significance $\alpha = 0.05$ and $\alpha = 0.01$, we obtain the critical values from Theorem 4.4.1 and (4.14), i.e.,

$$c_{\text{crit};1-\alpha} = N^{d+\frac{1}{2}} T_0^{d-\frac{1}{2}} \cdot \left(\sqrt{c_f v(d)} \sum_{j=1}^{k} p_Z \left(c - \mu - S_{t_j} \right) \right) \cdot q_{1-\alpha, B^0_{H_Z}}, \tag{4.15}$$

where c_f, $v(d)$, p_Z and $q_{1-\alpha, B^0_{H_Z}}$ are as before, and H_0 is rejected if the value of the test statistic exceeds this critical value.

Several unknown parameters first have to be estimated. For each location, the long-memory

parameters c_f and d are estimated from the seasonally adjusted residual process $\hat{Z}_t = Y_t - \hat{S}_{t_j}$ by use of Whittle estimation together with the BIC selection criterion applied to fractional ARIMA(p,d,0)-models, $p = 0, 1, \ldots, [\log(n)]$. This approach is discussed in Chapter 3.2.

The long-memory estimates and the corresponding 95%-confidence intervals are given in Table 4.2.

	Konstanz	Düsseldorf	Brocken	Zugspitze	Rostock	Norderney
$\hat{d}$	0.12 [0.08, 0.17]	0.1 [0.05, 0.15]	0.1 [0.05, 0.15]	0.08 [0.03, 0.13]	0.24 [0.2, 0.28]	0.26 [0.22, 0.3]

Table 4.2: Location-specific long-memory estimates together with a 95%-confidence interval

Since all lower confidence bounds are greater than zero, the data at all six locations show long-range dependent behavior according to Remark 1.1. Interestingly, the long-memory parameters at the maritime locations are comparable and, in particular, are much larger than for the other locations. This might possibly be explained by the typical maritime climate. The ocean serves as a temperature buffer and yields only small variations in temperature over the whole year. The effect increases for islands, such as Norderney. On the other hand, the long-memory parameters for the mountainous locations, especially for Zugspitze, are low and similar to the parameters presented in the data example of our Paper (Beran, Steffens and Ghosh, 2021b [39]) with four locations in Switzerland.

Finally, the unknown $(1-\alpha)$-quantiles $q_{1-\alpha,B^0_{H_Z}}$ of (4.15) are obtained by simulating a fractional Brownian Bridge ($n = 1000$) with corresponding location-specific long-memory estimate $\hat{H} = \hat{d} + \frac{1}{2}$ given in Table 4.2, and taking the supremum.

We first consider the results for April temperatures. Here, the seasonal time points are between $t = 91$ and $t = 120$ per period, formally $t_i = t + (i-1) \cdot T_0$, with $t \in [91, 120]$ and $i = \mathbb{N}$, where $i = 1$ corresponds to the year 1973.

Figure 4.5 shows $V(\tau)$ (black points) for April temperatures at the different locations over the years with the critical limits $c^{\text{Apr}}_{crit;0.05}$ (solid blue line) and $c^{\text{Apr}}_{crit;0.01}$ (dashed red line). Each year x_i on the y-axis corresponds to the portion $\tau = \frac{x_i - 1973}{2021-1973}$.

We can observe that the expected number of exceedances increases throughout almost the whole period at the cities and the mountainous locations, with a maximum between 1985 and 1995 for the cities Konstanz (A) and Düsseldorf (B), and Zugspitze (D). For Brocken location (C) the maximum prevails a few years later than for the other locations. For the maritime locations Rostock-Warnemünde (E) and Norderney (F) an increase at 1% significance level is only observed over a time period of 20 and 15 years, respectively. This can possibly be related to the prevailing maritime climate.

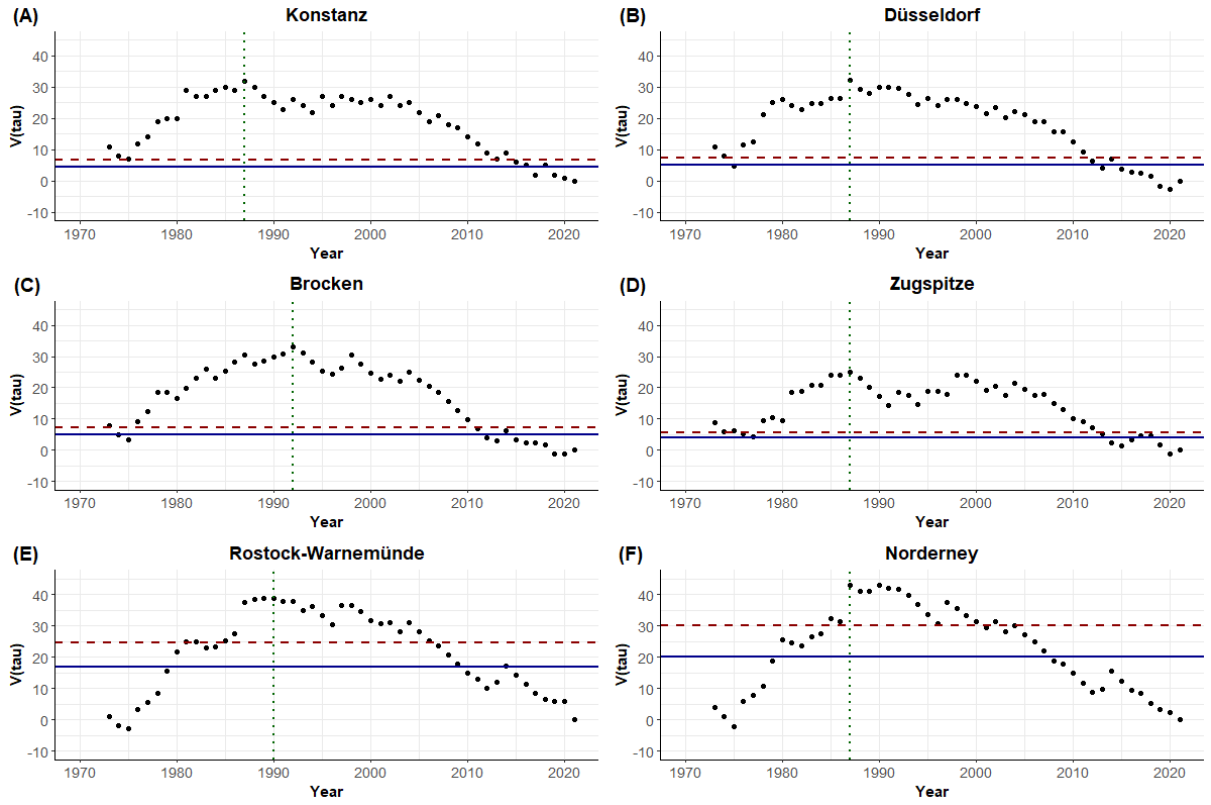

Figure 4.5: $V(\tau)$ (τ corresponds to the portion of elapsed years since 1973) as black points for daily April temperatures at the six locations with critical limit at 5% significance level in horizontal solid blue line and critical limit at 1% significance level in horizontal dashed red line. M as the supremum over τ is displayed with vertical dotted green line.

The results for December temperatures are quite different. Here, the time points are between and $t = 335$ and $t = 365$. Figure 4.6 displays $V(\tau)$ (black points) for December temperatures with critical limits $c^{\text{Dec}}_{crit;0.05}$ (solid blue line) and $c^{\text{Dec}}_{crit;0.01}$ (dashed red line).

For all locations the main increase of the expected number of exceedances seems to have happened between 2010 and 2013. But there are differences in the course: For Konstanz in panel (A), a first significant change is already observed around 1980 and the expected number of exceedances almost continuously increases until the maximum change in 2011. All other five locations show some wavelike behavior with a first peak between 1980 and 1985, followed by a second peak between 1995 and 2005, and a third peak after 2010. With the exception of Düsseldorf and Brocken (panels (B) and (C)), a significant increase (at 1% significance level) of the exceedance probability can be observed not before 1995 at the maritime locations (panels (E) and (F)), and after 2000 and 2010 at Zugspitze (panel (D)).

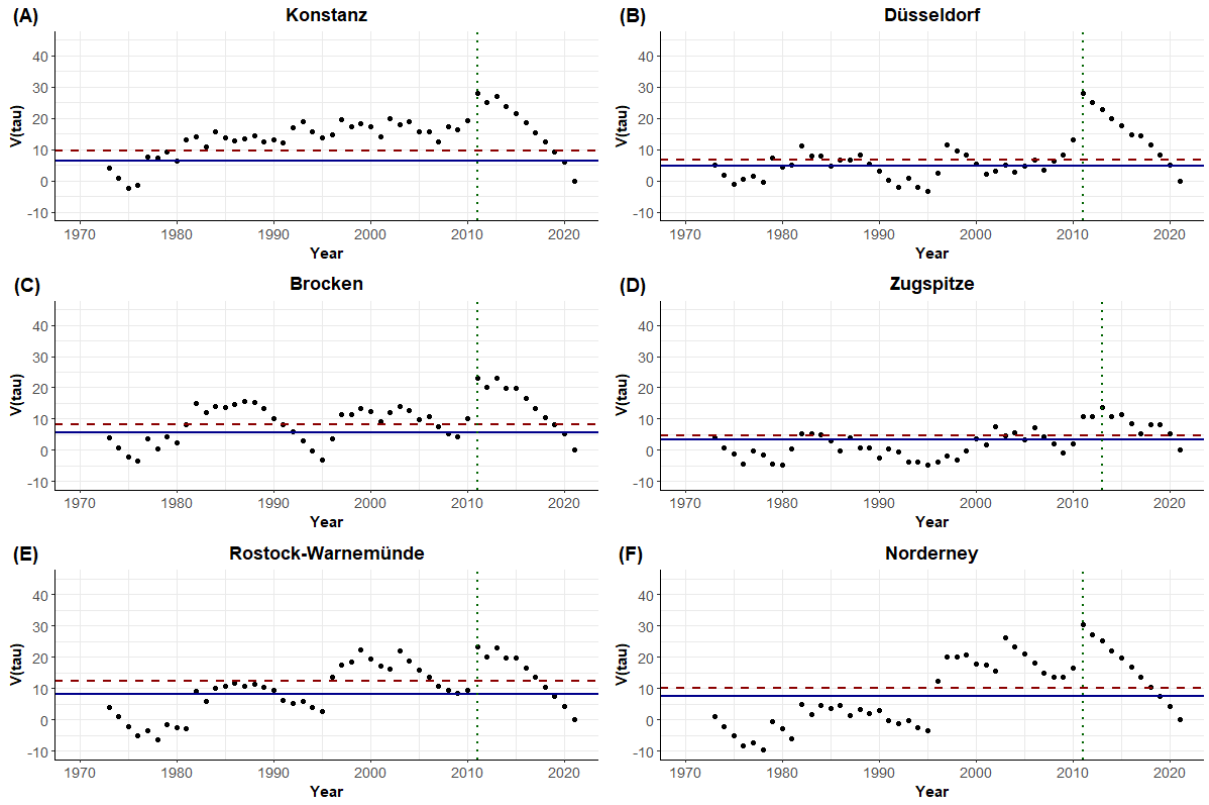

Figure 4.6: $V(\tau)$ (τ corresponds to the portion of elapsed years since 1973) for daily December temperatures at the six locations with critical limit at 5% significance level in horizontal solid blue line and critical limit at 1% significance level in horizontal dashed red line. M as the supremum over τ is displayed with vertical dotted green line.

From a climate change perspective we might conclude the following: The occurrence of the significant change not before 1995 could be an indication of a delayed effect of global warming on in general low December temperatures at higher altitudes (Zugspitze, 2965 m) and at locations that are influenced by the maritime climate. In addition, although, Düsseldorf is a city and should behave similar to Konstanz or the two cities considered in our paper, the course is somewhat special. This might be due to the fact that Düsseldorf also has a somewhat maritime climate.

In our paper (Beran, Steffens and Ghosh, 2021b) an extended data set of December temperatures at 15 locations divided into two categories is analyzed (but not presented here) to verify the presumption for higher altitudes. A similar analysis would be necessary for confirming the presumption for maritime locations.

Part II

Inference for circular long-memory processes

Chapter 5

Circular statistics and processes

This first introductory chapter deals with models describing circular data, i.e., data that can be represented as points on the unit circle. Well-known examples for circular data are directions that are given as angles and measured clockwise or counterclockwise from the zero direction. Other examples of circular data that can be transformed into angles, and thus can be regarded as directions, are, e.g., daily time of sunrise or the first day of spring in the year, orientation of bees or angles of molecular bindings. Actually, there are numerous areas of application where circular data occur, among others meteorology, astronomy, medicine, oceanography, geology and bioinformatics. In environmental sciences, circular statistics has gained interest during the last decades, especially in the context of climate change and biodiversity, e.g., to detect shifts in the phenology of species (Morellato et al., 2010 [301]).

There are many textbooks that give an introduction to circular statistics including Mardia (1972, 2014) [285], [287], Batschelet (1981) [18], Upton and Fingleton (1985) [395], Fisher et al. (1993) [142], Fisher (1995) [137], Mardia and Jupp (1999, 2009) [288], [289], Jammalamadaka and SenGupta (2001) [224], and Pewsey et al. (2013) [320]. Two books of Ley and Verdebout (2017, 2018) [265], [266] discuss modern methods in directional statistics. A summary of literature and recent advances in directional statistics is given in Pewsey and García-Portugués (2021) [319]. However, most methods described in these references deal with iid observations, whereas our interest lies in strongly dependent observations.

In the following, we review main definitions and notions in circular statistics and introduce different classes of models, in particular, a special class of circular long-memory processes that are generated via Gaussian subordination.

5.1 Basic definitions and notations

Throughout this section, we follow the definitions and notions in Mardia (1972) [285] and Pewsey et al. (2013) [320]. For more details, see also the textbooks of Mardia and Jupp (1999, 2009) [288], [289], Jammalamadaka and SenGupta (2001) [224], and Mardia (2014) [287].

5.1.1 Properties of the state space in circular statistics

As mentioned above, the idea in circular statistics is to represent data as points on the unit sphere $S^1 := \{x \in \mathbb{R}^2 : \ ||x|| = 1\}$. Each point in S^1 can be uniquely identified either with a *unit complex number* or with the *angle of its position vector* with respect to the x-axis. Let z denote the complex number corresponding to position vector x with angle θ (see Figure 5.1), then

$$x = (\cos(\theta), \sin(\theta))^T, \qquad z = e^{i\theta} = \cos(\theta) + i\sin(\theta), \qquad \text{and} \qquad \theta = arg\{z\},$$

where $\arg\{z\}$ denotes the argument of the complex number z.

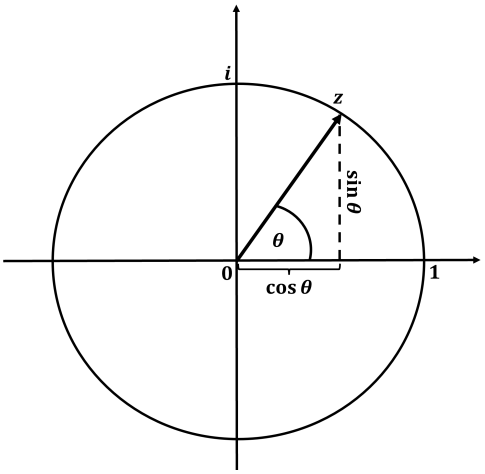

Figure 5.1: Polar representation of a point z in the complex plane with corresponding position vector x and angle θ.

Note that, due to periodicity, angles θ and $\theta + 2\pi k$, $k \in \mathbb{Z}$, correspond to the same point on the unit sphere. Hence, standard statistical methods such as the usual arithmetic mean of angles as a measure of location, are no longer appropriate, since these measures do not reflect this identification.

We define the **canonical mapping** q with respect to the equivalence relation

$$x \sim y :\Leftrightarrow y = x \ mod \ 2\pi$$

by

$$\begin{aligned} q: \mathbb{R} &\rightarrow \mathbb{R}/2\pi\mathbb{Z} \\ x &\mapsto x \; mod \; 2\pi. \end{aligned} \tag{5.1}$$

The mapping $z = e^{i\theta} \mapsto \theta$ defines a homeomorphism of S^1 equipped with the *Euclidean* $\mathbb{R}^2$-*topology* onto the quotient space $\mathbb{R}/2\pi\mathbb{Z}$ equipped with the *(Euclidean) quotient topology.*

For two points x and y on the unit sphere S^1 with angles θ and ψ, respectively, the **Euclidean metric (Euclidean distance)** given by

$$\begin{aligned} d_2(x,y) &= ||x-y||_2 = \sqrt{(\cos(\theta)-\cos(\psi))^2 + (\sin(\theta)-\sin(\psi))^2} \\ &= \sqrt{2\cdot(1-\cos(\theta)\cos(\psi)-\sin(\theta)\sin(\psi))} \\ &= \sqrt{2\cdot(1-\cos(\theta-\psi))} \end{aligned} \tag{5.2}$$

induces the *Euclidian topology* restricted to S^1. Since the metric defined by

$$d(\theta,\psi) := \sqrt{1-\cos(\theta-\psi)} = \frac{1}{\sqrt{2}}\cdot d_2(x,y) \tag{5.3}$$

differs from (5.2) only by a factor, the S^1-topology is as well induced by the metric d.

The latter one is commonly used in circular statistics as a *"circular metric"*. Its square, which is half the squared Euclidean distance, i.e.,

$$D(\theta,\psi) := d^2(\theta,\psi) = 1-\cos(\theta-\psi), \tag{5.4}$$

is called a "**distance measure**" in the relevant literature, although it is not a metric itself. It describes the "*being close to each other*" for circular data in the following sense:

$$\begin{aligned} 1-\cos(\theta-\psi) &= 1-(\cos(\theta)\cos(\psi)+\sin(\theta)\sin(\psi)) \\ &= 1-x^Ty, \end{aligned} \tag{5.5}$$

where x^Ty is the scalar product of x and y. This term becomes minimal, if and only if the angles θ and ψ coincide, and maximal if and only if they differ by π. In terms of position vectors x and y, this means that the measure becomes minimal if and only if x and y coincide, and maximal if and only if x and y lie exactly opposite to each other.

Furthermore, as can be seen from Figure 5.1, applying the Pythagorean theorem to the right-

hand triangle, we have

$$2\left(1-\cos\left(\theta-\psi\right)\right)=\left(1-\cos\left(\theta-\psi\right)\right)^2+\sin^2\left(\theta-\psi\right).$$

Since either term gets small when $\theta-\psi$ is small, the state of "*being close each other*" can be expressed by any of the three terms. This will be of interest again when we define dispersion and association measures.

We refer to the circular measure (5.4) again in the definition of circular kernels (see Definition 6.2.1) in the next chapter.

To motivate the notions in circular statistics, we have a look at a sample $x_1,\ldots,x_n$ of unit vectors with angles θ_j, $j=1,\ldots,n$. In contrast to the arithmetic mean of the angles, the commonly used *measure of location* is the **(sample) mean direction** $\bar{\theta}$ defined as the direction of the **mean resultant vector** $\bar{x}$ (provided it exists) or equivalently as the direction of the **center of mass** $\bar{x}$ of $x_1,\ldots,x_n$ with Cartesian coordinates $(\bar{C},\bar{S})$, where

$$\bar{C}=\frac{1}{n}\sum_{j=1}^{n}\cos\left(\theta_j\right)\quad\text{and}\quad\bar{S}=\frac{1}{n}\sum_{j=1}^{n}\sin\left(\theta_j\right). \tag{5.6}$$

The mean direction then is given as

$$\bar{\theta}=arg\left\{\bar{C}+i\cdot\bar{S}\right\}.$$

In other words, $\bar{\theta}$ is the solution of

$$\bar{C}=\bar{R}\cos\left(\bar{\theta}\right),\quad\bar{S}=\bar{R}\sin\left(\bar{\theta}\right),$$

where $\bar{R}=\sqrt{\bar{C}^2+\bar{S}^2}\in[0,1]$ is the **length of the center of mass vector** $\bar{x}$, and we obtain in terms of complex numbers:

$$\bar{x}=\bar{R}\cdot e^{i\cdot\bar{\theta}}=\frac{1}{n}\sum_{j=1}^{n}e^{i\theta_j}. \tag{5.7}$$

In case of unimodal circular data, $\bar{R}$ serves as a *measure of concentration* in the following sense: If the data are evenly or uniformly spread around the circle, $\bar{R}\approx 0$, and if all points are concentrated close to each other (around $\bar{\theta}$), $\bar{R}\approx 1$.

Remark. *Summation in* $\mathbb{C}$ *corresponds to summation in the quotient space* $\mathbb{R}/2\pi\mathbb{Z}$*, i.e., for angles* θ_j*,* $j=1,\ldots,n$*,*

$$arg\left\{\frac{1}{n}\sum e^{i\theta_j}\right\}=\frac{1}{n}\left[\sum\theta_j\ (mod\ 2\pi)\right].$$

Proof for $n = 2$: Let θ and ψ, $\theta \neq \psi$, be two angles with mean direction μ. Then

$$\mu = arg\left\{\frac{1}{2}\left(e^{i\theta} + e^{i\psi}\right)\right\} \tag{5.8}$$

and we need to prove

$$\mu \stackrel{!}{=} \frac{1}{2}\left[(\theta + \psi) \ mod\, 2\pi\right], \qquad or \qquad 2\mu - \theta - \psi \stackrel{!}{=} 0 \ mod\, 2\pi, \qquad or \qquad e^{i(2\mu-\theta-\psi)} \stackrel{!}{=} 1,$$

respectively. From (5.8) we obtain

$$\begin{aligned} z := e^{i(2\mu-\theta-\psi)} &= \underbrace{\frac{1}{2}\left(e^{i\theta} + e^{i\psi}\right)}_{=e^{i\mu}} \cdot e^{-i\theta} \cdot \underbrace{\frac{1}{2}\left(e^{i\theta} + e^{i\psi}\right)}_{=e^{i\mu}} \cdot e^{-i\psi} \\ &= \frac{1}{4}\left(2 + e^{i(\psi-\theta)} + e^{i(\theta-\psi)}\right). \end{aligned}$$

Now

$$\sin(\psi - \theta) + \sin(\theta - \psi) = 0$$

and

$$2 + \cos(\psi - \theta) + \cos(\theta - \psi) \geq 0,$$

imply

$$arg\{z\} = 0.$$

The proof for n angles now results by mathematical induction.

To illustrate the idea behind this, let $\theta_1 = \epsilon$ and $\theta_2 = 2\pi - \epsilon$. Then the usual arithmetic mean of these angles results in $\frac{1}{2}(\theta_1 + \theta_2) = \pi$. However, the mean direction is given by $\bar{\theta} = 0$ which is the same as $2 \cdot [(\epsilon + (2\pi - \epsilon)) \, mod\, 2\pi]$.

5.1.2 Mean direction and circular correlation

In this section, we introduce *circular random variables* and *basic measures of location, concentration or dispersion, and association*, such as the *mean direction* and the *circular correlation.*

Note that the main difference between distributions on the real line and circular distributions is the periodic nature.

Definition 5.1.1. *Let Θ be a circular random variable, i.e., $\Theta \in [0, 2\pi)$. The **circular distribution function** F of Θ on $[0, 2\pi)$ is right-continuous and defined by*

$$F(\alpha) = P(0 < \Theta \leq \alpha), \qquad \alpha \in [0, 2\pi),$$

with

$$F(0) = 0 \qquad \textit{and} \qquad \lim_{\alpha \to 2\pi} F(\alpha) = 1.$$

The circular distribution function can be extended on the whole real line as follows (see, e.g., Mardia, 1972 [285]): For $\alpha \in [k, k + 2\pi)$, $k \in \mathbb{Z}$,

$$\bar{F}(\alpha) := F(\alpha \; mod \; 2\pi) + k$$

with the consequence, that for any $\alpha \in \mathbb{R}$ the following two properties hold:

$$\bar{F}(\alpha + 2\pi) - \bar{F}(\alpha) = 1, \qquad -\infty < \alpha < \infty$$

and (in contrast to the real-valued situation)

$$\lim_{\alpha \to -\infty} \bar{F}(\alpha) = -\infty \qquad \text{and} \qquad \lim_{\alpha \to +\infty} \bar{F}(\alpha) = \infty.$$

In applications, it is commonly assumed that the distribution function F of the circular random variable Θ is absolutely continuous with respect to the Lebesgue measure on $[0, 2\pi)$. Hence, it can be specified through its probability density function f such that

$$\int_{\alpha}^{\beta} f(\theta) d\theta = F(\beta) - F(\alpha), \qquad 0 \leq \alpha < \beta < 2\pi.$$

If $\bar{f}$ is the *periodical continuation* of the probability density function f, $\bar{f}$ has the following properties:

(i) $\bar{f}(\theta) \geq 0$ almost everywhere on $\mathbb{R}$,

(ii) $\bar{f}(\theta) = \bar{f}(\theta + k \cdot 2\pi)$ for any $k \in \mathbb{Z}$ and almost everywhere on $\mathbb{R}$,

(iii) $\int_{\alpha}^{2\pi+\alpha} \bar{f}(\theta) d\theta = 1$ for any $\alpha \in \mathbb{R}$.

Based on this continuation $\bar{f}$ we obtain:

$$\bar{F}(\beta) - \bar{F}(\alpha) = \int_{\alpha}^{\beta} \bar{f}(\theta) d\theta.$$

Assume that a real-valued random variable X with probability distribution function F_X and probability density function $f_X = F_X'$ is given. By the canonical mapping q defined in (5.1) of X onto its quotient space, we obtain a circular random variable $\Theta = X \; mod \; 2\pi$, and the image

distribution $P_{q\circ X}$ has the probability distribution function F_Θ given by

$$F_\Theta(\theta) = \sum_{k=-\infty}^{\infty} (F_X(\theta + 2\pi k) - F_X(2\pi k)), \qquad \theta \in [0, 2\pi), \tag{5.9}$$

provided F_X is the distribution function of X. Hence, a corresponding probability density function is

$$f_\Theta(\theta) = \sum_{k=-\infty}^{\infty} f_X(\theta + 2\pi k), \qquad \theta \in [0, 2\pi), \tag{5.10}$$

provided f_X is a probability density function of X. Such distributions are called *wrapped distributions*, examples of which will be given Chapter 5.2.1.

Recall the introductory remarks with respect to circular data on S^1 and the associated mean direction. Accordingly, we now define parameters of circular distributions. Due to the periodic nature of a random angle Θ, we obtain for the characteristic function of Θ

$$\begin{aligned} \phi_\Theta(t) = E\left[e^{it\Theta}\right] &= E\left[e^{it(\Theta+2\pi)}\right] \\ &= e^{it2\pi} \cdot E\left[e^{it\Theta}\right], \end{aligned}$$

which implies $\phi_\Theta(t) = 0$ only if $e^{it2\pi} = 0$ or, equivalently, if $t \in \mathbb{Z}$. Hence, the characteristic function ϕ_Θ of Θ has non-zero values only for integers $t \in \mathbb{Z}$. Consequently, ϕ_Θ can be written as a doubly infinite sequence of complex numbers, the so-called **pth trigonometric moments of Θ about the zero direction**. These trigonometric moments $\tau_{p,0}$ are the Fourier coefficients of the circular distribution F, i.e.,

$$\begin{aligned} \tau_{p,0} &= E\left[e^{ip\Theta}\right] \\ &= \int_0^{2\pi} e^{ip\theta} dF(\theta) \\ &=: \alpha_p + i \cdot \beta_p, \end{aligned} \tag{5.11}$$

where

$$\tau_{0,0} = 1, \qquad \bar{\tau}_{-p,0} = \tau_{p,0}, \qquad \text{and} \quad |\tau_{p,0}| \leq 1,$$

and where

$$\alpha_p = E\left[\cos\left(p\Theta\right)\right] \qquad \text{and} \qquad \beta_p = E\left[\sin\left(p\Theta\right)\right], \qquad p \in \mathbb{Z}, \tag{5.12}$$

and obviously

$$\alpha_{-p} = \alpha_p, \qquad \beta_{-p} = -\beta_p, \qquad |\alpha_p| \leq 1 \qquad \text{and} \qquad |\beta_p| \leq 1.$$

This means that any circular distribution is uniquely determined by its p-th sine and cosine moments. Based on this, $R_p = \sqrt{\alpha_p^2 + \beta_p^2} \in [0,1]$ defines the length of the mean resultant vector $\tau_{0,p}$ with direction

$$\mu_p = atan2(\beta_p, \alpha_p) := \begin{cases} \arctan\left(\frac{\beta_p}{\alpha_p}\right), & \text{if } \alpha_p > 0, \\ \arctan\left(\frac{\beta_p}{\alpha_p}\right) + \pi, & \text{if } \alpha_p < 0, \beta_p \geq 0 \\ \arctan\left(\frac{\beta_p}{\alpha_p}\right) + 2\pi, & \text{if } \alpha_p < 0, \beta_p < 0, \\ \frac{\pi}{2}, & \text{if } \alpha_p = 0, \beta_p > 0, \\ -\frac{\pi}{2}, & \text{if } \alpha_p = 0, \beta_p < 0, \\ \text{undefined}, & \text{if } \alpha_p = 0, \beta_p = 0. \end{cases} \tag{5.13}$$

Moreover,

$$\tau_{p,0} = R_p e^{i\mu_p}, \tag{5.14}$$

and consequently

$$\alpha_p = R_p \cos(\mu_p) \qquad \text{and} \qquad \beta_p = R_p \sin(\mu_p), \qquad p \in \mathbb{Z}. \tag{5.15}$$

If the sequence $\left(R_p^2\right)$ is summable, Θ is absolutely continuous and the corresponding circular density function $f = F'$ can be defined (almost everywhere on $[0, 2\pi)$) by the Fourier expansion

$$\begin{aligned} f(\theta) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \tau_{p,0} e^{-ip\theta} &= \frac{1}{2\pi}\left[1 + \sum_{p=-\infty}^{\infty} \left(\alpha_p \cos(p\theta) + \beta_p \sin(p\theta)\right)\right] \\ &= \frac{1}{2\pi}\left[1 + \sum_{p=-\infty}^{\infty} R_p \cos(\mu_p - p\theta)\right], \end{aligned} \tag{5.16}$$

which is similar to the inversion formula for continuous real-valued random variables. This directly follows from (5.15) by use of trigonometric identities. We will later refer to this representation to show uniform convergence of our circular kernel estimator.

In an analogous manner as in (5.11), we define the **pth trigonometric moment about the mean direction** μ,

$$\tau_{p,\mu} = E\left[e^{ip(\Theta-\mu)}\right] = \bar{\alpha}_p + i \cdot \bar{\beta}_p$$

where

$$\bar{\alpha}_p = E\left[\cos(p(\Theta - \mu))\right] \qquad \text{and} \qquad \bar{\beta}_p = E\left[\sin(p(\Theta - \mu))\right], \qquad p \in \mathbb{Z}.$$

From this, together with (5.14), we further obtain

$$\begin{aligned}\tau_{p,\mu} &= E\left[e^{ip\Theta}\right]e^{-ip\mu} \\ &= \tau_{p,0}e^{-ip\mu} \\ &= R_p e^{i(\mu_p - p\mu)},\end{aligned}$$

where

$$\bar{\alpha}_p = R_p\cos(\mu_p - p\mu) \qquad \text{and} \qquad \bar{\beta}_p = R_p\sin(\mu_p - p\mu), \qquad p \in \mathbb{Z}.$$

This corresponds to the Fourier expansion of f evaluated at $\theta = \mu$.

The basic measures of location and concentration are defined by the first trigonometric moment about the mean direction. This means μ_1 defines the **mean direction** μ, and R_1 defines the **concentration** $\boldsymbol{R}$. In other words (and analogously as in (5.7)), if $R > 0$, the mean direction μ is defined through the characteristic function of Θ, i.e.,

$$E\left[e^{i\Theta}\right] = \alpha_1 + i\cdot\beta_1 =: Re^{i\mu}. \tag{5.17}$$

By dividing both sides of equation (5.17) by $e^{i\mu}$, we see that the imaginary part of the left-hand side has to be zero and consequently, we obtain

$$E[\sin(\Theta-\mu)] = 0 \quad \text{and} \quad E[\cos(\Theta-\mu)] = R. \tag{5.18}$$

Remark. *Assertion (5.18) can as well be directly shown using the representation of the trigonometric moments:*

If $R > 0$, applying the trigonometric identities, we obtain for the first cosine moment

$$\begin{aligned}\bar{\alpha}_1 &= E[\cos(\Theta-\mu)] \\ &= E[\cos(\Theta)\cos(\mu) + \sin(\Theta)\sin(\mu)] \\ &\underbrace{=}_{(5.15)} \frac{\alpha_1}{R_1}E[\cos(\Theta)] + \frac{\beta_1}{R_1}E[\sin(\Theta)] \\ &\underbrace{=}_{(5.12)} \frac{\alpha_1}{R_1}\alpha_1 + \frac{\beta_1}{R_1}\beta_1 \\ &\underbrace{=}_{R=R_1=\sqrt{\alpha_1^2+\beta_1^2}} \frac{R_1^2}{R_1} = R_1 = R.\end{aligned}$$

For the first sine moment, we obtain

$$\begin{aligned}\bar{\beta}_1 &= E\left[\sin\left(\Theta-\mu\right)\right] \\ &= E\left[\sin\left(\Theta\right)\cos\left(\mu\right)-\cos\left(\Theta\right)\sin\left(\mu\right)\right] \\ &\underbrace{=}_{(5.15)} \frac{\alpha_1}{R_1}E\left[\sin\left(\Theta\right)\right]-\frac{\beta_1}{R_1}E\left[\cos\left(\Theta\right)\right] \\ &\underbrace{=}_{(5.12)} 0.\end{aligned}$$

In accordance to the real-valued analogue, a circular distribution with density f is called **reflectively symmetric about any value** ξ if and only if for all $\theta \in [0, 2\pi)$

$$f([\theta-\xi] \; mod \; 2\pi) = f([\xi-\theta] \; mod \; 2\pi).$$

In case of *reflectively symmetric circular distributions* with $R > 0$, assertion (5.18) holds for the pth cosine and sine moments, i.e., $\bar{\alpha_p} = R_p$ and $\bar{\beta_p} = 0$. In the following, we will write "*symmetric distributions*" for short. As a consequence, for symmetric random variables, (5.16) reduces to

$$f(\theta) = \frac{1}{2\pi}\left[1+2\sum_{p=1}^{\infty}\left(\alpha_p\cos\left(p\theta\right)\right)\right], \tag{5.19}$$

which is equivalent to the fact that symmetric random variables possess real characteristic functions.

Remark. *Note that, if $R = 0$, the mean direction μ is undefined (which in the real-valued case corresponds to the fact that the expected value does not exist). This is the case for the continuous circular uniform distribution given in (5.26) or any cyclically symmetric circular distribution.*

Examples for circular distributions that are (reflectively) symmetric about their unique mode are the *von Mises distribution*, the *Cardioid distribution* or the *wrapped normal distribution*. The *circular uniform distribution* is a special symmetric distribution that is (reflectively) symmetric about any axis. Since symmetric distributions show many inferential advantages, there are several tests for symmetry in literature, mainly about a specified direction, see, e.g., Batschelet (1965) [17], Schach (1969) [348], Pewsey (2002, 2004) [317], [318], Ley and Verdebout (2014) [264], Ameijeiras-Alonso et al (2021) [7]. However, since in applications symmetric distributions rarely occur, in recent years, a main focus was on non-symmetric or skewed distributions (see, e.g., Kato and Jones, 2010 [239], Abe and Pewsey, 2011a [2], and Jones and Pewsey, 2012 [233]). These distributions are briefly reviewed in the subsequent sections.

Turning to *measures of dispersion*: In literature, a **circular variance** as some "*inverse*" of the

measure of concentration R, is defined by

$$V = 1 - R = E\left[1 - \cos\left(\Theta - \mu\right)\right] \tag{5.20}$$

with **standard deviation** $\sigma = \sqrt{V}$ and values in $[0, 1]$. This measure is defined in accordance to the circular distance given in (5.4). Taking into account that $e^{i\mu}$ is the point on S^1 corresponding to the mean direction μ, we could calculate a dispersion measure of variance type as follows:

$$\mathrm{var}(\Theta) := E[|e^{i\Theta} - e^{i\mu}|^2] \underbrace{=}_{\text{trig. identities}} 2\left(1 - E\left[\cos\left(\Theta - \mu\right)\right]\right) \underbrace{=}_{(5.18)} 2 \cdot (1 - R) = 2V.$$

Some references refer to this $\mathrm{var}(\Theta)$ as a measure of dispersion.

Further alternative measures of location and dispersion are given in the following remark.

Remarks. *(i) The **population median** $\tilde{\mu}$ is defined as any angle minimizing*

$$E\left[\pi - |\pi - |\Theta - \mu||\right]$$

with respect to μ. This means $\tilde{\mu}$ is such that

$$P\left(\Theta \in [\tilde{\mu}, \tilde{\mu} + \pi)\right] \geq \frac{1}{2} \qquad \textit{and} \qquad P\left(\Theta \in [\tilde{\mu} - \pi, \tilde{\mu})\right] \geq \frac{1}{2}.$$

As a robust measure of location, it is, e.g., used for skewed distributions. This is in accordance with the population median used in the real-valued setting.

(ii) A monotone transformation of the circular variance that has values in $[0, \infty)$ and reflects the property that the variance of highly concentrated data is approximately 0 and it is approximately ∞ in case of evenly spread data is given by

$$V^* := \log\left(\frac{1}{R^2}\right) = -2\log\left(R\right) = -2\log\left(1 - V\right) \tag{5.21}$$

with circular standard deviation $\sigma^* = \sqrt{V^*}$. This measure is strongly related to the standard deviation on the real line, since in case of the wrapped normal distribution with parameters μ and σ^* defined in (5.28), $R = e^{-\frac{1}{2}(\sigma^*)^2}$.

Moreover, if V in (5.21) is small, $\log\left(1 - V\right) \approx -V$, and thus (together with (5.20))

$$\sigma^* \approx \sqrt{2V} = \sqrt{2 \cdot (1 - R)} = \sqrt{2} \cdot \sigma.$$

*E.g., in Batschelet (1981) [18], this so-called **angular deviation** is used as an alternative*

measure of dispersion.

(iii) The length of the smallest arc that contains all observations is denoted as the ***circular range****.*

(iv) For measures of skewness and kurtosis see, e.g., Mardia (1972) and Pewsey et al. (2013).

Finally, turning to *measures of association*: At first sight, it seems reasonable to define a measure of association of two circular variables in the following way:

$$\begin{aligned} cov(\Theta, \Psi) &:= E\left[|e^{i\Theta} - e^{i\mu}| \cdot |e^{i\Psi} - e^{i\nu}|\right] \\ &= E\left[\sqrt{2\left(1 - \cos\left(\Theta - \mu\right)\right)} \cdot \sqrt{2\left(1 - \cos\left(\Psi - \mu\right)\right)}\right] \end{aligned}$$

and accordingly the *circular correlation.*

However, a measure of association of covariance-type should reflect the "*direction of variation*" of the two variables. Roughly speaking, the covariance should be positive if both variables exhibit the "*same rotational behavior*", else it should be negative. But since cosine is an even function it does not distinguish between positive and negative rotation, hence it is not appropriate. On the other hand, sine would reflect this since it is odd. This explains why, as a *measure of association*, the *circular correlation coefficient* defined by Jammalamadaka and Sarma (1988) [223] is frequently in use.

Definition 5.1.2. *Let Θ, Ψ be circular random variables with corresponding mean directions μ and ν, respectively. Then the* ***circular correlation coefficient*** *ϱ_{JS} is defined by*

$$\varrho_{JS}(\Theta, \Psi) = \frac{E\left[\sin\left(\Theta - \mu\right)\sin\left(\Psi - \nu\right)\right]}{\sqrt{E\left[\sin^2\left(\Theta - \mu\right)\right] E\left[\sin^2\left(\Psi - \nu\right)\right]}}. \tag{5.22}$$

The **circular autocovariance** for circular random variables Θ and Ψ with corresponding mean directions μ and ν, which becomes crucial when considering long-range dependent variables, can analogously be defined as

$$\operatorname{cov}(\Theta, \Psi) = E\left[\sin\left(\Theta - \mu\right)\sin\left(\Psi - \nu\right)\right]. \tag{5.23}$$

Moreover, if Θ and Ψ exhibit small dispersion, i.e., if $\sin\left(\Theta - \mu\right)) \approx \Theta - \mu$, the circular correlation coefficient ϱ_{JS} coincides with the classical (real-valued) correlation, and indeed, it retains many of the known properties of the *classical (real-valued) product moment correlation coefficient.*

This correlation coefficient ϱ_{JS} (in terms of the autocovariance function) will be later (see Chapter 5.3) utilized when introducing the long-memory property in the context of circular processes.

In view of the remarks concerning the "*being close to each other*" and the different approaches to define a dispersion measure, using the sine in the denominator in (5.22) as "*variance*" makes sense as well. Figure 5.2 provides a simple illustration of all mentioned "*distance measures*" applied to two points y and z and the corresponding angles ψ and θ, where, without loss of generality, $y = (1, 0)$ and $\psi \equiv 0$.

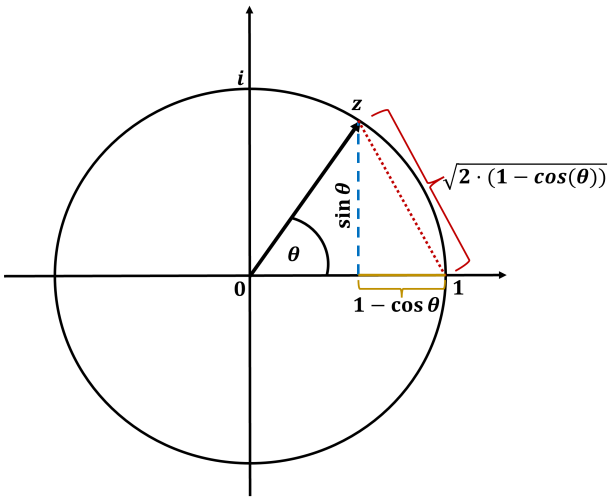

Figure 5.2: Summary of circular distance measures. The yellow solid distance displays the circular measure (5.4), the red dotted distance corresponds to the Euclidean metric (5.2) and the blue dashed distance represents the sine as used in (5.22).

An example of a *measure of rotational dependence* of the form $\Theta = \Psi + \xi \ (mod\ 2\pi)$ or $\Theta = -\Psi + \xi \ (mod\ 2\pi)$, where ξ is some constant angle, is the *circular correlation coefficient* defined by Fisher and Lee (1983) [138]. This coefficient corresponds to the usual real-valued correlation coefficient of a bivariate random variable as standard measure of linear dependence. Its derivation is analogous as described above.

Definition 5.1.3. *Let* (Θ_1, Ψ_1) *and* (Θ_2, Ψ_2) *be independently distributed as* (Θ, Ψ)*. Then a* ***circular correlation coefficient*** *is given by*

$$\varrho_{FL}(\Theta, \Psi) = \frac{E\left[\sin\left(\Theta_1 - \Theta_2\right)\sin\left(\Psi_1 - \Psi_2\right)\right]}{\sqrt{E\left[\sin^2\left(\Theta_1 - \Theta_2\right)\right] E\left[\sin^2\left(\Psi_1 - \Psi_2\right)\right]}}. \tag{5.24}$$

In literature, there are several approaches to define a correlation coefficient for bivariate circular random variables. Product moment type correlation coefficients similar to ϱ_{JS} or ϱ_{FL} given above are, e.g., introduced by Downs (1974) [123], Mardia (1975) [286], Johnson and Wehrly (1977) [225], Mardia and Puri (1978) [290], Jupp and Mardia (1980) [234]. For a review on

measures of circular association, see also the textbooks of Madia and Jupp (1999) [288] and of Jammalamadaka and SenGupta (2001) [224].

5.2 Examples of circular distributions and the corresponding processes

In this section we have a look at different examples of circular distributions and define the corresponding circular processes. Clearly, a *circular process* is a family of circular random variables. We will denote circular processes by $(\Theta_t)_{t\in\mathbb{Z}} \in [0, 2\pi)$ and will write (Θ_t) for short.

In the circular setting, properties such as stationarity can be defined analogously as in the real-valued setting. This means, a circular process (Θ_t) is **strongly stationary** if

$$(\Theta_1, \ldots, \Theta_n) \stackrel{d}{=} (\Theta_{1+k}, \ldots, \Theta_{n+k}), \qquad k \geq 1.$$

We restrict our considerations to **(second-order) stationary circular processes** (Θ_t) with (existing) mean direction μ, i.e., to processes for which the circular autocovariance given in (5.23) can be rewritten as

$$\gamma(t, s) = \gamma(k) = \operatorname{cov}(\Theta_t, \Theta_{t+k}) = E\left[\sin(\Theta_1 - \mu)\sin(\Theta_{1+k} - \mu)\right], \qquad k \in \mathbb{Z}.$$

As a consequence, for stationary processes, the circular correlation ϱ_{JS} (5.22) reduces to

$$\varrho_{JS}(k) = \frac{\gamma(k)}{E\left[\sin^2(\Theta_1 - \mu)\right]}, \qquad k \in \mathbb{Z}. \tag{5.25}$$

Most literature deals with weakly dependent or Markov-type circular processes, see, e.g., Wehrly and Johnson (1980) [413], Breckling (1989, 2012) [62], [63], Fisher and Lee (1994) [140], Kato (2010) [238], Lee (2010a) [257], Modlin et al. (2012) [300], and Wang and Gelfand (2014) [407]. Here, weak dependence is defined by $0 < \sum_{k\in\mathbb{Z}} \varrho_{JS}(k) < \infty$, i.e., a weak dependent circular process has summable autocorrelations. In contrast to this, and again analogously as in the real-valued case, a circular process is said to exhibit long-range dependence if the autocorrelations are non-summable, i.e., if

$$\sum_{k\in\mathbb{Z}} \varrho_{JS}(k) = \infty.$$

In literature, to the best of our knowledge, there are two approaches for constructing long-range dependent circular processes. Di Marzio et al. (2012) [114] construct such a process via the *autocorrelation of bivariate processes*. A more general approach is given by Beran and Ghosh (2020) [34], who apply the method of *Gaussian subordination*. We review both approaches in

Chapter 5.3.

The most basic circular distribution is the **(continuous) circular uniform distribution** with density

$$f_U(\theta) = \frac{1}{2\pi} \mathbb{1}_{\{[0,2\pi)\}}(\theta). \tag{5.26}$$

This is a symmetric distribution with unit circular variance, and consequently $R = 0$. It is often used as the *null model* in testing for circular uniformity, circular randomness or isotropy.

Frequently used circular distributions to model circular data can be constructed either out of already existing circular distributions using *suitable link functions*, by *perturbation* or out of real-valued distributions. Classical approaches for the latter are *radial projection* of distributions in the plane and the *wrapping* of distributions on the real line onto the circle. For the latter two classes, distributions can be viewed as image measures under a suitable canonical mapping (with respect to an equivalence relation) onto the associated quotient space, see (5.1).

Classical examples of circular distributions such as the *wrapped normal* or the *von Mises distribution* are representatives of symmetric distributions. In case of circular data which are asymmetric or multimodal, there are many further approaches to provide more flexible models, e.g., the *Möbius transformation* (Kato and Jones, 2010 [239]), approaches based on *maximum likelihood* or *maximum entropy methods* (von Mises, 1981 [401] and Mardia, 1972 [285]), or approaches that are based on extensions or rather mixtures of von Mises distributions (Batschelet, 1981 [18] and Spurr and Koutbeiy, 1991 [370]). Skewed distributions can also be generated by *perturbation* of existing symmetric circular models (Abe and Pewsey, 2011a,b [2], [3] and Azzalini, 1985, 2005 [12], [13]) or by wrapping skewed distributions onto the unit circle (Pewsey, 2000 [316]). A new approach, proposed recently by Taniguchi et al. (2020) [380], is based on the normalization of the spectra of complex-valued stationary processes. This construction, in particular, allows for generating circular models from the spectra of ARMA(p,q)-processes, which can be applied to wind direction data and is also appropriate for multivariate circular models.

In the following, we review four classes of circular distributions, *wrapped distributions*, *direct linked distributions*, *projected* or *offset distributions*, and *perturbed distributions*. For an introduction and explicit analysis of these examples, we refer to the monographs of Mardia and Jupp (1999, 2009) [288], [289], Jammalamadaka and SenGupta (2001) [224], and Pewsey et al. (2013) [320].

5.2.1 Wrapped distributions and processes

Recall the construction of *wrapped circular random variables* from (5.9) or (5.10), i.e., Θ is obtained by wrapping a real-valued random variable X onto $[0, 2\pi)$ or onto S^1.

There are two wrapped distributions that are mostly in use: the *wrapped normal distribution* and the *wrapped Cauchy distribution*.

The wrapped normal distribution corresponds to the normal distribution on the real line and is strongly related to the von Mises distribution (see Example 5.2.8). In addition, as we will see in Example 5.2.4, this distribution has useful applications in the theory of *circular Brownian Motion*.

Note that in literature μ commonly denotes the *mean direction* and κ denotes the *concentration*. We will use this notation in the following examples.

Example 5.2.1. *The **wrapped normal distribution** is obtained by wrapping the normal distribution $N(\mu, \sigma^2)$ onto the circle. The resulting probability density function (see 5.10) of $WN(\mu, \sigma^2)$ is given by*

$$\varphi_\Theta(\theta; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{j=-\infty}^{\infty} \exp\left(\frac{-(\theta - \mu + 2\pi j)^2}{2\sigma^2}\right). \tag{5.27}$$

This distribution is unimodal and symmetric about its mode $\theta = \mu \bmod 2\pi$. In addition, when κ denotes the concentration of Θ, i.e.,

$$\kappa = e^{-\frac{1}{2}\sigma^2} \tag{5.28}$$

then

$$\sigma^2 = -2\log(\kappa).$$

Thus, an equivalent representation of this density is given by

$$\varphi_\Theta(\theta;\ \mu,\ R) = \frac{1}{2\pi}\left(1 + 2\sum_{j=1}^{\infty} \kappa^{j^2} \cos\left(j(\theta - \mu)\right)\right),$$

and the corresponding wrapped normal distribution is usually denoted by $WN(\mu, \kappa)$.

We have the following relationships with other distributions:

- *As $\kappa \to 0$, $WN(\mu, \kappa)$ tends to the circular uniform distribution.*
- *As $\kappa \to 1$, Θ is the unit mass at $\mu \bmod 2\pi$.*
- *This distribution is related to the Brownian motion on the circle defined in (5.30) in Example 5.2.4 in the following sense: If a particle starting at $\theta = \mu$ at time $\tau = 0$ has mean zero and variance $\sigma^2 = ct$ at time $\tau = t$, the distribution of the position of the particle at time $\tau = t$ is wrapped normal, precisely it is $WN\left(\mu, e^{-\frac{1}{2}ct}\right)$-distributed. For more details, see, e.g., Mardia (1972) [285] and Mardia and Jupp (1999) [288].*

The second example, the *wrapped Cauchy distribution*, is a useful distribution in the field of spectroscopy to analyze diffraction patterns and fiber orientation, in image processing, and in earth science, see, e.g., Schulgasser (1985) [349] and Borradaile (2003) [58]. This distribution was first introduced by Levy (1939) [263] and has been studied later, e.g., by Wintner (1947) [417] and by McCullagh (1992, 1996) [293], [294].

Example 5.2.2. *By wrapping the probability density function f of the real-valued Cauchy distribution, i.e.,*

$$f(x;\mu,a) = \frac{1}{\pi}\frac{a}{a^2+(x-\mu)^2},$$

onto the unit circle, where $a > 0$ is the scale factor and $\mu \in \mathbb{R}$ is the location parameter, we obtain the corresponding ***wrapped Cauchy distribution*** *$WC(\mu,\kappa)$ with density*

$$c(\theta;\mu,\kappa) = \sum_{j=-\infty}^{\infty} f(\theta+2\pi j;\mu,\kappa) = \frac{1}{2\pi}\frac{1-\kappa^2}{1+\kappa^2-2\kappa\cos(\theta-\mu)},$$

and concentration parameter $\kappa = e^{-a}$.

This distribution is unimodal and symmetric, and we have the following relationships:

- *As $\kappa \to 0$ (i.e., $a \to \infty$), the wrapped Cauchy distribution reduces to a circular uniform distribution with density (5.26). In addition, as discussed in McCullagh (1996) [294], a wrapped Cauchy random variable can be generated by Möbius transformation of the continuous circular uniform random variable.*

- *The wrapped Cauchy distribution is related to Brownian Motion in $\mathbb{R}^2$ in the following sense: The consecutive positions of a Brownian particle at exit times from the unit disc are distributed according to the wrapped Cauchy distribution, see, e.g., Kato and Jones (2013) [240].*

A **wrapped circular process** is obtained in an analogous way as the wrapped distributions

Definition 5.2.3. *Let (X_t) be a real-valued process, then the corresponding* ***wrapped process*** *(Θ_t) is obtained by wrapping (X_t) onto the unit circle, i.e.,*

$$\Theta_t = X_t \ mod\ 2\pi. \tag{5.29}$$

Example 5.2.4. *(i) The* ***wrapped Brownian Motion*** *on the circle is obtained by wrapping a real-valued Brownian motion onto the circle. This means, if $(B(t))$ is a Brownian Motion as in Definition 1.2.18, then*

$$\Theta_t = B(t)\ mod\ 2\pi \tag{5.30}$$

*is a **wrapped Brownian Motion**. This process appears as limiting distribution in limit theorems on the circle, and clearly (as in the real-valued case), the wrapped normal distribution from Example 5.2.1 is the one-dimensional distribution of a wrapped Brownian Motion.*

For more details, see de Haas-Lorentz (1913, 2013) [103], [104], Stephens (1963) [371], Bingham (1972) [55], and Mardia and Jupp (1999) [288].

*(ii) Another example is the **wrapped AR(p)-process (WAR(p)-process)**, which was introduced by Breckling for modeling stationary time series of wind directions (see, e.g., Breckling (1984, 1989) [61], [62], Fisher and Lee (1983) [138], and Coles (1998) [81]). Here, (X_t) in (5.29) is a stationary AR(p)-process as defined in (1.4).The circular correlation coefficient of the WAR(p)-process then depends both, on the parameters of the AR(p)-polynomial and on the correlation of (X_t).*

5.2.2 Direct linked distributions and processes

Wrapped distributions can be generalized by using a suitable link function and linking the $[0, 2\pi)$-valued variables with those of a real-valued process, see Fisher and Lee (1992) [139]. Let X be a real-valued random variable and let $\Psi : \mathbb{R} \to [0, 2\pi)$ be a strictly increasing and odd function, then, for $\delta \in [0, 2\pi)$, a **direct linked distribution** F_Θ of Θ can be defined by

$$\Theta = (\Psi(X) + \delta) \; mod \; 2\pi.$$

In fact, if $\Psi = id$ and $\delta = 0$, we obtain a wrapped distribution.

Useful functions Ψ (transformed to the interval $[0, 2\pi)$) are, e.g., the *inverse tan link function* $\Psi(x) = 2 \arctan(x) + \pi$ and the *scaled probit link function* $\Psi(x) = 2\pi\Phi(x)$, where Φ is the standard normal distribution function, or, more generally, $\Psi(x) = 2\pi F(x)$, where F is some arbitrary cumulative distribution function (see Johnson and Wehrly, 1978 [226]).

Conversely, the real-valued random variable can be written as

$$X = \Psi^{-1}((\Theta - \delta) \; mod \; 2\pi). \tag{5.31}$$

Such distributions are useful for analyzing concentrated data or data with a low dispersion, since in this case, although the values of Θ are similar, the corresponding values of X can differ substantially.

A **direct-linked process** (with $\delta = 0$) can be regarded as a **circular subordinated process**. Such processes are constructed in an analogous manner as in the real-valued case. Let (X_t) be

a real-valued process and let $\Psi : \mathbb{R} \to [0, 2\pi)$ be a suitable link function, then the corresponding circular process is obtained by

$$\Theta_t = \Psi(X_t). \tag{5.32}$$

In the special case of a real-valued zero-mean Gaussian process one obtains a *circular Gaussian subordinated process*, analogously defined as in Definition 1.2.7 in the real-valued setting.

Definition 5.2.5. *Let (Z_t) be a real-valued zero-mean Gaussian process and let $G : [0, 2\pi) \to [0, 1]$ be an absolutely continuous circular distribution function with density $g = G'$. Then a circular process (Θ_t) with marginal distribution function G given by*

$$\Theta_t = G^{-1}\left(\Phi(Z_t)\right),$$

where Φ denotes the standard normal distribution function, is called ***circular Gaussian subordinated process****.*

Using notation (5.32), we have $\Psi := G^{-1} \circ \Phi$.

The direct-link approach with inverse construction (5.31) is, e.g., used to define *linked ARMA-processes*, where (X_t) is an ARMA process and g is a scaled probit link function. This is, e.g., discussed in detail in Fisher and Lee (1994) [140].

Example 5.2.6. *Let (Θ_t) be a stationary circular process with mean direction μ. Such a process is called* ***linked ARMA(p,q)-process****, if the linked linear process $\Psi^{-1}(\Theta_t - \mu)$ is a real-valued ARMA(p,q)-process. The circular autocorrelation function is then given by*

$$\varrho_\Theta(k) = \varrho_{FL}(\Psi(X_t), \Psi(X_{t+k})), \qquad k \in \mathbb{Z},$$

where ϱ_{FL} is the circular correlation coefficient from Fisher and Lee (1983) [138] defined in (5.24).

These linked linear models have the advantage of transferring the property of stationarity from the real-valued process onto the circular process.

5.2.3 Projected distributions and processes

Another class of circular distributions is the class of **projected** or **offset distributions**. These circular distributions are generated by projecting a real-valued bivariate random vector X onto the unit sphere S^1. If $X = (Y, Z)^T \in \mathbb{R}^2$ is a bivariate real-valued random variable, then using polar coordinates,

$$Y = R\cos(\Theta) \quad \text{and} \quad Z = R\sin(\Theta),$$

where $R > 0$, we obtain a circular random variable

$$\Theta = \Theta(X) = arg\left\{||X||^{-1}X\right\} \in [0, 2\pi)$$

or equivalently

$$\Theta = arg(X) = arg\left\{Y + i \cdot Z\right\}.$$

If X has density f_X on $\mathbb{R}^2$, then the *projected density function* is given by

$$f_\Theta(\theta) = \int_0^\infty \left(f_X\left(r \cdot \cos(\theta), r \cdot \sin(\theta)\right) \cdot r\right)\, dr.$$

We present two examples of projected distributions: the *projected normal distribution* and the *von Mises distribution.*

Example 5.2.7. *Let X be a bivariate normally distributed random variable, i.e., $X \sim N_2(\mu, \Sigma)$, then the corresponding random angle Θ is said to have a* ***projected normal*** *or* ***angular Gaussian distribution*** *$PN_2(\mu, \Sigma)$ with mean direction $\mu \in [0, 2\pi)$ and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}. \tag{5.33}$$

For the explicit expression of the density of Θ, see, Mardia and Jupp (1999) [288].

In the special case of $\mu = 0$ and diagonal matrix Σ with $\sigma_{12} = \sigma_{21} = 0$, i.e., in case of $\Theta \sim PN(0, \Sigma)$, the corresponding probability density function reduces to

$$f_\Theta(\theta; 0, \Sigma) = \frac{(1 - b^2)^{\frac{1}{2}}}{2\pi\left(1 - b\cos(2\theta)\right)},$$

where

$$b = \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

This distribution is, e.g., used in meteorology for modeling wind velocities and directions, where (Y_t) and (Z_t) are assumed to be independently normally distributed with zero mean and variances σ_1^2 and σ_2^2, respectively.

If further $X \sim N_2(0, I)$, i.e., Σ is the unit matrix, then $b = 0$, and Θ is uniformly distributed.

Remark. *A useful property of the class of projected normal distributions is the following: Let*

A be an invertible linear transformation of the plane, then ψ_A *with*

$$\psi_A(x) = \frac{1}{||Ax||}Ax$$

is an invertible transformation of the unit circle, and $\Theta \sim PN_2(\mu, \Sigma)$ *implies*

$$A\Theta \sim PN_2(A\mu, A\Sigma A^T).$$

Since the set of projected normal distributions is closed under these transformations and, since for $\Sigma = I$, Θ *is uniformly distributed, each angular central Gaussian distribution can be obtained from the uniform distribution by a suitable transformation* ψ_A*, see Mardia and Jupp (1999) [288]. Thus, these circular distributions has useful applications in image analysis, see, e.g., Blake and Marinos (1990) [57].*

Another important representative of the class of projected distributions is the *von Mises distribution* $vM(\mu, \kappa)$. This distribution was first introduced by von Mises (1918) [400] when studying deviations of measured atomic weights from integral values. Due to its similarity to the normal distribution on the real line, it is also called the *circular normal distribution* $CN(\mu, \kappa)$. It plays a key role among circular distributions and has been widely studied in literature. Here, we will only mention its main properties.

Example 5.2.8. *Let* $X = (Y, Z)$ *have a bivariate normal distribution* $N_2(\mu, \Sigma)$*,* Y *and* Z *independent, with mean direction* $\mu \in [0, 2\pi)$ *and covariance matrix*

$$\Sigma = \begin{pmatrix} \kappa^{-1} & 0 \\ 0 & \kappa^{-1} \end{pmatrix},$$

where the concentration parameter κ *corresponds to the reciprocal of a measure of dispersion, i.e.,* $\kappa = \sigma_1^{-2} = \sigma_2^{-2}$*. Then the generated circular random variable* Θ *is said to have a* ***von Mises distribution*** $vM(\mu, \kappa)$ *with probability density given by*

$$f_\Theta(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\theta - \mu)} = \frac{1}{2\pi I_0(\kappa)} \left(I_0(\kappa) + 2 \sum_{p=1}^{\infty} I_p(\kappa) \cos(p(\theta - \mu)) \right), \tag{5.34}$$

where $I_p(\kappa)$ *denotes the* ***modified Bessel function of the first kind of order*** p *(see, e.g., Robert, 1990 [338]) defined by*

$$I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) e^{\kappa \cos(\theta)} d\theta. \tag{5.35}$$

In contrast to the normal distribution on the real line, here, for $\theta \in [0, 2\pi)$, the cumulative distribution function has an analytic representation given by

$$F(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \left(\theta \cdot I_0(\kappa) + 2 \sum_{p=1}^{\infty} \frac{I_p(\kappa) \sin\left(p(\theta - \mu)\right)}{p} \right).$$

Since

$$E\left[e^{i\Theta}\right] = \frac{I_1(\kappa)}{I_0(\kappa)} e^{i\mu},$$

the first trigonometric moment R_1 is given by

$$R_1 = R = \frac{I_1(\kappa)}{I_0(\kappa)} =: A(\kappa) \tag{5.36}$$

with Fourier coefficients

$$\alpha = A(\kappa) \cos(\mu) \qquad \text{and} \qquad \beta = A(\kappa) \sin(\mu),$$

respectively, and the circular variance (5.20) accordingly reads

$$V = 1 - A(\kappa).$$

The von Mises distribution is related to other distributions in the following way:

- *The von Mises distribution can be regarded as a further special case of the projected normal distribution with uncorrelated random variables with same variance, i.e., $\sigma_{12} = \sigma_{21} = 0$ and $\sigma_1^2 = \sigma_2^2$ in (5.33).*
- *Any von Mises distribution can be approximated by a wrapped normal distribution defined in (5.27) via $A(\kappa) = e^{-\frac{1}{2}\sigma^2}$ (see (5.28)), i.e., the distribution function of $WN(\mu, \kappa^{-1})$ as defined in Example 5.2.1 provides an approximation of the distribution function $vM(\mu, \kappa)$.*
- *As $\kappa = 0$, the von Mises distribution $vM(\mu, \kappa)$ reduces to the circular uniform distribution.*
- *Since for small κ, $e^{\kappa \cos(\theta - \mu)} \approx 1 + \kappa \cos(\theta - \mu)$, the von Mises distribution $vM(\mu, \kappa)$ can be approximated by the Cardioid distribution $C\left(\mu, \frac{\kappa}{2}\right)$ defined in Example 5.2.9.*
- *As $\kappa \to \infty$, the von Mises distribution $vM(\mu, \kappa)$ becomes concentrated at the point $\theta = \mu$, i.e., the distribution tends to a point distribution centered around μ.*
- *For large κ, the random variable $\beta = \sqrt{\kappa}(\Theta - \mu)$ converges in distribution to $N(0, 1)$. In other words, for large κ, the von Mises distribution can be approximated by a real-valued normal distribution with mean μ and variance $\sigma^2 = \kappa^{-1}$.*

- *The von Mises distribution is related to a circular AR(p)-process, see Example 5.2.4 (ii) and Fisher and Lee (1994) [140], in the sense that the conditional distribution of Θ_t given the past is a von Mises distribution with mean direction μ_t depending on the coefficients of the corresponding real-valued AR(p)-polynomial.*

- *The von Mises distribution is related to diffusion processes such as the Brownian Motion in the plane or the Ornstein-Uhlenbeck process on the line. This is, e.g., addressed by Mardia and Jupp (1999) [288], pp. 41 - 43.*

Moreover, there are several approaches for constructing a von Mises distribution including maximum entropy method or maximum likelihood characterization. Asymmetric transformations and multimodal models, e.g., the generalized von Mises distribution, the spherical von Mises-Fisher distribution, or multiplicative mixtures of those, are mentioned in Pewsey et al. (2013) [320].

5.2.4 Perturbed distributions and processes

The *perturbation approach* is a more direct approach. A circular probability density is multiplied by a function such that the resulting product again is a circular probability density. This means, such an approach can be applied on either of the above introduced distributions resulting in a new circular distribution. In other words, the perturbation approach can be seen as a generalization of the direct linked approach in the following sense:

Let X be a real-valued random variable, let $\Psi : \mathbb{R} \to [0, 2\pi)$ be a strictly increasing and odd function, and let $\delta \in [0, 2\pi)$. Let Θ be the resulting (circular) direct linked random variable of the form (5.2.2) with respective probability density function f_Θ. If a function h is such that the product $h \cdot f_\Theta$ again becomes a circular probability density, the resulting random variable $\tilde{\Theta}$ is called a **perturbed random variable** with circular density $f_{\tilde{\Theta}}$.

A classical example of a perturbed distribution is the *Cardioid distribution*, which is, e.g., used to model directional spectra of ocean waves.

Example 5.2.9. *For a circular random variable Θ the **Cardioid distribution**, first mentioned by Jeffreys (1961) [221], p. 328, is defined by multiplying the circular uniform density given in (5.26) with $h(\theta) := (1 + 2\kappa \cos(\theta - \mu))$, $\kappa \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. The notion Cardioid comes from the fact that the polar representation describes a Cardioid curve. The resulting probability density function $f_{\tilde{\Theta}}$ is of the form*

$$f_{\tilde{\Theta}}(\theta; \mu, \kappa) = \frac{1}{2\pi} (1 + 2\kappa \cos(\theta - \mu))$$

with location parameter $\mu \in [0, 2\pi)$ and concentration parameter $\kappa \in \left(-\frac{1}{2}, \frac{1}{2}\right)$.

This distribution is unimodal and symmetric around its mode $\theta = \mu$. Obviously, for $\kappa = 0$, the Cardioid distribution coincides with the circular uniform distribution defined in (5.26).

Here, perturbation was used to generate a symmetric circular distribution. However, this approach is frequently used to provide skewed distributions as well. As an example, we have a look at *sine-skewed distributions.*

Example 5.2.10. *Let Θ be a circular random variable with mean direction $\mu \in [-\pi, \pi)$ and let f_0 be any symmetric circular density function. If we define*

$$h(\theta) := 1 + \lambda \sin(\theta - \mu), \qquad \theta \in [-\pi, \pi),$$

where $\lambda \in [-1, 1]$ is some skewing parameter, the density f_Θ obtained via

$$f_\Theta(\theta) = f_0(\theta - \mu) \cdot h(\theta) = f_0(\theta - \mu) \cdot (1 + \lambda \sin(\theta - \mu))$$

*is referred to as a **sine-skewed circular distribution**. The corresponding distribution function is given by*

$$F_\Theta(\theta) = F_0(\theta) + \lambda \int_{-\pi}^{\theta} \sin(\alpha) f_0(\alpha) d\alpha,$$

where F_0 denotes the distribution function of the initial circular symmetric distribution with circular density function $f_0 = F_0'$. Clearly, for $\lambda = 0$, this distribution reduces to the initial circular symmetric distribution F_0.

*A well-known representative is the **sine-skewed von Mises distribution**, where f_0 is the von Mises density (5.34). Let $\Theta \sim vM(0, \kappa)$, then the corresponding sine-skewed von Mises density is given by*

$$f_\Theta(\theta) = \frac{1}{2\pi} e^{\kappa \cos(\theta)} \cdot (1 + \lambda \sin(\theta)).$$

This distribution is not necessarily unimodal. The first trigonometric moments read

$$\mu = arg(\kappa + i\lambda), \qquad \text{and} \qquad R = \frac{I_1(\kappa)}{\kappa I_0(\kappa)} \sqrt{\kappa^2 + \lambda^2},$$

which, for $\lambda = 0$, are the first trigonometric moments of the von Mises distribution given in (5.36).

This distribution is related to other distributions in the following way:

- *If f_0 is the density of the circular uniform distribution, the resulting density f_Θ corresponds to the Cardioid distribution with mode $\theta = \mu + \frac{\pi}{2}$ $(mod\ 2\pi)$ and concentration parameter*

$\kappa = \frac{\lambda}{2}$, *since*

$$\sin(\theta - \mu) = \cos\left((\theta - \mu) - \frac{\pi}{2}\right) = \cos\left(\theta - \left(\mu + \frac{\pi}{2}\right)\right).$$

- *A family of skew-symmetric circular distributions that on the one hand incorporates modeling flexibility and on the other hand includes either representative of the above mentioned classes of distributions is the* ***(sine-skewed) Jones-Pewsey family****. Precisely, the Jones-Pewsey family comprises the von Mises distribution, the Cardioid distribution and the wrapped Cauchy distribution. For more details on sine-skewed distributions in general and this special family of skewed distributions, see, e.g., Jones and Pewsey (2005) [232], Umbach and Jammalamadaka (2009, 2010) [393], [394], and Abe and Pewsey (2011a) [2].*

A **perturbed process** (Θ_t) now is a family of perturbed random variables with marginal distribution given by f_Θ.

5.3 Circular long-memory processes

In this last section, we introduce two different classes of *circular long-memory processes*. As mentioned above, most papers deal with circular Markov-type or at least weakly dependent processes. To the best of our knowledge, circular processes that exhibit long-memory behavior (besides our paper and our book chapter Beran, Steffens and Ghosh, 2021a, 2022 [38], [40], respectively) have only been considered in Di Marzio et al. (2012) [114] and in Beran and Ghosh (2020) [34], who used different approaches for constructing circular long-memory processes.

Di Marzio et al. (2012) [114] introduced the long-range dependence via the autocovariance function according to the characterization of long memory for bivariate processes (see, e.g., Leipus and Viano (2000) [259] and Marinucci (2005) [291]). To sketch their approach shortly, let (ϵ_t) be a $[0, 2\pi)$-valued stationary process such that $E[\sin(\epsilon_t)] = 0$ and such that the autocovariance function satisfies the following three conditions:

(C1) $\gamma_{cc}(k) := E[\cos(\epsilon_t)\cos(\epsilon_{t+k})] \underset{k\to\infty}{\sim} C_{cc}|k|^{2d-1}$,

(C2) $\gamma_{cs}(k) := E[\cos(\epsilon_t)\sin(\epsilon_{t+k})] \underset{k\to\infty}{\sim} C_{cs}|k|^{2d-1}$,

(C3) $\gamma_{ss}(k) := E[\sin(\epsilon_t)\sin(\epsilon_{t+k})] \underset{k\to\infty}{\sim} C_{ss}|k|^{2d-1}$,

where $C_{\cdot\cdot} \in \mathbb{R}\setminus\{0\}$ are constants each.

Then, for $d \in \left(0, \frac{1}{2}\right)$, the process (ϵ_t) is said to exhibit long-range dependence. This characterization of long-range dependence is based on autocovariance functions regularly varying at ∞ with exponent $\alpha = 1 - 2d \in (0, 1)$, see, e.g., Bingham et al. (1989) [56].

Remark. *Note that for small ϵ_t, $\sin(\epsilon_t) \approx \epsilon_t$, hence condition (C3) amounts to the same condition as in the real-valued case. Conditions (C1) and (C2) do not have real-valued correspondents, though.*

In contrast to the definition via the autocovariance function, Beran and Ghosh (2020) [34] define a new and more general class of circular long-memory processes by Gaussian subordination. Analogously as in the case of real-valued Gaussian subordinated processes (see Chapter 1.2.2, Corollary 1.2.12), the long-memory property of the subordinated process is inherited. Such a process is already mentioned in Definition 5.2.5 and in (5.32). In the circular setting, the function G in Definition 5.2.5 can be either of the circular distribution functions discussed in the previous sections.

To this end, let (Z_t) be a Gaussian process with zero mean, unit variance, autocovariance function γ_Z of the form

$$\gamma_Z(k) \underset{k\to\infty}{\sim} c_{\gamma,Z}|k|^{2d-1}, \tag{5.37}$$

and spectral density f_Z of the form

$$f_Z(\lambda) \underset{\lambda\to 0}{\sim} c_{f,Z}|\lambda|^{-2d}$$

with $d \in \left(0, \frac{1}{2}\right)$ and $0 < c_{f,Z}, c_{\gamma,Z} < \infty$. Recall from Part I (1.10) and (1.11), respectively, that such a process exhibits long-range dependence. Moreover, let $G : [0, 2\pi) \to [0, 1]$ be an absolutely continuous circular distribution function with density $g = G'$, and let Φ and $\varphi = \Phi'$ denote the standard normal distribution and its density function, respectively.

From Definition 5.2.5 we know that the respective circular Gaussian subordinated process (Θ_t) is defined by

$$\Theta_t = G^{-1}(\Phi(Z_t)), \qquad t \in \mathbb{Z}. \tag{5.38}$$

This means, we can write

$$\Theta_t = \Psi(Z_t),$$

with $\Psi := G^{-1} \circ \Phi \in L^2(\mathbb{R})$.

Recall the concept of Hermite polynomial expansion introduced in Chapter 1.2.2 for any function $H \in L^2(\mathbb{R})$:

$$H(Z) = E[Z] + \sum_{k=m}^{\infty} \frac{J(k)}{k!} H_k(Z) \tag{5.39}$$

with Hermite coefficients

$$J(k) = \langle H, H_k \rangle = E\left[H(Z)H_k(Z)\right] = \int_{-\infty}^{\infty} H(z)H_k(z)\varphi(z)dz \tag{5.40}$$

and Hermite rank $m \geq 1$ (see Lemma 1.2.9).

For the circular Gaussian subordinated process $(\Theta_t) = \left(G^{-1}\left(\Phi(Z_t)\right)\right)$ with (constant) mean direction μ, we consider the function $H(z) := \sin\left(\Theta_t - \mu\right)$ with Hermite polynomial expansion given by (5.39). Then, applying Lemma 1.2.11, the autocovariance function of H asymptotically behaves as

$$\gamma_H(k) \underset{k\to\infty}{\sim} \sum_{l=1}^{\infty} \frac{J^2(l)}{l!}\gamma_Z^l(k).$$

Based on the circular correlation ϱ_{JS} of a stationary process given in (5.25), it is shown in Beran and Ghosh (2020) [34], Lemma 1, that the long-memory property of (Z_t) is inherited by (Θ_t). We reformulate this Lemma in terms of covariances:

Lemma 5.3.1. ***(Beran and Ghosh, 2020)***

Let (Z_t) with autocovariance function γ_Z and (Θ_t) be defined as in (5.38). Suppose that the Hermite rank $m \geq 1$ of $H(z) = \sin\left(G^{-1}(\Phi(z)) - \mu\right)$ is such that $0 < m(1-2d) < 1$. Then

$$\gamma_H(k) \underset{k\to\infty}{\sim} c_m k^{2d_m - 1} \tag{5.41}$$

with

$$d_m = \frac{1}{2}\left(1 + m(2d-1)\right) \in \left(0, \frac{1}{2}\right) \quad \text{and} \quad c_m := \frac{J^2(m)}{m!}c_{\gamma,Z}^m$$

where $c_{\gamma,Z}$ is as in (5.37) and the Hermite coefficients $J(m)$ are as in (5.40).

Note that this is the same result as in the real-valued case, see Beran et al. (2016) [35], and Lemma 1.2.11 and Corollary 1.2.12 in Chapter 1.2.2.

Remark 5.3.2. *From Lemma 5.3.1, we can (analogously as in the real-valued case, see Corollary 1.2.12)) conclude the following: Under the assumption of Hermite rank $m = 1$, we have*

$$d_1 = \frac{1}{2}\left(1 + (2d-1)\cdot 1\right) = d,$$

which means that the memory parameter d_1 of the circular process (Θ_t) coincides with the memory parameter d of the corresponding real-valued process (Z_t). Consequently, (5.41) becomes

$$\begin{aligned} \gamma_H(k) &\underset{k\to\infty}{\sim} J^2(1)\gamma_Z(k) \\ &\underset{k\to\infty}{\sim} J^2(1)c_{\gamma,Z}|k|^{2d-1}, \end{aligned}$$

where $J(1) = \int zH(z)\varphi(z)dz$*, and the asymptotic behavior of the spectral density of* H *reads*

$$\begin{aligned} f_{\lambda,H} &\underset{\lambda\to 0}{\sim} c_{f,H}|\lambda|^{-2d} \\ &= J^2(1)c_{f,Z}|\lambda|^{-2d}. \end{aligned}$$

Chapter 6

Estimation of the mean direction

In this chapter, we discuss circular regression models and the estimation of time series components, mainly of the mean direction. The setting is essentially the same as in the real-valued case (see Chapter 3) but in the state space $[0, 2\pi)$. We consider the general *circular nonparametric regression model*, a random-design regression model, introduced by Gould (1969) [169]:

$$\Theta_t = [\nu(\Psi_t) + Z_t] \ \ mod\ 2\pi, \tag{6.1}$$

where (Θ_t, Ψ_t) is a bivariate circular process, (Ψ_t) is a stationary circular process independent of the error process (Z_t) and ν is the unknown mean direction.

Of interest, for instance, is estimation of this unknown mean direction ν. Nonparametric estimation on the circle in general is, e.g., discussed in Presnell et al. (1998) [327], Downs and Mardia (2002) [125] and Kato et al. (2008) [241].

The general random-design model captures three different types of models, where either the explanatory variables Ψ_t and/or the response variables Θ_t are $[0, 2\pi)$-valued and realizations can be identified with angles. These models are called *linear-circular* (real-valued explanatory variables and circular response variables), *circular-linear* (vice versa) and *circular-circular models* (both variables are $[0, 2\pi)-$valued), respectively.

Our focus in this chapter is on the circular-circular regression model given in (6.1). This is the circular analogue of the real-valued random-design model defined in (3.1). We will introduce both, the *circular kernel density estimation* and the *kernel estimation of the mean direction* using *circular kernel functions*, which are the circular analogues of the real-valued kernels defined in Part I. In addition, for the sake of completeness, we will review the results for parametric estimation in the fixed-design setting and discuss possible estimators for the two circular mixed models.

6.1 Parametric estimation of the mean direction - literature review

The general *circular fixed-design time series model* is of the form

$$\Theta_t = [\nu_t + Z_t] \ mod\, 2\pi, \tag{6.2}$$

where ν_t is the unknown mean direction and (Z_t) is a stationary zero-mean error process. In the simplest case of a circular time series model with a constant mean direction $\nu_t = \mu$, (6.2) reduces to

$$\Theta_t = [\mu + Z_t] \ mod\, 2\pi. \tag{6.3}$$

Recall from (5.17) that the *mean direction μ of a random angle Θ* is defined by

$$E\left[e^{i\Theta}\right] = Re^{i\mu}.$$

In case of (6.3), assertion (5.18) on the first trigonometric moments about the mean direction implies

$$E\left[\sin\left(\Theta_t - \mu\right)\right] = E\left[\sin\left(Z_t\right)\right] = 0 \qquad \text{and} \qquad E\left[\cos\left(\Theta_t - \mu\right)\right] = E\left[\cos\left(Z_t\right)\right] = R > 0. \tag{6.4}$$

For the second moments we obtain from (5.15)

$$E\left[\sin\left(2(\Theta_t - \mu)\right)\right] = \beta_2 \qquad \text{and} \qquad E\left[\cos\left(2(\Theta_t - \mu)\right)\right] = \alpha_2. \tag{6.5}$$

For parametric approaches on estimating the mean direction in general, see, e.g., Fisher and Lee (1992) [139] and Presnell et al. (1998) [327], and see Kim and SenGupta (2017) [244] in a multivariate setting. Asymptotic results on parametric regression of the mean direction can, e.g., be found in Johnson and Wehrly (1978) [226], Fisher and Lee (1992) [139], Mardia and Jupp (1999) [288], Jammalamadaka and SenGupta (2001) [224], Pewsey et al. (2013) [320], Kim and SenGupta (2016) [243], and references therein. However, these references only discuss iid or weakly dependent residuals, whereas we are interested in long-range dependent residuals.

Given a random sample $\Theta_1, \ldots, \Theta_n$ of (Θ_t), a standard estimator of the mean direction μ is defined through

$$\bar{R}\exp\left(i\hat{\mu}\right) = \left(\bar{C} + i\bar{S}\right), \tag{6.6}$$

i.e., $\hat{\mu} = arg\left\{\bar{C} + i\bar{S}\right\}$, where $\bar{C}$ and $\bar{S}$ are as given in (5.6),

$$\left(\bar{C},\ \bar{S}\right) = \left(\frac{1}{n}\sum_{t=1}^{n}\cos\left(\Theta_t\right),\ \frac{1}{n}\sum_{t=1}^{n}\sin\left(\Theta_t\right)\right)$$

and $\bar{R} = \sqrt{\bar{C}^2 + \bar{S}^2}$. As a consequence, we obtain

$$\bar{R}\sin\left(\hat{\mu} - \mu\right) = \frac{1}{n}\sum_{t=1}^{n}\sin\left(Z_t\right) \qquad \text{and} \qquad \bar{R}\cos\left(\hat{\mu} - \mu\right) = \frac{1}{n}\sum_{t=1}^{n}\cos\left(Z_t\right). \tag{6.7}$$

Under the assumption of iid residuals, by applying the CLT for $\bar{C}$ and $\bar{S}$, Fisher and Lewis (1983) [141] derive the asymptotic distribution of the estimator as follows:

$$n^{\frac{1}{2}}\bar{R}\sin\left(\hat{\mu} - \mu\right) \xrightarrow{d} \sqrt{\frac{1-\alpha_2}{2}}W_{\sin} \tag{6.8}$$

and

$$n^{\frac{1}{2}}\bar{R}\cos\left(\hat{\mu} - \mu\right) \xrightarrow{d} \left(\sqrt{\frac{1+\alpha_2}{2} - R^2}\right)W_{\cos}.$$

where $W_{\sin} \sim N(0,1)$ and $W_{\cos} \sim N\left(R, 1\right)$, and α_2 is as defined in (6.5).

This follows from (6.7) together with (6.4), (6.5), and since

$$\begin{aligned} \operatorname{var}\left(\sin\left(\Theta_t - \mu\right)\right) = \operatorname{var}\left(\sin\left(Z_t\right)\right) &= E\left[\sin^2\left(Z_t\right)\right] \\ &= E\left[\frac{1}{2}\left(1 - \cos\left(2Z_t\right)\right)\right] \\ &= \frac{1}{2}\left(1 - \alpha_2\right) \end{aligned} \tag{6.9}$$

and

$$\begin{aligned} \operatorname{var}\left(\cos\left(\Theta_t - \mu\right)\right) = \operatorname{var}\left(\cos\left(Z_t\right)\right) &= E\left[\cos^2\left(Z_t\right)\right] - R^2 \\ &= E\left[\frac{1}{2}\left(\cos\left(2Z_t\right) + 1\right)\right] - R^2 \\ &= \frac{1}{2}\left(1 + \alpha_2\right) - R^2. \end{aligned} \tag{6.10}$$

Since (6.9) does not involve the nuisance parameter R (in contrast to (6.10)), the first one is used in Fisher and Lewis (1983) [141] to deduce a simple approximate asymptotic $(1-\alpha)$-confidence interval for μ (based on the CLT for the sine moments (6.8)) of the form

$$\hat{\mu} \pm \arcsin\left(n^{-\frac{1}{2}}\frac{1}{\bar{R}}\sqrt{\frac{1-\alpha_2}{2}}q_{1-\alpha/2}\right), \tag{6.11}$$

where $q_{1-\alpha/2}$ denotes the $(1-\alpha/2)$-quantile of the standard normal distribution. As it is in the real-valued case, the same convergence holds for short-range dependent residuals.

To the best of our knowledge (besides Beran, Steffens and Ghosh, 2021a, 2022 [38], [40]), there are only two more references that consider long-range dependent residuals, Di Marzio et al. (2012) [114] and Beran and Ghosh (2020) [34]. Either discuss estimation of the mean direction in the fixed-design setting, but, while Di Marzio et al. (2012) apply nonparametric regression methods, Beran and Ghosh (2020) consider a parametric approach based on (6.6) and thereby extend the results from Fisher and Lewis (1983) [141] to long-memory time series models.

More precisely, Beran and Ghosh (2020) show that under the assumption of long-range dependence, the limiting distribution of the sine moments in (6.7) again is a degenerate Gaussian random variable but with a lower convergence rate. Let therefore $(\Theta_t) = (G^{-1}(\Phi(Z_t)))$ be a circular Gaussian subordinated long-memory process defined by (5.38) and let $H(z) = \sin\left(G^{-1}(\Phi(z)) - \mu\right)$ has a Hermite representation of the form (5.39). Recall from Lemma 5.3.1 that the covariance of a circular Gaussian subordinated long-range dependent random variable is associated with the covariance of the subordinated process through the Hermite polynomials. Based on the CLT for the sine moments, Beran and Ghosh, derive the asymptotic distribution and an asymptotic $(1-\alpha)$-confidence interval for $\hat{\mu}$ as follows:

Theorem 6.1.1. ***(Beran and Ghosh, 2020)***

Let $(\Theta_t) = (G^{-1}(\Phi(Z_t)))$ be defined by (5.38). Suppose that $H(z) = \sin\left(G^{-1}(\Phi(z)) - \mu\right)$ has Hermite rank $m = 1$, i.e., $J(1) \neq 0$ in (5.40), then

$$n^{\frac{1}{2}-d}\bar{R}\sin\left(\hat{\mu}-\mu\right) \xrightarrow{d} J(1)\sqrt{c_{f,Z}\cdot v(d)}W,$$

where $W \sim N(0,1)$ and

$$v(d) = \frac{2\sin(\pi d)}{d(2d+1)}\Gamma(1-2d), \qquad d \in \left(0, \frac{1}{2}\right).$$

The corresponding asymptotic $(1-\alpha)$-confidence interval is given by

$$\hat{\mu} \pm \arcsin\left(n^{d-\frac{1}{2}}\frac{J(1)}{\bar{R}}\sqrt{c_{f,Z}\cdot v(d)}q_{1-\alpha/2}\right), \tag{6.12}$$

where $q_{1-\alpha/2}$ again is the $(1-\alpha/2)$-quantile of the standard normal distribution.

This theorem shows both, that under long-range dependence the convergence rate is slower than under independence or weak dependence and that the convergence is essentially the same as in the real-valued long-memory situation for convergence of the sample mean (given in (2.10)).

In addition, for $d = 0$, i.e., in the short-memory case (setting $v(0) = \lim_{d\to 0} v(d) = 2\pi$), or under independence, (6.12) is asymptotically equivalent to the asymptotic confidence interval (6.11) derived by Fisher and Lewis (1983) [141].

However, note that this theorem assumes that the Hermite rank of $H(z)$ is $m = 1$. In the case of Hermite rank $m \geq 2$ the limiting distribution is (with the same arguments as in the real-valued situation) no longer Gaussian. For an explanation on the assumption of Hermite rank $m = 1$ in this context, see the discussion on higher Hermite ranks in Part I and Remark 2 in Beran and Ghosh (2020) [34].

Remark. *As already discussed in Part I in the context of long-range dependence, in practice, (6.12) is not directly applicable, since it involves unknown long-memory parameters d and $c_{f,Z}$, and, in general, an unknown first Hermite coefficient $J(1)$. For consistent estimation of the long-memory parameters, we can apply the methods discussed in Chapter 3 on $\hat{H}(z) = \sin(\Theta_t - \hat{\mu})$, since, according to Remark 5.3.2, the spectral density of H is given by*

$$f_H(\lambda) \sim c_{f,H}|\lambda|^{-2d} = J^2(1)c_{f,Z}|\lambda|^{-2d}.$$

The resulting approximate confidence interval then reads

$$\hat{\mu} \pm \arcsin\left(n^{\hat{d}-\frac{1}{2}} \frac{1}{\bar{R}} \sqrt{\hat{c}_{f,H} v(\hat{d})} q_{1-\alpha/2}\right). \tag{6.13}$$

In addition, since

$$f_Z\left(n^{-1}\right) \sim c_{f,Z} \cdot n^{2d},$$

we obtain an equivalent approximate confidence interval

$$\hat{\mu} \pm \arcsin\left(n^{-\frac{1}{2}} \frac{1}{\bar{R}} \sqrt{\hat{f}_Z\left(n^{-1}\right) \cdot v(\hat{d})} q_{1-\alpha/2}\right).$$

For more details, see Beran and Ghosh (2020) [34].

Beran and Ghosh (2020) [34] further extend this result to non-stationary time series models, where $\nu_t = \nu(x_t; \beta)$ in (6.2) with explanatory variables x_t and an unknown parameter vector β. The error process (Z_t) is such that (6.4) holds, and the associated stationary process $(Y_t) = (\sin(Z_t))$ has an autocovariance function and a spectral density such that (Y_t) exhibits long-range dependence, i.e., such that (1.10) and (1.11) hold. In the circular setting, the additive time series model is of the form

$$\Theta_t = [\nu_t + Z_t] \ mod\ 2\pi = [g_t + S_t + Z_t] \ mod\ 2\pi,$$

where the mean direction ν_t is decomposed into a non-periodic trend g_t and a seasonal component S_t. In an analogous manner as in the real-valued situation, the non-periodic trend can be modeled by

$$g_t = g(x_t; \beta_g) = \left[\beta_{g,0} + \sum_{j=1}^{p} \beta_{g,j} \cdot t^j \right] \ mod\ 2\pi,$$

and the seasonal component with period $T_0 \geq 2$ can be modeled by

$$S_t = S(x_t; \beta_S) = \left[\sum_{j=1}^{q} \beta_{S,j} \sin(\lambda_j t) + \beta_{S,q+j} \cos(\lambda_j t) \right] \mod 2\pi, \tag{6.14}$$

where, for fixed $T_0 \geq 2$, $\lambda_j = \frac{2\pi j}{T_0}$, $j = 1, \ldots, q$.

Since the estimator $\hat{\nu}$ of the mean direction defined by (6.6) minimizes the risk function

$$Q_n(\nu) := \sum_{t=1}^{n} (1 - \cos(\Theta_t - \nu_t)), \tag{6.15}$$

defined via the circular distance (5.4), replacing ν_t in (6.15) by $\nu(x_t; \beta)$, $\beta = (\beta_g, \beta_S)$ generalizes this minimization procedure to parametric estimation of β. Beran and Ghosh (2020) [34] derive asymptotic confidence intervals for both, the non-periodic trend component and the seasonal component. For both components, the distribution of the coefficient vector β is asymptotically normal. But, analogously as in the real-valued case, only the convergence rate of $\hat{\beta}_g$ depends on the long-memory parameter d. The convergence rate of $\hat{\beta}_S$ is the same as under independence or weak dependence, see the respective results for real-valued seasonal time series in Beran, Steffens and Ghosh (2018) [37]. Both results are based on the conditions introduced in Grenander and Rosenblatt (1984) [174] and investigations of Yajima (1988, 1991) [423], [424] in the context of linear time series regression. Using analogous arguments as in Yajima (1991) [424] and in Beran, Steffens and Ghosh (2018) [37], it is further shown that the estimated coefficients for the trend component are asymptotically independent of those for the seasonal components and hence, a simultaneous confidence interval can be defined.

6.2 Circular kernel density estimation

In this chapter we focus on *circular kernel density estimation* and introduce the so-called *circular kernels*, which in some way are similarily defined as in the real-valued situation. The kernel density estimation in the circular state space was introduced by Beran (1979) [43] and further discussed in various references including Watson (1983) [411], Hall et al. (1987) [187], Bai et al. (1989) [15], Fisher (1989) [136], Fisher et al. (1993) [142], Jammalamadaka and SenGupta

(2001) [224], and Taylor (2008) [387]. For properties on higher order kernels on the circle see Tsuruta and Sagae (2017) [391]. However, all these references consider iid residuals only.

We consider a $[0, 2\pi)$-valued stationary process (Ψ_t) and we suppose that (Ψ_t) has a marginal distribution function F_Ψ with probability density function $f_\Psi = F'_\Psi$. As in the real-valued case, a simple way to gain knowledge about the underlying distribution is the circular kernel density estimation. The main techniques are essentially the same as those on the real line. For definitions and notions in this chapter, we refer to Hall et al. (1987) [187] and Tsuruta and Sagae (2017) [391], and adapt it to our situation:

Definition 6.2.1. *Given a function $L \geq 0$ and $\kappa > 0$, we define a circular kernel by*

$$\begin{aligned} K_\kappa : [0, 2\pi) &\to \mathbb{R}^+ \\ \psi &\mapsto c_\kappa^{-1}(L) \cdot L\left(\kappa(1 - \cos(\psi))\right), \end{aligned}$$

where $c_\kappa(L)$ is a normalizing constant such that K_κ integrates to unity.

*Instead of the **concentration parameter** κ, we define the so-called **circular bandwidth** as $b := \kappa^2$, the circular normalizing constant as c_b and redefine circular kernels as*

$$K_b(\psi) = c_b^{-1}(L) \cdot L\left(\frac{1 - \cos(\psi)}{b^2}\right), \qquad \psi \in [0, 2\pi). \tag{6.16}$$

This means, the concentration parameter κ plays the same role as the inverse of the squared kernel bandwidth b of real-valued kernels.

By definition, we have

$$\int K_b(\psi) d\psi = 1,$$

consequently, since we assume $L \geq 0$, K_b is a circular probability density function.

Definition 6.2.2. *The **l-th moment of L** with even integer $l \geq 0$ is defined as*

$$m_l(L) = \int_0^\infty L(r) r^{\frac{l-1}{2}} dr.$$

For asymptotic considerations, the following assumptions on L from Tsuruta and Sagae (2017) [391] are needed:

(CK1) The first derivative $L'(r) := \frac{dL(r)}{dr}$ is continuous.

(CK2) $L(r) r^{\frac{v+1}{2}} \to 0$ as $r \to \infty$.

(CK3) For $t = 0, 1$:

$$\int_0^\infty u^{2t} L^2\left(\frac{u^2}{2}\right) du < \infty.$$

(CK4) For any even number $v \geq 0$, the l-th moment of L $m_l(L)$ is bounded for $0 \leq l \leq v$, and it can be decomposed into

$$m_l(L) = m_{y,l}(L) + \mathcal{O}\left(y^{-\frac{v+1}{2}}\right),$$

where $m_{y,l}(L) := \int_0^y L(r) r^{\frac{l-1}{2}} dr$.

For circular as opposed to real-valued kernels, these assumptions are satisfied for all smooth and rapidly varying functions L that are asymptotically equivalent to the kernel corresponding to $L(r) = e^{-r}$. One popular representative of this class of circular kernels is the *von Mises kernel*, since with $r = \kappa(1 - \cos(\psi))$ the resulting circular kernel K_b is asymptotically equivalent to the von Mises density with $\mu = 0$ and $\kappa = b^{-2}$. Further circular densities that fulfill these 4 assumptions are the wrapped normal density and the wrapped Cauchy density, although the latter is not of the form $L(\kappa(1 - \cos(\psi)), \psi \in [0, 2\pi)$.

We define the higher-order kernels as follows (see Tsuruta and Sagae, 2017 [391]):

Definition 6.2.3. *Let $p \geq 2$ be even. K_b is called a **circular kernel of order p** if conditions (CK1) to (CK4) hold with $v \geq p + 2$ and*

$$m_l(L) \begin{cases} \neq 0, & l = 0, p \\ = 0, & l = 2, 4, \ldots, p - 2. \end{cases}$$

Remark. *Methods for constructing higher-order kernels from lower-order kernels are discussed in Tsuruta and Sagae (2017) [391]. In our paper (Beran, Steffens and Ghosh, 2021b [39]), we present one example to construct higher-order kernels coming from $L_{\kappa,2}(\psi) = e^{-\kappa(1-\cos(\psi))}$ by use of the additive method in Tsuruta and Sagae (2017) [391], i.e., by using the kernel and its derivatives. Note that $L_{\kappa,2}(\psi)$ is the main part of the so-called von Mises kernel function given in (6.20). More precisely, setting $\kappa = b^{-2}$,*

$$K_{b,2}(\psi) = C_b^{-1} L_{\kappa,2}(\psi)$$

is a second-order kernel, since with $L(r) = e^{-r}$, the l-th moment of L, $l = 2k$, $k \geq 1$, reads

$$m_l(L) = \int e^{-r} r^{\frac{l-1}{2}} dr = \Gamma\left(l + \frac{1}{2}\right),$$

and consequently,

$$m_0(L) = \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \neq 0 \qquad \textit{and} \qquad m_2(L) = \frac{3}{4}\sqrt{\pi} \neq 0.$$

Then, circular kernels of order $p+2$ can be recursively constructed by

$$L_{p+2}(r) = \frac{p+1}{2} L_p(r) + \frac{2}{p} r L_p'(r).$$

A general circular kernel estimator is defined in an analogous way as in the real-valued case:

Definition 6.2.4. *Let $\Psi_1, \ldots, \Psi_t$ be a random sample of angles from the unknown circular distribution F_Ψ with density $f_\Psi = F_\Psi'$. For $\psi^0 \in [0, 2\pi)$, given a circular kernel K_b with circular bandwidth b the* ***circular kernel density estimator*** *of f_Ψ is defined by*

$$\hat{f}_\Psi(\psi^0) = \frac{1}{n} \sum_{j=1}^{n} K_b\left(\Psi_j - \psi^0\right), \qquad \psi^0 \in [0, 2\pi). \tag{6.17}$$

This kernel density estimator is similar to the PC-estimator on the real line defined in (3.9). However, for circular kernels, the notion of „being in the neighborhood“ is incorporated via the respective circular distance measure given in (5.4).

Remark. *Note that in some references, cf. Hall et al. (1987) [187] and Jammalamadaka and SenGupta (2001) [224], the term $\Psi_j - \psi$ in (6.17) is replaced by $1 - x^T x_j$, where x corresponds to the point on the unit circle. In view of (5.5), this yields an equivalent representation. The resulting kernel estimator reads:*

$$\hat{f}_\Psi(x) = \frac{1}{n} \sum_{i=1}^{n} K_b(1 - x^T x_j).$$

Recall that a kernel density estimator coincides with the density estimator of the corresponding characteristic function, and any circular density has a Fourier representation as given in (5.16). Hence, under assumptions (CK1) to (CK4), the kernel K_b can be expanded into a Fourier series of the form

$$K_b(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \alpha_s(b) e^{-isu}, \quad u \in [0, 2\pi), \tag{6.18}$$

with Fourier coefficients

$$\alpha_{-s}(b) = \alpha_s(b) = \int_0^{2\pi} K_b(u) e^{isu} du. \tag{6.19}$$

This representation will be later (see Chapter 7.4) utilized to prove uniform convergence of our circular kernel estimator.

Remarks. *(i) Note that in case of symmetric circular densities (see (5.19)), the Fourier series reduces to*

$$K_b(u) = \frac{1}{2\pi}\left(1 + 2\sum_{s=1}^{\infty}\alpha_s(b)\cos(su)\right), \qquad u \in [0, 2\pi).$$

(ii) Assuming a symmetric function L, Condition (CK4) can be replaced by a condition on so-called sine-type moments as discussed in Di Marzio et al. (2009, 2011) [112], [113], and Tsuruta and Sagae (2017) [391]. This condition is similar to the moment condition on symmetric kernels on the real line.

(iii) The class of sine-type kernels of order 2 include the frequently used von Mises, wrapped normal and wrapped Cauchy kernels.

(iv) By use of higher-order kernels, one obtains the respective higher-order kernel estimators. Further higher-order kernel estimators can be constructed by combining two kernel density estimators with different kernel bandwidths, as it is, e.g., done for Terrell and Scott's kernel density estimator, see Terrell and Scott (1980) [388] for the real-valued case. This multiplicative method is proposed in Tsuruta and Sagae (2017) [391].

Let's look at two typical circular kernels:

Example 6.2.5. *(i) Recall that for a random variable $\Theta \sim vM(\mu, \kappa)$ the probability density function is given by*

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(\theta-\mu)}.$$

Hence, the corresponding von Mises kernel with $\kappa = b^{-2}$ is of the form

$$K_b(u) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(u)} \tag{6.20}$$

with Fourier expansion

$$K_b(u) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\left|\frac{I_s(\kappa)}{I_0(\kappa)}\right| e^{-isu},$$

where $I_p(\kappa)$, $p = 0, 1, \ldots$, is the modified Bessel function of order p defined in (5.35). This

follows from the symmetric nature of the distribution, since

$$\begin{aligned}\alpha_s(b) &= \int_0^{2\pi} \frac{1}{2\pi I_0(\kappa)} e^{\kappa\cos(u)} e^{isu} du \\ &= \frac{1}{I_0(\kappa)} \cdot \frac{1}{2\pi} \int_0^{2\pi} e^{\kappa\cos(u)} \cdot \cos(su) du \\ &= \frac{1}{I_0(\kappa)} I_s(\kappa).\end{aligned}$$

(ii) The wrapped normal kernel obtained from the probability density function of a wrapped normal distribution $WN(\mu, \sigma^2)$ reads

$$f(\theta; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{j=-\infty}^{\infty} \exp\left(-\frac{(\theta - \mu + 2\pi j)^2}{2\sigma^2}\right).$$

With $\sigma = b$, the corresponding wrapped normal kernel is of the form

$$K_b(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{j=-\infty}^{\infty} \exp\left(-\frac{(u + 2\pi j)^2}{2\sigma^2}\right) \tag{6.21}$$

with Fourier representation

$$K_b(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \exp\left(-\frac{\sigma^2}{2} s^2\right) e^{-isu}.$$

This results from the Fourier transform of the real-valued normal distribution.

In circular kernel density estimation, the concentration parameter κ as the squared inverse linear bandwidth controls the smoothing of the estimator: As κ increases, the estimator undersmoothes the underlying density, and, as κ decreases, the estimator oversmoothes local features and approaches a global average. Thus, as well as in the real-valued setting, the crucial problem in kernel estimation on the circle is the choice of the optimal circular bandwidth or the optimal concentration parameter.

This issue is not discussed here, however, in the literature there are different approaches that are similar to the real-valued ones. For instance, Hall et al. (1987) [187] propose a least squares cross-validation (LSCV) method, a rule of thumb based on the concentration of the von Mises distribution as a reference density is introduced by Taylor (2008) [387], and Di Marzio et al. (2011) [113] propose a direct plug-in rule (DPI). The plug-in rule procedure introduced by Oliveira et al. (2012) [310] extends the approach by Taylor (2008) [387] to more flexible

reference densities such as mixtures of von Mises distributions. Tsuruta and Sagae (2020) [392] derive asymptotic properties of the LSCV and the DPI selector.

6.3 Circular kernel regression

As mentioned in the beginning of Chapter 6, there are three different random-design models that contain $[0, 2\pi)$-valued components: the *linear-circular*, the *circular-linear* and the *circular-circular regression model*. The two mixed models apply to bivariate random variables, where one variable is $[0, 2\pi)$-valued and the other variable is real-valued. Our main focus in this thesis is on the third model, where both, the explanatory and the response variables, are circular, but first of all, we will shortly have a look at the two mixed models.

In contrast to the model given in (6.1), the **circular-linear regression model** is a circular model with linear response of the form

$$X_t = \nu(\Psi_t) + Z_t,$$

where X_t are real-valued, $\nu : [0, 2\pi) \to \mathbb{R}$ is the mean function, and Ψ_t are circular random variables. For such models, there exist different approaches, since the standard (real-valued) theory only needs adaptations. There are, for instance, circular-linear distributions generated based on the *principle of maximum entropy* or on *a Bayesian approach*, or a variant of the NW-estimator (see (3.8) and (6.24)) for circular-linear data. For more details, see, e.g., Johnson and Wehrly (1978) [226], George and Ghosh (2006) [149], and Oliveira et al. (2013) [311] for an univariate circular-linear regression model, see SenGupta and Ugwuowo (2006) [354] for circular-linear models in a multivariate setting, and see García-Portugués et al. (2013) [144] for kernel density estimation in the context of circular-linear data. For an extension of the real-valued local polynomial fitting to the circular-linear case, see Di Marzio et al. (2009) [112].

The **linear-circular regression model** as a model with circular response variables Θ_t and real-valued explanatory variables X_t is of the form

$$\Theta_t = [\nu(X_t) + Z_t] \ \ mod\, 2\pi, \tag{6.22}$$

where $\nu : [0, 1] \to (-\pi, \pi]$ is the mean function. Under the assumption of long-range dependence for the respective fixed-design model, such a model is considered by Di Marzio et al. (2012, 2013) [114], [115]. Their approach is shortly reviewed at the end of this chapter as they consider nonparametric estimation under long-range dependence.

These mixed regression models, in particular, arise in the environmental sciences, meteorology

or oceanography. Typical examples are animal movements, ocean velocities and directions, or the relationship between wind direction and wind speed, or wind direction and level of ozone concentration or pollutants, see, e.g., Johnson and Wehrly (1978) [226], Fisher and Lee (1992) [139], Jammalamadaka and Lund (2006) [222], SenGuptaUgwuowo (2006) [354], Di Marzio et al. (2013) [115], and Oliveira et al. (2013) [311].

Turning to the **circular-circular regression model** given in (6.1): Let (Θ_t, Ψ_t) be a bivariate circular process, where

$$\Theta_t = [\nu(\Psi_t) + Z_t] \ mod\ 2\pi \tag{6.23}$$

with stationary process (Ψ_t) and a stationary error process (Z_t) such that

$$E\left[\sin\left(Z_t\right)\right] = 0 \quad \text{and} \quad E\left[\cos\left(Z_t\right)\right] = R(\Psi_t) > 0.$$

For such a model, nonparametric estimation amounts to conditional estimation of the mean direction of (Θ_t) given $\Psi_t = \psi^0$, denoted by $\nu(\psi^0)$, which is defined by

$$E\left[e^{i\Theta} \mid \Psi_t = \psi^0\right] = R(\psi^0)e^{i\cdot\nu(\psi^0)}.$$

Nonparametric circular-circular regression with iid residuals can be found in various references including Gould (1969) [169], Johnson and Wehrly (1978) [226], Fisher and Lee (1992) [139], Mardia and Jupp (1999) [288], Jammalamadaka and SenGupta (2001) [224], Kato et al. (2008) [241], Pewsey et al. (2013) [320], Di Marzio et al. (2013, 2014) [115], [116], Polsen and Taylor (2015) [326], and Kim and SenGupta (2016) [243].

In this chapter, we only focus on circular kernel estimation of the mean direction and will review recent results in both, the iid and the long-memory situation.

A circular kernel estimator for the conditional mean direction ν can be defined in an analogous manner as in the real-valued situation.

Definition 6.3.1. *Let K_b be a circular kernel function as defined in (6.16). Given a sample $\Psi_1, \ldots, \Psi_n$, and $\psi^0 \in [0, 2\pi)$ fixed, the **circular NW-estimator** is defined by*

$$\hat{\nu}(\psi_0) = \frac{\frac{1}{n}\sum\limits_{t=1}^{\infty} K_b(\Psi_t - \psi_0)\Theta_t}{\hat{f}_\Psi(\psi_0)}, \tag{6.24}$$

where $\hat{f}_\Psi$ as given in (6.17) is the kernel density estimator of the marginal density f_Ψ.

This circular kernel estimator is the circular version of the (real-valued) *NW-estimator* introduced in (3.8).

In the iid situation, e.g., Qin et al. (2011) [331] considered local least squares kernel regression and derived asymptotic properties of the conditional bias and the conditional variance. Di Marzio et al. (2013) [115] derived asymptotic formulas for the asymptotic mean squared error (AMSE), the asymptotic mean integrated squared error (AMISE), and the optimal smoothing degree for both, the linear-circular and the circular-circular model. Nonparametric regression in higher-order state spaces is, e.g., considered in Chang (1986) [74], Downs (2003) [124], Di Marzio et al. (2014) [116], and, in the context of spherical-spherical regression, in Rivest et al. (2016) [337], Di Marzio et al. (2019) [117], and Meilán-Vila et al. (2019, 2021) [295], [296].

Di Marzio et al. (2013) [115] extend the results of Di Marzio et al. (2009, 2012) [112], [114] to nonparametric regression for circular responses and residuals with zero mean direction and finite concentration that are independent of the explanatory variables Ψ_t. They consider the extension to both, circular-circular regression models of the form (6.23) and linear-circular regression models of the form (6.22). Hence, the kernel estimator, again given by the atan2 function (5.13), is either defined using a circular or a real-valued kernel function.

Recall that none of the two papers that addresses circular time series models under the assumption of long-range dependence consider a circular-circular regression model. Beran and Ghosh (2020) [34] address parametric estimation of the mean direction of Gaussian subordinated long-memory processes defined in Definition 5.2.5 (see Chapter 6.1). A nonparametric approach is discussed by Di Marzio et al. (2012) [114]. They consider a model of the form

$$\Theta_t = \left[\nu\left(\frac{t}{n}\right) + Z_t\right] \ mod\, 2\pi$$

with rescaled time points $\frac{t}{n} \in [0,1]$. Here, $\nu : [0,1] \to (-\pi, \pi]$ is an unknown mean function, and (Z_t) is a $(-\pi, \pi]$-valued stationary error process such that $E\left[\sin\left(Z_t\right)\right] = 0$, and with an autocovariance function satisfying conditions (1) to (3) given in Chapter 5.3. Recall that, for $d \in \left(0, \frac{1}{2}\right)$, this process exhibits long-range dependence.

Since such a regression model corresponds to the class of *linear-circular fixed-design models*, estimating the mean function ν requires real-valued kernels only. Di Marzio et al. (2012) [114] define a special kernel estimator, which separately smooths the sine and the cosine series and recombines them using the inverse tangent function, i.e., ν can be written as

$$\nu\left(\frac{t}{n}\right) = \left(\nu_1\left(\frac{t}{n}\right),\ \nu_2\left(\frac{t}{n}\right)\right)^T,$$

with $\nu_1\left(\frac{t}{n}\right) := E\left[\cos\left(\Theta_t\right)\right]$ and $\nu_2\left(\frac{t}{n}\right) := E\left[\sin\left(\Theta_t\right)\right]$. They use the same risk measure (6.15) but yield a minimizer of the form $atan2(\nu_1, \nu_2) = arg\left\{\nu_1 + i \cdot \nu_2\right\}$ as defined in (5.13).

The corresponding kernel estimators then result in

$$\hat{\nu}_1\left(\frac{t}{n}\right) := \frac{1}{n}\sum_{j=1}^{\nu} \cos\left(\Theta_j\right) W\left(\frac{j-t}{n}\right)$$

and

$$\hat{\nu}_2\left(\frac{t}{n}\right) := \frac{1}{n}\sum_{j=1}^{n} \sin\left(\Theta_j\right) W\left(\frac{j-t}{n}\right),$$

where W, e.g., is a symmetric real-valued kernel K_b with kernel bandwidth b satisfying the conditions (K1) to (K3) on the kernel given in Chapter 3.1.1. Hence, kernel estimation of the mean function for such a circular equidistant fixed-design model reduces to kernel estimation in the real-valued situation, and the results for the convergence rate of both, the estimator and the optimal bandwidth, are the same as in the case of linear equidistant fixed-design models (3.11) discussed in Chapter 3.1.2.

Chapter 7

Bivariate circular long-memory time series models

In this chapter we present our theoretical results on bivariate circular long-range dependent time series models. These results are recently published in Beran, Steffens and Ghosh (2021a) [38], and are summarized in the book chapter Beran, Steffens and Ghosh (2022) [40].

Our focus is on a circular-circular regression model of the form (6.1) under the assumption of long-range dependence of both, the explanatory variables and the error process. We estimate the conditional mean direction using the circular NW-estimator defined in (6.24) and prove asymptotic normality and uniform convergence in probability of our estimator. We can show that the convergence rate of our estimator depends on the order of the kernel and on the strength of both long-memory parameters. Based on this convergence rate, we obtain a sufficient condition for the existence of an (in some sense) "*optimal*" sequence of bandwidths and derive simple confidence intervals for the mean direction. Furthermore, we propose a hypothesis test for testing for such an "*optimal*" sequence of bandwidths and derive an asymptotic formula for the rejection regions of our test. Finally, we illustrate our technical results by applying them to a data example of wind directions at three locations in the United States.

7.1 The phenomenon of smoothing dichotomy

In the context of kernel estimation, one question concerns the choice of an optimal bandwidth. However, as mentioned in Part I, Chapter 3.1, one important phenomenon in the context of long-memory regression models is the *smoothing dichotomy*. Commonly, "*optimal*" is meant in the sense of minimizing the asymptotic MISE. However, in the context of long-memory models, this procedure would possibly lead to a too small bandwidth, and consequently, the long-range

dependence is no longer reflected in the asymptotic behavior yielding the same convergence rate as in the iid case. In contrast to this, a large bandwidth entails that the asymptotic behavior of the process is determined by the strength of the long-memory parameter. Hence, in our special situation, we do not determine an explicit optimal sequence of bandwidths but formulate a sufficient condition to avoid this dichotomous effect and to ensure "*optimality*" in the sense of being independent of the size of the bandwidth.

In the case of bivariate long-memory processes, this dichotomy can as well be seen in the limiting distribution, which gives rise to the notion of *smoothing trichotomy* introduced by Mielniczuk and Wu (2004) [298] in the real-valued situation. For linear processes in general, the smoothing dichotomy is discussed in Ho (1996) [205]. As one could have expected, we observe the same dichotomous behavior in the context of circular bivariate long-memory processes, and, on behalf of this, we extend the results from Mielniczuk and Wu (2004) [298] to the circular situation. Therefore, we will shortly review their framework and their results.

Mielniczuk and Wu (2004) [298] consider a real-valued bivariate long-memory process of the form

$$Y_t = m(X_t) + G(Z_t, X_t), \tag{7.1}$$

where (X_t) and (Z_t) are weakly stationary linear long-memory processes with $E\left[G(Z_t, X_t) \mid X_t\right] = 0$, and unknown mean function m. m is estimated using the real-valued NW-estimator given in (3.8). Analysis of the convergence of the estimator $\hat{m}$ is based on the so-called *martingale/long-memory decomposition (M/L-decomposition)*, which decomposes the difference $\hat{m} - m_n$, where m_n is a sequence of kernel estimators, into a martingale difference term, a sum of conditional expectations, and a term related to the bias of the estimator. To avoid that the bias of the estimator is the dominating term, they introduce a technical condition on the bandwidth, which is similar to the condition we will introduce in our (circular) situation. In case of weakly dependent or iid residuals, the limiting behavior is determined by the martingale term. Under the assumption of long-range dependence, all three terms make a contribution depending on their order of convergence. However, Wu and Mielniczuk (2002) [420] show that if the bandwidth tends to 0 sufficiently fast, the martingale term dominates in the long-memory situation as well. This is not the case if the convergence rate is appropriately slow. More precisely, the convergence rate of the estimator is determined by the amount of smoothing and the larger one of the long-memory parameters of the two processes (X_t) and (Z_t). Since the error term in (7.1) depends on both processes, Mielniczuk and Wu (2004) [298] do a case-by-case analysis for both, establishing limit laws for the mean function and deriving asymptotic confidence intervals for the estimator $\hat{m}$. As we will see in Chapter 7.4, our restriction on the error process yields an explicit condition on the relation between the strength of dependence of the two linear processes

(X_t) and (Z_t) such that the asymptotic convergence rate of the estimator is independent of the bandwidth. This allows for constructing simultaneous confidence bands.

7.2 Framework and notations

Let $((\Theta_t),(\Psi_t))$ be a bivariate $([0,2\pi)\times[0,2\pi))$-valued process where (Ψ_t) is stationary. According to (6.1) a circular-circular regression model is of the form

$$\Theta_t = [\nu(\Psi_t) + Z_t] \, mod \, 2\pi, \qquad t \in \mathbb{Z}, \tag{7.2}$$

with a stationary zero-mean error process (Z_t) such that

$$E\left[\sin\left(Z_t\right)\right] = 0 \quad \text{and} \quad E\left[\cos\left(Z_t\right)\right] = R_Z > 0. \tag{7.3}$$

Equations (7.2) and (7.3) imply that the *conditional mean direction* of Θ_t given $\Psi_t = \psi^0$, $\psi^0 \in [0,2\pi)$, is $\nu(\psi^0)$, i.e.,

$$E\left[e^{i\Theta_t} \mid \Psi_t = \psi^0\right] = R(\Psi)e^{i\nu(\psi^0)}, \qquad t \in \mathbb{Z}.$$

For asymptotic considerations we need the following assumptions:

(B1) (Z_t) and (X_t) are two stationary long-range dependent Gaussian processes, independent of each other, with zero mean, and variances $0 < \sigma_Z^2 = \mathrm{var}(Z_t) < \infty$ and $\sigma_X^2 = \mathrm{var}(X_t) = 1$, respectively.

(B2) The circular process (Ψ_t) is defined by Gaussian subordination, i.e., (Ψ_t) is subordinated to the Gaussian process (X_t), which means (see (5.38)):

$$\Psi_t = G_\Psi^{-1}\left(\Phi(X_t)\right), \qquad t \in \mathbb{Z},$$

where Φ denotes the standard normal distribution function, and G_Ψ is an absolutely continuous circular distribution function with density $g_\Psi = G_\Psi'$.

(B3) $\nu : [0,2\pi) \to \mathbb{R}$ is a twice continuously differentiable function.

The subsequent three remarks summarize important properties which directly result from assumptions (B1) to (B3).

Remark 7.2.1. *From assumption (B1) we obtain the following properties:*

(i) From (1.10) and (1.11) we know that assumption (B1) means that the autocovariance functions γ_Z of (Z_t) and γ_X of (X_t) behave as

$$\gamma_Z(k) = cov\left(Z_t, Z_{t+k}\right) \underset{k\to\infty}{\sim} c_{\gamma,Z}|k|^{2d_Z-1},$$

and

$$\gamma_X(k) = cov\left(X_t, X_{t+k}\right) \underset{k\to\infty}{\sim} c_{\gamma,X}|k|^{2d_X-1},$$

respectively, and the spectral densities are of the form

$$f_Z(\lambda) \underset{\lambda\to 0}{\sim} c_{f,Z}|\lambda|^{-2d_Z}, \tag{7.4}$$

and

$$f_X(\lambda) \underset{\lambda\to 0}{\sim} c_{f,X}|\lambda|^{-2d_X},$$

respectively, with constants $c_{\gamma,Z}, c_{\gamma,X}, c_{f,Z}, c_{f,X} > 0$ and long-memory parameters $d_X, d_Z \in \left(0, \frac{1}{2}\right)$.

(ii) Since (X_t) and (Z_t) are stationary processes, (Θ_t) is a stationary process with constant mean direction ν.

(iii) Moreover, using the Wold representation (see Theorem 1.2.1), (Z_t) and (X_t) can be written as

$$Z_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}, \qquad \text{and} \qquad X_t = \sum_{j=0}^{\infty} c_j \eta_{t-j}, \qquad t \in \mathbb{Z},$$

respectively, where (ε_t) and (η_t) are iid zero mean normal random variables with finite variances σ_ε^2 and σ_η^2, respectively. The coefficients a_j and c_j, $j \in \mathbb{N}_0$, have to satisfy condition (1.17), i.e., $a_0 = c_0 = 1$, and

$$a_j \underset{j\to\infty}{\sim} C_a j^{d_Z-1}, \qquad \text{and} \qquad c_j \underset{j\to\infty}{\sim} C_c j^{d_X-1},$$

respectively, where $d_Z, d_X \in \left(0, \frac{1}{2}\right)$ and $C_a, C_c > 0$ are finite constants.

Combining the assumptions, we obtain two more properties:

Remark 7.2.2. *(i) Assumptions (B1) and (B2) imply, in addition to (7.3), that*

$$E\left[\sin\left(X_t\right)\right] = 0, \qquad \text{and} \qquad E\left[\cos\left(X_t\right)\right] = R_X > 0.$$

(ii) Assumptions (B2) and (B3) imply that

$$\Theta_t = [\nu(\Psi_t) + Z_t] \ mod\, 2\pi = [\mu(X_t) + Z_t] \ mod\, 2\pi, \tag{7.5}$$

where

$$\mu(x) = \nu\left(G_\Psi^{-1}\left(\Phi(x)\right)\right).$$

For $\psi^0 \in [0, 2\pi)$, the conditional mean direction of Θ_t given $\Psi_t = \psi^0$ then is defined through

$$\begin{aligned} R(\Psi_t)e^{i\nu(\psi^0)} &= E\left[e^{i\Theta_t} \mid \Psi_t = \psi^0\right] \\ &= E\left[e^{iZ_t}\right] \cdot e^{i\cdot\nu(\psi^0)}. \end{aligned}$$

Thus, since Z_t is a Gaussian process with variance σ_Z^2,

$$R(\Psi_t) = E\left[e^{iZ_t}\right] = e^{-\frac{1}{2}\sigma_Z^2} = R.$$

This subsequent remark concerns the dependence structure of the two circular processes. Due to Gaussian subordination and long-range dependence of the real-valued processes (X_t) and (Z_t), we derive long-range dependence of the circular processes (Θ_t) and (Ψ_t) as follows:

Remark 7.2.3. *(i) According to Lemma 5.3.1, assumptions (B1) and (B2) imply that long-range dependence in (X_t) is inherited by (Ψ_t).*

More precisely, let ν_Ψ be the constant mean direction of Ψ_t and define

$$H(x) := \sin\left(G_\Psi^{-1}(\Phi(x)) - \nu_\Psi\right).$$

Assume that this L^2-function H has an orthogonal Hermite polynomial expansion with Hermite rank m and Hermite coefficients a_m as given in (5.39). If $m \geq 1$ is such that $d_X > \frac{1}{2}\left(1 - m^{-1}\right)$, then the autocovariance function of (Ψ_t) asymptotically behaves as

$$\gamma_\Psi(k) \underset{k\to\infty}{\sim} c_{\gamma,m}|k|^{2d_m-1},$$

where $d_m := \frac{1}{2}\left(1 + m(2d_X - 1)\right) \in \left(0, \frac{1}{2}\right)$, and $c_{\gamma,m} := \frac{a_m^2}{m!}c_{\gamma,X}^m$ with $c_{\gamma,X}$ obtained from the autocovariance function of (X_t).

(ii) By definition of the mean direction ν_Ψ, see (5.18), together with assumption (B2), we have

(applying trigonometric identities)

$$E\left[H(X_t)\right] = E\left[\sin\left(G_\Psi^{-1}(\Phi(X_t)) - \nu_\Psi\right)\right] = E\left[\sin\left(\Psi_t - \nu_\Psi\right)\right] = 0.$$

(iii) Assumption (B1) together with (7.5) imply that long-range dependence in (Z_t) or in (X_t) leads to long-range dependence in (Θ_t).

Concerning the estimation procedure, the mean direction ν is estimated using circular kernel regression. Let therefore K_b be a nonnegative, symmetric and continuous circular kernel for which the conditions (CK1) to (CK4) on the kernel given in Chapter 6.2 are satisfied.

Analogously as in the real-valued situation, for asymptotic considerations, we need the following *consistency condition* on the sequence of circular kernel bandwidths (b_n):

(BW) (b_n) is a sequence of circular bandwidths such that

$$b_n \searrow 0 \quad \text{and} \quad nb_n \to \infty \quad \text{as } n \to \infty,$$

where, in this context, $b_n \searrow 0$ means, that $b_n > 0$ and $b_n \to 0$ as $n \to \infty$.

Based on random samples $\Theta_1, \ldots, \Theta_n \in [0, 2\pi)$ and $\Psi_1, \ldots, \Psi_n \in [0, 2\pi)$, given $\psi^0 \in [0, 2\pi)$ and any $n \in \mathbb{N}$, the *circular NW-estimator* for the conditional mean direction defined in (6.24) is of the form

$$\hat{\nu}_n(\psi^0) = \frac{\frac{1}{n}\sum_{t=1}^{n} K_b(\Psi_t - \psi^0)\Theta_t}{\hat{f}_n(\psi^0)}, \tag{7.6}$$

where

$$\hat{f}_n(\psi^0) = \frac{1}{n}\sum_{t=1}^{n} K_b(\Psi_t - \psi^0) \tag{7.7}$$

is the kernel density estimator of the marginal density f_Ψ at ψ^0 as defined in (6.17).

We use the following notations for fixed $\psi^0 \in [0, 2\pi)$:

$$\nu_n(\psi^0) \;:=\; \frac{E\left[K_b(\Psi_1 - \psi^0)\Theta_1\right]}{E\left[K_b(\Psi_1 - \psi^0)\right]}, \tag{7.8}$$

which, due to stationarity, is independent of t,

$$J_t(\psi^0) \;:=\; Z_t \cdot K_b(\Psi_t - \psi^0), \tag{7.9}$$

and

$$D_n(\psi^0) := \hat{\nu}_n(\psi^0) - \nu_n(\psi^0). \tag{7.10}$$

Furthermore, let f_η, f_ε, f_X and f_Z denote the marginal densities of (η_t), (ε_t), (X_t) and (Z_t), respectively.

7.3 Auxiliary results

In order to derive asymptotic convergence of our kernel estimator, we decompose the difference $\hat{\nu}_n(\psi^0) - \nu(\psi^0)$ into a deterministic part, the bias part, and a stochastic part. For the asymptotic behavior of the deterministic part, we use the same arguments as in the real-valued case and hence, only have to study the stochastic part. Before we state our first results on this decomposition, we recall the definitions of $\mathcal{O}_p$- and o_p-notation:

Remark. *Let (f_n) and (g_n) be random sequences. Then $f_n = o_p(g_n)$ means that $\frac{f_n}{g_n} \xrightarrow{P} 0$, and $f_n = \mathcal{O}_p(g_n)$ means that for any $\epsilon > 0$ there exists $C, n_0 > 0$ such that $P\left(\left|\frac{f_n}{g_n}\right| \geq C\right) \leq \epsilon$ for all $n \geq n_0$. Note that $\xrightarrow{P}$ denotes convergence in probability.*
Clearly, if (f_n) and (g_n) are deterministic sequences, one writes $\mathcal{O}(g_n)$ and $o(g_n)$, respectively, meaning the usual convergence, with the convention that $\frac{0}{0} := 0$.

Proposition 7.3.1. *Under the assumptions (B1) to (B3) on the model, the assumptions (CK1) to (CK4) on the kernel, and the consistency condition (BW) on the sequence of bandwidths, the asymptotic behavior of $\hat{\nu}_n(\psi^0) - \nu(\psi^0)$ is determined by the asymptotic behavior of D_n.*

Proof. The difference can be decomposed into:

$$\begin{aligned} \hat{\nu}_n(\psi^0) - \nu(\psi^0) &= \hat{\nu}_n(\psi^0) - \nu_n(\psi^0) + \nu_n(\psi^0) - \nu(\psi^0) \\ &=: D_n(\psi^0) + \nu_n(\psi^0) - \nu(\psi^0). \end{aligned} \tag{7.11}$$

Using the same arguments as in the real-valued case (Taylor's expansion) we know that the deterministic term $\nu_n(\psi^0) - \nu(\psi^0)$ is of order $\mathcal{O}(b_n^2)$, and hence, is independent of the dependence structure (see, e.g., Taylor, 2008 [387]). Thus, in order to investigate the asymptotic behavior of $\hat{\nu}_n(\psi^0) - \nu(\psi^0)$ in (7.11), it is sufficient to study the asymptotic behavior of $D_n(\psi^0)$. ■

For this, we use a random-design version of the M-L-decomposition, which is, e.g., used in Beran and Schumm (2017) [36] in the real-valued fixed-design setting, and in Mielniczuk and Wu (2002) [420] and Wu and Mielniczuk (2004) [298] in the real-valued random-design setting.

Lemma 7.3.2. *Suppose that the assumptions of Proposition 7.3.1 hold. Then $D_n \cdot \hat{f}_n(\psi^0)$ can be decomposed into*

$$D_n \cdot \hat{f}_n(\psi^0) = A_n(\psi^0) + B_n(\psi^0) + C_n(\psi^0) + r_n,$$

where

$$\begin{aligned} A_n(\psi^0) &= \frac{1}{n}\sum_{t=1}^{n}\Big(J_t(\psi^0) - E\Big[J_t(\psi^0) \mid (\varepsilon_s,\eta_s), s \le t-1\Big]\Big), \\ B_n(\psi^0) &= \frac{1}{n}\sum_{t=1}^{n} E\Big[J_t(\psi^0) \mid (\varepsilon_s,\eta_s), s \le t-1\Big], \\ C_n(\psi^0) &= \frac{1}{n}\sum_{t=1}^{n} K_b(\Psi_t - \psi^0)\cdot(\nu(\Psi_t) - \nu_n(\psi^0)), \end{aligned}$$

and

$$A_n(\psi^0) = \mathcal{O}_p\left((nb_n)^{-\frac{1}{2}}\right), \qquad B_n(\psi^0) = \mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right), \qquad C_n(\psi^0) = \mathcal{O}_p\left(b_n^2 n^{d_X-\frac{1}{2}}\right),$$

and $r_n = o_p\left(\max\left\{A_n(\psi^0), B_n(\psi^0), C_n(\psi^0)\right\}\right)$.

For proving this, we use the fact that $A_n(\psi^0)$ admits a martingale structure, $B_n(\psi^0)$ is a sum of conditional expectations, and $C_n(\psi^0)$ is a kernel weighted bias term.

Proof. The decomposition is obtained by writing out the definitions of $\hat{\nu}_n$, $\hat{f}_n$, ν_n and J_t given in (7.6) - (7.10):

$$\begin{aligned} D_n\cdot\hat{f}_n(\psi^0) &= \Big(\hat{\nu}_n(\psi^0) - \nu_n(\psi^0)\Big)\cdot\hat{f}_n(\psi^0) \qquad (7.12) \\ &= \frac{n^{-1}\sum_{t=1}^{n} K_b(\Psi_t-\psi^0)\Theta_t}{\hat{f}_n(\psi^0)}\cdot\hat{f}_n(\psi^0) - \nu_n(\psi^0)\cdot n^{-1}\sum_{t=1}^{n} K_b(\Psi_t-\psi^0) \\ &= n^{-1}\sum_{t=1}^{n} K_b(\Psi_t-\psi^0)\Big(\Theta_t - \nu_n(\psi^0)\Big) \\ &= n^{-1}\sum_{t=1}^{n}\left[K_b(\Psi_t-\psi^0)\Big(\nu(\psi_t) - \nu_n(\psi^0)\Big) + \underbrace{K_b\Big(\Psi_t-\psi^0\Big)Z_t}_{=J_t(\psi^0)}\right] \\ &= n^{-1}\sum_{t=1}^{n}\Big(J_t(\psi^0) - E\Big[J_t(\psi^0)\mid(\varepsilon_s,\eta_s), s\le t-1\Big]\Big) \\ &\quad + n^{-1}\sum_{t=1}^{n} E\Big[J_t(\psi^0)\mid(\varepsilon_s,\eta_s), s\le t-1\Big] \\ &\quad + n^{-1}\sum_{t=1}^{n} K_b(\Psi_t-\psi^0)\cdot(\nu(\Psi_t)-\nu_n(\psi^0)) \\ &=: A_n(\psi^0) + B_n(\psi^0) + C_n(\psi^0) + r_n, \qquad (7.13) \end{aligned}$$

where $r_n = o_p\left(\max\left\{A_n(\psi^0), B_n(\psi^0), C_n(\psi^0)\right\}\right)$.

The results on the asymptotic behavior of the three parts now follow along the same lines as in Mielniczuk and Wu (2004) [298], Lemma 3, Lemma 4 and Lemma 6, by adapting their results to our situation.

As $A_n(\psi^0)$ forms a martingale difference sequence, $(nb_n)^{-\frac{1}{2}} A_n(\psi^0)$ is asymptotically normally distributed and thus, $A_n(\psi^0)$ is of order $\mathcal{O}_p\left((nb_n)^{-\frac{1}{2}}\right)$. The proof for term $B_n(\psi^0)$ is based on the projection method developed in Wu (2003) [419] and Wu and Woodroofe (2002) [421]. Analysis of the third term $C_n(\psi^0)$ is based on the results in Wu and Mielniczuk (2002) [420]. ■

7.4 Uniform consistency of the kernel estimator

Using the auxiliary results of the previous chapter, we can now prove our asymptotic results in order to derive uniform convergence in probability of our kernel estimator. In the real-valued setting, this is shown, e.g., by Silverman (1978) [367] and Bierens (1983) [50]. The convergence rate is later used to address the phenomenon of smoothing dichotomy in the context of circular random-design long-memory models. Based on the asymptotic behavior of the estimator, we obtain a condition on the long-memory parameter d_Z which avoids that the bandwidth has any impact on the convergence rate.

Theorem 7.4.1 combines the individual results in Lemma 7.3.2 with the behavior of the deterministic term in (7.11):

Theorem 7.4.1. *Suppose that assumptions (B1) to (B3) on the model and assumptions (CK1) to (CK4) on the kernel hold, that $f_\psi(\omega) > 0$ for $\omega \in [0, 2\pi)$, and that (b_n) is a sequence of bandwidths satisfying the consistency condition (BW). Then the difference term (7.11) exhibits the following asymptotic behavior*

$$\hat{\nu}_n(\psi^0) - \nu(\psi^0) \quad = \quad \mathcal{O}(b_n^2) + \mathcal{O}_p\left((nb_n)^{-\frac{1}{2}}\right) + \mathcal{O}_p\left(n^{d_Z - \frac{1}{2}}\right) + \mathcal{O}_p\left(b_n^2 n^{d_X - \frac{1}{2}}\right). \tag{7.14}$$

Proof. From Lemma 7.3.2 we have

$$D_n \cdot \hat{f}_n(\psi^0) = \mathcal{O}_p\left((nb_n)^{-\frac{1}{2}}\right) + \mathcal{O}_p\left(n^{d_Z - \frac{1}{2}}\right) + \mathcal{O}_p\left(b_n^2 n^{d_X - \frac{1}{2}}\right).$$

Thus, recalling that the deterministic term leads to

$$\nu_n(\psi^0) - \nu(\psi^0) = \mathcal{O}\left(b_n^2\right),$$

we obtain the result.

■

Note, that $\mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right)$ in equation (7.14) is the only term that is independent of the sequence of bandwidths (b_n). Hence, if this term dominates the other terms, the convergence would be independent of the choice of the bandwidth and as a consequence, we would achieve "*optimal convergence*". We use this fact to define a notion of "*optimality*" for a sequence of bandwidths, where "*optimal*" is meant in the sense of *having no impact on the convergence rate*:

Definition 7.4.2. *In the situation of Theorem 7.4.1, a sequence of bandwidths (b_n) is called* ***optimal*** *if $\mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right)$ in equation (7.14) is the dominant part.*

Remark 7.4.3. *This condition amounts to*

$$\max\left\{b_n^2,\ (nb_n)^{-\frac{1}{2}},\ b_n^2 n^{d_X-\frac{1}{2}}\right\} = o\left(n^{d_Z-\frac{1}{2}}\right), \tag{7.15}$$

since, if the term $\mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right)$ has to dominate the other three terms, the following conditions must be satisfied:

$$\begin{aligned} b_n^2 &= o\left(n^{d_Z-\frac{1}{2}}\right), \\ (nb_n)^{-\frac{1}{2}} &= o\left(n^{d_Z-\frac{1}{2}}\right), \\ b_n^2 n^{d_X-\frac{1}{2}} &= o\left(n^{d_Z-\frac{1}{2}}\right), \end{aligned}$$

and we obtain the requirement

$$\max\left\{b_n^2,\ (nb_n)^{-\frac{1}{2}},\ b_n^2 n^{d_X-\frac{1}{2}}\right\} = o_p\left(n^{d_Z-\frac{1}{2}}\right)\left(= o\left(n^{d_Z-\frac{1}{2}}\right)\right).$$

We obtain the following characterizations:

Theorem 7.4.4. *Suppose that the assumptions of Theorem 7.4.1 hold. Then the following statements are equivalent:*

(i) The set of bandwidth sequences (b_n) such that the consistency condition (BW) and (7.15) holds, is not empty.

(ii) The set of bandwidth sequences (b_n) that satisfies the consistency condition (BW) and the following three conditions

$$\lim_{n\to\infty} n^{\frac{1-2d_Z}{4}} b_n = 0, \quad \lim_{n\to\infty} n^{\frac{d_X-d_Z}{2}} b_n = 0, \quad \lim_{n\to\infty} n^{2d_Z} b_n = \infty$$

is not empty.

(iii) The following relation holds for the two memory parameters d_X and d_Z:

$$d_Z > \max\left\{\frac{1}{5}d_X,\ \frac{1}{10}\right\}. \tag{7.16}$$

Proof. To proof this, we need equivalent transformations using the definition of o-notation given in Remark 7.3. From Definition 7.4.2, we obtain the condition

$$\max\left\{b_n^2,\ (nb_n)^{-\frac{1}{2}},\ b_n^2 n^{d_X-\frac{1}{2}}\right\} = o\left(n^{d_Z-\frac{1}{2}}\right),$$

which is (using o-calculations) equivalent to the following conditions on the sequence (b_n):

$$b_n = o\left(\min\left\{n^{\frac{2d_Z-1}{4}},\ n^{\frac{d_Z-d_X}{2}}\right\}\right) \tag{7.17}$$

and

$$n^{-2d_Z} = o(b_n). \tag{7.18}$$

Rewriting these two conditions using the definition of the o-notation yields (ii).

Finally, assuming that (7.17) and (7.18) hold simultaneously, we obtain

$$n^{-2d_Z} = o\left(\min\left\{n^{\frac{2d_Z-1}{4}},\ n^{\frac{d_Z-d_X}{2}}\right\}\right),$$

which, in terms of the exponents, means

$$d_Z > \max\left\{\frac{1}{5}d_X,\ \frac{1}{10}\right\}.$$

■

Analogously as in the real-valued case, one can consider kernels of some even order greater or equal 2. Since higher order kernels affect particularly the bias of an estimator, this has an impact on the asymptotic behavior of the difference (7.11), and hence condition (7.16) slightly changes.

Theorem 7.4.5. *Suppose that the assumptions of Theorem 7.4.1 hold but let K_{b_n} be a circular kernel of order $2k$, $k \geq 1$. Then, the statements in Theorem 7.4.1 and Theorem 7.4.4 read as follows:*

(i) The deterministic term $\hat{\nu}_n(\psi^0) - \nu_n(\psi^0)$ can be generalized to $\mathcal{O}(b_n^{2k})$, $k \geq 1$, which results in an asymptotic behavior of the form

$$\hat{\nu}_n(\psi^0) - \nu(\psi^0) \quad = \quad \mathcal{O}\left(b_n^{2k}\right) + \mathcal{O}_p\left((nb_n)^{-\frac{1}{2}}\right) + \mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right) + \mathcal{O}_p\left(b_n^2 n^{d_X-\frac{1}{2}}\right). \tag{7.19}$$

(ii) The respective "optimality condition", analogously to condition (7.15), reads

$$\max\left\{b_n^{2k},\ (nb_n)^{-\frac{1}{2}},\ b_n^2 n^{d_X-\frac{1}{2}}\right\} = o\left(n^{d_Z-\frac{1}{2}}\right).$$

(iii) Condition (7.16) in Lemma 7.4.4 becomes

$$d_Z > \max\left\{\frac{1}{5}d_X,\ \frac{1}{2(4k+1)}\right\}. \tag{7.20}$$

Proof. The proof of (i) follows with the same arguments as for real-valued kernels (see Remark 3.1.3), and, based on this, we need to rewrite Definition 7.4.2 in an analogous manner as shown in Remark 7.4.3.

We prove (iii) along the same lines as in Lemma 7.4.4. Here, the following conditions have to be satisfied:

$$\begin{aligned} b_n^{2k} &= o\left(n^{d_Z-\frac{1}{2}}\right), \\ (nb_n)^{-\frac{1}{2}} &= o\left(n^{d_Z-\frac{1}{2}}\right), \\ b_n^2 n^{d_X-\frac{1}{2}} &= o\left(n^{d_Z-\frac{1}{2}}\right), \end{aligned}$$

and from this we obtain

$$n^{-2d_Z} = o\left(\min\left\{n^{\frac{2d_Z-1}{4k}},\ n^{\frac{d_Z-d_X}{2}}\right\}\right), \tag{7.21}$$

which (in terms of the exponents) means

$$d_Z > \max\left\{\frac{1}{5}d_X,\ \frac{1}{2(4k+1)}\right\}.$$

■

Obviously, condition (iii) in Theorem 7.4.5 can be simplified in certain situations (without proof):

Corollary 7.4.6. *Suppose that the assumptions of Theorem 7.4.5 hold.*

(i) For a circular kernel of order 2, i.e., for $k = 1$, condition (iii) in Theorem 7.4.5 coincides with condition (iii) in Theorem 7.4.4, and since $d_X < \frac{1}{2}$ by definition, can be further reduced to

$$d_Z > \frac{1}{10}.$$

Hence, this condition is independent of the dependence structure of the process (X_t).

(ii) For $k \to \infty$, we have

$$\frac{1}{2(4k+1)} \searrow 0 \tag{7.22}$$

and condition (7.20) reduces to

$$d_Z > \max\left\{\frac{1}{5}d_X,\ 0\right\}.$$

If further $d_X \leq 0$ *(in case of weak dependent or antipersistent* (X_t)*), condition (7.20) becomes*

$$d_Z > 0,$$

which always holds for a long-range dependent process (Z_t)*.*

(iii) If, in addition to (7.22), $d_X > 0$ *(as it is in case of a long-memory process* (X_t)*), condition (7.20) becomes*

$$d_Z > \frac{1}{5}d_X.$$

In order to obtain the overall convergence rate of the estimator, we first need to ensure that there exists such an optimal sequence of bandwidths:

Corollary 7.4.7. *Suppose that the assumptions of Theorem 7.4.1 and condition (7.16) hold. Then the set of optimal sequences is not empty.*

Moreover, for any such sequence (b_n)*, we have for the estimator* $\hat{\nu}$ *based on the associated kernel* K_{b_n} *the following convergence:*

$$\hat{\nu}_n(\psi^0) - \nu(\psi^0) = \mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right). \tag{7.23}$$

Proof. From Theorem 7.4.1 and Lemma 7.3.2 we know that for the existence of an optimal sequence of bandwidths, the term $\mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right)$ in (7.14) must be the dominating term, which means

$$\hat{\nu}_n(\psi^0) - \nu(\psi^0) = \mathcal{O}_p\left(n^{d_Z-\frac{1}{2}}\right).$$

■

Analysis of (7.23) with a suitable normalizing constant leads to convergence in distribution, which we need in order to derive asymptotic confidence intervals for the mean direction ν. Instead of working with $B_n(\psi^0)$, we define a new random variable $\tilde{B}_n(\psi^0)$, which is asymptotically equivalent to $B_n(\psi^0)$ and can be handled more easily. For this, we need a useful asymptotic expression for $B_n(\psi^0)$:

Lemma 7.4.8. *Suppose that the assumptions of Corollary 7.4.7 hold. Then* $B_n(\psi^0)$ *can be expressed as*

$$B_n(\psi^0) = n^{-1}\sum_{s=1}^{n} Z_s \cdot E\left[K_b(\Psi_s - \psi^0) \mid \eta_t,\ t \leq s-1\right] + o_p(B_n).$$

Proof. Note that independence of (Z_s) and (ψ_t) leads to

$$\begin{aligned} B_n(\psi^0) &= n^{-1}\sum_{s=1}^{n} E\left[J_s(\psi^0) \mid (\varepsilon_t,\eta_t), t \le s-1\right] \\ &= n^{-1}\sum_{s=1}^{n} E\left[Z_s \cdot K_b(\Psi_s - \psi^0) \mid (\varepsilon_t,\eta_t), t \le s-1\right] \\ &= n^{-1}\sum_{s=1}^{n} \Big(E\left[Z_s \mid \varepsilon_t, t \le s-1\right] E\left[K_b(\Psi_s - \psi^0) \mid \eta_t, t \le s-1\right]\Big). \end{aligned}$$

Thus, since

$$\begin{aligned} E\left[Z_s \mid \varepsilon_t, t \le s-1\right] &= \sum_{j=1}^{\infty} a_j \varepsilon_{s-j} \\ &= Z_s - \varepsilon_s, \end{aligned} \tag{7.24}$$

which is the one-step forecast of (Z_s) given (ε_t), $t \le s-1$, we obtain

$$\begin{aligned} B_n(\psi^0) &= n^{-1}\sum_{s=1}^{n} (Z_s - \varepsilon_s) \cdot E\left[K_b(\Psi_s - \psi^0) \mid \eta_t, t \le s-1\right] \\ &= n^{-1}\sum_{s=1}^{n} Z_s \cdot E\left[K_b(\Psi_s - \psi^0) \mid \eta_t, t \le s-1\right] + o_p(B_n). \end{aligned}$$

■

Proposition 7.4.9. *Suppose that the assumptions of Corollary 7.4.7 hold and let*

$$\tilde{B}_n(\psi^0) := \frac{1}{n}\sum_{t=1}^{n} E\left[Z_t \mid \varepsilon_s, s \le t-1\right] \cdot \Big(K_b * f_\Psi(\psi^0)\Big), \tag{7.25}$$

where f_Ψ is the marginal density of (Ψ_t) and $*$ denotes convolution. Then the second term in (7.12), $B_n(\psi^0)$, may asymptotically be replaced by $\tilde{B}_n$ with

$$E\left[B_n - \tilde{B}_n\right] = 0$$

and

$$var\left(B_n - \tilde{B}_n\right) = o\left(n^{2d_Z - 1}\right).$$

Proof. Note that

$$\begin{aligned} K_b * f_\Psi(\psi^0) &= \int K_b(y - \psi^0) f_\Psi(y) dy \\ &= E[K_b(\Psi_s - \psi^0)], \end{aligned}$$

where f_Ψ is the marginal density of (Ψ_t), and, with the same argument as in Lemma 7.4.8, the conditional expectation reads

$$E\left[Z_s \mid \varepsilon_t, t \leq s-1\right] = Z_s - \varepsilon_s.$$

Hence (7.25) reduces to

$$\tilde{B}_n(\psi^0) = n^{-1}\sum_{t=1}^{n} Z_s \cdot E\left[K_b(\Psi_s - \psi^0)\right] + o\left(\tilde{B}_n\right).$$

Defining

$$\tilde{K}_s(\psi^0) := E\left[K_b(\Psi_s - \psi^0) \mid \eta_t,\ t \leq s-1\right],$$

we obtain

$$\begin{aligned} E\left[\tilde{K}_s(\psi^0)\right] &= E\left[E\left[K_b(\Psi_s - \psi^0) \mid \eta_t,\ \eta_t, t \leq s-1\right]\right] \\ &= E[K_b(\Psi_s - \psi^0)]. \end{aligned}$$

Consequently,

$$\begin{aligned} B_n(\psi^0) - \tilde{B}_n(\psi^0) &= n^{-1}\sum_{s=1}^{n}\left\{E\left[Z_s \mid \varepsilon_t, t \leq s-1\right]\left(E\left[K_b(\Psi_s - \psi^0) \mid \eta_t, t \leq s-1\right] - E\left[K_b(\Psi_s - \psi^0)\right]\right)\right\} \\ &\quad + o_p\left(\max\left\{B_n, \tilde{B}_n\right\}\right) \\ &=: n^{-1}\sum_{s=1}^{n}\left(Z_s \cdot \left(\tilde{K}_s(\psi^0) - E\left[\tilde{K}_s(\psi^0)\right]\right)\right) + o\left(\max\left\{B_n, \tilde{B}_n\right\}\right) \end{aligned}$$

and, since (Z_t) and (Ψ_t) are independent, the expected value becomes

$$E\left[B_n - \tilde{B}_n\right] = 0.$$

Again, due to independence of (Z_t) and (Ψ_t), the variance is of the form

$$\begin{aligned} \mathrm{var}(\tilde{B}_n - B_n) &= n^{-2}\mathrm{var}\left(\sum_{s=1}^{n}\left(Z_s \cdot \left(\tilde{K}_s(\psi^0) - E\left[\tilde{K}_s(\psi^0)\right]\right)\right)\right) \\ &= n^{-2}\sum_{s,r=1}^{n}\mathrm{cov}\left(Z_s \cdot \left(\tilde{K}_s(\psi^0) - E\left[\tilde{K}_s(\psi^0)\right]\right), Z_r \cdot \left(\tilde{K}_r(\psi^0) - E\left[\tilde{K}_r(\psi^0)\right]\right)\right) \\ &= n^{-2}\sum_{s,r=1}^{n}\mathrm{cov}(Z_s, Z_r)\mathrm{cov}\left(\tilde{K}_s(\psi^0), \tilde{K}_r(\psi^0)\right). \end{aligned}$$

Due to Lipschitz continuity of the kernel K_b, we obtain

$$\operatorname{cov}\Big(\tilde{K}_s(\psi^0), \tilde{K}_r(\psi^0)\Big) \leq c \cdot \operatorname{cov}\Big(E\left[\Psi_s - \psi^0 \mid \eta_t,\ t \leq s-1\right], E\left[\Psi_r - \psi^0 \mid \eta_t,\ t \leq r-1\right]\Big),$$

where c is a suitable constant. Let $\gamma_{\tilde{K}}$ denote the autocovariance function of $\tilde{K}_s(\psi^0)$, then, according to Lemma 6.1.1 and since $m = 1$, the autocovariance function is asymptotically of the form:

$$\gamma_{\tilde{K}}(k) \leq c \cdot \gamma_\Psi(k) \underset{k\to\infty}{\sim} c_\gamma k^{2d_X-1},$$

with some positive constant c_γ.

Using the Riemann sum and the asymptotic behavior of the autocovariances of (Z_t) and (Ψ_t), the variance asymptotically reads

$$\begin{aligned}
\operatorname{var}(\tilde{B}_n - B_n) &\leq n^{-2} \cdot c \cdot \sum_{s,r=1}^{n} \gamma_Z(s-r)\gamma_\Psi(s-r) \\
&= n^{-2} \cdot c \cdot \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \gamma_Z(k)\gamma_\Psi(k) \\
&\underset{k\to\infty}{\sim} c_1 \cdot n^{2(d_Z+d_X-1)} \cdot \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \left(\frac{|k|}{n}\right)^{2(d_Z+d_X-1)} \cdot \frac{1}{n} \\
&\underset{n\to\infty}{\sim} c_2 \cdot n^{2(d_Z+d_X-1)},
\end{aligned}$$

where c_1, c_2 are suitable positive constants.

Finally, since by definition $d_X < \frac{1}{2}$, we obtain

$$n^{2(d_Z+d_X-1)} = o\left(n^{2d_Z-1}\right).$$

■

Based on this, we can show weak convergence of our estimator.

Theorem 7.4.10. *Under the assumptions of Corollary 7.4.7, we obtain*

$$n^{\frac{1}{2}-d_Z}\left[\hat{\nu}_n(\psi^0) - \nu(\psi^0)\right] \xrightarrow{d} \sqrt{c_{f,Z} \cdot v(d_Z)} \cdot W =: c_\nu(\psi^0) \cdot W, \tag{7.26}$$

where

$$v(d_Z) = \frac{2\sin(\pi d_Z)}{d_Z(2d_Z+1),}\Gamma(1-2d_Z), \tag{7.27}$$

is a constant $c_{f,Z}$ corresponding to the spectral density of (Z_t) given in (7.4), and $W \sim N(0,1)$.

More generally, for $\psi_1^0 < \psi_2^0 < \ldots < \psi_k^0$, $\psi_j^0 \in [0, 2\pi)$, *we obtain*

$$n^{\frac{1}{2}-d_Z}\left[\hat{\nu}_n(\psi_1^0) - \nu(\psi_1^0), \ldots, \hat{\nu}_n(\psi_k^0) - \nu(\psi_k^0)\right] \xrightarrow{d} \left[c_\nu(\psi_1^0), \ldots, c_\nu(\psi_k^0)\right] \cdot W. \tag{7.28}$$

Proof. From Lemma 7.4.9 we have

$$B_n(\psi^0) = \tilde{B}_n(\psi^0) + o\left(n^{d_X+d_Z-1}\right)$$

and, since $d_X < \frac{1}{2}$ by definition, this leads to

$$B_n(\psi^0) = \tilde{B}_n(\psi^0) + o\left(n^{d_Z-\frac{1}{2}}\right).$$

Note that $\hat{f}_n$ is a consistent estimator of the marginal density f_ψ. Hence, with the fact that $K_b * f_\psi(\psi^0)$ converges to $f_\psi(\psi^0)$ as $n \to \infty$ and given the decompositions (7.12) and (7.24), we further obtain

$$\begin{aligned} n^{\frac{1}{2}-d_Z}\left[\hat{\nu}_n(\psi^0) - \nu(\psi^0)\right] &= n^{\frac{1}{2}-d_Z}\frac{\tilde{B}_n(\psi^0)}{f_\psi(\psi^0)} + o_p\left(n^{d_Z-\frac{1}{2}}\right) \\ &= n^{\frac{1}{2}-d_Z}\frac{1}{n}\sum_{t=1}^{n} E\left[Z_t \mid \varepsilon_s, s \le t-1\right] + o_p\left(n^{d_Z-\frac{1}{2}}\right) \\ &= n^{\frac{1}{2}-d_Z}\frac{1}{n}\sum_{t=1}^{n}(Z_t - \varepsilon_t) + o_p\left(n^{d_Z-\frac{1}{2}}\right). \end{aligned}$$

Finally, applying the limit theorems for sample means under long-range dependence given in (2.10) yields the following asymptotic behavior:

$$n^{\frac{1}{2}-d_Z}\left[\hat{\nu}_n(\psi^0) - \nu(\psi^0)\right] \xrightarrow{d} c_\nu(\psi^0) \cdot W,$$

where $W \sim N(0,1)$ and c_ν is as defined in (7.26).

The convergence (7.28) then follows from the multivariate version of weak convergence of sample means under the assumption of long-range dependence given in (2.10). ■

This means that weak convergence of our circular kernel estimator can be reduced to weak convergence of sample means in the long-range dependent real-valued setting. In other words, this theorem states that under long-range dependence, the circular NW-estimator converges weakly to a degenerate standard normal random variable, which is essentially the same as for the real-valued NW-estimator.

Remark. *Equation (7.28), in particular, means that the standardized random deviations of estimates at different values* ψ_j^0, ψ_l^0 $(j \neq l)$ *are asymptotically perfectly correlated, which, on the*

one hand, is the same as in the real-valued case and, on the other hand, is in contrast to the usual behavior in nonparametric smoothing (see Part I, Chapters 2 and 3).

Recall from (6.18) and (6.19) that under conditions (CK1) to (CK4) any circular kernel estimator K_b has a Fourier series representation of the form

$$K_b(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \alpha_s(b) e^{-isu}, \quad u \in [0, 2\pi), \tag{7.29}$$

with Fourier coefficients

$$\alpha_{-s} = \alpha_s = \int_0^{2\pi} K_b(u) e^{isu} du.$$

For inference for the kernel estimators $\hat{\nu}_n$ we require uniform constistency. Based on the Fourier series representation, we derive a simple condition for the sum of the Fourier coefficients that implies uniform convergence in probability of $\hat{\nu}_n$. This condition is related to Lemma 3.2 and Corollary 3.2 in Ghosh (2014) [152].

Theorem 7.4.11. *Suppose that the assumptions of Corollary 7.4.7 hold and that*

$$\sum_{s=-\infty}^{\infty} |\alpha_s(b_n)| = o\left(n^{1-d_Z-d_X}\right), \tag{7.30}$$

where $\alpha_s(b_n)$ are the coefficients in the Fourier series representation of the kernel estimator K_b given in (7.29).

Then

$$\sup_{\psi^0 \in [0,2\pi)} |\hat{\nu}_n(\psi^0) - \nu(\psi^0)| \xrightarrow{P} 0.$$

Proof. For the proof we follow Parzen (1962) [312], Ghosh (2014) [152] and Ghosh (2018) [153]. Recall that uniform convergence in probability means that for all $\epsilon > 0$:

$$\lim_{n\to\infty} P\left(\sup_{\psi^0 \in [0,2\pi)} |\hat{\nu}_n(\psi^0) - \nu(\psi^0)| > \varepsilon\right) = 0.$$

Using Markov's inequality we obtain

$$P\left(\sup_{\psi^0 \in [0,2\pi)} |\hat{\nu}_n(\psi^0) - \nu(\psi^0)| > \varepsilon\right) \leq \frac{E\left[\sup_{\psi^0 \in [0,2\pi)} |\hat{\nu}_n(\psi^0) - \nu(\psi^0)|\right]}{\varepsilon}.$$

Further, with

$$E\left[\sup_{\psi^0\in[0,2\pi)}|\hat{\nu}_n(\psi^0)-\nu(\psi^0)|\right]\leq E\left[\sup_{\psi^0\in[0,2\pi)}|\hat{\nu}_n(\psi^0)-E\left[\hat{\nu}_n(\psi^0)\right]|\right]+\sup_{\psi^0\in[0,2\pi)}|E\left[\hat{\nu}_n(\psi^0)\right]-\nu(\psi^0)|$$

and the fact that the latter converges to 0 with the same arguments used for the asymptotic representation of the bias, a sufficient condition for uniform convergence in probability is

$$E\left[\sup_{\psi^0\in[0,2\pi)}\left|\hat{\nu}_n(\psi^0)-E\left[\hat{\nu}_n(\psi^0)\right]\right|\right]\longrightarrow 0\quad\text{as } n\to\infty.$$

Along the same lines as in the proof of Lemma 3.2 in Ghosh (2014) [152], we define *kernel smoothed errors* by

$$S_n(\psi^0)=\frac{1}{n}\sum_{t=1}^{n}E\left[K_b(\Psi_t-\psi^0)\mid\eta_r,\ r\leq t-1\right]\cdot Z_t.$$

Under the given conditions we know that

$$\hat{\nu}_n(\psi_0)-\nu(\psi^0)=\mathcal{O}\left(B_n(\psi^0)\right)=\mathcal{O}\left(n^{d_Z-\frac{1}{2}}\right).$$

Hence, applying Lemma 7.4.8, it suffices to prove that $S_n(\psi^0)-E\left[S_n(\psi^0)\right]$ converges uniformly to zero in probability. And further, since $E\left[S_n(\psi^0)\right]=0$, we only need to show that

$$\lim_{n\to\infty}E\left[\sup_{\psi^0\in[0,2\pi)}|S_n(\psi^0)|\right]=0.$$

For this, we use the Fourier series representation of the kernel (7.29) (see, e.g., Parzen, 1962 [312]).

Since by assumption (A2), (Z_t) is independent of (η_s), we can write

$$\begin{aligned}
S_n(\psi^0) &= \frac{1}{n}\sum_{t=1}^{n}E\left[K_b(\Psi_t-\psi^0)\mid\eta_r,\ r\leq t-1\right]Z_t\\
&= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\alpha_s(b_n)\left(\frac{1}{n}\sum_{t=1}^{n}E\left[\exp\left(-is(\Psi_t-\psi^0)\right)\mid\eta_r,r\leq t-1\right]\cdot Z_t\right)\\
&= \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\alpha_s(b_n)\left(\frac{1}{n}\sum_{t=1}^{n}E\left[\exp\left(-is\Psi_t\right)\mid\eta_r,r\leq t-1\right]\cdot Z_t\right)\cdot\exp\left(is\psi^0\right)
\end{aligned}$$

and we obtain

$$\left|S_n(\psi^0)\right|\leq\frac{1}{2\pi}\sum_{s=-\infty}^{\infty}|\alpha_s(b_n)|\left|\frac{1}{n}\sum_{t=1}^{n}E\left[\exp\left(-is\Psi_t\right)\mid\eta_r,r\leq t-1\right]\cdot Z_t\right|. \tag{7.31}$$

Further, since the right-hand side of (7.31) is independent of ψ^0:

$$\sup_{\psi^0\in[0,2\pi)}\left|S_n(\psi^0)\right| \leq \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}|\alpha_s(b_n)|\left|\frac{1}{n}\sum_{t=1}^{n}E\left[\exp\left(-is\Psi_t\right)\mid \eta_r, r\leq t-1\right]\cdot Z_t\right|. \tag{7.32}$$

Using the notations

$$\xi_{s,t;\cos} := E\left[\cos\left(s\Psi_t\right)\mid \eta_r, r\leq t-1\right] \quad \text{and} \quad \xi_{s,t;\sin} := E\left[\sin\left(s\Psi_t\right)\mid \eta_r, r\leq t-1\right],$$

we rewrite (7.32) as

$$\sup_{\psi^0\in[0,2\pi)}\left|S_n(\psi^0)\right| \leq \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}|\alpha_s(b_n)|\left|\frac{1}{n}\sum_{t=1}^{n}\left(\xi_{s,t;\cos}Z_t - i\xi_{s,t;\sin}Z_t\right)\right|.$$

The expected value then results in

$$E\left[\sup_{\psi^0\in[0,2\pi)}\left|S_n(\psi^0)\right|\right] \leq \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}|\alpha_s(b_n)|E\left[\left|\frac{1}{n}\sum_{t=1}^{n}\left(\xi_{s,t;\cos}Z_t - i\xi_{s,t;\sin}Z_t\right)\right|\right] \tag{7.33}$$

Noting that

$$E\left[|X+Y|\right] \leq \sqrt{E\left[|X|^2\right]+E\left[|Y|^2\right]}$$

for two random variables X and Y, we obtain

$$E\left[\left|\frac{1}{n}\sum_{t=1}^{n}\left(\xi_{s,t;\cos}Z_t - i\xi_{s,t;\sin}Z_t\right)\right|\right] \leq \sqrt{E\left[\left|\frac{1}{n}\sum_{t=1}^{n}\xi_{s,t;\cos}Z_t\right|^2\right]+E\left[\left|\frac{1}{n}\sum_{t=1}^{n}\xi_{s,t;\sin}Z_t\right|^2\right]}$$

Since $E\left[\xi_{s,t;\cos}Z_t\right]=0=E\left[\xi_{s,t;\sin}Z_t\right]$ and since, according to Beran and Ghosh (2020) [34], both, the covariances of $\sin\left(\Psi\right)$ and $\cos\left(\Psi_t\right)$, have the same asymptotic behavior as the covariance of $\left(\Psi_t\right)$, (7.33) becomes (analogously as in the proof of Proposition 7.4.9)

$$\begin{aligned} E\left[\sup_{\psi^0\in[0,2\pi)}\left|S_n(\psi^0)\right|\right] &\leq \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\left(|\alpha_s(b_n)|\sqrt{\operatorname{var}\left(n^{-1}\sum_{t=1}^{n}\xi_{s,t;\cos}Z_t\right)+\operatorname{var}\left(n^{-1}\sum_{t=1}^{n}\xi_{s,t;\sin}Z_t\right)}\right) \\ &\leq C\cdot n^{d_Z+d_X-1}\sum_{s=-\infty}^{\infty}|\alpha_s(b_n)|, \end{aligned}$$

where $C>0$ is a suitable constant. Hence, condition (7.30) on the sum of Fourier coefficients implies

$$\lim_{n\to\infty}E\left[\sup_{\psi^0\in[0,2\pi)}|S_n(\psi^0)|\right]=0.$$

■

A further assumption on the sum of Fourier coefficients which is strongly related to assumption (7.30) is given in terms of the sequence of circular kernel bandwidths (b_n):

Corollary 7.4.12. *Suppose that the assumptions of Corollary 7.4.7 hold. Suppose further that for some $\beta > 0$,*

$$\sum_{s=-\infty}^{\infty} |\alpha_s(b_n)| = \mathcal{O}\left(b_n^{-\beta}\right), \qquad as\ n \to \infty, \tag{7.34}$$

and that (b_n) is an optimal sequence of bandwidths satisfying the consistency condition (BW) and

$$b_n^{-\beta} = o\left(n^{1-d_Z-d_X}\right). \tag{7.35}$$

Then

$$\sup_{\psi^0 \in [0,2\pi)} |\hat{\nu}_n(\psi^0) - \nu(\psi^0)| \xrightarrow{P} 0.$$

Proof. The result directly follows from Theorem 7.4.11, since combining condition (7.34) with condition (7.35) yields condition (7.30).

■

In order to investigate how restrictive the assumptions of Corollary 7.4.12 and 7.4.11 are, we have a look at two main circular distributions, the von Mises distribution and the wrapped normal distribution, and their associated kernels.

Example 7.4.13. *From Example 6.2.5 (i) we know that the von Mises kernel defined in (6.20) with $\kappa = b^{-2}$,*

$$K_b(u) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(u)}, \tag{7.36}$$

has the Fourier representation

$$K_b(u) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |\alpha_s(\kappa)| e^{-isu},$$

where

$$\alpha_s(\kappa) = \frac{I_s(\kappa)}{I_0(\kappa)}$$

and $I_s(\kappa)$ is the modified Bessel function of order s defined in (5.35).

Since $I_s(\kappa) > 0$ for all κ (from $K_b(0)$ in (7.36)), we obtain for the coefficients:

$$\begin{aligned} \sum_{s=-\infty}^{\infty} |\alpha_s(\kappa)| &= 2\pi K_b(0) \\ &= \frac{e^{\kappa}}{I_0(\kappa)}, \end{aligned}$$

and consequently

$$I_0(\kappa) \underset{\kappa\to\infty}{\sim} \frac{e^{\kappa}}{\sqrt{2\pi\kappa}},$$

which implies

$$\sum_{s=-\infty}^{\infty} |\alpha_s(\kappa)| \underset{\kappa\to\infty}{\sim} \sqrt{2\pi\kappa} = \sqrt{2\pi} b^{-1}.$$

Hence, it follows that $\beta = 1$ in (7.34) and condition (7.35) reduces to

$$\lim_{n\to\infty} n^{1-d_Z-d_X} b_n = \infty.$$

Example 7.4.14. *The wrapped normal kernel defined in (6.21) from Example 6.2.5 (ii) with $\sigma = b$,*

$$K_b(u) = \left(\sqrt{2\pi}b\right)^{-1} \sum_{s=-\infty}^{\infty} \exp\left(-\frac{(u+2\pi s)^2}{2b^2}\right), \tag{7.37}$$

has the Fourier series representation

$$K_b(u) = (2\pi)^{-1} \sum_{s=-\infty}^{\infty} |\alpha_s(b)| e^{-isu},$$

where $\alpha_s(b) = \exp\left(-\frac{b^2}{2}s^2\right)$.

Thus, the absolute sum of the coefficients (from $K_b(0)$ in (7.37)) is given by

$$\begin{aligned} \sum_{s=-\infty}^{\infty} |\alpha_s(b)| &= 2\pi K_b(0) \\ &= \sqrt{2\pi} b^{-1} \sum_{s=-\infty}^{\infty} \exp\left(-\frac{2\pi^2}{b^2}s^2\right) \end{aligned}$$

and for $b \leq 1$

$$\begin{aligned} \sum_{s=-\infty}^{\infty} |\alpha_s(b)| &\leq \sqrt{2\pi} b^{-1} \sum_{s=-\infty}^{\infty} \exp\left(-2\pi^2 s^2\right) \\ &= Cb^{-1}. \end{aligned}$$

Again, condition (7.34) is satisfied for $\beta = 1$, and condition (7.35) becomes

$$\lim_{n\to\infty} n^{1-d_Z-d_X} b_n = \infty.$$

This means that the kernel estimators $\hat{\nu}_n$ with both, the von Mises kernel considered in Example 7.4.13 and the wrapped normal kernel considered in Example 7.4.14 serve as a consistent esti-

mator of the mean direction yielding uniform convergence in probability. These two examples show that the assumptions are satisfied in standard situations.

7.5 Asymptotic inference

In this chapter, we discuss asymptotic inference for the mean function. On the one hand, we use the asymptotic distribution of the difference $\nu_n(\psi^0) - \nu(\psi^0)$ given in Theorem 7.4.10, mainly the second statement in this theorem, to define asymptotic confidence intervals for the mean direction ν. On the other hand, we formulate a hypothesis test for condition (7.16) from Theorem 7.4.4.

As to the confidence bands, we obtain the following corollary:

Corollary 7.5.1. *Under the same assumptions as in Theorem 7.4.10, a two-sided asymptotic confidence interval for ν is given by*

$$\hat{\nu}_n(\psi^0) \pm n^{d_Z - \frac{1}{2}} \cdot c_\nu(\psi^0) \cdot q_{1-\alpha/2},$$

where $c_\nu(\psi^0)$ is as in (7.26) and $q_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution.

More generally, for a sample $\psi_1^0 < \ldots < \psi_k^0$, $\psi_j^0 \in [0, 2\pi)$, $j = 1, \ldots, k$, we obtain a simultaneous asymptotic confidence interval of the form

$$\left[\hat{\nu}_n(\psi_1^0), \ldots, \hat{\nu}_n(\psi_k^0)\right] \pm n^{d_Z - \frac{1}{2}} \cdot \left[c_\nu(\psi_1^0), \ldots, c_\nu(\psi_k^0)\right] \cdot q_{1-\alpha/2}, \tag{7.38}$$

where $c_\nu(\psi_j^0)$, $j = 1, \ldots, k$, are as in (7.28) and $q_{1-\alpha/2}$ is as before.

Proof. Both results directly follow from Theorem 7.4.10. ■

Remarks. *(i) This asymptotic confidence interval is simultaneous due to the assumption of a finite number of values $\psi_1^0, \ldots, \psi_k^0$.*

(ii) In practice, the value of $c_\nu(\psi_j^0)$, $j = 1, \ldots, k$, in (7.38) is unknown and has to be estimated, e.g., by local Whittle estimation (see Chapter 3.2.1) and, in addition, d_Z and $c_{f,Z}$ have to be replaced by consistent estimates.

An asymptotic confidence interval can also be specified in the case that the equivalent statements in Theorem 7.4.10 are not satisfied, which in turn means that there exists *no such optimal sequence of bandwidths* and, moreover, the convergence rate of the kernel estimator always depends on the bandwidth:

Remark. *A general confidence interval can even be defined by considering the asymptotic distribution of the terms in (7.12).*

Note that

$$C_n = \mathcal{O}_p\left(b_n^2 \bar{X}_n\right),$$

then Lemma 7.3.2 and Theorem 7.4.10 yield

$$\begin{aligned}(nb_n)^{-\frac{1}{2}} A_n &\xrightarrow{d} W,\\ n^{\frac{1}{2}-d_Z}\tilde{B}_n &\xrightarrow{d} W,\\ n\sigma_{n,X}^{-1} C_n &\xrightarrow{d} b_n^2 \cdot W,\end{aligned}$$

where $W \sim N(0,1)$, $\sigma_{n,X}^2 = var(\bar{X})$, *and*

$$\left(n\sigma_{n,X}^{-1}\right)^{-1} = \mathcal{O}_p\left(n^{d_X-\frac{1}{2}}\right).$$

Assume that the sequence of bandwidths (b_n) *is such that*

$$\frac{b_n^{2k}}{(nb_n)^{-\frac{1}{2}} + n^{d_Z-\frac{1}{2}} + b_n^2 n^{d_X-\frac{1}{2}}} \xrightarrow[n\to\infty]{} C \tag{7.39}$$

where $C > 0$ *is some constant, or equivalently that*

$$b_n^{2k} = \mathcal{O}\left((nb_n)^{-\frac{1}{2}} + n^{d_Z-\frac{1}{2}} + b_n^2 n^{d_X-\frac{1}{2}}\right).$$

This condition (7.39) ensures that the sequence of bandwidths does not converge to zero too fast and, furthermore, that the bias is not the dominating term.

A general asymptotic confidence interval for ν *is then given by*

$$\hat{\nu}_n(\psi^0) \pm c_n \cdot q_{1-\alpha/2},$$

where c_n *is such that*

$$c_n = \begin{cases} \frac{b_n^2 n^{d_X-\frac{1}{2}}}{f_\Psi(\psi^0)}, & \text{if } A_n + \tilde{B}_n = o(C_n),\\ n^{d_Z-\frac{1}{2}}, & \text{if } A_n + C_n = o(\tilde{B}_n),\\ \frac{(nb_n)^{-\frac{1}{2}}}{f_\Psi(\psi^0)}, & \text{if } \tilde{B}_n + C_n = o(A_n),\end{cases}$$

with marginal density f_Ψ *of* (Ψ_t). *Here, the second case corresponds to the situation of Corollary 7.5.1, which is our main interest in the following testing procedure.*

Moreover, for a sequence $\psi_1^0 < \ldots < \psi_k^0$, *a general simultaneous asymptotic confidence interval*

can be defined analogously as in Corollary 7.5.1.

Note, that (7.39) is an analogue of the condition given in Mielniczuk and Wu (2004) [298] in the real-valued situation, and this remark displays the *smoothing trichotomy*, which is already addressed in the same paper.

Now, we want to test for the existence of such an optimal sequence of bandwidths. This means, we test if the term B_n in (7.14) or the third term in (7.19), respectively, is the dominating term, and use the last of the three equivalent statements from Theorem 7.4.4 and Theorem 7.4.5, respectively.

We consider a random sequence $\Theta_1, \ldots, \Theta_n$ generated by a process (Θ_t) of the form

$$\Theta_t = [\nu(\Psi_t) + Z_t] \ mod\, 2\pi = [\mu(X_t) + Z_t] \ mod\, 2\pi,$$

for which the assumptions (B1) to (B3) hold. Now, Theorems 7.4.4 and 7.4.5 state that an optimal sequence of bandwidths exists if and only if the memory parameter d_Z satisfies

$$d_Z > \max\left\{\frac{1}{5}d_X,\ \frac{1}{2(4k+1)}\right\}, \qquad k \geq 1.$$

The testing procedure can be divided into two main steps. In a first step, we estimate the unknown parameters of the long-memory processes and in a second step, we define the test statistic and derive asymptotic rejection regions.

Step 1: Estimation of the memory parameters d_Z and d_X

Under the given assumptions and, following Beran and Ghosh (2020) [34], we know that $(\sin(\Psi_t))$ and (Ψ_t) have the same memory parameter d_X. Thus, we can estimate d_X based on

$$S_{\Psi,t} = \sin(\Psi_t), \qquad t = 1, \ldots, n.$$

Estimation of the memory parameter d_Z, is based on the estimated residuals

$$\hat{Z}_t = [\Theta_t - \hat{\nu}(\Psi_t)] \mod 2\pi, \qquad t = 1, \ldots, n,$$

and we obtain $\hat{d}_Z$ from $S_{\hat{Z},t} = \sin(\hat{Z}_t)$, $t = 1, \ldots, n$.

A consistent method for estimating long-memory parameters is the *local Whittle estimation*, which is discussed in Part I, Chapter 3.2.1. The estimators for $\hat{d}_X$ and $\hat{d}_Z$ are based on the periodogram of the series $S_{\Psi,t}$ and $S_{\hat{Z},t}$, respectively, at the lowest m frequencies. Recall from (3.19) that, as $m \to \infty$ and $\frac{m}{n} \to 0$, under mild regularity conditions (see Robinson (1995b) [341])

$\hat{d}_Z$ and $\hat{d}_X$ are consistent and asymptotically normal estimators, i.e.,

$$\sqrt{m}\left(\hat{d}_Z - d_Z\right) \xrightarrow{d} \frac{1}{2}W_1, \qquad \text{and} \qquad \sqrt{m}\left(\hat{d}_X - d_X\right) \xrightarrow{d} \frac{1}{2}W_2, \tag{7.40}$$

where W_1 and W_2 are standard normal random variables, independent of each other.

Step 2: Defining the test statistic

This second step now depends on the order of the circular kernel. First, we have a look at the commonly used second-order kernels. According to Corollary 7.4.6 (i), in this case, i.e., for $k = 1$, provided that the assumptions (B1) to (B3) hold, we wish to test the null hypothesis

$$H_0 : \; d_Z \leq \frac{1}{10} \tag{7.41}$$

against the alternative

$$H_1 : \; d_Z > \frac{1}{10}. \tag{7.42}$$

This means, we do not need any information on the strength of dependence of the processes (X_t) or (Ψ_t), and only need a consistent estimate for d_Z.

Hence, given a level of significance $\alpha \in (0, 1)$, we obtain with (7.40) the following asymptotic rejection region

$$T > \frac{1}{10} + \frac{1}{2\sqrt{m}} \cdot q_{1-\alpha},$$

where $q_{1-\alpha}$ is the $(1-\alpha)$-quantile of the standard normal distribution.

This testing procedure is more complicated when considering higher order kernels, i.e., when considering a general circular kernel of order $2k$ with $k > 1$. Here, according to Theorem 7.4.5 (iii), we wish to test the null hypothesis

$$H_0 : \; d_Z \leq \max\left\{\frac{1}{5}d_X, \; \frac{1}{2(4k+1)}\right\} \tag{7.43}$$

against the alternative

$$H_1 : \; d_Z > \max\left\{\frac{1}{5}d_X, \; \frac{1}{2(4k+1)}\right\},$$

which requires knowledge about both memory parameters d_X and d_Z. Let $\hat{d}_X$ and $\hat{d}_Z$ be local Whittle estimates obtained as given in Step 1. Since under H_0, (7.43) leads to

$$d_Z - \frac{1}{5}d_X \leq \max\left\{0, \; \frac{1}{2(4k+1)} - \frac{1}{5}d_X\right\}$$

and to

$$d_Z - \frac{1}{2(4k+1)} \leq \max\left\{\frac{1}{5}d_X - \frac{1}{2(4k+1)}, \; 0\right\},$$

we define

$$T_1 = \hat{d}_Z - \frac{1}{5}\hat{d}_X, \qquad \text{and} \qquad T_2 = \hat{d}_Z - \frac{1}{2(4k+1)}.$$

Thus, taking the minimum of T_1 and T_2 and defining the test statistic

$$T := \min\{T_1,\ T_2\}$$

yields the equivalent hypothesis test

$$H_0:\ T \leq 0 \quad \text{vs.} \quad H_1:\ T > 0.$$

Then, still under H_0, and accounting for asymptotic normality of the estimators $\hat{d}_Z$ and $\hat{d}_X$ given in (7.40), we obtain the following asymptotic inequalities

$$T_1 \leq d_Z + \frac{1}{2\sqrt{m}}W_1 - \frac{1}{5}d_X - \frac{1}{10\sqrt{m}}W_2 + o_p\left(m^{-\frac{1}{2}}\right)$$

and

$$T_2 \leq d_Z + \frac{1}{2\sqrt{m}} \cdot W_1 - \frac{1}{2(4k+1)} + o_p\left(m^{-\frac{1}{2}}\right).$$

Consequently,

$$T \leq d_Z + \frac{1}{2\sqrt{m}}W_1 - \max\left\{\frac{1}{5}d_X + \frac{1}{10\sqrt{m}}W_2,\ \frac{1}{2(4k+1)}\right\} + o_p\left(m^{-\frac{1}{2}}\right),$$

and we have equality if

$$d_Z = \frac{1}{5}d_X \qquad \text{or} \qquad d_Z = \frac{1}{2(4k+1)}.$$

Under H_1, in contrast,

$$T > d_Z + \frac{1}{2\sqrt{m}}W_1 - \max\left\{\frac{1}{5}d_X + \frac{1}{10\sqrt{m}}W_2,\ \frac{1}{2(4k+1)}\right\} + o_p\left(m^{-\frac{1}{2}}\right),$$

and the rejection region at an asymptotic level of significance $\alpha \in (0,1)$ is given by

$$T > q_{1-\alpha}(d_Z, d_X),$$

where $q_{1-\alpha}(d_Z, d_Z)$ is such that

$$P\left(d_Z + \frac{1}{2\sqrt{m}}W_1 - \max\left\{\frac{1}{5}d_X + \frac{1}{10\sqrt{m}}W_2,\ \frac{1}{2(4k+1)}\right\} > q_{1-\alpha}(d_Z, d_X)\right) = \alpha.$$

This means, the quantile $q_{1-\alpha}(d_Z, d_X)$ depends on both memory parameters d_X and d_Z. Values

for this quantile can, e.g., be obtained by numerical integration or simulation.

Remark. *In practice, as discussed in Step 1,* d_X *and* d_Z *are unknown and have to be replaced by consistent estimates* $\hat{d}_X$ *and* $\hat{d}_Z$*, respectively, yielding approximate asymptotic* $(1-\alpha)$*-quantiles* $q_{1-\alpha}\left(\hat{d}_Z, \hat{d}_X\right)$.

7.6 Data example: Daily wind directions

In this last chapter, we apply our theoretical results to a data example from meteorology. We consider daily average wind directions recorded between January 1st and December 31st of 2020 at three meteorological stations in the United States, namely Milwaukee (WI), Chicago (IL), and Alpena (MI). The first two cities are located around the Lake Michigan, Milwaukee in the South-West and Chicago in the South, and the latter one, Alpena, is located at Thunder Bay along Lake Huron's shore. These data are obtained from the homepage of the NOAA Great Lakes Environmental Research Laboratory (`https://www.glerl.noaa.gov/`).

Figures, simulations and the test procedure were conducted in R 4.1.0, R Core Team.

Figure 7.1 displays wind rose plots of the wind directions at the three locations.

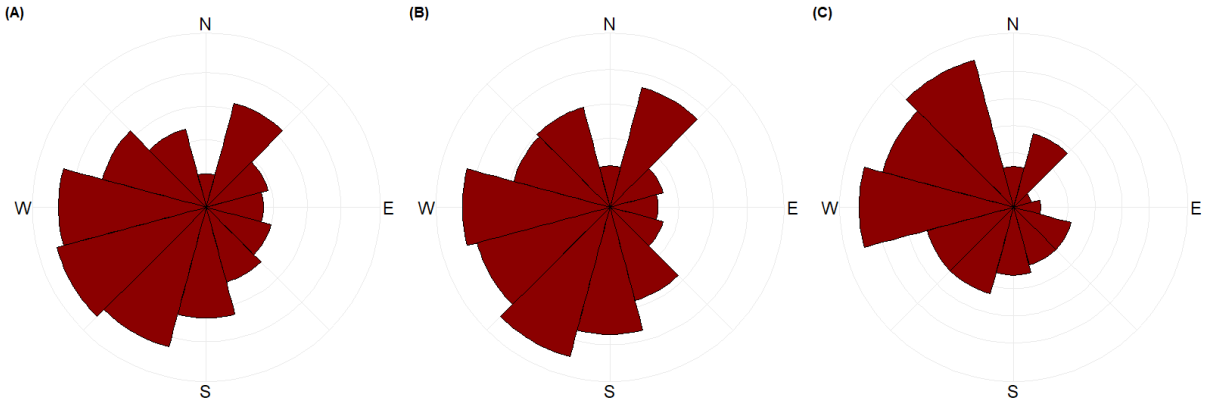

Figure 7.1: Wind rose plots of daily wind directions: (A) Milwaukee, (B) Chicago, (C) Alpena.

Assuming stationarity, we obtain estimates for a constant mean direction according to Beran and Ghosh (2020) [34], (6.13). Interestingly, for both, Milwaukee and Chicago, the mean direction is approx. 257 degrees which corresponds to west-southwest wind and, in each case, is related to a dominating land-breeze. For Alpena, we obtain a mean direction of approx. 283 degrees which corresponds to west-northwest wind, and still relates to a dominating land-breeze.

However, in general, one would expect a seasonal pattern. We therefore consider a seasonal

model with period $T_0 = 365$ of the form

$$S_t(\beta) = \left[\beta_0 + \sum_{j=1}^{3} \beta_j \sin(\lambda_j t) + \beta_{3+j} \cos(\lambda_j t) \right] \mod 2\pi,$$

where $\lambda_j = \frac{2\pi j}{T_0}$, $j = 1, \ldots, 6$, are the Fourier frequencies. The coefficients β are estimated by minimizing the circular distance measure (5.4), i.e., by minimizing

$$Q_n(\beta) = \sum_{t=1}^{n} \{1 - \cos(\Theta_t - S_t(\beta))\}.$$

This deseasonalization method was proposed by Beran and Ghosh (2020) [34] and is addressed in Chapter 6.1. Figure 7.2 displays the daily wind directions together with estimated seasonal series in red solid line (panels (A), (C), (E)) and the corresponding seasonally adjusted wind directions (panels (B), (D), (F)) for the three locations.

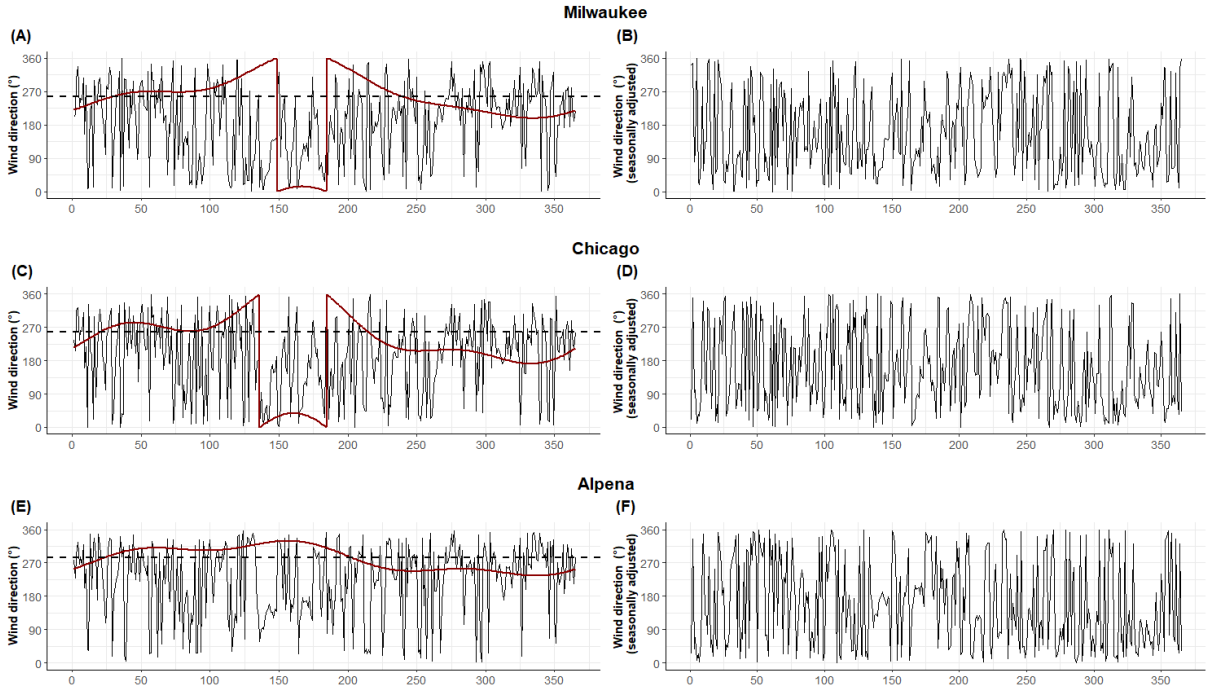

Figure 7.2: Daily average and seasonally adjusted wind directions: Left three panels show daily average wind directions together with fitted seasonal trends (solid red line) and estimated mean directions (dashed black line) for Milwaukee (A), Chicago (C) and Alpena (E). The estimated seasonally adjusted residual series are displayed in the right three panels, Milwaukee (B), Chicago (D) and Alpena (F).

We define (Θ_t) and (Ψ_t) as the respective deseasonalized series of the wind directions and have

to confirm the assumption of a circular-circular regression model between (Θ_t) and (Ψ_t). The following three relationships are studied and displayed in Figure 7.3: Chicago vs. Milwaukee (panel A), Alpena vs. Milwaukee (panel B), and Alpena vs. Chicago (panel C). The conditional

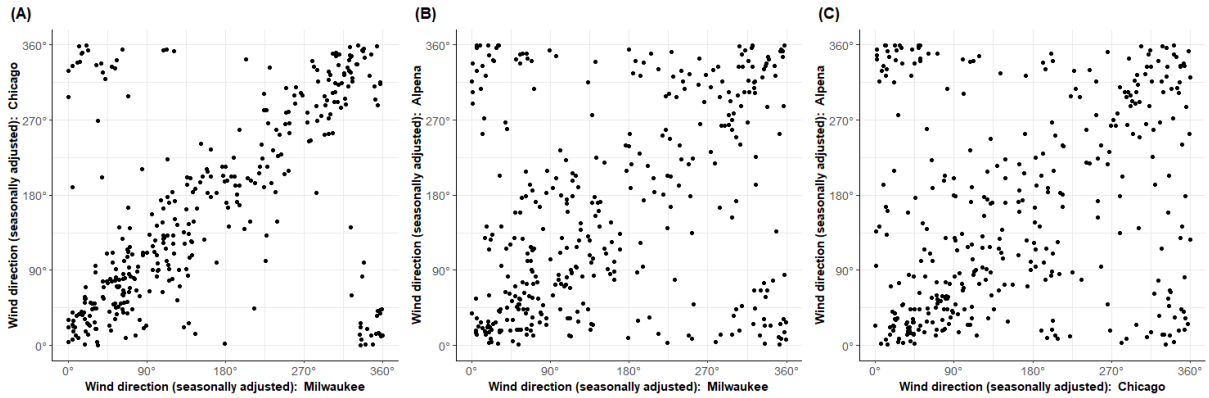

Figure 7.3: Scatterplot of seasonally adjusted wind directions: (A) Chicago vs. Milwaukee, (B) Alpena vs. Milwaukee, (C) Alpena vs. Chicago.

mean direction is estimated utilizing the circular NW-estimator (6.24) with von Mises kernel (6.20). However, we observe boundary effects which need some boundary correction modulo 2π. See here the discussion in Chapter 3 for real-valued kernels. To overcome this issue, in addition, the *kernel smoother* with von Mises kernel proposed by Di Marzio et al. (2012) [114] (and mentioned in Chapter 6.3), is applied. This kernel smoothes the sine part and the cosine part separately. The two approaches are displayed in Figure 7.4 for Chicago vs. Milwaukee, and in Figure 7.5 for all three relationships.

The assumption of long-range dependence of both series is proven by Whittle estimation together with fitting a fractional ARIMA(p,d,0)-model up to order $p_{\max} = [\log(n)]$ combined with the BIC model selection criterion, see Chapter 3.2. Table 7.1 shows the long-memory parameter for the deseasonalized series estimated from $(\sin(\Psi_t))$:

	Milwaukee	Chicago
$\hat{d}_\Psi$	0.2 [0.12, 0.28]	0.14 [0.06, 0.22]

Table 7.1: Estimated long-memory parameters for seasonally adjusted explanatory series together with 95%-confidence intervals.

For the residual series $\hat{Z}_t = [\Theta_t - \hat{\nu}_n(\Psi_t)] \ \ mod\ 2\pi$, where $\hat{\nu}_n$ is estimated utilizing the NW-estimator or the kernel smoother, in each case with von Mises kernel, we obtain the long-memory parameter estimate from $\sin(\hat{Z})$ by local Whittle estimation with $m = [n^{0.6}]$ frequencies. The

estimated parameters are given in Table 7.2.

	Chicago vs. Milwaukee	Alpena vs. Milwaukee	Alpena vs. Chicago
NW-estimator	0.23 [0.06, 0.4]	0.19 [0.02, 0.36]	0.22 [0.05, 0.38]
kernel smoother	0.22 [0.05, 0.38]	0.26 [0.09, 0.42]	0.25 [0.08, 0.42]

Table 7.2: Local Whittle estimates $\hat{d}_Z$ of the long-memory parameters together with 95%-confidence interval for the residual series.

Utilizing the Whittle estimation together with fitting a fractional ARIMA(p,d,0)-model up to order $p_{\max} = [\log(n)]$ combined with the BIC model selection criterion only slightly changes the long-memory estimates. The resulting Whittle estimates are given in Table 7.3.

	Chicago vs. Milwaukee	Alpena vs. Milwaukee	Alpena vs. Chicago
NW-estimator	0.27 [0.19, 0.35]	0.25 [0.17, 0.33]	0.22 [0.14, 0.3]
kernel smoother	0.27 [0.19, 0.35]	0.3 [0.22, 0.38]	0.31 [0.23, 0.39]

Table 7.3: Whittle estimates $\hat{d}_Z$ of the long-memory parameters together with 95%-confidence interval for the residual series.

Since the von Mises kernel is a kernel of order 2, we are in the reduced hypothesis test situation given in (7.41) and (7.42), i.e.,

$$H_0 : \ d_Z \leq \frac{1}{10}$$

is tested against the alternative

$$H_1 : \ d_Z > \frac{1}{10},$$

and H_0 is rejected at the 5% level of significance, if, for $m = [n^{0.6}]$,

$$T > \frac{1}{10} + \frac{1}{2\sqrt{m}} \cdot 1.96 = 0.24.$$

According to Tables 7.2 and 7.3, the upper confidence limits of all long-memory parameters exceed 0.24. Thus, we can reject H_0 in all situations and consequently can calculate 95%-confidence bands for the mean direction based on Theorem 7.4.10 and (7.38).

Figure 7.4 shows the seasonally adjusted wind directions for Chicago vs. Milwaukee together with the NW-estimates and the corresponding 95%-confidence band based on the Whittle estimates from Table 7.3, and illustrates the issue concerning the boundary effect.

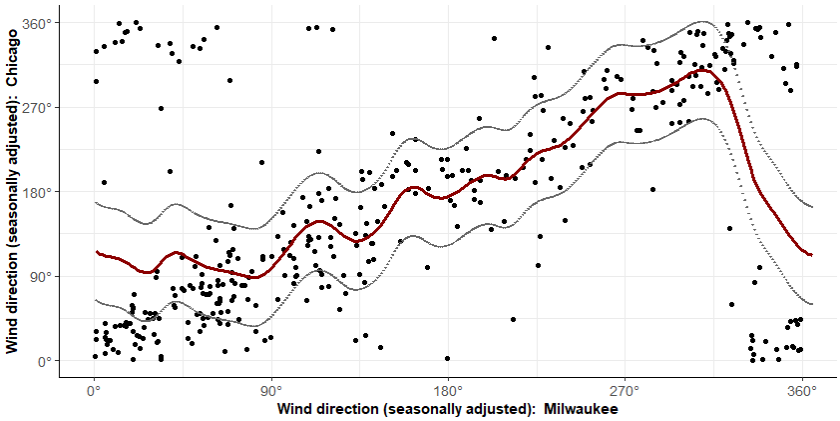

Figure 7.4: Scatterplot of seasonally adjusted wind directions from Chicago and Milwaukee as black points together with the NW-estimates (without boundary correction) as solid red line and a 95%-confidence band as dotted grey line.

The results for all three relationships based on the kernel smoother (and on the Whittle estimates) are displayed in Figure 7.5.

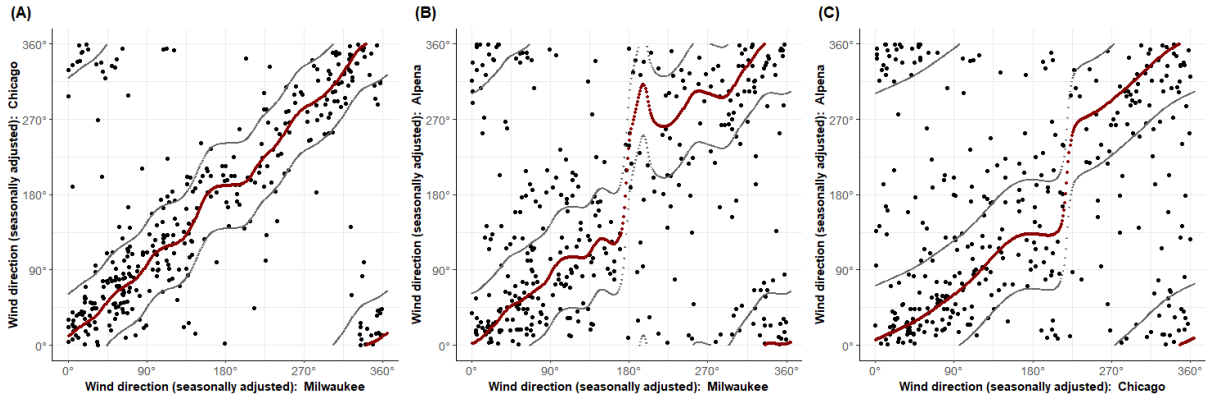

Figure 7.5: Scatterplot of seasonally adjusted wind directions together with estimated kernel smoothed mean direction as dotted red line and a 95%-confidence band as dotted grey line: (A) Chicago vs. Milwaukee, (B) Alpena vs. Milwaukee, (C) Alpena vs. Chicago.

For the two relationships with wind directions at Alpena as dependent variable (panels (B) and (C)), there seems to be a distinct structural break at approx. 180 degrees for Alpena vs. Milwaukee (B) and at approx. 225 degrees Alpena vs. Chicago (C). Interestingly, this is not the case for the estimated mean direction of Chicago vs. Milwaukee (A). This finding might possibly be related to the different locations at the Lake Michigan and the Lake Huron and the dominance of land-breeze.

In order to visualize the behavior of the estimated mean direction, a three-dimensional plot for the circular-circular regression model of Chicago vs. Milwaukee is presented in Figure 7.6.

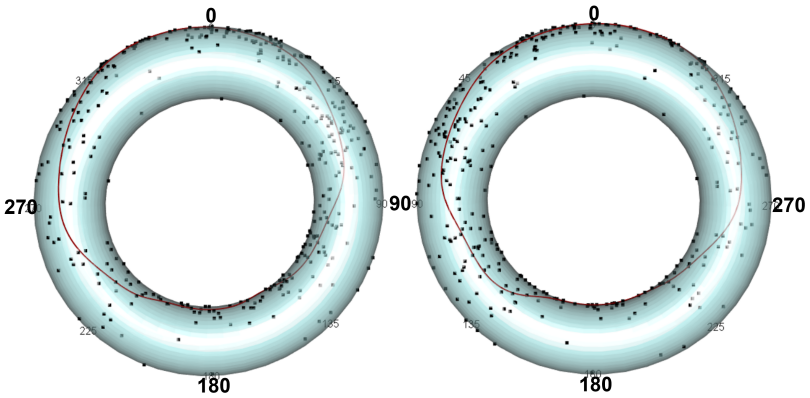

Figure 7.6: 3D-plot of seasonally adjusted wind directions (Chicago vs. Milwaukee) together with estimated mean direction as solid red line: Left torus shows the front side and the right torus shows the back side (horizontally rotated by 180 degrees).

Since the state space is $[0, 2\pi) \times [0, 2\pi)$ with quotient topology, it is homeomorphic to a torus, on which the estimated mean direction can be displayed. Here, the mean direction is estimated utilizing the kernel smoother with von Mises kernel. The seasonally adjusted wind directions from Milwaukee as independent variable are displayed in *toroidal direction* ("large circular ring") and those from Chicago as dependent variable are depicted in *poloidal direction* ("small circular ring"). To explore the evolution of the wind directions and the characteristics of the estimated mean direction the front side (left panel) and the back side (right panel) of the torus are shown.

Chapter 8

Final remarks

This final chapter contains some thoughts and addresses some open questions in the context of our investigations.

- In Part I and Part II, one could ask if similar results can be obtained for long-memory processes in continuous time. In the real-valued state space the definition of long-range dependence can be directly extended to continuous-time processes (see Beran et al., 2013 [33]) and even to spatio-temporal random fields (see, e.g., Leonenko and Ruiz-Medina, 2022 [262]). In an analogous manner, it should be possible to define circular continuous-time processes. Furthermore, one could be interested whether one obtains similar results in case of time-varying long-memory parameters, i.e., when considering locally stationary processes (see, e.g., Beran, 2009 [24]).

- Part I considers an equidistant (fixed-design) seasonal time series. Here, the theory might be extended to non-equidistant time series and potentially to randomly-spaced time series (see, e.g., Haye et al., 2019 [197]).

- As we considered linear long-memory processes, it could be worthwhile to broaden this assumption, e.g., to processes with mixing conditions.

- We considered the expected number of seasonal exceedances, which corresponds to sojourn times of continuous-time processes, see, e.g., Berman (2017) [46]. Here, one could as well be interested in multiple changes and the number of upcrossings (e.g., when (X_t) models flood levels).

- A further question concerns first-passage times T in general and seasonal first-passage times such as, e.g., the first day in the year the temperature exceeds a certain threshold. As known from stopping times of Markov processes, here one would have to determine the

joint distribution of (T, X_T). This question could also be of interest when modeling flood levels.

- Furthermore, one could extend the theory of seasonal exceedances to multivariate seasonal time series. This, in particular, requires the extension of the multivariate weak reduction principle (as studied, e.g., in Beran and Telkmann, 2018 [41]) to multivariate periodic empirical processes.

- Regarding Part II, one could extend the theory of locally weighted regression to the circular state space by combining the results by Beran and Ghosh (2020) [34] on parametric estimation of the mean direction and circular kernel estimation as it is done in the real-valued situation, for instance, by Beran, Steffens and Ghosh (2018) [37].

- A further topic, which is widely discussed in the real-valued setting, is testing for a shift in the mean function or for structural breaks in the context of long-range dependence. As we have seen in our data example with the wind directions in Chapter 7.6, such structural breaks can occur in this circular setting as well. Thus, one could investigate hypotheses tests for structural breaks or for shifts in the mean direction, tentatively by defining a circular Kolmogorov-Smirnov-type test statistic and extending the results from the real-valued to the circular state space.

- Considering examples of oceanographic velocity measurements or wind direction measurements in two dimensions, one would be interested in circular-circular regression models in the product state space $[0, 2\pi)^2 \times [0, 2\pi)^2$ under the assumption of long-range dependence.

- Our circular state space $\mathbb{R}\backslash[0, 2\pi)$ is a quotient space (precisely, a topological vector space) with respect to the equivalence relation modulo 2π, and we deal with kernel functions on this quotient space using a suitable distance measure. Following the arguments in Chapters 5 and 6, one could develop a general theory of kernel estimation where the crucial issue is a distance measure that induces the quotient topology and that is related to a meaningful notion of covariance.

- Last but not least, the question arises how to determine a general condition on the equivalence relation for which the long-memory property is passed on to the quotient space - so far, this is known for the equivalence relation associated with Gaussian subordinated processes.

Bibliography

[1] Abadir, K.M., Distaso, W. and Giraitis, L. (2007). Nonstationary-extended local Whittle estimation. *Journal of Econometrics*, **141(2)**, 1353-1384, Elsevier.

[2] Abe, T. and Pewsey, A. (2011a). Sine-skewed circular distributions. *Statistical Papers*, **52**, 683-707, Springer.

[3] Abe, T. and Pewsey, A. (2011b). Symmetric circular models through duplication and cosine perturbation. *Computational Statistics & Data Analysis*, **55**, 3271-3282, Elsevier.

[4] Akaike, H. (1969a). Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics*, **21**, 243-247, Springer.

[5] Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, **60**, 255-265, Oxford University Press.

[6] Altman, N.S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association*, **85**, 749-759, Taylor & Francis Group.

[7] Ameijeiras-Alonso, J., Ley, C., Pewsey, A. and Verdebout, T. (2021). On optimal tests for circular reflective symmetry about an unknown central direction. *Statistical Papers*, 1-24, Springer.

[8] Appell, P. (1880). Sur une classe de polynômes. In: *Annales scientifiques de l'École normale supérieure*, **9**, 119-144.

[9] Arcones, M.A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *The Annals of Probability*, 2242-2274, JSTOR.

[10] Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, **4**, 40-79.

[11] Avram, F. and Taqqu, M.S. (1987). Noncentral limit theorems and Appell polynomials. *The Annals of Probability*, 767-775, JSTOR.

[12] Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 171-178, JSTOR.

[13] Azzalini, A. (2005). The skew-normal distribution and related multivariate families. *Scandinavian Journal of Statistics*, **32(2)**, 159-188, Wiley Online Library.

[14] Baek, C. and Pipiras, V. (2012). Statistical tests for a single change in mean against long-range dependence. *Journal of Time Series Analysis*, **33(1)**, 131-151, Wiley Online Library.

[15] Bai, Z.D., Rao, C.R. and Zhao, L.C. (1989). Kernel estimators of density function of directional data. In: *Multivariate Statistics and Probability*, 24-39, Elsevier.

[16] Bartlett, M.S. (1963). Statistical estimation of density functions. *Sankhyā: The Indian Journal of Statistics, Series A*, 245-254, JSTOR.

[17] Batschelet, E. (1965). Statistical methods for the analysis of problems in animal orientation and certain biological rhythms. *American Institute of Biological Sciences*.

[18] Batschelet, E. (1981). *Circular statistics in biology*. New York Academic Press.

[19] Bender, C. and Sottinen, T. and Valkeila, E. (2007). Arbitrage with fractional Brownian motion?. *Theory of Stochastic Processes*, **13**, 23-34.

[20] Beran, J. (1986). *Estimation, testing and prediction for self-similar and related processes*. PhD thesis, ETH Zürich.

[21] Beran, J. (1989). A test of location for data with slowly decaying serial correlations. *Biometrika*, **76**, 261-269, Oxford University Press.

[22] Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

[23] Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society: Series B (Methodological)*, **57**, 659-672, Wiley Online Library.

[24] Beran, J. (2009). On parameter estimation for locally stationary long-memory processes. *Journal of Statistical Planning and Inference*, **139**, 900-915.

[25] Beran, J. (2017). *Mathematical Foundations of Time Series Analysis: A Concise Introduction*. Springer.

[26] Beran, J., Bhansali, R.J. and Ocker, D. (1998). On unified model selection for stationary and nonstationary short-and long-memory autoregressive processes. *Biometrika*, **85**, 921-934, Oxford University Press.

[27] Beran, J. and Feng, Y. (2001). Local polynomial estimation with a FARIMA-GARCH error process. *Bernoulli*, **7**, 733-750.

[28] Beran, J. and Feng, Y. (2002a). Local polynomial fitting with long-memory, short-memory and antipersistent errors. *Annals of the Institute of Statistical Mathematics*, **54**, 291-311.

[29] Beran, J. and Feng, Y. (2002b). SEMIFAR models - a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis*, **40**, 393-419.

[30] Beran, J. and Feng, Y. (2002c). Iterative plug-in algorithms for SEMIFAR models - definition, convergence and asymptotic properties. *Journal of Computational and Graphical Statistics*, **11**, 690-713, Taylor & Francis Group.

[31] Beran, J., Feng, Y. and Ocker, D. (1999). *SEMIFAR models.* Technical Report.

[32] Beran, J., Feng, Y., Ghosh, S. and Sibbertsen, P. (2002). On robust local polynomial estimation with long-memory errors. *International Journal of Forecasting*, **18**, 227-241.

[33] Beran, J., Feng, Y., Ghosh, S. and Kulik, R. (2013). *Long-Memory Processes: Probabilistic Properties and Statistical Methods.* Springer-Verlag Berlin Heidelberg.

[34] Beran, J. and Ghosh, S. (2020). Estimating the mean direction of strongly dependent circular time series. *Journal of Time Series Analysis*, **41**, 210-228, Wiley Online Library.

[35] Beran, J., Möhrle, S. and Ghosh, S. (2016). Testing for Hermite rank in Gaussian subordination processes. *Journal of Computational and Graphical Statistics*, **25**, 917-934, Taylor & Francis.

[36] Beran, J. and Schumm, N. (2017). On nonparametric statistical inference for densities under long-range dependence. *Communications in Statistics - Theory and Methods*, **46**, 11296-11314, Taylor & Francis Group.

[37] Beran, J., Steffens, B. and Ghosh, S. (2018). On local trigonometric regression under dependence. *Journal of Time Series Analysis*, **39(4)**, 592-617, Wiley Online Library.

[38] Beran, J., Steffens, B. and Ghosh, S. (2021a). On nonparametric regression for bivariate circular long-memory time series. *Statistical Papers*, 1-24, Springer.

[39] Beran, J., Steffens, B. and Ghosh, S. (2021b). Testing for the expected number of exceedances in strongly dependent seasonal time series. *Journal of Nonparametric Statistics*, **33(3-4)**, 417-434, Taylor & Francis.

[40] Beran, J., Steffens, B. and Ghosh, S. (2022). Long-Range Dependence in Directional Data. In: *Directional Statistics for Innovative Applications: A Bicentennial Tribute to Florence Nightingale*, Springer Nature Singapore, 395-406.

[41] Beran, J. and Telkmann, K. (2018). On nonparametric density estimation for multivariate linear long-memory processes. *Communications in Statistics - Theory and Methods*, **47**, 5460-5473, Taylor & Francis Group.

[42] Beran, J. and Terrin, N. (1996). Testing for a change of the long-memory parameter. *Biometrika*, **83**, 627-638, Oxford University Press.

[43] Beran, R. (1979). Exponential models for directional data. *The Annals of Statistics*, 1162-1178, JSTOR.

[44] Berkes, I. and Philipp, W. (1977). An almost sure invariance principle for the empirical distribution function of mixing random variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **41**, 115-137, Springer.

[45] Berkes, I., Horváth, L., Kokoszka, P. and Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics*, **34(3)**, 1140-1165, Institute of Mathematical Statistics.

[46] Berman, S. (2017). *Sojourns and extremes of stochastic processes.* CRC Press.

[47] Betken, A. (2016). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *Journal of Time Series Analysis*, **37(6)**, 785-809, Wiley Online Library.

[48] Bhansali, R.J. (1982). The evaluation of certain quadratic forms occurring in autoregressive model fitting. *The Annals of Statistics*, 121-131, JSTOR.

[49] Bickel, P.J. and Wichura, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *The Annals of Mathematical Statistics*, 1656-1670, JSTOR.

[50] Bierens, H.J. (1983). Uniform consistency of kernel estimators of a regression function under generalized conditions. *Journal of the American Statistical Association*, **78(383)**, 699-707, Taylor & Francis.

[51] Billingsley, P. (1968). *Convergence of Probability Measures.*

[52] Billingsley, P. (1969). Maxima of partial sums. In: *Probability and Information Theory*, Springer, 64-76.

[53] Billingsley, P. (1999). *Convergence of Probability Measures.*. John Wiley & Sons.

[54] Billingsley, P. (2013). *Convergence of probability measures.* John Wiley & Sons.

[55] Bingham, N.H. (1972). Random walk on spheres. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **22**, 169-192, Springer.

[56] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1989). *Regular variation. Cambridge University Press.*

[57] Blake, A. and Marinos, C. (1990). Shape from texture: Estimation, isotropy and moments. *Artificial Intelligence*, **45**, 323-380, Elsevier.

[58] Borradaile, G.J. (2003). *Statistics of earth science data: their distribution in time, space and orientation.* Springer Science & Business Media.

[59] Bowman, A.W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, **71(2)**, 353-360, Oxford University Press.

[60] Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis: Forecasting and Control.* Wiley Series in Probability and Statistics.

[61] Breckling, J. (1984). Eine Analyse von Windrichtungen und Geschwindigkeiten. In: *Simulationstechnik*, 564-569, Springer.

[62] Breckling, J. (1989). *The analysis of directional time series: applications to wind speed and direction.*

[63] Breckling, J. (2012). *The analysis of directional time series: applications to wind speed and direction.* Springer Science & Business Media.

[64] Breuer, P. and Major, P. (1983). Central limit theorems for non-linear functionals of Gaussian fields. *Journal of Multivariate Analysis*, **13**, 425-441, Elsevier.

[65] Brockwell, P.J. and Davis, R.A. (2016). *Introduction to Time Series and Forecasting.* Springer.

[66] Brodsky, E. and Darkhovsky, B.S. (2013a). *Nonparametric Methods in Change-Point Problems.* Springer Science & Business Media.

[67] Browne, M.W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*.

[68] Bryk, A. (2014). Smoothing dichotomy in randomized fixed-design regression with strongly dependent errors based on a moving average. *Applicationes Mathematicae*, **1**, 67-80.

[69] Bryk, A. and Mielniczuk, J. (2007). Using randomization to improve the performance of regression estimators under dependence. *Acta Scientiarum Mathematicarum*, **73**, 817-838, Citeseer.

[70] Bryk, A. and Mielniczuk, J. (2008). Randomized fixed design regression under long-range-dependent errors. *Communications in Statistics - Theory and Methods*, **37**, 520-531, Taylor & Francis Group.

[71] Burnham, K.P. and Anderson, D.R. (2002). A practical information-theoretic approach. *Model Selection and Multimodel Inference*, **2**, 70-71, Springer New York.

[72] Burnham, K.P. and Anderson, D.R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, **33(2)**, 261-304, Sage Publications Sage CA: Thousand Oaks, CA.

[73] Cantelli, F.P. (1933). Sulla determinazione empirica delle leggi di probabilità. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 421-424.

[74] Chang, T. (1986). Spherical regression. *The Annals of Statistics*, 907-924, JSTOR.

[75] Cheng, B. and Robinson, P.M. (1991). Density estimation in strongly dependent non-linear time series. *Statistica Sinica*, 335-359, JSTOR.

[76] Cheng, B. and Robinson, P.M. (1994). Semiparametric estimation from time series with long-range dependence. *Journal of Econometrics*, **64**, 335-353, Elsevier.

[77] Cheung, Y.L. and Hassler, U. (2020). Whittle-type estimation under long memory and nonstationarity. *AStA - Advances in Statistical Analysis*, **104(3)**, 363-383, Springer.

[78] Chiu, S.T. (1989). Bandwidth selection for kernel estimate with correlated noise. *Statistics & Probability Letters*, **8(4)**, 347-354, Elsevier.

[79] Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, **74**, 829-836, Taylor & Francis Group.

[80] Cleveland, R.B., Cleveland, W.S., McRae, J.E. and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on LOESS. *Journal of Official Statistics*, **6**, 3-73, The AAAI Press.

[81] Coles, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, **8**, 105-113, Springer.

[82] Coursol, J. and Dacunha-Castelle, D. (1982). Remarks on the approximation of the likelihood function of a stationary Gaussian process. *Theory of Probability & Its Applications*, **27(1)**, 162-167, SIAM.

[83] Csörgő, M. (1979). Brownian Motion - Wiener Process. *Canadian Mathematical Bulletin*, **22**, 257-279, Cambridge University Press.

[84] Csörgő, S. (2002). The smoothing dichotomy in nonparametric regression under long-memory errors. *Statistica neerlandica*, **56**, 132-142, Wiley Online Library.

[85] Csörgő, M. and Horváth, L. (1988). Nonparametric methods for change point problems. *Handbook of Statistics*, **7**, 403-425.

[86] Csörgő, M. and Horváth, L. (1997). *Limit theorems in change-point analysis*. John Wiley & Sons.

[87] Csörgő, M. and Kulik, R. (2008a). Reduction principles for quantile and Bahadur-Kiefer processes of long-range dependent linear sequences. *Probability Theory and Related Fields*, **142**, 339-366, Springer.

[88] Csörgő, M. and Kulik, R. (2008b). Weak convergence of Vervaat and Vervaat error processes of long-range dependent sequences. *Journal of Theoretical Probability*, **21**, 672-686, Springer.

[89] Csörgő, S. and Mielniczuk, J. (1995a) Density estimation under long-range dependence. *The Annals of Statistics*, 990-999, JSTOR.

[90] Csörgő, S. and Mielniczuk, J. (1995b). Nonparametric regression under long-range dependent normal errors. *The Annals of Statistics*, 1000-1014.

[91] Csörgő, S. and Mielniczuk, J. (1995c). Close short-range dependent sums and regression estimation. *Acta Scientiarum Mathematicarum*, **60**, 177-196, Szeged [Hungary]: M. Kir. Ferencz Jozsef-Tudomanyegyetem Baratai.

[92] Csörgő, S. and Mielniczuk, J. (1996). The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probability Theory and Related Fields*, **104(1)**, 15-25, Springer.

[93] Csörgő, S. and Mielniczuk, J. (1999). Random-design regression under long-range dependent errors. *Bernoulli*, **5**, 209-224, Bernoulli Society for Mathematical Statistics and Probability.

[94] Csörgő, S. and Mielniczuk, J. (2000). The smoothing dichotomy in random-design regression with long-memory errors based on moving averages. *Statistica Sinica*, 771-787, JSTOR.

[95] Csörgő, M. and Szyszkowicz, B. (1994). Weighted multivariate empirical processes and contiguous change-point analysis. *Lecture Notes-Monograph Series*, 93-98.

[96] Csörgő, M., Szyszkowicz, B. and Wang, L. (2006). Strong invariance principles for sequential Bahadur-Kiefer and Vervaat error processes of long-range dependent sequences. *The Annals of Statistics*, **34**, 1013-1044, Institute of Mathematical Statistics.

[97] Dahlhaus, R. (1988). Small sample effects in time series analysis: a new asymptotic theory and a new estimate. *The Annals of Statistics*, 808-841, JSTOR.

[98] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 1749-1766, JSTOR.

[99] Dahlhaus, R. (1990). Approximations for the inverse of Toeplitz matrices with applications to stationary processes. *Linear Algebra and its Applications*, **127**, 27-40, Elsevier.

[100] Dahlhaus, R. (2006). Correction: Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, **34(2)**, 1045-1047, JSTOR.

[101] Davison, A.C. and Smith, R.L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, **52**, 393-425, Wiley Online Library.

[102] Davydov, Y.A. (1970). The invariance principle for stationary processes. *Theory of Probability & Its Applications*, **15**, 487-498, SIAM.

[103] de Haas-Lorentz, G.L. (1913). *Die Brownsche Bewegung und einige verwandte Erscheinungen*, Vol. 52, F. Vieweg.

[104] de Haas-Lorentz, G.L. (2013). *Die Brownsche Bewegung und einige verwandte Erscheinungen*, Springer-Verlag.

[105] Dehling, H. and Taqqu, M.S. (1988). The functional law of the iterated logarithm for the empirical process of some long-range dependent sequences. *Statistics & Probability Letters*, **7**, 81-85, Elsevier.

[106] Dehling, H. and Taqqu, M.S. (1989). The empirical process of some long-range dependent sequences with an application to U-statistics. *The Annals of Statistics*, 1767-1783, JSTOR.

[107] Dehling, H., Rooch, A. and Taqqu, M.S. (2013). Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics*, **40**, 153-173, Wiley Online Library.

[108] Deshayes, J. and Picard, D. (1985). Off-line statistical analysis of change-point models using non parametric and likelihood methods. In: *Detection of Abrupt Changes in Signals and Dynamical Systems*, 103-168, Springer.

[109] Ding, J., Tarokh, V. and Yang, Y. (2018). Optimal variable selection in regression models.

[110] Ding, J., Tarokh, V. and Yang, Y. (2017). Bridging AIC and BIC: a new criterion for autoregression. *IEEE Transactions on Information Theory*, **64(6)**, 4024-4043, IEEE.

[111] Ding, J., Tarokh, V. and Yang, Y. (2018). Model selection techniques: An overview. *IEEE Signal Processing Magazine*, **35(6)**, 16-34, IEEE.

[112] Di Marzio, M., Panzera, A. and Taylor, C.C. (2009). Local polynomial regression for circular predictors. *Statistics & Probability Letters*, **79**, 2066-2075, Elsevier.

[113] Di Marzio, M., Panzera, A. and Taylor, C.C. (2011). Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, **141**, 2156-2173, Elsevier.

[114] Di Marzio, M., Panzera, A. and Taylor, C.C. (2012). Non-parametric smoothing and prediction for nonlinear circular time series. *Journal of Time Series Analysis*, **33**, 620-630, Wiley Online Library.

[115] Di Marzio, M., Panzera, A. and Taylor, C.C. (2013). Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, **40**, 238-255, Wiley Online Library.

[116] Di Marzio, M., Panzera, A. and Taylor, C.C. (2014). Nonparametric regression for spherical data. *Journal of the American Statistical Association*, **109**, 748-763, Taylor & Francis.

[117] Di Marzio, M., Panzera, A. and Taylor, C.C. (2019). Nonparametric rotations for sphere-sphere regression. *Journal of the American Statistical Association*, **114**, 466-476, Taylor & Francis Group.

[118] Dobrushin, R.L. (1980). Gaussian random fields - Gibbsian point of view. *Multicomponent Random Systems*, **6**, 119-151, New York: Marcel Dekker.

[119] Dobrushin, R.L. and Major, P. (1979). Non-central limit theorems for non-linear functional of Gaussian fields. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **50**, 27-52, Springer.

[120] Donsker, M.D. (1951). *An invariance principle for certain probability limit theorems.*

[121] Doukhan, P. and Surgailis, D. (1998). Functional central limit theorem for the empirical process of short memory linear processes. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, **326**, 87-92, Elsevier.

[122] Doukhan, P., Massart, P. and Rio, E. (1995). Invariance principles for absolutetly regular empirical processes. *Annales de l'IHP Probabilités et Statistiques*, **31**(2), 393-427.

[123] Downs, T.D. (1974). Rotational angular correlation. In: *Biorhythyms and Human Reproduction*, 97-104, John Wiley & Sons, New York.

[124] Downs, T.D. (2003). Spherical regression. *Biometrika*, **90**, 655-668, Oxford University Press.

[125] Downs, T.D. and Mardia, K.V. (2002). Circular regression. *Biometrika*, **89**, 683-698, Oxford University Press.

[126] Dumbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *The Annals of Statistics*, **19**, 1471-1495, Institute of Mathematical Statistics.

[127] Einstein, A. (1905). Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, **4**.

[128] Embrechts, P. and Maejima, M. (2002). Selfsimilar Processes. *Applied Mathematics*, Princeton University Press.

[129] Eubank, R.L. (1988). *Nonparametric regression and spline smoothing.* Statistics: Textbooks and Monographs.

[130] Falconer, K. (2004). *Fractal geometry: mathematical foundations and applications.* John Wiley & Sons.

[131] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman and Hall, London.

[132] Fan, J. and Yao, Q. (2008). *Nonlinear time series: nonparametric and parametric methods.* Springer Science & Business Media.

[133] Feng, Y. (1999). *Kernel- and Locally Weighted Regression.* PhD thesis, Universität Konstanz.

[134] Feng, Y. (2013). An iterative plug-in algorithm for decomposing seasonal time series using the Berlin Method. *Journal of Applied Statistics*, **40**, 266-281, Taylor & Francis Group.

[135] Feng, Y. and McNeil, A.J. (2008). Modelling of scale change, periodicity and conditional heteroskedasticity in return volatility. *Economic Modelling*, **25**, 850-867, Elsevier.

[136] Fisher, N.I. (1989). Smoothing a sample of circular data. *Journal of Structural Geology*, **11**, 775-778, Elsevier.

[137] Fisher, N.I. (1995). *Statistical analysis of circular data*. Cambridge University Press.

[138] Fisher, N.I. and Lee, A.J. (1983). A correlation coefficient for circular data. *Biometrika*, **70**, 327-332, Oxford University Press.

[139] Fisher, N.I. and Lee, A.J. (1992). Regression models for an angular response. *Biometrics*, 665-677, JSTOR.

[140] Fisher, N.I. and Lee, A.J. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **56**, 327-339, Wiley Online Library.

[141] Fisher, N.I. and Lewis, T. (1983). Estimating the common mean direction of several circular or spherical distributions with differing dispersions. *Biometrika*, **70**, 333-341, Oxford University Press.

[142] Fisher, N.I., Lewis, T. and Embleton, B.J.J. (1993). *Statistical analysis of spherical data*. Cambridge University Press.

[143] Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *The Annals of Statistics*, 517-532, JSTOR.

[144] García-Portugués, E., Crujeiras, R.M. and González-Manteiga, W. (2013). Kernel density estimation for directional-linear data. *Journal of Multivariate Analysis*, **121**, 152-175, Elsevier.

[145] Gasser, T. and Müller, H.-G. (1979). *Kernel Estimation of Regression Functions*. Springer-Verlag Berlin Heidelberg.

[146] Gasser, T. and Müller, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics*, 171-185, JSTOR

[147] Gasser, T., Müller, H.-G. and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, **47**, 238-252.

[148] Gasser, T., Kneip, A. and Köhler, W. (1991). A flexible and fast method for automatic smoothing. *Journal of the American Statistical Association*, **86**, 643-652, Taylor & Francis Group.

[149] George, B.J. and Ghosh, K. (2006). A semiparametric Bayesian model for circular-linear regression. *Communications in Statistics-Simulation and Computation*, **35**, 911-923, Taylor & Francis.

[150] Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, **4**, 221-238, Wiley Online Library.

[151] Ghosh, S. (2001). Nonparametric trend estimation in replicated time series. *Journal of Statistical Planning and Inference*, **97**, 263-274, Elsevier.

[152] Ghosh, S. (2014). On local slope estimation in partial linear models under Gaussian subordination. *Journal of Statistical Planning and Inference*, **155**, 42-53, Elsevier.

[153] Ghosh, S. (2018). *Kernel Smoothing: Principles, Methods and Applications*. John Wiley & Sons.

[154] Ghysels, E. and Osborn, D.R. (2001). *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press.

[155] Giraitis, L. (1985). Central limit theorem for functionals of a linear process. *Lithuanian Mathematical Journal*, **25**, 25-35, Springer.

[156] Giraitis, L., Kokoszka, P., Leipus, R., Teyssiére, G. (2003) Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, **112(2)**, 265-294, Elsevier.

[157] Giraitis, L., Koul, H.L. and Surgailis, D. (1996a). Asymptotic normality of regression estimators with long memory errors. *Statistics & Probability Letters*, **29**, 317-335, Elsevier.

[158] Giraitis, L., Koul, H.L. and Surgailis, D. (2012) Large Sample Inference for Long-memory Processes. *World Scientific Publishing Company*.

[159] Giraitis, L. and Leipus, R. (1992). Testing and estimating in the change-point problem of the spectral function. *Lithuanian Mathematical Journal*, **32**, 15-29, Springer.

[160] Giraitis, L., Leipus, R. and Surgailis, D. (1996b). The change-point problem for dependent observations. *Journal of Statistical Planning and Inference*, **53**, 297-310, Elsevier.

[161] Giraitis, L. and Surgailis, D. (1985). CLT and other limit theorems for functionals of Gaussian processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **70**, Springer.

[162] Giraitis, L. and Surgailis, D. (1986). Multivariate Appell polynomials and the central limit theorem. In: *Dependence in Probability and Statistics*, 21-71, Springer.

[163] Giraitis, L. and Surgailis, D. (1989). Limit theorem for polynomials of a linear process with long-range dependence. *Lithuanian Mathematical Journal*, **29**, 128-145, Springer.

[164] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probability Theory and Related Fields*, **86**, 87-104, Springer.

[165] Giraitis, L. and Surgailis, D. (1999). Central limit theorem for the empirical process of a linear sequence with long memory. *Journal of Statistical Planning and Inference*, **80**, 81-93, Elsevier.

[166] Giraitis, L. and Taqqu, M.S. (1999). Whittle estimator for finite-variance non-Gaussian time series with long memory. *The Annals of Statistics*, **27**, 178-203, Institute of Mathematical Statistics.

[167] Glivenko, V. (1933). Sulla determinazione empirica delle leggi di probabilità'. *Giornale dell'Istituto Italiano degli Attuari*, **4**, 92-99.

[168] Gorodetskii, V.V. (1978). On convergence to semi-stable Gaussian processes. *Theory of Probability & Its Applications*, **22**, 498-508, SIAM.

[169] Gould, A.L. (1969). A regression technique for angular variates. *Biometrics*, 683-700, JSTOR.

[170] Granger, C.W.J. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, **73(1)**, 61-77, Elsevier.

[171] Granger, C.W.J. and Hyung, N. (1999). Occasional structural breaks and long memory. *Journal of Econometrics*, **73(1)**, 61-77, Elsevier.

[172] Granger, C.W.J. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, **11(3)**, 399-421, Elsevier.

[173] Granger, C.W.J. and Joyeux, R. (1980). An introduction to long-range time series models and fractional differencing. *Journal of Time Series Analysis*.

[174] Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*. American Mathematical Society.

[175] Grenander, U. and Szegö, G. (1958). *Toeplitz forms and their applications*. Univ of California Press.

[176] Gumbel, E.J. (1958). Statistics of Extremes. In: *Statistics of Extremes*, Columbia University Press.

[177] Gumbel, E.J. (2004). *Statistics of extremes*. Courier Corporation.

[178] Gumbel, E.J. and von Schelling, H. (1950). The distribution of the number of exceedances. *The Annals of Mathematical Statistics*, **21**, 247-262, JSTOR.

[179] Guo, H. and Koul, H.L. (2008). Asymptotic inference in some heteroscedastic regression models with long memory design and errors. *The Annals of statistics*, **36**, 458-487, Institute of Mathematical Statistics.

[180] Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). Regression Estimation and Time Series Analysis. In: *Nonparametric Curve Estimation from Time Series*,15-51, Springer.

[181] Hall, P. and Hart, J.D. (1990). Nonparametric regression with long-range dependence. *Stochastic Processes and their Applications*, **36**, 339-351.

[182] Hall, P. and Marron, J.S. (1987). Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probability Theory and Related Fields*, **74**, 567-581, Springer.

[183] Hall, P. and Marron, J.S. (1988). Choice of kernel order in density estimation. *The Annals of Statistics*, 161-173, JSTOR.

[184] Hall, P., Marron, J.S. and Park, B.U. (1992). Smoothed cross-validation. *Probability Theory and Related Fields*, **92**, 1-20, Springer.

[185] Hall, P., Lahiri, S.N. and Polzehl, J. (1995a). On bandwidth choice in nonparametric regression with both short-and long-range dependent errors. *The Annals of Statistics*, **23**, 1921-1936, Institute of Mathematical Statistics.

[186] Hall, P., Lahiri, S.N. and Truong, Y.K. (1995b). On bandwidth choice for density estimation with dependent data. *The Annals of Statistics*, **23**, 2241-2263, Institute of Mathematical Statistics.

[187] Hall, P., Watson, G.S. and Cabrera, J. (1987). Kernel density estimation with spherical data. *Biometrika*, **74**, 751-762, Oxford University Press.

[188] Hannan, E.J. (1973). Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **26**, 157-170, Springer.

[189] Hannan, E.J. (1980). The estimation of the order of an ARMA process. *The Annals of Statistics*, 1071-1081, JSTOR.

[190] Hannan, E.J. and Quinn, B.G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, **41**, 190-195, Wiley Online Library.

[191] Härdle, W. (1990). *Applied nonparametric regression*. Econometric society monographs.

[192] Härdle, W., Lütkepohl, H. and Chen, R. (1997). A review of nonparametric time series analysis. *International Statistical Review*, **65**(1), 49-72, Wiley Online Library.

[193] Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*. Springer Series in Statistics.

[194] Hart, J.D. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society*, **53**, 173-187.

[195] Hart, J.D. and Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *The Annals of Statistics*, 873-890, JSTOR.

[196] Hastie, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry. *Statistical Science*, 120-129, JSTOR.

[197] Haye, M.O., Philippe, A. and Robet, C. (2019). Inference for continuous-time long memory randomly sampled processes.

[198] Hawkins, D.L. (1988). Retrospective and sequential tests for a change in distribution based on kolmogorov-smirnov-type statistics: A change in distribution based on Kolmogorov-Smirnov-type statistics. *Sequential Analysis*, **7**, 23-51, Taylor & Francis Group.

[199] Heidenreich, N.-B., Schindler, A. and Sperlich, S. (2013). Bandwidth selection for kernel density estimation: a review of fully automatic selectors. *AStA - Advances in Statistical Analysis*, **97**, 403-433, Springer.

[200] Heiler, S. (1970). Theoretische Grundlagen des "'Berliner Verfahrens"'. In: *Neuere Entwicklungen auf dem Gebiet der Zeitreihenanalyse*, Sonderheft, **1**, 67-93.

[201] Heiler, S. and Feng, Y. (2000). Data-driven decomposition of seasonal time series. *Journal of Statistical Planning and Inference*, **91**, 351-363.

[202] Hermann, E. and Gasser, T. (1994). Iterative plug-in algorithm for bandwidth selection in kernel regression estimation. *Discussion Paper*.

[203] Hermann, E., Gasser, T. and Kneip, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika*, **79**, 783-795.

[204] Hidalgo, J. (1997). Non-parametric estimation with strongly dependent multivariate time series. *Journal of Time Series Analysis*, **18**, 95-122, Wiley Online Library.

[205] Ho, H.-C. (1996). On central and non-central limit theorems in density estimation for sequences of long-range dependence. *Stochastic processes and their Applications*, **63**, 153-174, Elsevier.

[206] Ho, H.-C. and Hsing, T. (1996). On the asymptotic joint distribution of the sum and maximum of stationary normal random variables. *Journal of applied probability*, 138-145, JSTOR.

[207] Ho, H.-C. and Hsing, T. (1997). Limit theorems for functionals of moving averages. *The Annals of Probability*, **25**, 1636-1669, Institute of Mathematical Statistics.

[208] Ho, H.-C. and Sun, T.-C. (1987). A central limit theorem for non-instantaneous filters of a stationary Gaussian process. *Journal of Multivariate Analysis*, **22**, 144-155, Elsevier.

[209] Ho, H.-C. and Sun, T.-C. (1990). Limiting distributions of nonlinear vector functions of stationary Gaussian processes. *The Annals of Probability*, **18**, 1159-1173, Institute of Mathematical Statistics.

[210] Honda, T. (2000a). Nonparametric density estimation for a long-range dependent linear process. *The Annals of the Institute of Statistical Mathematics*, **52**, 599-611, Springer.

[211] Honda, T. (2000b). Nonparametric estimation of the conditional median function for long-range dependent processes. *Journal of the Japan Statistical Society*, **30**, 129-142, The Japan Statistical Society.

[212] Horváth, L., Kokoszka, P. and Steinebach, J. (1999). Testing for changes in multivariate dependent oberbations with an application to temperature changes. *Journal of Multivariate Analysis*, **68(1)**, 96-119, Elsevier.

[213] Horváth, L. and Shao, Q.-M. (1999). Limit theorems for quadratic forms with applications to Whittle's estimate. *The Annals of Applied Probability*, 146-187, JSTOR.

[214] Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, **68**, 165-176.

[215] Hsing, T. (2000). Linear processes, long-range dependence and asymptotic expansions. *Statistical inference for stochastic processes*, **3**, 19-29, Springer.

[216] Hurst, H.E. (1951). Long-term Storage Capacity of Reservoirs. *American Society of Civil Engineers*.

[217] Hurst, H.E. (1965). *Long term storage: An experimental study*.

[218] Hutchinson, J.E. (1981). Fractals and self similarity. *Indiana University Mathematics Journal*, **30**, 713-747, JSTOR.

[219] Hylleberg, S. (1986). Seasonality in regression. *Academic Press*, Incorporated.

[220] Itô, K. (1951). Multiple wiener integral. *Journal of the Mathematical Society of Japan*, **3**, 157-169, The Mathematical Society of Japan.

[221] Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

[222] Jammalamadaka, S.R. and Lund, U.J. (2006). The effect of wind direction on ozone levels: a case study. *Environmental and Ecological Statistics*, **13**, 287-298, Springer.

[223] Jammalamadaka, S.R. and Sarma, Y.R. (1988). A correlation coefficient for angular variables. *Statistical theory and data analysis II*, 349-364, Amsterdam North Holland.

[224] Jammalamadaka, S.R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific.

[225] Johnson, R.A. and Wehrly, T.E. (1977). Measures and models for angular correlation and angular-linear correlation. *Journal of the Royal Statistical Society: Series B (Methodological)*, **39**, 222-229, Wiley Online Library.

[226] Johnson, R.A. and Wehrly, T.E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, **73**, 602-606, Taylor & Francis Group..

[227] Johnstone, I.M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society: series B (statistical methodology)*, **59**, 319-351, Wiley Online Library.

[228] Jones, M.C. (1993). Simple boundary correction for kernel density estimation. *Statistics and computing*, **3**, 135-146, Springer.

[229] Jones, M.C. (1995). On higher order kernels. *Journal of Nonparametric Statistics*, **5**, 215-221, Taylor & Francis Group.

[230] Jones, M.C., Marron, J.S. and Park, B.U. (1991). A simple root n bandwidth selector. *The Annals of Statistics*, 1919-1932, JSTOR.

[231] Jones, M.C., Marron, J.S. and Sheather, S.J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American statistical association*, **91**, 401-407, Taylor & Francis Group.

[232] Jones, M.C. and Pewsey, A. (2005). A family of symmetric distributions on the circle. *Journal of the American Statistical Association*, **100(472)**, 1422-1428, Taylor & Francis.

[233] Jones, M.C. and Pewsey, A. (2012). Inverse Batschelet distributions for circular data. *Biometrics*, **68**, 183-193, Wiley Online Library.

[234] Jupp, P.E. and Mardia, K.V. (1980). A general correlation coefficient for directional data and related regression problems. *Biometrika*, **67**, 163-173, Oxford University Press.

[235] Karamata, J. (1930a). Sur certains "Tauberian Theorems" de M.M. Hardy et Littlewood. *Mathematica (Cluj)*, **3**, 33-48.

[236] Karamata, J. (1930b). Sur un mode de croissance rigulilre des fonctions. *Mathematica (Cluj)*, **4**, 38-53.

[237] Karamata, J. (1933). Sur un mode de croissance régulière. Théorèmes fondamentaux. *Bulletin de la Société Mathématique de France*, **61**, 55-62.

[238] Kato, S. (2010). A Markov process for circular data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **72**, 655-672, Wiley Online Library.

[239] Kato, S. and Jones, M.C. (2010). A family of distributions on the circle with links to, and applications arising from, Möbius transformation. *Journal of the American Statistical Association*, **105**, 249-262, Taylor & Francis Group.

[240] Kato, S. and Jones, M.C. (2013). An extended family of circular distributions related to wrapped Cauchy distributions via Brownian motion. *Bernoulli*, **19**, 154-171, Bernoulli Society for Mathematical Statistics and Probability.

[241] Kato, S., Shimizu, K. and Shieh, G.S. (2008). A circular-circular regression model. *Statistica Sinica*, 633-645, JSTOR.

[242] Kiefer, J. (1972). Skorohod embedding of multivariate random variables and the sample distribution function. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **24**, 1-35, Springer.

[243] Kim, S. and SenGupta, A. (2016). Regressions involving circular variables: An overview. In: *Platinum Jubilee International Conference on Applications of Statistics*, 25-33, Springer.

[244] Kim, S. and SenGupta, A. (2017). Multivariate-multiple circular regression. *Journal of Statistical Computation and Simulation*, **87**, 1277-1291, Taylor & Francis Group.

[245] Kitagawa, G. (2020). *Introduction to Time Series Modeling with Applications in R.* CRC Press.

[246] Köhler, M., Schindler, A. and Sperlich, S. (2014). A review and comparison of bandwidth selection methods for kernel regression. *International Statistical Review*, **82**, 243-274, Wiley Online Library.

[247] Kokoszka, P. and Leipus, R. (1998). Change-point in the mean of dependent observations. *Statistics & Probability Letters*, **40**, 385-393, Elsevier.

[248] Kolmogorov, A.N. (1941). The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. In: *Dokl. Akad. Nauk SSSR*, **30**, 299-303.

[249] Koul, H.L., Baillie, R.T. and Surgailis, D. (2004). Regression model fitting with a long memory covariate process. *Econometric Theory*, 485-512, JSTOR.

[250] Kulik, R. and Lorek, P. (2011). Some results on random design regression with long memory errors and predictors. *Journal of statistical planning and inference*, **141**, 508-523, Elsevier.

[251] Kulik, R. and Wichelhaus, C. (2012). Conditional variance estimation in regression models with long memory. *Journal of Time Series Analysis*, **33**, 468-483, Wiley Online Library.

[252] Künsch, H., Beran, J. and Hampel, F. (1993). Contrasts under long-range correlations. *The Annals of Statistics*, 943-964, JSTOR.

[253] Lang, G. and Soulier, P. (2000). Convergence de mesures spectrales aléatoires et applications à des principes d'invariance. *Statistical Inference for Stochastic Processes*, **3**, 41-51, Springer.

[254] Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, **21(1)**, 33-59, Wiley Online Library.

[255] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

[256] Leadbetter, M.R. and Rootzén, H. (1988). Extremal theory for stochastic processes. *The Annals of Probability*, 431-478, JSTOR.

[257] Lee, A. (2010a). Circular data. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 477-486, Wiley Online Library.

[258] Leipus, R. (1988) Weak convergence of two-parameter empirical fields in change-point problems. *Lithuanian Mathematical Journal*, **28**, 348-352, Springer.

[259] Leipus, R. and Viano, M.-C. (2000). Modelling long-memory time series with finite or infinite variance: a general approach. Journal of Time Series Analysis, **21**, 61-74, Wiley Online Library.

[260] Lejeune, M. (1985). Estimation non-paramétrique par noyaux: régression polynomiale mobile. *Revue de Statistique Appliquée*, **33**, 43-67.

[261] Lejeune, M. and Sarda, P. (1992). Smooth estimators of distribution and density functions. *Computational Statistics & Data Analysis*, **14**, 457-471.

[262] Leonenko, N. and Ruiz-Medina, M.D. (2022). Sojourn functionals for spatiotemporal Gaussian fields with long memory. *Journal of Applied Probability*, Applied Probability Trust.

[263] Lévy, P. (1939). L'addition des variables aléatoires définies sur une circonférence. *Bulletin de la Société mathématique de France*, **67**, 1-41.

[264] Ley, C. and Verdebout, T. (2014). Simple optimal tests for circular reflective symmetry about a specified median direction. *Statististica Sinica*, 1319-1339, JSTOR.

[265] Ley, C. and Verdebout, T. (2017). *Modern Directional Statistics*. CRC Press.

[266] Ley, C. and Verdebout, T. (2018). *Applied Directional Statistics: Modern Methods and Case Studies*. CRC Press.

[267] Lindvall, T. (1973). Weak convergence of probability measures and random functions in the function space $D[0, \infty)$. *Journal of Applied Probability*, 109-121, JSTOR.

[268] Major, P. (1981). Multiple Wiener-Itô Integrals. *Lecture Notes in Mathematics*, **849**, Springer Berlin.

[269] Mandelbrot, B.B. (1965). Une classe de processus stochastiques homothétiques à soi; application à la loi climatologique de H.E. Hurst, 260: 3274-3277 (1965). *Comptes Rendus de L'Académie des Sciences de Paris*, **260**, 3274-3277.

[270] Mandelbrot, B.B. (1972). Statistical methodology for nonperiodic cycles: from the covariance to R/S analysis. In: *Annals of Economic and Social Measurement*, **1**, 259-290, NBER.

[271] Mandelbrot, B.B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **31**, 271-285, Springer.

[272] Mandelbrot, B.B. (1982). *The Fractal Geometry of Nature.* W.H. Freeman & Company, New York.

[273] Mandelbrot, B.B. (1989). Fractal geometry: what is it, and what does it do?. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, **423**, 3-16, The Royal Society London.

[274] Mandelbrot, B.B. (2013a). *Fractals and Chaos: The Mandelbrot Set and Beyond.* Springer Science & Business Media.

[275] Mandelbrot, B.B. (2013b). *Fractals and scaling in finance: Discontinuity, concentration, risk. Selecta volume E.* Springer Science & Business Media.

[276] Mandelbrot, B.B. and McCamy, K. (1970). On the secular pole motion and the Chandler wobble. *Geophysical Journal International*, **21**, 217-232, Blackwell Publishing Ltd Oxford, UK.

[277] Mandelbrot, B.B. and Taqqu, M.S. (1979). *Robust R/S analysis of long run serial correlation.* IBM Thomas J. Watson Research Division.

[278] Mandelbrot, B.B. and van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM review*, **10**, 422-437, SIAM.

[279] Mandelbrot, B.B. and Wallis, J.R. (1968). Noah, Joseph, and operational hydrology. *Water Resources Research*, **4**, 909-918, Wiley Online Library.

[280] Mandelbrot, B.B. and Wallis, J.R. (1969a). Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research*, **5**, 967-988, Wiley Online Library.

[281] Mandelbrot, B.B. and Wallis, J.R. (1969b). Some long-run properties of geophysical records. *Water Resources Research*, **5**, 321-340, Wiley Online Library.

[282] Mandelbrot, B.B. and Wallis, J.R. (1969c). Computer experiments with fractional Gaussian noises: Part 1, Averages and Variances. *Water Resources Research*, **5**, 228-241, Wiley Online Library.

[283] Mandelbrot, B.B. and Wallis, J.R. (1969d). Computer experiments with fractional Gaussian noises: Part 2, Rescaled ranges and Spectra. *Water Resources Research*, **5**, 242-259, Wiley Online Library.

[284] Mandelbrot, B.B. and Wallis, J.R. (1969e). Computer experiments with fractional Gaussian noises: Part 3, Mathematical Appendix. *Water Resources Research*, **5**, 260-267, Wiley Online Library.

[285] Mardia, K.V. (1972). Statistics of Directional Data. *Academic Press*, **357**.

[286] Mardia, K.V. (1975). Statistics of Directional Data. *Journal of the Royal Statistical Society: Series B (Methodological)*, **37**, 349-371, Wiley Online Library.

[287] Mardia, K.V. (2014). *Statistics of Directional Data*. Academic Press.

[288] Mardia, K.V. and Jupp, P.E. (1999). *Directional statistics*. Wiley.

[289] Mardia, K.V. and Jupp, P.E. (2009). *Directional statistics*. John Wiley & Sons.

[290] Mardia, K.V. and Puri, M.L. (1978). A spherical correlation coefficient robust against scale. *Biometrika*, **65**, 391-395, Oxford University Press.

[291] Marinucci, D. (2005). The empirical process for bivariate sequences with long memory. *Statistical Inference for Stochastic Processes*, **8**, 205-223, Springer.

[292] Marron, J.S. and Nolan, D. (1988). Canonical kernels for density estimation. *Statistics & Probability Letters*, **7**, 195-199, Elsevier.

[293] McCullagh, P. (1992). Conditional inference and Cauchy models. *Biometrika*, **79**, 247-259, Oxford University Press.

[294] McCullagh, P. (1996). Möbius transformation and Cauchy parameter estimation. *The Annals of Statistics*, **24**, 787-808, Institute of Mathematical Statistics.

[295] Meilán-Vila, A., Francisco-Fernández, M., Crujeiras, R.M. and Panzera, A. (2019). Nonparametric Regression Estimation for Circular Data. In: *Multidisciplinary Digital Publishing Institute Proceedings*, **21**.

[296] Meilán-Vila, A., Francisco-Fernández, M., Crujeiras, R.M. and Panzera, A. (2021). Nonparametric multiple regression estimation for circular response. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, **30(3)** 650-672, Springer.

[297] Menéndez, P., Ghosh, S., Künsch, H.R. and Tinner, W. (2013). On trend estimation under monotone Gaussian subordination with long-memory: application to fossil pollen series. *Journal of Nonparametric Statistics*, **25**, 765-785, Taylor & Francis Group.

[298] Mielniczuk, J. and Wu, W.B. (2004). On random-design model with dependent errors. *Statistica Sinica*, 1105-1126, JSTOR.

[299] Mishura, Y.S. (2008). Wiener integration with respect to fractional brownian motion. *Stochastic Calculus for Fractional Brownian Motion and Related Processes*, 1-121, Springer.

[300] Modlin, D., Fuentes, M. and Reich, B. (2012). Circular conditional autoregressive modeling of vector fields. *Environmetrics*, **23**, 46-53, Wiley Online Library.

[301] Morellato, L.P.C., Alberti, L..F and Hudson, I.L. (2010). Applications of circular statistics in plant phenology: a case studies approach. In: *Phenological Research*, 339-359, Springer.

[302] Moulines, E. and Soulier, P (2003). Semiparametric spectral estimation for fractional processes. *Theory and Applications of Long-Range Dependence*, 251-301.

[303] Müller, H.-G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics*, 766-774, JSTOR.

[304] Müller, H.-G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting. *Journal of the American Statistical Association*, **82**, 231-238.

[305] Müller, H.-G. (1988). *Nonparametric Analysis of Longitudinal Data*. Springer-Verlag Berlin Heidelberg.

[306] Müller, H.-G. (1991). Smooth optimum kernel estimators near endpoints. *Biometrika*, **78**, 521-530, Oxford University Press.

[307] Müller, H.-G. and Wang, J.-L. (1994). Hazard rate estimation under random censoring with varying kernels and bandwidths. *Biometrics*, 61-76, JSTOR.

[308] Nadaraya, E.A. (1964). On estimating regression. *Theory of Probability & Its Applications*, **9**, 141-142, SIAM.

[309] Nordin, C.F. and Rosbjerg, D.M. (1970). Applications of crossing theory in hydrology. *Hydrological Sciences Journal*, **15**, 27-43, Taylor & Francis Group.

[310] Oliveira, M., Crujeiras, R.M. and Rodríguez-Casal, A. (2012). A plug-in rule for bandwidth selection in circular density estimation. *Computational Statistics & Data Analysis*, **56**, 3898-3908, Elsevier.

[311] Oliveira, M., Crujeiras, R.M. and Rodríguez-Casal, A. (2013). Nonparametric circular methods for exploring environmental data. *Environmental and Ecological Statistics*, **20(1)**, 1-17, Springer.

[312] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, **33**, 1065-1076, JSTOR.

[313] Parzen, E. (1974). Some recent advances in time series modeling. *IEEE Transactions on Automatic Control*, **19**, 723-730, IEEE.

[314] Pearson, K. (1905). The problem of the random walk. *Nature*, **72**, 342-342, Nature Publishing Group.

[315] Perron, P. (2006). Dealing with structural breaks. *Palgrave Handbook of Econometrics*, **1(2)**, 278-352.

[316] Pewsey, A. (2000). The wrapped skew-normal distribution on the circle. *Communications in Statistics-Theory and Methods*, **29(11)**, 2459-2472, Taylor & Francis.

[317] Pewsey, A. (2002). Testing circular symmetry. *Canadian Journal of Statistics*, **30(4)**, 591-600, Wiley Online Library.

[318] Pewsey, A. (2004). Testing for circular reflective symmetry about a known median axis . *Journal of Applied Statistics*, **31(5)**, 575-585, Taylor & Francis.

[319] Pewsey, A. and García-Portugués, E. (2021). Recent advances in directional statistics. *TEST: An Official Journal of the Spanish Society of Statistics and Operations Research*, **30(1)**, 1-58, Springer.

[320] Pewsey, A., Neuhäuser, M. and Ruxton, G.D. (2013). *Circular Statistics in R.* Oxford University Press.

[321] Picard, D. (1985). Testing and estimating change-points in time series. *Advances in Applied Probability*, 841-867, JSTOR.

[322] Pipiras, V. and Taqqu, M.S. (2000a). Convergence of weighted sums of random variables with long-range dependence. *Stochastic Processes and their Applications*, **90**, 157-174, Elsevier.

[323] Pipiras, V. and Taqqu, M.S. (2000b). Integration questions related to fractional Brownian motion. *Probability Theory and Related Fields*, **118**, 251-291, Springer.

[324] Pipiras, V. and Taqqu, M.S. (2003). Fractional calculus and its connections to fractional Brownian motion. *Theory and Applications of Long-Range Dependence*, 165-201, Birkhäuser.

[325] Pipiras, V. and Taqqu, M.S. (2017). *Long-range dependence and self-similarity.* Cambridge University Press.

[326] Polsen, O. and Taylor, C.C. (2015). Parametric circular-circular regression and diagnostic analysis. *Geometry Driven Statistics*, **121**, 115-128, Wiley Online Library.

[327] Presnell, B., Morrison, S.P. and Littell, R.C. (1998). Projected multivariate linear models for directional data. *Journal of the American Statistical Association*, **93**, 1068-1077, Taylor & Francis Group.

[328] Priestley, M.B. (1981). *Spectral analysis and time series: probability and mathematical statistics.* New York Academic Press.

[329] Priestley, M.B. and Chao, M.T. (1972). Nonparametric function fitting. *Journal of the Royal Statistical Society: Series B (Methodological).*

[330] Prokhorov, Y.V. (1956). Convergence of random processes and limit theorems in probability theory. *Theory of Probability & Its Applications*, **1**, 157-214, SIAM.

[331] Qin, X., Zhang, J.-S. and Yan, X.-D. (2011). Local linear least squares kernel regression for linear and circular predictors. *Communications in Statistics - Theory and Methods*, **40**, 3812-3823, Taylor & Francis Group.

[332] Quinn, B.G. (1980). Order determination for a multivariate autoregression. *Journal of the Royal Statistical Society: Series B (Methodological)*, **42**, 182-185, Wiley Online Library.

[333] Rao, S.S. and Yang, J. (2021). Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. *The Annals of Statistics*, **49(5)**, 2774-2802, Institute of Mathematical Statistics.

[334] Ray, B.K. and Tsay, R.S. (1997). Bandwidth selection for kernel regression with long-range dependent errors. *Biometrika*, **84**, 791-802.

[335] Ray, B.K. and Tsay, R.S. (2002). Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis*, **23**, 687-705, Wiley Online Library.

[336] Rayleigh, L. (1880). On the resultant of a large number of vibrations of the same pitch and of arbitrary phase. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **10**, 73-78, Taylor & Francis Group.

[337] Rivest, L.-P., Duchesne, T., Nicosia, A. and Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 445-463, JSTOR.

[338] Robert, C. (1990). Modified Bessel functions and their applications in probability and statistics. *Statistics & Probability Letters*, **9**, 155-161, Elsevier.

[339] Robinson, P.M. (1989). Nonparametric estimation of time-varying parameters. In: *Statistical Analysis and Forecasting of Economic Structural Change*, 253-264, Springer.

[340] Robinson, P.M. (1995a). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, **23**, 1630-1661, Institute of Mathematical Statistics.

[341] Robinson, P.M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 1048-1072, JSTOR.

[342] Robinson, P.M. (1997). Large-sample inference for nonparametric regression with dependent errors. *The Annals of Statistics*, **25**, 2054-2083, Institute of Mathematical Statistics.

[343] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, **42**, National Academy of Sciences.

[344] Rosenblatt, M. (1961). Independence and dependence. In: *Proc. 4th Berkeley Symposium of Mathematical Statistics and Probability*, **2**, 431-443.

[345] Rosenblatt, M. (1979). Some limit theorems for partial sums of quadratic forms in stationary Gaussian variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **49**, 125-132, Springer.

[346] Ruppert, D. and Wand, M.P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics*, **22**, 1346-1370.

[347] Sarkadi, K. (1957). On the distribution of the number of exceedances. *The Annals of Mathematical Statistics*, **28**, 1021-1023, JSTOR.

[348] Schach, S. (1969). Nonparametric symmetry tests for circular distributions . *Biometrika*, **56(3)**, 571-577, Oxford University Press.

[349] Schulgasser, K. (1985). Fibre orientation in machine-made paper. *Journal of Materials Science*, **20**, 859-866, Springer.

[350] Schützner, M. (2006). *Appell-Polynome und Grenzwertsätze*. Diploma thesis, Universität Konstanz.

[351] Schützner, M. (2009). *Asymptotic statistical theory for long memory volatility models*. PhD thesis, Universität Konstanz.

[352] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464, Institute of Mathematical Statistics.

[353] Scott, D.W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

[354] SenGupta, A. and Ugwuowo, F.I. (2006). Asymmetric circular-linear multivariate regression models with applications to environmental data. *Environmental and Ecological Statistics*, **13**, 299-309, Springer.

[355] Shaman, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *The Annals of Statistics*, **3(2)**, 532-538, Institute of Mathematical Statistics.

[356] Shaman, P. (1976). Approximations for stationary covariance matrices and their inverses with application to ARMA models. *The Annals of Statistics*, 292-301, JSTOR.

[357] Shao, Q.-M. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability*, 2098-2127, JSTOR.

[358] Shao, X. (2010). Nonstationary-extended Whittle estimation. *Econometric Theory*, **26(4)**, 1060-1087, Cambridge University Press.

[359] Shao, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis*, **32(6)**, 598-606, Wiley Online Library.

[360] Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika*, **63**, 117-126, Oxford University Press.

[361] Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics*, 147-164, JSTOR.

[362] Shimotsu, K. (2005). Simple (but effective) tests of long memory versus structural breaks. *Queen's Economics Department Working Paper*.

[363] Shimotsu, K. and Phillips, P.C.B. (2005). Exact local Whittle estimation of fractional integration. *The Annals of Statistics*, **33(4)**, 1890-1933, Institute of Mathematical Statistics.

[364] Shorack, G.R. and Wellner, J.A. (2009). *Empirical Processes with Applications to Statistics*. SIAM.

[365] Sibbertsen, P. (2004). Long memory versus structural breaks: An overview. *Statistical Papers*, **45(4)**, 465-515, Springer.

[366] Sibbertsen, P., Wenger, K. and Wingert, S. (2020). Testing for multiple structural breaks in multivariate long memory time series. *Hannover Economic Papers (HEP)*.

[367] Silverman, B.W. (1978). Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *The Annals of Statistics*, 177-184, JSTOR.

[368] Silverman, B.W. (1986). *Density estimation for statistics and data analysis*. CRC Press.

[369] Skorokhod, A.V. (1956). Limit theorems for stochastic processes. *Theory of Probability & Its Applications*, **1**, 261-290, SIAM.

[370] Spurr, B.D. and Koutbeiy, M.A. (1991). A comparison of various methods for estimating the parameters in mixtures of von Mises distributions . *Communications in Statistics-Simulation and Computation*, **20(2-3)**, 725-741, Taylor & Francis.

[371] Stephens, M.A. (1963). Random walk on a circle. *Biometrika*, **50**, 385-390, JSTOR.

[372] Stone, C.J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, 595-620, JSTOR.

[373] Surgailis, D. (1981). Convergence of sums of nonlinear functions of moving averages to selfsimilar processes. In: *Doklady Akademii Nauk*, **257**, 51-54, Russian Academy of Sciences.

[374] Surgailis, D. (1982). Zones of attraction of self-similar multiple integrals. *Lithuanian Mathematical Journal*, **22**, 327-340, Springer.

[375] Surgailis, D. (2000). Long-range dependence and Appell rank. *The Annals of Probability*, **28**, 478-497, Institute of Mathematical Statistics.

[376] Surgailis, D. (2003). CLTs for polynomials of linear sequences: Diagram formula with illustrations. *Theory and Applications of Long-Range Dependence*, 111-128, Birkhäuser Boston, Berlin.

[377] Surgailis, D. and Vaičiulis, M. (1999). Convergence of Appell polynomials of long range dependent moving averages in martingale differences. *Acta Applicandae Mathematica*, **58**, 343-357, Springer.

[378] Sykulski, A.M., Olhede, S.C., Guillaumin, A.P., Lilly, J.M. and Early, J.J. (2019). The debiased Whittle likelihood. *Biometrika*, **106(2)**, 251-266, Oxford University Press.

[379] Szyszkowicz, B. (1994). Weak convergence of weighted empirical type processes under contiguous and changepoint alternatives. *Stochastic Processes and their Applications*, **50**, 281-313, Elsevier.

[380] Taniguchi, M., Kato, S., Ogata, H. and Pewsey, A. (2020). Models for circular data from time series spectra. *Journal of Time Series Analysis*, **41**, 808-829, Wiley Online Library.

[381] Taqqu, M.S. (1970). Note on evaluations of R/S for fractional noises and geophysical records. *Water Resources Research*, **6**, 349-350, Wiley Online Library.

[382] Taqqu, M.S.(1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **31**, 287-302, Springer.

[383] Taqqu, M.S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **40**, 203-238, Springer.

[384] Taqqu, M.S. (1978). A representation for self-similar processes. *Stochastic Processes and their Applications*, **7**, 55-64, Elsevier.

[385] Taqqu, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **50**, 53-83, Springer.

[386] Taqqu, M.S. (1986). A bibliographical guide to self-similar processes and long-range dependence. In: *Dependence in Probability and Statistics*, 137-162, Springer.

[387] Taylor, C.C. (2008). Automatic bandwidth selection for circular density estimation. *Computational Statistics & Data Analysis*, **52**, 3493-3500, Elsevier.

[388] Terrell, G.R. and Scott, D.W. (1980). On improving convergence rates for nonnegative kernel density estimators. *The Annals of Statistics*, 1160-1163, JSTOR.

[389] Tewes, J. (2017). Change-point tests under local alternatives for long-range dependent processes. *Electronic Journal of Statistics*, **11(1)**, 2461-2498, Institute of Mathematical Statistics and Bernoulli Society.

[390] Tewes, J. (2018). Block bootstrap for the empirical process of long-range dependent data. *Journal of Time Series Analysis*, **39**, 28-53, Wiley Online Library.

[391] Tsuruta, Y. and Sagae, M. (2017). Higher order kernel density estimation on the circle. *Statistics & Probability Letters*, **131**, 46-50, Elsevier.

[392] Tsuruta, Y. and Sagae, M. (2020). Theoretical properties of bandwidth selectors for kernel density estimation on the circle. *Annals of the Institute of Statistical Mathematics*, **72**, 511-530, Springer.

[393] Umbach, D. and Jammalamadaka, S.R. (2009). Building asymmetry into circular distributions. *Statistics & Probability Letters*, **79(5)**, 659-663, Elsevier.

[394] Umbach, D. and Jammalamadaka, S.R. (2010). Some moment properties of skew-symmetric circular distributions. *Metron*, **68(3)**, 265-273, Springer.

[395] Upton, G.J.G. and Fingleton, B. (1985). *Spatial data analysis by example: categorical and directional data*. Wiley.

[396] van der Vaart, A.W. and Wellner, J.A. (1996). Weak convergence. In: *Weak convergence and empirical processes*, 16-28, Springer.

[397] Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, **20(1)**, 87-127, Wiley Online Library.

[398] Vogelsang, T.J. (1998). Trend function hypothesis testing in the presence of serial correlation. *Econometrica*, 123-148, JSTOR.

[399] Vogelsang, T.J. (1999). *Testing for a shift in trend when serial correlation is of unknown form*. Tinbergen Institute.

[400] von Mises, R.V. (1918). Über Kurven gleichmäßigster Krümmung. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, **27**, 67-71.

[401] von Mises, R.V. (1981). Über die "Ganzzahligkeit" der Atomgewichte und verwandte Fragen. *Physikalische Zeitschrift*, **19**, 490-500.

[402] von Smoluchowski, M. (1906). Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Annalen der Physik*, **326**, 756-780, Wiley Online Library.

[403] Wand, M.P. and Jones, M.C. (1994). *Kernel Smoothing*. CRC Press.

[404] Wang, L. (2007). Change-point estimation in long memory nonparametric models with applications. *Communications in Statistics - Simulation and Computation*, **37**, 48-61, Taylor & Francis Group.

[405] Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *The Annals of Statistics*, **24**, 466-484, Institute of Mathematical Statistics.

[406] Wang, L. (2008). Change-point detection with rank statistics in long-memory time-series models. *Australian & New Zealand Journal of Statistics*, **50**, 241-256, Wiley Online Library.

[407] Wang, F. and Gelfand, A.E. (2014). Modeling space and space-time directional data using projected Gaussian processes. *Journal of the American Statistical Association*, **109**, 1565-1580, Taylor & Francis Group.

[408] Wang, J., Guan, Y., Wu, L., Guan, X., Cai, W., Huang, J., Dong, W., Zhang, B. (2021). Changing lengths of the four seasons by global warming. *Geophysical Research Letters*, **48(6)**, Wiley Online Library.

[409] Wang, Q., Lin, Y.-X. and Gulati, C.M. (2003). Strong approximation for long memory processes with applications. *Journal of Theoretical Probability*, **16**, 377-389, Springer.

[410] Watson, G.S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359-372, JSTOR.

[411] Watson, G.S. (1983). *Statistics on spheres*. Wiley-Interscience.

[412] Watson, G.S. and Williams, E.J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika*, **43**, 344-352, JSTOR.

[413] Wehrly, T.E. and Johnson, R.A. (1980). Bivariate models for dependence of angular observations and a related Markov process. *Biometrika*, **67**, 255-256, Oxford University Press.

[414] Wenger, K., Leschinski, C. and Sibbertsen, P. (2019). Change-in-mean tests in long-memory time series: a review of recent developments. *AStA - Advances in Statistical Analysis*, **103(2)**, 237-256, Springer.

[415] Whittle, P. (1951). *Hypothesis testing in time series analysis*. Almqvist & Wiksells boktr., **4**.

[416] Whittle, P. (1953). Estimation and information in stationary time series. *Arkiv för matematik*, **2**, 423-434, Springer.

[417] Wintner, A. (1947). On the shape of the angular case of Cauchy's distribution curves. *The Annals of Mathematical Statistics*, **18**, 589-593, JSTOR.

[418] Wold, H. (1954). Causality and econometrics. *Econometrica: Journal of the Econometric Society*, 162-177, JSTOR.

[419] Wu, W.B. (2003). Empirical processes of long-memory sequences. *Bernoulli*, **9**, 809-831, Bernoulli Society for Mathematical Statistics and Probability.

[420] Wu, W.B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *The Annals of Statistics*, **30**, 1441-1459, Institute of Mathematical Statistics.

[421] Wu, B.W. and Woodroofe, M. (2002). *Martingale approximation for sums of stationary processes*. University of Chicago, Dept. of Statistics.

[422] Yajima, Y. (1985). On estimation of long-memory time series models. *Australian Journal of Statistics*, **27**, 303-320, Wiley Online Library.

[423] Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *The Annals of Statistics*, **16**, 791-807.

[424] Yajima, Y. (1991). Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *The Annals of Statistics*, **19**, 158-177.

[425] Yao, Y.-C. (1987). Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables. *The Annals of Statistics*, **15**, 1321-1328, Institute of Mathematical Statistics.

[426] Zhao, Z. and Wu, W.B. (2008). Confidence bands in nonparametric time series regression. *The Annals of Statistics*, **36**, 1854-1878, Institute of Mathematical Statistics.

[427] Zygmund, A. (1968). *Trigonometric series*. Cambridge University Press.

[428] Zygmund, A. (2002). *Trigonometric series*. Cambridge University Press.