INAUGURAL-DISSERTATION

zur
Erlangung der Doktorwürde
der
Gesamtfakultät für Mathematik, Ingenieur- und
Naturwissenschaften
der
Ruprecht - Karls - Universität
Heidelberg

vorgelegt von

Nathawut Phandoidaen
M.Sc. Mathematik

aus Chiangmai, Thailand

Datum der mündlichen Prüfung:

Statistical analysis of dependent data

Statistical analysis of dependent data

Topics in empirical process theory with applications to neural networks, and survival analysis

Betreuer: Prof. Dr. Rainer Dahlhaus

Acknowledgments

I would like to express my gratitude towards my supervisor Prof. Dr. Rainer Dahlhaus for giving me the opportunity to broaden my knowledge in the field of Probability Theory and Statistics. His lectures and expertise were of fundamental value and were greatly appreciated. On the same note I would also like to thank my co-supervisor Prof. Dr. Leif Döring as part of the Research Training Group 1953.

My sincerest thanks also goes to Dr. Stefan Richter with whom I had the pleasure to collaborate. He provided me with inspiring ideas, believed in my abilities and encouraged me to stay on track whenever I was in doubt. My work would not have been as it is without his input.

Thank you Sergio Filipe Brenner Miguel for a wonderful time. I am glad to have worked with you. Our research findings enhance this thesis.

The four years of my graduate studies have been delightful. Thank you Alexander Kreiß, Christof Schötz, Joseph Meyer, Marilena Müller, Sandra Schluttenhofer, Xavier Loizeau for all the fun activities and the enlightening conversations during break times and outside of office hours.

I would like to give kind regards to Elke Carlow, Sanja Juric, Dagmar Neubauer, Marion Münster and Jutta Wiech, as well, for handling all the administrative work.

To my family and my husband: Thank you for supporting me throughout my life and through the hardships of a PhD, especially during the final stretches of writing up this thesis. Thank you Eric, for being there with a cup of coffee and lending me your ears.

I am grateful to the administration of the *Faculty of Natural Sciences and Mathematics* of the Ruprecht-Karls-Universität Heidelberg (Heidelberg University).

This work was financially supported by the Deutsche Forschungsgemeinschaft (German Research Foundation) within the program "Statistical Modeling of Complex Systems and Processes", Research Training Group 1953.

Zusammenfassung

In der vorliegenden Disssertation beschäftigen wir uns mit abhängigen Daten in drei verschiedenen Situationen. Als erstes untersuchen wir die emprisiche Prozesstheorie für (lokal) stationäre Prozesse bezüglich Klassen von glatten bzw. nichtglatten Funktionen. Dabei durchleuchten wir unsere Theorie unter dem funktionalen Abhängigkeitsmaß (*functional dependence measure*) und führen eine zusätzliche Abhängigkeit in der Zeit ein. Wir formulieren funktionale zentrale Grenzwertsätze und nichtasymptotische Maximalungleichungen. Unsere Resultate erweitern bereits bekannte Ergebnisse auf dem Gebiet der stationären Markovketten und mischenden Prozesse (*mixing sequences*). Als Anwendung unserer Theorie leiten wir gleichmäßige Konvergenzraten für nichtparametrische Regression mit lokal stationärem Rauschen sowie die funktionale Konvergenz der empirischen Verteilungsfunktion her. Weiterhin folgern wir gleichmäßige Konvergenzraten für den Kerndichte-Schätzer im (lokal) stationären Fall. Sämtliche Ergebnisse werden in der bestehenden Literatur eingeordnet und verglichen.

In einer daran anschließenden Abhandlung wenden wir uns dem Gebiet des statistischen Lernens zu. Wir betrachten dabei hoch-dimensionale stationäre Daten, die aus einer verrauschten Transformation vergangener Beobachtungen hervorgehen. Basierend auf unseren vorherigen Resultaten und ausgehend von Realisierungen eines absolut regulären mischenden Prozesses oder eines Bernoulli-Shift-Prozesses unter dem funktionalen Abhängigkeitsmaß leiten wir Orakelungleichungen für den empirischen Risiko-Minimierer her. Wenn wir davon ausgehen, dass die Daten einer Kodierung-Dekodierung-Struktur folgen, so sind wir in der Lage einen Neuronalen-Netzwerk-Schätzer zu konstruieren, der eine Vorhersage für zukünfitge Zeitpunkte erlaubt. Unter spezifischen strukturellen Bedingungen und Spärlichkeitsannahmen (*sparsity*) an die zugehörigen Netzwerke lässt sich der erwartete Vorhersagefehler nach oben abschätzen. Über quantitative Simulationen untersuchen wir das Verhalten von Netzwerk-Schätzern unter verschiedenen Modellannahmen. Wir stellen abschließend eine praktische Anwedung durch die Wettervorhersage von deutschen Städten mit den Daten des Deutschen Wetterdienstes vor.

In einer weiteren Untersuchung von abhängigen Daten widmen wir uns der nichtparametrischen Schätzung der Überlebensfunktion auf der positiven reellen Achse durch Stichproben mit mulitplikativen Messabweichungen. Das vorgeschlagene datengetriebene Verfahren in dieser Arbeit basiert auf der Schätzung der entsprechenden Mellin-Transformierten und einer Regularisierung ihrer Inversen durch einen spektralen Cut-Off; die datengetriebene Wahl des Cut-Off-Parameters gleicht üblicherweise den Bias und die Varianz aus (*bias-variance trade-off*). Für die Analyse des Bias-Terms führen wir sogenannte *Mellin-Sobolev-Räume* ein, welche die Regularität der Überlebensfunktion durch das Zerfallverhalten ihrer Mellin-Transformierten charakterisiert. Den Varianz-Term werden wir anhand von unabhängig, identisch verteilten (i.i.d.) Beobachtungen und abhängigen Daten durchleuchten. Wie zuvor spezialisieren wir uns auf Bernoulli-Shift-Prozesse unter dem funktionalen Abhängigkeitsmaß und auf absolut regulär mischende Prozesse. Im i.i.d.-Fall erreichen wir Minimax-Optimalität des sprektralen Cut-Off-Schätzers auf Mellin-Sobolev-Räumen.

Abstract

In this doctoral dissertation we will investigate dependence structures in three different cases.

We first provide a framework for empirical process theory of (locally) stationary processes for classes of either smooth or nonsmooth functions. The theory is approached by using the so-called functional dependence measure in order to quantify dependence. This work extends known results for stationary Markov chains and mixing sequences while accounting for additional time dependence. The main contributions consist of functional central limit theorems and nonasymptotic maximal inequalities. These can be employed to show, for example, uniform convergence rates for nonparametric regression with locally stationary noise. We further derive rates for kernel density estimators in the case of stationary and locally stationary observations. A special focus is placed on the functional convergence of the empirical distribution function (EDF). Comparisons with results based on other measures of dependence are carried out, as well.

In a subsequent step, we consider high-dimensional stationary processes where new observations are generated by a noisy transformation of past observations. By means of our previous results we prove oracle inequalities for the empirical risk minimizer if the data is generated by either an absolutely regular mixing sequence (β-mixing) or a Bernoulli shift process under functional dependence. Assuming that the underlying transformation of our data follows an encoder-decoder structure, we construct an encoder-decoder neural network estimator for the prediction of future time steps. We give upper bounds for the expected forecast error under specific structural and sparsity conditions on the network architecture. In a quantitative simulation we discuss the behavior of network estimators under different model assumptions and provide a weather forecast for German cities using data available by the German Meteorological Service (Deutsche Wetterdienst).

Moving onto a different setting, we study the nonparametric estimation of an unknown survival function with support on the positive real line based on a sample with multiplicative measurement errors. The proposed fully data-driven procedure involves an estimation step of the survival function's Mellin transform and a regularization of the Mellin transform's inverse by a spectral cut-off. A data-driven choice of the cut-off parameter balances bias and variance. In order to discuss the bias term, we consider *Mellin-Sobolev spaces* which characterize the regularity of the unknown survival function by the decay behavior of its Mellin transform. When analyzing the variance term we consider the standard i.i.d. case and incorporate dependent observations in form of Bernoulli shift processes and absolutely regular mixing sequences. In the i.i.d. setting we are able to show minimax-optimality over Mellin-Sobolev spaces for the spectral cut-off estimator.

Contents

Chapter 1

Introduction

In this thesis we explore stochastic processes based on two dependence structures, Bernoulli shift processes under the functional dependence measure and absolutely regular mixing sequences. Beginning this thesis with empirical process theory for locally stationary processes, Chapter 2 and Chapter 3, we wind our way through its consequences in forecasting with neural networks, Chapter 4, and bring it to an end with survival analysis, Chapter 5. Let us briefly introduce these topics and place them in their literary context.

Empirical process theory. Empirical process theory is a powerful tool to prove uniform convergence rates and weak convergence of functionals. The main concern revolves around the probabilistic behavior of processes $\{\mathbb{G}_n(f) : f \in \mathcal{F}\}$, where

$$\mathbb{G}_n(f) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{ f(X_i) - \mathbb{E} f(X_i) \big\},$$

is given for a random sample X_i, $i = 1, ..., n$, and a function f in the class $\mathcal{F}$ of measurable functions. The theory for independent variables is well-studied (cf. Dudley [2014], Giné and Nickl [2016], van der Vaart and Wellner [1996] or van der Vaart [1998] for an overview, based on the original ideas of Donsker [1952], Dudley [1966], Dudley [1978], Pollard [1982] and Ossiander [1987], among others). For dependent observations various approaches have been discussed. There exist a well-developed empirical process theory and large deviation results for Harris-recurrent Markov chains based on regenerative schemes (cf. Levental [1988], Samur [2004], Ellis and Wyner [1989] and Adamczak [2008], among others) or geometric ergodicity (cf. Kulik et al. [2019]). To quantify the speed of convergence in maximal inequalities, additional assumptions like β-recurrence (cf. Karlsen and Tjø stheim [2001]) have to be imposed. The theory covers a rich class of Markov chains, but for instance, fails to discuss linear processes.

In the context of stationary processes, an empirical process theory under complex assumptions on the moments of means is formulated in Dehling et al. [2009] and further discussion papers. In the paradigm of weak dependence, in which the size of covariances of Lipschitz continuous functions of random variables is measured, Doukhan and Neumann [2007] provides Bernstein-type inequalities.

Focusing on the analysis of the empirical distribution function (EDF), much more techniques have been developed. Regarding the functional weak convergence of the EDF, more specific conditions are stated in the literature for stationary observations. The work of Durieu and Tusche [2014] provides functional convergence of the EDF (Theorem 4 therein) using bounds for covariances of Hölder functions of random variables. Another abstract concept was introduced by Berkes et al. [2009] via stationary mixing (S-mixing), which imposes the existence of m-dependent approximations of the original observations. They derived strong approximations and uniform central limit theorems for the EDF. More approaches are presented in Borovkova et al. [2001] and Dedecker [2010].

Another idea to measure dependence of random variables is given by mixing coefficients. In this case, several concepts were established, the most common (in order of increasing strength) being α-, β- and ϕ-mixing. For an overview about mixing coefficients we refer to Doukhan [1994]. Large deviation results and uniform central limit theorems for general classes of functions (not only for EDFs) can be derived by using coupling techniques, cf. Rio [1995], Liebscher [1996], Arcones and Yu [1994], Yu [1994] for α-mixing, successively refined by Doukhan et al. [1995], Rio [1998], Dedecker and Prieur [2007] and Dedecker [2010] (the last two for EDFs only); Rio [2017] for β-mixing and Dedecker and Louhichi [2002], Borovkova et al. [2001] for β- and ϕ-mixing. For a comprehensive overview also consider Andrews and Pollard [1994], Dedecker and Louhichi [2002] and Rio [2017].

In Dedecker and Louhichi [2002] it is argued that β-mixing is the weakest mixing assumption that allows for a “complete” empirical process theory which incorporates maximal inequalities and uniform central limit theorems. There exist explicit upper bounds for β-mixing coefficients for Markov chains (cf. Heinrich [1992]) and for so-called V-geometric mixing coefficients (cf. Meyn and Tweedie [2009]). For several stationary time series models like linear processes (cf. Pham and Tran [1985] for α-mixing), ARMA (cf. Mokkadem [1988]), nonlinear AR (cf. Karlsen and Tjø stheim [2001]) and GARCH processes (cf. Francq and Zakoïan [2006]) there are upper bounds on mixing coefficients available. A common assumption in these results is that the observed process or, more often, the innovations of the corresponding process, have a continuous distribution. This is a crucial assumption in order to handle the rather complicated mixing coefficients defined over a supremum over two different σ-algebras. A relaxation of β-mixing coefficients was investigated by [Dedecker and Prieur, 2007, Theorem 1] and is specifically designed for the analysis of the EDF. They provide coefficients that are defined by conditional expectations of certain classes of functions and are easier to upper bound for a wide range of time series models.

Recently, another measure of dependence, the so-called functional dependence measure, became popular (cf. Wu [2005]). It uses a Bernoulli shift representation (see equation (1.2.1) below) and a decomposition into martingales and m-dependent sequences. It has been shown in various applications that the functional dependence measure, when combined with the rich theory of martingales, allows for sharp large deviation inequalities (cf. Wu et al. [2013] or Zhang and Wu [2017]). In Wu [2008] and Mayer et al. [2020], uniform central limit theorems for the EDF were derived for stationary and piecewise locally stationary processes.

So far, to the best of our knowledge, no *general* empirical process theory that allows for general classes of functions using the functional dependence measure is available. This dissertation intends to fill this gap and prove maximal inequalities as well as functional central limit theorems under functional dependence. We are going to consider classes of smooth functions and additionally allow for nonsmooth functions in a separate discussion. In particular, our framework includes, but is by far not limited to, the empirical distribution function (EDF). Furthermore, we will draw connections and compare our results to the already existing empirical process theory for dependent data we mentioned earlier.

While the empirical process theory for Markov chains and mixing sequences cited above was developed for stationary processes, we will work in the framework of locally stationary processes and thus provide the first general empirical process theory in this setting. Locally stationary processes allow for a smooth change of the distribution over time but can locally be approximated by stationary processes. Therefore, they provide more flexible time series models (cf. Dahlhaus et al. [2019] for an introduction). The work of Dahlhaus and Polonik [2009] investigates spectral empirical processes for linear processes and Mayer et al. [2020] shows a functional central limit theorem for a localized empirical distribution function.

The definition of the functional dependence measure for locally stationary processes is similar to its stationary version and is easy to calculate for many time series models. It does not rely on the stationarity assumption but on the representation of the process as a Bernoulli shift. Therefore, many upper bounds for stationary time series given in Wu [2011], including recursively defined models and linear models, directly carry over to the locally stationary case. It seems reasonable to use it as a starting point to generalize empirical process theory for stationary processes to the more general setting of locally stationary processes. Contrary to the concept of β-mixing, the functional dependence measure can effortlessly deal with transformed observations by using Hölder-type assumptions. On the same note, it can easily be calculated in many situations and is not restricted to continuously distributed data. It is worth pointing out that linear processes are covered, as well.

However, there are some peculiarities that come along when using the functional dependence measure (1.2.2). While for Harris-recurrent Markov chains and β-mixing, the empirical process theory is independent of the function class considered, the situation for the functional dependence measure is more involved. For example, let us consider some (possibly high-dimensional) random variables $X_1, ..., X_n$ and a general function class $f \in \mathcal{F}$. In order to quantify the dependence of a transformation $f(X)$, $X = (X_1, ..., X_n)$, we have to impose certain smoothness conditions on the function f itself. Therefore, certain measures of distances that are necessary to derive weak convergence results will change according to the dependence structure of X. They also have to be "compatible" with the function class $\mathcal{F}$. However, requiring smoothness conditions will pose challenging issues when considering chaining procedures where rare events are excluded by (nonsmooth) indicator functions.

Oracle inequalities. In a next step, we would like to apply the results from the empirical process theory for classes of smooth functions. They help us to better understand the performance of certain time series forecasting regimes that involve empirical risk minimizers. Let us consider such a (high-dimensional) time series and assume that the observed data X_i, $i = 1, ..., n$, is a realization of a stationary stochastic process which follows an autoregressive regression model with Subgaussian innovations ε_i,

$$X_i = f_0(X_{i-1}, ..., X_{i-r}) + \varepsilon_i, \quad i = r + 1, ..., n,$$

for number of lags r, made precise later in (4.1.1). First, we develop oracle inequalities for observations that come from Bernoulli shift processes under the functional dependence measure and on a sample that is drawn from β-mixing sequences. Oracle inequalities provide a useful tool when trying to measure an estimator's accuracy.

Based on the prediction error

$$R(f) = \frac{1}{d}\mathbb{E}[|X_{r+1} - f(X_r, ..., X_1)|_2^2]$$

(cf. (4.1.2)), we will define its empirical counterpart and asses the excess Bayes risk of the empirical risk minimizer over a certain class of measurable functions.

Specifically, we investigate the statistical behavior of structured neural network estimators in high-dimensional time series forecasting. Here, we harness the approximation results obtained by Schmidt-Hieber [2017] and derive upper bounds for the excess Bayes risk. The general idea to use networks for forecasting was already described in Tang and Fishwick [1993], Kline [2004], Zhang [2012]. However, up to now, no theoretical results for convergence rates on dependent data in this setting seem to exist. They are of utmost value since the conditions needed can shed light on the choice of a network structure. Furthermore, quantifying the impact of the underlying dependence in the data can yield information on the number of training samples (or observation length, in a time series context) required to bound the prediction error.

A general overview for oracle inequalities and its use cases can be found in Bühlmann and Van De Geer [2011]. Related studies can be found in, for example, Blanchard et al. [2008] and Schmidt-Hieber [2017].

Let us emphasize that our stochastic results can be seen as a generalization of Schmidt-Hieber [2017] which deals with independent and identically distributed (i.i.d.) data X_i and one-dimensional outputs Y_i, in particular. Furthermore, we will be imposing an encoder-decoder structure which is crucial when transferring the strong convergence rates from Schmidt-Hieber [2017] to the setting of high-dimensional outputs, especially in the case of recursively defined time series.

Survival analysis. So far, many statistical procedures under dependent data have yet to be developed. For our next topic, we focus on an estimation procedure for the unknown survival function of a positive random variable X, that is,

$$S : \mathbb{R}_+ \to [0, 1], \quad x \mapsto \mathbb{P}(X > x),$$

given identically distributed copies of $Y = XU$, where X and U are independent of each other and U has a known density $g : \mathbb{R}_+ \to \mathbb{R}_+$. In this setting, the density $f_Y : \mathbb{R}_+ \to \mathbb{R}_+$ of Y is given by

$$f_Y(y) = [f * g](y) := \int_0^\infty f(x)g(y/x)x^{-1}dx \quad \forall y \in \mathbb{R}_+.$$

Here, "$*$" denotes multiplicative convolution. The estimation of S using an observable sample $Y_1, \dots, Y_n$ from f_Y becomes thus an inverse problem called multiplicative deconvolution. We account for certain dependence structures on the sample $Y_1, \dots, Y_n$. More precisely, we assume $X_1, \dots, X_n$ to be a stationary process while the error terms $U_1, \dots, U_n$ will be i.i.d.
Multiplicative deconvolution models are, for example, treated in the recent work of Brenner Miguel et al. [2020], which uses the Mellin transform to construct a density estimator under multiplicative measurement errors, and Brenner Miguel [2021], where they consider the multivariate case of density estimation. The model of multiplicative measurement errors was motivated in the work of Belomestny and Goldenshluger [2020] as a generalization of several models, for instance, the multiplicative censoring model or the stochastic volatility model.

To the best of our knowledge, the estimation for the survival function of a positive random variable for general multiplicative measurement errors has not been studied yet.

The investigations in Vardi [1989] and Vardi and Zhang [1992] introduce *multiplicative censoring* and focus on multiplicative deconvolution problems with uniformly distributed multiplicative error U on $[0, 1]$. This model is applied often in survival analysis and was motivated in Van Es et al. [2000]. The estimation of the cumulative distribution function of X is discussed in Asgharian and Wolfson [2005] and Vardi and Zhang [1992]. Series expansion methods are covered in Andersen and Hansen [2001], treating the model as an inverse problem. The survival function estimation in a multiplicative censoring model is considered in Brunel et al. [2016] using a kernel estimator and a convolution power kernel estimator. Assuming an uniform error distribution on an interval $[1 - \alpha, 1 + \alpha]$ for $\alpha \in (0, 1)$, Comte and Dion [2016] analyzes a projection survival function estimator with respect to the Laguerre basis. On the other hand, Belomestny et al. [2016] sheds light on the theory with beta-distributed error U.

In the work of Belomestny and Goldenshluger [2020], the authors use the Mellin transform to construct a kernel estimator for the pointwise density estimation. Moreover, they point out that the following widely used naive approach is a special case of their estimation strategy. Applying a log-transformation, the model $Y = XU$ is equivalent to $\log(Y) = \log(X) + \log(U)$, which in turn is just a convolution of the log-transformed data. As a consequence, the density of $\log(X)$ is eventually estimated by employing the usual strategies for nonparametric deconvolution problems (cf. Meister [2009]) and then transformed back to an estimator of f. However, it is difficult to interpret regularity conditions on the density of $\log(X)$. Furthermore, the global risk analysis of an estimator using this naive approach is challenging as Comte and Dion [2016] pointed out.

In the last chapter, we extend the results of Brenner Miguel et al. [2020] for the estimation of the survival function. To do so, we introduce the Mellin transform for positive random variables and revise necessary properties. The key to the multiplicative deconvolution problem is the multiplication theorem of the Mellin transform, which roughly states $\mathcal{M}[f_Y] = \mathcal{M}[f]\mathcal{M}[g]$ for a density $f_Y = f * g$. Exploiting the multiplication theorem and applying a spectral cut-off regularization on the inverse of the Mellin transform we then define a survival function estimator. The accuracy of the estimator is measured by introducing a global risk in terms of the $\mathbb{L}^2$-norm. We borrow ideas from, for example, Engl et al. [2000] when characterizing the underlying inverse problem. The regularity conditions will be expressed in the form of *Mellin-Sobolev spaces.* The proposed estimator, however, involves a tuning parameter which is selected by a data-driven method. We will first establish an oracle inequality for the fully-data driven spectral cut-off estimator under fairly mild assumptions on the error density g and then show that uniformly over *Mellin-Sobolev spaces* the proposed data-driven estimator in the i.i.d. case is minimax-optimal. More precisely, we state both an upper bound for the mean integrated squared error of the fully-data driven spectral cut-off estimator and a general lower bound for the estimation of the density f based on copies from $f_Y = f * g$.

Besides the discussion of i.i.d. samples, we also examine the estimator's behavior when certain dependence structures are present, similar to Comte et al. [2008] whose work considers density estimation for general ARCH models using the log-transformed data approach. We base our theory on the two concepts of dependence that we are by now familiar with, namely absolutely regular mixing sequences and Bernoulli shift processes under the functional dependence measure. The merits of the latter were already highlighted in the above paragraphs.

1.1 Contribution and outline

General display of results. For overview purposes, the subsequent chapters will display main results first. Thereafter, proofs and technicalities are provided in a separate section found towards the end of each chapter, respectively.

We would like to give a short overview of how this thesis and each chapter is structured.

Chapter 2 - Empirical process theory for smooth functions

Our main contributions are the following.

- We derive maximal inequalities for $\mathbb{G}_n(f)$ for classes of smooth functions $\mathcal{F}$,
- a chaining device which preserves smoothness during the chaining procedure and
- conditions to ensure asymptotic tightness and functional convergence of $\mathbb{G}_n(f)$, $f \in \mathcal{F}$.

In Section 2.2, we present our main result Theorem 2.2.5, the functional central limit theorem under minimal moment conditions. As a special case, we derive a version for stationary processes. We include a discussion on the custom distance that we are going to use in this new context and compare our result with the empirical process theory for β-mixing sequences. Some assumptions are postponed to Section 2.3, where a novel multivariate central limit theorem for locally stationary processes is shown. In Section 2.4 we provide new maximal inequalities for $\mathbb{G}_n(f)$ over a class $\mathcal{F}$ of either finitely or infinitely many functions. We apply our theory in Section 2.5 to prove uniform convergence rates and weak convergence of several estimators. The aim of the last section is to highlight the wide range of applicability of our theory and to display the typical conditions which have to be imposed, as well as some discussion. We derive further large deviation inequalities in Section 2.6 as they might be of interest for some future applications. In Section 2.7, a conclusion is drawn. A detailed account on the proofs can be found in Section 2.8.

This chapter has been published independently as Phandoidaen and Richter [2022].

Chapter 3 - Empirical process theory for nonsmooth functions

The results obtained are similar to Chapter 2 but now consider nonsmooth functions.

Even though our theory allows for general function classes, we will focus on the empirical distribution function (EDF). In particular, we derive functional convergence of the EDF under weak conditions on the moments and the dependence structure of the process itself. We will see that our results typically require weaker conditions on the underlying dependence structure than comparable results for the stationary case mentioned above.

In Section 3.1, we present our main result, Theorem 3.1.2, the functional central limit theorem under minimal moment conditions, now formulated in its nonsmooth context. We then derive a version for stationary processes and discuss its application on empirical distribution functions where the underlying process is either stationary or locally stationary. Some assumptions are postponed to Section 3.2, where we state a reformulation of the multivariate central limit theorem for locally stationary processes presented in Chapter 2. In the new setting, Section 3.3 provides new maximal inequalities for $\mathbb{G}_n(f)$ in case of a finite and infinite function class $\mathcal{F}$. It is the aim of Subsection 3.3.3 to show the wide range our theory can be applied to. Section 3.4 accommodates a conclusion for this chapter. We postpone all detailed proofs to Section 3.5.

Chapter 4 - Oracle inequalities applied to neural network estimators

Based on the empirical process theory for smooth functions we present

- oracle inequalities, i.e. upper bounds for the excess Bayes risk

of the empirical risk minimizer. They can be found as Theorem 4.1.3 in case of Bernoulli shift processes under the functional dependence measure and in Theorem 4.1.6 for β-mixing sequences. As a consequence we then have

- convergence rates for encoder-decoder neural network estimators,

found in Theorem 4.2.6 and Theorem 4.2.5 for each dependence case, respectively.

In Section 4.1 we briefly reintroduce the two aforementioned dependence structures and provide the corresponding oracle inequalities. In Section 4.2 we describe the structural assumptions on the true regression function, formulate the neural network estimator and provide upper bounds for the excess Bayes risk. A small simulation study is provided in Section 4.3 and sheds light on the behavior of neural network estimators from a practical point of view. We also include an investigation on real-world temperature data for German cities. The arising approximation error when using a certain class of neural networks to approximate the true regression function is discussed in Section 4.4. In Section 4.5, a conclusion is drawn. Most of the proofs are deferred to Section 4.6.

Chapter 5 - Survival function estimation as a deconvolution problem

We summarize the novel contributions in survival function estimation with multiplicative errors via Mellin transforms as follows. In the setting of i.i.d. and dependent observations we provide

- an upper bound for the spectral cut-off estimator's global risk and the rates in the Mellin-Sobolev space (cf. Theorem 5.3.2 and corollary),
- minimax-optimality for i.i.d. data (cf. Theorem 5.3.8) as well as
- an upper bound and respective rates when a data-driven method is employed for selecting an appropriate tuning parameter (cf. Theorem 5.4.2 and corollary).

Our discourse is organized as follows. After a brief motivation and introduction, Section 5.3 revises the Mellin transform, briefly offers an overview of its main properties and gives examples. Thereafter, we establish our estimation strategy for the survival function. We provide oracle inequalities for independent and dependent data and with respect to the MISE derive upper bounds with parametric as well as nonparametric rates in an appropriate Mellin-Sobolev space. We conclude the section by deriving the estimator's minimax optimality in the i.i.d. case. Since our theory depends on a spectral cut-off parameter, we propose in Section 5.4 a data-driven method based on a penalized contrast approach for an optimal choice. As before, we state an oracle inequality and derive an upper bound, accordingly. To illustrate our results, we showcase numerical studies in Section 5.5. A conclusion is drawn in Section 5.6. The proofs can be found in Section 5.7.

1.2 Preliminaries and notation

Functional dependence. The functional dependence measure is a key concept of this thesis. It uses a representation of the given process as a Bernoulli shift process and quantifies dependence with a norm defined on moments. More precisely, we assume that $X_i = (X_{ij})_{j=1,\dots,d}$, $i = 1, \dots, n$, is a d-dimensional process of the form

$$X_i := X_{i,n} := J_{i,n}(\mathcal{A}_i) \tag{1.2.1}$$

where $\mathcal{A}_i = (\varepsilon_i, \varepsilon_{i-1}, ...)$ for ε_i, $i \in \mathbb{Z}$, a sequence of i.i.d. random variables in $\mathbb{R}^{\tilde{d}}$ ($d, \tilde{d} \in \mathbb{N}$) and some measurable function $J_{i,n} : (\mathbb{R}^{\tilde{d}})^{\mathbb{N}_0} \to \mathbb{R}^d$, $i = 1, ..., n$, $n \in \mathbb{N}$. For a real-valued random variable W and some $\nu > 0$, we define the norm $\|W\|_\nu := \mathbb{E}[|W|^\nu]^{1/\nu}$. If ε_k^* is an independent copy of ε_k, independent of $\varepsilon_i, i \in \mathbb{Z}$, we define $\mathcal{A}_i^{*(i-k)} := (\varepsilon_i, ..., \varepsilon_{i-k+1}, \varepsilon_{i-k}^*, \varepsilon_{i-k-1}, ...)$ and $X_i^{*(i-k)} := X_{i,n}^{*(i-k)} := J_{i,n}(\mathcal{A}_i^{*(i-k)})$. Then, the uniform functional dependence measure is given by

$$\delta_\nu^X(k) = \sup_{i=1,...,n} \sup_{j=1,...,d} \big\| X_{ij} - X_{ij}^{*(i-k)} \big\|_\nu. \tag{1.2.2}$$

Graphically, δ_ν^X can be interpreted to measure the impact of ε_0 in X_k. Although representation (1.2.1) appears to be rather restrictive, it does cover a large variety of processes. In Borkar [1993] it was motivated that the set of all processes of the form $X_i = J(\varepsilon_i, \varepsilon_{i-1}, ...)$ should be equal to the set of all stationary and ergodic processes. However, this conjecture was proven not to hold true by Rosenblatt [2009]. We additionally allow J to vary with i and n in order to cover processes which change their stochastic behavior over time; as suggested by the double index of $J_{i,n}$ or $X_{i,n}$. This is exactly the form of the so-called locally stationary processes discussed in Dahlhaus et al. [2019]. As it is quite common to omit the double index in $X_{i,n}$ we will denote the process simply by X_i if no confusion arises.

Since we are working in the time series context, many applications ask for functions f that not only depend on the actual observation of the process but on the whole (infinite) past $Z_i := (X_i, X_{i-1}, X_{i-2}, ...)$. During the course of this thesis, one fundamental aim is to derive asymptotic properties of the empirical process

$$\mathbb{G}_n(f) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \big\{ f(Z_i, \frac{i}{n}) - \mathbb{E} f(Z_i, \frac{i}{n}) \big\}, \quad f \in \mathcal{F}, \tag{1.2.3}$$

where

$$\mathcal{F} \subset \{ f : (\mathbb{R}^d)^{\mathbb{N}_0} \times [0,1] \to \mathbb{R} \text{ measurable} \}.$$

Let $\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote the bracketing entropy, that is, the logarithm of the number of ε-brackets with respect to some distance $\|\cdot\|$ that is necessary to cover $\mathcal{F}$; a precise formulation is given below. If the corresponding bracketing entropy integral $\int_0^1 \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)} d\varepsilon$ on a custom seminorm V_n is finite, the weak convergence of the empirical process in (1.2.3) can be guaranteed.

We now introduce some basic notation grouped into the chapters they first appear in.

Chapter 2 and Chapter 3

For $a, b \in \mathbb{R}$, let $a \wedge b := \min\{a, b\}$, $a \vee b := \max\{a, b\}$. For $s \in (0, 1]$, a sequence $z = (z_i)_{i \in \mathbb{N}_0}$ of elements of $\mathbb{R}^d$ (equipped with the maximum norm $|\cdot|_\infty$) and an absolutely summable sequence $\chi = (\chi_i)_{i \in \mathbb{N}_0}$ of nonnegative real numbers, we set

$$|z|_{\chi,s} := \Big(\sum_{i=0}^\infty \chi_i |z_i|_\infty^s \Big)^{1/s}$$

and abbreviate $|z|_\chi := |z|_{\chi,1}$.

For $k \in \mathbb{N}$,

$$H(k) := 1 \vee \log(k), \tag{1.2.4}$$

which naturally appears in large deviation inequalities. For a given finite class $\mathcal{F}$, let $|\mathcal{F}|$ denote its cardinality. We use the abbreviation

$$H = H(|\mathcal{F}|) = 1 \vee \log|\mathcal{F}| \tag{1.2.5}$$

if no confusion arises. For some norm $\|\cdot\|$, let $\mathbb{N}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote the bracketing numbers, that is, the smallest number of ε-brackets $[l_j, u_j] := \{f \in \mathcal{F} : l_j \leq f \leq u_j\}$ needed to cover $\mathcal{F}$. In more detail, an ε-bracket contains measurable functions $l_j, u_j : (\mathbb{R}^d)^{\mathbb{N}_0} \times [0,1] \to \mathbb{R}$ for which $\|u_j - l_j\| \leq \varepsilon$ for all j. Let $\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|) := \log \mathbb{N}(\varepsilon, \mathcal{F}, \|\cdot\|)$ denote the bracketing entropy.

For $\nu \geq 1$, let

$$\|f\|_{\nu,n} := \Big(\frac{1}{n}\sum_{i=1}^n \|f(Z_i, \frac{i}{n})\|_\nu^\nu\Big)^{1/\nu}.$$

Chapter 4

We additionally make use of the following notations.

For $q > 0$, let $|v|_q := (\sum_{j=1}^r |v_j|^q)^{1/q}$ denote the q-norm of a vector $v \in \mathbb{R}^r$ with the convention $|v|_\infty := \max_{j=1,\dots,r} |v_j|$ and $|v|_0 := |\{j \in \{1,...,r\} : v_j \neq 0\}|$.

For a matrix $W \in \mathbb{R}^{r\times s}$, let $|W|_\infty := \max_{j=1,\dots,r,k=1,\dots,s} |W_{jk}|$ and $|W|_0 := |\{j \in \{1,...,r\}, k \in \{1,...,s\} : W_{jk} \neq 0\}|$.

For mappings $f : \mathbb{R}^t \to \mathbb{R}$, we denote by $\|f\|_\infty := \sup_{x\in\mathbb{R}^t} |f(x)|$ the supremum norm. If $f : \mathbb{R}^t \to \mathbb{R}^d$, we use $\|f\|_\infty := \||f|_\infty\|_\infty$.

For sequences x_n, y_n we write $x_n \lesssim y_n$ if there exists a constant $C > 0$ independent of n such that $x_n \leq C y_n$ for $n \in \mathbb{N}$. We write $x_n \asymp y_n$ if $x_n \lesssim y_n$ and $y_n \lesssim x_n$.

Chapter 5

The flavor of this chapter will differ from the preceding chapters. We therefore need to establish a new notation habit which might, however, be overloaded. Please keep that in mind, when considering this chapter.

We define for any weight function $\omega : \mathbb{R} \to \mathbb{R}_+$ the corresponding weighted norm by $\|h\|_\omega^2 := \int_0^\infty |h(x)|^2 \omega(x) dx$ for a measurable, complex-valued function h. Denote by $\mathbb{L}^2(\mathbb{R}_+, \omega)$ the set of all measurable, complex-valued functions with finite $\| \,.\, \|_\omega$-norm and by $\langle h_1, h_2\rangle_\omega := \int_0^\infty h_1(x)\overline{h_2}(x)\omega(x)dx$ for $h_1, h_2 \in \mathbb{L}^2(\mathbb{R}_+, \omega)$ the corresponding weighted scalar product. On the other hand, $\mathbb{L}^1(\Omega, \omega) := \{h : \Omega \to \mathbb{C} : \|h\|_{\mathbb{L}^1(\Omega,\omega)} < \infty\}$, where $\|h\|_{\mathbb{L}^1(\Omega,\omega)} := \int_\Omega |h(x)|\omega(x)dx$, for a normed space Ω. Similarly, we set $\mathbb{L}^2(\mathbb{R}) := \{h : \mathbb{R} \to \mathbb{C} \text{ measurable } : \|h\|_{\mathbb{R}}^2 < \infty\}$ for $\|h\|_{\mathbb{R}}^2 := \int_{-\infty}^\infty h(t)\overline{h(t)}dt$ and $\mathbb{L}^1(\mathbb{R}) := \{h : \mathbb{R} \to \mathbb{C} \text{ measurable } : \|h\|_{\mathbb{R}} < \infty\}$ for $\|h\|_{\mathbb{R}} := \int_{-\infty}^\infty |h(t)|dt$, if the constant unit weight is applied. We use the shorthand $\|h\|^2 := \|h\|_{\mathbb{R}}^2$ on $\mathbb{L}^2(\mathbb{R})$, sometimes written as $\mathbb{L}^2(\mathbb{R}, x^0)$ depending on the context.

Chapter 2

Empirical process theory for smooth functions under functional dependence

At the heart of this thesis lies the empirical process theory, spanning the next two chapters and impacting the one after. In this chapter we are going to focus on locally stationary process for classes of smooth functions. As we consider Bernoulli shift processes under the functional dependence measure, novel approaches have to be found or existing ideas modified in order to provide asymptotic tightness. From a purely technical perspective, one of the main issues will be finding an alternative to the chaining procedure, which is crucial for similar results in the setting of i.i.d. and mixing observations, but does not work for the setting studied here.

2.1 Motivation

One of the crucial concepts for empirical process theory is chaining. Chaining describes a procedure for dealing with stochastic processes that are indexed by an uncountable set. It allows us to handle suprema of such processes or a transformation thereof by breaking down the index set into a sequence of finite subsets that is able to "approximate" the original set. We then decompose the process into a sum of increments based on the carefully chosen sequence of finite subsets. These contributions can be bounded in a way that guarantees convergence of the sum. We will eventually, in Subsection 2.8.3, provide a novel chaining technique for Bernoulli shift processes under the functional dependence measure for classes of sufficiently smooth functions as standard approaches as found, for example, in van der Vaart [1998] fail to work. An introduction to chaining is also available by Pollard [2012].

A basic property that a seminorm $V_n(\cdot)$ has to fulfill when using a chaining procedure is that its square has to be an upper bound of the variance of $\mathbb{G}_n(f)$, that is,

$$\mathrm{Var}(\mathbb{G}_n(f)) \leq V_n(f)^2.$$

We therefore first derive an expression for the left hand side. Let $k \in \mathbb{N}_0$. For a sequence $W_i = J_{i,n}(\mathcal{A}_i)$ with $\|W_i\|_1 < \infty$, let $P_{i-k}W := \mathbb{E}[W_i \mid \mathcal{A}_{i-k}] - \mathbb{E}[W_i \mid \mathcal{A}_{i-k-1}]$. Then, $(P_{i-k}W_i)_{i\in\mathbb{N}}$ is a martingale difference sequence with respect to $(\mathcal{A}_i)_{i\in\mathbb{N}}$, and $W_i - \mathbb{E}W_i = \sum_{k=0}^{\infty} P_{i-k}W_i$.

Our theory is mainly based on the case $\nu = 2$. By the projection property of the conditional expectation and an elementary property of δ_2^W (cf. [Wu, 2005, Theorem 1]), we have

$$\|P_{i-k}W_i\|_2 \le \min\{\|W_i\|_2, \delta_2^W(k)\}. \tag{2.1.1}$$

Since $\min\{a_1, b_1\} + \min\{a_2, b_2\} \le \min\{a_1 + a_2, b_1 + b_2\}$ for nonnegative real numbers $a_1, b_1, a_2, b_2 \ge 0$, we obtain

$$\begin{aligned}
\mathrm{Var}(\mathbb{G}_n(f))^{1/2} &\le \sum_{k=0}^{\infty} \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} P_{i-k} f(Z_i, \frac{i}{n}) \Big\|_2 \\
&= \sum_{k=0}^{\infty} \Big(\frac{1}{n} \sum_{i=1}^{n} \|P_{i-k} f(Z_i, \frac{i}{n})\|_2^2 \Big)^{1/2} \\
&\le \sum_{k=0}^{\infty} \Big(\frac{1}{n} \sum_{i=1}^{n} \min\{\big\|f(Z_i, \frac{i}{n})\big\|_2^2, \delta_2^{f(Z,\frac{i}{n})}(k)^2\} \Big)^{1/2} \\
&\le \sum_{k=0}^{\infty} \min\Big\{ \|f\|_{2,n}, \Big(\frac{1}{n} \sum_{i=1}^{n} \delta_2^{f(Z,\frac{i}{n})}(k)^2 \Big)^{1/2} \Big\}.
\end{aligned} \tag{2.1.2}$$

To further bound (2.1.2), we therefore have to investigate for $u \in [0, 1]$,

$$\delta_2^{f(Z,u)}(k) = \sup_{i\in\mathbb{Z}} \big\| f(Z_i, u) - f(Z_i^{*(i-k)}, u) \big\|_2, \tag{2.1.3}$$

the functional dependence measure accommodated for f (cf. 1.2.2). Due to the linear nature of the functional dependence measure, it is necessary to impose smoothness assumption on $f \in \mathcal{F}$ in order to derive upper bounds for $\delta_2^{f(Z,u)}(k)$ in terms of the functional dependence measure of X_i from (1.2.2). When doing so, we "lose" the properties of f and especially of $\|f\|_{2,n}$, whose information we would like to retain. Therefore, our goal should be to bound (2.1.3) by some quantity which is completely independent of a specific f. To obtain a rich enough theory for our setting (of local stationarity), we have to allow f to depend on n and include classes $\mathcal{F}$ where parts of f change the convergence rate of $\mathbb{G}_n(f)$. In an abstract way, we would like for each $f \in \mathcal{F}$ to factorize as

$$f(z, u) = D_{f,n}(u) \cdot \bar{f}(z, u), \qquad z \in (\mathbb{R}^d)^{\mathbb{N}_0}, u \in [0, 1],$$

where $D_{f,n}(u) \in \mathbb{R}$ does not depend on z.

Given some decreasing sequence $\Delta(k) \ge 0$ and $\mathbb{D}_n \ge 0$ which fulfill

$$\sup_{u\in[0,1]} \sup_{f\in\mathcal{F}} \delta_2^{\bar{f}(Z,u)} \le \Delta(k), \qquad \sup_{f\in\mathcal{F}} \Big(\frac{1}{n} \sum_{i=1}^{n} D_{f,n}\big(\frac{i}{n}\big)^2 \Big)^{1/2} \le \mathbb{D}_n, \tag{2.1.4}$$

we obtain from (2.1.2),

$$\mathrm{Var}(\mathbb{G}_n(f))^{1/2} \le \sum_{k=0}^{\infty} \min\{\|f\|_{2,n}, \mathbb{D}_n \Delta(k)\}.$$

This motivates the definition of our actual $V_n(\cdot)$ in the next section.

2.2 A functional central limit theorem

Roughly speaking, a process X_i, $i = 1, ..., n$, is called locally stationary if for each $u \in [0,1]$, there exists a stationary process $\tilde{X}_i(u)$, $i = 1, ..., n$ such that $X_i \approx \tilde{X}_i(u)$ if $|u - \frac{i}{n}|$ is small (cf. Dahlhaus et al. [2019]). The exact form needed is stated in Assumption 2.3.1. Thus, X_i behaves stationary around each fixed (rescaled) time point $u \in [0,1]$, but over the whole time period $i = 1, ..., n$ its distribution can change drastically. Deterministic properties of the process like expectation, covariance, spectral density or empirical distribution functions therefore also depend on the rescaled time $u \in [0,1]$. Typical estimators are of the form

$$\frac{1}{nh} \sum_{i=1}^{n} K\Big(\frac{i/n - u}{h}\Big) \bar{f}\big(Z_i, \frac{i}{n}\big)$$

where K is a kernel function and $h = h_n > 0$ is a bandwidth. Such a localization will certainly have an impact on the convergence rate. To cover these cases, suppose that any $f \in \mathcal{F}$ has a representation

$$f(z,u) = D_{f,n}(u) \cdot \bar{f}(z,u), \qquad z \in (\mathbb{R}^d)^{\mathbb{N}_0}, u \in [0,1], \tag{2.2.1}$$

where $\bar{f}$ is independent of n and $D_{f,n}(u)$ is independent of z. We put

$$\bar{\mathcal{F}} := \{\bar{f} : f \in \mathcal{F}\}. \tag{2.2.2}$$

Definition 2.2.1. The function class $\bar{\mathcal{F}}$ is called a $(L_{\mathcal{F}}, s, R, C)$-class if there exist $L_{\mathcal{F}} = (L_{\mathcal{F},i})_{i \in \mathbb{N}_0}$, a sequence of nonnegative real numbers, $s \in (0,1]$ and $R : (\mathbb{R}^d)^{\mathbb{N}_0} \times [0,1] \to [0,\infty)$ such that for all $u \in [0,1]$, $z, z' \in (\mathbb{R}^d)^{\mathbb{N}_0}$, $\bar{f} \in \bar{\mathcal{F}}$,

$$|\bar{f}(z,u) - \bar{f}(z',u)| \le |z - z'|_{L_{\mathcal{F}},s}^s \cdot \big[R(z,u) + R(z',u)\big].$$

Furthermore, $C = (C_R, C_{\bar{f}}) \in (0,\infty)^2$ fulfills $\sup_u |\bar{f}(0,u)| \le C_{\bar{f}}$, $\sup_u |R(0,u)| \le C_R$.

We can consider the function class $\bar{\mathcal{F}}$ as consisting of Hölder continuous functions in direction of z.

Remark 2.2.2. The condition on $\bar{\mathcal{F}}$ to be a $(L_{\mathcal{F}}, s, R, C)$-class poses a smoothness condition on any $\bar{f} \in \bar{\mathcal{F}}$ separately. There is no need for any connection between the different $\bar{f} \in \bar{\mathcal{F}}$. Moreover, it should not be confused with the important example of so-called parametric Lipschitz classes in empirical process theory (cf. [van der Vaart, 1998, Example 19.7]), where it is assumed that there exists some parameter space $\Theta \subset \mathbb{R}^p$ such that $\bar{\mathcal{F}} = \{\bar{f}_\theta : \theta \in \Theta\}$ and for two $\theta_1, \theta_2 \in \Theta$, $|\bar{f}_{\theta_1}(z,u) - \bar{f}_{\theta_2}(z,u)| \le m(z,u) \cdot |\theta_1 - \theta_2|_\infty$ for some measurable function m.

The basic assumption for our main results is the following compatibility condition on $\mathcal{F}$. We will later provide a slightly adapted version of it.

Assumption 2.2.3. *The class $\bar{\mathcal{F}}$ is a $(L_{\mathcal{F}}, s, R, C)$-class. There exist $p \in (1, \infty]$, $C_X > 0$ such that*

$$\textit{(i)} \ \sup_{i,u} \|R(Z_i, u)\|_{2p} \le C_R, \qquad \textit{(ii)} \ \sup_{i,j} \|X_{ij}\|_{\frac{2sp}{p-1}} \le C_X. \tag{2.2.3}$$

Let $\mathbb{D}_n \ge 0$, $\Delta(k) \ge 0$ be such that for all $k \in \mathbb{N}_0$,

$$2dC_R \cdot \sum_{j=0}^{k} L_{\mathcal{F},j}\big(\delta^X_{\frac{2sp}{p-1}}(k-j)\big)^s \le \Delta(k), \qquad \sup_{f \in \mathcal{F}} \Big(\frac{1}{n}\sum_{i=1}^{n} \Big|D_{f,n}\big(\frac{i}{n}\big)\Big|^2\Big)^{1/2} \le \mathbb{D}_n.$$

While (2.2.3) summarizes moment assumptions on X_{ij} which are balanced by p, the sequence $\Delta(k)$ reflects the intrinsic dependence of $f(Z_i, \frac{i}{n})$. The value $\mathbb{D}_n$ measures the influence of the factor $D_{f,n}(u)$ to the convergence rate of $\mathbb{G}_n(f)$.

Based on Assumption 2.2.3, we define for $f \in \mathcal{F}$,

$$V_n(f) := \|f\|_{2,n} + \sum_{k=1}^{\infty} \min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(k)\}. \tag{2.2.4}$$

The following lemma collects some properties of V_n and in particular shows that V_n is a seminorm. The proof is straightforward and thus omitted.

Lemma 2.2.4. *Let $f, g \in \mathcal{F}$ and $\lambda \in \mathbb{R}$. Then*

(i) $V_n(0) = 0$, $V_n(f+g) \le V_n(f) + V_n(g)$ *and* $V_n(\lambda \cdot f) = |\lambda| V_n(f)$,

(ii) $|f| \le g \implies V_n(f) \le V_n(g)$,

(iii) $\|f\|_{1,n}, \|f\|_{2,n} \le V_n(f)$, *and* $V_n(f) \le V_n(\|f\|_\infty) < \infty$ *if* $\|f\|_\infty < \infty$.

Therefore, $V_n(f-g)$ can be interpreted as a (pseudo) distance between $f, g \in \mathcal{F}$.

Based on the fact that we will later assume that $\mathcal{F}$ fulfills (2.1.4) or Assumption 2.2.3 (and thus $\mathbb{G}_n(f)$ is properly standardized), it is reasonable to suppose that $\mathbb{D}_n \in (0, \infty)$ is independent of $n \in \mathbb{N}$. In this case, simpler forms of V_n can be derived for special cases of $\Delta(k)$ which are given in Table 2.1. Note that if $f(Z_i, \frac{i}{n})$, $i = 1, ..., n$, are independent, $\delta_\nu^{f(Z,u)}(k) = 0$, $\nu \ge 1$, for $k > 0$ and thus $V_n(f) = \|f\|_{2,n}$. We therefore exactly recover the case of independent variables with our theory.

We are now able to state our main result, a weak convergence statement that takes place in the normed space

$$\ell^\infty(\mathcal{F}) = \{\mathbb{G} : \mathcal{F} \to \mathbb{R} \,|\, \|\mathbb{G}\|_\infty := \sup_{f \in \mathcal{F}} |\mathbb{G}(f)| < \infty\}, \tag{2.2.5}$$

cf. van der Vaart [1998] for a detailed discussion of this space. The proof of the following Theorem 2.2.5 consists of two ingredients, convergence of the finite-dimensional distributions (cf. Theorem 2.3.4) and asymptotic tightness (cf. Corollary 2.4.5). We will have to impose some complex conditions on the process' behavior and certain regularity conditions on the function class, in order for the statements to be valid. For overview purposes, we postpone the exact wordings of the assumptions necessary. They will appear right before the two key theorems, accordingly. For the sake of completeness we have to include them here, already.

Theorem 2.2.5. *Let $\mathcal{F}$ satisfy Assumption 2.2.3, 2.3.1, 2.3.2 and 2.3.3. Suppose that*

$$\sup_{n\in\mathbb{N}} \int_0^1 \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)} d\varepsilon < \infty.$$

Then, in $\ell^\infty(\mathcal{F})$,

$$\left[\mathbb{G}_n(f)\right]_{f\in\mathcal{F}} \xrightarrow{d} \left[\mathbb{G}(f)\right]_{f\in\mathcal{F}}$$

where $(\mathbb{G}(f))_{f\in\mathcal{F}}$ is a centered Gaussian process with covariances

$$\operatorname{Cov}(\mathbb{G}(f), \mathbb{G}(g)) = \lim_{n\to\infty} \operatorname{Cov}(\mathbb{G}_n(f), \mathbb{G}_n(g)) = \Sigma^{(\mathbb{K})}$$

and $\Sigma^{(\mathbb{K})}$ is from Assumption 2.3.3.

The more challenging part will be the proof for asymptotic tightness; it only relies on Assumption 2.2.3 and consists of a new maximal inequality presented in Theorem 2.4.1. To ensure convergence of the finite-dimensional distributions, we have to formalize local stationarity (Assumption 2.3.1) and impose conditions in time direction on $\bar{f}(z,\cdot)$ (cf. Assumption 2.3.2) and $D_{f,n}(\cdot)$ (cf. Assumption 2.3.3), which is done in Section 2.3. In particular, we require that $D_{f,n}(u)$ is properly normalized.

Let us note that in the case where X_i is stationary, $\bar{f}(z,u) = \bar{f}(z)$ and $D_{f,n}(u) = 1$. Hence, Assumption 2.3.1, 2.3.2 and 2.3.3 are automatically fulfilled. In other words, Assumption 2.2.3 is sufficient for Theorem 2.2.5 in the stationary case. We formulate this finding as a simple corollary. Let $X_i = J(\mathcal{A}_i)$, $i = 1, ..., n$, be a stationary process and

$$\tilde{\mathbb{G}}_n(h) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{h(X_i) - \mathbb{E}h(X_i)\right\}$$

where the functions h are contained in

$$\mathcal{H} \subset \{h : \mathbb{R}^d \to \mathbb{R} \text{ measurable}\}$$

such that for all $x, y \in \mathbb{R}^d$, $|h(x) - h(y)| \le L_{\mathcal{H}}|x-y|_\infty^s$.

Corollary 2.2.6. *Suppose that $\|X_1\|_{2s} < \infty$. Let $\Delta(k) := 2dL_{\mathcal{H}}\delta_{2s}^X(k)^s$ and $\mathbb{D}_n := 1$. Assuming that*

$$\sup_{n\in\mathbb{N}} \int_0^1 \sqrt{\mathbb{H}(\varepsilon, \mathcal{H}, V_n)} d\varepsilon < \infty, \tag{2.2.6}$$

we have in $\ell^\infty(\mathcal{H})$,

$$\left[\tilde{\mathbb{G}}_n(h)\right]_{h\in\mathcal{H}} \xrightarrow{d} \left[\tilde{\mathbb{G}}(h)\right]_{h\in\mathcal{H}}$$

where $(\tilde{\mathbb{G}}(h))_{h\in\mathcal{H}}$ is a centered Gaussian process with covariances

$$\operatorname{Cov}(\tilde{\mathbb{G}}(h_1), \tilde{\mathbb{G}}(h_2)) = \sum_{k\in\mathbb{Z}} \operatorname{Cov}(h_1(X_0), h_2(X_k)).$$

2.2.1 Form of V_n and discussion on $\Delta(k)$

Form of V_n

Suppose that $\mathbb{D}_n \in (0,\infty)$ is independent of $n \in \mathbb{N}$. Based on the decay rates of $\Delta(k)$, simpler forms of V_n can be derived and are given in Table 2.1. These results are elementary and are proven in Lemma 2.8.14 and Lemma 2.8.15 in Section 2.8.

	$\Delta(j)$	
	$cj^{-\alpha}$, $\alpha > 1, c > 0$	$c\rho^j$, $\rho \in (0,1)$, $c > 0$
$V_n(f)$	$\|f\|_{2,n} \max\{\|f\|_{2,n}^{-\frac{1}{\alpha}}, 1\}$	$\|f\|_{2,n} \max\{\log(\|f\|_{2,n}^{-1}), 1\}$
$\int_0^\sigma \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)}d\varepsilon$	$\int_0^{\tilde{\sigma}} \varepsilon^{-\frac{1}{\alpha}} \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,n})}d\varepsilon$	$\int_0^{\tilde{\sigma}} \log(\varepsilon^{-1}) \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,n})}d\varepsilon$

Table 2.1: Equivalent expressions of V_n and the corresponding entropy integral under the condition that $\mathbb{D}_n \in (0,\infty)$ is independent of n. We omitted the lower and upper bound constants which are only depending on c, ρ, α and $\mathbb{D}_n$. Furthermore, $\tilde{\sigma} = \tilde{\sigma}(\sigma)$ fulfills $\tilde{\sigma} \to 0$ for $\sigma \to 0$.

Discussion on $\Delta(k)$

Assumption 2.2.3 provides an upper bound $\Delta(k)$ for

$$\sum_{j=0}^{k} L_{\mathcal{F},j}\big(\delta^X_{\frac{2sp}{p-1}}(k-j)\big)^s,$$

which is a convolution of the uniform Hölder constants $L_{\mathcal{F},j}$ of $f \in \mathcal{F}$ and the dependence measure $\delta^X_{\frac{2sp}{p-1}}(k)$ of $X = (X_1, ..., X_n)$. Therefore, the specific form of $f \in \mathcal{F}$ has an impact on the dependence structure, which is then introduced via V_n. This is quite different to other typical chaining approaches for Harris-recurrent Markov chains or β-mixing sequences where the dependence structure of X_i simply transfers to functions $f(X_i)$ without further conditions.

Furthermore, in contrast to other chaining approaches, we *have to* ask for the existence of moments of X_i in Assumption 2.2.3, even though $\mathbb{G}_n(f)$ only involves $f(X_i)$. This is due to the linear nature of the functional dependence measure (1.2.2). If f is Lipschitz continuous with respect to its first argument ($s = 1$ in Assumption 2.2.3), we have to impose $\sup_{i,j} \|X_{ij}\|_2 < \infty$. However, these moment assumptions can be relaxed at the cost of larger $\Delta(k)$, as follows. Let us consider the special case that $f(Z_i, \frac{i}{n})$ only depends on X_i, that is, $f(z,u) = f(z_0,u)$. If f is bounded and Lipschitz continuous with respect to its first argument with Lipschitz constant L, for any $s \in (0,1]$,

$$|f(z_0,u) - f(z_0',u)| \le \min\{2\|f\|_\infty, L|z_0 - z_0'|\} \le (2\|f\|_\infty)^{1-s} L^s |z_0 - z_0'|^s.$$

Thus, $\Delta(k)$ can be chosen proportionally to $\delta_{2s}^X(k)^s$. This means that we can reduce the moment assumption to $\sup_{i,j} \|X_{ij}\|_{2s} < \infty$ at the cost of having a larger norm V_n.

2.2.2 Comparison to empirical process theory with β-mixing

In this section, we compare our functional central limit theorem for stationary processes from Corollary 2.2.6 under functional dependence with similar results obtained under β-mixing. Unfortunately, it does not seem to be straightforward to find a general setting under which the functional dependence measure δ_2^X can be compared with the β-mixing coefficients β^X of X_i, $i = 1, ..., n$. However, in some special cases, both quantities can be upper bounded.

Upper bounds for dependence coefficients of linear processes

Consider the linear process

$$X_i = \sum_{k=0}^{\infty} a_k \varepsilon_{i-k}, \quad i = 1, ..., n,$$

with an absolutely summable sequence a_k, $k \in \mathbb{N}_0$, and i.i.d. ε_k, $k \in \mathbb{Z}$, with $\mathbb{E}\varepsilon_1 = 0$. Then it is immediate that

$$\delta_2^X(k) \le 2|a_k| \cdot \|\varepsilon_1\|_2.$$

From Pham and Tran [1985] (cf. also [Doukhan, 1994, Section 2.3.1]) we have the following result. If for some $\nu \ge 1$, $\|\varepsilon_1\|_\nu < \infty$, ε_1 has a Lipschitz continuous Lebesgue density and the process X_i is invertible, then for some constant $\zeta > 0$,

$$\beta^X(k) \le \zeta \cdot \Big(\sum_{m=k}^{\infty} |A_{m,\nu}|^{\frac{1}{1+\nu}} \Big) \vee \Big(\sum_{m=k}^{\infty} L(A_{m,2}) \Big)$$

where $A_{m,s} := \sum_{k=m}^{\infty} |a_k|^s$ and $L(u) = \sqrt{u(1 \vee |\log(u)|)}$. If $a_k = O(k^{-\alpha})$ for some $\alpha > 1$,

$$\delta_2^X(k) = O(k^{-\alpha}), \qquad \beta^X(k) = O\big(k^{-\alpha + \frac{1+\alpha}{1+\nu} + 1} \vee (k^{-\alpha + \frac{3}{2}} \log(k)^{1/2})\big). \tag{2.2.7}$$

Note that the calculation of the functional dependence measure is much easier. Moreover, bounds for $\beta^X(k)$ are typically larger than $\delta_2^X(k)$; the reason being that δ_2^X is of simpler structure than the more involved formulation of dependence via σ-algebras for the β-mixing coefficients. For recursively defined processes with a finite number of lags, δ_2^X are typically upper bounded by geometric decaying coefficients (cf. Wu [2011], Dahlhaus et al. [2019]); the same holds true for $\beta^X(k)$ under additional continuity assumptions (cf. [Doukhan, 1994, Section 2.4], or Kulik et al. [2019], Heinrich [1992] among others).

Entropy integral

In Doukhan et al. [1995] (and also Dedecker and Louhichi [2002]), it was shown that if X_i, $i = 1, ..., n$, is stationary and β-mixing with coefficients $\beta(k)$, $k \in \mathbb{N}_0$, then

$$\int_0^1 \sqrt{\mathbb{H}(\varepsilon, \mathcal{H}, \|\cdot\|_{2,\beta})} d\varepsilon < \infty \tag{2.2.8}$$

implies weak convergence of $(\mathbb{G}_n(h))_{h\in\mathcal{H}}$ in $\ell^\infty(\mathcal{H})$. Here, the $\|\cdot\|_{2,\beta}$-norm is defined as follows. If β^{-1} denotes the inverse cadlag of the decreasing function $t \mapsto \beta(\lfloor t \rfloor)$ and Q_h the inverse cadlag of the tail function $t \mapsto \mathbb{P}(h(X_1) > t)$, then

$$\|h\|_{2,\beta} := \int_0^1 \beta^{-1}(u) Q_h(u)^2 du.$$

Condition (2.2.8) was later relaxed in [Rio, 2017, Theorem 8.3]. It was shown that if $\mathcal{F}$ consists of indicator functions of specific classes of sets (in particular, $\mathcal{F}$ corresponds to the empirical distribution function), weak convergence can be obtained under less restrictive conditions than (2.2.8). At the moment, our theory does not allow us to analyze indicator functions directly because $\mathcal{F}$ has to be a $(L_{\mathcal{F}}, s, R, C)$-class, meaning that smoothness assumptions have to be satisfied. This will be the topic of Chapter 3.

In the special cases of polynomial and geometric decay, simple upper bounds for $\|h\|_{2,\beta}$ are available (cf. Dedecker and Louhichi [2002]). If $\sum_{k=0}^\infty k^{b-1}\beta(k) < \infty$ for some $b \geq 1$, then $\|\cdot\|_{2,\beta}$ is upper bounded by $\|\cdot\|_{\frac{2b}{b-1}}$.

Generally speaking, (2.2.8) asks for $\frac{2b}{b-1}$ moments of the process $f(X_i)$ to exist while our condition in (2.2.6) only requires second moments of $f(X_i)$. However, the additional factors given in the entropy integral (cf. Table 2.1) reduce the function classes' size. In specific examples (cf. (2.2.7)) it may occur that the entropy integral (2.2.6) is finite while (2.2.8) is infinite due to missing summability of $\beta^X(k)$.

To give a precise comparison, consider the situation of linear processes from (2.2.7). If $\nu > 2\alpha + 1$, we can choose $b = \alpha - \frac{3}{2}$. Then, the two entropy integrals from Corollary 2.2.6 (left) and (2.2.8) read as

$$\int_0^1 \varepsilon^{-\frac{1}{\alpha}} \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|_2)} d\varepsilon \qquad \text{vs.} \qquad \int_0^1 \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|_{\frac{4\alpha-6}{2\alpha-5}})} d\varepsilon.$$

Here, the entropy integral for mixing only exists if $\alpha > \frac{5}{2}$. The difference in the behavior can be explained by the different bounds used for the variance of $\mathbb{G}_n(f)$.

2.3 A general central limit theorem for locally stationary processes

In this section, we introduce the remaining assumptions needed in Theorem 2.2.5 which pose regularity conditions on the process X_i and the function class $\mathcal{F}$ in time direction. They are used to derive a multivariate central limit theorem for $(\mathbb{G}_n(f_1), ..., \mathbb{G}_n(f_k))$

under minimal moment conditions, Theorem 2.3.4. Comparable results in different and more specific contexts were shown in Dahlhaus et al. [2019] or Truquet [2020].

We first formalize the property of X_i to be locally stationary (cf. Dahlhaus et al. [2019]). Recall the quantities $R(\cdot), s, p$ from Assumption 2.2.3.

Assumption 2.3.1. *For each $u \in [0,1]$, there exists a process $\tilde{X}_i(u) = J(\mathcal{A}_i, u)$, $i \in \mathbb{Z}$, where J is a measurable function. Furthermore, there exists some $C_X > 0$, $\varsigma \in (0,1]$ such that for every $i \in \{1, ..., n\}$, $u_1, u_2 \in [0,1]$,*

$$\Big\|X_i - \tilde{X}_i(\frac{i}{n})\Big\|_{\frac{2sp}{p-1}} \leq C_X n^{-\varsigma}, \qquad \|\tilde{X}_i(u_1) - \tilde{X}_i(u_2)\|_{\frac{2sp}{p-1}} \leq C_X |u_1 - u_2|^{\varsigma}.$$

For $\tilde{Z}_i(u) = (\tilde{X}_i(u), \tilde{X}_{i-1}(u), ...)$ we require $\sup_{v,u} \|R(\tilde{Z}_0(v), u)\|_{2p} < \infty$.

The behavior of the functions $f(z,u) = D_{f,n}(u) \cdot \bar{f}(z,u)$ of the class $\mathcal{F}$ in the direction of time $u \in [0,1]$ is controlled by the following two continuity assumptions which impose conditions on $\bar{f}(z, \cdot)$ and $D_{f,n}(\cdot)$ separately.

Assumption 2.3.2. *There exists some $\varsigma \in (0,1]$ such that for every $\bar{f} \in \bar{\mathcal{F}}$,*

$$\sup_{v,u_1,u_2} \Big\| \frac{|\bar{f}(\tilde{Z}_0(v), u_1) - \bar{f}(\tilde{Z}_0(v), u_2)|}{|u_1 - u_2|^{\varsigma}} \Big\|_2 < \infty.$$

For $f \in \mathcal{F}$, let $D_{f,n}^{\infty} := \sup_{i=1,\dots,n} D_{f,n}(\frac{i}{n})$.

Assumption 2.3.3. *For all $f \in \mathcal{F}$, the function $\frac{D_{f,n}(\cdot)}{D_{f,n}^{\infty}}$ has bounded variation uniformly in n, and*

$$\sup_{n \in \mathbb{N}} \frac{1}{n} \sum_{i=1}^{n} D_{f,n}(\frac{i}{n})^2 < \infty, \qquad \frac{D_{f,n}^{\infty}}{\sqrt{n}} \to 0. \tag{2.3.1}$$

One of the two following cases holds true.

(i) Case $\mathbb{K} = 1$ (global). For all $f, g \in \mathcal{F}$,

$$u \mapsto \mathbb{E}\big[\mathbb{E}[\bar{f}(\tilde{Z}_{j_1}(u), u) \mid \mathcal{A}_0] \cdot \mathbb{E}[\bar{g}(\tilde{Z}_{j_2}(u), u) \mid \mathcal{A}_0]\big]$$

has bounded variation for all $j_1, j_2 \in \mathbb{N}_0$ and the following limit exists:

$$\Sigma_{fg}^{(1)} := \lim_{n \to \infty} \int_0^1 D_{f,n}(u) D_{g,n}(u) \cdot \sum_{j \in \mathbb{Z}} \mathrm{Cov}(\bar{f}(\tilde{Z}_0(u), u), \bar{g}(\tilde{Z}_j(u), u)) du.$$

(ii) Case $\mathbb{K} = 2$ (local). There exists a sequence $h_n > 0$ and $v \in [0,1]$ such that $\mathrm{supp}(D_{f,n}(\cdot)) \subset [v - h_n, v + h_n]$. *It holds true that*

$$h_n \to 0, \qquad \sup_{n \in \mathbb{N}} (h_n^{1/2} \cdot D_{f,n}^{\infty}) < \infty.$$

The following limit exists for all $f, g \in \mathcal{F}$:

$$\Sigma_{fg}^{(2)} := \lim_{n \to \infty} \int_0^1 D_{f,n}(u) D_{g,n}(u) du \cdot \sum_{j \in \mathbb{Z}} \mathrm{Cov}(\bar{f}(\tilde{Z}_0(v), v), \bar{g}(\tilde{Z}_j(v), v)).$$

Assumption 2.3.3 looks rather technical. The first part including (2.3.1) guarantees the right normalization of $D_{f,n}(\cdot)$. The second part ensures the convergence of the asymptotic variances $\mathrm{Var}(\mathbb{G}_n(f))$ and covariances $\mathrm{Cov}(\mathbb{G}_n(f), \mathbb{G}_n(g))$.

We obtain the following central limit theorem.

Theorem 2.3.4. *Let $\mathcal{F}$ satisfy Assumption 2.2.3, 2.3.1, 2.3.2 and 2.3.3. Let $m \in \mathbb{N}$ and $f_1, ..., f_m \in \mathcal{F}$ and $\Sigma^{(\mathbb{K})} = (\Sigma^{(\mathbb{K})}_{f_k f_l})_{k,l=1,...,m}$. Then for $\mathbb{K} \in \{1, 2\}$,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \begin{pmatrix} f_1(Z_i, \frac{i}{n}) \\ \vdots \\ f_m(Z_i, \frac{i}{n}) \end{pmatrix} - \mathbb{E} \begin{pmatrix} f_1(Z_i, \frac{i}{n}) \\ \vdots \\ f_m(Z_i, \frac{i}{n}) \end{pmatrix} \right\} \xrightarrow{d} N(0, \Sigma^{(\mathbb{K})}).$$

Theorem 2.3.4 generalizes the one-dimensional central limit theorem from Dahlhaus et al. [2019]. We now comment on the assumptions.

Remark 2.3.5. Assumptions 2.3.1, 2.3.2 and 2.3.3 allow for very general structures of $f \in \mathcal{F}$. However, in many special cases, a subset of them is automatically fulfilled:

- If X_i is stationary, then Assumption 2.2.3 already implies Assumption 2.3.1.
- If $\bar{f}(z, u) = \bar{f}(z)$ does not depend on u, Assumption 2.3.2 is fulfilled.

Regarding Assumption 2.3.3 we have:

- If $D_{f,n}(u) = 1$, X_i is stationary and $\bar{f}(z, u) = \bar{f}(z)$, then Assumption 2.2.3 already implies Assumption 2.3.3(i) with $\Sigma^{(1)}_{fg} = \sum_{j\in\mathbb{Z}} \mathrm{Cov}(\bar{f}(Z_0), \bar{f}(Z_j))$.
- If $D_{f,n}(u) = 1$, Assumption 2.2.3 and 2.3.2 are satisfied with $s = \varsigma = 1$, then Assumption 2.3.3(i) holds true with $\Sigma^{(1)}_{fg} = \int_0^1 \sum_{j\in\mathbb{Z}} \mathrm{Cov}(\bar{f}(\tilde{Z}_0(u), u), \bar{f}(\tilde{Z}_j(u), u))du$.
- If $h_n \to 0$, $nh_n \to \infty$ and $D_{f,n}(u) = \frac{1}{\sqrt{h_n}} K(\frac{u-v}{h_n})$ for some Lipschitz continuous kernel $K : \mathbb{R} \to \mathbb{R}$ with support $\subset [-1, 1]$ and fixed $v \in (0, 1)$, then Assumption 2.3.3(ii) holds true with $\Sigma^{(2)}_{fg} = \int_0^1 K(x)^2 dx \cdot \sum_{j\in\mathbb{Z}} \mathrm{Cov}(\bar{f}(\tilde{Z}_0(v), v), \bar{f}(\tilde{Z}_j(v), v))$.

2.4 Maximal inequalities and asymptotic tightness under functional dependence

In this section, we provide the necessary ingredients for the proof of asymptotic tightness of $\mathbb{G}_n(f)$. We first derive a new maximal inequality for finite $\mathcal{F}$ under functional dependence in Theorem 2.4.1 and then generalize this bound to arbitrary $\mathcal{F}$ in Section 2.4.2 using a modified chaining technique.

2.4.1 Maximal inequalities

We first derive a maximal inequality which is a main ingredient for chaining devices but also is of independent interest. To state the result, let

$$\beta(q) := \sum_{j=q}^{\infty} \Delta(k).$$

and define

$$q^*(x) := \min\{q \in \mathbb{N} : \beta(q) \le q \cdot x\}.$$

Set $D_n^\infty(u) := \sup_{f\in\mathcal{F}} |D_{f,n}(u)|$. For $\nu \ge 2$, choose $\mathbb{D}_{\nu,n}^\infty$ such that

$$\Big(\frac{1}{n}\sum_{i=1}^n D_n^\infty\big(\frac{i}{n}\big)^\nu\Big)^{1/\nu} \le \mathbb{D}_{\nu,n}^\infty. \tag{2.4.1}$$

Put $\mathbb{D}_n^\infty = \mathbb{D}_{2,n}^\infty$. Recall that $H = H(|\mathcal{F}|) = 1 \vee \log|\mathcal{F}|$ as in (1.2.5).

Theorem 2.4.1. *Suppose that $\mathcal{F}$ satisfies $|\mathcal{F}| < \infty$ and Assumption 2.2.3. Then there exists some universal constant $c > 0$ such that the following holds: If $\sup_{f\in\mathcal{F}} \|f\|_\infty \le M$ and $\sup_{f\in\mathcal{F}} V_n(f) \le \sigma$, then*

$$\mathbb{E}\max_{f\in\mathcal{F}}\big|\mathbb{G}_n(f)\big| \le c\cdot \min_{q\in\{1,\dots,n\}}\Big[\sigma\sqrt{H} + \sqrt{H}\cdot\mathbb{D}_n^\infty\beta(q) + \frac{qMH}{\sqrt{n}}\Big] \tag{2.4.2}$$

and

$$\mathbb{E}\max_{f\in\mathcal{F}}\big|\mathbb{G}_n(f)\big| \le 2c\cdot\Big(\sigma\sqrt{H} + q^*\big(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty}\big)\frac{MH}{\sqrt{n}}\Big). \tag{2.4.3}$$

Clearly, the second bound (2.4.3) is a corollary of (2.4.2) which balances the two terms where q is involved. Values of $q^*(\cdot)$ for the two prominent cases that $\Delta(\cdot)$ is polynomially or exponentially decaying can be found in Table 2.2. The proof of Theorem 2.4.1 relies on a decomposition of $\mathbb{G}_n(f)$ in i.i.d. parts and a residual term of martingale structure. Similar decompositions are also the core of empirical process results for Harris-recurrent Markov chains (cf. Li et al. [2016]) and mixing sequences (cf. Dedecker and Louhichi [2002]).

In the next subsections, we will prove asymptotic tightness for $\mathbb{G}_n(f)$ under the condition that $\mathbb{D}_n^\infty$, $\mathbb{D}_n$ do not depend on n. However, uniform convergence rates of $\mathbb{G}_n(f)$ for finite $\mathcal{F}$ (growing with n) can be obtained without this condition but with additional moment assumptions, which is done in the following Corollary 2.4.3. To incorporate the additional moment assumptions, we use a slightly stronger assumption than Assumption 2.2.3.

Assumption 2.4.2. *Let $\bar{\mathcal{F}}$ be a $(L_{\mathcal{F}}, s, R, C)$-class. There exist $\nu \ge 2$, $p \in (1,\infty]$, $C_X > 0$ such that*

$$\sup_{i,u}\|R(Z_i,u)\|_{\nu p} \le C_R, \qquad \sup_{i,j}\|X_{ij}\|_{\frac{\nu s p}{p-1}} \le C_X. \tag{2.4.4}$$

	$\Delta(j)$	
	$Cj^{-\alpha},\ \alpha > 1$	$C\rho^j,\ \rho \in (0,1)$
$q^*(x)$	$\max\{x^{-\frac{1}{\alpha}}, 1\}$	$\max\{\log(x^{-1}), 1\}$
$r(\delta)$	$\min\{\delta^{\frac{\alpha}{\alpha-1}}, \delta\}$	$\min\{\frac{\delta}{\log(\delta^{-1})}, \delta\}$

Table 2.2: Equivalent expressions of $q^*(\cdot)$ and $r(\cdot)$ taken from Lemma 2.8.13 in Section 2.8.7. We omitted the lower and upper bound constants which are only depending on C, ρ, α.

Let $\mathbb{D}_n \geq 0$, $\Delta(k) \geq 0$ be such that for all $k \in \mathbb{N}_0$,

$$2dC_R \cdot \sum_{j=0}^{k} L_{\mathcal{F},j}\big(\delta^X_{\frac{\nu s p}{p-1}}(k-j)\big)^s \leq \Delta(k), \qquad \sup_{f\in\mathcal{F}}\Big(\frac{1}{n}\sum_{i=1}^{n}\Big|D_{f,n}\big(\frac{i}{n}\big)\Big|^2\Big)^{1/2} \leq \mathbb{D}_n.$$

Note that Assumption 2.2.3 is obtained by taking $\nu = 2$. For $\delta > 0$, let

$$r(\delta) := \max\{r > 0 : q^*(r)r \leq \delta\},$$

cf. Table 2.2 for values of $r(\cdot)$ in special cases.

Corollary 2.4.3 (Uniform convergence rates)**.** *Suppose that $\mathcal{F}$ satisfies $|\mathcal{F}| < \infty$ and that Assumption 2.4.2 is fulfilled for some $\nu \geq 2$. Let $C_\Delta := 4d \cdot |L_{\mathcal{F}}|_1 \cdot C_X^s C_R + C_{\bar{f}}$. Furthermore, suppose that*

$$\sup_{n\in\mathbb{N}}\sup_{f\in\mathcal{F}} V_n(f) < \infty, \qquad \sup_{n\in\mathbb{N}} \frac{\mathbb{D}^\infty_{\nu,n}}{\mathbb{D}^\infty_n} < \infty, \qquad \sup_{n\in\mathbb{N}} \frac{C_\Delta H}{n^{1-\frac{2}{\nu}} r\big(\frac{\sigma}{\mathbb{D}^\infty_n}\big)^2} < \infty. \tag{2.4.5}$$

Then,

$$\max_{f\in\mathcal{F}} |\mathbb{G}_n(f)| = O_p(\sqrt{H}).$$

The first condition in (2.4.5) guarantees that $\mathbb{G}_n(f)$ is properly normalized. The second and third condition are needed to prove that “rare events”, where $|f(Z_i, \frac{i}{n})|$ exceeds some threshold $M_n \in (0,\infty)$, are of the same order as $\sqrt{H}$. For this, we may need more than two moments of $f(Z_i, \frac{i}{n})$, that is, $\nu > 2$, depending on $\sqrt{H}$ and the behavior of $\mathbb{D}^\infty_n$.

Corollary 2.4.3 can be used to prove (optimal) convergence rates for kernel density and regression estimators as well as maximum likelihood estimators under dependence. We give an example in Section 2.5.

2.4.2 Asymptotic tightness

In this subsection, we extend the maximal inequality from Theorem 2.4.1 to arbitrary (infinite) classes $\mathcal{F}$. Since Assumption 2.2.3 forces $f \in \mathcal{F}$ to be Hölder continuous with respect to its first argument z, classical chaining approaches which use indicator functions do not apply here. For example, a standard approach in [van der Vaart, 1998, Lemma 19.34] is to split the function of interest into a sequence of segments on a cleverly chosen sequence of nested partition of $\mathcal{F}$, using the noncontinuous indicator function. Each segment can then be bounded in a way such that their sum is finite.

We provide a new chaining technique based on a special truncation method which preserves continuity in Subsection 2.8.3 in its full detail.

For $n \in \mathbb{N}$, $\delta > 0$ and $k \in \mathbb{N}$ define $H(k) = 1 \vee \log(k)$ and

$$m(n, \delta, k) := r\big(\frac{\delta}{\mathbb{D}_n}\big) \cdot \frac{\mathbb{D}_n^{\infty} n^{1/2}}{H(k)^{1/2}}. \tag{2.4.6}$$

Here, $m(n, \delta, k)$ represents the threshold for rare events in the chaining procedure. We have the following result.

Theorem 2.4.4. *Let $\mathcal{F}$ satisfy Assumption 2.2.3 and let F be some envelope function of $\mathcal{F}$, that is, for each $f \in \mathcal{F}$, $|f| \leq F$. Let $\sigma > 0$ and assume that $\sup_{f\in\mathcal{F}} V_n(f) \leq \sigma$. Then there exists some universal constant $\tilde{c} > 0$ such that*

$$\begin{aligned}
&\mathbb{E} \sup_{f\in\mathcal{F}} \big|\mathbb{G}_n(f)\big| \\
\leq\ & \tilde{c}\Big[\Big(1 + \frac{\mathbb{D}_n^{\infty}}{\mathbb{D}_n} + \frac{\mathbb{D}_n}{\mathbb{D}_n^{\infty}}\Big) \int_0^{\sigma} \sqrt{\mathbb{H}\big(\varepsilon, \mathcal{F}, V_n\big)}\, \mathrm{d}\varepsilon + \sqrt{n}\big\|F \mathbb{1}_{\{F > \frac{1}{4} m(n,\sigma,\mathbb{N}(\frac{\sigma}{2},\mathcal{F},V_n))\}}\big\|_{1,n}\Big].
\end{aligned}$$

As a corollary, we obtain asymptotic equicontinuity of $\mathbb{G}_n(f)$. Here, we use Assumption 2.3.1 and 2.3.2 only to discuss the remainder term in Theorem 2.4.4 without imposing the existence of additional moments.

Corollary 2.4.5. *Let $\mathcal{F}$ satisfy Assumption 2.2.3, 2.3.1 and 2.3.2. Suppose that*

$$\sup_{n\in\mathbb{N}} \int_0^1 \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)} d\varepsilon < \infty. \tag{2.4.7}$$

Furthermore, assume that $\mathbb{D}_n, \mathbb{D}_n^{\infty} \in (0, \infty)$ are independent of n, and

$$\sup_{i=1,\dots,n} \frac{D_n^{\infty}(\frac{i}{n})}{\sqrt{n}} \to 0. \tag{2.4.8}$$

Then, the process $\mathbb{G}_n(f)$ is equicontinuous with respect to V_n, that is, for every $\eta > 0$,

$$\lim_{\sigma\to 0} \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{f,g\in\mathcal{F}, V_n(f-g)\leq\sigma} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| \geq \eta\Big) = 0.$$

2.5 Applications

In this section, we provide some applications of some main results (Corollary 2.4.3 and Corollary 2.2.5). We will focus on locally stationary processes and therefore use a localization in our functionals, but the results also hold true for stationary processes, accordingly.

Let $K : \mathbb{R} \to \mathbb{R}$ be some bounded kernel function which is Lipschitz continuous with Lipschitz constant L_K, $\int K(u)du = 1$, $\int K(u)^2 du \in (0, \infty)$ and support $\subset [-\frac{1}{2}, \frac{1}{2}]$. For some bandwidth $h := h_n > 0$, put $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$.

In the first example we consider the nonparametric kernel estimator in the context of nonparametric regression with fixed design and locally stationary noise. We show that under conditions on the bandwidth h, which are common in the presence of dependence (cf. Hansen [2008] or Vogt [2012]), we obtain the optimal uniform convergence rate $\sqrt{\frac{\log(n)}{nh}}$. Recall that for sequences a_n, b_n we have $a_n \gtrsim b_n$ whenever there exist $c > 0$ such that $a_n \geq cb_n$ for all $n \in \mathbb{N}$.

Example 2.5.1 (Nonparametric Regression)**.** Let X_i be some arbitrary process of the form (1.2.1) with $\sum_{k=0}^{\infty} \delta_2^X(k) < \infty$ fulfilling $\sup_{i=1,\dots,n} \|X_i\|_\nu \leq C_X \in (0, \infty)$ for some $\nu > 2$. Suppose that we observe Y_i, $i = 1, ..., n$, given by

$$Y_i = g\big(\frac{i}{n}\big) + X_i,$$

where $g : [0,1] \to \mathbb{R}$ is some function. Estimation of g is performed via

$$\hat{g}_{n,h}(v) := \frac{1}{n}\sum_{i=1}^{n} K_h\Big(\frac{i}{n} - v\Big)Y_i.$$

Suppose that either

- $\delta_2^X(j) \leq \kappa j^{-\alpha}$ with some $\kappa > 0, \alpha > 1$, and $h \gtrsim \big(\frac{\log(n)}{n^{1-\frac{2}{\nu}}}\big)^{\frac{\alpha-1}{\alpha}}$ or
- $\delta_2^X(j) \leq \kappa\rho^j$ with some $\kappa > 0, \rho \in (0,1)$ and $h \gtrsim \frac{\log(n)^3}{n^{1-\frac{2}{\nu}}}$.

Equations (2.5.1) and (2.5.2) (cf. below) imply

$$\sup_{v\in[0,1]} |\hat{g}_{n,h}(v) - \mathbb{E}\hat{g}_{n,h}(v)| = O_p\Big(\sqrt{\frac{\log(n)}{nh}}\Big).$$

First note that due to Lipschitz continuity of K with Lipschitz constant L_K, we have

$$\begin{aligned}&\sup_{|v-v'|\leq n^{-3}} \Big|(\hat{g}_{n,h}(v) - \mathbb{E}\hat{g}_{n,h}(v)) - (\hat{g}_{n,h}(v') - \mathbb{E}\hat{g}_{n,h}(v'))\Big| \\ \leq\ & \frac{L_K n^{-3}}{nh^2}\sum_{i=1}^{n}\big(|X_i| + \mathbb{E}|X_i|\big) = O_p(n^{-1}). \end{aligned} \tag{2.5.1}$$

For the grid $\mathcal{V}_n = \{in^{-3}, i = 1, ..., n^3\}$, which discretizes $[0,1]$ up to distances n^{-3}, we obtain by Corollary 2.4.3,

$$\sqrt{nh} \sup_{v \in \mathcal{V}_n} |\hat{g}_{n,h}(v) - \mathbb{E}\hat{g}_{n,h}(v)| = \sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)| = O_p\big(\sqrt{\log |\mathcal{V}_n|}\big) = O_p\big(\log(n)^{1/2}\big) \quad (2.5.2)$$

where

$$\mathcal{F} = \Big\{ f_v(x,u) = \frac{1}{\sqrt{h}} K\big(\frac{u-v}{h}\big) x : v \in V_n \Big\}.$$

The conditions of Corollary 2.4.3 are easily verified: First, observe that $f_v(x,u) = D_{f,n}(u) \cdot \bar{f}_v(x,u)$ with $D_{f,n}(u) = \frac{1}{\sqrt{h}} K(\frac{u-v}{h})$ and $\bar{f}_v(x,u) = x$. Thus, Assumption 2.4.2 is satisfied with $\Delta(k) = 2\delta_2^X(k)$, $p = \infty$, $R(\cdot) = C_R = 1$. Furthermore, $\mathbb{D}_n = |K|_\infty$, $\mathbb{D}_{\nu,n} = \frac{|K|_\infty}{\sqrt{h}}$ and

$$\|f_v\|_{2,n} \le \frac{1}{\sqrt{h}} \Big(\frac{1}{n} \sum_{i=1}^n K\big(\frac{v-u}{h}\big)^2 \|X_i\|_2^2 \Big)^{1/2} \le C_X |K|_\infty,$$

which shows that $\sup_{f \in \mathcal{F}} V_n(f) = O(1)$. The conditions on h emerge from the last condition in (2.4.5) and the bounds for $r(\cdot)$ from Table 2.2.

For the following two examples we assume that the underlying process X_i is locally stationary in the sense of Assumption 2.3.1. Similar assumptions are stated in Dahlhaus et al. [2019] and are fulfilled for a large variety of locally stationary processes.

In the same spirit as Example 2.5.1, it is possible to derive uniform rates of convergence for M-estimators of parameters θ in models of locally stationary processes. Furthermore, weak Bahadur representations can be obtained. The following results apply for instance to maximum likelihood estimation of parameters in tvARMA or tvGARCH processes. The main tool is to prove uniform convergence of the corresponding objective functions and its derivatives. Since the rest of the proof is standard, the details are postponed to Section 2.8, Subsection 2.8.5. Let ∇_θ^j denote the j-th derivative with respect to θ. To apply empirical process theory, we ask for the objective functions to be contained in $(L_{\mathcal{F}}, 1, R, C)$-classes in (A1) (see lemma below) and Lipschitz continuous with respect to θ in (A2) (see lemma below).

Lemma 2.5.2 (M-estimation, uniform results). *Let $\Theta \subset \mathbb{R}^d$ be compact and $\theta_0 : [0,1] \to$ interior(Θ). For each $\theta \in \Theta$, let $\ell_\theta : \mathbb{R}^k \to \mathbb{R}$ be some measurable function which is twice continuously differentiable. Let $Z_i = (X_i, ..., X_{i-k+1})$ and define for $v \in [0,1]$,*

$$\hat{\theta}_{n,h}(v) := \arg\min_{\theta \in \Theta} L_{n,h}(v, \theta), \qquad L_{n,h}(v,\theta) := \frac{1}{n} \sum_{i=k}^n K_h\big(\frac{i}{n} - v\big) \cdot \ell_\theta(Z_i).$$

Suppose that there exists $C_\Theta > 0$ such that for $j \in \{0,1,2\}$,

(A1) $\bar{\mathcal{F}}_j = \{\nabla_\theta^j \ell_\theta : \theta \in \Theta\}$ is an $(L_{\mathcal{F}}, 1, R, C)$-class with $R(z) = 1 + |z|_1^{M-1}$ for some $M \ge 1$ and Assumption 2.3.1 for $\bar{\mathcal{F}}_j$ is fulfilled with $s = 1$, $p = \frac{M}{M-1}$.

(A2) for all $z \in \mathbb{R}^k$, $\theta, \theta' \in \Theta$,

$$\left|\nabla_\theta^j \ell_\theta(z) - \nabla_\theta^j \ell_{\theta'}(z)\right|_\infty \le C_\Theta(1 + |z|_1^M) \cdot |\theta - \theta'|_2.$$

(A3) $\theta \mapsto \mathbb{E}\ell_\theta(\tilde{Z}_0(v))$ attains its global minimum in $\theta_0(v)$ with positive definite $I(v) := \mathbb{E}\nabla_\theta^2 \ell_\theta(\tilde{Z}_0(v))$.

Furthermore, suppose that either

- *$\delta_{2M}^X(j) \le \kappa j^{-\alpha}$ with some $\kappa > 0, \alpha > 1$, and $h \gtrsim \big(\frac{\log(n)}{n^{1-\frac{2}{\nu}}}\big)^{\frac{\alpha-1}{\alpha}}$ or*
- *$\delta_{2M}^X(j) \le \kappa \rho^j$ with some $\kappa > 0, \rho \in (0,1)$ and $h \gtrsim \frac{\log(n)^3}{n^{1-\frac{2}{\nu}}}$.*

Define $\tau_n := \sqrt{\frac{\log(n)}{nh}}$ and $B_h := \sup_{v\in[0,1]} |\mathbb{E}\nabla_\theta L_{n,h}(v, \theta_0(v))|$ (the bias). Then, $B_h = O(h^\varsigma)$, and as $nh \to \infty$,

$$\sup_{v\in[\frac{h}{2}, 1-\frac{h}{2}]} \left|\hat{\theta}_{n,h}(v) - \theta_0(v)\right| = O_p\big(\tau_n + B_h\big)$$

and

$$\sup_{v\in[\frac{h}{2}, 1-\frac{h}{2}]} \left|\{\hat{\theta}_{n,h}(v) - \theta_0(v)\} - I(v)^{-1}\nabla_\theta L_{n,h}(v, \theta_0(v))\right| = O_p((\tau_n + h^\varsigma)(\tau_n + B_h)).$$

Remark 2.5.3.

- In the tvAR(1) case $X_i = a(i/n)X_{i-1} + \varepsilon_i$, we can use for instance

$$\ell_\theta(x_1, x_0) = (x_1 - a x_0)^2,$$

  which for $a \in (-1, 1)$ is a $((1, a), 1, |x_0| + |x_1|, (0, 1))$-class.
- With more smoothness assumptions on $\nabla_\theta \ell$ or using a local linear estimation method for $\hat{\theta}_{n,h}$, the bias term B_h can be shown to be of smaller order, for instance $O(h^2)$ (cf. Dahlhaus et al. [2019]).
- The theory derived here can also be used to prove asymptotic properties of M-estimators based on objective functions ℓ_θ which are only almost everywhere differentiable in the Lebesgue sense by following the theory of [van der Vaart, 1998, Chapter 5]. This is of utmost interest for ℓ_θ that have additional analytic properties, such as convexity. Since these properties are also needed in the proofs, we will not discuss this in detail.

We give an easy application of the functional central limit theorem from Theorem 2.2.5 by inspecting a local stationary version of Example 19.25 in van der Vaart [1998].

Example 2.5.4 (Local mean absolute deviation)**.** For fixed $v \in (0,1)$, put $\overline{X_n}(v) := \frac{1}{n}K_h\big(\frac{i}{n} - v\big)X_i$ and define the mean absolute deviation

$$\text{mad}_n(v) := \frac{1}{n}\sum_{i=1}^{n} K_h\big(\frac{i}{n} - v\big)|X_i - \overline{X_n}(v)|.$$

Let Assumption 2.3.1 hold true with $s = 1, p = \infty$. Suppose that $\mathbb{P}(\tilde{X}_0(v) = \mathbb{E}\tilde{X}_0(v)) = 0$ and that for some $\kappa > 0, \alpha > 1$, $\delta_2^X(j) \le \kappa j^{-\alpha}$. We show that if $nh \to \infty$ and $nh^{1+2\varsigma} \to 0$,

$$\sqrt{nh}\big(\text{mad}_n(v) - \mathbb{E}|\tilde{X}_0(v) - \mu|\big) \xrightarrow{d} N(0, \sigma^2) \tag{2.5.3}$$

where $\mu = \mathbb{E}\tilde{X}_0(v)$, G denotes the distribution function of $\tilde{X}_0(v)$ and

$$\begin{aligned}\sigma^2 \quad = \quad \int K(u)^2 du \cdot \sum_{j=0}^{\infty} \text{Cov}\big(&|\tilde{X}_0(v) - \mu| + (2G(\mu) - 1)\tilde{X}_0(v),\\ &|\tilde{X}_j(v) - \mu| + (2G(\mu) - 1)\tilde{X}_j(v)\big).\end{aligned}$$

The result is obtained by using the decomposition

$$\begin{aligned}\sqrt{nh}\big(\text{mad}_n(v) - \mathbb{E}|\tilde{X}_0(v) - \mu|\big) &= \mathbb{G}_n(f_{\overline{X_n}(v)} - f_\mu) + \mathbb{G}_n(f_\mu) + A_n,\\ A_n &= \frac{\sqrt{nh}}{n}\sum_{i=1}^{n} K_h\big(\frac{i}{n} - v\big)\big\{\mathbb{E}|X_i - \theta| - \mathbb{E}|\tilde{X}_0(v) - \mu|\big\}\Big|_{\theta = \overline{X_n}(v)}\end{aligned}$$

where $\Theta = \{\theta \in \mathbb{R} : |\theta - \mu| \le 1\}$ and

$$\mathcal{F} = \{f_\theta(x, u) = \sqrt{h}K_h(u - v)|x - \theta| : \theta \in \Theta\}.$$

By the triangle inequality, $\mathcal{F}$ satisfies Assumption 2.2.3 with $\bar{f}_\theta(x, u) = |x - \theta|$, $R(\cdot) = C_R = 1$, $p = \infty$, $s = 1$ and $\Delta(k) = 2\delta_2^X(k)$. Assumption 2.3.2 is trivially fulfilled since $\bar{f}$ does not depend on u. Since $\mathcal{F}$ is a one-dimensional Lipschitz class, we have $\sup_{n\in\mathbb{N}} \mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,n}) = O(\log(\varepsilon^{-1} \vee 1))$. By Corollary 2.2.5, we obtain that there exists some process $[\mathbb{G}(f_\theta)]_{\theta\in\Theta}$ such that for $h \to 0$, $nh \to \infty$,

$$\big[\mathbb{G}_n(f_\theta)\big]_{\theta\in\Theta} \xrightarrow{d} \big[\mathbb{G}(f_\theta)\big]_{\theta\in\Theta} \qquad \text{in } \ell^\infty(\Theta). \tag{2.5.4}$$

Furthermore by Assumption 2.3.1,

$$\begin{aligned}&\|f_{\overline{X_n}(v)}(X_i) - f_\mu(X_i)\|_2\\ \le\;& \|\overline{X_n}(v) - \mu\|_2 \le \|\overline{X_n}(v) - \mathbb{E}\overline{X_n}(v)\|_2 + \|\mathbb{E}\overline{X_n}(v) - \mu\|_2\\ \le\;& \frac{1}{\sqrt{nh}}\Big(\frac{1}{nh}\sum_{i=1}^{n} K\big(\frac{\frac{i}{n} - v}{h}\big)^2\Big)^{1/2}\sum_{j=0}^{\infty}\delta_2^X(j) + \frac{1}{n}\sum_{i=1}^{n} K_h\big(\frac{i}{n} - v\big)\big|\mathbb{E}X_i - \mathbb{E}\tilde{X}_0(v)\big|\\ =\;& O((nh)^{-1/2} + h^\varsigma).\end{aligned} \tag{2.5.5}$$

By [van der Vaart, 1998, Lemma 19.24], we conclude from (2.5.4) and (2.5.5) that

$$\mathbb{G}_n(f_{\overline{X_n}(v)} - f_\mu) \xrightarrow{p} 0. \tag{2.5.6}$$

in probability.

By Assumption 2.3.1 and bounded variation of K,

$$A_n = \sqrt{nh}\big\{\mathbb{E}|\tilde{X}_0(v) - \theta|\big|_{\theta=\overline{X_n}(v)} - \mathbb{E}|\tilde{X}_0(v) - \mu|\big\} + O_p((nh)^{-1/2} + (nh)^{1/2}h^{-\varsigma}). \tag{2.5.7}$$

Due to $\mathbb{P}(\tilde{X}_0(v) = \mu) = 0$, the function $g(\theta) = \mathbb{E}|\tilde{X}_0(v) - \theta|$ is differentiable in $\theta = \mu$ with derivative $2G(\mu) - 1$. The Delta method delivers

$$\begin{aligned} & \sqrt{nh}\big\{\mathbb{E}|\tilde{X}_0(v) - \theta|\big|_{\theta=\overline{X_n}(v)} - \mathbb{E}|\tilde{X}_0(v) - \mu|\big\} \\ = \;& (2G(\mu) - 1)\sqrt{nh}(\overline{X_n}(v) - \mu) + o_p(1). \end{aligned} \tag{2.5.8}$$

From (2.5.6), (2.5.7) and (2.5.8) we obtain

$$\sqrt{nh}\big(\mathrm{mad}_n(v) - \mathbb{E}|\tilde{X}_0(v) - \mu|\big) = \mathbb{G}_n(f_\mu + (2G(\mu) - 1)\mathrm{id}) + o_p(1).$$

Theorem 2.3.4 now yields (2.5.3).

2.6 Excursus: Large deviation inequalities

A variety of large deviation inequalities using the functional dependence measure have been derived, see for instance Zhang and Wu [2017] and Wu et al. [2013] for Nagaev- and Rosenthal-type inequalities. Here, we present a Bernstein-type inequality for $\mathbb{G}_n(f)$ which can be extended to a large deviation inequality for $\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|$ using a combination of our novel chaining scheme and results in Alexander [1984]. We provide these results to complete the picture of empirical process theory for the functional dependence measure and to show the power of the decomposition (2.8.23), a key approach when proving the maximal inequality (of finite $\mathcal{F}$). In general, however, the derived inequalities are weaker than a combination of Markov's inequality and Theorem 2.4.1. The reason for this mainly lies in the treatment of the first summand in (2.8.23) and the fact that the functional dependence measure is formulated with a moment based norm instead of probabilities. This leads to an additional factor for $V_n(\cdot)$ and $\beta(\cdot)$.

For $q \in \mathbb{N}$, $\nu \geq 2$, define

$$\omega(q) := q^{1/\nu}\log(eq)^{3/2}, \qquad \mathcal{L}(q) = \log\log(e^e q), \qquad \Phi(q) = q\mathcal{L}(q)$$

as well as

$$\tilde{\beta}(q) = \sum_{j=q}^{\infty}\Delta(j)\omega(j)\mathcal{L}(j), \qquad \tilde{V}_n(f) = \|f\|_{2,n} + \sum_{j=1}^{\infty}\min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(j)\omega(j)\}\mathcal{L}(j).$$

With the above quantities, we can formulate the following result.

Theorem 2.6.1 (Bernstein-type large deviation inequality)**.** *Let $\mathcal{F}$ satisfy Assumption 2.4.2. Then there exist universal constants $c_0, c_1 > 0$ such that the following holds true: For each $q \in \{1, ..., n\}$ there exists a set $B_n(q)$ independent of $f \in \mathcal{F}$ such that for all $x > 0$,*

$$\mathbb{P}\Big(\big|\mathbb{G}_n(f)\big| > x, B_n(q)\Big) \le c_0 \exp\Big(-\frac{1}{c_1}\frac{x^2}{\tilde{V}_n(f)^2 + \frac{M\Phi(q)}{\sqrt{n}}x}\Big) \tag{2.6.1}$$

and

$$\mathbb{P}(B_n(q)^c) \le 4\Big(\frac{\mathbb{D}_n^\infty \tilde{\beta}(q)\sqrt{n}}{M\Phi(q)}\Big)^2.$$

Define $\tilde{q}^*(z) := \min\{q \in \mathbb{N} : \tilde{\beta}(q) \le \Phi(q)z\}$. Then for any $y > 0$, $x > 0$,

$$\mathbb{P}\Big(|\mathbb{G}_n(f)| > x, B_n\big(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y})\big)\Big) \le c_0 \exp\Big(-\frac{1}{c_1}\frac{x^2}{\tilde{V}_n(f)^2 + \Phi(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y}))\frac{Mx}{\sqrt{n}}}\Big) \tag{2.6.2}$$

and $\mathbb{P}(B_n(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y}))^c) \le \frac{4}{y^2}$.

Remark 2.6.2. (i) Theorem 2.6.1 mimics the well-known large deviation inequalities from [Rio, 1995, Theorem 5] or Liebscher [1996] in the case of α-mixing sequences.

(ii) The reason for the change of V_n, β, q to $\tilde{V}_n, \tilde{\beta}, \Phi(q)$ in Theorem 2.6.1 compared to Theorem 2.4.1 is due to the arising sums over $l = 1, ..., L$ in the second term and $j = q, q+1, ...$ in the first term $\sum_{j=q}^{\infty} \max_{f\in\mathcal{F}} \frac{1}{\sqrt{n}}\big|S_{n,j+1}(f) - S_{n,j}(f)\big|$ in the decomposition (2.8.23), which forces us to include additional log-factors to obtain convergence. The additional factor $j^{1/\nu}$ that appears in $\tilde{\beta}$ is due to an application of Markov's inequality. It can be argued that this is a relict of the fact that the dependence conditions are stated with moments and not with probabilities as in the case of mixing.

(iii) Theorem 2.6.1 can be seen as an improvement of the Bernstein inequalities given in Doukhan and Neumann [2007] which are only available for random variables with exponential decay (in our setting, the conditions are comparable to $\Delta(k) = O(\exp(k^{-a}))$ for some $a > 0$).

A similar statement is valid in the case of classes $\mathcal{F}$ of noncontinuous functions. We decide to include its discussion here, before we study noncontinuous functions formally. We then need the following analogue of Assumption 3.3.3 (an additional submultiplicativity statement for $\beta(\cdot)$) where $\beta(\cdot)$ is replaced by $\tilde{\beta}(\cdot)$ and q is replaced by $\Phi(q)$.

Assumption 2.6.3. *The sequence $j \mapsto \Delta(j)\omega(j)\mathcal{L}(j)$ is decreasing. There exists some constant $C_{\tilde{\beta}} > 0$ such that $\tilde{\beta}_{norm}(q) := \frac{\tilde{\beta}(q)}{\Phi(q)}$ fulfills for all $q_1, q_2 \in \mathbb{N}$,*

$$\tilde{\beta}_{norm}(q_1 q_2) \le C_{\tilde{\beta}}\tilde{\beta}_{norm}(q_1)\tilde{\beta}_{norm}(q_2).$$

In the following, a constant C_Δ will appear. This constant will be defined later on in Lemma 3.5.1, Chapter 3.

Theorem 2.6.4. *Let $\mathcal{F}$ satisfy the Assumption 3.1.1, 2.6.3. Then there exist universal constants $c_0^\circ, c_1^\circ > 0$ such that the following holds true: For each $q \in \{1, ..., n\}$ there exists a set $B_n^\circ(q)$ independent of $f \in \mathcal{F}$ such that for all $x > 0$,*

$$\mathbb{P}\Big(\big|\mathbb{G}_n(f)\big| > x, B_n^\circ(q)\Big) \le c_0^\circ \exp\Big(-\frac{1}{c_1^\circ}\frac{x^2}{\tilde{V}_n(f)^2 + \frac{M\Phi(q)}{\sqrt{n}}x}\Big) \tag{2.6.3}$$

and

$$\mathbb{P}(B_n^\circ(q)^c) \le [4 + C_\Delta C_{\tilde{\beta}}]\Big(\frac{\sqrt{n}\mathbb{D}_n^\infty}{M}\frac{\tilde{\beta}(q)}{\Phi(q)}\Big)^2.$$

Furthermore, for any $x > 0$, $y > 0$,

$$\mathbb{P}\Big(\big|\mathbb{G}_n(f)\big| > x, B_n^\circ\big(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y})\big)\Big) \le c_0^\circ \exp\Big(-\frac{1}{c_1^\circ}\frac{x^2}{\tilde{V}_n(f)^2 + \tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y})\frac{Mx}{\sqrt{n}}}\Big) \tag{2.6.4}$$

and $\mathbb{P}(B_n^\circ(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y})^c) \le \frac{4+C_\Delta C_{\tilde{\beta}}}{y^2}$.

It is possible to extend Theorem 2.6.1 to an exponential inequality for $\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|$ using a chaining scheme from Alexander [1984], which incorporates an entropy integral of the form $\int_0^\sigma \psi(\varepsilon)\mathcal{W}(1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, \tilde{V}_n))d\varepsilon$ where ψ includes a log-factor (cf. (3.1.5)) and $\mathcal{W}: \mathbb{R} \to \mathbb{R}$ fulfills $\mathbb{H}^{1/2} \le \mathcal{W}(\mathbb{H}) \le \mathbb{H}$, depending on the decay of $\Delta(\cdot)$. Details can be found in Subsection 2.8.6, Theorem 2.8.11. The larger entropy integral comes from the fact that in the proof of Theorem 2.6.1, we can only recover the $\exp(-x)$-part of the Bernstein inequality in the discussion of the first summand in (2.8.23) (cf. (2.8.87) in Section 2.8).

2.7 Concluding remarks

In this chapter, we developed a new empirical process theory for locally stationary processes with the functional dependence measure on classes of smooth functions. We have proven a functional central limit theorem and maximal inequalities. A general empirical process theory for locally stationary processes is a key step in deriving asymptotic and nonasymptotic results for M-estimates or testing based on $\mathbb{L}^2$- or $\mathbb{L}^\infty$-statistics. We provided an example in nonparametric estimation where our theory is applicable. Due to the possibility to analyze the size of the function class and the stochastic properties of the underlying process separately, we conjecture that our theory also permits an extension of various results from i.i.d. to dependent data, such as empirical risk minimization (which will be a later chapter's main topic).

From a technical point of view, the linear and moments based nature of the functional dependence measure has forced us to modify several approaches from empirical process theory for i.i.d. or mixing variables. A main issue was given by the fact that the dependence measure only transfers decay rates for continuous functions. We therefore have provided a new chaining technique which preserves continuity of the arguments of the empirical process.

2.8 Lemmata and proofs of Chapter 2

2.8.1 Proofs of Section 2.3

In the following we provide a proof for the central limit theorem. All lemmata used are given thereafter.

Proof of Theorem 2.3.4. Denote $W_i(f) := f(Z_i, \frac{i}{n})$ and $\mathbb{W}_i := (W_i(f_1), ..., W_i(f_m))'$. Let $a = (a_1, ..., a_m)' \in \mathbb{R}^m \backslash \{0\}$. We use the decomposition

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a'(\mathbb{W}_i - \mathbb{E}\mathbb{W}_i) = \sum_{j=0}^{\infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} a' P_{i-j} \mathbb{W}_i.$$

For fixed $J \in \mathbb{N} \cup \{\infty\}$, set

$$(S_n(J))_{k=1,...,m} := S_n(J) := \sum_{j=0}^{J-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} P_{i-j} \mathbb{W}_i.$$

Then, since $P_{i-j} W_i(f_k)$, $i = 1, ..., n$, is a martingale difference sequence and by Lemma 2.8.3(i),

$$\begin{aligned} \|S_n(\infty)_k - S_n(J)_k\|_2 &\leq \sum_{j=J}^{\infty} \Big\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} P_{i-j} W_i(f_k) \Big\|_2 = \sum_{j=J}^{\infty} \Big(\frac{1}{n} \sum_{i=1}^{n} \|P_{i-j} W_i(f_k)\|_2^2 \Big)^{1/2} \\ &\leq \Big(\frac{1}{n} \sum_{i=1}^{n} D_{f_k,2,n}(\frac{i}{n})^2 \Big)^{1/2} \cdot \sum_{j=J}^{\infty} \Delta(j), \end{aligned}$$

thus

$$\limsup_{J,n\to\infty} \|S_n(\infty)_k - S_n(J)_k\|_2 \leq \sup_{n\in\mathbb{N}} \Big(\frac{1}{n} \sum_{i=1}^{n} D_{f_k,2,n}(\frac{i}{n})^2 \Big)^{1/2} \cdot \limsup_{J\to\infty} \sum_{j=J}^{\infty} \Delta(j) = 0. \quad (2.8.1)$$

Define

$$(S_n^{\circ}(J)_k)_{k=1,...,m} := S_n^{\circ}(J) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n-J+1} \sum_{j=0}^{J-1} P_i \mathbb{W}_{i+j}.$$

Then, we have

$$\begin{aligned} \|S_n^{\circ}(J)_k - S_n(J)_k\|_2 &\leq \sum_{j=0}^{J-1} \| \frac{1}{\sqrt{n}} \sum_{i=1}^{j} P_{i-j} W_i(f_k)\|_2 + \frac{1}{\sqrt{n}} \sum_{j=0}^{J-1} \| \sum_{i=n-J+j+1}^{n} P_{i-j} W_i(f_k)\|_2 \\ &\leq \frac{2J^2}{\sqrt{n}} \cdot \sup_{i=1,...,n+j} \|P_{i-j} W_i(f_k)\|_2 \\ &\leq \frac{2J^2}{\sqrt{n}} \cdot \sup_{i=1,...,n+j} \|f_k(Z_i, \frac{i}{n})\|_2. \end{aligned}$$

By Lemma 2.8.3(i),

$$\sup_{i=1,\dots,n+j} \|f_k(Z_i, \frac{i}{n})\|_2 \le C_{\Delta,2} \cdot D_{2,n}(\frac{i}{n}),$$

which gives

$$\lim_{n\to\infty} \|S_n^\circ(J)_k - S_n(J)_k\|_2 = 0. \tag{2.8.2}$$

Stationary approximation: Put $\tilde{S}_n^\circ(J) = (\tilde{S}_n^\circ(J)_k)_{k=1,\dots,m}$ where

$$\tilde{S}_n^\circ(J)_k := \frac{1}{\sqrt{n}} \sum_{i=1}^{n-J+1} \sum_{j=0}^{J-1} P_i f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n}).$$

Then, we have

$$\begin{aligned} & \|S_n^\circ(J)_k - \tilde{S}_n^\circ(J)_k\|_2 \\ \le & \sum_{j=0}^{J-1} \Big(\frac{1}{n} \sum_{i=1}^{n-J+1} \Big\| P_i f_k(Z_{i+j}, \frac{i+j}{n}) - P_i f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})\Big\|_2^2\Big)^{1/2}. \end{aligned}$$

For each j, k,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n-J+1} \|P_i f_k(Z_{i+j}, \frac{i+j}{n}) - P_i f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})\|_2^2 \\ \le & \frac{2}{n} \sum_{i=1}^{n-J+1} \Big(D_{f_k,n}(\frac{i+j}{n}) - D_{f_k,n}(\frac{i}{n})\Big)^2 \cdot \sup_i \|\bar{f}(Z_{i+j}, \frac{i+j}{n})\|_2^2 \\ & + \frac{2}{n} \sum_{i=1}^{n-J+1} D_{f,n}(\frac{i}{n})^2 \cdot \sup_i \Big\| \bar{f}_k(Z_{i+j}, \frac{i+j}{n}) - \bar{f}_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})]\|_2^2. \end{aligned}$$

By Lemma 2.8.3, we have $\sup_i \|\bar{f}(Z_{i+j}, \frac{i+j}{n})\|_2^2 < \infty$. Since $\frac{1}{\sqrt{n}} D_{f_k,n}(\cdot)$ has bounded variation uniformly in n,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^{n-J+1} \Big(D_{f_k,n}(\frac{i+j}{n}) - D_{f_k,n}(\frac{i}{n})\Big)^2 \\ & \le \sup_{i=1,\dots,n} \frac{1}{\sqrt{n}} D_{f_k,n}(\frac{i}{n}) \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n-J+1} \Big| D_{f_k,n}(\frac{i+j}{n}) - D_{f_k,n}(\frac{i}{n})\Big| \to 0. \end{aligned}$$

By Lemma 2.8.3(ii),

$$\sup_i \Big\| \bar{f}_k(Z_{i+j}, \frac{i+j}{n}) - \bar{f}_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})\Big\|_2 \to 0.$$

We therefore obtain

$$\|S_n^\circ(J)_k - \tilde{S}_n^\circ(J)_k\|_2 \to 0. \tag{2.8.3}$$

Note that

$$M_{i,k} := \frac{1}{\sqrt{n}} \sum_{j=0}^{J} P_i f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n}), \quad i = 1, ..., n,$$

is a martingale difference sequence with respect to $\mathcal{A}_{i-1} := \sigma(\varepsilon_{i-1}, \varepsilon_{i-2}, ...)$ (the σ-algebra generated accordingly) and

$$\tilde{S}_n^\circ(J)_k = \sum_{i=1}^{n-J+1} M_{i,k}.$$

We can therefore apply a central limit theorem for martingale difference sequences to $a'\tilde{S}_n^\circ(J) = \sum_{i=1}^{n-J+1}(\sum_{k=1}^m a_k M_{i,k})$.

The Lindeberg condition: Let $\varsigma > 0$. Iterated application of Lemma 2.8.1(i) yields that there are constants $c_1, c_2 > 0$ only depending on m, J such that

$$\begin{aligned} &\sum_{i=1}^{n-J+1} \mathbb{E}[(\sum_{k=1}^{m} a_k M_{i,k})^2 \mathbb{1}_{\{|\sum_{k=1}^m a_k M_{i,k}| > \varsigma\sqrt{n}\}}] \\ \leq\ & c_1 \sum_{l=0,1} \sum_{j=0}^{J-1} \sum_{k=1}^{m} |a_k|^2 \cdot \frac{1}{n} \sum_{i=1}^{n-J} \mathbb{E}\Big[\mathbb{E}[f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})|\mathcal{A}_{i-l}]^2 \mathbb{1}_{\{|\mathbb{E}[f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})|\mathcal{A}_{i-l}]| > \sqrt{n}\frac{\varsigma}{c_2|a|_\infty}\}}\Big]. \end{aligned}$$

For each l, j, k we have

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n-J} \mathbb{E}\Big[\mathbb{E}[f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})|\mathcal{A}_{i-l}]^2 \mathbb{1}_{\{|\mathbb{E}[f_k(\tilde{Z}_{i+j}(\frac{i}{n}), \frac{i}{n})|\mathcal{A}_{i-l}]| > \sqrt{n}\frac{\varsigma}{c_2|a|_\infty}\}}\Big] \\ =\ & \frac{1}{n} \sum_{i=1}^{n-J} D_{f_k,n}(\frac{i}{n})^2 \mathbb{E}\Big[\mathbb{E}[\bar{f}_k(\tilde{Z}_i(\frac{i}{n}), \frac{i}{n})|\mathcal{A}_{i-l}]^2 \mathbb{1}_{\{|\mathbb{E}[\bar{f}_k(\tilde{Z}_i(\frac{i}{n}), \frac{i}{n})|\mathcal{A}_{i-l}]| > \frac{\sqrt{n}}{\sup_{i=1,\dots,n}|D_{f,n}(\frac{i}{n})|}\frac{\varsigma}{c_2|a|_\infty}\}}\Big] \\ =\ & \frac{1}{n} \sum_{i=1}^{n-J} D_{f_k,n}(\frac{i}{n})^2 \mathbb{E}\Big[\tilde{W}_i(\frac{i}{n})^2 \mathbb{1}_{\{|\tilde{W}_i(\frac{i}{n})| > c_n\}}\Big], \end{aligned} \tag{2.8.4}$$

where we have put

$$\tilde{W}_i(u) := \mathbb{E}[\bar{f}_k(\tilde{Z}_i(u), u)|\mathcal{A}_{i-l}], \qquad c_n := \frac{\sqrt{n}}{\sup_{i=1,\dots,n}|D_{f,n}(\frac{i}{n})|}\frac{\varsigma}{c_2|a|_\infty}.$$

By Lemma 2.8.3(ii), $\tilde{W}_i(u)$ satisfies the assumptions (2.8.8) of Lemma 2.8.2. By assumption, $c_n \to \infty$. With $a_n(u) := D_{f_k,n}(u)^2$, we obtain from Lemma 2.8.2 that (2.8.4) converges to 0, which shows that the Lindeberg condition is satisfied.

Convergence of the variance: We have

$$\begin{aligned} &\sum_{i=1}^{n-J+1} \mathbb{E}[(\sum_{k=1}^{m} M_{i,k})^2|\mathcal{A}_{i-1}] \\ =\ & \sum_{j_1,j_2=0}^{J-1} \sum_{k_1,k_2=1}^{m} a_k a_l \cdot \frac{1}{n} \sum_{i=1}^{n-J+1} D_{f_k,n}(\frac{i}{n}) D_{f_l,n}(\frac{i}{n}) \\ &\qquad \cdot \mathbb{E}\big[P_i \bar{f}_k(\tilde{Z}_{i+j_1}(\frac{i}{n}), \frac{i}{n}) \cdot P_i \bar{f}_l(\tilde{Z}_{i+j_2}(\frac{i}{n}), \frac{i}{n})|\mathcal{A}_{i-1}\big]. \end{aligned}$$

For each j_1, j_2, k_1, k_2 we define

$$\tilde{W}_i(u) := \mathbb{E}\big[P_i\bar{f}_k(\tilde{Z}_{i+j_1}(u),u)\cdot P_i\bar{f}_l(\tilde{Z}_{i+j_2}(u),u)|\mathcal{A}_{i-1}\big], \qquad a_n(u) := D_{f_k,n}(u)D_{f_l,n}(u).$$

Then

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n-J+1} D_{f_k,n}(\frac{i}{n})D_{f_l,n}(\frac{i}{n})\cdot\mathbb{E}\big[P_i\bar{f}_k(\tilde{Z}_{i+j_1}(\frac{i}{n}),\frac{i}{n})\cdot P_i\bar{f}_l(\tilde{Z}_{i+j_2}(\frac{i}{n}),\frac{i}{n})|\mathcal{A}_{i-1}\big] \\ = & \frac{1}{n}\sum_{i=1}^{n-J+1} a_n(\frac{i}{n})\tilde{W}_i(\frac{i}{n}). \end{aligned}$$

By Lemma 2.8.3(i),(ii), we have

$$\begin{aligned} \|\tilde{W}_0(u)-\tilde{W}_0(v)\|_1 &\le \|\bar{f}_k(\tilde{Z}_0(u),u)-\bar{f}_k(\tilde{Z}_0(v),v)\|_2\cdot\|\bar{f}_l(\tilde{Z}_0(u))\|_2 \\ &\qquad +\|\bar{f}_l(\tilde{Z}_0(u),u)-\bar{f}_l(\tilde{Z}_0(v),v)\|_2\cdot\|\bar{f}_k(\tilde{Z}_0(v))\|_2 \\ &\le 2C_{cont}C_{\bar{f}}\cdot|u-v|^{\varsigma s/2}. \end{aligned}$$

Let $A_n := \sup_{i=1,\dots,n}|a_n(\frac{i}{n})|$. Since $\frac{D_{f,n}(\cdot)}{D_{f,n}^{\infty}}$ has bounded variation uniformly in n, it follows that $\frac{a_n(\cdot)}{A_n}$ has bounded variation uniformly in n. From $\frac{D_{f,n}^{\infty}}{\sqrt{n}}\to 0$ we conclude $\frac{A_n}{n}\to 0$.

By assumption and the Cauchy-Schwarz inequality,

$$\sup_n\Big[\frac{1}{n}\sum_{i=1}^n|a_n(\frac{i}{n})|\Big] \le \sup_n\Big(\frac{1}{n}\sum_{i=1}^n D_{f_k,n}(\frac{i}{n})^2\Big)^{1/2}\cdot\Big(\frac{1}{n}\sum_{i=1}^n D_{f_l,n}(\frac{i}{n})^2\Big)^{1/2}<\infty.$$

We have $\sup_n(h_n\cdot A_n)\le\sup_n(h_n^{1/2}D_{f_k,n}^{\infty})\cdot\sup_n(h_n^{1/2}D_{f_l,n}^{\infty})<\infty$, and

$$|v-u|>h_n \quad\Rightarrow\quad D_{f_k,n}(u)=0, D_{f_l,n}(u)=0, \quad\Rightarrow\quad a_n(u)=0.$$

Thus, Lemma 2.8.2(ii) is applicable.
Case $\mathbb{K}=1$: If $u\mapsto\mathbb{E}[P_0\bar{f}_k(\tilde{Z}_{j_1}(u),u)\cdot P_0\bar{f}_l(\tilde{Z}_{j_2}(u),u)]$ has bounded variation, we have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n-J+1} D_{f_k,n}(\frac{i}{n})D_{f_l,n}(\frac{i}{n})\cdot\mathbb{E}\big[P_i\bar{f}_k(\tilde{Z}_{i+j_1}(\frac{i}{n}),\frac{i}{n})\cdot P_i\bar{f}_l(\tilde{Z}_{i+j_2}(\frac{i}{n}),\frac{i}{n})|\mathcal{A}_{i-1}\big] \\ \xrightarrow{p} & \lim_{n\to\infty}\int_0^1 D_{f_k,n}(u)D_{f_l,n}(u)\cdot\mathbb{E}[P_0\bar{f}_k(\tilde{Z}_{j_1}(u),u)\cdot P_0\bar{f}_l(\tilde{Z}_{j_2}(u),u)]du. \end{aligned}$$

and thus

$$\begin{aligned} & \sum_{i=1}^{n-J+1}\mathbb{E}[(\sum_{k=1}^m M_{i,k})^2|\mathcal{A}_{i-1}] \\ \xrightarrow{p} & \sum_{k,l=1}^m a_ka_l\cdot\lim_{n\to\infty}\int_0^1 D_{f_k,n}(u)D_{f_l,n}(u)\cdot\sum_{j_1,j_2=0}^{J-1}\mathbb{E}[P_0\bar{f}_k(\tilde{Z}_{j_1}(u),u)\cdot P_0\bar{f}_l(\tilde{Z}_{j_2}(u),u)]du \\ = & a'\Sigma_{kl}^{(1)}(J)a \end{aligned}$$

Here, for $f, g \in \mathcal{F}$, $\mathbb{E}[P_0\bar{f}(\tilde{Z}_{j_1}(u),u) \cdot P_0\bar{g}(\tilde{Z}_{j_2}(u),u)]$ can be written as

$$\begin{aligned}
& \mathbb{E}[P_0\bar{f}(\tilde{Z}_{j_1}(u),u) \cdot P_0\bar{g}(\tilde{Z}_{j_2}(u),u)] \\
= \; & \mathbb{E}[\mathbb{E}[\bar{f}(\tilde{Z}_{j_1}(u),u)|\mathcal{A}_0]] \cdot \mathbb{E}[\bar{g}(\tilde{Z}_{j_2}(u),u)|\mathcal{A}_0]] \\
& \qquad -\mathbb{E}[\mathbb{E}[\bar{f}(\tilde{Z}_{j_1}(u),u)|\mathcal{A}_{-1}]] \cdot \mathbb{E}[\bar{g}(\tilde{Z}_{j_2}(u),u)|\mathcal{A}_{-1}]],
\end{aligned}$$

which shows that the condition stated in the assumption guarantees the bounded variation of $u \mapsto \mathbb{E}[P_0\bar{f}(\tilde{Z}_{j_1}(u),u) \cdot P_0\bar{g}(\tilde{Z}_{j_2}(u),u)]$.
Case $\mathbb{K} = 2$: If $h_n \to 0$, we obtain similarly

$$\begin{aligned}
& \sum_{i=1}^{n-J+1} \mathbb{E}[(\sum_{k=1}^{m} M_{i,k})^2|\mathcal{A}_{i-1}] \\
\xrightarrow{p} \; & \sum_{k,l=1}^{m} a_k a_l \cdot \lim_{n\to\infty} \int_0^1 D_{f_k,n}(u) D_{f_l,n}(u) du \cdot \sum_{j_1,j_2=0}^{J-1} \mathbb{E}[P_0\bar{f}_k(\tilde{Z}_{j_1}(v),v) \cdot P_0\bar{f}_l(\tilde{Z}_{j_2}(v),v)]du \\
= \; & a'\Sigma_{kl}^{(2)}(J)a.
\end{aligned}$$

By the martingale central limit theorem and (2.8.2), (2.8.3),

$$a'S_n(J) \xrightarrow{d} N(0, a'\Sigma_{kl}^{(\mathbb{K})}(J)a). \tag{2.8.5}$$

Conclusion: For $\mathbb{K} \in \{1, 2\}$, we have

$$a'\Sigma_{kl}^{(\mathbb{K})}(J)a \to a'\Sigma_{kl}^{(\mathbb{K})}(\infty)a \qquad (J \to \infty) \tag{2.8.6}$$

due to

$$\begin{aligned}
& \sum_{j_1,j_2:\max\{j_1,j_2\}\geq J} \|P_0\bar{f}_k(\tilde{Z}_{j_1}(u),u) \cdot P_0\bar{f}_l(\tilde{Z}_{j_2}(u),u)\|_1 \\
\leq \; & \sum_{j_1,j_2:\max\{j_1,j_2\}\geq J} \|P_0\bar{f}_k(\tilde{Z}_{j_1}(u),u)\|_2 \|P_0\bar{f}_l(\tilde{Z}_{j_2}(u),u)\|_2 \to 0 \quad (J \to \infty)
\end{aligned}$$

uniformly in n and

$$\sup_n \int_0^1 |D_{f_k,n}(u) D_{f_l,n}(u)| du \leq \sup_n \Big(\int_0^1 D_{f_k,n}(u)^2 du\Big)^{1/2} \Big(\int_0^1 D_{f_l,n}(u)^2 du\Big)^{1/2} < \infty.$$

By (2.8.1), (2.8.5) and (2.8.6) we have

$$\sum_{j\in\mathbb{Z}} \mathrm{Cov}(\bar{f}_k(\tilde{Z}_0(u),u), \bar{f}_l(\tilde{Z}_j(u),u)) = \sum_{j_1,j_2=0}^{\infty} \mathbb{E}[P_0\bar{f}_k(\tilde{Z}_{j_1}(u),u) \cdot P_0\bar{f}_l(\tilde{Z}_{j_2}(u),u)].$$

Via the Cramer-Wold device the proof is completed.

□

Lemma 2.8.1. *Let $c \in \mathbb{R}$, $c > 0$.*

(i) For $x, y \in \mathbb{R}$, it holds true that

$$(x+y)^2 \mathbb{1}_{\{|x+y|>c\}} \le 8x^2 \mathbb{1}_{\{|x|>\frac{c}{2}\}} + 8y^2 \mathbb{1}_{\{|y|>\frac{c}{2}\}}.$$

(ii) For random variables $W, \tilde{W}$, it holds true that

$$\mathbb{E}[W^2 \mathbb{1}_{\{|W|>c\}}] \le 4\mathbb{E}[(W-\tilde{W})^2] + 4\mathbb{E}[\tilde{W}^2 \mathbb{1}_{\{|\tilde{W}|>\frac{c}{2}\}}].$$

Proof of Lemma 2.8.1. (i) We have

$$\begin{aligned}
(x+y)^2 \mathbb{1}_{\{|x+y|>c\}} &\le 2\big[x^2+y^2\big]\mathbb{1}_{\{|x|>\frac{c}{2} \text{ or } |y|>\frac{c}{2}\}} \\
&\le 2\big[x^2+y^2\big]\big\{2\mathbb{1}_{\{|x|>\frac{c}{2},|y|>\frac{c}{2}\}} + \mathbb{1}_{\{|x|>\frac{c}{2},|y|\le\frac{c}{2}\}} + \mathbb{1}_{\{|x|\le\frac{c}{2},|y|>\frac{c}{2}\}}\big\} \\
&\le 4\big[x^2\mathbb{1}_{\{|x|>\frac{c}{2}\}} + y^2\mathbb{1}_{\{|y|>\frac{c}{2}\}}\big] + 4x^2\mathbb{1}_{\{|x|>\frac{c}{2}\}} + 4y^2\mathbb{1}_{\{|y|>\frac{c}{2}\}} \\
&\le 8x^2\mathbb{1}_{\{|x|>\frac{c}{2}\}} + 8y^2\mathbb{1}_{\{|y|>\frac{c}{2}\}}.
\end{aligned}$$

(ii) We have

$$\begin{aligned}
\mathbb{E}[W^2\mathbb{1}_{\{|W|>c\}}] &\le 2\mathbb{E}[(|W|-\tilde{W})^2\mathbb{1}_{\{|W|>c\}}] + 2\mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|W|>c\}}] \\
&\le 2\mathbb{E}[(W-\tilde{W})^2] + 2\mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|W-\tilde{W}|+|\tilde{W}|>c\}}]. \qquad (2.8.7)
\end{aligned}$$

Furthermore, by Markov's inequality,

$$\begin{aligned}
&\mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|W-\tilde{W}|+|\tilde{W}|>c\}}] \\
\le\;& \mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|W-\tilde{W}|>\frac{c}{2}\}}] + \mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|\tilde{W}|>\frac{c}{2}\}}] \\
\le\;& (\frac{c}{2})^2\mathbb{P}(|W-\tilde{W}|>\frac{c}{2}) + \mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|W-\tilde{W}|>\frac{c}{2}\}}\mathbb{1}_{\{|\tilde{W}|>\frac{c}{2}\}}] + \mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|\tilde{W}|>\frac{c}{2}\}}] \\
\le\;& \mathbb{E}[(W-\tilde{W})^2] + 2\mathbb{E}[\tilde{W}^2\mathbb{1}_{\{|\tilde{W}|>\frac{c}{2}\}}].
\end{aligned}$$

Inserting this inequality into (2.8.7), we obtain the assertion.

□

The following lemma generalizes some results from Dahlhaus et al. [2019] using similar techniques as therein.

Lemma 2.8.2. *Let $q \in \{1,2\}$. Let $\tilde{W}_i(u)$ be a stationary sequence with*

$$\sup_{u\in[0,1]} \|\tilde{W}_0(u)\|_q < \infty, \qquad \|\tilde{W}_0(u) - \tilde{W}_0(v)\|_q \le C_W|u-v|^\varsigma. \qquad (2.8.8)$$

Let $a_n : [0,1] \to \mathbb{R}$ be some sequence of functions with $\limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n |a_n(\frac{i}{n})| < \infty$.

(i) Let $q = 2$ and c_n be some sequence with $c_n \to \infty$. Then,

$$\frac{1}{n}\sum_{i=1}^n |a_n(\frac{i}{n})| \cdot \mathbb{E}[\tilde{W}_i(\frac{i}{n})^2\mathbb{1}_{\{|\tilde{W}_i(\frac{i}{n})|>c_n\}}] \to 0,$$

(ii) *Let $q = 1$. Suppose that there exists $h_n > 0, v \in [0,1]$ such that for all $u \in [0,1]$, $|v-u| > h_n$ implies $a_n(u) = 0$. Set $A_n = \sup_{i=1,\dots,n} |a_n(\frac{i}{n})|$ and suppose that*

$$\sup_{n\in\mathbb{N}}(h_n \cdot A_n) < \infty, \qquad \frac{A_n}{n} \to 0, \qquad \frac{a_n(\cdot)}{A_n} \text{ has bounded variation uniformly in } n.$$

Under the assumption that the limits of the following right hand sides exist, if $u \mapsto \mathbb{E}\tilde{W}_0(u)$ has bounded variation, then

$$\frac{1}{n}\sum_{i=1}^n a_n(\frac{i}{n})\tilde{W}_i(\frac{i}{n}) \xrightarrow{p} \lim_{n\to\infty}\int_0^1 a_n(u)\mathbb{E}\tilde{W}_0(u)du.$$

If $h_n \to 0$, then

$$\frac{1}{n}\sum_{i=1}^n a_n(\frac{i}{n})\tilde{W}_i(\frac{i}{n}) \xrightarrow{p} \lim_{n\to\infty}\int_0^1 a_n(u)du \cdot \mathbb{E}\tilde{W}_0(v).$$

Proof of Lemma 2.8.2. Let $J \in \mathbb{N}$ be fixed and assume that $n \ge 2\cdot 2^J$. For $j \in \{1,...,2^J\}$, Define $I_{j,J,n} := \{i \in \{1,...,n\} : \frac{i}{n} \in (\frac{j-1}{2^J}, \frac{j}{2^J}]\}$. Then $(I_{j,J,n})_j$ forms a decomposition of $\{1,...,n\}$ in the sense that $\sum_{j=1}^{2^J} I_{j,J,n} = \{1,...,n\}$. Since $\frac{i}{n} \in (\frac{j-1}{2^J}, \frac{j}{2^J}] \iff \frac{j-1}{2^J}\cdot n < i \le n\cdot\frac{j-1}{2^J} \le \frac{n}{2^J}$, we conclude that $\frac{n}{2^J} - 1 \le |I_{j,J,n}| \le \frac{n}{2^J}$. Thus, since $n \ge 2\cdot 2^J$,

$$\Big|\frac{I_{j,J,n}|}{n} - \frac{1}{2^J}\Big| \le \frac{1}{n}, \qquad |I_{j,J,n}| \ge \frac{1}{2}\frac{n}{2^J}. \tag{2.8.9}$$

Let w_i, $i \in \mathbb{N}$ be an arbitrary sequence. Then,

$$\begin{aligned}\Big|\frac{1}{n}\sum_{i=1}^n w_i - \frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}} w_i\Big| &\le \sum_{j=1}^{2^J}\Big|\frac{|I_{j,J,n}|}{n} - \frac{1}{2^J}\Big|\cdot\Big|\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}} w_i\Big|\\ &\le \frac{1}{n}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}|w_i|\\ &\le \frac{2^J}{n^2}\sum_{i=1}^n |w_i|. \end{aligned}\tag{2.8.10}$$

(i) Application of (2.8.10) with $w_i = a_n(\frac{i}{n})\mathbb{E}[\tilde{W}_i(\frac{i}{n})^2\mathbb{1}_{\{|\tilde{W}_i(\frac{i}{n})|>c_n\}}]$ yields

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \mathbb{E}[\tilde{W}_i(\frac{i}{n})^2\mathbb{1}_{\{|\tilde{W}_i(\frac{i}{n})|>c_n\}}]\\ \le\;& \frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}\Big[\mathbb{E}[\tilde{W}_i(\frac{i}{n})^2\mathbb{1}_{\{|\tilde{W}_i(\frac{i}{n})|>c_n\}}]\\ &\qquad + \frac{2^J}{n}\cdot\frac{1}{n}\sum_{i=1}^n a_n(\frac{i}{n})\cdot\sup_u\|\tilde{W}_0(u)\|_2^2\Big]. \end{aligned}\tag{2.8.11}$$

By Lemma 2.8.1(ii),

$$
\begin{aligned}
&\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}|a_n(\frac{i}{n})|\cdot\mathbb{E}[\tilde{W}_i(\frac{i}{n})^2\mathbb{1}_{\{|\tilde{W}_i(\frac{i}{n})|>c_n\}}]\\
\leq\ &\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}|a_n(\frac{i}{n})|\cdot\mathbb{E}[\tilde{W}_0(\frac{j}{2^J})^2\mathbb{1}_{\{|\tilde{W}_0(\frac{j}{2^J})|>c_n\}}]\\
&\qquad+\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}|a_n(\frac{i}{n})|\cdot\big\|\tilde{W}_0(\frac{i}{n})-\tilde{W}_0(\frac{j}{2^J})\big\|_2^2\\
\leq\ &\Big[\sup_{j=1,\dots,2^J}\mathbb{E}[\tilde{W}_0(\frac{j}{2^J})^2\mathbb{1}_{\{|\tilde{W}_0(\frac{j}{2^J})|>c_n\}}]+C_W(2^{-J})^{\varsigma}\Big]\\
&\qquad\cdot\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}|a_n(\frac{i}{n})|. \qquad (2.8.12)
\end{aligned}
$$

By (2.8.9),

$$
\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}|a_n(\frac{i}{n})|\leq\frac{2}{n}\sum_{i=1}^{n}|a_n(\frac{i}{n})|.
$$

The dominated convergence theorem delivers

$$
\limsup_{n\to\infty}\mathbb{E}[\tilde{W}_0(\frac{j}{2^J})^2\mathbb{1}_{\{|\tilde{W}_0(\frac{j}{2^J})|>c_n\}}].
$$

Furthermore, $\limsup_{n\to\infty}\frac{2^J}{n}\cdot\sup_u\|\tilde{W}_0(u)\|_2^2=0$. Inserting (2.8.12) into (2.8.11) and applying $\limsup_{n\to\infty}$ as well as $\limsup_{J\to\infty}$ afterwards, yields the assertion.

(ii) Since (2.8.8) also holds true for $\tilde{W}_0(u)$ replaced by $\tilde{W}_0(u)-\mathbb{E}\tilde{W}_0(u)$, we may assume in the following that w.l.o.g. $\mathbb{E}\tilde{W}_0(u)=0$.

By (2.8.10) applied to $w_i=a(\frac{i}{n})W_i(\frac{i}{n})$ we obtain

$$
\begin{aligned}
&\Big\|\frac{1}{n}\sum_{i=1}^{n}a_n(\frac{i}{n})\tilde{W}_i(\frac{i}{n})-\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}a_n(\frac{i}{n})\tilde{W}_i(\frac{i}{n})\Big\|_1\\
\leq\ &\frac{2^J}{n}\cdot\frac{1}{n}\sum_{i=1}^{n}|a_n(\frac{i}{n})|\cdot\sup_u\|W_0(u)\|_1\to 0\quad(n\to\infty). \qquad (2.8.13)
\end{aligned}
$$

We furthermore have

$$\begin{aligned}
&\Big\|\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}a_n(\frac{i}{n})\tilde{W}_i(\frac{i}{n}) \\
&\qquad\qquad -\frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}a_n(\frac{i}{n})\tilde{W}_i(\frac{j-1}{2^J})\Big\|_1 \\
\le\ & \frac{1}{2^J}\sum_{j=1}^{2^J}\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}|a_n(\frac{i}{n})|\cdot\big\|\tilde{W}_0(\frac{i}{n})-\tilde{W}_0(\frac{j-1}{2^J})\big\|_1 \\
\le\ & \frac{2}{n}\sum_{i=1}^{n}|a_n(\frac{i}{n})|\cdot C_W(2^{-J})^{\varsigma}. \qquad (2.8.14)
\end{aligned}$$

Fix $j\in\{1,...,2^J\}$. Put $u_j:=\frac{j-1}{2^J}$ and, for a real-valued positive x, define $[x]:=\max\{k\in\mathbb{N}:k>x\}$. By stationarity, the following equality is valid in distribution,

$$\frac{1}{|I_{j,J,n}|}\sum_{i\in I_{j,J,n}}a_n(\frac{i}{n})\tilde{W}_i(u_j)\overset{d}{=}\frac{1}{|I_{j,J,n}|}\sum_{i=1}^{|I_{j,J,n}|}a_n(\frac{i}{n}+\frac{[u_jn]-1}{n})\tilde{W}_i(u_j). \qquad (2.8.15)$$

Set $\tilde{W}_i(u)^\circ:=\tilde{W}_i(u)\mathbb{1}_{\{\frac{i}{n}+\frac{[u_jn]-1}{n}\in[\underline{r}_n,\overline{r}_n]\}}$. By partial summation and since $\frac{a_n(\cdot)}{A_n}$ has bounded variation B_a uniformly in n,

$$\begin{aligned}
&\frac{1}{|I_{j,J,n}|}\sum_{i=1}^{|I_{j,J,n}|}a_n(\frac{i}{n}+[u_jn]-1)\tilde{W}_i(u_j) \\
=\ & \frac{1}{|I_{j,J,n}|}\sum_{i=1}^{|I_{j,J,n}|-1}\big\{a_n(\frac{i}{n}+[u_jn]-1)-a_n(\frac{i+1}{n}+[u_jn]-1)\big\}\sum_{l=1}^{i}\tilde{W}_l(u_j)^\circ \\
&\quad+\frac{1}{|I_{j,J,n}|}A_n\cdot\sum_{l=1}^{|I_{j,J,n}|}\tilde{W}_l(u_j)^\circ \\
\le\ & \frac{B_a+1}{|I_{j,J,n}|}A_n\cdot\sup_{i=1,...,|I_{j,J,n}|}\Big|\sum_{l=1}^{i}\tilde{W}_l(u_j)^\circ\Big| \qquad (2.8.16)
\end{aligned}$$

By stationarity, we have

$$\begin{aligned}
&\sup_{i=1,...,|I_{j,J,n}|}\Big|\sum_{l=1}^{i}\tilde{W}_l(u_j)^\circ\Big| \\
=\ & \sup_{i=1,...,|I_{j,J,n}|}\Big|\sum_{l=1\vee(\lceil n(v+h_n)\rceil-[u_jn]+1)}^{i\wedge(\lfloor n(v-h_n)\rfloor-[u_jn]+1)}\tilde{W}_l(u_j)\Big|\overset{d}{=}\sup_{i=1,...,m_n}\Big|\sum_{l=1}^{i}\tilde{W}_l(u_j)\Big|,
\end{aligned}$$

since $(|I_{j,J,n}| \wedge (\lfloor n(v+h_n) \rfloor - [u_j n]+1)) - (1 \vee (\lceil n(v-h_n) \rceil - [u_j n]+1)) \leq m_n := 2nh_n$. By assumption, $m_n = \frac{2n}{A_n} \cdot A_n h_n \to \infty$.

Exploiting the ergodic theorem, we have

$$\lim_{m\to\infty} \Big| \frac{1}{m} \sum_{l=1}^{m} \tilde{W}_l(u_j) \Big| = 0 \quad a.s.$$

and especially $(\frac{1}{m}\sum_{l=1}^{m} \tilde{W}_l(u_j))_m$ is bounded almost surely. We derive that

$$\begin{aligned} & \frac{1}{m_n} \sup_{i=1,\ldots,m_n} \Big| \sum_{l=1}^{i} \tilde{W}_l(u_j) \Big| \\ \leq \; & \frac{1}{\sqrt{m_n}} \sup_{i=1,\ldots,\sqrt{m_n}} \Big| \frac{1}{i} \sum_{l=1}^{i} \tilde{W}_l(u_j) \Big| + \sup_{i=\sqrt{m_n}+1,\ldots,m_n} \Big| \frac{1}{i} \sum_{l=1}^{i} \tilde{W}_l(u_j) \Big| \to 0. \end{aligned}$$

We conclude from (2.8.16) that

$$\begin{aligned} & \frac{1}{|I_{j,J,n}|} \sum_{i=1}^{|I_{j,J,n}|} a_n(\frac{i}{n} + [u_j n] - 1) \tilde{W}_i(u_j) \\ \leq \; & 2 \cdot 2^J (B_a + 1) \cdot A_n \cdot \frac{m_n}{n} \cdot \frac{1}{m_n} \sup_{i=1,\ldots,|I_{j,J,n}|} \Big| \sum_{l=1}^{i} \tilde{W}_l(u_j)^{\circ} \Big| \to 0. \quad (2.8.17) \end{aligned}$$

Combining (2.8.13), (2.8.14), (2.8.15) and (2.8.17), taking $\limsup_{n\to\infty}$, $\limsup_{J\to\infty}$, successively, we obtain

$$\frac{1}{n} \sum_{i=1}^{n} a_n(\frac{i}{n}) \big\{ \tilde{W}_i(\frac{i}{n}) - \mathbb{E}\tilde{W}_0(\frac{i}{n}) \big\} \xrightarrow{p} 0.$$

If $u \mapsto \mathbb{E}\tilde{W}_0(u)$ has bounded variation, we have for some intermediate value $\xi_{i,n} \in [\frac{i-1}{n}, \frac{i}{n}]$,

$$\begin{aligned} & \Big| \frac{1}{n} \sum_{i=1}^{n} a_n(\frac{i}{n}) \mathbb{E}\tilde{W}_0(\frac{i}{n}) - \int_0^1 a_n(u) \mathbb{E}\tilde{W}_0(u) du \Big| \\ \leq \; & \frac{1}{n} \sum_{i=1}^{n} \big| a_n(\frac{i}{n}) \mathbb{E}\tilde{W}_0(\frac{i}{n}) - a_n(\xi_{i,n}) \mathbb{E}\tilde{W}_0(\xi_{i,n}) \big| \\ \leq \; & \frac{A_n}{n} \cdot \frac{1}{A_n} \sum_{i=1}^{n} |a_n(\frac{i}{n}) - a_n(\xi_{i,n})| \cdot \sup_u \|\tilde{W}_0(u)\|_1 \\ & + \frac{A_n}{n} \sum_{i=1}^{n} \big| \mathbb{E}\tilde{W}_0(\frac{i}{n}) - \mathbb{E}\tilde{W}_0(\xi_{i,n}) \big| \to 0. \end{aligned}$$

If instead $h_n \to 0$, we have for some intermediate value $\xi_{i,n} \in [\frac{i-1}{n}, \frac{i}{n}]$,

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^{n} a_n(\frac{i}{n})\mathbb{E}\tilde{W}_0(\frac{i}{n}) - \frac{1}{n}\sum_{i=1}^{n} a_n(\frac{i}{n})\mathbb{E}\tilde{W}_0(v)\Big| \\
\leq\ & \frac{1}{n}\sum_{i=1}^{n}|a_n(\frac{i}{n})| \cdot \sup_{|u-v|\leq h_n}\|\tilde{W}_0(u) - \tilde{W}_0(v)\|_1 \to 0.
\end{aligned}$$

Since $\frac{a_n(\cdot)}{A_n}$ has bounded variation uniformly in n,

$$\Big|\frac{1}{n}\sum_{i=1}^{n} a_n(\frac{i}{n}) - \int_0^1 a_n(u)du\Big| \leq \frac{A_n}{n}\cdot\frac{1}{A_n}\sum_{i=1}^{n}|a_n(\frac{i}{n}) - a_n(\xi_{i,n})| \to 0.$$

□

Lemma 2.8.3. *Let $\mathcal{F}$ satisfy Assumptions 2.3.1, 2.3.2 and 2.2.3. Then there exist constants $C_{cont} > 0, C_{\bar{f}} > 0$ such that for any $f \in \mathcal{F}$,*

(i) for any $j \geq 1$,

$$\begin{aligned}
\|P_{i-j}f(Z_i,u)\|_2 &\leq D_{f,n}(u)\Delta(j), \\
\sup_{i=1,\dots,n}\|f(Z_i,u)\|_2 &\leq C_\Delta \cdot D_{f,n}(u),
\end{aligned}$$

$$\sup_{i,u}\|\bar{f}(Z_i,u)\|_2 \leq C_{\bar{f}}, \qquad \sup_{v,u}\|\bar{f}(\tilde{Z}_0(v),u)\|_2 \leq C_{\bar{f}}.$$

(ii)

$$\|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(\frac{i}{n}),u)\|_2 \leq C_{cont}n^{-\varsigma s}, \tag{2.8.18}$$

$$\|\bar{f}(\tilde{Z}_i(v_1),u_1) - \bar{f}(\tilde{Z}_i(v_2),u_2)\|_2 \leq C_{cont}\big(|v_1-v_2|^{\varsigma s} + |u_1-u_2|^{\varsigma s}\big). \tag{2.8.19}$$

Proof of Lemma 2.8.3. (i) If Assumption 2.2.3 is satisfied, we have by Lemma 2.8.4,

$$\|P_{i-j}f(Z_i,u)\|_2 \leq \|f(Z_i,u) - f(Z_i^{*(i-j)},u)\|_2 = \delta_2^{f(Z,u)}(j) \leq D_{f,n}(u)\Delta(j).$$

The second assertion follows from Lemma 2.8.4.

(ii) Let $\bar{C}_R := \sup_{v,u_1,u_2}\|\frac{|\bar{f}(\tilde{Z}_0(v),u_1) - f(\tilde{Z}_0(v),u_2)|)}{|u_1-u_2|^\varsigma}\|_2 < \infty$ (by Assumption 2.3.2) and $C_R := \max\{\sup_{i,u}\|R(Z_i,u)\|_2, \sup_{u,v}\|R(\tilde{Z}_0(v),u)\|_2\}$. Then,

$$\|\bar{f}(\tilde{Z}_i(v),u_1) - \bar{f}(\tilde{Z}_i(v),u_2)\|_2 \leq \bar{C}_R|u_1-u_2|^\varsigma. \tag{2.8.20}$$

We derive

$$\begin{aligned}
\|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)\|_2 &\leq \||Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F},s}}^s(R(Z_i,u) + R(\tilde{Z}_i(v),u)\|_2 \\
&\leq \||Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F},s}}^s\|_{\frac{2p}{p-1}} \\
&\qquad \times\big(\|R(Z_i,u)\|_{2p} + \|R(\tilde{Z}_i(v),u)\|_{2p}\big) \\
&\leq 2C_R\||Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F},s}}^s\|_{\frac{2p}{p-1}}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\| |Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F}},s}^s \|_{\frac{2p}{p-1}} &\le \sum_{l=0}^{\infty} L_{\mathcal{F},l} \| |X_{i-l} - \tilde{X}_{i-l}(v)|^s \|_{\frac{2p}{p-1}} \\
&= \sum_{l=0}^{i} L_{\mathcal{F},l} \| X_{i-l} - \tilde{X}_{i-l}(v) \|_{\frac{2ps}{p-1}}^s \\
&\le \sum_{l=0}^{i} L_{\mathcal{F},l} C_X^s \big(|v - \frac{i}{n}|^{\varsigma} + l^{\varsigma} n^{-\varsigma} \big)^s \\
&\le |v - \frac{i}{n}|^{\varsigma} \cdot C_X |L_{\mathcal{F}}|_1 + n^{-\varsigma} \cdot C_X \sum_{l=0}^{\infty} L_{\mathcal{F},l} l^{\varsigma s} \big\}.
\end{aligned}$$

We obtain with $C_{cont} := 2\bar{C}_R + 2C_R C_X \big\{ |L_{\mathcal{F}}|_1 + \sum_{j=0}^{\infty} L_{\mathcal{F},j} j^{\varsigma s} \big\}$ that

$$\| \bar{f}(Z_i, u) - \bar{f}(\tilde{Z}_i(v), u) \|_2 \le C_{cont} \cdot \Big[|v - \frac{i}{n}|^{\varsigma s} + n^{-\varsigma s} \Big]. \tag{2.8.21}$$

Furthermore, as above,

$$\begin{aligned}
\| f(\tilde{Z}_i(v_1), u) - f(\tilde{Z}_i(v_2), u) \|_2 &\le 2C_R \| |\tilde{Z}_0(v_1) - \tilde{Z}_0(v_2)|_{L_{\mathcal{F}},s}^s \|_{\frac{2p}{p-1}} \\
&\le 2C_R \sum_{l=0}^{i} L_{\mathcal{F},l} \| \tilde{X}_0(v_1) - \tilde{X}_0(v_2) \|_{\frac{2ps}{p-1}}^s \\
&\le 2C_R C_X |L_{\mathcal{F}}|_1 \cdot |v_1 - v_2|^{\varsigma s} \qquad (2.8.22)
\end{aligned}$$

Equation (2.8.21) yields (2.8.18) with $v = \frac{i}{n}$. By (2.8.20) and (2.8.22), we conclude equation (2.8.19).

□

2.8.2 Proofs of Section 2.4.1

We provide an approach to obtain maximal inequalities for sums of random variables $W_i(f)$, $i = 1, ..., n$, indexed by $f \in \mathcal{F}$, by using a decomposition into independent random variables. An approach with similar intentions is presented in [Dedecker and Louhichi, 2002, Section 4.3] for absolutely regular sequences and in Li et al. [2016] for Harris-recurrent Markov chains. For convenience, we abbreviate

$$W_i(f) := f(Z_i, \frac{i}{n})$$

and put $S_n(f) := \sum_{i=1}^n W_i(f)$.

To approximate $W_i(f)$ by independent variables, we use a technique from Wu et al. [2013] which was refined in Zhang and Wu [2017]. This decomposition is much more involved then the ones for Harris-recurrent Markov chains or mixing sequences since no direct coupling method is available. Define

$$W_{i,j}(f) := \mathbb{E}[W_i(f) | \varepsilon_{i-j}, \varepsilon_{i-j+1}, ..., \varepsilon_i], \qquad j \in \mathbb{N},$$

and

$$S_n(f) := \sum_{i=1}^n \{W_i(f) - \mathbb{E}W_i(f)\}, \qquad S_{n,j}(f) := \sum_{i=1}^n \{W_{i,j}(f) - \mathbb{E}W_{i,j}(f)\}.$$

Let $q \in \{1,...,n\}$ be arbitrary. Put $L := \lfloor \frac{\log(q)}{\log(2)} \rfloor$ and $\tau_l := 2^l$ $(l = 0,...,L-1)$, $\tau_L := q$. Then we have

$$W_i(f) = W_i(f) - W_{i,q}(f) + \sum_{l=1}^L (W_{i,\tau_l}(f) - W_{i,\tau_{l-1}}(f)) + W_{i,1}(f)$$

(in the case $q = 1$, the sum in the middle does not appear) and thus

$$S_n(f) = \big[S_n(f) - S_{n,q}(f)\big] + \sum_{l=1}^L \big[S_{n,\tau_l}(f) - S_{n,\tau_{l-1}}(f)\big] + S_{n,1}(f).$$

We write

$$S_{n,\tau_l}(f) - S_{n,\tau_{l-1}}(f) = \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} T_{i,l}(f), \qquad T_{i,l}(f) := \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \big[W_{k,\tau_l}(f) - W_{k,\tau_{l-1}}(f)\big].$$

The random variables $T_{i,l}(f), T_{i',l}(f)$ are independent if $|i - i'| > 1$. This leads to the decomposition

$$\begin{aligned}
\max_{f\in\mathcal{F}} \big|\mathbb{G}_n(f)\big| \quad &\leq \quad \max_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \big|S_n(f) - S_{n,q}(f)\big| \\
&\quad + \sum_{l=1}^L \Big[\max_{f\in\mathcal{F}} \Big|\frac{1}{\sqrt{\frac{n}{\tau_l}}} \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{\sqrt{\tau_l}} T_{i,l}(f)\Big| + \max_{f\in\mathcal{F}} \Big|\frac{1}{\sqrt{\frac{n}{\tau_l}}} \sum_{\substack{i=1 \\ i \text{ odd}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{\sqrt{\tau_l}} T_{i,l}(f)\Big|\Big] \\
&\quad + \max_{f\in\mathcal{F}} \frac{1}{\sqrt{n}} \big|S_{n,1}^W(f)\big| \qquad (2.8.23) \\
=: \quad & A_1 + A_2 + A_3.
\end{aligned}$$

While the first term in (2.8.23) can be made small by assumptions on the dependence of $W_i(f)$ and by the use of a large deviation inequality for martingales in Banach spaces from Pinelis [1994], the second and third term allow for an application of Rosenthal-type bounds due to the independency of the summands $T_{i,l}(f)$ and $W_{i,1}(f)$, respectively. Since the first term in (2.8.23) allows for a stronger bound in terms of n than it is the case for mixing, we can obtain a theory which only needs second moments of $W_i(f) = f(X_i, \frac{i}{n})$. By Assumption 2.2.3, we can show the following results (cf. Lemma 2.8.4 and recall (2.4.1) for the definition of D_n^∞). For each $i = 1,...,n$, $j \in \mathbb{N}$, $s \in \mathbb{N} \cup \{\infty\}$, $f \in \mathcal{F}$,

$$\Big\|\sup_{f\in\mathcal{F}} \big|W_i(f) - W_i(f)^{*(i-j)}\big|\Big\|_2 \quad \leq \quad D_n^\infty(\frac{i}{n})\Delta(j), \qquad (2.8.24)$$

$$\big\|W_i(f) - W_i(f)^{*(i-j)}\big\|_2 \quad \leq \quad |D_{f,n}(\frac{i}{n})| \cdot \Delta(j), \qquad (2.8.25)$$

$$\big\|W_i(f)\big\|_s \quad \leq \quad \big\|f(Z_i, \frac{i}{n})\big\|_s. \qquad (2.8.26)$$

Proof of Theorem 2.4.1. We denote the three terms on the right hand side of (2.8.23) by A_1, A_2, A_3. We now discuss the three terms separately. First, we have

$$\mathbb{E}A_1 \le \sum_{j=q}^{\infty} \frac{1}{\sqrt{n}} \mathbb{E} \max_{f\in\mathcal{F}} \Big| \sum_{i=1}^{n} (W_{i,j+1}(f) - W_{i,j}(f)) \Big|.$$

For fixed j, the sequence

$$\begin{aligned} E_{i,j} := (E_{i,j}(f))_{f\in\mathcal{F}} &= \big((W_{i,j+1}(f) - W_{i,j}(f))\big)_{f\in\mathcal{F}} \\ &= (\mathbb{E}[W_i(f)|\varepsilon_{i-j}, ..., \varepsilon_i] - \mathbb{E}[W_i(f)|\varepsilon_{i-j+1}, ..., \varepsilon_i])_{f\in\mathcal{F}} \end{aligned}$$

is a $|\mathcal{F}|$-dimensional martingale difference vector with respect to $\mathcal{A}^i = \sigma(\varepsilon_{i-j}, \varepsilon_{i-j+1}, ...)$. For a vector $x = (x_f)_{f\in\mathcal{F}}$ and $s \ge 1$, we write $|x|_s := (\sum_{f\in\mathcal{F}} |x_f|^s)^{1/s}$. By [Pinelis, 1994, Theorem 4.1] there exists an absolute constant $c_1 > 0$ such that for $s > 1$,

$$\Big\|\Big| \sum_{i=1}^{n} E_{i,j} \Big|_s \Big\|_2 \le c_1 \Big\{ 2 \Big\| \sup_{i=1,...,n} |E_{i,j}|_s \Big\|_2 + \sqrt{2(s-1)} \Big\| \Big(\sum_{i=1}^{n} \mathbb{E}[|E_{i,j}|_s^2 | \mathcal{A}^{i-1}] \Big)^{1/2} \Big\|_2 \Big\}. \tag{2.8.27}$$

We have

$$\Big\| \sup_{i=1,...,n} |E_{i,j}|_s \Big\|_2 = \Big\| \big(\sup_{i=1,...,n} |E_{i,j}|_s^2 \big)^{1/2} \Big\|_2 \le \Big\| \big(\sum_{i=1}^{n} |E_{i,j}|_s^2 \big)^{1/2} \Big\|_2,$$

therefore both terms in (2.8.27) are of the same order and it is enough to bound the second term in (2.8.27). We have

$$\begin{aligned} \Big\| \Big(\sum_{i=1}^{n} \mathbb{E}[|E_{i,j}|_s^2 | \mathcal{A}^{i-1}] \Big)^{1/2} \Big\|_2 &= \Big\| \sum_{i=1}^{n} \mathbb{E}[|E_{i,j}|_s^2 | \mathcal{A}^{i-1}] \Big\|_1^{1/2} \\ &\le \Big(\sum_{i=1}^{n} \big\| \mathbb{E}[|E_{i,j}|_s^2 | \mathcal{A}^{i-1}] \big\|_1 \Big)^{1/2} \\ &\le \Big(\sum_{i=1}^{n} \big\| |E_{i,j}|_s \big\|_2^2 \Big)^{1/2}. \end{aligned} \tag{2.8.28}$$

Note that

$$\begin{aligned} E_{i,j}(f) &= W_{i,j+1}(f) - W_{i,j}(f) = \mathbb{E}[W_i(f)|\varepsilon_{i-j}, ..., \varepsilon_i] - \mathbb{E}[W_i(f)|\varepsilon_{i-j+1}, ..., \varepsilon_i] \\ &= \mathbb{E}[W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)} | \mathcal{A}_i], \end{aligned} \tag{2.8.29}$$

where $H(\mathcal{F}_i)^{**(i-j)} := H(\mathcal{F}_i^{**(i-j)})$ and $\mathcal{F}_i^{**(i-j)} = (\varepsilon_i, \varepsilon_{i-1}, ..., \varepsilon_{i-j}, \varepsilon_{i-j-1}^*, \varepsilon_{i-j-2}^*, ...)$.

By Jensen's inequality, Lemma 2.8.4 and the fact that $(W_i(f)^{**(i-j)}, W_i(f)^{**(i-j+1)})$

has the same distribution as $(W_i(f), W_i(f)^{*(i-j)})$,

$$\begin{aligned}
\||E_{i,j}|_s\|_2 &= \Big\|\Big(\sum_{f\in\mathcal{F}} |E_{i,j}(f)|^s\Big)^{1/s}\Big\|_2 \\
&\le s^{1/s}\Big\|\sup_{f\in\mathcal{F}}\big|\mathbb{E}[W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}|\mathcal{A}_i]\big|\Big\|_2 \\
&\le e\cdot\Big\|\mathbb{E}\big[\sup_{f\in\mathcal{F}}\big|W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}\big|\,\big|\mathcal{A}_i\big]\Big\|_2 \\
&\le e\cdot\Big\|\sup_{f\in\mathcal{F}}\big|W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}\big|\,\Big\|_2 \\
&= e\cdot\Big\|\sup_{f\in\mathcal{F}}\big|W_i(f) - W_i(f)^{*(i-j)}\big|\,\Big\|_2 \\
&\le e\cdot D_n^\infty(\frac{i}{n})\Delta(j). \qquad (2.8.30)
\end{aligned}$$

Inserting (2.8.30) into (2.8.28) delivers

$$\Big(\sum_{i=1}^n \||E_{i,j}|_s\|_2^2\Big)^{1/2} \le e\Big(\sum_{i=1}^n D_n^\infty(\frac{i}{n})^2\Big)^{1/2}\Delta(j).$$

Inserting this bound into (2.8.27), we obtain

$$\Big\|\Big|\sum_{i=1}^n E_{i,j}\Big|_s\Big\|_2 \le 4ec_1 s^{1/2}n^{1/2}\Big(\frac{1}{n}\sum_{i=1}^n D_n^\infty(\frac{i}{n})^2\Big)^{1/2}\Delta(j).$$

We conclude with $s := 2\vee\log|\mathcal{F}|$ that

$$\begin{aligned}
\mathbb{E}A_1 &\le \frac{1}{\sqrt{n}}\sum_{k=q}^\infty\Big\|\Big|\sum_{i=1}^n E_{i,j}\Big|_s\Big\|_2 \\
&\le 4ec_1\cdot\sqrt{2\vee\log|\mathcal{F}|}\cdot\Big(\frac{1}{n}\sum_{i=1}^n D_n^\infty(\frac{i}{n})^2\Big)^{1/2}\sum_{j=q}^\infty\Delta_p(j) \\
&\le 8ec_1\cdot\sqrt{H}\cdot\mathbb{D}_n^\infty\beta(q). \qquad (2.8.31)
\end{aligned}$$

We now discuss $\mathbb{E}A_2$. If $M_Q, \sigma_Q > 0$ are constants and $Q_i(f)$, $i = 1, ..., m$, mean-zero independent variables (depending on $f\in\mathcal{F}$) with $|Q_i(f)| \le M_Q$ and the upper bound $(\frac{1}{m}\sum_{i=1}^m \|Q_i(f)\|_2^2)^{1/2} \le \sigma_Q$, then there exists some universal constant $c_2 > 0$ such that

$$\mathbb{E}\max_{f\in\mathcal{F}}\frac{1}{\sqrt{m}}\Big|\sum_{i=1}^m\big[Q_i(f) - \mathbb{E}Q_i(f)\big]\Big| \le c_2\cdot\Big(\sigma_Q\sqrt{H} + \frac{M_Q H}{\sqrt{m}}\Big), \qquad (2.8.32)$$

(see e.g. Dedecker and Louhichi [2002] equation (4.3) in Section 4.1 therein).

Note that $(W_{k,j} - W_{k,j-1})_k$ is a martingale difference sequence and $W_{k,\tau_l} - W_{k,\tau_{l-1}} = \sum_{j=\tau_{l-1}+1}^{\tau_l}(W_{k,j} - W_{k,j-1})$. Furthermore, we have

$$\|W_{k,j} - W_{k,j-1}\|_2 \le \|W_k - \mathbb{E}[W_k|\varepsilon_{k-j+1}]\|_2 \le \|W_k\|_2$$

and

$$\begin{aligned}\|W_{k,j}-W_{k,j-1}\|_2 &= \|\mathbb{E}[W_k^{**(k-j+1)}-W_k^{**(k-j+2)}|\mathcal{A}_k]\|_2\\ &\le \|W_k^{**(k-j+1)}-W_k^{**(k-j+2)}\|_2\\ &= \|W_k-W_k^{*(k-j+1)}\|_2=\delta_2^{W_k}(j-1),\end{aligned}$$

thus

$$\|W_{k,j}-W_{k,j-1}\|_2\le\min\{\|W_k\|_2,\delta_2^{W_k}(j-1)\}.$$

We conclude by the elementary inequality $\min\{a_1,b_1\}+\min\{a_2,b_2\}\le\min\{a_1+a_2,b_1+b_2\}$ that

$$\begin{aligned}\|T_{i,l}\|_2 &= \Big\|\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}(W_{k,\tau_l}-W_{k,\tau_{l-1}})\Big\|_2\\ &= \Big\|\sum_{j=\tau_{l-1}+1}^{\tau_l}\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}(W_{k,j}-W_{k,j-1})\Big\|_2\\ &\le \sum_{j=\tau_{l-1}+1}^{\tau_l}\Big\|\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}(W_{k,j}-W_{k,j-1})\Big\|_2\\ &\le \sum_{j=\tau_{l-1}+1}^{\tau_l}\Big(\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}\|W_{k,j}-W_{k,j-1}\|_2^2\Big)^{1/2}\\ &\le \sum_{j=\tau_{l-1}+1}^{\tau_l}\min\Big\{\Big(\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}\|W_k\|_2^2\Big)^{1/2},\Big(\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)}(\delta_2^{W_k}(j-1))^2\Big)^{1/2}\Big\}.\end{aligned}$$

Let us set

$$\sigma_{i,l}:=\Big(\frac{1}{\tau_l}\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}\|W_k\|_2^2\Big)^{1/2},\qquad \Delta_{i,j,l}:=\Big(\frac{1}{\tau_l}\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}\delta_2^{W_k}(j-1)^2\Big)^{1/2}.$$

Then,

$$\begin{aligned}
& \Big(\frac{1}{\frac{n}{\tau_l}} \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{\tau_l} \|T_{i,l}(f)\|_2^2\Big)^{1/2} \\
\leq \; & \Big(\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Big(\sum_{j=\tau_{l-1}+1}^{\tau_l} \min\Big\{\Big(\frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \|W_k\|_2^2\Big)^{1/2}, \\
& \qquad \Big(\frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \delta_2^{W_k}(j-1)^2\Big)^{1/2}\Big\}\Big)^2\Big)^{1/2} \\
= \; & \Big(\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Big(\big(\tau_l - \tau_{l-1}\big)^2 \min\{\sigma_i^2, \Delta_{i,\tau_{l-1}+1,l}^2\}\Big)^{1/2} \\
= \; & \Big(\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \big(\tau_l - \tau_{l-1}\big)^2 \min\{\sigma_{i,l}^2, \Delta_{i,\tau_{l-1}+1,l}^2\}\Big)^{1/2} \\
\leq \; & \big(\tau_l - \tau_{l-1}\big) \cdot \Big(\min\{\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \sigma_{i,l}^2, \frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Delta_{i,\tau_{l-1}+1,l}^2\}\Big)^{1/2} \\
\leq \; & \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\Big(\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \sigma_{i,l}^2\Big)^{1/2}, \Big(\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Delta_{i,\tau_{l-1}+1,l}^2\Big)^{1/2}\} \\
\leq \; & \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\|f\|_{2,n}, \Big(\frac{1}{n} \sum_{i=1}^{n} \delta_2^{W_i}(\tau_{l-1})^2\Big)^{1/2}\} \\
\leq \; & \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\|f\|_{2,n}, \mathbb{D}_n \Delta(\lfloor \frac{j}{2} \rfloor)\}. \qquad (2.8.33)
\end{aligned}$$

With $\frac{1}{\sqrt{\tau_l}}\big|T_{i,l}(f)\big| \leq 2\sqrt{\tau_l}\|f\|_\infty \leq 2\sqrt{\tau_l}M$ and (2.8.32), we obtain

$$\begin{aligned}
& \sum_{l=1}^{L} \Big[\mathbb{E} \max_{f \in \mathcal{F}} \Big| \frac{1}{\sqrt{\frac{n}{\tau_l}}} \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{\sqrt{\tau_l}} T_{i,l}(f)\Big|\Big] \\
\leq \; & c_2 \sum_{l=1}^{L} \Big[\sup_f \Big(\frac{1}{\frac{n}{\tau_l}} \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Big\|\frac{1}{\sqrt{\tau_l}} T_{i,l}(f)\Big\|_2^2\Big)^{1/2} \sqrt{H} + \frac{2\sqrt{\tau_l} M H}{\sqrt{\frac{n}{\tau_l}}}\Big]
\end{aligned}$$

and a similar assertion for the second term (i odd) in A_2. With (2.8.33), we conclude

$$\begin{aligned}\mathbb{E}A_2 &\le \sum_{l=1}^{L}\Big[\mathbb{E}\max_{f\in\mathcal{F}}\frac{1}{\sqrt{\frac{n}{\tau_l}}}\Big|\sum_{1\le i\le\lfloor\frac{n}{\tau_l}\rfloor+1, i\text{ odd}}\frac{1}{\sqrt{\tau_l}}T_{i,l}(f)\Big|\\ &\qquad +\mathbb{E}\max_{f\in\mathcal{F}}\frac{1}{\sqrt{\frac{n}{\tau_l}}}\Big|\sum_{1\le i\le\lfloor\frac{n}{\tau_l}\rfloor+1, i\text{ even}}\frac{1}{\sqrt{\tau_l}}T_{i,l}(f)\Big|\Big]\\ &\le 4c_2\sum_{l=1}^{L}\Big[\Big(\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\max_{f\in\mathcal{F}}\|f\|_{2,n},\mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\Big)\cdot\sqrt{H}\\ &\qquad +\frac{\sqrt{\tau_l}MH}{\sqrt{\lfloor\frac{n}{\tau_l}\rfloor+1}}\Big]. \qquad (2.8.34)\end{aligned}$$

Note that

$$\sum_{l=1}^{L}\frac{\sqrt{\tau_l}}{\sqrt{\lfloor\frac{n}{\tau_l}\rfloor+1}}\le\sum_{l=1}^{L}\frac{\sqrt{\tau_l}}{\sqrt{\frac{n}{\tau_l}}}=\frac{1}{\sqrt{n}}\sum_{l=0}^{L}\tau_l=\frac{1}{\sqrt{n}}\sum_{l=1}^{L-1}2^l\le\frac{1}{\sqrt{n}}(2^L+q)\le\frac{2q}{\sqrt{n}}. \qquad (2.8.35)$$

Furthermore, we have by Lemma 2.8.5 that

$$\begin{aligned}&\sum_{l=1}^{L}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\max_{f\in\mathcal{F}}\|f\|_{2,n},\mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\le\sum_{j=2}^{\infty}\min\{\max_{f\in\mathcal{F}}\|f\|_{2,n},\mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\\ \le\;& 2\bar{V}_n(\max_{f\in\mathcal{F}}\|f\|_{2,n})\\ =\;& 2\max_{f\in\mathcal{F}}\bar{V}_n(\|f\|_{2,n})=2\max_{f\in\mathcal{F}}V_n(f) \qquad (2.8.36)\end{aligned}$$

where

$$\bar{V}_n(x)=x+\sum_{j=1}^{\infty}\min\{x,\mathbb{D}_n\Delta(j)\} \qquad (2.8.37)$$

and the second to last equality holds true due to $x\mapsto\bar{V}_n(x)$ being increasing.

Inserting (2.8.35) and (2.8.36) into (2.8.34), we conclude that for some universal $c_3>0$,

$$\mathbb{E}A_2\le c_3\Big(\sup_{f\in\mathcal{F}}V_n(f)\sqrt{H}+\frac{qMH}{\sqrt{n}}\Big)\le c_2\Big(\sigma\sqrt{H}+\frac{qMH}{\sqrt{n}}\Big). \qquad (2.8.38)$$

Since $S_{n,1}^W=\sum_{i=1}^n W_{i,1}(f)$ is a sum of independent variables with $|W_{i,1}(f)|\le\|f\|_\infty\le M$ and $\|W_{i,0}(f)\|_2\le 2\|f\|_2\le 2V_n(f)\le 2\sigma$, we obtain from (2.8.32) again

$$\mathbb{E}A_3\le c_2\Big(\sigma\sqrt{H}+\frac{MH}{\sqrt{n}}\Big). \qquad (2.8.39)$$

If we insert the bounds (2.8.31), (2.8.38) and (2.8.39) into (2.8.23), we obtain the result (2.4.2).

We now show (2.4.3). If $q^*(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty})\frac{H}{n} \le 1$, we have $q^*(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty}) \in \{1,...,n\}$ and thus by (2.4.2),

$$\begin{aligned}
\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{\sqrt{n}}S_n(f)\Big| &\le c\Big(\sqrt{H}\mathbb{D}_n^\infty\beta\Big(q^*\Big(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty}\Big)\Big) + q^*\Big(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty}\Big)\frac{MH}{\sqrt{n}} + \sigma\sqrt{H}\Big)\\
&\le 2c\Big(q^*\Big(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty}\Big)\frac{MH}{\sqrt{n}} + \sigma\sqrt{H}\Big)\\
&= 2c\Big(\sqrt{n}M\cdot\min\Big\{q^*\Big(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty}\Big)\frac{H}{n},1\Big\} + \sigma\sqrt{H}\Big). \qquad (2.8.40)
\end{aligned}$$

If $q^*(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty})\frac{H}{n} \ge 1$, we note that the simple bound

$$\begin{aligned}
\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{\sqrt{n}}S_n(f)\Big| &\le 2\sqrt{n}M\\
&\le 2c\Big(\sqrt{n}M\min\Big\{q^*\Big(\frac{M\sqrt{H}}{\sqrt{n}\mathbb{D}_n^\infty}\Big)\frac{H}{n},1\Big\} + \sigma\sqrt{H}\Big) \qquad (2.8.41)
\end{aligned}$$

holds true. Putting the two bounds (2.8.40) and (2.8.41) together, we obtain the result (2.4.3).

□

Lemma 2.8.4. *Let Assumption 2.4.2 be fulfilled for some $\nu \ge 2$. Then,*

$$\begin{aligned}
\delta_\nu^{f(Z,u)}(k) &\le |D_{f,n}(u)|\cdot\Delta(k),\\
\sup_i\Big\|\sup_{f\in\mathcal{F}}\big|f(Z_i,u) - f(Z_i^{*(i-j)},u)\big|\Big\|_\nu &\le D_n^\infty(u)\cdot\Delta(k),\\
\sup_i\|f(Z_i,u)\|_\nu &\le |D_{f,n}(u)|\cdot C_\Delta,
\end{aligned}$$

where $C_\Delta := 4d\cdot|L_\mathcal{F}|_1\cdot C_X^s C_R + C_{\bar{f}}$.

Proof of Lemma 2.8.4. We have for each $f\in\mathcal{F}$ and $\nu\ge 2$ that

$$\begin{aligned}
&\sup_i\Big\|\bar{f}(Z_i,u) - \bar{f}(Z_i^{*(i-k)},u)\Big\|_\nu\\
\le\ &\sup_i\Big\||Z_i - Z_i^{*(i-k)}|_{L_{\mathcal{F},s}}^s\big(R(Z_i,u) + R(Z_i^{*(i-k)},u)\big)\Big\|_\nu\\
\le\ &\sup_i\Big\||Z_i - Z_i^{*(i-k)}|_{L_{\mathcal{F},s}}^s\Big\|_{\frac{p}{p-1}\nu}\Big\|R(Z_i,u) + R(Z_i^{*(i-k)},u)\Big\|_{p\nu}\\
\le\ &\sup_i\Big\|\sum_{j=0}^\infty L_{\mathcal{F},j}\big|X_{i-j} - X_{i-j}^{*(i-k)}\big|_\infty^s\Big\|_{\frac{p}{p-1}\nu}\Big(\|R(Z_i,u)\|_{p\nu} + \Big\|R(Z_i^{*(i-k)},u)\Big\|_{p\nu}\Big)\\
\le\ &2dC_R\sum_{j=0}^k L_{\mathcal{F},j}(\delta_{\frac{p}{p-1}\nu s}^X(k-j))^s.
\end{aligned}$$

This shows the first assertion. Due to

$$\sup_{f\in\mathcal{F}} \big|\bar{f}(Z_i,u) - \bar{f}(Z_i^{*(i-k)},u)\big| \le |Z_i - Z_i^{*(i-k)}|_{L_{\mathcal{F}},s}^s \big(R(Z_i,u) + R(Z_i^{*(i-k)},u)\big),$$

the second assertion follows similarly. The last assertion can be derived from

$$|\bar{f}(z,u)| \le |\bar{f}(z,u) - \bar{f}(0,u)| + |\bar{f}(0,u)| \le |z|_{L_{\mathcal{F}},s}^s \cdot (R(z,u) + R(0,u)) + |\bar{f}(0,u)|,$$

which implies

$$\begin{aligned}
\|\bar{f}(Z_i,u)\|_\nu &\le \Big\|\sum_{j=0}^{\infty} L_{\mathcal{F},j}|Z_{i-j}|_\infty^s\Big\|_{\frac{p}{p-1}\nu}\big(\big\|R(Z_i,u)\big\|_{pq} + R(0,u)\big) + |\bar{f}(0,u)| \\
&\le 2d\cdot|L_{\mathcal{F}}|_1\cdot C_X^s\cdot(C_R + |R(0,u)|) + |\bar{f}(0,u)| \\
&\le 4d\cdot|L_{\mathcal{F}}|_1\cdot C_X^s\cdot C_R + C_{\bar{f}}.
\end{aligned}$$

□

Lemma 2.8.5. *Let $\omega(k)$ be an increasing sequence in k. Then, for any $x>0$,*

$$\sum_{j=2}^{\infty} \min\{x, \mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\omega(j) \le 2\sum_{j=1}^{\infty}\min\{x,\mathbb{D}_n\Delta(j)\}\omega(2j+1).$$

Especially in the case $\omega(k)=1$,

$$\sum_{j=2}^{\infty} \min\{x, \mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\} \le 2\sum_{j=1}^{\infty}\min\{x,\mathbb{D}_n\Delta(j)\}.$$

Proof of Lemma 2.8.5. We have

$$\begin{aligned}
&\sum_{j=2}^{\infty} \min\{x, \mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\omega(j) \\
=\ & \sum_{k=1}^{\infty} \min\{x, \mathbb{D}_n\Delta(\lfloor\frac{2k}{2}\rfloor)\}\omega(2k) + \sum_{k=1}^{\infty} \min\{x, \mathbb{D}_n\Delta(\lfloor\frac{2k+1}{2}\rfloor)\}\omega(2k+1) \\
=\ & \sum_{k=1}^{\infty} \min\{x, \mathbb{D}_n\Delta(k)\}\cdot\{\omega(2k)+\omega(2k+1)\} \\
\le\ & 2\sum_{k=1}^{\infty} \min\{x, \mathbb{D}_n\Delta(k)\}\cdot\omega(2k+1).
\end{aligned}$$

□

Proof of Corollary 2.4.3. Let $\sigma := \sup_{n\in\mathbb{N}}\sup_{f\in\mathcal{F}} V_n(f) < \infty$. For $Q\ge 1$, define

$$M_n = \frac{\sqrt{n}}{\sqrt{H}} r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})\mathbb{D}_n^\infty.$$

Let $\bar{F} = \sup_{f\in\mathcal{F}} \bar{f}$, and $F(z,u) = D_n^{\infty}(u)\cdot\bar{F}(z,u)$. Then F is an envelope function of $\mathcal{F}$. We furthermore have

$$\mathbb{P}(\sup_{i=1,\dots,n} F(Z_i,\frac{i}{n}) > M_n) \le \mathbb{P}\Big(\big(\frac{1}{n}\sum_{i=1}^n F(Z_i,\frac{i}{n})^{\nu}\big)^{1/\nu} > \frac{M_n}{n^{1/\nu}}\Big) \le \frac{n}{M_n^{\nu}}\cdot\|F\|_{\nu,n}^{\nu}. \quad (2.8.42)$$

Inserting the bound

$$\|F\|_{\nu,n}^{\nu} = \frac{1}{n}\sum_{i=1}^n D_n^{\infty}(\frac{i}{n})^{\nu}\|\bar{F}(Z_i,\frac{i}{n})\|_{\nu}^{\nu} \le C_{\Delta}^{\nu}\cdot\frac{1}{n}\sum_{i=1}^n D_n^{\infty}(\frac{i}{n})^{\nu} \le C_{\Delta}^{\nu}\cdot(\mathbb{D}_{\nu,n}^{\infty})^{\nu}$$

into (2.8.42) and using $r(\gamma a) \ge \gamma r(a)$ for $\gamma \ge 1, a > 0$ (this is similarly proven as in Lemma 2.8.6), we obtain

$$\begin{aligned}\mathbb{P}(\sup_{i=1,\dots,n} F(Z_i,\frac{i}{n}) > M_n) &\le \Big(\frac{H}{n^{1-\frac{2}{\nu}}r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})^2}\Big)^{\nu/2}\cdot\Big(\frac{C_{\Delta}\mathbb{D}_{\nu,n}^{\infty}}{\mathbb{D}_n^{\infty}}\Big)^{\nu}\\ &\le \frac{1}{Q^{\nu/2}}\Big(\frac{H}{n^{1-\frac{2}{\nu}}r(\frac{\sigma}{\mathbb{D}_n^{\infty}})^2}\Big)^{\nu/2}\cdot\Big(\frac{C_{\Delta}\mathbb{D}_{\nu,n}^{\infty}}{\mathbb{D}_n^{\infty}}\Big)^{\nu}. \quad (2.8.43)\end{aligned}$$

Using the rough bound $\|f\|_{\nu,n} \le \|F\|_{\nu,n}$ and $r(a) \le a$ for $a > 0$ from Lemma 2.8.6, we obtain

$$\begin{aligned}&\max_{f\in\mathcal{F}}\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}[f(Z_i,\frac{i}{n})\mathbb{1}_{\{|f(Z_i,\frac{i}{n})|>M_n\}}] \le \frac{1}{\sqrt{n}M_n^{\nu-1}}\max_{f\in\mathcal{F}}\sum_{i=1}^n \mathbb{E}[|f(Z_i,\frac{i}{n})|^{\nu}]\\ \le\ & \frac{n}{M_n^{\nu}}\cdot\frac{M_n}{\sqrt{n}}\max_{f\in\mathcal{F}}\|f\|_{\nu,n}^{\nu}\\ \le\ & \Big(\frac{C_{\Delta}^2 H}{n^{1-\frac{2}{\nu}}r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^{\infty}})^2}\Big)^{\nu/2}\cdot\frac{\sigma Q^{1/2}}{\sqrt{H}}\cdot\Big(\frac{\mathbb{D}_{\nu,n}^{\infty}}{\mathbb{D}_n^{\infty}}\Big)^{\nu}\\ \le\ & \frac{\sigma}{Q^{\frac{\nu-2}{2}}\sqrt{H}}\Big(\frac{C_{\Delta}^2 H}{n^{1-\frac{2}{\nu}}r(\frac{\sigma}{\mathbb{D}_n^{\infty}})^2}\Big)^{\nu/2}\cdot\Big(\frac{\mathbb{D}_{\nu,n}^{\infty}}{\mathbb{D}_n^{\infty}}\Big)^{\nu}. \quad (2.8.44)\end{aligned}$$

Let us abbreviate

$$C_n := \Big(\frac{C_{\Delta}^2 H}{n^{1-\frac{2}{\nu}}r(\frac{\sigma}{\mathbb{D}_n^{\infty}})^2}\Big)^{\nu/2}\cdot\Big(\frac{\mathbb{D}_{\nu,n}^{\infty}}{\mathbb{D}_n^{\infty}}\Big)^{\nu}.$$

By assumption, $\sup_{n\in\mathbb{N}} C_n < \infty$. By Theorem 2.4.1, (2.8.43) and (2.8.44),

$$\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}}\big|\mathbb{G}_n(f)\big| > Q\sqrt{H}\Big)\\
\leq\;& \mathbb{P}\Big(\max_{f\in\mathcal{F}}\big|\mathbb{G}_n(f)\big| > Q\sqrt{H}, \sup_{i=1,\dots,n}\bar{F}(Z_i,\frac{i}{n}) \leq M\Big)\\
&\qquad +\mathbb{P}(\sup_{i=1,\dots,n} F(Z_i,\frac{i}{n}) > M)\\
\leq\;& \mathbb{P}\Big(\max_{f\in\mathcal{F}}\big|\mathbb{G}_n(\max\{\min\{f,M\},-M\})\big| > Q\sqrt{H}/2\Big)\\
&\qquad +\mathbb{P}\Big(\max_{f\in\mathcal{F}}\big|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}[f(Z_i,\frac{i}{n})\mathbb{1}_{\{|f(Z_i,\frac{i}{n})|>M\}}] > Q\sqrt{H}/2\Big)\\
&\qquad +\mathbb{P}(\sup_{i=1,\dots,n} F(Z_i,\frac{i}{n}) > M)\\
\leq\;& \frac{2c}{Q\sqrt{H}}\Big[\sigma\sqrt{H} + q^*\Big(r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})\Big)r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})\mathbb{D}_n^\infty\Big] + \Big(\frac{1}{Q^{\frac{\nu}{2}}} + \frac{2\sigma}{Q^{\frac{\nu}{2}}H}\Big)C_n\\
\leq\;& \frac{4c\sigma}{Q^{1/2}} + \Big(\frac{1}{Q^{\frac{\nu}{2}}} + \frac{2\sigma}{Q^{\frac{\nu}{2}}H}\Big)C_n.
\end{aligned}$$

Since $\sup_{n\in\mathbb{N}} C_n < \infty$ and σ is independent of n, the assertion follows for $Q \to \infty$. □

Lemma 2.8.6 (Properties of $r(\cdot)$)**.** *The quantity $r(\cdot)$ is well-defined and for each $a > 0$, $\frac{r(a)}{2} \geq r(\frac{a}{2})$ and $r(a) \leq a$.*

Proof. The quantities $q^*(\cdot)$ and $r(\cdot)$ are well-defined since $\beta_{norm}(\cdot)$ is decreasing (at a rate $\ll q^{-1}$), $r \mapsto q^*(r)r$ is increasing (at a rate $\ll r$) and $\lim_{r\downarrow 0} q^*(r)r = 0$.

Let $a > 0$. We show that $r = 2r(\frac{a}{2})$ fulfills $q^*(r)r \leq a$. By definition of $r(a)$, we obtain $r(a) \geq r = 2r(\frac{a}{2})$ which gives the result. Since β_{norm} is decreasing, q^* is decreasing. We conclude that

$$q^*(r)r = 2\cdot q^*(2r(\frac{a}{2}))r(\frac{a}{2}) \leq 2\cdot q^*(r(\frac{a}{2}))r(\frac{a}{2}) \leq 2\cdot\frac{a}{2} = a.$$

The second inequality $r(a) \leq a$ follows from the fact that $q^*(r)r$ is increasing and $q^*(a)a \geq a$. □

2.8.3 A chaining approach which preserves continuity

In this section we provide a chaining approach which preserves continuity of the functions inside the empirical process. Typical chaining approaches work with indicator functions which are not suitable for the application of Theorem 2.4.1. Instead, we replace the indicator functions by suitably chosen truncations. For $m > 0$, define $\varphi_m^{\wedge} : \mathbb{R} \to \mathbb{R}$ and the corresponding "peaky" residual $\varphi_m^{\vee} : \mathbb{R} \to \mathbb{R}$ via

$$\varphi_m^{\wedge}(x) := (x \vee (-m)) \wedge m, \qquad \varphi_m^{\vee}(x) := x - \varphi_m^{\wedge}(x).$$

In the following, assume that for each $j \in \mathbb{N}_0$ there exists a decomposition $\mathcal{F} = \bigcup_{k=1}^{N_j} \mathcal{F}_{jk}$, where $(\mathcal{F}_{jk})_{k=1,\dots,N_j}$, $j \in \mathbb{N}_0$, is a sequence of nested partitions. For each $j \in \mathbb{N}_0$ and $k \in \{1,...,N_j\}$ we choose a fixed element $f_{jk} \in \mathcal{F}_{jk}$. For $j \in \mathbb{N}_0$ we define $\pi_j f := f_{jk}$ if $f \in \mathcal{F}_{jk}$.

Assume furthermore that there exists a sequence $(\Delta_j f)_{j\in\mathbb{N}}$ such that for all $j \in \mathbb{N}_0$, $\sup_{f,g\in\mathcal{F}_{jk}} |f-g| \le \Delta_j f$. Finally, let $(m_j)_{j\in\mathbb{N}_0}$ be a decreasing sequence which will serve as a truncation sequence.

For $j \in \mathbb{N}_0$ we use the decomposition

$$f - \pi_j f = \varphi^{\wedge}_{m_j}(f - \pi_j f) + \varphi^{\vee}_{m_j}(f - \pi_j f).$$

Since

$$\begin{aligned} f - \pi_j f &= f - \pi_{j+1} f + \pi_{j+1} f - \pi_j f \\ &= \varphi^{\wedge}_{m_{j+1}}(f - \pi_{j+1} f) + \varphi^{\vee}_{m_{j+1}}(f - \pi_{j+1} f) \\ &\quad + \varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f) + \varphi^{\vee}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f), \end{aligned} \tag{2.8.45}$$

we can write

$$\varphi^{\wedge}_{m_j}(f - \pi_j f) = \varphi^{\wedge}_{m_{j+1}}(f - \pi_{j+1} f) + \varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f) + R(j), \tag{2.8.46}$$

where

$$R(j) := \varphi^{\wedge}_{m_j}(f - \pi_j f) - \varphi^{\wedge}_{m_j}(\varphi^{\wedge}_{m_{j+1}}(f - \pi_{j+1} f)) - \varphi^{\wedge}_{m_j}(\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f)).$$

To bound $R(j)$, we use (i) of the following elementary Lemma 2.8.7 which is proven at the end of this subsection.

Lemma 2.8.7. *Let y, x, x_1, x_2, x_3 and $m, m' > 0$ be real numbers. Then the following assertions hold true:*

(i) If $|x_1| + |x_2| \le m$, then

$$\left|\varphi^{\wedge}_m(x_1 + x_2 + x_3) - \varphi^{\wedge}_m(x_1) - \varphi^{\wedge}_m(x_2)\right| \le \min\{|x_3|, 2m\}.$$

(ii) $|\varphi^{\wedge}_m(x)| \le \min\{|x|, m\}$ and if $|x| < y$,

$$|\varphi^{\vee}_m(x)| \le \varphi^{\vee}_m(y) \le y \mathbb{1}_{\{y>m\}}.$$

(iii) If $\mathcal{F}$ fulfills Assumption 2.4.2, then Assumption 2.4.2 also holds true for $\{\varphi^{\wedge}_m(f) : f \in \mathcal{F}\}$ and $\{\varphi^{\vee}_m(f) : f \in \mathcal{F}\}$.

Because the partitions are nested, we have $|\pi_{j+1} f - \pi_j f| \le \Delta_j f$. By Lemma 2.8.7 and (2.8.45), we have

$$\begin{aligned} |R(j)| &\le \min\left\{\left|\varphi^{\vee}_{m_{j+1}}(f - \pi_{j+1} f) + \varphi^{\vee}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f)\right|, 2m_j\right\} \\ &\le \min\left\{\left|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1} f)\right|, 2m_j\right\} + \min\left\{\left|\varphi^{\vee}_{m_j - m_{j+1}}(\Delta_j f)\right|, 2m_j\right\}. \end{aligned} \tag{2.8.47}$$

Let $\tau \in \mathbb{N}$. Then, by iterated application of (2.8.46) and linearity of $f \mapsto W_i(f)$, we obtain

$$\begin{aligned}
& \mathbb{G}_n(\varphi^{\wedge}_{m_0}(f - \pi_0 f)) \\
= \; & \mathbb{G}_n(\varphi^{\wedge}_{m_1}(f - \pi_1 f)) + \mathbb{G}_n(\varphi^{\wedge}_{m_0 - m_1}(\pi_1 f - \pi_0 f)) + \mathbb{G}_n(R(0)) \\
= \; & \mathbb{G}_n(\varphi^{\wedge}_{m_\tau}(f - \pi_\tau f)) \\
& + \sum_{j=0}^{\tau-1} \mathbb{G}_n(\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f)) + \sum_{j=0}^{\tau-1} \mathbb{G}_n(R(j)),
\end{aligned} \tag{2.8.48}$$

which in combination with (2.8.47) can now be used for chaining. The following lemma provides the necessary balancing between the truncated versions of $\mathbb{G}_n(f)$ and the rare events excluded. Recall that $H(k) = 1 \vee \log(k)$ as in (1.2.4).

Lemma 2.8.8 (Compatibility lemma)**.** *If $\mathcal{F}$ fulfills $|\mathcal{F}| \leq k$ and Assumption 2.4.2, then* $\sup_{f \in \mathcal{F}} V_n(f) \leq \delta$, $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq m(n, \delta, k)$ *imply*

$$\mathbb{E} \max_{f \in \mathcal{F}} \big|\mathbb{G}_n(f)\big| \leq c(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})\delta\sqrt{H(k)}, \tag{2.8.49}$$

and $\sup_{f \in \mathcal{F}} V_n(f) \leq \delta$ *implies that for each* $\gamma > 0$,

$$\sqrt{n}\|f \mathbb{1}_{\{f > \gamma \cdot m(n,\delta,k)\}}\|_{1,n} \leq \frac{1}{\gamma}\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\delta\sqrt{H(k)}. \tag{2.8.50}$$

Proof of Lemma 2.8.8. For $q \in \mathbb{N}$, put $\beta_{norm}(q) := \frac{\beta(q)}{q}$. By Theorem 2.4.1 and the definition of $r(\cdot)$,

$$\begin{aligned}
\mathbb{E} \max_{f \in \mathcal{F}} \big|\mathbb{G}_n(f)\big| \; & \leq \; c\Big(\delta\sqrt{H(k)} + q^*\Big(\frac{m(n,\delta,k)\sqrt{H(k)}}{\sqrt{n}\mathbb{D}_n^\infty}\Big)\frac{m(n,\delta,k)H(k)}{\sqrt{n}}\Big) \\
& = \; c\Big(\delta\sqrt{H(k)} + \mathbb{D}_n^\infty q^*(r(\frac{\delta}{\mathbb{D}_n}))r(\frac{\delta}{\mathbb{D}_n})\sqrt{H(k)}\Big) \\
& = \; c(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})\delta\sqrt{H(k)}
\end{aligned}$$

which shows (2.8.49). Since

$$\|f(Z_i, \frac{i}{n})\mathbb{1}_{\{f(Z_i,\frac{i}{n}) > \gamma m(n,\delta,k)\}}\|_1 \leq \frac{1}{\gamma m(n,\delta,k)}\|f(Z_i, \frac{i}{n})^2\|_1 = \frac{1}{\gamma m(n,\delta,k)}\|f(Z_i, \frac{i}{n})\|_2^2,$$

for all $f \in \mathcal{F}$ with $V_n(f) \leq \delta$,

$$\sqrt{n}\|f \mathbb{1}_{\{f > \gamma m(n,\delta,k)\}}\|_{1,n} \leq \frac{\sqrt{n}}{\gamma m(n,\delta,k)}\|f\|_{2,n}^2 \leq \frac{1}{\gamma}\frac{\|f\|_{2,n}^2}{\mathbb{D}_n^\infty r(\frac{\delta}{\mathbb{D}_n})}\sqrt{H(k)}. \tag{2.8.51}$$

If $\|f\|_{2,n} \geq \mathbb{D}_n \Delta(1)$, we have

$$V_n(f) = \|f\|_{2,n} + \mathbb{D}_n \sum_{j=1}^\infty \Delta(j) \geq \|f\|_{2,n} + \mathbb{D}_n \beta(1). \tag{2.8.52}$$

In the case $\|f\|_{2,n} < \mathbb{D}_n\Delta(1)$, the fact that $\Delta(\cdot)$ is decreasing implies that $a^* = \max\{j \in \mathbb{N} : \|f\|_{2,n} < \mathbb{D}_n\Delta(j)\}$ is well-defined. We conclude that

$$\begin{aligned} V_n(f) &= \|f\|_{2,n} + \sum_{j=0}^{\infty} \|f\|_{2,n} \wedge (\mathbb{D}_n\Delta(j)) = \|f\|_{2,n} + \sum_{j=1}^{a^*} \|f\|_{2,n} + \mathbb{D}_n \sum_{j=a^*+1}^{\infty} \Delta(j) \\ &= \|f\|_{2,n}(a^*+1) + \mathbb{D}_n\beta(a^*) \geq \|f\|_{2,n}a^* + \beta(a^*). \end{aligned} \tag{2.8.53}$$

Summarizing the results (2.8.52) and (2.8.53), we have

$$V_n(f) \geq \|f\|_{2,n}(a^* \vee 1) + \mathbb{D}_n\beta(a^* \vee 1).$$

We conclude that

$$V_n(f) \geq \min_{a\in\mathbb{N}} \big[\|f\|_{2,n}a + \mathbb{D}_n\beta(a)\big] \geq \|f\|_{2,n}\hat{a} + \mathbb{D}_n\beta(\hat{a}),$$

where $\hat{a} = \arg\min_{j\in\mathbb{N}} \big\{\|f\|_{2,n} \cdot j + \mathbb{D}_n\beta(j)\big\}$.

Since $\delta \geq V_n(f)$, we have $\delta \geq \mathbb{D}_n\beta(\hat{a}) = \mathbb{D}_n\beta_{norm}(\hat{a})\hat{a}$. Thus, $\beta_{norm}(\hat{a}) \leq \frac{\delta}{\mathbb{D}_n\hat{a}}$. By definition of q^*, $q^*(\frac{\delta}{\mathbb{D}_n\hat{a}}) \leq \hat{a}$. Hence, $q^*(\frac{\delta}{\mathbb{D}_n\hat{a}})\frac{\delta}{\mathbb{D}_n\hat{a}} \leq \frac{\delta}{\mathbb{D}_n}$. By definition of $r(\cdot)$, $r(\frac{\delta}{\mathbb{D}_n}) \geq \frac{\delta}{\mathbb{D}_n\hat{a}}$. We conclude with $\|f\|_{2,n} \leq V_n(f) \leq \delta$ that

$$\frac{\|f\|_{2,n}^2}{\mathbb{D}_n^{\infty} r(\frac{\delta}{\mathbb{D}_n})} \leq \frac{\mathbb{D}_n\hat{a}\|f\|_{2,n}^2}{\mathbb{D}_n^{\infty}\delta} \leq \frac{\mathbb{D}_n V_n(f)\|f\|_{2,n}}{\mathbb{D}_n^{\infty}\delta} \leq \frac{\mathbb{D}_n}{\mathbb{D}_n^{\infty}}\|f\|_{2,n} \leq \frac{\mathbb{D}_n}{\mathbb{D}_n^{\infty}}\delta. \tag{2.8.54}$$

Inserting the result into (2.8.51), we finally obtain that for all $f \in \mathcal{F}$ with $V_n(f) \leq \delta$,

$$\sqrt{n}\|f\mathbb{1}_{\{f>\gamma m(n,\delta,k)\}}\|_{1,n} \leq \frac{\sqrt{n}}{\gamma m(n,\delta,k)}\|f\|_{2,n}^2 \leq \frac{1}{\gamma}\frac{\|f\|_{2,n}^2}{\mathbb{D}_n^{\infty} r(\frac{\delta}{\mathbb{D}_n})}\sqrt{H(k)} \leq \frac{1}{\gamma}\frac{\mathbb{D}_n}{\mathbb{D}_n^{\infty}}\delta\sqrt{H(k)}$$

which shows (2.8.50). □

Proof of Lemma 2.8.7. (i) Since $|x_1| + |x_2| \leq m$ implies $|x_1|, |x_2| \leq m$, we have

$$I := \big|\varphi_m^{\wedge}(x_1 + x_2 + x_3) - \varphi_m^{\wedge}(x_1) - \varphi_m^{\wedge}(x_2)\big| = \big|\varphi_m^{\wedge}(x_1 + x_2 + x_3) - x_1 - x_2\big|.$$

Case 1: $x_1 + x_2 + x_3 > m$. Then, since $|x_1| + |x_2| \leq m$, we have $I = |m - x_1 - x_2| = m - x_1 - x_2 < x_3 \leq |x_3|$.
Case 2: $x_1 + x_2 + x_3 \in [-m, m]$. Then $I = |x_1 + x_2 + x_3 - x_1 - x_2| = |x_3|$.
Case 3: $x_1 + x_2 + x_3 < -m$. Then, since $|x_1| + |x_2| \leq m$, we have $I = |-m - x_1 - x_2| = m + x_1 + x_2 < -x_3 \leq |x_3|$.
Furthermore, $I \leq |\varphi_m(x_1 + x_2 + x_3)| + |x_1 + x_2| \leq m + m = 2m$.

(ii) The first assertion is obvious. If $|x| \le y$, we have

$$\begin{aligned}
|\varphi_m^\vee(x)| &= \begin{cases} x - m, & x > m \\ 0, & x \in [-m, m] \\ -x - m, & x < -m \end{cases} \\
&= \begin{cases} |x| - m, & x > m \\ 0, & x \in [-m, m] = (|x| - m)\mathbb{1}_{|x|>m} \\ |x| - m, & x < -m \end{cases} \\
&\le (y - m)\mathbb{1}_{y>m} = (y - m) \vee 0 = (y - m)\mathbb{1}_{\{y-m>0\}} \le y\mathbb{1}_{y>m},
\end{aligned}$$

which shows the second assertion.

(iii) We will show that for all $z, z' \in \mathbb{R}^{\mathbb{N}}$,

$$|\varphi_m^\wedge(f)(z) - \varphi_m^\wedge(f)(z')| \le |f(z) - f(z')|, \quad |\varphi_m^\vee(f)(z) - \varphi_m^\vee(f)(z')| \le |f(z) - f(z')| \tag{2.8.55}$$

from which the assertion follows. For real numbers a_i, b_i, we have

$$\max_i \{a_i\} = \max_i \{a_i - b_i + b_i\} \le \max_i \{a_i - b_i\} + \max_i \{b_i\},$$

which gives $|\max_i\{a_i\} - \max_i\{b_i\}| \le \max_i |a_i - b_i|$. This implies the inequality $|\max\{a, y\} - \max\{a, y'\}| \le |y - y'|$ and therefore

$$\begin{aligned}
|\varphi_m^\wedge(f)(z) - \varphi_m^\wedge(f)(z')| &= |(-m) \vee (f(z) \wedge m) - (-m) \vee (f(z') \wedge m)| \\
&\le |f(z) \wedge m - f(z') \wedge m| \\
&= |(-f(z')) \vee (-m) - (-f(z)) \vee (-m)| \\
&\le |f(z) - f(z')|.
\end{aligned}$$

For the second inequality in (2.8.55), note that

$$\varphi_m^\vee(f)(z) = (f(z) - m) \vee 0 + (f(z) + m) \wedge 0.$$

Therefore,

$$|\varphi_m^\vee(f)(z) - \varphi_m^\vee(f)(z')| = \big|(f(z) - m) \vee 0 - (f(z') - m) \vee 0 + (f(z) + m) \wedge 0 - (f(z') + m) \wedge 0\big|.$$

If $f(z), f(z') \ge m$, then

$$|\varphi_m^\vee(f)(z) - \varphi_m^\vee(f)(z')| \le \big|(f(z) - m) \vee 0 - (f(z') - m) \vee 0\big| \le |f(z) - f(z')|.$$

A similar result is obtained for $f(z), f(z') \le -m$. If $f(z) \ge m$, $f(z') < m$, then

$$\begin{aligned}
& |\varphi_m^\vee(f)(z) - \varphi_m^\vee(f)(z')| \\
\le\ & \big|(f(z) - m) - (f(z') + m) \wedge 0\big| \\
=\ & \begin{cases} |f(z) - f(z') - 2m| = f(z) - f(z') - 2m \le f(z) - f(z'), & f(z') \le -m, \\ |f(z) - m| = f(z) - m \le f(z) - f(z'), & f(z') > -m \end{cases}.
\end{aligned}$$

A similar result is obtained for $f(z) \ge m$, $f(z') \le m$, which proves (2.8.55).

□

2.8.4 Proofs of Section 2.4.2

Proof of Theorem 2.4.4. In the following, we abbreviate $\mathbb{H}(\delta) = \mathbb{H}(\delta, \mathcal{F}, V_n)$ and $\mathbb{N}(\delta) = \mathbb{N}(\delta, \mathcal{F}, V_n)$. We set $\delta_0 = \sigma$ and $\delta_j = 2^{-j}\delta_0$.

For each $j \in \mathbb{N}_0$, we choose a covering by brackets $\mathcal{F}_{jk}^{pre} := [l_{jk}, u_{jk}] \cap \mathcal{F}$, $k = 1, ..., \mathbb{N}(\delta_j)$, such that $V_n(u_{jk} - l_{jk}) \le \delta_j$ and $\sup_{f,g\in\mathcal{F}_{jk}} |f - g| \le u_{jk} - l_{jk} =: \Delta_{jk}$. We may assume w.l.o.g. that $l_{jk}, u_{jk}, \Delta_{jk} \in \mathcal{F}$.

If l_{jk}, u_{jk} do not belong to $\mathcal{F}$, we can simply define new brackets by

$$\tilde{l}_{jk}(z, u) := \inf_{f\in[l_{jk},u_{jk}]} f(z, u), \qquad \tilde{u}_{jk}(z, u) := \sup_{f\in[l_{jk},u_{jk}]} f(z, u)$$

which fulfill $[l_{jk}, u_{jk}] \cap \mathcal{F} = [\tilde{l}_{jk}, \tilde{u}_{jk}] \cap \mathcal{F}$, and

$$|\tilde{l}_{jk}(z, u) - \tilde{l}_{jk}(z', u)| \le \sup_{f\in[l_{jk},u_{jk}]} |f(z, u) - f(z', u)|.$$

Thus, we can add $\tilde{l}_{jk}, \tilde{u}_{jk}$ to $\mathcal{F}$ without changing the bracketing numbers $\mathbb{N}(\varepsilon, \mathcal{F}, \|\cdot\|)$ and the validity of Assumption 2.4.2.

We now inductively construct a new nested sequence of partitions $(\mathcal{F}_{jk})_k$ of $\mathcal{F}$ from $(\mathcal{F}_{jk}^{pre})_k$ in the following way: For each fixed $j \in \mathbb{N}_0$ define

$$\{\mathcal{F}_{jk} : k\} := \{\bigcap_{i=0}^{j} \mathcal{F}_{ik_i}^{pre} : k_i \in \{1, ..., \mathbb{N}(\delta_i)\}, i \in \{0, ..., j\}\}$$

as the intersections of all previous partitions and the j-th partition. Then, $|\{\mathcal{F}_{jk} : k\}| \le N_j := \mathbb{N}(\delta_0) \cdot ... \cdot \mathbb{N}(\delta_j)$. By monotonicity of V_n we have

$$\sup_{f,g\in\mathcal{F}_{jk}} |f - g| \le \Delta_{jk}, \qquad V_n(\Delta_{jk}) \le \delta_j.$$

In each $\mathcal{F}_{jk}$, fix some $f_{jk} \in \mathcal{F}$, and define $\pi_j f := f_{j,\psi_j f}$ where $\psi_j f := \min\{i \in \{1, ..., N_j\} : f \in \mathcal{F}_{ji}\}$. Setting $\Delta_j f := \Delta_{j,\psi_j f}$ and

$$I(\sigma) := \int_0^\sigma \sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V_n)} d\varepsilon,$$

we construct

$$\tau := \min\Big\{j \ge 0 : \delta_j \le \frac{I(\sigma)}{\sqrt{n}}\Big\} \vee 1. \tag{2.8.56}$$

Let

$$m_j := \frac{1}{2} m(n, \delta_j, N_{j+1}),$$

($m(\cdot)$ from Lemma 2.8.8). Choose $M_n = \frac{1}{2} m_0$. We then have

$$\mathbb{E} \sup_{f\in\mathcal{F}} \big|\mathbb{G}_n(f)\big| \le \mathbb{E} \sup_{f\in\mathcal{F}(M_n)} \big|\mathbb{G}_n(f)\big| + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}\big[W_i(F\mathbb{1}_{\{F>M_n\}})\big]$$

where $\mathcal{F}(M_n) := \{\varphi^{\wedge}_{M_n}(f) : f \in \mathcal{F}\}$. Due to Lemma 2.8.7(iii), $\mathcal{F}(M_n)$ still fulfills Assumption 2.4.2.

Since $|f| \le g$ implies $|W_i(f)| \le W_i(g)$ and $\|W_i(g)\|_1 \le \|g(Z_i, \frac{i}{n})\|_1$, it holds true that

$$\begin{aligned}
|\mathbb{G}_n(f)| &\le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big|W_i(f) - \mathbb{E}W_i(f)\big| \\
&\le \mathbb{G}_n(g) + \frac{2}{\sqrt{n}}\sum_{i=1}^{n}\|W_i(g)\|_1 \le \mathbb{G}_n(g) + 2\sqrt{n}\|g\|_{1,n}.
\end{aligned}$$

By (2.8.47) and (2.8.48) and the fact that $\|f - \pi_0 f\|_\infty \le 2M_n \le m_0$, we have the decomposition

$$\begin{aligned}
\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)| &\le \sup_{f\in\mathcal{F}}|\mathbb{G}_n(\pi_0 f)| \\
&\qquad + \sup_{f\in\mathcal{F}}|\mathbb{G}_n(\varphi^{\wedge}_{m_\tau}(f - \pi_\tau f))| \\
&\qquad\qquad + \sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1}f - \pi_j f))\Big| \\
&\qquad\qquad\quad + \sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}|\mathbb{G}_n(R(j))| \\
&\le \sup_{f\in\mathcal{F}}|\mathbb{G}_n(\pi_0 f)| \\
&\qquad + \Big\{\sup_{f\in\mathcal{F}}|\mathbb{G}_n(\varphi^{\wedge}_{m_\tau}(\Delta_\tau f))| + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_\tau f\|_{1,n}\Big\} \\
&\qquad + \sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1}f - \pi_j f))\Big| \\
&\qquad + \sum_{j=0}^{\tau-1}\Big\{\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1}f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad\qquad + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_{j+1}f\mathbb{1}_{\{\Delta_{j+1}f > m_{j+1}\}}\|_{1,n}\Big\} \\
&\qquad + \sum_{j=0}^{\tau-1}\Big\{\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{m_j - m_{j+1}}(\Delta_j f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad\qquad + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_j f\mathbb{1}_{\{\Delta_j f > m_j - m_{j+1}\}}\|_{1,n}\Big\} \\
&=: R_1 + R_2 + R_3 + R_4 + R_5. \qquad (2.8.57)
\end{aligned}$$

We now discuss the terms R_i, $i \in \{1, ..., 5\}$ from (2.8.57). Therefore, put $C_n := c(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}) + \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}$.

Since $\Delta_{jk} = u_{jk} - l_{jk}$ with $l_{jk}, u_{jk} \in \mathcal{F}$, the class $\{\frac{1}{2}\Delta_{jk} : k \in \{1, ..., \mathbb{N}(\delta_j)\}\}$ still fulfills Assumption 2.4.2. We conclude by Lemma 2.8.7(iii) that for arbitrary $m, \tilde{m} > 0$, the classes

$$\{\frac{1}{2}\varphi_m^{\wedge}(\Delta_{jk}) : k \in \{1, ..., \mathbb{N}(\delta_j)\}\},$$
$$\{\frac{1}{2}\min\{\varphi_m^{\vee}(\Delta_{jk}), 2\tilde{m}\} : k \in \{1, ..., \mathbb{N}(\delta_j)\}\},$$
$$\{\frac{1}{2}\varphi_m^{\wedge}(\pi_{j+1}f - \pi_j f) : k \in \{1, ..., \mathbb{N}(\delta_j)\}\}$$

fulfill Assumption 2.4.2.

- Since $|\{\pi_0 f : f \in \mathcal{F}(M_n)\}| \le \mathbb{N}(\delta_0) = \mathbb{N}(\sigma)$, $\|\pi_0 f\|_\infty \le M_n \le m(n, \delta_0, \mathbb{N}(\delta_1))$ and $V_n(\pi_0 f) \le \sigma = \delta_0$ (by assumption, every $f \in \mathcal{F}$ fulfills $V_n(f) \le \sigma$), we have by (2.8.49):
$$\mathbb{E}R_1 = \mathbb{E} \sup_{f \in \mathcal{F}(M_n)} |\mathbb{G}_n(\pi_0 f)| \le C_n \delta_0 \sqrt{1 \vee \log \mathbb{N}(\delta_1)}.$$

- Note that $|\{\varphi_{m_\tau}^{\wedge}(\Delta_\tau f) : f \in \mathcal{F}(M_n)\}| \le N_\tau$. If $g := \varphi_{m_\tau}^{\wedge}(\Delta_\tau f)$, then $\|g\|_\infty \le m_\tau \le m(n, \delta_\tau, N_{\tau+1})$ and $V_n(g) \le V_n(\Delta_\tau f) \le \delta_\tau$. We conclude by (2.8.49) that:
$$\mathbb{E} \sup_{f \in \mathcal{F}(M_n)} |\mathbb{G}_n(\varphi_{m_\tau}^{\wedge}(\Delta_\tau f))| \le C_n \delta_\tau \cdot \sqrt{1 \vee \log N_{\tau+1}}. \tag{2.8.58}$$
As for the second term, we have by definition of τ in (2.8.56) and the Cauchy-Schwarz inequality:
$$\sqrt{n}\|\Delta_\tau f\|_{1,n} \le \sqrt{n}\|\Delta_\tau f\|_{2,n} \le \sqrt{n} V_n(\Delta_\tau f) \le \sqrt{n}\delta_\tau \le I(\sigma). \tag{2.8.59}$$
From (2.8.58) and (2.8.59) we obtain
$$\mathbb{E}R_2 \le C_n \delta_\tau \sqrt{1 \vee \log N_{\tau+1}} + 2 \cdot I(\sigma).$$

- Since the partitions are nested, it holds true that $|\{\varphi_{m_j - m_{j+1}}^{\wedge}(\pi_{j+1}f - \pi_j f) : f \in \mathcal{F}(M_n)\}| \le N_{j+1}$. If $g := \varphi_{m_j - m_{j+1}}^{\wedge}(\pi_{j+1}f - \pi_j f)$, we have $\|g\|_\infty \le m_j - m_{j+1} \le m_j \le m(n, \delta_j, N_{j+1})$ and
$$|g| \le |\pi_{j+1}f - \pi_j f| \le \Delta_j f.$$
Furthermore, $V_n(g) \le V_n(\Delta_j f) \le \delta_j$. We conclude by (2.8.49) that:
$$\mathbb{E}R_3 \le \sum_{j=0}^{\tau-1} \mathbb{E} \sup_{f \in \mathcal{F}(M_n)} |\mathbb{G}_n(\varphi_{m_j - m_{j+1}}^{\wedge}(\pi_{j+1}f - \pi_j f))| \le C_n \sum_{j=0}^{\tau-1} \delta_j \sqrt{1 \vee \log N_{j+1}}.$$

- Note that $|\{\min\{\varphi_{m_{j+1}}^{\vee}(\Delta_{j+1}f), 2m_j\} : f \in \mathcal{F}(M_n)\}| \le N_{j+1}$. If we set $g := \min\{\varphi_{m_{j+1}}^{\vee}(\Delta_{j+1}f), 2m_j\}$, we have $\|g\|_\infty \le 2m_j = m(n, \delta_j, N_{j+1})$ and
$$|g| \le \Delta_{j+1}f.$$

By monotonicity of V_n, $V_n(g) \leq V_n(\Delta_{j+1}f) \leq \delta_{j+1} \leq \delta_j$. We conclude by (2.8.49) that:

$$\sum_{j=0}^{\tau-1} \mathbb{E} \sup_{f\in\mathcal{F}(M_n)} |\mathbb{G}_n(\min\{\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1}f), 2m_j\})| \leq C_n \sum_{j=0}^{\tau-1} \delta_j \sqrt{1 \vee \log N_{j+1}}. \tag{2.8.60}$$

Furthermore, $V_n(\Delta_{j+1}f) \leq \delta_{j+1}$ and $m_{j+1} = \frac{1}{2}m(n, \delta_{j+1}, N_{j+2})$. By (2.8.50), we have

$$\sqrt{n}\|\Delta_{j+1}f\mathbb{1}_{\{\Delta_{j+1}f>m_{j+1}\}}\|_1 \leq 2\delta_{j+1}\sqrt{1 \vee \log N_{j+2}}. \tag{2.8.61}$$

From (2.8.60) and (2.8.61) we obtain

$$\mathbb{E}R_4 \leq (C_n + 4) \sum_{j=0}^{\tau} \delta_j \sqrt{1 \vee \log N_{j+1}}.$$

- It holds true that $|\{\min\{\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f), 2m_j\} : f \in \mathcal{F}(M_n)\}| \leq N_{j+1}$. If $g := \min\{\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f), 2m_j\}$, we have $\|g\|_\infty \leq 2m_j = m(n, \delta_j, N_{j+1})$ and

$$|g| \leq \Delta_j f.$$

Thus, $V_n(g) \leq V_n(\Delta_j f) \leq \delta_j$. We conclude by (2.8.49) that:

$$\sum_{j=0}^{\tau-1} \mathbb{E} \sup_{f\in\mathcal{F}(M_n)} |\mathbb{G}_n(\min\{\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_{j+1}f), 2m_j\})| \leq C_n \sum_{j=0}^{\tau-1} \delta_j \cdot \sqrt{1 \vee \log N_{j+1}}. \tag{2.8.62}$$

Note that $V_n(\Delta_j f) \leq \delta_j$ and

$$\begin{aligned} 2(m_j - m_{j+1}) &= m(n, \delta_j, N_{j+1}) - m(n, \delta_{j+1}, N_{j+2}) \\ &= \mathbb{D}_n^\infty n^{1/2}\Big[\frac{r(\frac{\delta_j}{\mathbb{D}_n})}{\sqrt{1 \vee \log N_{j+1}}} - \frac{r(\frac{\delta_{j+1}}{\mathbb{D}_n})}{\sqrt{1 \vee \log N_{j+2}}}\Big] \\ &\geq \frac{\mathbb{D}_n^\infty n^{1/2}}{\sqrt{1 \vee \log N_{j+1}}}\big[r(\frac{\delta_j}{\mathbb{D}_n}) - r(\frac{\delta_{j+1}}{\mathbb{D}_n})\big] \\ &\geq \frac{1}{2}\frac{\mathbb{D}_n^\infty n^{1/2}}{\sqrt{1 \vee \log N_{j+1}}} r(\frac{\delta_j}{\mathbb{D}_n}) = m_j, \end{aligned}$$

where the last inequality is due to Lemma 2.8.6. By (2.8.50) we have

$$\sqrt{n}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>m_j-m_{j+1}\}}\|_{1,n} \leq \sqrt{n}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>\frac{m_j}{2}\}}\|_{1,n} \leq 4\delta_j\sqrt{1 \vee \log N_{j+1}} \tag{2.8.63}$$

keeping in mind that $m_j = \frac{1}{2}m(n, \delta_j, N_{j+1})$ in the last inequality. From (2.8.62) and (2.8.63) we obtain

$$R_5 \leq (C_n + 8) \sum_{j=0}^{\tau-1} \delta_j \sqrt{1 \vee \log N_{j+1}}.$$

Summarizing the bounds for R_i, $i = 1, ..., 5$, we obtain that for some universal constant $\tilde{c} > 0$,

$$\mathbb{E} \sup_{f \in \mathcal{F}(M_n)} \Big|\mathbb{G}_n(f)\Big| \le \tilde{c} \cdot C_n \Big[\sum_{j=0}^{\tau} \delta_j \sqrt{1 \vee \log N_{j+1}} + I(\sigma)\Big]. \tag{2.8.64}$$

We have $(1 \vee \log N_j)^{1/2} = \Big(1 \vee \sum_{i=0}^{j} \log \mathbb{N}(\delta_i)\Big)^{1/2} \le \Big(\sum_{i=0}^{j}(1 \vee \mathbb{H}(\delta_i))\Big) \le \sum_{i=0}^{j}(1 \vee \mathbb{H}(\delta_i))^{1/2}$, whence

$$\begin{aligned}\sum_{j=0}^{\tau} \delta_j \sqrt{1 \vee \log N_{j+1}} &\le \sum_{j=0}^{\infty} \delta_j \sum_{i=0}^{j} \sqrt{1 \vee \mathbb{H}(\delta_{i+1})} \le \sum_{i=0}^{\infty} \Big(\sum_{j=i}^{\infty} \delta_j\Big) \sqrt{1 \vee \mathbb{H}(\delta_{i+1})} \\ &= 2\sum_{i=0}^{\infty} \delta_i \sqrt{1 \vee \mathbb{H}(\delta_{i+1})} \le 4 \sum_{i=0}^{\infty} \delta_{i+1} \sqrt{1 \vee \mathbb{H}(\delta_{i+1})}. \end{aligned} \tag{2.8.65}$$

Since H is increasing, we obtain

$$\begin{aligned} &\sum_{i=0}^{\infty} \delta_{i+1} \sqrt{1 \vee \mathbb{H}(\delta_{i+1})} \le \sum_{i=0}^{\infty} \delta_i \sqrt{1 \vee \mathbb{H}(\delta_i)} = 2 \sum_{i=0}^{\infty} \delta_{i+1} \sqrt{1 \vee \mathbb{H}(\delta_i)} \\ = \;& 2 \sum_{i=0}^{\infty} \int_{\delta_{i+1}}^{\delta_i} \sqrt{1 \vee \mathbb{H}(\delta_i)} d\varepsilon \\ \le \;& 2 \sum_{i=0}^{\infty} \int_{\delta_{i+1}}^{\delta_i} \sqrt{1 \vee \mathbb{H}(\varepsilon)} d\varepsilon = 2 \int_0^{\sigma} \sqrt{1 \vee \mathbb{H}(\varepsilon)} d\varepsilon = 2 \cdot I(\sigma). \end{aligned} \tag{2.8.66}$$

Inserting (2.8.66) into (2.8.65) and then into (2.8.64), yields the statement. □

Proof of Corollary 2.4.5. Define $\tilde{\mathcal{F}} := \{f - g : f, g \in \mathcal{F}\}$. It can easily be seen that $\mathbb{N}(\varepsilon, \tilde{\mathcal{F}}, V_n) \le \mathbb{N}(\frac{\varepsilon}{2}, \mathcal{F}, V_n)^2$ (cf. van der Vaart [1998], Theorem 19.5), thus

$$\mathbb{H}(\varepsilon, \tilde{\mathcal{F}}, V_n) \le 2\mathbb{H}(\frac{\varepsilon}{2}, \mathcal{F}, V_n) \tag{2.8.67}$$

Let $\sigma > 0$. Define

$$F(z, u) := 2D_n^{\infty}(u) \cdot \bar{F}(z, u), \qquad \bar{F}(z, u) := \sup_{f \in \mathcal{F}} |\bar{f}(z, u)|.$$

Then obviously, F is an envelope function of $\tilde{\mathcal{F}}$.

By Markov's inequality, Theorem 2.4.4 and (2.8.67),

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{V_n(f-g)\le\sigma,\, f,g\in\mathcal{F}} |\mathbb{G}_n(f)-\mathbb{G}_n(g)| \ge \eta\Big)\\
\le\;& \frac{1}{\eta}\mathbb{E}\sup_{V_n(f-g)\le\sigma,\, f,g\in\mathcal{F}} |\mathbb{G}_n(f)-\mathbb{G}_n(g)|\\
=\;& \frac{1}{\eta}\mathbb{E}\sup_{\tilde f\in\tilde{\mathcal{F}}, V_n(\tilde f)\le\sigma} |\mathbb{G}_n(\tilde f)|\\
\le\;& \frac{\tilde c}{\eta}\Big[(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}+\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty})\int_0^\sigma \sqrt{1\vee\mathbb{H}(\varepsilon,\tilde{\mathcal{F}},V_n)}d\varepsilon + \sqrt{n}\big\|F\mathbb{1}_{\{F>\frac{1}{4}m(n,\sigma,\mathbb{N}(\frac{\sigma}{2}))\}}\big\|_1\Big]\\
\le\;& \frac{\tilde c}{\eta}\Big[2\sqrt{2}(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}+\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty})\int_0^{\sigma/2} \sqrt{1\vee\mathbb{H}(u,\mathcal{F},V_n)}du\\
&\qquad + \frac{4\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}{r(\frac{\sigma}{\mathbb{D}_n})}\big\|F^2\mathbb{1}_{\{F>\frac{1}{4}n^{1/2}\frac{r(\sigma)}{\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}\}}\big\|_{1,n}\Big].
\end{aligned}$$

The first term converges to 0 by (2.4.7) and (2.4.8) for $\sigma\to 0$ (uniformly in n).

We now discuss the second term. The continuity conditions from Assumption 2.4.2 and Assumption 2.3.2 transfer to $\bar F$ by the inequality

$$|\bar F(z_1,u_1)-\bar F(z_2,u_2)| = |\sup_{f\in\mathcal{F}}\bar f(z_1,u_1)-\sup_{f\in\mathcal{F}}\bar f(z_2,u_2)| \le \sup_{f\in\mathcal{F}}|f(z_1,u_1)-f(z_2,u_2)|$$

We therefore have, as in Lemma 2.8.3(ii), for all $u,u_1,u_2,v_1,v_2\in[0,1]$,

$$\|\bar F(Z_i,u)-\bar F(\tilde Z_i(\frac{i}{n}),u)\|_2 \le C_{cont}\cdot n^{-\alpha s}, \tag{2.8.68}$$

$$\|\bar F(Z_i(v_1),u_1)-\bar F(\tilde Z_i(v_2),v_2)\|_2 \le C_{cont}\cdot\big(|v_1-v_2|^{\alpha s}+|u_1-u_2|^{\alpha s}\big). \tag{2.8.69}$$

Put $c_n = \frac{1}{8}\frac{n^{1/2}}{\sup_{i=1,\dots,n}D_n^\infty(\frac{i}{n})}\frac{r(\sigma)}{\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}$. Then by Lemma 2.8.1(ii) and (2.8.68),

$$\begin{aligned}
&\|F^2\mathbb{1}_{\{F>\frac{1}{4}n^{1/2}\frac{r(\sigma)}{\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}\}}\|_{1,n}\\
\le\;& \frac{4}{n}\sum_{i=1}^n D_n^\infty(\frac{i}{n})^2\cdot\mathbb{E}\Big[\bar F(Z_i,\frac{i}{n})^2\mathbb{1}_{\{|\bar F(Z_i,\frac{i}{n})|>c_n\}}\Big]\\
\le\;& \frac{16}{n}\sum_{i=1}^n D_n^\infty(\frac{i}{n})^2\cdot\mathbb{E}\Big[\bar F(\tilde Z_i(\frac{i}{n}),\frac{i}{n})^2\mathbb{1}_{\{|\bar F(\tilde Z_i(\frac{i}{n}),\frac{i}{n})|>c_n\}}\Big]\\
&\qquad +16C_{cont}\cdot n^{-\alpha s}\cdot(\mathbb{D}_n^\infty)^2.
\end{aligned} \tag{2.8.70}$$

Put $\tilde W_i(u) := \bar F(\tilde Z_i(u),u)$ and $a_n(u) := (D_n^\infty(u))^2$. By (2.8.69), $\|\tilde W_i(u_1)-\tilde W_i(u_2)\|_2 \le 2C_{cont}|u_1-u_2|^{\alpha s}$. By the assumptions on $D_{f,n}(\cdot)$ we can derive that $c_n\to\infty$ and $\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n|a_n(\frac{i}{n})| = \limsup_{n\to\infty}(\mathbb{D}_n^\infty)^2<\infty$. We conclude with Lemma 2.8.2(i) that

$$\frac{16}{n}\sum_{i=1}^n D_n^\infty(\frac{i}{n})^2\cdot\mathbb{E}\Big[\bar F(\tilde Z_i(\frac{i}{n}),\frac{i}{n})^2\mathbb{1}_{\{|\bar F(\tilde Z_i(\frac{i}{n}),\frac{i}{n})|>c_n\}}\Big]\to 0,$$

that is, the first summand in (2.8.70) tends to 0. Since $\limsup_{n\to\infty} \mathbb{D}_n^\infty < \infty$, we obtain that (2.8.70) tends to 0. $\square$

2.8.5 Proofs of Section 2.5

Proof of Lemma 2.5.2. Put $D_{v,n}(u) := \sqrt{h}K_h(u-v)$. From (A1) and Assumption 2.3.1 we obtain $\Delta(k) = O(\delta_{2M}^X(k))$, $C_R = 1 + k\max\{C_X, 1\}^{2M}$.

Since K is Lipschitz continuous and (A2) holds true, we have

$$\begin{aligned}
& \sup_{|v-v'|\le n^{-3}, |\theta-\theta'|_2\le n^{-3}} \big|\big(\nabla_\theta^j L_{n,h}(v,\theta) - \mathbb{E}\nabla_\theta^j L_{n,h}(v,\theta)\big) \\
& \qquad\qquad -\big(\nabla_\theta^j L_{n,h}(v',\theta') - \mathbb{E}\nabla_\theta^j L_{n,h}(v',\theta')\big)\big|_\infty \\
\le\ & \sup_{|v-v'|\le n^{-3}, |\theta-\theta'|_2\le n^{-3}} \frac{C_R}{h^2}\big[L_K|v-v'| + C_\Theta|\theta-\theta'|_2\big] \\
& \qquad\qquad \times\frac{1}{n}\sum_{i=k}^n \big(1 + |Z_i|_1^M + \mathbb{E}|Z_i|_1^M\big) \\
=\ & O_p(n^{-1}).
\end{aligned}$$

Let Θ_n be a grid approximation of Θ such that for any $\theta \in \Theta$, there exists some $\theta' \in \Theta_n$ such that $|\theta - \theta'|_2 \le n^{-3}$. Since $\Theta \subset \mathbb{R}^d$, it is possible to choose Θ_n such that $|\Theta_n| = O(n^{-6d})$. Furthermore, define $\mathcal{V}_n := \{in^{-3} : i = 1, ..., n\}$ as an approximation of $[0,1]$.

As in Example 2.5.1, Corollary 2.4.3 applied to

$$\mathcal{F}_j' = \{f_{v,\theta} : \theta \in \Theta_n, v \in \mathcal{V}_n\}$$

yields for $j \in \{0,1,2\}$ that

$$\sup_{v\in[\frac{h}{2},1-\frac{h}{2}]} \big|\nabla_\theta^j L_{n,h}(v,\theta) - \mathbb{E}\nabla_\theta^j L_{n,h}(v,\theta)\big|_\infty = O_p\big(\tau_n\big). \tag{2.8.71}$$

Put $\tilde{L}_{n,h}(v,\theta) := \frac{1}{n}\sum_{i=1}^n K_h(i/n - v)\ell_\theta(\tilde{Z}_i(v))$. With (A1) it is easy to see that

$$\begin{aligned}
& \big|\mathbb{E}\nabla_\theta^j L_{n,h}(v,\theta) - \mathbb{E}\nabla_\theta^j \tilde{L}_{n,h}(v,\theta)\big|_\infty \\
\le\ & \frac{d_\Theta^j C_R}{n}\sum_{i=1}^n |K_h(i/n - v)| \cdot \||Z_i - \tilde{Z}_i(v)|_1\|_M \\
& \qquad\qquad \times\big(1 + \||Z_i|_1\|_M^{M-1} + \||\tilde{Z}_i(v)|_1\|_M^{M-1}\big) \\
\le\ & d_\Theta^j C_R |K|_\infty C_X(1 + 2C_X^{M-1})\big(n^{-1} + h\big).
\end{aligned} \tag{2.8.72}$$

Finally, since K has bounded variation and $\int K(u)du = 1$, uniformly in $v \in [\frac{h}{2}, 1-\frac{h}{2}]$ it holds true that

$$\mathbb{E}\nabla_\theta^j \tilde{L}_{n,h}(v,\theta) = \frac{1}{n}\sum_{i=1}^n K_h(i/n - v)\mathbb{E}\nabla_\theta^j \ell_\theta(\tilde{Z}_1(v)) = \mathbb{E}\nabla_\theta^j \ell_\theta(\tilde{Z}_1(v)) + O((nh)^{-1}). \tag{2.8.73}$$

From (2.8.71), (2.8.72) and (2.8.73) we obtain

$$\sup_{v\in[\frac{h}{2},1-\frac{h}{2}]}\sup_{\theta\in\Theta}\big|\nabla_\theta^j L_{n,h}(v,\theta)-\mathbb{E}\nabla_\theta^j \ell_\theta(\tilde{Z}_1(v))\big|_\infty = O_p(\tau_n^{(j)}) \tag{2.8.74}$$

where

$$\tau_n^{(j)} := \tau_n + (nh)^{-1} + h, \quad j\in\{0,2\}, \qquad \tau_n^{(1)} := \tau_n + (nh)^{-1} + B_h.$$

By (A3) and (2.8.74) for $j=0$, we obtain via standard arguments that if $\tau_n^{(0)} = o(1)$,

$$\sup_{v\in[\frac{h}{2},1-\frac{h}{2}]}\big|\hat{\theta}_{n,h}(v)-\theta_0(v)\big|_\infty = o_p(1).$$

Since $\hat{\theta}_{n,h}(v)$ is a minimizer of $\theta\mapsto L_{n,h}(v,\theta)$ and ℓ_θ is twice continuously differentiable, we have the representation

$$\hat{\theta}_{n,h}(v)-\theta_0(v) = -\nabla_\theta^2 L_{n,h}(v,\bar{\theta}_v)^{-1}\nabla_\theta L_{n,h}(v,\theta_0(v)) \tag{2.8.75}$$

where $\bar{\theta}_v\in\Theta$ fulfills $|\bar{\theta}_v-\theta_0(v)|_\infty \le |\hat{\theta}_{n,h}(v)-\theta_0(v)|_\infty = o_p(1)$.

By (A2), we have

$$\big|\mathbb{E}\nabla_\theta^2\ell_\theta(\tilde{Z}_0(v))\big|_{\theta=\theta_0(v)} - \mathbb{E}\nabla_\theta^2\ell_\theta(\tilde{Z}_0(v))\big|_{\theta=\bar{\theta}_v}\big|_\infty = O(|\theta_0(v)-\bar{\theta}_v|_2) = o_p(1)$$

and thus with (2.8.74),

$$\sup_{v\in[\frac{h}{2},1-\frac{h}{2}]}\big|\nabla_\theta^2 L_{n,h}(v,\bar{\theta}_v)-\mathbb{E}\nabla_\theta^2\ell_\theta(\tilde{Z}_1(v))\big|_{\theta=\theta_0(v)}\big|_\infty = O_p(\tau_n^{(2)}) + o_p(1). \tag{2.8.76}$$

By (A3) and the dominated convergence theorem, $\mathbb{E}\nabla_\theta\ell(\tilde{Z}_0(v)) = \nabla_\theta\mathbb{E}\ell(\tilde{Z}_0(v)) = 0$. By (2.8.74),

$$\begin{aligned}\sup_{v\in[\frac{h}{2},1-\frac{h}{2}]}\big|\nabla_\theta L_{n,h}(v,\theta_0(v))\big|_\infty &= \sup_{v\in[\frac{h}{2},1-\frac{h}{2}]}\big|\nabla_\theta L_{n,h}(v,\theta_0(v))-\mathbb{E}\nabla_\theta\ell(\tilde{Z}_0(v))\big|_\infty\\ &= O_p(\tau_n^{(1)}). \end{aligned}\tag{2.8.77}$$

Inserting (2.8.76) and (2.8.77) into (2.8.75), we obtain

$$\sup_{v\in[\frac{h}{2},1-\frac{h}{2}]}\big|\hat{\theta}_{n,h}(v)-\theta_0(u)\big|_\infty = O_p(\tau_n^{(1)}).$$

This yields an improved version of (2.8.76):

$$\sup_{v\in[\frac{h}{2},1-\frac{h}{2}]}\big|\nabla_\theta^2 L_{n,h}(v,\bar{\theta}_v)-\mathbb{E}\nabla_\theta^2\ell_\theta(\tilde{Z}_1(v))\big|_{\theta=\theta_0(v)}\big|_\infty = O_p(\tau_n^{(2)}). \tag{2.8.78}$$

Inserting (2.8.77) and (2.8.78) into (2.8.75), we obtain the assertion. □

2.8.6 Proofs of Section 2.6

We now provide the proofs for the large deviation inequalities. We generally consider $\mathbb{G}_n(f) = \frac{1}{\sqrt{n}}S_n(f)$, see Subsection 2.8.2.

Proof of Theorem 2.6.1. The statement of the theorem is obtained for $W_i(f) = f(Z_i, \frac{i}{n})$.

Let $V_n^{\circ}(f) = \|f\|_{2,n} + \sum_{j=1}^{\infty}\min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(j)\}\varphi(j)^{1/2}$, where $\varphi(j) = \log\log(e^e j)$. The term $V_n^{\circ}(f)$ serves as a lower bound for $\tilde{V}_n(f)$.

For $q \in \{1, ..., n\}$, we use decomposition (2.8.23) without the maximum. The set $B_n(q)$ is defined below in (2.8.85). We then have

$$\begin{aligned}
&\mathbb{P}\big(\big|\frac{1}{\sqrt{n}}S_n(f)\big| > x, B_n(q)\big)\\
\leq\ & \mathbb{P}\Big(\frac{1}{\sqrt{n}}\big|S_n(f) - S_{n,q}(f)\big| > x/4, B_n(q)\Big)\\
&+\mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\sqrt{\frac{n}{\tau_l}}}\sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\sqrt{\tau_l}}T_{i,l}(f)\Big| > x/4\Big) + \mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\sqrt{\frac{n}{\tau_l}}}\sum_{\substack{i=1\\ i \text{ odd}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\sqrt{\tau_l}}T_{i,l}(f)\Big| > x/4\Big)\\
&+\mathbb{P}\Big(\frac{1}{\sqrt{n}}\big|S_{n,1}(f)\big| > x/4\Big)\\
=:\ & A_1 + A_2 + A_3.
\end{aligned}$$

Define for $l \in \mathbb{N}$,

$$g_1(l) = \sqrt{\log(l+1)+1}, \qquad g_2(l) = \log(l+1)+1, \qquad a(l) = l^{1/2}\log(el)^{1/2}\varphi(l),$$

and for $j \in \mathbb{N}$, $\gamma(j) = \log_2(j) + 1$. By elementary calculations, we see that there exists a universal constant $c \geq 1$ such that

$$\begin{aligned}
\sum_{l=1}^{L}\tau_l g_2(l) &\leq \sum_{l=1}^{L}2\sum_{j=\tau_{l-1}+1}^{\tau_l}g_2(l) \leq 2\sum_{j=1}^{q}\sum_{l=1}^{L}\mathbb{1}_{\{\tau_{l-1}+1\leq j\leq\tau_l\}}g_2(l)\\
&\leq 2\sum_{j=1}^{q}g_2(\gamma(j)) \leq 2q\cdot g_2(\gamma(q)) \leq 8\Phi(q).
\end{aligned}$$

The third to last inequality is due to $2^{l-1}+1 = \tau_{l-1}+1 \leq j \Longleftrightarrow l \leq \log_2(j-1)+1 \leq \gamma(j)$ and the monotonicity of g. In a similar fashion,

$$\begin{aligned}
&\sum_{l=1}^{L}g_1(l)\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\\
\leq\ & \sum_{j=1}^{q}\min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\cdot\sum_{l=1}^{L}\mathbb{1}_{\{\tau_{l-1}+1\leq j\leq\tau_l}g_1(l)\\
\leq\ & \sum_{j=1}^{q}\min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}g_1(\gamma(j)) \leq 4V_n^{\circ}(f)
\end{aligned}$$

by $g_1(\gamma(j)) \le 2\varphi(j)^{1/2}$ and Lemma 2.8.5.

Therefore,

$$\begin{aligned}
\frac{x}{4} &= \frac{x}{8} + \frac{x}{8} = \frac{x}{32V_n^\circ(f)}\sum_{l=1}^{L} g_1(l) \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(\lfloor \frac{j}{2}\rfloor)\} + \frac{x}{64\Phi(q)}\sum_{l=1}^{L}\tau_l g_2(l) \\
&= \sum_{l=1}^{L} y_1(l) + \sum_{l=1}^{L} y_2(l),
\end{aligned}$$

where $y_1(l) := \frac{x}{32V^\circ(f)} g_1(l)\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n}, \mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}$, $y_2(l) = \frac{x}{64\Phi(q)}\tau_l g_2(l)$.

We now use a standard Bernstein inequality for independent random variables: If $M_Q, \sigma_Q > 0$ are constants and Q_i, $i = 1, ..., m$ mean-zero independent variables with $|Q_i| \le M_Q$, $(\frac{1}{m}\sum_{i=1}^m \|Q_i\|_2^2)^{1/2} \le \sigma_Q$, then for any $z > 0$,

$$\mathbb{P}\Big(\frac{1}{\sqrt{m}}\Big|\sum_{i=1}^{m}\big[Q_i - \mathbb{E}Q_i\big]\Big| > z\Big) \le 2\cdot\exp\Big(-\frac{1}{2}\frac{z^2}{\sigma_Q^2 + \frac{2}{3}\frac{M_Q z}{\sqrt{m}}}\Big). \tag{2.8.79}$$

Using the bound (2.8.33), $\frac{1}{\sqrt{\tau_l}}|T_{i,l}(f)| \le 2\sqrt{\tau_l}\|f\|_\infty \le 2\sqrt{\tau_l}M$ and the elementary inequality $\frac{1}{2}\min\{\frac{a}{b},\frac{a}{c}\} \le \frac{a}{b+c} \le \min\{\frac{a}{b},\frac{a}{c}\}$ we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\sqrt{\frac{n}{\tau_l}}}\sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\sqrt{\tau_l}}T_{i,l}(f)\Big| > \frac{x}{4}\Big) \\
\le\;& \sum_{l=1}^{L}\mathbb{P}\Big(\Big|\frac{1}{\sqrt{\frac{n}{\tau_l}}}\sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\sqrt{\tau_l}}T_{i,l}(f)\Big| > y_1(l) + y_2(l)\Big) \\
\le\;& 2\sum_{l=1}^{L}\exp\Big(-\frac{1}{4}\min\Big\{\frac{(y_1(l)+y_2(l))^2}{\Big(\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n},\mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\Big)^2}, \frac{y_1(l)+y_2(l)}{\frac{2}{3}\frac{\sqrt{\tau_l}M}{\sqrt{\frac{n}{\tau_l}}}}\Big\}\Big) \\
\le\;& 2\sum_{l=1}^{L}\exp\Big(-\frac{1}{4}\min\Big\{\frac{y_1(l)^2}{\Big(\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n},\mathbb{D}_n\Delta(\lfloor\frac{j}{2}\rfloor)\}\Big)^2}, \frac{y_2(l)}{\frac{2}{3}\frac{\sqrt{\tau_l}M}{\sqrt{\frac{n}{\tau_l}}}}\Big\}\Big) \\
=\;& 2\sum_{l=1}^{L}\exp\Big(-\frac{1}{4}\min\Big\{\frac{x^2 g_1(l)^2}{2^{10}V_n^\circ(f)^2}, \frac{x g_2(l)}{\frac{2^7}{3}\frac{M\Phi(q)}{\sqrt{n}}}\Big\}\Big) \\
\le\;& 2\sum_{l=1}^{L}\exp\Big(-\frac{x^2 g_1(l)^2}{2^{12}V_n^\circ(f)^2}\Big) + 2\sum_{l=1}^{L}\exp\Big(-\frac{x g_2(l)}{\frac{2^9}{3}\frac{M\Phi(q)}{\sqrt{n}}}\Big).
\end{aligned} \tag{2.8.80}$$

We now discuss the summands in (2.8.80). If $x > \sqrt{2} \cdot 2^6 V_n^\circ(f)$,

$$\begin{aligned}\Big(\sum_{l=1}^{L} \exp\Big(-\frac{x^2 g_1(l)^2}{2^{12} V_n^\circ(f)^2}\Big)\Big) \exp\Big(\frac{x^2}{2^{12} V_n^\circ(f)^2}\Big) &= \sum_{l=1}^{L} \exp\Big(-\log(l+1) \cdot \Big(\frac{x}{2^6 V_n^\circ(f)}\Big)^2\Big) \\ &\leq \sum_{l=1}^{L} (l+1)^{-(\frac{x}{2^6 V^\circ(f)})^2} \leq \frac{\pi^2}{6}.\end{aligned}$$

Similarly, if $x > \frac{2^{10}}{3} \frac{M\Phi(q)}{\sqrt{n}}$,

$$\Big(\sum_{l=0}^{L} \exp\Big(-\frac{x g_2(l)}{\frac{2^9}{3}\frac{M\Phi(q)}{\sqrt{n}}}\Big)\Big) \exp\Big(\frac{x}{\frac{2^9}{3}\frac{M\Phi(q)}{\sqrt{n}}}\Big) \leq \frac{\pi^2}{6}.$$

We conclude from (2.8.80): If

$$x > \max\{\sqrt{2} \cdot 2^6 V_n^\circ(f), \frac{2^{10}}{3}\frac{M\Phi(q)}{\sqrt{n}}\}, \tag{2.8.81}$$

then

$$\begin{aligned}&\mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\sqrt{\frac{n}{\tau_l}}} \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{\sqrt{\tau_l}} T_{i,l}(f)\Big| > \frac{x}{4}\Big) \\ \leq\ & \frac{\pi^2}{3}\Big[\exp\Big(-\frac{x^2}{2^{12} V_n^\circ(f)^2}\Big) + \exp\Big(-\frac{x}{\frac{2^9}{3}\frac{M\Phi(q)}{\sqrt{n}}}\Big)\Big] \\ \leq\ & \frac{2\pi^2}{3} e^2 \exp\Big(-\min\Big\{\frac{x^2}{2^{12} V_n^\circ(f)^2}, \frac{x}{\frac{2^9}{3}\frac{M\Phi(q)}{\sqrt{n}}}\Big\}\Big), \end{aligned} \tag{2.8.82}$$

where in the last step we added the factor e^2 for convenience with regards to the next step of the proof. If (2.8.81) is not fulfilled, then either $x \leq \sqrt{2} \cdot 2^6 V_n^\circ(f)$ or $x \leq \frac{2^{10}}{3}\frac{M\Phi(q)}{\sqrt{n}}$. The upper bound (2.8.82) then still holds true since $x \leq \sqrt{2} \cdot 2^6 V_n^\circ(f)$ implies

$$\exp\Big(-\min\Big\{\frac{x^2}{2^{12} V_n^\circ(f)^2}, \frac{x}{\frac{2^9}{3}\frac{M\Phi(q)}{\sqrt{n}}}\Big\}\Big) \geq \exp\Big(-\frac{x^2}{2^{12} V_n^\circ(f)^2}\Big) \geq \exp(-2)$$

and the left hand side of (2.8.82) is ≤ 1 since it is a probability. A similar bound is valid for $x \leq \frac{2^{10}}{3}\frac{M\Phi(q)}{\sqrt{n}}$. Thus, (2.8.82) holds true for all $x > 0$.

We therefore obtain

$$A_2 \leq \frac{4\pi^2 e^2}{3} \cdot \exp\Big(-\frac{1}{2}\frac{x^2}{2^{11} V_n^\circ(f)^2 + \frac{2^8}{3}\frac{M\Phi(q)x}{\sqrt{n}}}\Big). \tag{2.8.83}$$

Since $\|W_i(f)\|_2 \leq \|f(Z_i, \frac{i}{n})\|_2$ and $\|W_i(f)\|_\infty \leq \|f\|_\infty \leq M$, we obtain from (2.8.79)

$$A_3 \leq 2\exp\Big(-\frac{1}{2}\frac{x^2}{2^4\|f\|_{2,n}^2 + \frac{2^3}{3}\frac{Mx}{\sqrt{n}}}\Big). \tag{2.8.84}$$

Since $1 \leq \Phi(q)$ and $\|f\|_{2,n} \leq V_n^\circ(f)$, this yields a similar bound as (2.8.83).

We now discuss A_1. Write

$$\frac{1}{\sqrt{n}}(S_n(f) - S_{n,q}(f)) = \sum_{j=q}^{\infty}\frac{1}{\sqrt{n}}(S_{n,j+1}(f) - S_{n,j}(f)) = \sum_{j=q}^{\infty}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(W_{i,j+1}(f) - W_{i,j}(f)).$$

Put

$$\begin{aligned}\Omega_n(j) \quad := \quad & \{\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[(W_{i,j+1}(f) - W_{i,j}(f))^2|\mathcal{A}_{i-1}] \leq (\frac{M\Phi(q)}{\tilde{\beta}(q)\sqrt{n}})^2\Delta(j)^2 a(j)^2 g_2(j)\} \\ & \cap\{\sup_{f\in\mathcal{F}}\sup_{i=1,\dots,n}|W_{i,j+1}(f) - W_{i,j}(f)| \leq 2\frac{M\Phi(q)}{\tilde{\beta}(q)}\Delta(j)a(j)\}\end{aligned}$$

and

$$B_n(q) := \bigcap_{j=q}^{\infty}\Omega_n(j). \tag{2.8.85}$$

Note that

$$A_1 \leq \mathbb{P}\Big(\frac{1}{\sqrt{n}}\big|S_n(f) - S_{n,q}(f)\big| > \frac{x}{4}, \bigcap_{j=q}^{\infty}\Omega_n(j)\Big). \tag{2.8.86}$$

Here, $W_{i,j+1}(f) - W_{i,j}(f)$ is a martingale difference with respect to $\mathcal{A}_i$. Furthermore,

$$\sum_{j=q}^{\infty}\Delta(j)a(j)g_2(j) \leq 4\sum_{j=q}^{\infty}\Delta(j)j^{1/2}\log(ej)^2 = 4\tilde{\beta}(q).$$

By Freedman's Bernstein-type inequality for martingales (cf. Freedman [1975]) we have for $x \geq 2^4 \frac{M\Phi(q)}{\sqrt{n}}$,

$$
\begin{aligned}
& \mathbb{P}\Big(\frac{1}{\sqrt{n}}\big|S_n(f)-S_{n,q}(f)\big| > \frac{x}{4}, \bigcap_{j=q}^{\infty}\Omega_n(j)\Big) \\
\leq \; & \sum_{k=q}^{\infty}\mathbb{P}\Big(\frac{1}{\sqrt{n}}\Big|\sum_{i=1}^{n}(W_{i,j+1}(f)-W_{i,j}(f))\Big| > \frac{x}{2^4}\frac{\Delta(j)a(j)g_2(j)}{\tilde{\beta}(q)}, \Omega_n(j)\Big) \\
\leq \; & 2\sum_{k=q}^{\infty}\exp\Big(-\frac{1}{2}\frac{(\frac{x}{2^4}\frac{\Delta(j)a(j)g_2(j)}{\tilde{\beta}(q)})^2}{(\frac{M\Phi(q)}{\tilde{\beta}(q)\sqrt{n}})^2\Delta(j)^2a(j)^2g_2(j)+\frac{2}{3}\frac{M\Phi(q)\Delta(j)a(j)}{\tilde{\beta}(q)\sqrt{n}}\cdot\frac{x}{2^4}\frac{\Delta(j)a(j)g_2(j)}{\tilde{\beta}(q)}}\Big) \\
= \; & 2\sum_{k=q}^{\infty}\exp\Big(-\frac{1}{2}\frac{x^2g_2(j)^2}{2^8(\frac{M\Phi(q)}{\sqrt{n}})^2g_2(j)+\frac{2^5}{3}\frac{M\Phi(q)xg_2(j)}{\sqrt{n}}}\Big) \\
= \; & 2\sum_{k=q}^{\infty}\exp\Big(-\frac{g_2(j)}{4}\min\Big\{\Big(\frac{x}{2^4\frac{M\Phi(q)}{\sqrt{n}}}\Big)^2, \Big(\frac{x}{2^4\frac{M\Phi(q)}{\sqrt{n}}}\Big)\Big\}\Big) \\
\leq \; & 2\sum_{k=q}^{\infty}\exp\Big(-\frac{g_2(j)x}{2^6\frac{M\Phi(q)}{\sqrt{n}}}\Big).
\end{aligned}
\tag{2.8.87}
$$

We conclude that for $x > 2^7 \frac{M\Phi(q)}{\sqrt{n}}$,

$$
\Big(\sum_{j=q}^{\infty}\exp\Big(-\frac{g_2(j)x}{2^6\frac{M\Phi(q)}{\sqrt{n}}}\Big)\Big)\cdot\exp\Big(\frac{x}{2^6\frac{M\Phi(q)}{\sqrt{n}}}\Big) \leq \sum_{j=q}^{\infty}(j+1)^{-(\frac{x}{2^4\frac{M\Phi(q)}{\sqrt{n}}})} \leq \frac{\pi^2}{6},
$$

and thus (with an additional factor e^2),

$$
A_1 \leq \mathbb{P}\Big(\frac{1}{\sqrt{n}}\big|S_n(f)-S_{n,q}(f)\big| > \frac{x}{4}, \bigcap_{j=q}^{\infty}\Omega_n(j)\Big) \leq \frac{\pi^2}{3}e^2\exp\Big(-\frac{x}{2^4\frac{M\Phi(q)}{\sqrt{n}}}\Big). \tag{2.8.88}
$$

In the case $x \leq 2^7 \frac{M\Phi(q)}{\sqrt{n}}$, we have

$$
\frac{\pi^2}{3}e^2\exp\Big(-\frac{x}{2^6\frac{M\Phi(q)}{\sqrt{n}}}\Big) \geq \frac{\pi^2}{3} \geq 1,
$$

thus (2.8.88) holds true for all $x > 0$.

Finally, since $g_2(j) \geq 1$, we have

$$
\begin{aligned}
\Omega_n(j) \;\subset\; & \{\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\sup_{f\in\mathcal{F}}\big|W_{i,j+1}(f)-W_{i,j}(f)\big|^2|\mathcal{A}_{i-1}] \leq (\frac{M\Phi(q)}{\tilde{\beta}(q)\sqrt{n}})^2\Delta(j)^2a(j)^2\} \\
& \cap\{\Big(\sum_{i=1}^{n}\sup_{f\in\mathcal{F}}|W_{i,j+1}(f)-W_{i,j}(f)|^2\Big)^{1/2} \leq 2\frac{M\Phi(q)}{\tilde{\beta}(q)}\Delta(j)a(j)\},
\end{aligned}
$$

thus we obtain by Markov's inequality as in (2.8.30) that

$$\begin{aligned}\mathbb{P}(\Omega_n(j)^c) &\leq \Big(\frac{\sqrt{n}\tilde{\beta}(q)}{M\Phi(q)}\Big)^2\frac{1}{\Delta(j)^2 a(j)^2}\frac{1}{n}\sum_{i=1}^{n}\Big\|\sup_{f\in\mathcal{F}}\big|W_{i,j+1}(f)-W_{i,j}(f)\big|\Big\|_2^2\\ &\qquad +\Big(\frac{\tilde{\beta}(q)}{2M\Phi(q)}\Big)^2\frac{1}{\Delta(j)^2 a(j)^2}\sum_{i=1}^{n}\Big\|\sup_{f\in\mathcal{F}}\big|W_{i,j+1}(f)-W_{i,j}(f)\big|\Big\|_2^2\\ &\leq 2\Big(\frac{\tilde{\beta}(q)\sqrt{n}}{M\Phi(q)}\Big)^2\frac{(\mathbb{D}_n^\infty)^2}{a(j)^2}.\end{aligned}$$

Therefore,

$$\mathbb{P}(B_n(q)^c) \leq \mathbb{P}\Big(\bigcup_{j=q}^{\infty}\Omega_n(j)^c\Big) \leq 2\Big(\frac{\mathbb{D}_n^\infty\tilde{\beta}(q)\sqrt{n}}{M\Phi(q)}\Big)^2\sum_{k=q}^{\infty}\frac{1}{a(j)^2}. \tag{2.8.89}$$

Note that

$$\begin{aligned}\sum_{j=q+1}^{\infty}\frac{1}{a(j)^2} &\leq \sum_{j=q+1}^{\infty}\int_{j-1}^{j}\frac{1}{a(j)^2}dx\\ &\leq \int_q^\infty\frac{1}{a(x)^2}dx \leq 2\int_q^\infty\frac{1}{x\log(e^e x)\log(\log(e^e x))}dx = \frac{2}{\log(\log(e^e q))},\end{aligned}$$

so that

$$\sum_{k=q}^{\infty}\frac{1}{a(j)^2} = \frac{1}{a(q)^2} + \sum_{j=q+1}^{\infty}\frac{1}{a(j)^2} \leq \frac{3}{\varphi(q)}.$$

Summarizing the bounds (2.8.83), (2.8.84), (2.8.88) and (2.8.89) and using the fact that

$$V_n^\circ(f) = \|f\|_{2,n} + \sum_{j=1}^{\infty}\min\{\|f\|_{2,n},\mathbb{D}_n\Delta(j)\}\varphi(j)^{1/2} \leq \tilde{V}_n(f),$$

we obtain (2.6.1).

We now show (2.6.2) by a case distinction. We abbreviate $\tilde{q}^* = \tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y})$. In the case $\Phi(\tilde{q}^*)\frac{1}{n} \leq 1$, we have $\tilde{q}^* \in \{1,...,n\}$, and thus by (2.6.1),

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}}|S_n(f)| > x, B_n(\tilde{q}^*)\Big) \leq c_0\exp\Big(-\frac{1}{c_1}\frac{x^2}{\tilde{V}_n(f)^2 + \frac{M\Phi(\tilde{q}^*)}{\sqrt{n}}x}\Big)$$

and, by definition of $\tilde{q}^*$,

$$\mathbb{P}(B_n(\tilde{q}^*)^c) \leq 4\Big(\frac{\tilde{\beta}(\tilde{q}^*)}{\Phi(\tilde{q}^*)}\cdot\frac{\mathbb{D}_n^\infty\sqrt{n}}{M}\Big)^2 \leq \frac{4}{y^2}.$$

In case of $\Phi(\tilde{q}^*)\frac{1}{n} > 1$, we obviously have

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{\sqrt{n}}|S_n(f)| > x\Big) &\leq \mathbb{P}(M\sqrt{n} > x) \leq c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{M\sqrt{n}}\Big)\\
&\leq c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{\frac{M\Phi(\tilde{q}^*)}{\sqrt{n}}}\Big) \leq c_0 \exp\Big(-\frac{1}{c_1}\frac{x^2}{\tilde{V}(f)^2 + \frac{M\Phi(\tilde{q}^*)}{\sqrt{n}}x}\Big),
\end{aligned}$$

and the assertion follows without any restricting set $B_n(q)$ We can therefore choose q arbitrarily. □

Lemma 2.8.9. *Let $\mathcal{F}$ be a class of functions in the setting of Lemma 3.5.2. Then there exist universal constants $c_0, c_1 > 0$ such that the following holds: For each $q \in \{1, ..., n\}$ there exists a set $B_n^{(2)}(q)$ independent of $f \in \mathcal{F}$ such that for all $x > 0$,*

$$\mathbb{P}\Big(\frac{1}{n}\big|S_n(f)\big| > x, B_n^{(2)}(q)\Big) \leq c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{\frac{M^2\Phi(q)}{n}} \cdot \min\Big\{\frac{x}{\|f\|_{2,n}V_n(f)}, 1\Big\}\Big) \quad (2.8.90)$$

and

$$\mathbb{P}(B_n^{(2)}(q)^c) \leq \frac{n(\mathbb{D}_n^\infty)^2}{M^2} \cdot C_\Delta \frac{\beta(q)}{\Phi(q)}.$$

Define $\tilde{q}^*(z) = \min\{q \in \mathbb{N} : \beta(q) \leq \Phi(q)x\}$. Then for any $x > 0$, $y > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\frac{1}{n}\big|S_n(f)\big| > x, B_n^{(2)}(\tilde{q}^*(\frac{M^2}{n(\mathbb{D}_n^\infty)^2y^2}))\Big)\\
\leq\ & c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{\frac{M^2}{n}\Phi(\tilde{q}^*(\frac{M^2}{n(\mathbb{D}_n^\infty)^2y^2}))} \cdot \min\Big\{\frac{x}{\|f\|_{2,n}V_n(f)}, 1\Big\}\Big) \quad (2.8.91)
\end{aligned}$$

and $\mathbb{P}(B_n^{(2)}(\tilde{q}^*(\frac{M^2}{n(\mathbb{D}_n^\infty)^2y^2}))^c) \leq \frac{C_\Delta}{y^2}$.

Proof of Lemma 2.8.9. We use a similar argument as in Theorem 2.6.1, especially we make use of the decomposition (2.8.23).

The set $B_n^{(2)}(q)$ is defined below in (2.8.97). We then have

$$\begin{aligned}
&\mathbb{P}(\big|\frac{1}{n}S_n(f)\big| > x, B_n^{(2)}(q))\\
\leq\ & \mathbb{P}\Big(\frac{1}{n}\big|S_n(f) - S_{n,q}(f)\big| > x/4, B_n^{(2)}(q)\Big)\\
&+\mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\frac{n}{\tau_l}}\sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\tau_l}T_{i,l}(f)\Big| > x/4\Big) + \mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\frac{n}{\tau_l}}\sum_{\substack{i=1\\ i \text{ odd}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\tau_l}T_{i,l}(f)\Big| > x/4\Big)\\
&+\mathbb{P}\Big(\frac{1}{n}\big|S_{n,1}(f)\big| > x/4\Big)\\
=:\ & A_1 + A_2 + A_3.
\end{aligned}$$

As in the proof of Theorem 2.6.1, we see that with $g_2(l) = \log(l+1)+1$,

$$\sum_{l=1}^{L} \tau_l g_2(l) \le 8\Phi(q).$$

Therefore,

$$\begin{aligned}\frac{x}{4} &= \frac{x}{8}+\frac{x}{8} = \frac{x}{8V_n(f)}\sum_{l=1}^{L}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n},\mathbb{D}_n\Delta(j)\} + \frac{x}{64\Phi(q)}\sum_{l=1}^{L}\tau_l g_2(l)\\ &= \sum_{l=1}^{L} y_1(l) + \sum_{l=1}^{L} y_2(l)\end{aligned}$$

where $y_1(l) := \frac{x}{8V_n(f)}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n},\mathbb{D}_n\Delta(j)\}$, $y_2(l) = \frac{x}{64\Phi(q)}\tau_l g_2(l)$.

By Lemma 3.5.3, if $M_Q, \sigma_Q > 0$ are constants and Q_i, $i = 1,...,m$ mean-zero independent variables with $|Q_i| \le M_Q$, $\frac{1}{m}\sum_{i=1}^m \|Q_i\|_1 \le \sigma_Q$, then for any $z > 0$,

$$\mathbb{P}\Big(\frac{1}{\sqrt{m}}\Big|\sum_{i=1}^{m}\big[Q_i - \mathbb{E}Q_i\big]\Big| > z\Big) \le 2\cdot\exp\Big(-\frac{1}{2}\frac{z^2}{\sigma_Q\frac{M_Q}{m}+\frac{2}{3}\frac{M_Q z}{m}}\Big). \tag{2.8.92}$$

Using the bound (3.5.10) combined with (3.5.13), $\frac{1}{\tau_l}|T_{i,l}(f)| \le 2\|f\|_\infty^2 \le 2M^2$ and the elementary inequalities $\frac{1}{2}\min\{\frac{a}{b},\frac{a}{c}\} \le \frac{a}{b+c} \le \min\{\frac{a}{b},\frac{a}{c}\}$ and $(a+b)^2 \ge 4ab$, we obtain

$$\begin{aligned}&\mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\frac{n}{\tau_l}}\sum_{\substack{i=1\\ i\text{ even}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\tau_l}T_{i,l}(f)\Big| > \frac{x}{4}\Big)\\
\le\ & \sum_{l=1}^{L}\mathbb{P}\Big(\Big|\frac{1}{\frac{n}{\tau_l}}\sum_{\substack{i=1\\ i\text{ even}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1}\frac{1}{\tau_l}T_{i,l}(f)\Big| > y_1(l)+y_2(l)\Big)\\
\le\ & 2\sum_{l=1}^{L}\exp\Big(-\frac{1}{4}\min\Big\{\frac{(y_1(l)+y_2(l))^2}{\|f\|_{2,n}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n},\mathbb{D}_n\Delta(j)\}\cdot\frac{M^2}{\frac{n}{\tau_l}}},\\
&\qquad\qquad\qquad\qquad\frac{y_1(l)+y_2(l)}{\frac{2}{3}\frac{M^2}{\frac{n}{\tau_l}}}\Big\}\Big)\\
\le\ & 2\sum_{l=1}^{L}\exp\Big(-\frac{1}{4}\cdot\frac{y_2(l)}{\frac{M^2}{\frac{n}{\tau_l}}}\min\Big\{\frac{4y_1(l)}{\|f\|_{2,n}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|f\|_{2,n},\mathbb{D}_n\Delta(j)\}},\frac{1}{\frac{2}{3}}\Big\}\Big)\\
\le\ & 2\sum_{l=1}^{L}\exp\Big(-\frac{x g_2(l)}{2^9\frac{M^2\Phi(q)}{n}}\cdot\min\Big\{\frac{x}{\|f\|_{2,n}V_n(f)},1\Big\}\Big).\end{aligned} \tag{2.8.93}$$

If x is such that $c(x) := \frac{x}{2^9 \frac{M^2\Phi(q)}{n}} \cdot \min\big\{\frac{x}{\|f\|_{2,n}V_n(f)}, 1\big\} \ge 2$, then

$$\Big(\sum_{l=1}^{L} \exp\big(-g_2(l)c(x)\big)\Big) \cdot \exp\big(-c(x)\big) = \sum_{l=1}^{L} \exp\Big(-\log(l+1)c(x)\Big) \le \sum_{l=1}^{L} (l+1)^{-c(x)} \le \frac{\pi^2}{6}.$$

Insertion into (2.8.93) leads to

$$\mathbb{P}\Big(\sum_{l=1}^{L}\Big|\frac{1}{\frac{n}{\tau_l}}\sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l}\rfloor+1}\frac{1}{\tau_l}T_{i,l}(f)\Big| > \frac{x}{4}\Big) \le \frac{\pi^2}{3}\cdot e^2\exp(-c(x)). \tag{2.8.94}$$

In case $c(x) < 2$, the right hand side of (2.8.94) is ≥ 1. Thus, (2.8.94) holds true for all $x > 0$.

We therefore obtain

$$A_2 \le \frac{2\pi^2 e^2}{3}\cdot\exp\Big(-\frac{1}{2^9}\cdot\frac{x}{\frac{M^2\Phi(q)}{n}}\cdot\min\big\{\frac{x}{\|f\|_{2,n}V_n(f)}, 1\big\}\Big). \tag{2.8.95}$$

Since $\|W_i(f)\|_1 \le \|f(Z_i, \frac{i}{n})\|_2^2$ and $\|W_i(f)\|_\infty \le \|f\|_\infty^2 \le M^2$, we obtain from (2.8.92),

$$A_3 \le 2\exp\Big(-\frac{1}{2}\frac{x^2}{\|f\|_{2,n}^2\cdot\frac{M^2}{n} + \frac{2}{3}\frac{M^2 x}{n}}\Big) \le 2\exp\Big(-\frac{1}{2}\frac{x}{\frac{M^2}{n}}\cdot\min\big\{\frac{x}{\|f\|_{2,n}^2}, 1\big\}\Big). \tag{2.8.96}$$

Since $1 \le \Phi(q)$ and $\|f\|_{2,n} \le V_n(f)$, this yields a similar bound as (2.8.95).

We now discuss A_1. Put

$$B_n^{(2)}(q) := \{\sup_{f\in\mathcal{F}}\frac{1}{n}|S_n(f) - S_{n,q}(f)| \le \frac{M^2\Phi(q)}{n}\}. \tag{2.8.97}$$

Then with Markov's inequality and using the same calculation as in (3.5.17),

$$\begin{aligned}\mathbb{P}(B_n^{(2)}(q)^c) &\le \frac{n}{M^2\Phi(q)}\cdot\big\|\sup_{f\in\mathcal{F}}\frac{1}{n}|S_n(f) - S_{n,q}(f)|\big\|_1\\ &\le \frac{n}{M^2\Phi(q)}\cdot\sum_{k=q}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\big\|\sup_{f\in\mathcal{F}}|W_{i,j+1}(f) - W_{i,j}(f)|\big\|_1\\ &\le \frac{n}{M^2\Phi(q)}\cdot(\mathbb{D}_n^\infty)^2 C_\Delta\beta(q). \end{aligned} \tag{2.8.98}$$

Furthermore,

$$\begin{aligned}A_1 &= \mathbb{P}\Big(\frac{1}{n}|S_n(f) - S_{n,q}(f)| > \frac{x}{4}, B_n^{(2)}(q)\Big)\\ &= \mathbb{1}_{\{\frac{M^2\Phi(q)}{n} > \frac{x}{4}\}} \le e\cdot\exp\Big(-\frac{x}{4\frac{M^2\Phi(q)}{n}}\Big).\end{aligned} \tag{2.8.99}$$

Summarizing the bounds (2.8.95), (2.8.96), (2.8.99) and (2.8.98), we obtain the result (2.8.90).

We now show (2.8.91) by a case distinction. Abbreviate $\tilde{q}^* = \tilde{q}^*(\frac{M^2}{n(\mathbb{D}_n^\infty)^2 y^2})$. In the case $\Phi(\tilde{q}^*)\frac{1}{n} \le 1$, we have $\tilde{q}^* \in \{1, ..., n\}$ and thus by (2.8.90),

$$\mathbb{P}\Big(\frac{1}{n}|S_n(f)| > x, B_n(\tilde{q}^*)\Big) \le c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{\frac{M^2\Phi(\tilde{q}^*)}{n}} \cdot \min\big\{\frac{x}{\|f\|_{2,n}V(f)}, 1\big\}\Big)$$

and, by definition of $\tilde{q}^*$,

$$\mathbb{P}(B_n(\tilde{q}^*)^c) \le \frac{(\mathbb{D}_n^\infty)^2 n}{M^2} \cdot C_\Delta \frac{\beta(\tilde{q}^*)}{\Phi(\tilde{q}^*)} \le \frac{C_\Delta}{y^2}.$$

The assertion follows with $B_n^{(2)}(M, y) = B_n^{(2)}(\tilde{q}^*)$.

In the case $\Phi(\tilde{q}^*)\frac{1}{n} > 1$, we obviously have

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{n}|S_n(f)| > x\Big) &\le \mathbb{P}(M^2 > x) \le c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{M^2}\Big) \\
&\le c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{\frac{M^2\Phi(\tilde{q}^*)}{n}}\Big) \\
&\le c_0 \exp\Big(-\frac{1}{c_1}\frac{x}{\frac{M^2\Phi(\tilde{q}^*)}{n}} \cdot \min\big\{\frac{x}{\|f\|_{2,n}V_n(f)}, 1\big\}\Big).
\end{aligned}$$

The assertion follows with $B_n^{(2)}(M, y)$ being the whole probability space. □

Proof of Theorem 2.6.4. Let $B_n(q)$ denote the set from Theorem 2.6.1 (applied to $W_i(f) = \mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]$ instead of $W_i(f) = f(Z_i, \frac{i}{n})$; the proof is similar for this situation). Let $B_n^{(2)}(q)$ denote the set from Lemma 2.8.9.

Put

$$B_n^\circ(q) = B_n(q) \cap B_n^{(2)}(q^2).$$

Then we have for $\mathbb{G}_n(f) = \mathbb{G}_n^{(1)}(f) + \mathbb{G}_n^{(2)}(f)$ as in Section 3.3,

$$\begin{aligned}
\mathbb{P}\big(|\mathbb{G}_n(f)| > x, B_n^\circ(q)\big) &\le \mathbb{P}\big(\mathbb{G}_n^{(1)}(f)| > \frac{x}{2}, B_n^{(2)}(q^2)\big) + \mathbb{P}\big(|\mathbb{G}_n^{(2)}(f)| > \frac{x}{2}, B_n(q)\big) \\
&\le \mathbb{P}\Big(|\mathbb{G}_n^{(1)}(f)| > \frac{x}{2}, R_n^2(f) \le \max\big\{\tilde{V}_n(f)^2, \frac{M\Phi(q)}{\sqrt{n}}x\big\}\Big) \\
&\quad + \mathbb{P}\Big(R_n^2(f) > \max\big\{\tilde{V}_n(f)^2, \frac{M\Phi(q)}{\sqrt{n}}x\big\}, B_n^{(2)}(q^2)\Big) \\
&\quad + \mathbb{P}\big(|\mathbb{G}_n^{(2)}(f)| > \frac{x}{2}, B_n(q)\big). \qquad (2.8.100)
\end{aligned}$$

We now discuss the three summands in (2.8.100) separately. By Theorem 2.6.1,

$$\mathbb{P}\big(|\mathbb{G}_n^{(2)}(f)| > \frac{x}{2}, B_n(q)\big) \le c_0 \exp\Big(-\frac{1}{c_1}\frac{(x/2)^2}{\tilde{V}_n(f)^2 + \frac{M\Phi(q)}{\sqrt{n}}(x/2)}\Big).$$

By Freedman's inequality for martingales, we have

$$\begin{aligned}
&\mathbb{P}\Big(|\mathbb{G}_n^{(1)}(f)| > \frac{x}{2}, R_n^2(f) \le \max\big\{\tilde{V}_n(f)^2, \frac{M\Phi(q)}{\sqrt{n}}x\big\}\Big)\\
\le\ & 2\exp\Big(-\frac{1}{2}\frac{(x/2)^2}{\max\{\tilde{V}_n(f)^2, \frac{M\Phi(q)}{\sqrt{n}}x\} + \frac{M}{\sqrt{n}}x}\Big)\\
\le\ & 2\exp\Big(-\frac{1}{4}\frac{(x/2)^2}{\tilde{V}_n(f)^2 + \frac{M\Phi(q)}{\sqrt{n}}x}\Big).
\end{aligned}$$

By Lemma 2.8.9 applied to $W_i(f) = \mathbb{E}[f(Z_i, \frac{i}{n})^2|Z_{i-1}]$ and using $\Phi(q^2) \le \Phi(q)^2$ (cf. (2.8.101)),

$$\begin{aligned}
&\mathbb{P}\Big(R_n^2(f) > \max\big\{\tilde{V}_n(f)^2, \frac{M\Phi(q)}{\sqrt{n}}x\big\}, B_n^{(2)}(q^2)\Big)\\
\le\ & c_0\exp\Big(-\frac{1}{c_1}\frac{\frac{M\Phi(q)}{\sqrt{n}}x}{\frac{M^2\Phi(q^2)}{n}}\cdot\min\Big\{\frac{\tilde{V}_n(f)^2}{\|f\|_{2,n}V_n(f)}, 1\Big\}\Big)\\
=\ & c_0\exp\Big(-\frac{1}{c_1}\frac{x^2}{\frac{M\Phi(q)}{\sqrt{n}}x}\Big).
\end{aligned}$$

Inserting the above estimates into (2.8.100), equation (2.6.3) is obtained. Furthermore by Assumption 2.6.3,

$$\mathbb{P}(B_n^{(2)}(q^2)^c) \le C_\Delta\frac{n(\mathbb{D}_n^\infty)^2}{M^2}\tilde{\beta}_{norm}(q^2) \le C_\Delta C_{\tilde{\beta}}\Big(\frac{\sqrt{n}\mathbb{D}_n^\infty}{M}\tilde{\beta}_{norm}(q)\Big)^2.$$

Thus,

$$\mathbb{P}(B_n^\circ(q)^c) \le \mathbb{P}(B_n(q)^c) + \mathbb{P}(B_n^{(2)}(q^2)^c) \le [4 + C_\Delta C_{\tilde{\beta}}]\Big(\frac{\sqrt{n}\mathbb{D}_n^\infty}{M}\tilde{\beta}_{norm}(q)\Big)^2.$$

The second assertion (2.6.4) is derived as in Theorem 2.6.1 with $q = \tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty y})$. □

Lemma 2.8.10 (A compatibility lemma)**.** *Let $n \in \mathbb{N}$, $\delta, a_M > 0$ and $k \in \mathbb{N}$. For $H > 0$, put*

$$\tilde{r}(\delta) := \max\{r > 0 : \tilde{q}^*(r)r \le \delta\}$$

and

$$w(H) := \min\{w > 0 : w\cdot\tilde{r}(w) \ge H^{-1}\}, \qquad \mathcal{W}(H) := Hw(H).$$

Define

$$\tilde{m}(n,\delta,k) := a_M\tilde{r}(\frac{\delta}{\mathbb{D}_n})\tilde{r}(w(H(k)))\cdot\mathbb{D}_n^\infty n^{1/2}.$$

Finally, put

$$\hat{C}_n := 8c_1(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})(1 + C_{\tilde{\beta}}^2(\tilde{\beta}(1)\vee 1))(1 + a_M).$$

(i) Then, $\mathcal{W}$ is subadditive.

(ii) If $\mathcal{F}$ fulfills Assumption 2.2.3 and Assumption 2.6.3, then $\sup_{f\in\mathcal{F}}\tilde{V}_n(f) \le \delta$, $\sup_{f\in\mathcal{F}}\|f\|_\infty \le \tilde{m}(n,\delta,k)$ implies that for any $\psi : (0,\infty) \to [1,\infty)$,

$$\begin{aligned}
\mathbb{P}\Big(\frac{1}{\sqrt{n}}\big|S_n(f)\big| > \hat{C}_n\psi(\delta)\delta\mathcal{W}(H(k)), B_n\Big) &\le c_0\exp\big(-2H(k)\big),\\
\sqrt{n}\|f\mathbb{1}_{\{f>\gamma\cdot\tilde{m}(n,\delta,k)\}}\|_{1,n} &\le \frac{4}{\gamma a_M}\cdot\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}\cdot\delta\mathcal{W}(H(k)),\\
\mathbb{P}\big(B_n^c\big) &\le \frac{4}{\psi(\delta)^2 a_M^2},
\end{aligned}$$

where $B_n = B_n(\tilde{q}^*(\frac{m(n,\delta,k)}{\sqrt{n}\mathbb{D}_n^\infty\psi(\delta)a_M}))$, c_0, c_1 are from Theorem 2.6.1.

Proof of Lemma 2.8.10. (i) Note that for $a, b > 0$, we have $w(a+b) \le w(a)$ since $w(a)\tilde{r}(w(a)) \ge a^{-1} \ge (a+b)^{-1}$. Thus,

$$\mathcal{W}(a+b) = (a+b)w(a+b) \le aw(a+b)+bw(a+b) \le aw(a)+bw(b) \le \mathcal{W}(a)+\mathcal{W}(b).$$

(ii) As in the proof of Lemma 3.5.5 (cf. (3.5.30)), we obtain that for $x_1, x_2 > 0$,

$$\tilde{q}^*(C_{\tilde{\beta}}x_1x_2) \le \tilde{q}^*(x_1)\tilde{q}^*(x_2).$$

Furthermore, for $q_1, q_2 \in \mathbb{N}$ we have due to $x_1 + x_2 \le x_1x_2 + 1$ that

$$\begin{aligned}
\log\log(e^e q_1q_2) &\le \log[\log(eq_1) + \log(eq_2)]\\
&\le \log[\log(eq_1)\cdot\log(eq_2) + 1] \le \log[\log(eq_1)\cdot\log(e^e q_2)]\\
&\le \log\log(eq_1) + \log\log(e^e q_2)\\
&\le \log\log(eq_1)\cdot\log\log(e^e q_2) + 1\\
&\le \log\log(e^e q_1)\cdot\log\log(e^e q_2),
\end{aligned}$$

and thus

$$\Phi(q_1q_2) \le \Phi(q_1)\Phi(q_2). \tag{2.8.101}$$

Furthermore, note that for $a \in (0, (\tilde{\beta}(1)\vee 1)]$, $q = \lceil\Phi^{-1}(\frac{(\tilde{\beta}(1)\vee 1)}{a})\rceil$ satisfies

$$\Phi(q)a = \Phi(\lceil\Phi^{-1}(\frac{(\tilde{\beta}(1)\vee 1)}{a})\rceil)a \ge (\tilde{\beta}(1)\vee 1) \ge \tilde{\beta}(q),$$

that is,

$$\begin{aligned}
\Phi(\tilde{q}^*(a)) &\le \Phi(\lceil\Phi^{-1}(\frac{(\tilde{\beta}(1)\vee 1)}{a})\rceil) \le \Phi(2\Phi^{-1}(\frac{(\tilde{\beta}(1)\vee 1)}{a}))\\
&\le 4\Phi(\Phi^{-1}(\frac{(\tilde{\beta}(1)\vee 1)}{a})) \le \frac{4(\tilde{\beta}(1)\vee 1)}{a}.
\end{aligned} \tag{2.8.102}$$

With $y = \psi(\delta)a_M$ we have

$$\tilde{q}^*(\frac{\tilde{m}(n,\delta,k)}{\sqrt{n}\mathbb{D}_n^\infty y}) = \tilde{q}^*(\frac{\tilde{m}(n,\delta,k)}{\sqrt{n}\mathbb{D}_n^\infty \psi(\delta)a_M}) = \tilde{q}^*(\frac{C_{\tilde{\beta}}^2(\tilde{\beta}(1)\vee 1)r_1 r_2}{\psi(\delta)}),$$

where $r_1 = \tilde{r}(\frac{\delta}{\mathbb{D}_n})$, $r_2 = \tilde{r}(w(H(k)))$, and thus with (2.8.101) and (2.8.102),

$$\begin{aligned}
\Phi(\tilde{q}^*)\frac{\tilde{m}(n,\delta,k)}{\sqrt{n}} &\le \Phi(\tilde{q}^*(\frac{C_{\tilde{\beta}}^2(\tilde{\beta}(1)\vee 1)r_1 r_2}{\psi(\delta)}))r_1 r_2 \mathbb{D}_n^\infty a_M \\
&\le \Phi\Big(\tilde{q}^*(\frac{(\tilde{\beta}(1)\vee 1)}{\psi(\delta)})\tilde{q}^*(r_1)\tilde{q}^*(r_2)\Big) r_1 r_2 \mathbb{D}_n^\infty a_M \\
&\le \Phi\Big(\tilde{q}^*(\frac{(\tilde{\beta}(1)\vee 1)}{\psi(\delta)})\Big)\Phi(\tilde{q}^*(r_1))\Phi(\tilde{q}^*(r_2)))r_1 r_2 \mathbb{D}_n^\infty a_M \\
&\le 4\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}\psi(\delta)\delta w(H(k))a_M.
\end{aligned}$$

By definition of $\mathcal{W}(\cdot)$ and Theorem 2.6.1, we obtain

$$\begin{aligned}
&\mathbb{P}\Big(\frac{1}{\sqrt{n}}\big|S_n(f)\big| > \hat{C}_n\psi(\delta)\delta\cdot\mathcal{W}(H(k)), B_n\Big) \\
&\le c_0 \exp\Big(-\frac{1}{c_1}\frac{\hat{C}_n^2\psi(\delta)^2\delta^2\mathcal{W}(H(k))^2}{\delta^2 + 4\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}a_M\hat{C}_n\delta^2\psi(\delta)^2 w(H(k))\mathcal{W}(H(k))}\Big) \\
&\le c_0 \exp\Big(-\frac{1}{c_1}\frac{\hat{C}_n^2}{1+4a_M\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}\hat{C}_n}H(k)\Big) \\
&\le c_0 \exp\big(-2H(k)\big).
\end{aligned}$$

Similar as in the proof of Lemma 2.8.8, we obtain due to Assumption 2.6.3 that

$$\tilde{V}_n(f) \ge \min_{a\in\mathbb{N}}\big[\|f\|_{2,n}\Big(1+\sum_{j=1}^{a}\varphi(j)\Big) + \mathbb{D}_n\tilde{\beta}(a)\big] \ge \|f\|_{2,n}\Big(1+\sum_{j=1}^{\hat{a}}\varphi(j)\Big) + \mathbb{D}_n\tilde{\beta}(\hat{a}),$$

where $\hat{a} = \arg\min_{a\in\mathbb{N}}\{\|f\|_{2,n}\cdot\big(1+\sum_{j=1}^{a}\varphi(j)\big)+\mathbb{D}_n\beta(a)\}$. Elementary calculations show that for $\hat{a}\ge 2$,

$$\begin{aligned}
\sum_{j=1}^{\hat{a}}\varphi(j) &= 1+\sum_{j=2}^{\hat{a}}\varphi(j) \ge 1+\int_1^{\hat{a}-1}\varphi(x)dx \\
&= 1+(\Phi(\hat{a}-1)-1) - \int_1^{\hat{a}-1}\frac{1}{\log(e^e x)}dx \ge \Phi(\hat{a}-1) - \frac{\hat{a}-2}{e} \ge \frac{1}{4}\Phi(\hat{a}).
\end{aligned}$$

Clearly, the same holds true for $\hat{a}=1$. We therefore have

$$\tilde{V}_n(f) \ge \frac{1}{4}\|f\|_{2,n}\Phi(\hat{a}). \qquad (2.8.103)$$

Now, $\delta \geq \tilde{V}_n(f) \geq \mathbb{D}_n\tilde{\beta}(\hat{a}) = \mathbb{D}_n\tilde{\beta}_{norm}(\hat{a})\Phi(\hat{a})$. Thus, $\tilde{\beta}_{norm}(\hat{a}) \leq \frac{\delta}{\mathbb{D}_n\Phi(\hat{a})}$. By definition of $\tilde{q}^*$, $\tilde{q}^*(\frac{\delta}{\mathbb{D}_n\Phi(\hat{a})}) \leq \hat{a}$. So, $\Phi(\tilde{q}^*(\frac{\delta}{\mathbb{D}_n\Phi(\hat{a})}))\frac{\delta}{\mathbb{D}_n\Phi(\hat{a})} \leq \frac{\delta}{\mathbb{D}_n}$. By definition of $\tilde{r}$, $\tilde{r}(\frac{\delta}{\mathbb{D}_n}) \geq \frac{\delta}{\mathbb{D}_n\Phi(\hat{a})}$.

Using this result, (2.8.103) and the definition of $w(\cdot)$ yield

$$\sqrt{n}\|f\mathbb{1}_{\{f>\gamma\cdot\tilde{m}(n,\delta,k)\}}\|_{1,n} \leq \frac{1}{\gamma}\frac{\sqrt{n}\|f\|_{2,n}^2}{\tilde{m}(n,\delta,k)} \leq \frac{1}{\gamma}\frac{1}{a_M\mathbb{D}_n^\infty}\frac{\|f\|_{2,n}^2}{\tilde{r}(\frac{\delta}{\mathbb{D}_n})\tilde{r}(w(H(k)))},$$

and

$$\begin{aligned}\frac{\|f\|_{2,n}^2}{\tilde{r}(\frac{\delta}{\mathbb{D}_n})\tilde{r}(w(H(k)))} &\leq \mathbb{D}_n\frac{\Phi(\hat{a})\|f\|_{2,n}^2}{\delta}\frac{1}{\tilde{r}(w(H(k))}\\ &\leq \mathbb{D}_n\frac{4\tilde{V}_n(f)\|f\|_{2,n}}{\delta}\frac{1}{\tilde{r}(w(H(k))} \leq 4\delta\cdot\frac{1}{\tilde{r}(w(H(k))}\\ &\leq 4\mathbb{D}_n\delta\mathcal{W}(H(k)),\end{aligned}$$

which provides

$$\sqrt{n}\|f\mathbb{1}_{\{f>\gamma\cdot\tilde{m}(n,\delta,k)\}}\|_{1,n} \leq \frac{4}{\gamma}\frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\frac{1}{a_M}\delta\mathcal{W}(H(k)).$$

Finally, Theorem 2.6.1 implies that

$$\mathbb{P}(B_n^c) \leq \frac{4}{y^2} = \frac{4}{\psi(\delta)^2a_M^2}.$$

□

Here, we would like to present a chaining version of the large deviation inequality from Theorem 2.6.1. For the sake of simplicity, we derive the result for some continuous strictly decreasing upper bound $\bar{\mathbb{H}}(\varepsilon)$ of $\mathbb{H}(\varepsilon, \mathcal{F}, V)$.

Theorem 2.8.11 (Chaining for large deviation inequalities)**.** *There exists a universal constant $c_3 > 0$ such that the following holds true.*

Let $a_M \geq 1$, $M, \sigma > 0$ be arbitrary, $\psi(x) := \sqrt{\log(x^{-1}\vee e)}\log\log(x^{-1}\vee e^e)$.

Let $\mathcal{F}$ be a class which satisfies Assumption 2.2.3 and 2.6.3, and $\sup_{f\in\mathcal{F}}\tilde{V}_n(f) \leq \sigma$, $\sup_{f\in\mathcal{F}}\|f\|_\infty \leq M$. Define

$$\tilde{I}(\sigma) := \int_0^\sigma \psi(\varepsilon)\mathcal{W}(1\vee\bar{\mathbb{H}}(\varepsilon))d\varepsilon.$$

Choose $\sigma^\circ, x > 0$ such that

$$\bar{\mathbb{H}}(\sigma^\circ) = \frac{1}{50c_1}\cdot\frac{x^2}{\sigma^2 + \Phi(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))\frac{Mx}{\sqrt{n}}}, \qquad x \geq c_3\hat{C}_n\tilde{I}(\sigma^\circ), \tag{2.8.104}$$

where $\hat{C}_n, \mathcal{W}$ is from Lemma 2.8.10. Then there exists a set Ω_n independent of x such that

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\big|\mathbb{G}_n(f)\big| > x, \Omega_n\Big) \leq 5\exp\Big(-\frac{1}{50c_1}\cdot\frac{x^2}{\sigma^2+\Phi(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))\frac{Mx}{\sqrt{n}}}\Big)$$

and

$$\mathbb{P}(\Omega_n^c) \leq \frac{16}{a_M}\int_0^{\sigma^\circ}\frac{1}{x\psi(x)^2}dx.$$

Proof of Theorem 2.8.11. We use the chaining technique from Alexander [1984], Theorem 2.3 therein.

We define $\delta_0 := \sigma^\circ$,

$$\delta_{j+1} := \max\{\delta \leq \frac{\delta_j}{2} : \bar{\mathbb{H}}(\delta) \geq 4\bar{\mathbb{H}}(\delta_j)\}.$$

Since $\bar{\mathbb{H}}(\cdot)$ is continuous, $\bar{\mathbb{H}}(\delta_{j+1}) = 4\bar{\mathbb{H}}(\delta_j)$. Put

$$\tau := \min\{j \geq 0 : \delta_j \leq \frac{\tilde{I}(\sigma^\circ)}{\sqrt{n}}\}.$$

Define

$$\eta_j := 4\hat{C}_n\psi(\delta_j)\delta_j\mathcal{W}(H(\bar{N}_{j+1})),$$

where $\hat{C}_n$ is from Lemma 2.8.10 and

$$\bar{N}_{j+1} := \prod_{k=0}^{j+1}\exp(\bar{\mathbb{H}}(\delta_k)) \geq \prod_{k=0}^{j+1}\exp(\mathbb{H}(\delta_k)) = \prod_{k=0}^{j+1}\mathbb{N}(\delta_k) =: N_{j+1}.$$

By Lemma 2.8.10(i), $\mathcal{W}(\cdot)$ is subadditive, whence

$$\begin{aligned}\sum_{j=0}^{\tau}\psi(\delta_j)\delta_j\mathcal{W}(H(\bar{N}_{j+1})) &\leq \sum_{j=0}^{\tau}\psi(\delta_j)\delta_j\mathcal{W}(1\vee\sum_{k=1}^{j+1}\bar{\mathbb{H}}(\delta_k))\\ &\leq \sum_{j=0}^{\tau}\psi(\delta_j)\delta_j\sum_{k=1}^{j+1}\mathcal{W}(1\vee\bar{\mathbb{H}}(\delta_k))\\ &\leq \sum_{k=0}^{\tau-1}\mathcal{W}(1\vee\bar{\mathbb{H}}(\delta_{k+1}))\sum_{j=k}^{\tau}\psi(\delta_j)\delta_j. \end{aligned} \quad (2.8.105)$$

Similar to (3.5.38), there exists some universal constant $c_\psi > 0$ such that

$$\begin{aligned}\sum_{j=k}^{\tau}\psi(\delta_j)\delta_j &\leq 2\sum_{j=k}^{\infty}\int_{\delta_j/2}^{\delta_j}\psi(\delta_j)dx \leq 2\sum_{j=k}^{\infty}\int_{\delta_{j+1}}^{\delta_j}\psi(x)dx\\ &\leq 2\int_0^{\delta_k}\psi(x)dx \leq 2c_\psi\delta_k\psi(\delta_k). \end{aligned} \quad (2.8.106)$$

Furthermore, by definition of the sequence $(\delta_j)_j$ and since $w(\cdot)$ is decreasing but $\mathcal{W}$ is increasing, we have

$$\mathcal{W}(1 \vee \bar{\mathbb{H}}(\delta_{j+1})) \leq \mathcal{W}(4(1 \vee \bar{\mathbb{H}}(\delta_j))) \leq 4\mathcal{W}(1 \vee \bar{\mathbb{H}}(\delta_j)). \tag{2.8.107}$$

Insertion of (2.8.106) and (2.8.107) into (2.8.105) yields

$$\begin{aligned}
\sum_{j=0}^{\tau} \eta_j &\leq 4\hat{C}_n \sum_{j=0}^{\tau} \psi(\delta_j)\delta_j \mathcal{W}(H(\bar{N}_{j+1})) \\
&\leq 32 c_\psi \hat{C}_n \sum_{k=0}^{\infty} \delta_k \psi(\delta_k)\mathcal{W}(1 \vee \bar{\mathbb{H}}(\delta_j)) \\
&\leq 64 c_\psi \hat{C}_n \sum_{k=0}^{\infty} \int_{\delta_k/2}^{\delta_k} \psi(\delta_k)\mathcal{W}(1 \vee \bar{\mathbb{H}}(\delta_j)) d\varepsilon \\
&\leq 64 c_\psi \hat{C}_n \sum_{k=0}^{\infty} \int_{\delta_{k+1}}^{\delta_k} \psi(\varepsilon)\mathcal{W}(1 \vee \bar{\mathbb{H}}(\varepsilon)) d\varepsilon \\
&\leq 64 c_\psi \hat{C}_n \int_0^{\sigma^\circ} \psi(\varepsilon)\mathcal{W}(1 \vee \bar{\mathbb{H}}(\varepsilon)) d\varepsilon = 64 c_\psi \hat{C}_n \tilde{I}(\sigma^\circ).
\end{aligned} \tag{2.8.108}$$

We set up the same decomposition as in the proof of Theorem 2.4.4. Define

$$\tilde{m}_j := \frac{1}{2}\tilde{m}(n, \delta_j, \bar{N}_{j+1}).$$

Note that

$$x \geq \frac{x}{5} + 3(\frac{x}{5} - 2\eta_\tau) + (\frac{x}{5} + 2\eta_\tau).$$

Set $c_3 := 5 \cdot 2^8 \cdot c_\psi$. Condition (2.8.104) implies

$$\frac{x}{5} \geq 2^8 c_\psi \hat{C}_n \tilde{I}(\sigma^\circ) \tag{2.8.109}$$

and thus with (2.8.108), we obtain

$$\frac{x}{5} - 2\eta_\tau \geq 2^8 c_\psi \hat{C}_n \tilde{I}(\sigma^\circ) - 2\eta_\tau \geq \sum_{j=0}^{\tau-1} 2\eta_j.$$

Let $\tilde{q}_j^* := \tilde{q}^*(\frac{m(n,\delta_j,N_{j+1})}{\sqrt{n}\mathbb{D}_n^\infty \psi(\delta_j) a_M})$ and

$$\Omega_n := B_n(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M})) \cap \bigcap_{j=0}^{\tau} B_n(\tilde{q}_j^*),$$

where $B_n(q)$ is from Theorem 2.6.1. From (2.8.57), we obtain the decomposition

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\big|\mathbb{G}_n(f)\big| > x, \Omega_n\Big)\\
\le\ & \mathbb{P}\Big(\sup_{f\in\mathcal{F}}|\mathbb{G}_n(\pi_0 f)| > \frac{x}{5}, \Omega_n\Big)\\
&\quad +\mathbb{P}\Big(\sup_{f\in\mathcal{F}}|\mathbb{G}_n(\varphi^{\wedge}_{\tilde{m}_\tau}(\Delta_\tau f))| + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_\tau f\|_{1,n} > \frac{x}{5} + 2\eta_\tau, \Omega_n\Big)\\
&\quad +\mathbb{P}\Big(\sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\varphi^{\wedge}_{\tilde{m}_j-\tilde{m}_{j+1}}(\pi_{j+1}f-\pi_j f))\Big| > \frac{x}{5} - 2\eta_\tau, \Omega_n\Big)\\
&\quad +\mathbb{P}\Big(\sum_{j=0}^{\tau-1}\Big\{\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{\tilde{m}_{j+1}}(\Delta_{j+1}f)\big|, 2\tilde{m}_j\big\})\Big|\\
&\qquad\qquad +2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_{j+1}f\mathbb{1}_{\{\Delta_{j+1}f>\tilde{m}_{j+1}\}}\|_{1,n}\Big\} > \frac{x}{5} - 2\eta_\tau, \Omega_n\Big)\\
&\quad +\mathbb{P}\Big(\sum_{j=0}^{\tau-1}\Big\{\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{\tilde{m}_j-\tilde{m}_{j+1}}(\Delta_j f)\big|, 2\tilde{m}_j\big\})\Big|\\
&\qquad\qquad +2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>\tilde{m}_j-\tilde{m}_{j+1}\}}\|_{1,n}\Big\} > \frac{x}{5} - 2\eta_\tau, \Omega_n\Big)\\
\le\ & \mathbb{P}\Big(\sup_{f\in\mathcal{F}}|\mathbb{G}_n(\pi_0 f)| > \frac{x}{5}, \Omega_n\Big)\\
&\quad +\mathbb{P}\Big(\sup_{f\in\mathcal{F}}|\mathbb{G}_n(\varphi^{\wedge}_{\tilde{m}_\tau}(\Delta_\tau f))| + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_\tau f\|_{1,n} > \frac{x}{5} + 2\eta_\tau, \Omega_n\Big)\\
&\quad +\sum_{j=0}^{\tau-1}\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\varphi^{\wedge}_{\tilde{m}_j-\tilde{m}_{j+1}}(\pi_{j+1}f-\pi_j f))\Big| > 2\eta_j, \Omega_n\Big)\\
&\quad +\sum_{j=0}^{\tau-1}\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{\tilde{m}_{j+1}}(\Delta_{j+1}f)\big|, 2\tilde{m}_j\big\})\Big|\\
&\qquad\qquad +2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_{j+1}f\mathbb{1}_{\{\Delta_{j+1}f>\tilde{m}_{j+1}\}}\|_{1,n} > 2\eta_j, \Omega_n\Big)\\
&\quad +\sum_{j=0}^{\tau-1}\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{\tilde{m}_j-\tilde{m}_{j+1}}(\Delta_j f)\big|, 2\tilde{m}_j\big\})\Big|\\
&\qquad\qquad +2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>\tilde{m}_j-\tilde{m}_{j+1}\}}\|_{1,n} > 2\eta_j, \Omega_n\Big)\\
=:\ & R_1^* + R_2^* + R_3^* + R_4^* + R_5^*. \qquad (2.8.110)
\end{aligned}$$

We now discuss the terms in (2.8.110) separately.

- We have by definition of Theorem 2.6.1,

$$\begin{aligned} R_1^* &\le \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n(\pi_0 f)| > \frac{x}{5}, B_n(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))\Big) \\ &\le \mathbb{N}(\sigma^\circ)\cdot \sup_{f\in\mathcal{F}} \mathbb{P}\Big(|\mathbb{G}_n(\pi_0 f)| > \frac{x}{5}, B_n(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))\Big) \\ &\le \exp(\mathbb{H}(\sigma^\circ))\cdot c_0 \exp\Big(-\frac{1}{c_1}\frac{(x/5)^2}{\sigma^2 + \Phi(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))\frac{M(x/5)}{\sqrt{n}}}\Big) \\ &\le c_0 \exp\Big(-\frac{1}{50c_1}\frac{x^2}{\sigma^2 + \Phi(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))\frac{M}{\sqrt{n}}}\Big). \end{aligned}$$

- We have by Lemma 2.8.10,

$$\begin{aligned} &\mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n(\varphi^{\wedge}_{\tilde{m}_\tau}(\Delta_\tau f))| > \eta_\tau, B_n(\tilde{q}_\tau^*)\Big) \\ &\le \exp(H(N_{\tau+1}))\cdot c_0 \exp(-2H(N_{\tau+1})) \\ &\le c_0 \sum_{j=0}^\infty \exp(-H(N_{j+1})) \le c_0 \exp(-\mathbb{H}(\sigma^\circ)) \end{aligned}$$

(for details on the last inequality look at the calculation for R_3^*, below). By the Cauchy-Schwarz inequality, the definition of τ and (2.8.109),

$$\sqrt{n}\sup_{f\in\mathcal{F}} \|\Delta_\tau f\|_{1,n} \le \sqrt{n}\|\Delta_\tau f\|_{2,n} \le \sqrt{n} V(\Delta_\tau f) \le \sqrt{n}\delta_\tau \le \tilde{I}(\sigma^\circ) < \frac{x}{5}.$$

We conclude that

$$R_2^* \le c_0 \exp(-\mathbb{H}(\sigma^\circ)).$$

- We have by Lemma 2.8.10(i),

$$\begin{aligned} R_3^* &\le \sum_{j=0}^{\tau-1} \mathbb{P}\Big(\sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n(\varphi^{\wedge}_{\tilde{m}_j - \tilde{m}_{j+1}}(\pi_{j+1} f - \pi_j f))\Big| > 2\eta_j, B_n(\tilde{q}_j^*)\Big) \\ &\le \sum_{j=0}^{\tau-1} N_{j+1}\cdot \sup_{f\in\mathcal{F}} \mathbb{P}\Big(\Big|\mathbb{G}_n(\varphi^{\wedge}_{\tilde{m}_j - \tilde{m}_{j+1}}(\pi_{j+1} f - \pi_j f))\Big| > \eta_j, B_n(\tilde{q}_j^*)\Big) \\ &\le \sum_{j=0}^{\tau-1} \exp(H(N_{j+1}))\cdot c_0 \exp(-2H(\bar{N}_{j+1})) \\ &\le c_0 \sum_{j=0}^{\tau-1} \exp(-H(\bar{N}_{j+1})) \le c_0 \sum_{j=0}^{\tau-1} \exp(-\bar{\mathbb{H}}(\delta_{j+1})) \\ &\le c_0 \sum_{j=0}^{\infty} \exp(-4^{j+1}\bar{\mathbb{H}}(\sigma^\circ)) \\ &\le c_0 \exp(-\bar{\mathbb{H}}(\sigma^\circ)). \end{aligned}$$

The last inequality is due to the fact that

$$\Big(\sum_{j=0}^{\infty}\exp(-4^{j+1}\bar{\mathbb{H}}(\sigma^\circ))\Big)\exp(\bar{\mathbb{H}}(\sigma^\circ)) = \sum_{j=0}^{\infty}\exp(-(4^{j+1}-1)\bar{\mathbb{H}}(\sigma^\circ))$$
$$\leq \sum_{j=0}^{\infty}\exp(-(4^{j+1}-1)) \leq 1. \tag{2.8.111}$$

- Similar to R_3^* we have by Lemma 2.8.10(ii),

$$\begin{aligned}
&\sum_{j=0}^{\tau-1}\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{\tilde{m}_{j+1}}(\Delta_{j+1}f)\big|, 2\tilde{m}_j\big\})\Big| > \eta_j, B_n(\tilde{q}_j^*)\Big)\\
\leq\ & \sum_{j=0}^{\tau-1} N_{j+1}\cdot c_0\exp(-2H(\bar{N}_{j+1}))\\
\leq\ & c_0\exp(-\bar{\mathbb{H}}(\sigma^\circ))
\end{aligned}$$

and, since $a_M \geq 1$,

$$\begin{aligned}
\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_{j+1}f\mathbb{1}_{\{\Delta_{j+1}f>\tilde{m}_{j+1}\}}\|_{1,n} &\leq 8\frac{\mathbb{D}_n^{\infty}}{\mathbb{D}_n}\delta_j\mathcal{W}(H(\bar{N}_{j+1}))\\
&< \hat{C}_n\cdot\delta_j\mathcal{W}(H(\bar{N}_{j+1})) \leq \frac{1}{2}\eta_j.
\end{aligned}$$

This shows

$$R_4^* \leq c_0\exp(-\mathbb{H}(\sigma^\circ)).$$

- Similar to R_4^* we obtain

$$\begin{aligned}
&\sum_{j=0}^{\tau-1}\mathbb{P}\Big(\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n(\min\big\{\big|\varphi^{\vee}_{\tilde{m}_j-\tilde{m}_{j+1}}(\Delta_j f)\big|, 2\tilde{m}_j\big\})\Big| > \eta_j, B_n(\tilde{q}_j^*\Big)\\
\leq\ & \sum_{j=0}^{\tau-1} N_{j+1}\cdot c_0\exp(-2H(\bar{N}_{j+1}))\\
\leq\ & c_0\exp(-\bar{\mathbb{H}}(\sigma^\circ)).
\end{aligned}$$

As in the proof of Theorem 2.4.4 (discussion of R_5 therein), we see that $2(\tilde{m}_j - \tilde{m}_{j+1}) \geq \tilde{m}_j$ due to the fact that the inequality

$$\tilde{r}(\frac{\delta_j}{\mathbb{D}_n}) - \tilde{r}(\frac{\delta_{j+1}}{\mathbb{D}_n}) \geq \tilde{r}(\frac{\delta_j}{\mathbb{D}_n}) - \tilde{r}(\frac{\delta_j}{2\mathbb{D}_n}) \geq \tilde{r}(\frac{\delta_j}{\mathbb{D}_n}) - \frac{1}{2}\tilde{r}(\frac{\delta_j}{\mathbb{D}_n}) \geq \frac{1}{2}\tilde{r}(\frac{\delta_j}{\mathbb{D}_n}).$$

only requires $\delta_{j+1} \leq \frac{\delta_j}{2}$. Thus, since $a_M \geq 1$,

$$\begin{aligned}
\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>\tilde{m}_j-\tilde{m}_{j+1}\}}\|_{1,n} &\leq \sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>\frac{1}{2}\tilde{m}_j\}}\|_{1,n}\\
&\leq 16\frac{\mathbb{D}_n^{\infty}}{\mathbb{D}_n}\delta_j\mathcal{W}(H(\bar{N}_{j+1}))\\
&< 2\hat{C}_n\delta_j\mathcal{W}(H(\bar{N}_{j+1})) \leq \frac{1}{2}\eta_j.
\end{aligned}$$

This shows

$$R_5^* \le c_0 \exp(-\bar{\mathbb{H}}(\sigma^\circ)).$$

By plugging in the above upper bounds for R_i^*, $i \in \{1, ..., 5\}$, into (2.8.110) and using (2.8.104), we obtain

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n(f)| > x, \Omega_n\Big) \le 5c_0 \exp\big(-\frac{1}{50c_1} \cdot \frac{x^2}{\sigma^2 + \Phi(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))\frac{Mx}{\sqrt{n}}}\big). \quad (2.8.112)$$

Discussion of the residual term: By Lemma 2.8.10(ii), we have

$$\begin{aligned}\mathbb{P}(\Omega_n^c) &\le \mathbb{P}(B_n(\tilde{q}^*(\frac{M}{\sqrt{n}\mathbb{D}_n^\infty a_M}))^c) + \sum_{j=0}^\infty \mathbb{P}(B_n(\tilde{q}_j^*)^c)\\ &\le \frac{4}{a_M^2} + \frac{4}{a_M^2}\sum_{j=0}^\infty \frac{1}{\psi(\delta_j)^2} \le \frac{8}{a_M^2}\sum_{j=0}^\infty \frac{1}{\psi(\delta_j)^2}.\end{aligned}$$

Due to

$$\begin{aligned}\sum_{j=0}^\infty \frac{1}{\psi(\delta_j)^2} &= \sum_{j=0}^\infty \frac{1}{\delta_j - \delta_{j+1}}\int_{\delta_{j+1}}^{\delta_j} \frac{1}{\psi(\delta_j)^2}dx\\ &\le 2\sum_{j=0}^\infty \int_{\delta_{j+1}}^{\delta_j} \frac{1}{\delta_j\psi(\delta_j)^2}dx \le 2\int_0^{\sigma^\circ} \frac{1}{x\psi(x)^2}dx \le \frac{4}{\log(\log((\sigma^\circ)^{-1} \vee e^e))},\end{aligned}$$

the proof is completed. □

2.8.7 Form of the V_n-norm and connected quantities

Lemma 2.8.12 (Summation of polynomial and geometric decay)**.** *Let $\alpha > 1$ and $q \in \mathbb{N}$. Then,*

(i)

$$\frac{1}{\alpha - 1}q^{-\alpha+1} \le \sum_{j=q}^\infty j^{-\alpha} \le \frac{\max\{\alpha, 2^{-\alpha+1}\}}{\alpha - 1}q^{-\alpha+1}.$$

(ii) For $\sigma > 0$, $\kappa_2 \ge 1$,

$$\begin{aligned}b_{\rho,\kappa_2,l} \cdot \sigma \cdot \log(\sigma^{-1}) \le \sum_{j=1}^\infty \min\{\sigma, \kappa_2\rho^j\} &\le b_{\rho,\kappa_2} \cdot \sigma \cdot \log(\sigma^{-1} \vee e),\\ b_{\alpha,\kappa_2,l} \cdot \sigma \cdot \sigma^{-\frac{1}{\alpha}} \le \sum_{j=1}^\infty \min\{\sigma, \kappa_2 j^{-\alpha}\} &\le b_{\alpha,\kappa_2} \cdot \sigma \cdot \max\{\sigma^{-\frac{1}{\alpha}}, 1\},\end{aligned}$$

where $b_{\rho,\kappa_2}, b_{\rho,\kappa_2,l}$, $b_{\alpha,\kappa_2}, b_{\alpha,\kappa_2,l}$ are constants only depending on ρ, κ_2, α.

Proof of Lemma 2.8.12. (i) Upper bound: If $q \geq 2$, then

$$\sum_{j=q}^{\infty} j^{-\alpha} = \sum_{j=q}^{\infty} \int_{j-1}^{j} j^{-\alpha} dx \leq \sum_{j=q}^{\infty} \int_{j-1}^{j} x^{-\alpha} dx = \int_{q-1}^{\infty} x^{-\alpha} dx = \frac{1}{-\alpha+1} x^{-\alpha+1}\Big|_{q-1}^{\infty}$$
$$= \frac{1}{\alpha-1}(q-1)^{-\alpha+1} = \frac{1}{\alpha-1} q^{-\alpha+1} \cdot (\frac{q-1}{q})^{-\alpha+1} \leq \frac{2^{-\alpha+1}}{\alpha-1} q^{-\alpha+1}.$$

If $q = 1$, then $\sum_{j=q}^{\infty} j^{-\alpha} = 1 + \sum_{j=q+1}^{\infty} j^{-\alpha} \leq 1 + \frac{1}{\alpha-1} q^{-\alpha+1} = \frac{\alpha}{\alpha-1}$.

Lower bound: Using similar decomposition arguments as above, we have

$$\sum_{j=q}^{\infty} j^{-\alpha} \geq \sum_{j=q}^{\infty} \int_{j}^{j+1} x^{-\alpha} dx = \int_{q}^{\infty} x^{-\alpha} dx = \frac{1}{-\alpha+1} x^{-\alpha+1}\Big|_{q}^{\infty} = \frac{1}{\alpha-1} q^{-\alpha+1}.$$

(ii) • *Exponential decay:* Upper bound: First let $a := \max\{\lfloor \frac{\log(\sigma/\kappa_2)}{\log(\rho)} \rfloor, 0\} + 1$. Then we have

$$\sum_{j=0}^{\infty} \min\{\sigma, \kappa_2 \rho^j\} \leq \sum_{j=0}^{a-1} \sigma + \kappa_2 \sum_{j=a}^{\infty} \rho^j = a\sigma + \kappa_2 \frac{\rho^a}{1-\rho}$$
$$\leq a\sigma + \frac{\kappa_2}{1-\rho} \min\{\frac{\sigma}{\kappa_2}, 1\} \leq a\sigma + \frac{\sigma}{1-\rho}$$
$$\leq \sigma \cdot \Big[\frac{1}{\log(\rho^{-1})} \max\{\log(\kappa_2/\sigma), 0\} + \frac{2}{1-\rho}\Big]$$
$$\leq \sigma \cdot \Big[\frac{1}{\log(\rho^{-1})} \max\{\log(\sigma^{-1}), 0\} + \frac{\log(\kappa_2) \vee 0}{\log(\rho^{-1})} + \frac{2}{1-\rho}\Big]$$
$$\leq b_{\rho,\kappa_2} \cdot \sigma \cdot \log(\sigma^{-1} \vee e),$$

where $b_{\rho,\kappa_2} := 2(\log(\kappa_2) \vee 1) \cdot \frac{1}{\log(\rho^{-1})}\big[1 + \frac{2\log(\rho^{-1})}{1-\rho}\big]$.

Lower Bound: Put $\beta(q) = \kappa_2 \sum_{j=q}^{\infty} \rho^j = \frac{\kappa_2}{1-\rho}\rho^q$. Then,

$$\sum_{j=1}^{\infty} \min\{\sigma, \kappa_2 \rho^j\} \geq \sigma(\hat{q} - 1) + \beta(\hat{q}),$$

where $\hat{q} = \min\{q \in \mathbb{N} : \frac{\sigma}{\kappa_2} \geq \rho^q\}$. We have $\hat{q} \geq \frac{\log(\sigma/\kappa_2)}{\log(\rho)} =: \underline{q}$ and $\hat{q} \leq \underline{q} + 1$. Thus,

$$\sum_{j=1}^{\infty} \min\{\sigma, \kappa_2 \rho^j\} \geq \sigma(\underline{q} - 1) + \beta(\underline{q} + 1).$$

Now consider the case $\frac{\sigma}{\kappa_2} < \rho^2$, that is, $\frac{\log(\sigma/\kappa_2)}{\log(\rho)} \geq 2$. Then, $\underline{q} - 1 \geq \frac{1}{2}\underline{q}$, and $\underline{q} \leq 2\frac{\log(\sigma/\kappa_2)}{\log(\rho)}$. We obtain

$$\sum_{j=1}^{\infty} \min\{\sigma, \kappa_2 \rho^j\} \geq \frac{1}{2}\sigma \frac{\log(\sigma/\kappa_2)}{\log(\rho)} + \frac{\kappa_2 \rho}{1-\rho} \rho^{\frac{\log(\sigma/\kappa_2)}{\log(\rho)}} = \frac{1}{2}\sigma \frac{\log(\sigma/\kappa_2)}{\log(\rho)} + \frac{\rho}{1-\rho}\sigma$$
$$\geq \frac{1}{2}\Big(\frac{\rho}{1-\rho} + \frac{1}{\log(\rho^{-1})}\Big)\sigma \log(\sigma^{-1}\kappa_2),$$

that is, the assertion holds true for $b_{\rho,\kappa_2,l} := \frac{1}{2}\big(\frac{\rho}{1-\rho} + \frac{1}{\log(\rho^{-1})}\big)$.

- *Polynomial decay:* Upper bound: Let $a := \lfloor (\frac{\sigma}{\kappa_2})^{-\frac{1}{\alpha}} \rfloor + 1 \geq (\frac{\sigma}{\kappa_2})^{-\frac{1}{\alpha}}$. Then we have by (i),

$$\begin{aligned}
\sum_{j=1}^{\infty} \min\{\sigma, \kappa_2 j^{-\alpha}\} &\leq \sum_{j=1}^{a} \sigma + \kappa_2 \sum_{j=a+1}^{\infty} j^{-\alpha} = a\sigma + \frac{\kappa_2}{\alpha-1} a^{-\alpha+1} \\
&\leq a\sigma + \frac{\kappa_2^{\frac{1}{\alpha}}}{\alpha-1} \sigma^{\frac{\alpha-1}{\alpha}} \\
&\leq \sigma \cdot \left[\kappa_2^{\frac{1}{\alpha}} \sigma^{-\frac{1}{\alpha}} + 1 + \frac{\kappa_2^{\frac{1}{\alpha}}}{\alpha-1} \sigma^{-\frac{1}{\alpha}}\right] \\
&\leq \sigma \cdot \left[\frac{\alpha}{\alpha-1} \kappa_2^{\frac{1}{\alpha}} \sigma^{-\frac{1}{\alpha}} + 1\right] \\
&\leq b_{\alpha,\kappa_2} \cdot \sigma \cdot \max\{\sigma^{-\frac{1}{\alpha}}, 1\},
\end{aligned}$$

where $b_{\alpha,\kappa_2} := 2\frac{\alpha}{\alpha-1}(\kappa_2 \vee 1)^{\frac{1}{\alpha}}$.

Lower Bound: Put $\beta(q) = \kappa_2 \sum_{j=q}^{\infty} j^{-\alpha}$. By (i), $\beta(q) \geq \frac{\kappa_2}{\alpha-1} q^{-\alpha+1}$. Then,

$$\begin{aligned}
\sum_{j=1}^{\infty} \min\{\sigma, \kappa_2 j^{-\alpha}\} &\geq \min_{q\in\mathbb{N}} \{\sigma q + \beta(q)\} \\
&\geq \min_{q\in\mathbb{N}} \{\sigma q + \frac{\kappa_2}{\alpha-1} q^{-\alpha+1}\}.
\end{aligned}$$

Elementary calculus yields that the minimum is achieved for $q = \kappa_2^{\frac{1}{\alpha}} \cdot \sigma^{-\frac{1}{a}} = (\frac{\kappa_2}{\sigma})^{\frac{1}{\alpha}}$, that is,

$$\sum_{j=1}^{\infty} \min\{\sigma, \kappa_2 j^{-\alpha}\} \geq \frac{\alpha}{\alpha-1} \kappa_2^{\frac{1}{\alpha}} \cdot \sigma^{\frac{\alpha-1}{\alpha}}.$$

The assertion holds with $b_{\alpha,\kappa_2,l} := \frac{\alpha}{\alpha-1}\kappa_2^{\frac{1}{\alpha}}$.

□

Lemma 2.8.13 (Values of q^*, $r(\delta)$)**.** *Depending on specific decay rates, the following statements holds true.*

- *Polynomial decay $\Delta(j) = \kappa j^{-\alpha}$ ($\alpha > 1$). There exist constants $c_{\alpha,\kappa}^{(i)}, C_{\alpha,\kappa}^{(i)} > 0$, $i = 1, 2$ only depending on κ, α such that*

$$c_{\alpha,\kappa}^{(1)} \max\{x^{-\frac{1}{\alpha}}, 1\} \leq q^*(x) \leq C_{\alpha,\kappa}^{(1)} \max\{x^{-\frac{1}{\alpha}}, 1\}$$

and

$$c_{\alpha,\kappa}^{(2)} \min\{\delta^{\frac{\alpha}{\alpha-1}}, \delta\} \leq r(\delta) \leq C_{\alpha,\kappa}^{(2)} \min\{\delta^{\frac{\alpha}{\alpha-1}}, \delta\}.$$

- *Geometric decay $\Delta(j) = \kappa\rho^j$ ($\rho \in (0,1)$). There exist constants $c^{(i)}_{\rho,\kappa}, C^{(i)}_{\rho,\kappa} > 0$, $i = 1,2$ only depending on κ, ρ such that*

$$c^{(1)}_{\rho,\kappa} \max\{\log(x^{-1}), 1\} \leq q^*(x) \leq C^{(1)}_{\rho,\kappa} \max\{\log(x^{-1}), 1\}$$

and

$$c^{(2)}_{\rho,\kappa} \frac{\delta}{\log(\delta^{-1} \vee e)} \leq r(\delta) \leq C^{(2)}_{\rho,\kappa} \frac{\delta}{\log(\delta^{-1} \vee e)}.$$

Proof of Lemma 2.8.13. (i) By Lemma 2.8.12(i), $\beta_{norm}(q) = \frac{\beta(q)}{q} \in [c_{\alpha,\kappa} q^{-\alpha}, C_{\alpha,\kappa} q^{-\alpha}]$ with $c_{\alpha,\kappa} = \frac{\kappa}{\alpha-1}$, $C_{\alpha,\kappa} = \kappa\frac{\max\{\alpha, 2^{-\alpha+1}\}}{\alpha-1}$. In the following we assume w.l.o.g. that $C_{\alpha,\kappa} > 1$ and $c_{\alpha,\kappa} < 1$.

- $q^*(x)$ Upper bound: For any $x > 0$,

$$q^*(x) = \min\{q \in \mathbb{N} : \beta_{norm}(q) \leq x\} \leq \min\{q \in \mathbb{N} : q \geq (\frac{x}{C_{\alpha,\kappa}})^{-\frac{1}{\alpha}}\} = \lceil (\frac{x}{C_{\alpha,\kappa}})^{-\frac{1}{\alpha}} \rceil.$$

Especially, $q^*(x) \leq (\frac{x}{C_{\alpha,\kappa}})^{-\frac{1}{\alpha}} + 1 \leq 2C_{\alpha,\kappa}^{\frac{1}{\alpha}} \max\{x^{-\frac{1}{\alpha}}, 1\}$. The assertion holds true for $C^{(1)}_{\alpha,\kappa} := 2\max\{C_{\alpha,\kappa}, 1\}^{\frac{1}{\alpha}}$.

- $q^*(x)$ Lower bound: Similar to the above calculations,

$$q^*(x) \geq \lceil (\frac{x}{c_{\alpha,\kappa}})^{-\frac{1}{\alpha}} \rceil \geq \big(\frac{x}{c_{\alpha,\kappa}}\big)^{-\frac{1}{\alpha}} = c_{\alpha,\kappa}^{\frac{1}{\alpha}} x^{-\frac{1}{\alpha}}.$$

On the other hand, $q^*(x) \geq 1 \geq c_{\alpha,\kappa}^{\frac{1}{\alpha}}$, which yields the assertion with $c^{(1)}_{\alpha,\kappa} = \min\{c_{\alpha,\kappa}, 1\}^{\frac{1}{\alpha}}$.

- $r(\delta)$ Upper bound: Put $r = 2^{\frac{\alpha}{\alpha-1}} c_{\alpha,\kappa}^{-\frac{1}{\alpha-1}} \delta^{\frac{\alpha}{\alpha-1}}$. Then we have

$$q^*(r)r \geq \lceil (\frac{r}{c_{\alpha,\kappa}})^{-\frac{1}{\alpha}} \rceil r = 2^{\frac{\alpha}{\alpha-1}} c_{\alpha,\kappa}^{-\frac{1}{\alpha-1}} \lceil 2^{-\frac{1}{\alpha-1}} c_{\alpha,\kappa}^{\frac{1}{\alpha-1}} \delta^{-\frac{1}{\alpha-1}} \rceil \delta^{\frac{\alpha}{\alpha-1}} \geq 2\delta > \delta.$$

By definition of $r(\cdot)$, $r(\delta) \leq r$. It was already shown in Lemma 2.8.6 that $r(\delta) \leq \delta$ holds true for all $\delta > 0$. We obtain the assertion with $C^{(2)}_{\alpha,\kappa} = 2^{\frac{\alpha}{\alpha-1}} c_{\alpha,\kappa}^{-\frac{1}{\alpha-1}}$.

- $r(\delta)$ Lower bound: First consider the case $\delta < C_{\alpha,\kappa}$.

Put $r = 2^{-\frac{\alpha}{\alpha-1}} C_{\alpha,\kappa}^{-\frac{1}{\alpha-1}} \delta^{\frac{\alpha}{\alpha-1}}$. Since $x := 2^{\frac{1}{\alpha-1}} C_{\alpha,\kappa}^{\frac{1}{\alpha-1}} \delta^{-\frac{1}{\alpha-1}} > 1$, $\lceil x \rceil \leq 2x$ and thus

$$q^*(r)r \leq \lceil (\frac{r}{C_{\alpha,\kappa}})^{-\frac{1}{\alpha}} \rceil r = 2^{-\frac{\alpha}{\alpha-1}} C_{\alpha,\kappa}^{-\frac{1}{\alpha-1}} \lceil 2^{\frac{1}{\alpha-1}} C_{\alpha,\kappa}^{\frac{1}{\alpha-1}} \delta^{-\frac{1}{\alpha-1}} \rceil \delta^{\frac{\alpha}{\alpha-1}} \leq 2 \cdot 2^{-1}\delta \leq \delta.$$

By definition of $r(\cdot)$, $r(\delta) \geq r = 2^{-\frac{\alpha}{\alpha-1}} \min\{(\frac{\delta}{C_{\alpha,\kappa}})^{\frac{1}{\alpha-1}}, 1\}\delta$.

In the case $\delta > C_{\alpha,\kappa}$, we have

$$q^*(\delta)\delta = \lceil (\frac{\delta}{C_{\alpha,\kappa}})^{-\frac{1}{\alpha}} \rceil \delta \leq 1 \cdot \delta \leq \delta.$$

Thus, $r(\delta) \geq \delta = \min\{(\frac{\delta}{C_{\alpha,\kappa}})^{\frac{1}{\alpha-1}}, 1\}\delta \geq 2^{-\frac{\alpha}{\alpha-1}} \min\{(\frac{\delta}{C_{\alpha,\kappa}})^{\frac{1}{\alpha-1}}, 1\}\delta$. We conclude that the assertion holds true for $c^{(2)}_{\alpha,\kappa} = 2^{-\frac{\alpha}{\alpha-1}} C_{\alpha,\kappa}^{-\frac{1}{\alpha-1}}$.

(ii) We have $\beta_{norm}(q) = \frac{\beta(q)}{q} = C_{\rho,\kappa}\frac{\rho^q}{q}$, where $C_{\rho,\kappa} = \frac{\kappa\rho}{1-\rho}$. In the following we assume w.l.o.g. that $C_{\rho,\kappa} > 8$.

- $q^*(x)$ Upper bound: Put $\psi(x) = \max\{\log(x^{-1}), 1\}$. Define $\tilde{q} = \lceil \frac{\psi(\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})}{\log(\rho^{-1})} \rceil$. Then we have

$$\beta_{norm}(\tilde{q}) \leq C_{\rho,\kappa}\frac{\rho^{\log\left(\left(\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})}\right)^{-1}\right)/\log(\rho^{-1})}}{\tilde{q}} \leq \frac{\frac{x}{\log(\rho^{-1})}}{\tilde{q}} \leq \frac{x}{\psi(\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})} \leq x,$$

whence

$$q^*(x) = \min\{q \in \mathbb{N} : \beta_{norm}(q) \leq x\} \leq \tilde{q} = \left\lceil \frac{\psi(\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})}{\log(\rho^{-1})} \right\rceil.$$

Especially,

$$q^*(x) \leq \frac{1}{\log(\rho^{-1})}\big(\psi(x) + \log(C_{\rho,\kappa}\log(\rho-1))\big) + 1 \leq \frac{2(1 + \log(C_{\rho,\kappa}\log(\rho^{-1})))}{\log(\rho^{-1})}\psi(x),$$

that is, the assertion holds true for $C^{(1)}_{\rho,\kappa} = \frac{2(1+\log(C_{\rho,\kappa}\log(\rho-1)))}{\log(\rho^{-1})}$.

- $q^*(x)$ Lower Bound: Case 1: Assume that $x < C_{\rho,\kappa}\log(\rho^{-1})\rho^4$. Define $\tilde{q} = \lceil \frac{1}{4}\frac{\log((\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})^{-1})}{\log(\rho^{-1})} \rceil \geq 1$. Then, $\tilde{q} \leq \frac{1}{2}\frac{\log((\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})^{-1})}{\log(\rho^{-1})}$ and thus

$$\beta_{norm}(\tilde{q}) \geq C_{\rho,k}\frac{\left(\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})}\right)^{1/2}}{\tilde{q}} \geq (C_{\rho,\kappa}\log(\rho^{-1}))^{1/2}\frac{x^{1/2}}{\log((\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})^{-1/2})} > x$$

since

$$(\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})^{-1/2} > \log((\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})})^{-1/2}).$$

We therefore have shown that for $x < C_{\rho,\kappa}\log(\rho^{-1})\rho^4$,

$$q^*(x) \geq \tilde{q} = \max\{1, \tilde{q}\}. \tag{2.8.113}$$

Case 2: If $x \geq C_{\rho,\kappa}\log(\rho^{-1})\rho^4$, then $\tilde{q} \leq 1$, that is,

$$q^*(x) \geq 1 = \max\{1, \tilde{q}\}.$$

We have shown that for all $x > 0$,

$$q^*(x) \geq \max\{1, \tilde{q}\}.$$

Since

$$\begin{aligned} \tilde{q} &\geq \frac{1}{4}\frac{\log\left(\left(\frac{x}{C_{\rho,\kappa}\log(\rho^{-1})}\right)^{-1}\right)}{\log(\rho^{-1})} \geq \frac{1}{4\log(\rho^{-1})}\big[\log(x^{-1}) + \log(C_{\rho,\kappa}\log(\rho^{-1}))\big] \\ &\geq \frac{1}{4\log(\rho^{-1})}\log(x^{-1}), \end{aligned}$$

the assertion follows for $c^{(1)}_{\rho,\kappa} = \frac{1}{4\log(\rho^{-1})}$.

- $r(\delta)$ Upper bound: Put $\tilde{r} = \frac{2(c^{(1)}_{\rho,\kappa})^{-1}\delta}{\log((2^{-1}c^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)}$. Then we have

$$\begin{aligned} q^*(\tilde{r})\tilde{r} &\geq c^{(1)}_{\rho,\kappa}\log(\tilde{r}^{-1}\vee e)\cdot\tilde{r} \\ &= \frac{2\delta}{\log((2^{-1}c^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)}\cdot\log([2^{-1}c^{(1)}_{\rho,\kappa}\delta^{-1}\log((2^{-1}c^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)]\vee e) \\ &\geq \frac{2\delta}{\log((2^{-1}c^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)}\cdot\log([2^{-1}c^{(1)}_{\rho,\kappa}\delta^{-1}]\vee e) = 2\delta > \delta. \end{aligned}$$

By definition of $r(\cdot)$ we obtain

$$r(\delta) \leq \tilde{r}.$$

For $a \in (0,1)$, the function $(0,\infty) \to (0,\infty), x \mapsto \frac{\log(x^{-1}\vee e)}{\log((ax^{-1})\vee e)}$ attains its maximum at $x = ae^{-1}$ with maximum value $1 + \log(a^{-1})$. Thus,

$$\tilde{r} \leq 2(c^{(1)}_{\rho,\kappa})^{-1}(1 + \log(2^{-1}(c^{(1)}_{\rho,\kappa})^{-1}))\cdot\frac{\delta}{\log(\delta^{-1}\vee e)}.$$

The assertion holds true for $C^{(2)}_{\rho,\kappa} = 2(c^{(1)}_{\rho,\kappa})^{-1}(1 + \log(2^{-1}(c^{(1)}_{\rho,\kappa})^{-1}))$.

- $r(\delta)$ Lower Bound: Put $\tilde{r} = \frac{2^{-1}(C^{(1)}_{\rho,\kappa})^{-1}\delta}{\log((2C^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)}$. Then

$$\begin{aligned} q^*(\tilde{r})\tilde{r} &\leq C^{(1)}_{\rho,\kappa}\log(\tilde{r}^{-1}\vee e)\cdot\tilde{r} \\ &= \frac{2^{-1}\delta}{\log((2C^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)}\cdot\log([2C^{(1)}_{\rho,\kappa}\delta^{-1}\log((2C^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)]\vee e) \\ &\leq \frac{2^{-1}\delta}{\log((C^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)}\cdot\big[\log((2C^{(1)}_{\rho,\kappa}\delta^{-1})\vee e) + \log\log((2C^{(1)}_{\rho,\kappa}\delta^{-1})\vee e)\big] \\ &\leq \delta, \end{aligned}$$

where the last step is due to $\log(x) + \log\log(x) \leq 2\log(x)$ for $x \geq e$. By definition of $r(\cdot)$ we obtain

$$r(\delta) \geq \tilde{r}.$$

For $a > 1$, the function $(0,\infty) \to (0,\infty)$, $x \mapsto \frac{\log(x^{-1}\vee e)}{\log((ax^{-1})\vee e)}$ attains its minimum at $x = e^{-1}$ with minimum value $\frac{1}{1+\log(a)}$. We therefore have

$$\tilde{r} \geq \frac{(C^{(1)}_{\rho,\kappa})^{-1}}{2(1+\log(2C^{(1)}_{\rho,\kappa}))}\frac{\delta}{\log(\delta^{-1}\vee e)}.$$

The assertion holds true for $c^{(2)}_{\rho,\kappa} = \frac{(C^{(1)}_{\rho,\kappa})^{-1}}{2(1+\log(2C^{(1)}_{\rho,\kappa}))}$.

□

Lemma 2.8.14 (Form of V_n)**.** *Depending on specific decay rates, the following statements hold true.*

(i) Polynomial decay $\Delta(j) = \kappa j^{-\alpha}$ (where $\alpha > 1$). There exist some constants $C^{(3)}_{\alpha,\kappa}, c^{(3)}_{\alpha,\kappa}$ only depending on $\kappa, \alpha, \mathbb{D}_n$ such that

$$c^{(3)}_{\alpha,\kappa}\|f\|_{2,n}\max\{\|f\|_{2,n}^{-\frac{1}{\alpha}}, 1\} \leq V_n(f) \leq C^{(3)}_{\alpha,\kappa}\|f\|_{2,n}\max\{\|f\|_{2,n}^{-\frac{1}{\alpha}}, 1\}.$$

(ii) Geometric decay $\Delta(j) = \kappa\rho^j$ (where $\rho \in (0,1)$). There exist some constants $c^{(3)}_{\rho,\kappa}, C^{(3)}_{\rho,\kappa}$ only depending on $\kappa, \rho, \mathbb{D}_n$ such that

$$c^{(3)}_{\rho,\kappa}\|f\|_{2,n}\max\{\log(\|f\|_{2,n}^{-1}), 1\} \leq V_n(f) \leq C^{(3)}_{\rho,\kappa}\|f\|_{2,n}\max\{\log(\|f\|_{2,n}^{-1}), 1\}.$$

Proof of Lemma 2.8.14. The assertions follow from Lemma 2.8.12(ii) by taking $\kappa_2 = \kappa\mathbb{D}_n$. The maximum in the lower bounds is obtained due to the additional summand $\|f\|_{2,n}$ in $V_n(f)$.

□

The following lemma formulates the entropy integral in terms of the well-known bracketing numbers with respect to the $\|\cdot\|_{2,n}$-norm in the case that $\sup_{n\in\mathbb{N}}\mathbb{D}_n < \infty$. We use the upper bounds of V_n given in Lemma 2.8.14.

Lemma 2.8.15. *Depending on specific decay rates, the following statements hold true.*

(i) Polynomial decay $\Delta(j) = \kappa j^{-\alpha}$ (where $\alpha > 1$). For any $\sigma \in (0, C^{(3)}_{\alpha,\kappa})$,

$$\int_0^\sigma \sqrt{\mathbb{H}(\varepsilon,\mathcal{F},V_n)}d\varepsilon \leq C^{(3)}_{\alpha,\kappa}\frac{\alpha-1}{\alpha}\int_0^{(\frac{\sigma}{C^{(3)}_{\alpha,\kappa}})^{\frac{\alpha}{\alpha-1}}} u^{-\frac{1}{\alpha}}\sqrt{\mathbb{H}(u,\mathcal{F},\|\cdot\|_{2,n})}du,$$

where $C^{(3)}_{\alpha,\kappa}$ is from lemma 2.8.14.

(ii) Exponential decay $\Delta(j) = \kappa\rho^j$ (where $\rho \in (0,1)$). For any $\sigma \in (0, e^{-1}C^{(3)}_{\rho,\kappa})$,

$$\int_0^\sigma \sqrt{\mathbb{H}(\varepsilon,\mathcal{F},V_n)}d\varepsilon \leq C^{(3)}_{\rho,\kappa}\int_0^{E^-(\frac{\sigma}{C^{(3)}_{\rho,\kappa}})} \log(u^{-1})\sqrt{\mathbb{H}(u,\mathcal{F},\|\cdot\|_{2,n})}du,$$

where $E^-(x) = \frac{x}{\log(x^{-1})}$ and $C^{(3)}_{\rho,\kappa}$ is from Lemma 2.8.14.

Proof of Lemma 2.8.15. (i) By Lemma 2.8.14, $V_n(f) \leq C^{(3)}_{\alpha,\kappa}\|f\|_{2,n}\max\{\|f\|_{2,n}^{-\frac{1}{\alpha}}, 1\}$. We abbreviate $c = C^{(3)}_{\alpha,\kappa}$ in the following.

Let $\varepsilon \in (0, c)$ and (l_j, u_j), $j = 1, ..., N$, brackets such that $\|u_j - l_j\|_{2,n} \leq (\frac{\varepsilon}{c})^{\frac{\alpha}{\alpha-1}}$. Then,

$$V_n(u_j - l_j) \leq c\max\{\|u_j - l_j\|_{2,n}, \|u_j - l_j\|_{2,n}^{\frac{\alpha-1}{\alpha}}\} \leq c\max\Big\{(\frac{\varepsilon}{c})^{\frac{\alpha}{\alpha-1}}, \frac{\varepsilon}{c}\Big\} \leq c\cdot\frac{\varepsilon}{c} = \varepsilon.$$

Therefore, the bracketing number fulfills the relation

$$\mathbb{N}(\varepsilon, \mathcal{F}, V_n) \leq \mathbb{N}\Big((\frac{\varepsilon}{c})^{\frac{\alpha}{\alpha-1}}, \mathcal{F}, \|\cdot\|_{2,n}\Big).$$

We conclude that for $\sigma \in (0, c)$,

$$\begin{aligned}\int_0^\sigma \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)}d\varepsilon &\leq \int_0^\sigma \sqrt{\mathbb{H}\Big((\frac{\varepsilon}{c})^{\frac{\alpha}{\alpha-1}}, \mathcal{F}, \|\cdot\|_{2,n}\Big)}d\varepsilon\\ &= c\frac{\alpha-1}{\alpha}\int_0^{(\frac{\sigma}{c})^{\frac{\alpha}{\alpha-1}}} u^{-\frac{1}{\alpha}}\sqrt{\mathbb{H}(u, \mathcal{F}, \|\cdot\|_{2,n})}du.\end{aligned}$$

In the last step we used the substitution $u = (\frac{\varepsilon}{c})^{\frac{\alpha}{\alpha-1}}$ which leads to $\frac{du}{d\varepsilon} = \frac{\alpha}{\alpha-1}\cdot\frac{1}{c}\cdot(\frac{\varepsilon}{c})^{\frac{1}{\alpha-1}} = \frac{\alpha}{\alpha-1}\cdot\frac{1}{c}\cdot u^{\frac{1}{\alpha}}$.

(ii) By Lemma 2.8.14, $V_n(f) \leq C^{(3)}_{\rho,\kappa}E(\|f\|_{2,n})$ with $E(x) = x\max\{\log(x^{-1}), 1\}$. We abbreviate $c = C^{(3)}_{\rho,\kappa}$ in the following.

We first collect some properties of E. Let $E^-(x) = \frac{x}{\log(x^{-1}\vee e)}$. In the case $x > e^{-1}$, we have $E(E^-(x)) = x$. In the case $x \leq e^{-1}$, we have

$$E(E^-(x)) = \frac{x}{\log(x^{-1})}\cdot\log\Big(\frac{x^{-1}}{\log(x^{-1})^{-1}}\Big) \leq \frac{x}{\log(x^{-1})}\log(x^{-1}) = x.$$

This shows that for all $x > 0$,

$$E(E^-(x)) \leq x. \tag{2.8.114}$$

Furthermore, for $x < e^{-1}$,

$$\log(E^-(x)^{-1}) = \log(x^{-1}\log(x^{-1})) \geq \log(x^{-1}). \tag{2.8.115}$$

Now let $\varepsilon \in (0, 1)$ and (l_j, u_j), $j = 1, ..., N$, brackets such that $\|u_j - l_j\|_{2,n} \leq E^-(\frac{\varepsilon}{c})$. Then by (2.8.114),

$$V_n(u_j - l_j) \leq cE(E^-(\frac{\varepsilon}{c})) \leq c\cdot\frac{\varepsilon}{c} = \varepsilon.$$

Therefore, we have the following relation between the bracketing numbers

$$\mathbb{N}(\varepsilon, \mathcal{F}, V_n) \leq \mathbb{N}\Big(E^-(\frac{\varepsilon}{c}), \mathcal{F}, \|\cdot\|_{2,n}\Big).$$

We conclude that for $\sigma \in (0, ce^{-1})$,

$$\begin{aligned}\int_0^\sigma \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)}d\varepsilon &\leq \int_0^\sigma \sqrt{\mathbb{H}\Big(E^-(\frac{\varepsilon}{c}), \mathcal{F}, \|\cdot\|_2\Big)}d\varepsilon \\ &\leq c\int_0^{E^-(\frac{\sigma}{c})} \log(u^{-1})\sqrt{\mathbb{H}(u, \mathcal{F}, \|\cdot\|_2)}du.\end{aligned}$$

In the last step we used the substitution $u = E^-(\frac{\varepsilon}{c})$ which leads to $\frac{du}{d\varepsilon} = \frac{1}{c} \cdot \frac{1+\log((\varepsilon/c)^{-1})}{\log((\varepsilon/c)^{-1})^2}$, and with (2.8.115) we obtain

$$d\varepsilon = c\frac{\log((\varepsilon/c)^{-1})^2}{1+\log((\varepsilon/c)^{-1})}du \leq c\log((\varepsilon/c)^{-1})du \leq c\log(E^-(\frac{\varepsilon}{c})^{-1})du = c\log(u^{-1})du.$$

□

Chapter 3

Empirical process theory for nonsmooth functions under functional dependence

In this chapter we would like to extend our current theory to nonsmooth functions, motivated by a proper discussion of the empirical distribution function. The previously obtained results come into play as we decompose our process into a smooth and a martingale type "contribution". The smooth contribution can be dealt with by the means of Chapter 2. It is the latter process that needs further investigation.

3.1 A functional central limit theorem under functional dependence and application to empirical distribution functions

Let us, as an example, consider the localized empirical distribution function of X_i,

$$\hat{G}_{n,h}(x,v) := \frac{1}{nh}\sum_{i=1}^{n} K\big(\frac{i/n-v}{h}\big)\mathbb{1}_{\{X_i\le x\}}, \tag{3.1.1}$$

where $K:\mathbb{R}\to\mathbb{R}$ is a kernel function and $h=h_n>0$ a bandwidth. Notice now that a noncontinuous indicator function is included. The goal of this chapter is to provide a general empirical process theory which allows us to show, for instance, a functional central limit theorem of $\hat{G}_{n,h}(x,v)$ for fixed $v\in[0,1]$ of the form

$$\big[\sqrt{nh}\big(\hat{G}_{n,h}(x,v)-G(x,v)\big)\big]_{x\in\mathbb{R}} \xrightarrow{d} \mathbb{G}(x)_{x\in\mathbb{R}} \tag{3.1.2}$$

where $(\mathbb{G}(x))_{x\in\mathbb{R}}$ is a centered Gaussian process and $G(x,v)=\mathbb{P}(\tilde{X}_0(v)\le x)$ denotes the distribution function of $\tilde{X}_0(v)$, the stationary approximation of X_0 according to Assumption 2.3.1.

As before, the additional localization via kernels changes the convergence rate of the empirical process. In order to discuss (3.1.1) with the general form (1.2.3), we assume that any $f \in \mathcal{F}$ has a representation as in (2.2.1),

$$f(z,u) = D_{f,n}(u) \cdot \bar{f}(z,u), \qquad z \in (\mathbb{R}^d)^{\mathbb{N}_0}, u \in [0,1], \tag{3.1.3}$$

where $\bar{f}$ is independent of n and $D_{f,n}(u)$ is independent of $z = (z_j)_{j\in\mathbb{N}_0}$. For the specific example given in (3.1.2), we consider

$$\mathcal{F} = \Big\{(z,u) \mapsto f_x(z,u) := \frac{1}{\sqrt{h}} K\big(\frac{u-v}{h}\big) \cdot \mathbb{1}_{\{z_0 \le x\}} : x \in \mathbb{R}\Big\},$$

and thus $D_{f_x,n}(u) = \frac{1}{\sqrt{h}} K(\frac{u-v}{h})$ and $\bar{f}_x(z,u) = \mathbb{1}_{\{z_0 \le x\}}$.

We now introduce the necessary assumptions for our empirical process theory in context of the functional dependence measure. Based on decomposition (3.1.3), we define the following two function classes with regard to $\bar{f}$, which mimic the one-step-ahead mean and variance forecast,

$$\begin{aligned} \bar{\mathcal{F}}^{(1)} &:= \{(z,u) \mapsto \mathbb{E}[\bar{f}(Z_i,u)|Z_{i-1} = z] : f \in \mathcal{F},\, i \in \mathbb{Z}\}, \\ \bar{\mathcal{F}}^{(2)} &:= \{(z,u) \mapsto \mathbb{E}[\bar{f}(Z_i,u)^2|Z_{i-1} = z]^{1/2} : f \in \mathcal{F},\, i \in \mathbb{Z}\}, \end{aligned}$$

where $Z_i := (X_i, X_{i-1}, X_{i-2}, ...)$, as in Subsection 1.2.

There are two key assumptions on $\bar{f}$ to show our main result, Theorem 3.1.2. The first is a compatibility condition which connects smoothness properties of $\bar{\mathcal{F}}^{(\kappa)}$, $\kappa \in \{1,2\}$, with corresponding moment assumptions on the process X_i, $i = 1, ..., n$.

Assumption 3.1.1 (Compatibility condition for function classes)**.** *The classes $\bar{\mathcal{F}}^{(\kappa)}$, $\kappa \in \{1,2\}$, are (L,s,R,C)-classes. There exists $p \in (1,\infty]$, $C_X > 0$ such that*

$$\text{(i') } \sup_{i,u} \|R(Z_{i-1},u)\|_{2p} \le C_R, \qquad \text{(ii) } \sup_{i,j} \|X_{ij}\|_{\frac{2sp}{p-1}} \le C_X.$$

Let $\mathbb{D}_n \ge 0$ and $\Delta(k) \ge 0$ such that

$$2dC_R \sum_{j=0}^{k-1} L_j \big(\delta^X_{\frac{2sp}{p-1}}(k-j-1)\big)^s \le \Delta(k), \qquad \sup_{f\in\mathcal{F}} \Big(\frac{1}{n}\sum_{i=1}^n \Big|D_{f,n}\big(\frac{i}{n}\big)\Big|^2\Big)^{1/2} \le \mathbb{D}_n.$$

Notice that the index in the variable Z_{i-1} of condition (i') is shifted in comparison to the earlier formulation of Assumption 2.2.3, condition (i).

First, let us summarize the dependence structure by the quantity

$$\beta(q) = \sum_{j=q}^{\infty} \Delta(j) \tag{3.1.4}$$

where $\Delta(\cdot)$ plays the same role in Assumption 3.1.1 as in Assumption 2.2.3. Recall $\mathbb{H}(\varepsilon, \mathcal{F}, \|\cdot\|)$ as the bracketing entropy and V_n from (2.2.4). Postponing some technicalities, we state our main result.

Theorem 3.1.2. *Suppose that $\mathcal{F}$ satisfies Assumption 3.1.1, 3.2.1, 3.3.3, 2.3.1, 3.2.2, 2.3.3. For*

$$\psi(\varepsilon) = \sqrt{\log(\varepsilon^{-1} \vee 1)} \log\log(\varepsilon^{-1} \vee e) \tag{3.1.5}$$

suppose that

$$\sup_{n \in \mathbb{N}} \int_0^1 \psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)} d\varepsilon < \infty.$$

Then in $\ell^\infty(\mathcal{F})$,

$$\big[\mathbb{G}_n(f)\big]_{f\in\mathcal{F}} \xrightarrow{d} \big[\mathbb{G}(f)\big]_{f\in\mathcal{F}}$$

where $(\mathbb{G}(f))_{f\in\mathcal{F}}$ is a centered Gaussian process with covariances

$$\mathrm{Cov}(\mathbb{G}(f), \mathbb{G}(g)) = \lim_{n\to\infty} \mathrm{Cov}(\mathbb{G}_n(f), \mathbb{G}_n(g)) = \Sigma^{(\mathbb{K})}$$

and $\Sigma^{(\mathbb{K})}$ is from Assumption 2.3.3.

As in the preceding chapter, the statement is a result of the convergence of the finite-dimensional distributions in Section 3.2, Theorem 3.2.3, and asymptotic tightness in Section 3.3.2, Corollary 3.3.6. They will be shown later on.

Suppose that $\mathbb{D}_n \in (0, \infty)$ is independent of $n \in \mathbb{N}$. Based on decay rates of $\Delta(k)$, we derive simpler forms of V_n which are shown below in Table 3.1. The rates are similar to Table 2.1 except for the additional factor $\psi(\varepsilon)$ in the entropy integral.

	$\Delta(j)$	
	$cj^{-\alpha}$, $\alpha > 1, c > 0$	$c\rho^j$, $\rho \in (0,1)$, $c > 0$
$V_n(f)$	$\\|f\\|_{2,n} \max\{\\|f\\|_{2,n}^{-\frac{1}{\alpha}}, 1\}$	$\\|f\\|_{2,n} \max\{\log(\\|f\\|_{2,n}^{-1}), 1\}$
$\int_0^\sigma \sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, V_n)} d\varepsilon$	$\int_0^{\tilde\sigma} \varepsilon^{-\frac{1}{\alpha}} \psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, \\|\cdot\\|_{2,n})} d\varepsilon$	$\int_0^{\tilde\sigma} \log(\varepsilon^{-1})\psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon, \mathcal{F}, \\|\cdot\\|_{2,n})} d\varepsilon$

Table 3.1: Equivalent expressions of V_n and the corresponding entropy integral under the condition that $\mathbb{D}_n \in (0, \infty)$ is independent of n. We omitted the lower and upper bound constants which are only depending on c, ρ, α and $\mathbb{D}_n$. Furthermore, $\tilde\sigma = \tilde\sigma(\sigma)$ fulfills $\tilde\sigma \to 0$ for $\sigma \to 0$.

The theorem significantly simplifies if X_i is stationary. In that case, $\bar f(z, u) = \bar f(z_0)$ depends only on one observation and no weighting is present, i.e. $D_{f,n}(u) = 1$. Assumption 3.2.1, 3.2.2, 2.3.1 and 2.3.3 then are fulfilled automatically. These assumptions are needed only to provide a (pointwise) central limit theorem for locally stationary processes. They basically enforce several smoothness properties of $\bar f$.

In more detail, let

$$\tilde{\mathbb{G}}_n(h) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \big\{h(X_i) - \mathbb{E}h(X_i)\big\},$$

where $X_i = J(\mathcal{A}_i)$, $i = 1, ..., n$, is a stationary Bernoulli shift process and $h \in \mathcal{H} \subset \{h : \mathbb{R}^d \to \mathbb{R} \text{ measurable}\}$, a function class with envelope function $\bar{h}$, i.e. for $h \in \mathcal{H}$ we have $|h(\cdot)| \leq \bar{h}(\cdot)$, such that

$$h^{(1)}(z_0) = \mathbb{E}[h(X_1) \mid X_0 = z_0], \qquad h^{(2)}(z_0) = \mathbb{E}[h(X_1)^2 \mid X_0 = z_0]^{1/2}$$

are Hölder continuous with exponent s and constant $L_{\mathcal{H}}$, that is, for all $z, z' \in \mathbb{R}$,

$$|h^{(1)}(z) - h^{(1)}(z')| \leq L_{\mathcal{H}}|z - z'|^s, \qquad |h^{(2)}(z) - h^{(2)}(z')| \leq L_{\mathcal{H}}|z - z'|^s.$$

Assumption 3.1.1 automatically holds true with $R(\cdot) = \frac{1}{2}$ and thus $C_R = \frac{1}{2}$, $L = L_{\mathcal{H}}$ as well as $C = \max\{h^{(1)}(0), h^{(2)}(0)\}$. Recall $\beta(\cdot)$ from (3.1.4). Then we have the following corollary of Theorem 3.1.2 in the stationary case.

Corollary 3.1.3. *Suppose that $\|X_1\|_{2s} < \infty$ and put $\mathbb{D}_n := 1$. Let $\Delta(k)$ fulfill $\Delta(k) \geq dL_{\mathcal{H}}\delta_{2s}^X(k-1)^s$ and there exists $C_\beta > 0$ such that for all $q_1, q_2 \in \mathbb{N}$,*

$$\beta(q_1 q_2) \leq C_\beta \beta(q_1)\beta(q_2). \tag{3.1.6}$$

Furthermore, $\|\bar{h}(X_1)\|_{2\bar{p}} < \infty$ for some $\bar{p} > 1$. Assume that

$$\sup_{n \in \mathbb{N}} \int_0^1 \psi(\varepsilon)\sqrt{\mathbb{H}(\varepsilon, \mathcal{H}, V_n)} d\varepsilon < \infty$$

where $\psi(\varepsilon)$ is from (3.1.5). Then in $\ell^\infty(\mathcal{H})$,

$$\left[\tilde{\mathbb{G}}_n(h)\right]_{h \in \mathcal{H}} \xrightarrow{d} \left[\tilde{\mathbb{G}}(h)\right]_{h \in \mathcal{H}}$$

where $(\tilde{\mathbb{G}}(h))_{h \in \mathcal{H}}$ is a centered Gaussian process with covariances

$$\operatorname{Cov}(\tilde{\mathbb{G}}(h_1), \tilde{\mathbb{G}}(h_2)) = \sum_{k \in \mathbb{Z}} \operatorname{Cov}(h_1(X_0), h_2(X_k)).$$

3.1.1 Application to empirical distribution functions of stationary processes

As an example, consider the family of indicators

$$\mathcal{H} = \{h_x(z_0) := \mathbb{1}_{\{z_0 \leq x\}} : x \in \mathbb{R}\},$$

which is the function class that corresponds to the empirical distribution function

$$\left[\hat{G}_n(x)\right]_{x \in \mathbb{R}} = \left[\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \leq x\}}\right]_{x \in \mathbb{R}} = \left[\tilde{\mathbb{G}}_n(h)\right]_{h \in \mathcal{H}}.$$

Suppose that X_i, $i = 1, ..., n$, is stationary. Define the conditional distribution function

$$G_z(x) = \mathbb{P}(X_1 \leq x \mid X_0 = z).$$

Then we have the following corollary.

Corollary 3.1.4. *Suppose that X_i is stationary and $z \mapsto G_z(x)$ is Lipschitz continuous with Lipschitz constant L_G for all $x \in \mathbb{R}$. Suppose that for some $s \in (0, \frac{1}{2}]$, $\|X_1\|_{2s} < \infty$ and $\delta_{2s}^X(k) \le ck^{-\alpha}$ with $\alpha > \frac{1}{s}$, $c > 0$. Then,*

$$\big[\hat{G}_n(x)\big]_{x\in\mathbb{R}} \xrightarrow{d} \big[\tilde{\mathbb{G}}(x)\big]_{x\in\mathbb{R}}$$

where $\tilde{\mathbb{G}}(x)$ is a Gaussian process with

$$\mathrm{Cov}(\tilde{\mathbb{G}}(x), \tilde{\mathbb{G}}(y)) = \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbb{1}_{\{X_0 \le x\}}, \mathbb{1}_{\{X_k \le y\}}).$$

As it is only a short discussion we include its proof here.

Proof of Corollary 3.1.4. Due to $\min\{1, w\} \le w^a$ for $a \in [0, 1]$, $w \ge 0$, we have for any $s \in (0, \frac{1}{2}]$,

$$|G_z(x) - G_{z'}(x)| \le \min\{1, L_G|z - z'|\} \le L_G^s|z - z'|^s$$

and

$$|G_z(x) - G_{z'}(x)|^{1/2} \le \min\{1, (L_G|z - z'|)^{1/2}\} \le L_G^s|z - z'|^s.$$

Choose $\Delta(k) = cL_G(k-1)^{-\alpha s}$, which is easily seen to satisfy (3.1.6) (in particular, $\beta(q) < \infty$ for $q \in \mathbb{N}$) for some $C_\beta = C_\beta(\alpha, s, c, L_G)$ chosen large enough.

Note that $\mathbb{H}(\varepsilon, \mathcal{H}, \|\cdot\|_{2,n}) = O(\log(\varepsilon^{-1}))$ for a given $\varepsilon > 0$ by [van der Vaart, 1998, Example 19.6], because in the stationary situation of the corollary, $\|h\|_{2,n} = \mathbb{E}[h(X_1)^2]^{1/2}$. Since $\alpha s > 1$, Table 3.1 implies

$$\int_0^1 \psi(\varepsilon)\sqrt{\mathbb{H}(\gamma, \mathcal{H}, V_n)}d\varepsilon = O\Big(\int_0^1 \psi(\varepsilon)\varepsilon^{-\frac{1}{\alpha s}}\sqrt{\log(\varepsilon^{-1})}d\varepsilon\Big) < \infty.$$

Corollary 3.1.3 now delivers the assertion. □

3.1.2 Comparison with other functional convergence results for the empirical distribution function of stationary processes

In the literature, several functional convergence results for the empirical distribution function were already provided. Here we list some approaches which are closely related to the functional dependence measure and compare the results to Corollary 3.1.4.

In Borovkova et al. [2001], stationary processes of the form $X_i = J(\mathcal{G}_i)$ are considered where $\mathcal{G}_i = (\varepsilon_i, \varepsilon_{i-1}, ...)$ and J is measurable. Therein, the function J itself is assumed to fulfill a (geometrically decaying) Lipschitz condition, i.e. for any sequences $(a_i), (a_i')$ with $a_i = a_i'$, $i \le k$,

$$\big|J((a_i)) - J((a_i'))\big| \le C\alpha^k \tag{3.1.7}$$

for some constants $C, \alpha > 0$. Based on this, 1-approximation coefficients a_k are defined as upper bounds on

$$\mathbb{E}\big\|X_0 - \mathbb{E}[X_0 \mid \sigma(\varepsilon_0, ..., \varepsilon_k)]\big\|_1 \le a_k.$$

There is a strong connection between $\delta_1^X(k)$ and a_k, since it is possible to choose $a_k \le \sum_{j=k+1}^{\infty} \delta_1^X(j)$. The work of [Borovkova et al., 2001, Theorem 5] shows that, under summability conditions on a_k, the β-mixing coefficients *and* monotonicity assumptions on $\mathcal{F} = \{f_t : t \in [0,1]\}$, a uniform central limit theorem for $(\mathbb{G}_n(f_t))_{t\in[0,1]}$ holds true. Compared to our setting, (3.1.7) would lead to a geometrically decaying functional dependence measure δ^X. Thus, the result in our Corollary 3.1.4 is much less restrictive regarding the dependence of the underlying process.

In [Dedecker, 2010, Theorem 2.1], a uniform central limit theorem for the empirical distribution function is shown under $\beta_2(k) = O(k^{-1-\gamma})$, $\gamma > 0$, by using specifically designed dependence coefficients $\beta_2(k)$, $k \in \mathbb{N}_0$, based on the idea of absolute regularity. We now compare this result to Corollary 3.1.4. In [Dedecker and Prieur, 2007, Section 6.1] it was shown that if $X_i = J(\mathcal{G}_i)$ is stationary and the distribution function of X_1 is Lipschitz continuous, then for any $\nu \in [0,1]$,

$$\beta_2(k) \le C \cdot \Big(\sum_{j=k+1}^{\infty} \delta_\nu^X(j)^{\nu'} \Big)^{\frac{\nu}{\nu'(\nu+1)}}, \qquad \nu' = \min\{\nu, 1\},$$

where $C > 0$ is a constant independent of k. The condition $\beta_2(k) = O(k^{-1-\gamma})$ now naturally provides a decay condition on $\delta_\nu^X(k)$. For $\nu = 2s$, which corresponds to the moments of the process we have given in Corollary 3.1.4, we see after a short calculation that $\beta_2(k) = O(k^{-1-\gamma})$ implies

$$\alpha \ge \frac{1}{s} + \frac{\gamma}{2s} + \gamma + 1.$$

In other words, if the results from Dedecker [2010], Dedecker and Prieur [2007] are transferred to the functional dependence measure setting, they need a more restrictive decay condition.

Meanwhile, Berkes et al. [2009] investigates strong approximations of the multivariate empirical distribution function process (that is, contrary to our approach, the results are limited to empirical distribution functions). They assume that the stationary process $X_i = J(\mathcal{G}_i)$ allows for approximations $(X_i^{(m)})$ such that for all m, i,

$$\mathbb{P}(|X_i - X_i^{(m)}| \ge m^{-A}) \le m^{-A} \tag{3.1.8}$$

for some $A > 4$, and that for any disjoint intervals $I_1, ..., I_r$ of integers and any positive integers $m_1, ..., m_r$, the vectors $\{X_i^{(m_1)} : i \in I_1\}, ..., \{X_i^{(m_r)} : j \in I_r\}$ are independent of each other provided the separation between I_k and I_l is greater than $m_k + m_l$. Under these assumptions, [Berkes et al., 2009, Theorem 1, Corollary 1] shows that the empirical distribution function with respect to X_i weakly converges to some Gaussian process.

When having knowledge about the functional dependence measure, $X_i^{(m)}$ could be chosen as $X_i^{(m)} = \mathbb{E}[X_i \mid \varepsilon_i, ..., \varepsilon_{i-m}]$. Then by Markov's inequality,

$$\mathbb{P}\big(|X_i - X_i^{(m)}| \ge m^{-A}\big) \le \frac{\|X_i - X_i^{(m)}\|_{2s}^{2s}}{m^{-2sA}} \le \big(m^A \cdot \sum_{j=m+1}^{\infty} \delta_{2s}^X(j)\big)^{2s},$$

so that (3.1.8) leads to a decay condition on $\delta_\nu^X(j)$. After a short calculation, we see that (3.1.8) is fulfilled if

$$\alpha \geq \Big(\frac{1}{2s}+1\Big)A+1,$$

again a more restrictive decay condition than given in Corollary 3.1.4.

The work of Durieu and Tusche [2014] discusses the functional convergence of the multivariate empirical distribution function under a general growth condition imposed on the moments of $\sum_{i=1}^n \{h(X_i) - \mathbb{E}h(X_i)\}$, where $h \in \mathcal{H}_\gamma$ is a Hölder continuous function with exponent $\gamma \in (0,1]$ approximating the indicator function. They relate their result to the functional dependence measure in a rather involved discussion, as well.

3.1.3 Application to empirical distribution functions of locally stationary processes

In this section, we apply our theory to the localized empirical distribution function $\hat{G}_{n,h}(x,v)$ from (3.1.1) on a locally stationary process, as motivated in the beginning of Section 3.1.

Suppose that X_i is locally stationary in the sense of Assumption 2.3.1 and recall $G(x,v) = \mathbb{P}(\tilde{X}_1(v) \leq x)$. Define the conditional distribution function of the stationary approximation of X_i,

$$G_z(x,v) = \mathbb{P}(\tilde{X}_1(v) \leq x \mid \tilde{X}_0(v) = z).$$

Next, we impose a regularity assumption on the distribution function $G_i(x) := \mathbb{P}(X_i \leq x)$ of the locally stationary process itself.

The following generalization of Corollary 3.1.4 holds true.

Corollary 3.1.5. *Let $v \in (0,1)$. Suppose that there exists some $L_G > 0$ such that*

- *$z \mapsto G_z(x,v)$ is Lipschitz continuous with constant L_G for all $x \in \mathbb{R}$,*
- *$x \mapsto G(x,v)$ is Lipschitz continuous with constant L_G,*
- *$x \mapsto G_i(x)$ is Lipschitz continuous with constant L_G and $\lim_{x\to-\infty}\sup_{i,n} G_i(x) = 0$, $\lim_{x\to+\infty}\inf_{i,n} G_i(x) = 1$.*

Assume that $K:\mathbb{R}\to\mathbb{R}$ is a Lipschitz continuous kernel function with $\int K(x)dx = 1$ and support $\subset [-\frac{1}{2},\frac{1}{2}]$.

Furthermore, for some $s \in (0,\frac{1}{2}]$ let $\sup_{i,n}\|X_i\|_{2s} < \infty$ and $\delta_{2s}^X(k) \leq ck^{-\alpha}$ with $\alpha > \frac{1}{s}$, $c > 0$.

Then for $hn \to \infty$, $h \to 0$,

$$\big[\hat{G}_{n,h}(x,v)\big]_{x\in\mathbb{R}} \xrightarrow{d} \big[\tilde{\mathbb{G}}(x,v)\big]_{x\in\mathbb{R}},$$

where $\tilde{\mathbb{G}}(x,v)$ is a Gaussian process with

$$\mathrm{Cov}(\tilde{\mathbb{G}}(x,v),\tilde{\mathbb{G}}(y,v)) = \int K(u)^2du \cdot \sum_{k\in\mathbb{Z}} \mathrm{Cov}(\mathbb{1}_{\{\tilde{X}_0(v)\leq x\}}, \mathbb{1}_{\{\tilde{X}_k(v)\leq y\}}).$$

The recently published work Mayer et al. [2020] considers functional convergence of the empirical distribution function of piece-wise locally stationary processes and retrieve similar results. We provide slightly less (weaker) assumptions and include polynomial decay of the dependence coefficients.

3.2 A general central limit theorem for locally stationary processes

In this section, we provide a multivariate central limit theorem for $\mathbb{G}_n(f)$. To guarantee a regular behavior of the asymptotic variance, we need the following assumptions.

Assumption 3.2.1. *Let $\bar{F}$ be an envelope function of $\{\bar{f} : f \in \mathcal{F}\}$, that is, $|\bar{f}(\cdot)| \le \bar{F}(\cdot)$ for all $f \in \mathcal{F}$. There exists $\bar{p} \in (1, \infty]$ such that $\sup_{i,u} \|\bar{F}(Z_i, u)\|_{2\bar{p}} < \infty$ and $\sup_{v,u} \|\bar{F}(\tilde{Z}_0(v), u)\|_{2\bar{p}} < \infty$. Furthermore, either*

- *X_i is stationary, or*
- *for all $c > 0$ and $f \in \mathcal{F}$,*

$$\sup_{u,v\in[0,1]} \frac{1}{c^s}\mathbb{E}\Big[\sup_{|a|_{L_{\mathcal{F}},s}\le c} \big|\bar{f}(\tilde{Z}_0(v), u) - \bar{f}(\tilde{Z}_0(v) + a, u)\big|^2\Big] < \infty. \tag{3.2.1}$$

Additionally, (3.2.1) also holds true for $\bar{F}$.

Assumption 3.2.2. *There exists some $\varsigma \in (0, 1]$ such that for every $f \in \mathcal{F}$,*

$$|\bar{f}(z, u_1) - \bar{f}(z, u_2)| \le |u_1 - u_2|^{\varsigma} \cdot \big(\bar{R}(z, u_1) + \bar{R}(z, u_2)\big),$$

and $\sup_{u,v} \|\bar{R}(\tilde{Z}_0(v), u)\|_2 < \infty$.

We comment on the assumptions after the following theorem.

Theorem 3.2.3. *Suppose that $\mathcal{F}$ satisfies Assumption 3.1.1, 3.2.1, 2.3.1, 3.2.2 and 2.3.3. Let $m \in \mathbb{N}$, $f_1, ..., f_m \in \mathcal{F}$ and $\Sigma^{\mathbb{K}} = \Sigma^{(\mathbb{K})}_{f_k,f_l})_{k,l=1,...,m}$. Then,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\begin{pmatrix} f_1(Z_i, \frac{i}{n}) \\ \vdots \\ f_m(Z_i, \frac{i}{n})\end{pmatrix} - \mathbb{E}\begin{pmatrix} f_1(Z_i, \frac{i}{n}) \\ \vdots \\ f_m(Z_i, \frac{i}{n})\end{pmatrix}\right\} \xrightarrow{d} N(0, \Sigma^{\mathbb{K}}),$$

where $\Sigma^{(\mathbb{K})}$ is from Assumption 2.3.3.

While Assumption 3.2.1 asks for smoothness of $\bar{f}$ in the $\mathbb{L}^2$-sense if X_i is nonstationary, Assumption 3.2.2 requires the function class $\mathcal{F}$ to behave smoothly in the second argument. The last Assumption 2.3.3 mainly controls the behavior of the part $D_{f,n}(u)$ of $f \in \mathcal{F}$ which does not depend on the observations.

In contrast to the continuous case where all conditions imposed on $f \in \mathcal{F}$ also transfer to $\sup_{f\in\mathcal{F}} f$ due to the purely analytic nature of Assumption 2.2.3 and 3.2.2, we here additionally require some envelope function $\bar{F}$ to fulfill Assumption 3.2.2 and (3.2.1) because the supremum over $f \in \mathcal{F}$ does not interchange with the expectation in (3.2.1).

Note that Assumptions 3.2.2, 2.3.1 and 2.3.3 are needed to allow for *very different* function classes $\mathcal{F}$. In many special cases, however, some of these assumptions are automatically fulfilled. We commented on similar cases in Remark 2.3.5. For example, if $\bar{f}(z,u) = \bar{f}(z)$ does not depend on u, Assumption 3.2.2 is fulfilled.

Theorem 3.2.3 is a version of Theorem 2.3.4 but for nonsmooth function classes. The proofs are very similar to each other. The only difference appears as Lemma 3.5.6, for which we supply the proof in Section 3.5.3. It can be considered as an analogue to Lemma 2.8.3 under the different Assumption 3.2.1, 3.2.2, 2.3.1 and 2.3.3.

3.3 Maximal inequalities and asymptotic tightness under functional dependence

We now provide an approach for empirical process theory if the class $\mathcal{F}$ consists of nonsmooth functions. Our approach is based on the decomposition

$$\mathbb{G}_n(f) = \mathbb{G}_n^{(1)}(f) + \mathbb{G}_n^{(2)}(f)$$

into a martingale

$$\mathbb{G}_n^{(1)}(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{ f\big(Z_i, \frac{i}{n}\big) - \mathbb{E}\big[f\big(Z_i, \frac{i}{n}\big) \mid Z_{i-1}\big]\Big\}$$

and a process

$$\mathbb{G}_n^{(2)}(f) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \Big\{ \mathbb{E}\big[f\big(Z_i, \frac{i}{n}\big) \mid Z_{i-1}\big] - \mathbb{E}f\big(Z_i, \frac{i}{n}\big)\Big\}$$

which is smooth with respect to the arguments Z_i if Assumption 3.1.1 is fulfilled. The second part $\mathbb{G}_n^{(2)}$ can then be controlled in a similar way as done in Section 2.4. Therefore, it is only discussed in Section 3.5, where all the proofs of this chapter are deferred to. The term $\mathbb{G}_n^{(1)}$ is dealt with by using a Bernstein-type inequality for martingales. Observe that the conditional variance of $\mathbb{G}_n^{(1)}(f)$ on Z_{i-1} is bounded from above by

$$R_n^2(f) := \frac{1}{n}\sum_{i=1}^n \mathbb{E}\Big[f\big(Z_i, \frac{i}{n}\big)^2 \mid Z_{i-1}\Big].$$

The first step is now to bound $R_n^2(f)$ uniformly over $f \in \mathcal{F}$.

3.3.1 Maximal inequalities

We recall that based on $\beta(\cdot)$ from (3.1.4), we define

$$q^*(x) := \min\{q \in \mathbb{N} : \beta(q) \leq q \cdot x\}$$

and for $\delta > 0$,

$$r(\delta) := \max\{r > 0 : q^*(r)r \leq \delta\}.$$

Again, $D_n^\infty(u) := \sup_{f\in\mathcal{F}} |D_{f,n}(u)|$. For $\nu \geq 2$, we choose $\mathbb{D}_{\nu,n}^\infty$ such that the inequality $\big(\frac{1}{n}\sum_{i=1}^n D_n^\infty(\frac{i}{n})^\nu\big)^{1/\nu} \leq \mathbb{D}_{\nu,n}^\infty$ is fulfilled. Put $\mathbb{D}_n^\infty = \mathbb{D}_{2,n}^\infty$. As before in (1.2.5), $H = H(|\mathcal{F}|) = 1 \vee \log|\mathcal{F}|$. The values for $q^*(\cdot)$ and $r(\cdot)$ under polynomial and exponential decaying $\Delta(\cdot)$ were given in Table 2.2.

We have the following theorem. Note that the below appearing constant C_Δ is quite complex, which is why we reduce it here to its formal existence. We refer to Lemma 3.5.1 for its exact form.

Theorem 3.3.1 (Controlling the variance)**.** *Let $\mathcal{F}$ satisfy $|\mathcal{F}| < \infty$ and Assumption 3.1.1. Then there exists some universal constant $c > 0$ such that the following holds true. If $\sup_{f\in\mathcal{F}}\|f\|_\infty \leq M$ and $\sup_{f\in\mathcal{F}} V_n(f) \leq \sigma$, then*

$$\mathbb{E}\max_{f\in\mathcal{F}}\Big|R_n^2(f) - \mathbb{E}R_n^2(f)\Big| \leq c \cdot \min_{q\in\{1,\dots,n\}}\Big[\mathbb{D}_n r\big(\frac{\sigma}{\mathbb{D}_n}\big)\sigma + C_\Delta(\mathbb{D}_n^\infty)^2\beta(q) + \frac{qM^2H}{n}\Big]. \quad (3.3.1)$$

Furthermore,

$$\mathbb{E}\max_{f\in\mathcal{F}}\Big|R_n^2(f) - \mathbb{E}R_n^2(f)\Big| \leq 2c \cdot \Big[\mathbb{D}_n r\big(\frac{\sigma}{\mathbb{D}_n}\big)\sigma + q^*\big(\frac{M^2H}{n(\mathbb{D}_n^\infty)^2C_\Delta}\big)\frac{M^2H}{n}\Big]. \quad (3.3.2)$$

Theorem 3.3.1 in conjunction with Theorem 2.4.1 can be used to provide uniform convergence rates for $\mathbb{G}_n(f)$.

Corollary 3.3.2 (Uniform convergence rates)**.** *Suppose that $\mathcal{F}$ satisfies $|\mathcal{F}| < \infty$, Assumption 3.1.1 for some $\nu \geq 2$ and Assumption 3.3.3. Let $\bar{F} := \sup_{f\in\mathcal{F}}\bar{f}$ and assume that for some $\nu_2 \in [2,\infty]$,*

$$C_{\bar{F},n} := \sup_{i,u}\|\bar{F}(Z_i,u)\|_{\nu_2} < \infty.$$

If

$$\sup_{n\in\mathbb{N}}\sup_{f\in\mathcal{F}} V_n(f) < \infty, \qquad \sup_{n\in\mathbb{N}}\frac{\mathbb{D}_{\nu_2,n}^\infty}{\mathbb{D}_n^\infty} < \infty, \qquad \sup_{n\in\mathbb{N}}\frac{C_{\bar{F},n}^2 H}{n^{1-\frac{2}{\nu_2}}r\big(\frac{\sigma}{\mathbb{D}_n^\infty}\big)^2} < \infty, \quad (3.3.3)$$

then

$$\max_{f\in\mathcal{F}}|\mathbb{G}_n(f)| = O_p(\sqrt{H}).$$

3.3.2 Asymptotic tightness

In this section, we extend the maximal inequality from Theorem 3.3.1 to arbitrary (infinite) classes $\mathcal{F}$. We need an additional submultiplicativity assumption on $\beta(\cdot)$ from (3.1.4).

Assumption 3.3.3. *There exists a constant $C_\beta > 0$ such that for each $q_1, q_2 \in \mathbb{N}$,*

$$\beta(q_1 q_2) \le C_\beta \cdot \beta(q_1)\beta(q_2).$$

It is easily seen that Assumption 3.3.3 is fulfilled if $\Delta(k)$ follows a polynomial ($\Delta(k) = ck^{-\alpha}$ for $c > 0, \alpha > 1$) or exponential decay ($\Delta(k) = c\rho^k$ for $c > 0$, $\rho \in (0,1)$), cf. Lemma 2.8.12. Assumption 3.3.3 is generally not fulfilled if $\Delta(k)$ contains a factor of the form $\frac{1}{\log(k)}$.

Recall $H(k) = 1 \vee \log(k)$. For $n \in \mathbb{N}$, $\delta > 0$, we defined (cf. (2.4.6))

$$m(n,\delta,k) := r\big(\frac{\delta}{\mathbb{D}_n}\big) \cdot \frac{\mathbb{D}_n^\infty n^{1/2}}{H(k)^{1/2}}.$$

Theorem 3.3.4. *Let $\mathcal{F}$ satisfy Assumption 3.1.1 and 3.3.3, F be some envelope function of $\mathcal{F}$. Furthermore, let $\sigma > 0$ and suppose that $\sup_{f\in\mathcal{F}} V_n(f) \le \sigma$. Let ψ be defined as in (3.1.5). Then there exists a universal constant $c > 0$ such that for each $\eta > 0$,*

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)| > \eta\Big)\\
\le\ & \frac{1}{\eta}\Big[c\Big(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n} + \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\Big) \cdot \int_0^\sigma \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}\big(\varepsilon, \mathcal{F}, V\big)}\,\mathrm{d}\varepsilon\\
&\qquad\qquad\qquad + \sqrt{n}\big\|F\mathbb{1}_{\{F > \frac{1}{4}m(n,\sigma,\mathbb{N}(\frac{\sigma}{2},\mathcal{F},V_n))\}}\big\|_1\Big]\\
&\qquad + c\Big(1 + q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)\Big(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}\Big)^2\Big)\int_0^\sigma \frac{1}{\varepsilon\psi(\varepsilon)^2} d\varepsilon. \qquad (3.3.4)
\end{aligned}$$

Remark 3.3.5. Let $m > 0$. The chaining procedure found in Nishiyama et al. [2000] for martingales uses the fact that for functions f, g with $|f| \le g$ and $g(\cdot) > m$,

$$|\mathbb{G}_n^{(1)}(f)| \le |\mathbb{G}_n^{(1)}(g)| + 2\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[g\big(Z_i, \frac{i}{n}\big) \mid Z_{i-1}\big] \le |\mathbb{G}_n^{(1)}(g)| + 2\sqrt{n}\frac{R_n^2(g)}{m}.$$

Afterwards, bounds for the conditional variance $R_n^2(g)$ are applied. In our case, these bounds are not sharp enough. We therefore employ the inequality

$$|\mathbb{G}_n^{(1)}(f)| \le |\mathbb{G}_n^{(1)}(g)| + 2|\mathbb{G}_n^{(2)}(g)| + 2\sqrt{n}\frac{\|g\|_{2,n}^2}{m}$$

and are forced to use the "smooth" chaining technique applied to $\mathbb{G}_n^{(2)}(g)$, as in Theorem 2.4.4, and discuss $R_n^2(g)$ with Theorem 3.3.1.

We now obtain asymptotic equicontinuity of the process $\mathbb{G}_n(f)$ by using Theorem 3.3.4 for $\mathbb{G}_n^{(1)}$ and Theorem 2.4.4 for $\mathbb{G}_n^{(2)}$.

Corollary 3.3.6. *Let $\mathcal{F}$ satisfy the Assumption 3.1.1, 3.3.3, 2.3.1, 3.2.2 and 3.2.1. For ψ from (3.1.5), suppose that*

$$\sup_{n\in\mathbb{N}} \int_0^\infty \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V_n)}d\varepsilon < \infty. \tag{3.3.5}$$

Furthermore, let $\mathbb{D}_n, \mathbb{D}_n^\infty \in (0, \infty)$ be independent of n, and

$$\sup_{i=1,\dots,n} \frac{D_n^\infty(\frac{i}{n})}{\sqrt{n}} \to 0. \tag{3.3.6}$$

Then, the process $\mathbb{G}_n(f)$ is equicontinuous with respect to V_n, that is, for every $\eta > 0$,

$$\lim_{\sigma\to 0} \limsup_{n\to\infty} \mathbb{P}\Big(\sup_{f,g\in\mathcal{F}, V_n(f-g)\le\sigma} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| \ge \eta\Big) = 0.$$

Remark 3.3.7. Compared to Corollary 2.4.5, the condition (3.3.5) of Corollary 3.3.6 is not optimal due to the additional $\psi(\varepsilon)$-factor or, in an explicitly calculated case, a log-factor; the reason here being that we do not approximate the distance $R_n^2(\cdot)$ uniformly over the class $\mathcal{F}$ in an external step but evaluate the needed bounds for $R_n^2(\cdot)$ during the chaining process. This is also the reason why our result does not include the i.i.d. version as a special case. However, in comparison to the results of Lemma 2.8.15 we do not lose much due to this factor in the case of polynomial dependence. Even in the case of exponential decay, the additional factor is of the same order as the factor already present due to dependence.

3.3.3 Further applications

Our theory allows for empirical process theory of general function classes. We illustrate the results with two short examples.

Example 1 (Distribution of residuals)

Consider the locally stationary time series model which is defined recursively via

$$X_i = m\big(X_{i-1}, \frac{i}{n}\big) + \sigma\big(X_{i-1}, \frac{i}{n}\big)\varepsilon_i, \qquad i = 1, ..., n,$$

where ε_i, $i \in \mathbb{Z}$, is an i.i.d. sequence of random variables and $\sigma, m : \mathbb{R} \times [0, 1] \to \mathbb{R}$.

Besides estimation of $m(\cdot), \sigma(\cdot)$, it may also be of interest to derive the distribution function G_ε of ε_i. Following the approach of Akritas and Van Keilegom [2001], we first have to specify estimators $\hat{m}$, $\hat{\sigma}$ for m, σ, respectively, and define empirical residuals $\hat{\varepsilon}_i = \frac{X_i - \hat{m}(X_{i-1}, i/n)}{\hat{\sigma}(X_{i-1}, i/n)}$. Then the convergence of $(\hat{G}_\varepsilon(x))_{x\in\mathbb{R}}$,

$$\hat{G}_\varepsilon(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{\hat{\varepsilon}_i \le x\}} = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\big\{\varepsilon_i \le x\cdot\frac{\hat{\sigma}(X_{i-1}, i/n)}{\sigma(X_{i-1}, i/n)} + \frac{\hat{m}(X_{i-1}, i/n) - m(X_{i-1}, i/n)}{\sigma(X_{i-1}, i/n)}\big\}},$$

can be discussed with empirical process theory and the rather involved analytic properties of $\hat{m}, \hat{\sigma}$ found in Akritas and Van Keilegom [2001].

In the following example we make use of the maximal inequality provided in Section 3.3, Corollary 3.3.2.

Example 2 (Kernel density estimation)

Let $K : \mathbb{R} \to \mathbb{R}$ be some bounded Lipschitz continuous kernel function that satisfies $\int K(u)du = 1$ and has support $\subset [-\frac{1}{2}, \frac{1}{2}]$. For some bandwidth $h := h_n > 0$, put $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$.

We consider the localized density estimate of the density $g_{\tilde{X}_1(v)}$ of the stationary approximation $\tilde{X}_1(v)$,

$$\hat{g}_{n,h}(x, v) = \frac{1}{n}\sum_{i=1}^{n} K_{h_1}\big(\frac{i}{n} - v\big)K_{h_2}(X_i - x)$$

where $h_1, h_2 > 0$ are bandwidths. Suppose that:

- For some $s \le \frac{1}{2}$, $\alpha > s^{-1}$, $\delta_{2s}^X(j) = O(j^{-\alpha})$ and $\sup_{i,n} \|X_i\|_{2s} < \infty$.
- There exist $p_K \ge 2s, C_K > 0$ such that for u large enough, $|K(u)| \le C_K|u|^{-p_K}$.
- There exist constants $C_\infty, L_G > 0$ such that the following holds true. The conditional density $g_{X_i|X_{i-1}=z}$ of X_i given $X_{i-1} = z$ satisfies $|g_{X_i|X_{i-1}=z}|_\infty \le C_\infty$ and for any $x \in \mathbb{R}$, $z \mapsto g_{X_i|X_{i-1}=z}(x)$ is Lipschitz continuous with constant L_G.

We show that if $\log(n)\big(nh_1h_2^{\frac{\alpha(s\wedge\frac{1}{2})}{\alpha(s\wedge\frac{1}{2})-1}}\big)^{-1} = O(1)$,

$$\sup_{x\in\mathbb{R},v\in[0,1]} \big|\hat{g}_{n,h}(x, v) - \mathbb{E}\hat{g}_{n,h}(x, v)\big| = O_p\Big(\sqrt{\frac{\log(n)}{nh_1h_2}}\Big). \tag{3.3.7}$$

To do so, note that

$$\sqrt{nh_1h_2}\big(\hat{g}_{n,h}(x, v) - \mathbb{E}\hat{g}_{n,h}(x, v)\big) = \mathbb{G}_n(f_{x,v})$$

with

$$\mathcal{F} = \{f_{x,v}(z, u) = \sqrt{h_1}K_{h_1}(u - v) \cdot \sqrt{h_2}K_{h_2}(z - x) : x \in \mathbb{R}, v \in [0, 1]\}.$$

To obtain (3.3.7), we use Corollary 3.3.2. We have for $\kappa \in \{1, 2\}$,

$$\begin{aligned} \mu^{(\kappa)}(z) &:= \frac{1}{h_2}\mathbb{E}[K_{h_2}(X_i - x)^\kappa \mid X_{i-1} = z]^\kappa \\ &= \frac{1}{\sqrt{h_2}}\Big(\int K\big(\frac{y - x}{h_2}^\kappa\big)f_{X_i|X_{i-1}=z}(y)dy\Big)^{1/\kappa} \\ &= h_2^{\frac{1}{\kappa}-\frac{1}{2}}\Big(\int K(w)^\kappa f_{X_i|X_{i-1}=z}(x + wh_2)dw\Big)^{1/\kappa}. \end{aligned}$$

Hence,

$$
\begin{aligned}
&|\mu^{(\kappa)}(z) - \mu^{(\kappa)}(z')| \\
&\quad \leq h_2^{\frac{1}{\kappa}-\frac{1}{2}} \Big(\int |K(w)|^{\kappa} |f_{X_i|X_{i-1}=z}(x+wh_2) - f_{X_i|X_{i-1}=z'}(x+wh_2)| dw \Big)^{1/\kappa}.
\end{aligned}
$$

On the other hand, $|f_{X_i|X_{i-1}=z}(x+wh_2) - f_{X_i|X_{i-1}=z'}(x+wh_2)| \leq \min\{L_G|z-z'|, C_\infty\}$. For $s \leq \frac{1}{\kappa}$, we obtain

$$
\begin{aligned}
|\mu^{(\kappa)}(z) - \mu^{(\kappa)}(z')| &\leq h_2^{\frac{1}{\kappa}-\frac{1}{2}} \Big(\int |K(w)|^{\kappa} dw \Big)^{1/\kappa} \cdot \Big[C_\infty \min\big\{1, \frac{L_G}{C_\infty}|z-z'|\big\} \Big]^{1/\kappa} \\
&\leq h_2^{\frac{1}{\kappa}-\frac{1}{2}} \Big(\int |K(w)|^{\kappa} \Big)^{1/\kappa} C_\infty^{\frac{1}{\kappa}-s} L_G^s |z-z'|^s.
\end{aligned}
$$

This shows that Assumption 3.1.1 is satisfied with $R(\cdot) = \frac{1}{2} = C_R$ and $L_{\mathcal{F}} = L_G$ and $\Delta(k) = L_G(k-1)^{-\alpha s}$. As before, it is easily seen that Assumption 3.3.3 is satisfied.

We apply Corollary 3.3.2 with $\bar{F} = \frac{|K|_\infty}{\sqrt{h_2}} =: C_{\bar{F},n}$. For the grids $\mathcal{V}_n = \{in^{-3} : i = 1, ..., n^3\}$, $\mathcal{X}_n = \{in^{-3} : i \in \{-2\lceil n^{3+\frac{1}{2s}} \rceil, ..., 2\lceil n^{3+\frac{1}{2s}} \rceil\}\}$, we obtain

$$
\sqrt{nh_1h_2} \sup_{x \in \mathcal{X}_n, v \in \mathcal{V}_n} \big| \hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}_{n,h}(x,v) \big| = \sup_{x \in \mathcal{X}_n, v \in \mathcal{V}_n} |\mathbb{G}_n(f_{x,v})| = O_p\big(\sqrt{\log(n)}\big).
$$

The discretization of (3.3.7) is rather standard and postponed to Section 3.5, Subsection 3.5.4.

3.4 Concluding remarks

We developed an empirical process theory for locally stationary processes on function classes of (possibly) nonsmooth functions. Here, the dependence was quantified again by the functional dependence measure. In this new setting, we provided maximal inequalities and functional central limit theorems.

Our theory can be applied to, for instance, empirical distribution functions (EDFs) and kernel density estimators. However, it allows for more models to be discussed, as well. Compared to earlier papers in the context of stationary processes and the EDFs, our results have remarkably weak conditions on the dependence structure of the process.

From a technical point of view, working with noncontinuous functions has forced us to modify several approaches from Chapter 2. One key step was to decompose our original process into a martingale and a conditional expectation part.

3.5 Lemmata and proofs of Chapter 3

3.5.1 Proofs of Section 3.1

Lemma 3.5.1. *Let Assumption 3.1.1 be satisfied for some $\nu \geq 2$. Then for all $u \in [0,1]$,*

$$\delta_\nu^{\mathbb{E}[f(Z_i,u)|Z_{i-1}]}(k) \leq |D_{f,n}(u)| \cdot \Delta(k), \tag{3.5.1}$$

$$\sup_i \Big\| \sup_{f\in\mathcal{F}} \big|\mathbb{E}[f(Z_i,u)|Z_{i-1}] - \mathbb{E}[f(Z_i,u)|Z_{i-1}]^{*(i-k)}\big| \Big\|_\nu \leq D_n^\infty(u) \cdot \Delta(k), \tag{3.5.2}$$

$$\sup_i \|f(Z_i,u)\|_2 \leq |D_{f,n}(u)| \cdot C_\Delta. \tag{3.5.3}$$

Furthermore,

$$\Big\|\mathbb{E}[f(Z_i,u)^2|Z_{i-1}] - \mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{*(i-k)}\Big\|_{\nu/2} \leq 2|D_{f,n}(u)| \cdot \|f(Z_i,u)\|_\nu \cdot \Delta(k), \tag{3.5.4}$$

$$\Big\|\sup_{f\in\mathcal{F}} \big|\mathbb{E}[f(Z_i,u)^2|Z_{i-1}] - \mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{*(i-k)}\big|\Big\|_{\nu/2} \leq D_n^\infty(u)^2 \cdot C_\Delta \cdot \Delta(k), \tag{3.5.5}$$

where $C_\Delta := 2\max\{d,\tilde{d}\}|L_\mathcal{F}|_1 C_X^s C_R + C_{\bar{f}}$.

Proof of Lemma 3.5.1. Let $\bar{\mu}_{f,i}^{(1)}(z,u) = \mathbb{E}[\bar{f}(Z_i,u)|Z_{i-1}=z]$ and accordingly $\bar{\mu}_{f,i}^{(2)}(z,u) = \mathbb{E}[\bar{f}(Z_i,u)^2|Z_{i-1}=z]$. We have by Assumption 3.1.1 that

$$\begin{aligned}
&\sup_i \big\|\mathbb{E}[f(Z_i,u)|Z_{i-1}] - \mathbb{E}[f(Z_i,u)|Z_{i-1}]^{*(i-k)}\big\|_\nu \\
=\ & |D_{f,n}(u)| \cdot \sup_i \big\|\bar{\mu}_{f,i}^{(1)}(Z_{i-1},u) - \bar{\mu}_{f,i}^{(1)}(Z_{i-1}^{*(i-k)},u)\big\|_\nu \\
\leq\ & |D_{f,n}(u)| \cdot \sup_i \Big\|\big|Z_{i-1} - Z_{i-1}^{*(i-k)}\big|_{L_\mathcal{F},s}^s\Big\|_{\frac{p\nu}{p-1}} \Big\|R(Z_{i-1},u) + R(Z_{i-1}^{*(i-k)},u)\Big\|_{p\nu} \\
\leq\ & |D_{f,n}(u)| \cdot \sup_i \Big\|\sum_{j=0}^\infty L_{\mathcal{F},j}\big|X_{i-1-j} - X_{i-1-j}^{*(i-k)}\big|_\infty^s\Big\|_{\frac{p\nu}{p-1}} \\
&\quad \times\Big(\|R(Z_{i-1},u)\|_{p\nu} + \Big\|R(Z_{i-1}^{*(i-k)},u)\Big\|_{p\nu}\Big) \\
\leq\ & |D_{f,n}(u)| \cdot 2dC_R \sum_{j=0}^{k-1} L_{\mathcal{F},j}\delta_{\frac{p\nu s}{p-1}}(k-j-1)^s,
\end{aligned}$$

that is, the assertion (3.5.1) holds true with the given $\Delta(k)$. The proof of (3.5.2) is similar.

We now prove (3.5.3). We have

$$\mathbb{E}[f(Z_i,u)^2] = \mathbb{E}[\mathbb{E}[f(Z_i,u)^2|Z_{i-1}]] = D_{f,n}(u)^2\mathbb{E}[\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)^2]$$

and thus $\|f(Z_i,u)\|_2 = |D_{f,n}(u)| \cdot \|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)\|_2$. Since

$$|\bar{\mu}_{f,i}^{(2)}(y,u)| \le |\bar{\mu}_{f,i}^{(2)}(y,u) - \bar{\mu}_{f,i}^{(2)}(0,u)| + |\bar{\mu}_{f,i}^{(2)}(0,u)|,$$

the proof now follows the same lines as in the proof of Lemma 2.8.4.

We now show (3.5.4) and (3.5.5). We have

$$\big|\bar{\mu}_{f,i}^{(2)}(z,u)^2 - \bar{\mu}_{f,i}^{(2)}(z',u)^2\big| = \big|\bar{\mu}_{f,i}^{(2)}(z,u) - \bar{\mu}_{f,i}^{(2)}(z',u)\big| \cdot \big[|\bar{\mu}_{f,i}^{(2)}(z,u)| + |\bar{\mu}_{f,i}^{(2)}(z',u)|\big].$$

We then have by the Cauchy-Schwarz inequality that

$$\begin{aligned}
&\Big\|\sup_{f\in\mathcal{F}}\big|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)^2 - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)^2\big|\Big\|_{\nu/2}\\
\le\ &\Big\|\sup_{f\in\mathcal{F}}\big|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u) - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)\big|\Big\|_{\nu} \cdot 2\Big\|\sup_{f\in\mathcal{F}}\big|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)\big|\Big\|_{\nu}. \qquad (3.5.6)
\end{aligned}$$

Since $\{\bar{\mu}_{f,i}^{(2)} : f \in \mathcal{F}, i \in \{1,...,n\}\}$ forms a $(L_{\mathcal{F}}, s, R, C)$-class, the first factor in (3.5.6) is bounded by $\Delta(k)$ as before. Furthermore,

$$\begin{aligned}
|\bar{\mu}_{f,i}^{(2)}(z,u)| &\le |\bar{\mu}_{f,i}^{(2)}(z,u) - \bar{\mu}_{f,i}^{(2)}(0,u)| + |\bar{\mu}_{f,i}^{(2)}(0,u)|\\
&\le |z|_{L_{\mathcal{F}},s}^s(R(z,u) + R(0,u)) + |\bar{\mu}_{f,i}^{(2)}(0,u)|.
\end{aligned}$$

Note that

$$\begin{aligned}
&\Big\||Z_{i-1}|_{L_{\mathcal{F}},s}^s \cdot \big[R(Z_{i-1},u) + R(0,u)\big]\Big\|_{\nu}\\
\le\ &\Big\|\sum_{j=0}^{\infty} L_{\mathcal{F},j}|Z_{i-1-j}|_\infty^s\Big\|_{\frac{p}{p-1}\nu} \cdot \Big(\|R(Z_{i-1},u)\|_{p\nu} + |R(0,u)|\Big)\\
\le\ &d|L_{\mathcal{F}}|_1 \sup_{i,j}\|X_{ij}\|_{\frac{\nu sp}{p-1}}^s \cdot (C_R + |R(0,u)|)\\
\le\ &2d|L_{\mathcal{F}}|_1 C_X^s C_R.
\end{aligned}$$

We now obtain (3.5.5) from (3.5.6) with the given C_Δ.

By the Cauchy-Schwarz inequality we have for $q \ge 2$,

$$\begin{aligned}
&\delta_{\nu/2}^{\mathbb{E}[f(Z_i,u)^2|Z_{i-1}]}(k)\\
=\ &\sup_i\Big\|\mathbb{E}[f(Z_i,u)^2|Z_{i-1}] - \mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{*(i-k)}\Big\|_{\nu/2}\\
=\ &|D_{f,n}(u)| \cdot \sup_i\Big\|D_{f,n}(u)\big(\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)^2 - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)^2\big)\Big\|_{\nu/2}\\
\le\ &|D_{f,n}(u)| \cdot \sup_i\Big\|\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u) - \bar{\mu}_{f,i}^{(2)}(Z_{i-1}^{*(i-k)},u)\Big\|_{\nu}\\
&\qquad\times 2\Big\|D_{f,n}(u)\bar{\mu}_{f,i}^{(2)}(Z_{i-1},u)\Big\|_{\nu} \qquad (3.5.7)
\end{aligned}$$

Furthermore,

$$\left\| D_{f,n}(u)\bar{\mu}^{(2)}_{f,i}(Z_{i-1},u)\right\|_\nu \le \|\mathbb{E}[f(Z_i,u)^2|Z_{i-1}]^{1/2}\|_\nu \le \|f(Z_i,u)\|_\nu. \tag{3.5.8}$$

Since Assumption 3.1.1 holds true for $\bar{\mu}^{(2)}_{f,i}$, the first factor in (3.5.7) is bounded by $D_{f,n}(u)\Delta(k)$ as in the proof of Lemma 2.8.4. Inserting this and (3.5.8) into (3.5.7), we obtain the result (3.5.4). □

Proof of Corollary 3.1.5. We verify the conditions of Theorem 3.1.2. By $\min\{1,w\} \le w^a$ for $a \in [0,1]$, $w \ge 0$, we have for any $s \in (0, \frac{1}{2}]$,

$$|G_z(x,v) - G_{z'}(x,v)| \le \min\{1, L_G|z-z'|\} \le L_G^s|z-z'|^s$$

and

$$|G_z(x,v) - G_{z'}(x,v)|^{1/2} \le \min\{1, (L_G|z-z'|)^{1/2}\} \le L_G^s|z-z'|^s.$$

This shows Assumption 3.1.1 with $p = \infty$, $R(\cdot) = \frac{1}{2} = C_R$.

Choose $\Delta(k) = cL_G(k-1)^{-\alpha s}$, which can easily be seen to satisfy Assumption 3.3.3 (in particular, $\beta(q) < \infty$ for $q \in \mathbb{N}$) for some $C_\beta = C_\beta(\alpha, s, c, L_G)$ chosen large enough. Regarding Assumption 3.2.1 we first have

$$\begin{aligned}
\frac{1}{c^s}\mathbb{E}\sup_{L_G|a|\le c}[|\mathbb{1}_{\{\tilde{Z}_0(v)\le x\}} - \mathbb{1}_{\{\tilde{Z}_0(v)+a\le x\}}|^2] &\le \frac{1}{c^s}\mathbb{E}|\mathbb{1}_{\{\tilde{Z}_0(v)\le x\}} - \mathbb{1}_{\{\tilde{Z}_0(v)\le x-\frac{c}{L_G}\}}| \\
&\le \frac{1}{c^s}(\mathbb{P}(\tilde{Z}_0(v)\le x) - \mathbb{P}(\tilde{Z}_0(v)\le x - \frac{c}{L_G})) \\
&\le \frac{1}{c^s}(G_z(x,v) - G_z(x-\frac{c}{L_G},v)) \\
&\le \frac{1}{c^s}\min\{1,c\} \le 1.
\end{aligned}$$

The envelope function is the constant 1-function and satisfies the required condition trivially. Therefore, Assumption 3.2.1 holds true. Assumption 3.2.2 is automatically satisfied for fixed $v \in (0,1)$. For Assumption 2.3.3, note that $D_{f,n}(u) = \frac{1}{\sqrt{h}}K(\frac{u-v}{h})$ satisfies

$$\frac{1}{n}\sum_{i=1}^n D_{f,n}(\frac{i}{n})^2 \le \frac{1}{nh}\sum_{i=1}^n K(\frac{i/n-v}{h})^2 \le |K|_\infty^2 < \infty,$$

and $D_{f,n}^\infty \le \frac{1}{\sqrt{h}}|K|_\infty$. Thus, $\frac{D_{f,n}^\infty}{\sqrt{n}} \le \frac{|K|_\infty}{\sqrt{nh}} \to 0$, and the support satisfies $\text{supp}[D_{f,n}(\cdot)] \subset [v-h, v+h]$. Finally, $h^{1/2}D_{f,n}^\infty \le |K|_\infty < \infty$ and, since $v \in (0,1)$,

$$\lim_{n\to\infty}\int_0^1 D_{f,n}(u)D_{g,n}(u)du = \lim_{n\to\infty}\frac{1}{h}\int_0^1 K(\frac{u-v}{h})^2du = \int K(u)^2du.$$

This shows all the conditions of Assumption 2.3.3 (ii).
Note that $\mathbb{H}(\varepsilon, \mathcal{H}, \|\cdot\|_{2,n}) = O(\log(\varepsilon^{-1}))$ which is proven subsequently.

Let $\varepsilon > 0$. Since $\lim_{x\to-\infty}\sup_{i,n} G_i(x) = 0$ and $\lim_{x\to+\infty}\inf_{i,n} G_i(x) = 1$, there exists $x_N = x_N(\varepsilon) > x_1 = x_1(\varepsilon) > 0$ such that $\sup_{i,n} G_i(x_1) \le \varepsilon$, $\inf_{i,n} G_i(x_1) \ge 1-\varepsilon$. Define $x_{j+1} := x_1 + j\cdot\frac{\varepsilon^2}{L_G}$, $j = 1,2,...,N-1$, with $N = 1 + \lceil\frac{(x_N-x_1)L_G}{\varepsilon^2}\rceil$. Put $x_0 = -\infty$ and $x_{N+1} = \infty$. Then for $j = 1,2,...,N-1$ we have

$$\begin{aligned}
&\|\mathbb{1}_{\{\cdot\le x_{j+1}\}} - \mathbb{1}_{\{\cdot\le x_{j+1}\}}\|_{2,n}^2\\
\le\ &\sup_{i=1,...,n}\mathbb{E}[(\mathbb{1}_{\{X_i\le x_{j+1}\}} - \mathbb{1}_{\{X_i\le x_j\}})^2] = \sup_{i=1,...,n}[G_i(x_{j+1}) - G_i(x_j)]\\
\le\ &L_G|x_{j+1} - x_j| \le \varepsilon^2,
\end{aligned}$$

which shows that $[\mathbb{1}_{\{\cdot\le x_j\}}, \mathbb{1}_{\{\cdot\le x_{j+1}\}}]$, $j = 0,...,N$, are ε-brackets with respect to $\|\cdot\|_{2,n}$. Hence, $\mathbb{H}(\varepsilon,\mathcal{H},\|\cdot\|_{2,n}) = O(\log(\varepsilon^{-1}))$.

Since $\alpha s > 1$, Table 3.1 implies that

$$\int_0^1 \psi(\varepsilon)\sqrt{\mathbb{H}(\gamma,\mathcal{H},V_n)}d\varepsilon = O\Big(\int_0^1 \psi(\varepsilon)\varepsilon^{-\frac{1}{\alpha s}}\sqrt{\log(\varepsilon^{-1})}d\varepsilon\Big) < \infty.$$

Theorem 3.1.2 now implies the assertion.

□

3.5.2 Proofs of Section 3.3.1

Proof of Theorem 3.3.1

In this section, we consider

$$W_i(f) = \mathbb{E}[f(Z_i,\frac{i}{n})^2|Z_{i-1}], \qquad S_n(f) := \sum_{i=1}^n\big\{W_i(f) - \mathbb{E}W_i(f)\big\}.$$

Then,

$$R_n^2(f) = \frac{1}{n}\sum_{i=1}^n W_i(f), \qquad R_n^2(f) - \mathbb{E}R_n^2(f) = \frac{1}{n}S_n(f).$$

We obtain from Lemma 3.5.1, (3.5.4) and (3.5.5) the following results for $\nu = 2$.

Lemma 3.5.2. *Suppose that Assumption 3.1.1 holds true. Then for each $i = 1,...,n$, $j\in\mathbb{N}$, $s\in\mathbb{N}\cup\{\infty\}$, $f\in\mathcal{F}$,*

$$\begin{aligned}
\Big\|\sup_{f\in\mathcal{F}}\big|W_i(f) - W_i(f)^{*(i-j)}\big|\Big\|_1 &\le C_\Delta D_n^\infty(\frac{i}{n})^2\Delta(j),\\
\big\|W_i(f) - W_i(f)^{*(i-j)}\big\|_1 &\le 2|D_{f,n}(\frac{i}{n})|\cdot\|f(Z_i,\frac{i}{n})\|_2\Delta(j),\\
\big\|W_i(f)\big\|_s &\le \|f(Z_i,\frac{i}{n})\|_{2s}^2.
\end{aligned}$$

We approximate $W_i(f)$ by independent variables as follows (cf. also Wu et al. [2013], Zhang and Wu [2017]). Let

$$W_{i,j}(f) := \mathbb{E}[W_i(f)|\varepsilon_{i-j}, \varepsilon_{i-j+1}, ..., \varepsilon_i], \qquad j \in \mathbb{N},$$

and

$$S_{n,j}(f) := \sum_{i=1}^{n}\{W_{i,j}(f) - \mathbb{E}W_{i,j}(f)\}.$$

We now follow the decomposition scheme that we already applied in equation (2.8.23).

The next result is a uniform bound on means of independent random variables.

Lemma 3.5.3. *Assume that $Q_i(f)$, $i = 1, ..., m$ are independent variables indexed by $f \in \mathcal{F}$ which fulfill $\mathbb{E}Q_i(f) = 0$, $\frac{1}{m}\sum_{i=1}^{m}\|Q_i(f)\|_1 \le \sigma_Q$ and $|Q_i(f)| \le M_Q$ almost surely ($i = 1, ..., n$). Then there exists some universal constant $c > 0$ such that*

$$\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{m}\sum_{i=1}^{m}Q_i(f)\Big| \le c\Big(\sigma_Q + \frac{M_Q H}{m}\Big), \tag{3.5.9}$$

where H is defined by (1.2.5).

Proof of Lemma 3.5.3. Let $Q_i = Q_i(f)$. By Bernstein's inequality, we have for each $f \in \mathcal{F}$,

$$\begin{aligned}\mathbb{P}\Big(\Big|\frac{1}{m}\sum_{i=1}^{m}Q_i\Big| \ge x\Big) &\le 2\exp\Big(-\frac{1}{2}\frac{x^2}{\frac{1}{m^2}\sum_{i=1}^{m}\|Q_i\|_2^2 + x\frac{M_Q}{m}}\Big)\\ &\le 2\exp\Big(-\frac{1}{2}\frac{x^2}{\frac{M_Q}{m}\cdot\sigma_Q + x\frac{M_Q}{m}}\Big),\end{aligned}$$

where we used in the last step that $\|Q_i\|_2^2 = \mathbb{E}[Q_i^2] \le M_Q\|Q_i\|_1$.

With standard arguments (cf. [van der Vaart, 1998, proof of Lemma 19.33]), we conclude that there exists some universal constant $c_1 > 0$ with

$$\mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{m}\sum_{i=1}^{m}Q_i(f)\Big| \le c_1\Big(\sqrt{H}(\frac{\sigma_Q M_Q}{m})^{1/2} + \frac{M_Q H}{m}\Big).$$

The result follows by using $(\frac{H\sigma_Q M_Q}{m})^{1/2} \le 2\frac{M_Q H}{m} + 2\sigma_Q$. □

We now prove Theorem 3.3.1 based on Lemma 3.5.2, Lemma 3.5.3 and the decomposition (2.8.23).

Proof of Theorem 3.3.1. We first discuss A_2. We have

$$\sum_{l=1}^{L}\mathbb{E}\max_{f\in\mathcal{F}}\frac{1}{\frac{n}{\tau_l}}\Big|\sum_{1\le i\le\lfloor\frac{n}{\tau_l}\rfloor+1, i\text{ odd}}\frac{1}{\tau_l}T_{i,l}(f)\Big|.$$

Since $\|W_{k,j}(f) - W_{k,j-1}(f)\|_1 \le 2\min\{\|W_k(f)\|_1, \delta_1^{W_k(f)}(j-1)\}$, we have for each $f \in \mathcal{F}$,

$$\begin{aligned}
\frac{1}{\tau_l}\|T_{i,l}\|_1 &\le \sum_{j=\tau_{l-1}+1}^{\tau_l} \frac{1}{\tau_l}\Big\| \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} (W_{k,j} - W_{k,j-1})\Big\|_1 \\
&\le \sum_{j=\tau_{l-1}+1}^{\tau_l} \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \Big\|W_{k,j} - W_{k,j-1}\Big\|_1 \\
&\le 2\sum_{j=\tau_{l-1}+1}^{\tau_l} \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \min\{\|W_k(f)\|_1, \delta_1^{W_k(f)}(j-1)\} \\
&\le 2\sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \|W_k(f)\|_1, \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \delta_1^{W_k(f)}(j-1)\} \\
&= 2\sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\sigma_{i,l}, \Delta_{i,j,l}\},
\end{aligned}$$

where

$$\sigma_{i,l} := \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \|W_k(f)\|_1, \qquad \Delta_{i,j,l} := \frac{1}{\tau_l} \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \delta_1^{W_k(f)}(j-1).$$

We conclude that

$$\begin{aligned}
\frac{1}{\lfloor \frac{n}{\tau_l}\rfloor + 1} \sum_{i=1}^{\lfloor \frac{n}{\tau_l}\rfloor+1} \frac{1}{\tau_l}\|T_{i,l}\|_1 &\le 2\sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l}\rfloor+1} \sigma_{i,l}, \frac{1}{\frac{n}{\tau_l}} \sum_{i=1}^{\lfloor \frac{n}{\tau_l}\rfloor+1} \Delta_{i,j,l}\} \\
&\le \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{n}\sum_{i=1}^{n} \|W_i(f)\|_1, \frac{1}{n}\sum_{i=1}^{n} \delta_1^{W_i}(j)\}. \qquad (3.5.10)
\end{aligned}$$

Furthermore,

$$\frac{1}{\tau_l}|T_{i,l}| \le 2\sup_i \|W_i(f)\|_\infty \le 2\|f\|_\infty^2 \le 2M^2. \qquad (3.5.11)$$

By Lemma 3.5.3, (3.5.9), we have for some universal constant $c_1 > 0$,

$$\begin{aligned}
\mathbb{E}A_2 &\le 2c_1 \sum_{l=1}^{L} \Big[\sup_{f\in\mathcal{F}} \Big(\frac{1}{\lfloor \frac{n}{\tau_l}\rfloor + 1} \sum_{i=1}^{\lfloor \frac{n}{\tau_l}\rfloor+1} \frac{1}{\tau_l}\|T_{i,l}(f)\|_1\Big) + \frac{2M^2 H}{\lfloor \frac{n}{\tau_l}\rfloor + 1}\Big] \\
&\le 2c_1 \Big(\sum_{l=1}^{L} \sup_{f\in\mathcal{F}} \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\frac{1}{n}\sum_{i=1}^{n} \|W_i(f)\|_1, \frac{1}{n}\sum_{i=1}^{n} \delta_1^{W_i(f)}(j)\} \\
&\qquad + \frac{qM^2 H}{n}\Big). \qquad (3.5.12)
\end{aligned}$$

By Lemma 3.5.2 and the Cauchy-Schwarz inequality for sums,

$$\begin{aligned}
&\sum_{l=1}^{L}\sup_{f\in\mathcal{F}}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\frac{1}{n}\sum_{i=1}^{n}\|W_i(f)\|_1,\frac{1}{n}\sum_{i=1}^{n}\delta_1^{W_i}(j)\}\\
\le\ &\sum_{l=1}^{L}\sup_{f\in\mathcal{F}}\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\frac{1}{n}\sum_{i=1}^{n}\|f(Z_i,\frac{i}{n})\|_2^2,\frac{2}{n}\sum_{i=1}^{n}D_{f,n}(\frac{i}{n})\|f(Z_i,\frac{i}{n})\|_2\cdot\Delta(j)\}\\
\le\ &\sum_{j=1}^{\infty}\min\{\sup_{f\in\mathcal{F}}\|f\|_{2,n}^2,2\mathbb{D}_n\sup_{f\in\mathcal{F}}\|f\|_{2,n}\cdot\Delta(j)\}\\
=\ &\sup_{f\in\mathcal{F}}\|f\|_{2,n}\cdot\bar{V}_n(\sup_{f\in\mathcal{F}}\|f\|_{2,n})\\
=\ &\sup_{f\in\mathcal{F}}\big(\|f\|_{2,n}\cdot\bar{V}_n(\|f\|_{2,n})\big)\\
\le\ &\sup_{f\in\mathcal{F}}\big[\|f\|_{2,n}V_n(f)\big], \qquad (3.5.13)
\end{aligned}$$

where

$$\bar{V}_n(x)=x+\sum_{j=1}^{\infty}\min\{x,\mathbb{D}_n\Delta(j)\}. \qquad (3.5.14)$$

The second-to-last equality uses the fact that $x\mapsto x\cdot\bar{V}_n(x)$ is increasing in x.

We also have $\|W_{i,0}(f)-\mathbb{E}W_{i,0}(f)\|_\infty\le 2\|f\|_\infty^2\le 2M^2$ and $\|W_{i,0}(f)-\mathbb{E}W_{i,0}(f)\|_1\le 2\|W_i(f)\|_1$. Thus by Lemma 3.5.3, (3.5.9),

$$\begin{aligned}
\mathbb{E}A_3 &\le \mathbb{E}\max_{f\in\mathcal{F}}\Big|\frac{1}{n}\sum_{i=1}^{n}(W_{i,0}(f)-\mathbb{E}W_{i,0}(f))\Big|\\
&\le 2c_1\Big(\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\|W_i(f)\|_1+\frac{M^2H}{n}\Big) \qquad (3.5.15)\\
&\le 2c_1\Big(\sup_{f\in\mathcal{F}}\|f\|_{2,n}^2+\frac{M^2H}{n}\Big). \qquad (3.5.16)
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbb{E}A_1 &\le \sum_{j=q}^{\infty}\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\frac{1}{n}\sum_{i=1}^{n}(W_{i,j+1}(f)-W_{i,j}(f))\Big|\\
&\le \sum_{j=q}^{\infty}\frac{1}{n}\sum_{i=1}^{n}\big\|\sup_{f\in\mathcal{F}}|W_{i,j+1}(f)-W_{i,j}(f)|\big\|_1.
\end{aligned}$$

Since $|W_{i,j+1}(f)-W_{i,j}(f)|=|\mathbb{E}[W_i(f)^{**(i-j)}-W_i(f)^{**(i-j+1)}|\mathcal{A}_i]|\le\mathbb{E}[|W_i(f)^{**(i-j)}-W_i(f)^{**(i-j+1)}|\,|\mathcal{A}_i]$, where we use the already seen notation $H(\mathcal{F}_i)^{**(i-j)}:=H(\mathcal{F}_i^{**(i-j)})$

and $\mathcal{F}_i^{**(i-j)} = (\varepsilon_i, \varepsilon_{i-1}, ..., \varepsilon_{i-j}, \varepsilon_{i-j-1}^*, \varepsilon_{i-j-2}^*, ...)$, we have

$$\begin{aligned} & \big\| \sup_{f\in\mathcal{F}} |W_{i,j+1}(f) - W_{i,j}(f)| \big\|_1 \\ \leq\ & \big\| \mathbb{E}[\max_{f\in\mathcal{F}} |W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}| \,|\mathcal{A}_i] \big\|_1 \\ \leq\ & \big\| \sup_{f\in\mathcal{F}} |W_i(f)^{**(i-j)} - W_i(f)^{**(i-j+1)}| \big\|_1 \\ =\ & \big\| \sup_{f\in\mathcal{F}} |W_i(f) - W_i(f)^{*(i-j)}| \big\|_1 \leq D_n^\infty(\frac{i}{n})^2 C_\Delta \Delta(j), \end{aligned} \tag{3.5.17}$$

which shows that

$$\mathbb{E}A_1 \leq (\mathbb{D}_n^\infty)^2 C_\Delta \beta(q). \tag{3.5.18}$$

Collecting the upper bounds (3.5.12), (3.5.13), (3.5.16) and (3.5.18), we obtain

$$\mathbb{E}\max_{f\in\mathcal{F}} \Big|\frac{1}{n}S_n(f)\Big| \leq (4c_1 + 1)\cdot \Big[\sup_{f\in\mathcal{F}} \big[\|f\|_{2,n} V_n(f)\big] + (\mathbb{D}_n^\infty)^2 C_\Delta \beta(q) + \frac{qM^2H}{n}\Big]. \tag{3.5.19}$$

By (3.5.28), $V_n(f) \leq \sigma$ implies $\|f\|_{2,n}^2 \leq \mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})\|f\|_{2,n}$, whence

$$\|f\|_{2,n} \leq \mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n}).$$

Thus,

$$\sup_{f\in\mathcal{F}} \big[\|f\|_{2,n} V_n(f)\big] \leq \mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma. \tag{3.5.20}$$

Inserting (3.5.20) into (3.5.19) yields the first equation (3.3.1) of the lemma.

We now show (3.3.2) with a case distinction. We abbreviate $q^* = q^*(\frac{M^2H}{n(\mathbb{D}_n^\infty)^2 C_\Delta})$. If $q^*\frac{H}{n} \leq 1$, we have $q^* \in \{1, ..., n\}$ and thus

$$\begin{aligned} P &\leq c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + (\mathbb{D}_n^\infty)^2 C_\Delta \beta(q^*) + q^*\frac{M^2H}{n}\Big) \\ &\leq 2c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + q^*\frac{M^2H}{n}\Big) \\ &= 2c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + M^2\cdot\min\Big\{q^*\frac{H}{n}, 1\Big\}\Big). \end{aligned} \tag{3.5.21}$$

If $q^*\frac{H}{n} \geq 1$, choose $q_0 = \lfloor \frac{n}{H} \rfloor \leq \frac{n}{H}$. By simply bounding each summand with M^2, we have

$$\begin{aligned} \mathbb{E}\max_{f\in\mathcal{F}} \Big|\frac{1}{n}S_n(f)\Big| &\leq M^2 \leq c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + M^2\Big) \\ &\leq 2c\Big(\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + M^2\cdot\min\Big\{q^*\frac{H}{n}, 1\Big\}\Big). \end{aligned} \tag{3.5.22}$$

Putting the two bounds (3.5.21) and (3.5.22) together, we obtain the result (3.3.2). □

The following lemma is an auxiliary result to prove Corollary 3.3.2 and Lemma 3.5.5.

Lemma 3.5.4. *Let $\mathcal{F}$ be some finite class of functions. Let $R > 0$ be arbitrary and assume that $\sup_{f\in\mathcal{F}} \|f\|_\infty \leq M$. Then there exists a universal constant $c > 0$ such that*

$$\mathbb{E} \max_{f\in\mathcal{F}} \big|\mathbb{G}_n^{(1)}(f)\big| \mathbb{1}_{\{R_n^2(f)\leq R^2\}} \leq c\Big\{R\sqrt{H} + \frac{MH}{\sqrt{n}}\Big\}, \tag{3.5.23}$$

where H is defined by (1.2.5).

Proof of Lemma 3.5.4. By Theorem 3.3 in Pinelis [1994], for $x, a > 0$ and a measurable function f,

$$\mathbb{P}\Big(\big|\mathbb{G}_n^{(1)}(f)\big| \geq x, R_n^2(f) \leq R^2\Big) \leq 2\exp\Big(-\frac{1}{2}\frac{x^2}{R^2 + \frac{2\|f\|_\infty x}{3\sqrt{n}})}\Big).$$

Using standard arguments (cf.[van der Vaart, 1998, proof of Lemma 19.33]), we obtain (3.5.23). □

Proof of Corollary 3.3.2. Let us, again, define the following functions first.

For $m > 0$, define $\varphi_m^\wedge : \mathbb{R} \to \mathbb{R}$ and the corresponding "peaky" residual function $\varphi_m^\vee : \mathbb{R} \to \mathbb{R}$ via

$$\varphi_m^\wedge(x) := (x \vee (-m)) \wedge m, \qquad \varphi_m^\vee(x) := x - \varphi_m^\wedge(x).$$

Now, let $Q \geq 1$, and $\sigma := \sup_{n\in\mathbb{N}} \sup_{f\in\mathcal{F}} V_n(f) < \infty$. Put

$$M_n = \frac{\sqrt{n}}{\sqrt{H}} r\big(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty}\big)\mathbb{D}_n^\infty.$$

Let $F(z, u) := D_n^\infty(u) \cdot \bar{F}(z, u)$ and recall $\bar{F} = \sup_{f\in\mathcal{F}} \bar{f}$. Then,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}} |\mathbb{G}_n(f)| > Q\sqrt{H}\Big) \\
&\leq \mathbb{P}\Big(\max_{f\in\mathcal{F}} |\mathbb{G}_n(f)| > Q\sqrt{H}, \sup_{i=1,\ldots,n} F(Z_i, \frac{i}{n}) \leq M_n\Big) \\
&\quad + \mathbb{P}\Big(\sup_{i=1,\ldots,n} F(Z_i, \frac{i}{n}) > M_n\Big) \\
\leq\ & \mathbb{P}\Big(\max_{f\in\mathcal{F}} |\mathbb{G}_n(\varphi_{M_n}^\wedge(f))| > \frac{Q\sqrt{H}}{2}\Big) \\
&\quad + \mathbb{P}\Big(\frac{1}{\sqrt{n}} \max_{f\in\mathcal{F}} \big|\sum_{i=1}^n \mathbb{E}[f(Z_i, \frac{i}{n})\mathbb{1}_{\{|f(Z_i,\frac{i}{n})|>M_n\}}]\big| > \frac{Q\sqrt{H}}{2}\Big) \\
&\quad + \mathbb{P}\Big(\sup_{i=1,\ldots,n} F(Z_i, \frac{i}{n}) > M_n\Big).
\end{aligned} \tag{3.5.24}$$

For the first summand in (3.5.24), we use the decomposition

$$\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{2}\Big)\\
\leq\ & \mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}^{(1)}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{4}\Big) + \mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}^{(2)}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{4}\Big)\\
\leq\ & \mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}^{(1)}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{4}, \max_{f\in\mathcal{F}} R_n^2(\varphi^{\wedge}_{M_n}(f)) \leq \sigma^2\Big)\\
&\quad +\mathbb{P}\Big(\max_{f\in\mathcal{F}} R_n^2(\varphi^{\wedge}_{M_n}(f)) > \sigma^2\Big)\\
&\quad +\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}^{(2)}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{4}\Big).
\end{aligned} \tag{3.5.25}$$

We now discuss the three terms separately. By Lemma 3.5.4, we have

$$\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}^{(1)}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{4}, \max_{f\in\mathcal{F}} R_n^2(\varphi^{\wedge}_{M_n}(f)) \leq Q^{3/2}\sigma^2\Big)\\
\leq\ & \frac{4c}{Q\sqrt{H}}\Big[\sigma Q^{3/4}\sqrt{H} + \frac{M_n H}{\sqrt{n}}\Big] \leq \frac{4c}{Q\sqrt{H}}\Big[\sigma Q^{3/4}\sqrt{H} + \sigma\sqrt{H}Q^{1/2}\Big] \leq \frac{8c}{Q^{1/4}}.
\end{aligned}$$

By Theorem 3.3.1 and (3.5.30),

$$\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}} R_n^2(\varphi^{\wedge}_{M_n}(f)) > Q^{3/2}\sigma^2\Big)\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\mathbb{D}_n r(\frac{\sigma}{\mathbb{D}_n})\sigma + q^*\Big(\frac{M^2H}{n(\mathbb{D}_n^\infty)^2 C_\Delta}\Big)\frac{M^2H}{n}\Big]\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\sigma^2 + q^*\Big(\frac{r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})^2}{C_\Delta}\Big) r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})^2(\mathbb{D}_n^\infty)^2\Big]\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\sigma^2 + q^*\Big(C_\Delta^{-1}C_\beta^{-2}\Big)\cdot\Big[q^*\Big(r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})\Big) r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})\Big]^2(\mathbb{D}_n^\infty)^2\Big]\\
\leq\ & \frac{2c}{\sigma^2 Q^{3/2}}\Big[\sigma^2 + q^*\Big(C_\Delta^{-1}C_\beta^{-2}\Big)\sigma^2 Q\Big]|\\
\leq\ & \frac{2c}{Q^{1/2}}\big[1 + q^*\Big(C_\Delta^{-1}C_\beta^{-2}\Big)\big]
\end{aligned}$$

for C_Δ defined in Lemma 3.5.1.

By Theorem 2.4.1 applied to $W_i(f) = \mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]$,

$$\begin{aligned}
&\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}^{(2)}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{4}\Big)\\
\leq\ & \frac{8c}{Q\sqrt{H}}\cdot\Big[\sigma\sqrt{H} + q^*\Big(r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})\Big) r(\frac{\sigma Q^{1/2}}{\mathbb{D}_n^\infty})\mathbb{D}_n^\infty\Big]\\
\leq\ & \frac{8c}{Q\sqrt{H}}\big[\sigma\sqrt{H} + \sigma Q^{1/2}\sqrt{H}\big] \leq \frac{16c\sigma}{Q^{1/2}}.
\end{aligned}$$

Inserting the upper bounds into (3.5.25), we obtain

$$\mathbb{P}\Big(\max_{f\in\mathcal{F}}|\mathbb{G}_n(\varphi^{\wedge}_{M_n}(f))| > \frac{Q\sqrt{H}}{2}\Big) \le \frac{8c}{Q^{1/4}} + \frac{2c}{Q^{1/2}}\big[1 + q^*\Big(C_{\Delta}^{-1}C_{\beta}^{-2}\Big)\big] + \frac{16c\sigma}{Q^{1/2}} \to 0$$

for $Q \to \infty$. The second and third summand in (3.5.24) were already discussed in the proof of Corollary 2.4.3 (equations (2.8.43), (2.8.44) therein) and converge to 0 for $Q \to \infty$ under the given assumptions; note especially that we only need $\|\bar{F}(Z_i, \frac{i}{n})\|_{\nu_2} \le C_{\bar{F},n}$ instead of C_Δ which is part of the assumptions). □

The following Lemma 3.5.5 is used to prove Theorem 3.3.4.

Lemma 3.5.5 (Compatibility lemma 2)**.** *Let $\psi : (0,\infty) \to [1,\infty)$ be some function and $k \in \mathbb{N}$, $\delta > 0$. If $\mathcal{F}$ fulfills $|\mathcal{F}| \le k$ and Assumption 3.1.1, 3.3.3, then there exists some universal constant $c > 0$ such that the following holds true: If $\sup_{f\in\mathcal{F}} V_n(f) \le \delta$ and $\sup_{f\in\mathcal{F}} \|f\|_\infty \le m(n,\delta,k)$, then*

$$\mathbb{E}\max_{f\in\mathcal{F}}\big|\mathbb{G}_n^{(1)}(f)\big|\mathbb{1}_{\{R_n(f)\le 2\delta\psi(\delta)\}} \le 2c(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})\cdot\psi(\delta)\delta\sqrt{H(k)}, \quad (3.5.26)$$

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}} R_n(f) > 2\delta\psi(\delta)\Big) \le \frac{2c(1+q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})^2)}{\psi(\delta)^2}. \quad (3.5.27)$$

Proof of Lemma 3.5.5. By Lemma 3.5.4 and since $r(a) \le a$ (cf. Lemma 2.8.6),

$$\begin{aligned}\mathbb{E}\max_{f\in\mathcal{F}}\big|\mathbb{G}_n^{(1)}(f)\big|\mathbb{1}_{\{R_n(f)\le 2\delta\psi(\delta)\}} &\le c\Big\{2\psi(\delta)\delta\sqrt{H(k)} + \frac{m(n,\delta,k)H(k)}{\sqrt{n}}\Big\}\\ &\le 2c\cdot\big[\psi(\delta)\cdot\delta + \mathbb{D}_n^\infty r(\frac{\delta}{\mathbb{D}_n})\big]\sqrt{H(k)}\\ &\le 2c\cdot(1+\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})\cdot\psi(\delta)\delta\sqrt{H(k)},\end{aligned}$$

which shows (3.5.26).

For $\hat{a} = \arg\min_{j\in\mathbb{N}}\big\{\|f\|_{2,n}\cdot j + \mathbb{D}_n\beta(j)\big\}$ and since $\|f\|_{2,n} \le V_n(f) \le \delta$ we have with $r(\frac{\delta}{\mathbb{D}_n}) \ge \frac{\delta}{\mathbb{D}_n\hat{a}}$,

$$\frac{\|f\|_{2,n}^2}{\mathbb{D}_n^\infty r(\frac{\delta}{\mathbb{D}_n})} \le \frac{\mathbb{D}_n\hat{a}\|f\|_{2,n}^2}{\mathbb{D}_n^\infty\delta} \le \frac{\mathbb{D}_n V_n(f)\|f\|_{2,n}}{\mathbb{D}_n^\infty\delta} \le \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\|f\|_{2,n}. \quad (3.5.28)$$

Therefore, $\|f\|_{2,n}^2 \le \mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})\|f\|_{2,n}$ and thus $\|f\|_{2,n} \le \mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})$. Note that due to $r(a) \le a$,

$$\mathbb{E}R_n^2(f) = \frac{1}{n}\sum_{i=1}^n\mathbb{E}[f(Z_i,\frac{i}{n})^2] \le \|f\|_{2,n}^2 \le (\mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n}))^2 \le \delta^2. \quad (3.5.29)$$

Recall that $\beta_{norm}(q) = \frac{\beta(q)}{q}$. By Assumption 3.3.3, we have that for any $x_1, x_2 > 0$, $\tilde{q} = q^*(x_1)q^*(x_2)$ satisfies

$$\beta_{norm}(\tilde{q}) \le C_\beta\beta_{norm}(q^*(x_1))\beta_{norm}(q^*(x_2)) \le C_\beta x_1 x_2.$$

Thus, by definition of q^*,

$$q^*(C_\beta x_1 x_2) \leq q^*(x_1) q^*(x_2). \tag{3.5.30}$$

We obtain

$$q^*\Big(r(\frac{\delta}{\mathbb{D}_n})^2 \frac{1}{C_\Delta}\Big) \leq q^*\Big(r(\frac{\delta}{\mathbb{D}_n})\Big)^2 q^*\big(C_\Delta^{-1} C_\beta^{-2}\big). \tag{3.5.31}$$

By (3.5.29), Markov's inequality, Theorem 3.3.1 and (3.5.31),

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{f\in\mathcal{F}} R_n^2(f) > 2\psi(\delta)^2\delta^2\Big) \\
\leq\ &\mathbb{P}\Big(\sup_{f\in\mathcal{F}} |R_n^2(f) - \mathbb{E}R_n^2(f)| > \psi(\delta)^2\delta^2\Big) \\
\leq\ &\frac{2c}{\psi(\delta)^2\delta^2}\cdot\Big[\mathbb{D}_n r(\frac{\delta}{\mathbb{D}_n})\delta + q^*\Big(r(\frac{\delta}{\mathbb{D}_n})^2\frac{1}{C_\Delta}\Big) r(\frac{\delta}{\mathbb{D}_n})^2 (\mathbb{D}_n^\infty)^2\Big] \\
\leq\ &\frac{2c}{\psi(\delta)^2\delta^2}\cdot\Big[\delta^2 + \Big[q^*\Big(r(\frac{\delta}{\mathbb{D}_n})\Big) r(\frac{\delta}{\mathbb{D}_n})\Big]^2 q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\mathbb{D}_n^\infty)^2\Big] \\
\leq\ &\frac{2c}{\psi(\delta)^2\delta^2}\cdot\Big[\delta^2 + \delta^2 q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})^2\Big] \\
\leq\ &\frac{2c(1 + q^*\big(C_\Delta^{-1}C_\beta^{-2}\big)(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})^2)}{\psi(\delta)^2},
\end{aligned}$$

which shows (3.5.27). □

Proof of Theorem 3.3.4. In the following, we abbreviate $\mathbb{H}(\delta) = \mathbb{H}(\delta, \mathcal{F}, V)$ and $\mathbb{N}(\delta) = \mathbb{N}(\delta, \mathcal{F}, V)$. The proof follows the lines of Theorem 2.4.4. We present it here for completeness. Recall one last time, $\varphi_m^\wedge : \mathbb{R} \to \mathbb{R}$ and the corresponding "peaky" residual function $\varphi_m^\vee : \mathbb{R} \to \mathbb{R}$ via

$$\varphi_m^\wedge(x) := (x \vee (-m)) \wedge m, \qquad \varphi_m^\vee(x) := x - \varphi_m^\wedge(x)$$

for $m > 0$.

We choose $\delta_0 = \sigma$ and $\delta_j = 2^{-j}\delta_0$, and

$$m_j = \frac{1}{2} m(n, \delta_j, N_{j+1})$$

as well as $M_n = \frac{1}{2} m_0$. We then use

$$\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(f)\Big| \leq \mathbb{E}\sup_{f\in\mathcal{F}(M_n)}\Big|\mathbb{G}_n^{(1)}(f)\Big| + \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}\big[F(Z_i)\mathbb{1}_{\{F(Z_i)>M_n\}}\big], \tag{3.5.32}$$

where $\mathcal{F}(M_n) := \{\varphi_{M_n}^\wedge(f) : f \in \mathcal{F}\}$.

We construct a nested sequence of partitions $(\mathcal{F}_{jk})_{k=1,\dots,N_j}$, $j \in \mathbb{N}$ of $\mathcal{F}(M_n)$ (where $N_j := \mathbb{N}(\delta_0) \cdot \ldots \cdot \mathbb{N}(\delta_j)$), and a sequence Δ_{jk} of measurable functions such that

$$\sup_{f,g\in\mathcal{F}_{jk}} |f - g| \leq \Delta_{jk}, \qquad V(\Delta_{jk}) \leq \delta_j.$$

In each $\mathcal{F}_{jk}$, we fix some $f_{jk} \in \mathcal{F}$, and define $\pi_j f := f_{j,\psi_j f}$ where $\psi_j f := \min\{i \in \{1, ..., N_j\} : f \in \mathcal{F}_{ji}\}$, and put $\Delta_j f := \Delta_{j,\psi_j f}$, and

$$I(\sigma) := \int_0^\sigma \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V)}d\varepsilon$$

as well as

$$\tau := \min\left\{j \geq 0 : \delta_j \leq \frac{I(\sigma)}{\sqrt{n}}\right\} \vee 1. \tag{3.5.33}$$

For functions f, g with $|f| \leq g$,

$$\begin{aligned}
|\mathbb{G}_n^{(1)}(f)| &\leq |\mathbb{G}_n^{(1)}(g)| + 2\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n \mathbb{E}[g(Z_i, \frac{i}{n})|Z_{i-1}] \\
&\leq |\mathbb{G}_n^{(1)}(g)| + 2|\mathbb{G}_n^{(2)}(g)| + 2\sqrt{n} \cdot \frac{1}{n}\sum_{i=1}^n \mathbb{E}[g(Z_i, \frac{i}{n})] \\
&\leq |\mathbb{G}_n^{(1)}(g)| + 2|\mathbb{G}_n^{(2)}(g)| + 2\sqrt{n}\|g\|_{1,n}.
\end{aligned}$$

Using a similar approach as in Subsection 2.8.3, equations (2.8.47) and (2.8.48) applied to $W_i(f) = f(Z_i, \frac{i}{n}) - \mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]$ and the fact that $\|f - \pi_0 f\|_\infty \leq 2M_n \leq m_0$, we have the decomposition

$$
\begin{aligned}
\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(f)| &\le \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\pi_0 f)| \\
&\qquad + \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_\tau}(f-\pi_\tau f))| \\
&\qquad\quad + \sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_j-m_{j+1}}(\pi_{j+1}f-\pi_j f))\Big| \\
&\qquad\quad\quad + \sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(R(j))| \\
&\le \sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\pi_0 f)| \\
&\qquad + \Big\{\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_\tau}(\Delta_\tau f))| + 2\sup_{f\in\mathcal{F}}|\mathbb{G}_n^{(2)}(\varphi^{\wedge}_{m_\tau}(\Delta_\tau f))| \\
&\qquad\qquad + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_\tau f\|_{1,n}\Big\} \\
&\qquad + \sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_j-m_{j+1}}(\pi_{j+1}f-\pi_j f))\Big| \\
&\qquad + \sum_{j=0}^{\tau-1}\Big\{\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1}f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad + 2\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(2)}(\min\big\{\big|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1}f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_{j+1}f\mathbb{1}_{\{\Delta_{j+1}f>m_{j+1}\}}\|_{1,n}\Big\} \\
&\qquad + \sum_{j=0}^{\tau-1}\Big\{\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad + 2\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(2)}(\min\big\{\big|\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f)\big|, 2m_j\big\})\Big| \\
&\qquad\qquad + 2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_j f\mathbb{1}_{\{\Delta_j f>m_j-m_{j+1}\}}\|_{1,n}\Big\}. \qquad (3.5.34)
\end{aligned}
$$

We have for $f\in\mathcal{F}(M_n)$,

$$
\begin{aligned}
\pi_0 f &= \varphi^{\wedge}_{2M_n}(\pi_0 f), \\
\varphi^{\wedge}_{m_\tau}(\Delta_\tau f) &\le \min\{\Delta_\tau f, 2m_\tau\}, \\
\varphi^{\wedge}_{m_j-m_{j-1}}(\pi_{j+1}f-\pi_j f) &\le \min\{\Delta_j f, 2m_j\}, \\
\min\{\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1}f), 2m_j\} &\le \min\{\Delta_j f, 2m_j\}, \\
\min\{\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f), 2m_j\} &\le \min\{\Delta_j f, 2m_j\}. \qquad (3.5.35)
\end{aligned}
$$

Let us define the event

$$\begin{aligned}\Omega_n &:= \{\sup_{f\in\mathcal{F}(M_n)} R_n(\varphi^{\wedge}_{2M_n}(\pi_0 f)) \le 2\sigma\psi(\sigma)\} \\ &\quad \cap \bigcap_{j=1}^{\tau}\big\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) \le 2\delta_j\psi(\delta_j)\big\}.\end{aligned}$$

From (3.5.34) and (3.5.35), we obtain

$$\begin{aligned}&\sup_{f\in\mathcal{F}(M_n)} |\mathbb{G}_n^{(1)}(f)|\mathbb{1}_{\Omega_n} \\ \le\ &\sup_{f\in\mathcal{F}(M_n)} |\mathbb{G}_n^{(1)}(\pi_0 f)|\mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\pi_0 f)\le 2\sigma\psi(\sigma)\}} \\ &+\Big\{\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_\tau}(\Delta_\tau f))| \\ &\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_\tau f, 2m_\tau\})\le 2\delta_\tau\psi(\delta_\tau)\}} + 2R_2\Big\} \\ &+\sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_j-m_{j+1}}(\pi_{j+1}f-\pi_j f))\Big| \\ &\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\})\le 2\delta_j\psi(\delta_j)\}} \\ &+\sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1}f)\big|, 2m_j\big\})\Big| \\ &\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\})\le 2\delta_j\psi(\delta_j)\}} + 2R_4 \\ &+\sum_{j=0}^{\tau-1}\sup_{f\in\mathcal{F}}\Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_j-m_{j+1}}(\Delta_j f)\big|, 2m_j\big\})\Big| \\ &\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\})\le 2\delta_j\psi(\delta_j)\}} + 2R_5 \\ =:\ &\tilde{R}_1 + \{\tilde{R}_2 + 2R_2\} + \tilde{R}_3 + \{\tilde{R}_4 + 2R_4\} + \{\tilde{R}_5 + 2R_5\}. \qquad (3.5.36)\end{aligned}$$

We now discuss the terms $\tilde{R}_i$, $i = 1, ..., 5$, separately. The terms R_i, $i \in \{2, 4, 5\}$ can be discussed similarly to the proof found in Theorem 2.4.4. Put

$$\tilde{C}_n := 2c(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n})$$

where c is a resulting constant from the bound in Theorem 2.4.1 or Lemma 2.8.8.

- Since $|\{\pi_0 f : f \in \mathcal{F}(M_n)\}| \le \mathbb{N}(\delta_0)$, $\|\pi_0 f\|_\infty \le M_n \le m(n, \delta_0, \mathbb{N}(\delta_1))$, we have by Lemma 3.5.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_1 &= \mathbb{E}\sup_{f\in\mathcal{F}(M_n)} |\mathbb{G}_n^{(1)}(\pi_0 f)|\mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\pi_0 f)\le 2\delta_0\psi(\delta_0)\}} \\ &\le \tilde{C}_n\psi(\delta_0)\delta_0\sqrt{1\vee\log\mathbb{N}(\delta_1)}.\end{aligned}$$

- It holds true that $|\{\varphi^{\wedge}_{m_\tau}(\Delta_\tau f) : f \in \mathcal{F}(M_n)\}| \leq N_\tau$. If $g := \varphi^{\wedge}_{m_\tau}(\Delta_\tau f)$, then $\|g\|_\infty \leq m_\tau \leq m(n, \delta_\tau, N_{\tau+1})$. We conclude by Lemma 3.5.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_2 &\leq \mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_\tau}(\Delta_\tau f))| \\ &\qquad\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_\tau f, 2m_\tau\}) \leq 2\delta_\tau \psi(\delta_\tau)\}} \\ &\leq \tilde{C}_n \psi(\delta_\tau)\delta_\tau \cdot \sqrt{1 \vee \log N_{\tau+1}}.\end{aligned}$$

- Since the partitions are nested, $|\{\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f) : f \in \mathcal{F}(M_n)\}| \leq N_{j+1}$. If $g := \varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f)$, we have $\|g\|_\infty \leq m_j - m_{j+1} \leq m_j \leq m(n, \delta_j, N_{j+1})$. We conclude by Lemma 3.5.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_3 &\leq \sum_{j=0}^{\tau-1} \mathbb{E}\sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\varphi^{\wedge}_{m_j - m_{j+1}}(\pi_{j+1} f - \pi_j f))\Big| \\ &\qquad\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) \leq 2\delta_j \psi(\delta_j)\}} \\ &\leq \tilde{C}_n \sum_{j=0}^{\tau-1} \psi(\delta_j)\delta_j \sqrt{1 \vee \log N_{j+1}}.\end{aligned}$$

- Note that $|\{\min\{\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1} f), 2m_j\} : f \in \mathcal{F}(M_n)\}| \leq N_{j+1}$. If we set $g := \min\{\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1} f), 2m_j\}$, we have $\|g\|_\infty \leq 2m_j = m(n, \delta_j, N_{j+1})$. We conclude by Lemma 3.5.5:

$$\begin{aligned}\mathbb{E}\tilde{R}_4 &\leq \sum_{j=0}^{\tau-1} \mathbb{E}\sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_{j+1}}(\Delta_{j+1} f)\big|, 2m_j\big\})\Big| \\ &\qquad\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) \leq 2\delta_j \psi(\delta_j)\}} \\ &\leq \tilde{C}_n \sum_{j=0}^{\tau-1} \psi(\delta_j)\delta_j \sqrt{1 \vee \log N_{j+1}}.\end{aligned}$$

- Furthermore, $|\{\min\{\varphi^{\vee}_{m_j - m_{j+1}}(\Delta_j f), 2m_j\} : f \in \mathcal{F}(M_n)\}| \leq N_{j+1}$. If $g := \min\{\varphi^{\vee}_{m_j - m_{j+1}}(\Delta_j f), 2m_j\}$ we have $\|g\|_\infty \leq 2m_j = m(n, \delta_j, N_{j+1})$. We conclude by Lemma 3.5.5 that:

$$\begin{aligned}\mathbb{E}\tilde{R}_5 &\leq \sum_{j=0}^{\tau-1} \mathbb{E}\sup_{f\in\mathcal{F}} \Big|\mathbb{G}_n^{(1)}(\min\big\{\big|\varphi^{\vee}_{m_j - m_{j+1}}(\Delta_j f)\big|, 2m_j\big\})\Big| \\ &\qquad\qquad \times \mathbb{1}_{\{\sup_{f\in\mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) \leq 2\delta_j \psi(\delta_j)\}} \\ &\leq \tilde{C}_n \sum_{j=0}^{\tau-1} \psi(\delta_j)\delta_j \cdot \sqrt{1 \vee \log N_{j+1}}.\end{aligned}$$

Inserting the bounds for $\mathbb{E}\tilde{R}_i$, $i = 1, ..., 5$, and the bounds for R_i, $i \in \{2, 4, 5\}$, from the proof of Theorem 2.4.4 into (3.5.36), we obtain for some universal constant $\tilde{c} > 0$,

$$\mathbb{E} \sup_{f\in\mathcal{F}(M_n)} \Big|\mathbb{G}_n^{(1)}(f)\Big| \mathbb{1}_{\Omega_n} \leq \tilde{c}\Big(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n} + \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\Big)\Big[\sum_{j=0}^{\tau+1} \psi(\delta_j)\delta_j \sqrt{1 \vee \log N_{j+1}} + I(\sigma)\Big]. \quad (3.5.37)$$

Note that

$$\sum_{j=k}^{\infty} \delta_j \psi(\delta_j) \leq 2 \sum_{j=k}^{\infty} \int_{\delta_{j+1}}^{\delta_j} \psi(\delta_j) dx \leq 2 \int_0^{\delta_k} \psi(x) dx.$$

By partial integration, it is easy to see that there exists some universal constant $c_\psi > 0$ such that

$$\Big| \int_0^{\delta_k} \psi(x) dx \Big| \leq c_\psi \delta_k \psi(\delta_k), \tag{3.5.38}$$

whence

$$\sum_{j=k}^{\infty} \delta_j \psi(\delta_j) \leq 2 c_\psi \delta_k \psi(\delta_k). \tag{3.5.39}$$

Using (3.5.39), we can argue as in the proof Theorem 2.4.4 (see equations (2.8.64), (2.8.65) and (2.8.66) therein) that there exists some universal constant $\tilde{c}_2 > 0$ such that

$$\sum_{j=0}^{\infty} \psi(\delta_j) \delta_j \sqrt{1 \vee \log N_{j+1}} \leq \tilde{c}_2 I(\sigma).$$

Insertion of the results into (3.5.37) yields

$$\mathbb{E} \sup_{f \in \mathcal{F}(M_n)} \big| \mathbb{G}_n^{(1)}(f) \big| \mathbb{1}_{\Omega_n} \leq \tilde{c} \cdot (3\tilde{c}_2 + 1)(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n} + \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}) I(\sigma). \tag{3.5.40}$$

Discussion of the event Ω_n: We have

$$\begin{aligned}
\mathbb{P}(\Omega_n^c) \quad &\leq \quad \mathbb{P}\Big(\sup_{f \in \mathcal{F}(M_n)} R_n(\varphi_{2M_n}^\wedge(\pi_0 f)) > 2\psi(\sigma)\sigma \Big) \\
&\qquad + \sum_{j=1}^{\tau+1} \mathbb{P}\Big(\sup_{f \in \mathcal{F}(M_n)} R_n(\min\{\Delta_j f, 2m_j\}) > 2\psi(\delta_j)\delta_j \Big) \\
&=: \quad R_1^\circ + R_2^\circ.
\end{aligned} \tag{3.5.41}$$

We now discuss R_i°, $i = 1, 2$. Put

$$C_n^\circ := 2c \Big\{ 1 + q^*\big(C_\Delta^{-1} C_\beta^{-2} \big) \big(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n} \big)^2 \Big\}$$

where c is from Lemma 3.5.5.

- Since $|\{\varphi_{2M_n}^\wedge(\pi_0 f) : f \in \mathcal{F}(M_n)\}| \leq \mathbb{N}(\delta_0) = \mathbb{N}(\sigma)$, $\|\varphi_{2M_n}^\wedge(\pi_0 f)\|_\infty \leq 2M_n \leq m(n, \sigma, \mathbb{N}(\sigma))$ and $V(\varphi_{2M_n}^\wedge(\pi_0 f)) \leq V(\pi_0 f) \leq \sigma$. We have by Lemma 3.5.5:

$$R_1^\circ \leq \frac{C_n^\circ}{\psi(\sigma)^2}.$$

- Note that $|\{\min\{\Delta_j f, 2m_j\} : f \in \mathcal{F}(M_n)\}| \le N_{j+1}$. So, $\|\min\{\Delta_j f, 2m_j\}\|_\infty \le 2m_j = m(n, \delta_j, N_{j+1})$ and $V(\min\{\Delta_j f, 2m_j\}) \le V(\Delta_j f) \le \delta_j$. We conclude by Lemma 3.5.5:
$$R_3^\circ \le C_n^\circ \sum_{j=0}^{\tau+1} \frac{1}{\psi(\delta_j)^2}.$$

Inserting the bounds for R_i°, $i = 1, 2$, into (3.5.41) yields

$$\mathbb{P}(\Omega_n^c) \le 2C_n^\circ \sum_{j=0}^{\infty} \frac{1}{\psi(\delta_j)^2}. \tag{3.5.42}$$

We now have

$$\sum_{j=0}^{\infty} \frac{1}{\psi(\delta_j)^2} \le 2\int_0^\sigma \frac{1}{\varepsilon\psi(\varepsilon)^2} d\varepsilon = \frac{2}{\log(\log(\sigma))}.$$

We conclude that for each $\eta > 0$,

$$\begin{aligned} \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)| > \eta\Big) &\le \mathbb{P}\Big(\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)| > \eta, \Omega_n\Big) + \mathbb{P}(\Omega_n^c) \\ &\le \frac{1}{\eta}\mathbb{E}\sup_{f\in\mathcal{F}} |\mathbb{G}_n^{(1)}(f)| \mathbb{1}_{\Omega_n} + \mathbb{P}(\Omega_n^c). \end{aligned}$$

Insertion of (3.5.32), (3.5.40) and (3.5.42) gives the result. □

Proof of Corollary 3.3.6. We will follow the proof of Corollary 2.4.5. Define $\tilde{\mathcal{F}}_: = \{f - g : f, g \in \mathcal{F}\}$. We obtain

$$\begin{aligned} &\mathbb{P}\Big(\sup_{V(f-g)\le\sigma,\, f,g\in\mathcal{F}} |\mathbb{G}_n(f) - \mathbb{G}_n(g)| \ge \eta\Big) \\ \le\ &\mathbb{P}\Big(\sup_{V(\tilde{f})\le\sigma,\, \tilde{f}\in\tilde{\mathcal{F}}} |\mathbb{G}_n^{(1)}(\tilde{f})| \ge \frac{\eta}{2}\Big) + \mathbb{P}\Big(\sup_{V(\tilde{f})\le\sigma,\, \tilde{f}\in\tilde{\mathcal{F}}} |\mathbb{G}_n^{(2)}(\tilde{f})| \ge \frac{\eta}{2}\Big). \end{aligned} \tag{3.5.43}$$

Now let $F(z, u) := 2D_n^\infty(u) \cdot \bar{F}(z, u)$, where $\bar{F}$ is from Assumption 3.2.1. Then obviously, F is an envelope function of $\tilde{\mathcal{F}}$.

We now discuss the second summand on the right hand side in (3.5.43). By Markov's inequality and Theorem 2.4.4 applied to $W_i(f) = \mathbb{E}[f(Z_i, \frac{i}{n})|Z_{i-1}]$, we obtain as in the proof of Corollary 2.4.5 that

$$\begin{aligned} &\mathbb{P}\Big(\sup_{V(\tilde{f})\le\sigma,\, \tilde{f}\in\tilde{\mathcal{F}}} |\mathbb{G}_n^{(2)}(\tilde{f})| \ge \frac{\eta}{2}\Big) \\ \le\ &\frac{\tilde{c}}{(\eta/2)}\Big[2\sqrt{2}\big(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n} + \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\big)\int_0^{\sigma/2} \sqrt{1 \vee \mathbb{H}(u, \mathcal{F}, V)}du \\ &\quad + \frac{4\sqrt{1 \vee \mathbb{H}(\frac{\sigma}{2})}}{r(\frac{\sigma}{\mathbb{D}_n})}\big\|F^2 \mathbb{1}_{\{F > \frac{1}{4}n^{1/2}\frac{r(\sigma)}{\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}\}}\big\|_{1,n}\Big]. \end{aligned} \tag{3.5.44}$$

The first summand in (3.5.44) converges to 0 for $\sigma \to 0$ (uniformly in n) since

$$\sup_{n\in\mathbb{N}} \int_0^{\sigma/2} \sqrt{1 \vee \mathbb{H}(u, \mathcal{F}, V)} du \le \sup_{n\in\mathbb{N}} \int_0^{\sigma} \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V)} d\varepsilon < \infty.$$

We now discuss the second summand in (3.5.44). The continuity conditions from Assumption 3.2.1 on $\bar{F}$ yield, as in the proof of Lemma 3.5.6(ii), that for all $u, u_1, u_2, v_1, v_2 \in [0, 1]$,

$$\|\bar{F}(Z_i, u) - \bar{F}(\tilde{Z}_i(\frac{i}{n}), u)\|_2 \le C_{cont} \cdot n^{-\alpha s/2}, \qquad (3.5.45)$$

$$\|\bar{F}(Z_i(v_1), u_1) - \bar{F}(\tilde{Z}_i(v_2), v_2)\|_2 \le C_{cont} \cdot \big(|v_1 - v_2|^{\alpha s/2} + |u_1 - u_2|^{\alpha s}\big). \qquad (3.5.46)$$

In the same manner of Corollary 2.4.5, we now obtain with (3.5.45) and (3.5.46) that

$$\big\|F^2 \mathbb{1}_{\{F > \frac{1}{4} n^{1/2} \frac{r(\sigma)}{\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}\}}\big\|_{1,n} \to 0 \qquad (3.5.47)$$

for $n \to \infty$ (this is obvious if Z_i is stationary, i.e. the first part of Assumption 3.2.1 is fulfilled), which shows that (3.5.44) converges to 0 for $\sigma \to 0$, $n \to \infty$.

We now consider the first term in (3.5.43). By Theorem 3.3.4, we have for some universal constant $c > 0$ that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{V(\tilde{f})\le\sigma,\, \tilde{f}\in\tilde{\mathcal{F}}} |\mathbb{G}_n^{(1)}(\tilde{f})| \ge \frac{\eta}{2}\Big)\\
\le\;& \frac{2}{\eta}\Big[c\Big(1 + \frac{\mathbb{D}_n^\infty}{\mathbb{D}_n} + \frac{\mathbb{D}_n}{\mathbb{D}_n^\infty}\Big) \cdot \int_0^\sigma \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}\big(\varepsilon, \tilde{\mathcal{F}}, V\big)}\, \mathrm{d}\varepsilon\\
&\quad + \frac{4\sqrt{1\vee\mathbb{H}(\frac{\sigma}{2})}}{r(\frac{\sigma}{\mathbb{D}_n})}\big\|F^2 \mathbb{1}_{\{F > \frac{1}{4} m(n,\sigma,\mathbb{N}(\frac{\sigma}{2}))\}}\big\|_1\Big]\\
&\quad\quad + c\Big(1 + q^*\big(C_\Delta^{-1} C_\beta^{-2}\big)\Big(\frac{\mathbb{D}_n^\infty}{\mathbb{D}_n}\Big)^2\Big)\int_0^\sigma \frac{1}{\varepsilon\psi(\varepsilon)^2} d\varepsilon. \qquad (3.5.48)
\end{aligned}$$

For the first summand in (3.5.48),

$$\begin{aligned}
&\int_0^\sigma \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \tilde{\mathcal{F}}, V)} d\varepsilon\\
\le\;& 2\sqrt{2}\int_0^{\sigma/2} \psi(2\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V)} d\varepsilon \le 2\sqrt{2}\int_0^{\sigma/2} \psi(\varepsilon)\sqrt{1 \vee \mathbb{H}(\varepsilon, \mathcal{F}, V)} d\varepsilon.
\end{aligned}$$

Equation (2.8.67) together with (3.3.5) and the uniform boundedness of $\mathbb{D}_n, \mathbb{D}_n^\infty$ show that the first summand in (3.5.48) converges to 0 for $\sigma \to 0$ (uniformly in n).

The third summand in (3.5.48) converges to 0 for $\sigma \to 0$ (uniformly in n) since $\int_0^\infty \varepsilon\psi(\varepsilon)^2 d\varepsilon < \infty$ and by the uniform boundedness of $\mathbb{D}_n, \mathbb{D}_n^\infty$.

The second summand in (3.5.48) converges to 0 for $n \to \infty$ by (3.5.47). □

3.5.3 Proofs of Section 3.2

Lemma 3.5.6. *Let $\mathcal{F}$ satisfy Assumption 2.3.1, 3.2.2. Suppose that Assumptions 3.1.1, 3.2.1 hold true. Then there exist constants $C_{cont} > 0, C_{\bar{f}} > 0$ such that for any $f \in \mathcal{F}$,*

(i) for any $j \geq 1$,

$$\begin{aligned} \|P_{i-j} f(Z_i, u)\|_2 &\leq D_{f,n}(u)\Delta(j), \\ \sup_{i=1,\dots,n} \|f(Z_i, u)\|_2 &\leq C_\Delta \cdot D_{f,n}(u), \end{aligned}$$
$$\sup_{i,u} \|\bar{f}(Z_i, u)\|_2 \leq C_{\bar{f}}, \qquad \sup_{v,u} \|\bar{f}(\tilde{Z}_0(v), u)\|_2 \leq C_{\bar{f}},$$

(ii) with $x = \frac{1}{2}$,

$$\|\bar{f}(Z_i, u) - \bar{f}(\tilde{Z}_i(\frac{i}{n}), u)\|_2 \leq \tilde{C} n^{-\varsigma s x}, \tag{3.5.49}$$
$$\|\bar{f}(\tilde{Z}_i(v_1), u_1) - \bar{f}(\tilde{Z}_i(v_2), u_2)\|_2 \leq \tilde{C}\big(|v_1 - v_2|^{\varsigma s x} + |u_1 - u_2|^{\varsigma s}\big). \tag{3.5.50}$$

Proof of Lemma 3.5.6. (i) If Assumption 3.1.1 is satisfied, we have by Lemma 3.5.1,

$$\begin{aligned} \|P_{i-j} f(Z_i, u)\|_2 &= \|P_{i-j}\mathbb{E}[f(Z_i, u)|\mathcal{A}_{i-1}]\|_2 \\ &\leq \|\mathbb{E}[f(Z_i, u)|\mathcal{A}_{i-1}] - \mathbb{E}[f(Z_i, u)|\mathcal{A}_{i-1}]^{*(i-j)}\|_2 \leq D_{f,n}(u)\Delta(j). \end{aligned}$$

The second assertion follows from Lemma 3.5.1.

(ii) Let $\bar{C}_R := \sup_{v,u} \|\bar{R}(\tilde{Z}_0(v), u)\|_2$ and

$$C_R := \max\{\sup_{i,u} \|R(Z_i, u)\|_2, \sup_{u,v} \|R(\tilde{Z}_0(v), u)\|_2\}.$$

We first use Assumption 3.2.2 and Hölder's inequality to obtain

$$\begin{aligned} &\|\bar{f}(\tilde{Z}_i(v), u_1) - \bar{f}(\tilde{Z}_i(v), u_2)\|_2 \qquad (3.5.51) \\ \leq\ & |u_1 - u_2|^\varsigma \cdot \big(\|\bar{R}(\tilde{Z}_i(v), u_1)\|_2 + \|R(\tilde{Z}_i(v), u_2)\|_2\big) \\ \leq\ & 2\bar{C}_R |u_1 - u_2|^\varsigma. \qquad (3.5.52) \end{aligned}$$

Assume w.l.o.g. that

$$\sup_{u,v} \frac{1}{c^s}\mathbb{E}\Big[\sup_{|a|_{L_{\mathcal{F}},s} \leq c} \big|\bar{f}(\tilde{Z}_0(v), u) - \bar{f}(\tilde{Z}_0(v) + a, u)\big|^2\Big] \leq C_R.$$

(which is obvious if Z_i is stationary, i.e. the first part of Assumption 3.2.1 is fulfilled; in this case $Z_i = \tilde{Z}_i(v)$ for all v). Let $c_n > 0$ be some sequence. Let

$C_{\bar{f}} := \max\{\sup_{i,u} \|f(Z_i,u)\|_{2\bar{p}}, \sup_{u,v} \|f(\tilde{Z}_0(v),u)\|_{2\bar{p}}\}$. Then we have by Jensen's inequality,

$$\begin{aligned}
&\big\|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)\big\|_2 \\
\leq\ & \mathbb{E}\Big[\big|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)\big|^2 \mathbb{1}_{\{|Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F}},s} \leq c_n\}}\Big]^{1/2} \\
&\quad + \mathbb{E}\Big[(\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)^2 \mathbb{1}_{\{|Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F}},s} > c_n\}}\Big]^{1/2} \\
\leq\ & \mathbb{E}\Big[\sup_{|a|_{L_{\mathcal{F}},s} \leq c_n} \big|\bar{f}(\tilde{Z}_i(v),u) - \bar{f}(\tilde{Z}_i(v)+a,u)\big|^2\Big]^{1/2} \\
&\quad + \{\big\|\bar{f}(Z_i,u)\big\|_{2\bar{p}} + \bar{f}(\tilde{Z}_i(v),u)\big\|_{2\bar{p}}\}\mathbb{P}(|Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F}},s} > c_n)^{\frac{\bar{p}-1}{2\bar{p}}} \\
\leq\ & C_R c_n^s + 2C_{\bar{f}}\Big(\frac{\||Z_i - \tilde{Z}_i(v)|_{L_{\mathcal{F}},s}\|_{\frac{2\bar{p}s}{\bar{p}-1}}}{c_n}\Big)^s \\
\leq\ & C_R c_n^s + 2C_{\bar{f}} C_X (|L_{\mathcal{F}}|_1 + \sum_{j=0}^{\infty} L_{\mathcal{F},j} j^{\varsigma s}) \cdot \frac{\{|v - \frac{i}{n}|^{\varsigma s} + n^{-\varsigma s}\}}{c_n^s}.
\end{aligned}$$

We obtain with $\tilde{C} := C_R + 2C_{\bar{f}} C_X (|L_{\mathcal{F}}|_1 + \sum_{j=0}^{\infty} L_{\mathcal{F},j} j^{\varsigma s})$ that

$$\|\bar{f}(Z_i,u) - \bar{f}(\tilde{Z}_i(v),u)\|_2 \leq \tilde{C} \cdot \Big[c_n^s + \frac{|v - \frac{i}{n}|^{\varsigma s} + n^{-\varsigma s}}{c_n^s}\Big]. \tag{3.5.53}$$

Furthermore, as above, for any $c > 0$,

$$\begin{aligned}
\|f(\tilde{Z}_i(v_1),u) - f(\tilde{Z}_i(v_2),u)\|_2 \ &\leq\ C_R c^s + 2C_{\bar{f}}\Big(\frac{\||\tilde{Z}_0(v_1) - \tilde{Z}_0(v_2)|_{L_{\mathcal{F}},s}^s\|_{\frac{2\bar{p}}{\bar{p}-1}}}{c}\Big)^s \\
&\leq\ C_R c^s + 2C_{\bar{f}} C_X |L_{\mathcal{F}}|_1 \cdot \frac{|v_1 - v_2|^{\varsigma s}}{c^s}. \qquad (3.5.54)
\end{aligned}$$

From (3.5.53), we obtain the first assertion with $v = \frac{i}{n}$. The second assertion follows from (3.5.54) and (3.5.52).

□

3.5.4 Details of Section 3.3.3

We first show that the supremum over $x \in \mathbb{R}$, $v \in [0,1]$ can be approximated by a supremum over grids $x \in \mathcal{X}_n$, $v \in \mathcal{V}_n$.

For some $Q > 0$, put $c_n = Qn^{\frac{1}{2s}}$. Define the event $A_n = \{\sup_{i=1,\dots,n} |X_i| \leq c_n\}$. Then by Markov's inequality,

$$\mathbb{P}(A_n^c) \leq n \cdot \frac{\|X_i\|_{2s}^{2s}}{Q^{2s} c_n^{2s}} \leq \frac{C_X^{2s} n}{c_n^{2s}} \tag{3.5.55}$$

is arbitrarily small for Q large enough.

Put $\hat{g}^{\circ}_{n,h}(x,v) := \frac{1}{n}\sum_{i=1}^{n} K_{h_1}(i/n - v)K_{h_2}(X_i - x)\mathbb{1}_{\{|X_i|\le c_n\}}$. Then on A_n,

$$\hat{g}^{\circ}_{n,h}(\cdot) = \hat{g}_{n,h}(\cdot). \tag{3.5.56}$$

Furthermore,

$$\begin{aligned}
\sqrt{nh_1h_2}\big|\mathbb{E}\hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h}(x,v)\big| &\le \frac{\sqrt{nh_1h_2}|K|_\infty}{nh_1}\sum_{i=1}^{n}\mathbb{E}[K_{h_2}(X_i - x)\mathbb{1}_{\{|X_i|>c_n\}}]\\
&\le \frac{\sqrt{nh_1h_2}}{h_1h_2}|K|_\infty c_n^{-2s}\sup_i \mathbb{E}[K(\frac{X_i - x}{h_2})|X_i|^{2s}]\\
&\le Q^{-2s}(nh_1h_2)^{-1/2}|K|_\infty^2 C_X^{2s} = o(1). \qquad (3.5.57)
\end{aligned}$$

For $|x| > 2c_n$ we have $K_{h_2}(X_i - x)\mathbb{1}_{\{|X_i|\le c_n\}} \le h^{-1}(\frac{c_n}{h})^{-p_K} = h^{p_K-1}c_n^{-p_K}$ and thus

$$\sqrt{nh}|\hat{g}^{\circ}_{n,h}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h}(x,v)| \le \frac{2|K|_\infty C_K}{h_1^{1/2}}(nh_2)^{1/2}h_2^{p_K-1}c_n^{-p_K} \le \frac{h_2^{p_K}}{Q^{p_K}(nh_1h_2)^{1/2}} = o(1). \tag{3.5.58}$$

By (3.5.56), (3.5.57) and (3.5.58) we have on A_n,

$$\begin{aligned}
&\sqrt{nh_1h_2}\sup_{x\in\mathbb{R},v\in[0,1]}|\hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}_{n,h}(x,v)|\\
&= \sqrt{nh_1h_2}\sup_{x\in\mathbb{R},v\in[0,1]}|\hat{g}^{\circ}_{n,h}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h}(x,v)| + o_p(1)\\
&= \sqrt{nh_1h_2}\sup_{|x|\le 2c_n,v\in[0,1]}|\hat{g}^{\circ}_{n,h}(x,v) - \mathbb{E}\hat{g}^{\circ}_{n,h}(x,v)| + o_p(1)\\
&= \sqrt{nh_1h_2}\sup_{|x|\le 2c_n,v\in[0,1]}|\hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}_{n,h}(x,v)| + o_p(1). \qquad (3.5.59)
\end{aligned}$$

Let $\mathcal{X}_n = \{in^{-3} : i \in \{-2\lceil c_n\rceil n^3, ..., 2\lceil c_n\rceil n^3\}\}$ be a grid that approximates each $x \in [-2c_n, 2c_n]$ with precision n^{-3}, and $\mathcal{V}_n = \{in^{-3} : i = 1, ..., n^3\}$. Since K is Lipschitz continuous with constant L_K,

$$\begin{aligned}
&\sqrt{nh_1h_2}\sup_{|x-x'|\le n^{-3},|v-v'|\le n^{-3}}\big|\big(\hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}_{n,h}(x,v)\big)\\
&\qquad\qquad - \big(\hat{g}_{n,h}(x',v) - \mathbb{E}\hat{g}_{n,h}(x',v)\big)\big|\\
&\le 2\frac{\sqrt{n}}{\sqrt{h_1h_2}}\sup_{|x-x'|\le n^{-3},|v-v'|\le n^{-3}}\Big[\frac{L_K|K|_\infty|x-x'|}{h_2} + \frac{L_K|K|_\infty|v-v'|}{h_1}\Big]\\
&= O(n^{-1}). \qquad (3.5.60)
\end{aligned}$$

We conclude from (3.5.55), (3.5.59) and (3.5.60) that

$$\begin{aligned}
&\sqrt{nh_1h_2}\sup_{x\in\mathbb{R},v\in[0,1]}|\hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}_{n,h}(x,v)|\\
&= \sqrt{nh_1h_2}\sup_{x\in\mathcal{X}_n,v\in\mathcal{V}_n}|\hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}_{n,h}(x,v)| + o_p(1). \qquad (3.5.61)
\end{aligned}$$

It was already shown that Assumption 3.1.1 is satisfied. Furthermore, we can choose $\mathbb{D}_n = |K|_\infty$, $\mathbb{D}^\infty_{\nu_2,n} = \frac{|K|_\infty}{\sqrt{h_1}}$ with $\nu_2 = \infty$, and $\bar{F}(z,u) = \sup_{f\in\mathcal{F}} \bar{f}(z,u) \leq \frac{|K|_\infty}{\sqrt{h_2}} =: C_{\bar{F},n}$. Note that

$$\begin{aligned}
\mathbb{E}[(\sqrt{h_2}K_{h_2}(X_i - x))^2] &= \mathbb{E}\big[\mathbb{E}[(\sqrt{h_2}K_{h_2}(X_i - x))^2|X_{i-1}]\big] \\
&= \int \Big(\int K(w)^\kappa f_{X_i|X_{i-1}=z}(x + wh_2)dw \Big)^{1/\kappa} d\mathbb{P}^{X_{i-1}}(z) \\
&\leq C_\infty \cdot (\int K(w)^2 dw)^{1/2}.
\end{aligned}$$

Therefore,

$$\|f_{x,v}\|_{2,n} \leq \mathbb{D}_n C_\infty \int K(w)^2 dw,$$

which implies $\sigma := \sup_{n\in\mathbb{N}} \sup_{f\in\mathcal{F}} V_n(f) < \infty$. Due to $\Delta(k) = O(k^{-\alpha s})$, the last condition in (3.3.3) is fulfilled if

$$\sup_{n\in\mathbb{N}} \frac{\log(n)}{nh_2 h_1^{\frac{\alpha s}{\alpha s - 1}}} < \infty.$$

By Corollary 3.3.2, we have

$$\sqrt{nh_1h_2} \sup_{x\in\mathcal{X}_n, v\in\mathcal{V}_n} \big|\hat{g}_{n,h}(x) - \mathbb{E}\hat{g}_{n,h}(x,v)\big| = \sup_{f\in\mathcal{F}} |\mathbb{G}_n(f)| = O_p(\sqrt{\log|\mathcal{F}|}) = O(\sqrt{\log(n)}).$$

Equation (3.5.61) yields

$$\sqrt{nh_1h_2} \sup_{x\in\mathbb{R}, v\in[0,1]} |\hat{g}_{n,h}(x,v) - \mathbb{E}\hat{g}_{n,h}(x,v)| = O_p\big(\sqrt{\log(n)}\big).$$

Chapter 4

Oracle inequalities for dependent data with applications to neural networks

In this chapter we want to highlight our theoretical findings from the previous chapters by giving an application which has been on the rise in popularity, but for which, yet, little profound knowledge is available – we are talking about neural networks. Their use cases are now a crucial part of modern technology. This black box has successfully been applied to many branches of science and relevant sectors in the industry. Only in the last couple of years have we come to understand parts of their mathematical workings. As mentioned in Chapter 1, Schmidt-Hieber [2017] introduces a theory based on independent and identically distributed observations for sparse neural networks. We will now generalize the approach and develop a theory for absolutely regular mixing sequences as well as Bernoulli shift processes under the functional dependence measure, quantifying a neural network estimator's performance by deriving convergence rates. Before that, we will first have to establish a proper mathematical background, beginning with oracle inequalities.

4.1 Oracle inequalities in the regression model under dependence

As motivated in the Introduction, Chapter 1, we observe a d-dimensional realization X_i, $i = 1, ..., n$, of a stationary stochastic process which follows the recurrence

$$X_i = f_0(X_{i-1}, ..., X_{i-r}) + \varepsilon_i, \quad i = r+1, ..., n, \tag{4.1.1}$$

where ε_i is an i.i.d. sequence of d-dimensional random variables, $r \in \mathbb{N}$ is the number of lags considered and $f_0 : \mathbb{R}^{dr} \to \mathbb{R}^d$ an unknown function. We abbreviate our notation by setting $\mathbb{X}_{i-1} := (X_{i-1}, ..., X_{i-r})$, suppressing the dependence on r in our notation.

We formalize the Subgaussianity assumption as follows.

Assumption 4.1.1. *The random variable ε_1 is Subgaussian, that is, for any $k \in \mathbb{N}$ and any component $j \in \{1, ..., d\}$,*

$$\mathbb{E}[|\varepsilon_{1j}|^k]^{1/k} \le C_\varepsilon \cdot \sqrt{k}.$$

Given some weight function $\mathcal{W} : \mathbb{R}^{dr} \to \mathbb{R}$ with compact support $\subset [0, 1]^{dr}$, the prediction error (or simply risk) of some function $f : \mathbb{R}^{dr} \to \mathbb{R}^d$ is defined by

$$R(f) := \frac{1}{d}\mathbb{E}[|X_{r+1} - f(\mathbb{X}_r)|_2^2 \mathcal{W}(\mathbb{X}_r)]. \tag{4.1.2}$$

Its empirical counterpart is

$$\hat{R}_n(f) = \frac{1}{n}\sum_{i=r+1}^{n} \frac{1}{d}\big|X_i - f(\mathbb{X}_{i-1})\big|_2^2 \mathcal{W}(\mathbb{X}_{i-1}). \tag{4.1.3}$$

We define by $\mathbb{E}R(\hat{f}) - R(f_0)$ the *excess Bayes risk* of the empirical risk minimizer

$$\hat{f} \in \arg\min_{f \in \mathcal{F}} \hat{R}_n(f)$$

over a class $\mathcal{F} \subset \{f : \mathbb{R}^{dr} \to \mathbb{R}^d \text{ measurable}\}$.

Let $\mathbb{N}(\delta, \mathcal{F}, \|\cdot\|_\infty)$ denote the smallest number of δ-brackets with respect to $\|f\|_\infty := \sup_{j \in \{1,...,d\}} \|f_j\|_\infty$ which is needed to cover a function class $\mathcal{F}$ of measurable functions, and let $\mathbb{H}(\delta) := \log \mathbb{N}(\delta, \mathcal{F}, \|\cdot\|_\infty)$ denote the corresponding bracketing entropy. We assume further that $\mathcal{F} \subset \{f : \mathbb{R}^{dr} \to \mathbb{R}^d \text{ measurable}\}$ satisfies

$$\sup_{f \in \mathcal{F}} \sup_{j \in \{1,...,d\}} \sup_{x \in \text{supp}(\mathcal{W})} |f_j(x)| \le F$$

for some constant $F > 0$

We now establish oracle inequalities under two different measures of dependence on X_i, $i = 1, ..., n$, namely absolute regularity and functional dependence. We shortly revise these concepts.

4.1.1 An oracle inequality under absolute regularity

Let $\beta^{mix}(k)$, $k \in \mathbb{N}_0$, denote the absolutely regular mixing coefficients of X_i, that is,

$$\beta^{mix}(k) := \beta^{mix}(\sigma(X_i : i \le 0), \sigma(X_i : i \ge k)), \tag{4.1.4}$$

where for two σ-algebras $\mathcal{U}, \mathcal{V}$ over some probability space Ω,

$$2\beta^{mix}(\mathcal{U}, \mathcal{V}) := \sup \sum_{(i,j) \in I \times J} |\mathbb{P}(U_i \cap V_i) - \mathbb{P}(U_i)\mathbb{P}(V_i)|$$

and the supremum is taken over all finite partitions $(U_i)_{i \in I}$, $(V_j)_{j \in J}$ of Ω such that $(U_i)_{i \in I} \subset \mathcal{U}$, $(V_j)_{j \in J} \subset \mathcal{V}$. Illustratively, $\beta^{mix}(k)$, $k \in \mathbb{N}_0$, measures the dependence between $\sigma(X_i : i \leq 0)$ and $\sigma(X_i : i \geq k)$, and decays to 0 for $k \to \infty$ if $\sigma(X_i : i \leq 0)$ contains no information about X_k for large k. We refer to [Rio, 2013, Section 1.3] or Doukhan et al. [1995] for a more detailed introduction. There are several results available which state that linear processes, GARCH or ARMA processes have absolutely summable $\beta^{mix}(k)$, cf. Bradley [2005], Fryzlewicz and Subba Rao [2011] or Doukhan [1994].

Based on $\beta^{mix}(\cdot)$, we define a new quantity $\Lambda^{mix}(\cdot)$ in the following assumption which then appears in the oracle inequality.

Assumption 4.1.2 (Compatibility assumptions)**.** *Let X_i have β-mixing coefficients $\beta^{mix}(k)$, $k \in \mathbb{N}_0$, which are submultiplicative, that is, there exists a constant $C_{\beta,sub} > 0$ such that for any $q_1, q_2 \in \mathbb{N}$,*

$$\beta^{mix}(q_1 q_2) \leq C_{\beta,sub}\beta^{mix}(q_1)\beta^{mix}(q_2). \tag{4.1.5}$$

Let $\phi : [0,\infty) \to [0,\infty)$ be a function which satisfies

(i) $\phi(0) = 0$, ϕ is convex and differentiable with $c_0 := \sup_{y\in\mathbb{R}} \frac{\phi'(y)y}{\phi(y)} < \infty$,

(ii) $(0,\infty) \to (0,\infty), y \mapsto \frac{y}{\phi(y)}$ is convex and decreasing,

(iii) $\sum_{k=0}^{\infty}(\phi^*(k+1) - \phi^*(k))\beta^{mix}(k) < \infty$ for the convex conjugate ϕ^* of ϕ, that is $\phi^*(x) = \sup_{t>0}\{xt - \phi(t)\}$.

We refer to Rockafellar [2015] for an introduction to convex analysis and convex conjugate functions. Based on ϕ, we define

$$\psi(x) := \phi^*(x)x, \qquad \Lambda^{mix}(x) := \lceil \psi^{-1}(x^{-1})\rceil x. \tag{4.1.6}$$

The proof of the following theorem is given in Section 4.6.1, restated there as Theorem 4.6.4.

Theorem 4.1.3. *Let Assumption 4.1.1 and 4.1.2 hold true and let $\Lambda^{mix}(\cdot)$ be the function defined in (4.1.6). Then, for any $\delta \in (0,1), \eta > 0$ there exists a constant $\mathbb{C} = \mathbb{C}(\eta, c_0, r, C_{\beta,sub}, C_\varepsilon, F)$ such that*

$$\mathbb{E}R(\hat{f}) - R(f_0) \leq (1+\eta)^2 \inf_{f\in\mathcal{F}}\{R(f) - R(f_0)\} + \mathbb{C}\cdot\Big\{\Lambda^{mix}\Big(\frac{\mathbb{H}(\delta)}{n}\Big) + \delta\Big\}.$$

In the special case of polynomial decay and exponential decay of $\beta^{mix}(\cdot)$, explicit representations of $\Lambda^{mix}(\cdot)$ are available via the following lemma.

Lemma 4.1.4. *Depending on specific decay rates, the following statements hold true.*

(i) Suppose that $\sum_{k=0}^{\infty} k^{\alpha-1}\beta^{mix}(k) < \infty$ for some $\alpha > 1$. Then Assumption 4.1.2 is fulfilled with $\phi(x) = x^{\frac{\alpha}{\alpha-1}}$ and

$$\Lambda^{mix}(x) \leq c_\alpha \cdot (x^{\frac{\alpha}{\alpha+1}} \vee x),$$

where $c_\alpha > 0$ is some constant only depending on α.

(ii) *Suppose that* $\beta^{mix}(k) \le \kappa\rho^k$ *for some* $\kappa > 0$, $\rho \in (0,1)$. *Then Assumption 4.1.2 is fulfilled with* $\phi(x) = x^{\frac{\log(x+1)}{\log(a)}}$ ($a = \frac{\rho+1}{2\rho}$) *and*

$$\Lambda^{mix}(x) \le c_\rho \cdot (1 \vee \log(x^{-1}))x,$$

where $c_\rho > 0$ *is some constant only depending on* ρ.

4.1.2 Oracle inequalities under functional dependence

We assume that $X_i = (X_{ij})_{j=1,...,d}$, $i = 1,...,n$, is a (stationary) Bernoulli shift process according to (1.2.1), that is,

$$X_i = J(\mathcal{A}_i). \tag{4.1.7}$$

The functional dependence measure of X_i, $i \in \mathbb{Z}$, for $q > 0$ is given by

$$\delta_q^X(k) = \sup_{j=1,\dots,d} \big\| X_{ij} - X_{ij}^{*(i-k)} \big\|_q. \tag{4.1.8}$$

In contrast to the case of absolutely regular mixing coefficients, the functional dependence measure in (4.1.8) requires the process X_i to have at least a q-th moment. To transfer the dependence structure from X_i to $g(X_i)$ for some function g, we have to impose smoothness assumptions on g, as it was done in Chapter 2, which also affect the dependence coefficients $\delta_q^{g(X)}$. Suppose that there exist $\varsigma, K, L_{\mathcal{F}} > 0$ such that for all $f \in \mathcal{F}$, $x, x' \in \mathbb{R}^{dr}$,

$$\begin{aligned} |\mathcal{W}(x) - \mathcal{W}(x')| &\le \frac{1}{\varsigma}|x - x'|_\infty, \\ |f_0(x) - f_0(x')|_\infty &\le K|x - x'|_\infty, \\ |f(x) - f(x')|_\infty &\le L_{\mathcal{F}}|x - x'|_\infty. \end{aligned}$$

That is, we assume a regularity condition (namely, Lipschitz continuity) on the weight and on the (true) regression function.

Assumption 4.1.5. *Let* X_i *be of the form (4.1.7). Given* $\mathbb{L} > 0$, *let* $\Delta(k)$, $k \in \mathbb{N}_0$, *be a decreasing sequence of real numbers such that for some* $\theta \in (0,1]$,

$$\mathbb{L} \cdot \sup_{l=1,\dots,r} \delta_{2\theta}^X(k-l)^\theta \le \Delta(k). \tag{4.1.9}$$

The parameter $\theta \in (0,1]$ in Assumption 4.1.5 can be chosen arbitrarily and regulates the number of moments which have to be imposed on X_i. A small θ coincides with a slower decay rate of $\Delta(k)$ due to the exponent θ in (4.1.9).

Based on $\Delta(\cdot)$, we define a new quantity $\Lambda^{dep}(\cdot)$ which will appear in our oracle inequality. For $x \in [0,\infty)$, let

$$\tilde{V}(x) = x^{1/2} + \sum_{j=0}^{\infty} \min\{x^{1/2}, \Delta(j)\}. \tag{4.1.10}$$

We have seen equations (4.1.9) and (4.1.10) in Chapter 2, already. Especially equation (4.1.9) is in line with Assumption 2.4.2 which is required for compatibility reasons when employing maximal inequalities. Let $\bar{y}(x) \in [0, \infty)$ be such that

$$\tilde{V}(\sqrt{x}\bar{y}(x)) \leq \bar{y}(x) \tag{4.1.11}$$

and put

$$\Lambda^{dep}(x) := \sqrt{x}\bar{y}(x). \tag{4.1.12}$$

We obtain the following oracle inequality under functional dependence (which is proven in Theorem 4.6.16, a restated version, of Section 4.6, Subsection 4.6.5).

Theorem 4.1.6. *Suppose that Assumption 4.1.1 and Assumption 4.1.5 hold true with* $\mathbb{L} = 2dr\big(\frac{2}{\varsigma} + \frac{(L_{\mathcal{F}}+K)}{F}\big)$*. Let* $\Lambda^{dep}(\cdot)$ *be the function defined in (4.1.12). Then, for any* $\delta \in (0,1), \eta > 0$ *there exists a constant* $\mathbb{C} = \mathbb{C}(\eta, C_\varepsilon, F)$ *such that*

$$\mathbb{E}R(\hat{f}) - R(f_0) \leq (1+\eta)^2 \inf_{f \in \mathcal{F}} \{R(f) - R(f_0)\} + \mathbb{C} \cdot \Big\{\Lambda^{dep}\Big(\frac{\mathbb{H}(\delta)}{n}\Big) + \delta\Big\}.$$

Remark 4.1.7. (i) While in Theorem 4.1.3 the parameter r is directly contained in the constant $\mathbb{C}$, in Theorem 4.1.6 it is contained in $\Lambda^{dep}(\cdot)$ via Assumption 4.1.5. Additionally, in the latter theorem the dimension d is incorporated through (4.1.9) and may be incorporated through $\mathbb{L}$. Besides these facts, both theorems are rather similar.

(ii) Theorems 4.1.3 and 4.1.6 are rather general and can be applied to any function class which allows for a measurement of their size via brackets with respect to the $\|\cdot\|_\infty$-norm.

(iii) Theorems 4.1.3 and 4.1.6 can be seen as generalizations of Lemma 4 in Schmidt-Hieber [2017] for dependent observations.

(iv) A specific example for $\mathcal{W}$ with support $[0,1]^{dr}$ is given by

$$\mathcal{W}(x) := 1 - \rho(\varsigma^{-1} d(x, [\varsigma, 1-\varsigma]^{dr})) = \begin{cases} 1, & x \in [\varsigma, 1-\varsigma]^{dr} \\ 0, & x \notin [0,1]^{dr}, \\ \text{linear}, & \text{else} \end{cases} \tag{4.1.13}$$

where $\rho(z) := \max\{\min\{z, 1\}, 0\}$ and $d_\infty(x, A) := \inf_{y \in A} |x - y|_\infty$.

In the special case of polynomial decay and exponential decay of $\Delta(\cdot)$, explicit representations of $\Lambda^{dep}(\cdot)$ are available via the following lemma.

Lemma 4.1.8 (Special cases)**.** *Depending on specific decay rates, the following statements hold true.*

(i) If $\Delta(j) \leq \kappa j^{-\alpha}$ *with some* $\kappa > 0, \alpha > 1$*, then*

$$\Lambda^{dep}(x) \leq c_{\kappa,\alpha} \max\{x^{\frac{\alpha}{\alpha+1}}, x\}$$

where $c_{\kappa,\alpha}$ *is a constant depending only on* κ, α*.*

(ii) If $\Delta(j) \leq \kappa\rho^j$ with some $\kappa > 0, \rho \in (0,1)$, then

$$\Lambda^{dep}(x) \leq c_{\kappa,\rho} x \log(x^{-1} \vee 1)^2$$

where $c_{\kappa,\rho}$ is a constant depending only on κ, ρ.

Remark 4.1.9. It seems to be quite challenging to establish appropriate lower bounds. Since we base our proof techniques on empirical process theory, the upper bounds suffer from an increased variance induced by the dependent observations. It is therefore not clear if our results in Theorem 4.1.3 and 4.1.6 are optimal (from a minimax point of view). We dedicate this problem to future research. In Hansen [2008] it has been shown that a kernel estimator applied to dependent observations is able to achieve the same convergence rates as in the i.i.d. case. However, the proof heavily relies on the estimator's explicit representation. The way we formalize our theory as a risk minimization problem does not allow for such a closed form of the estimator.

It should be noted that the recursion (4.1.1) is only used in the fashion of a regression model and we do not impose any contraction condition on f_0. Thus, it is not necessary that X_i itself has geometric decaying dependence coefficients. Moreover, for the same reason, our theory allows us to discuss the more general d-variate regression model

$$Y_i = f_0(\mathbb{X}_{i-1}) + \varepsilon_i, \quad i = r+1, ..., n,$$

where we do not impose a direct connection between input X_i and output Y_i.

4.2 Application to time series forecasting with neural networks

4.2.1 Encoder-decoder structure and smoothness assumptions

We require that f_0 in (4.1.1) has a specific "sparse" form, which we model through several structural assumptions.

Assumption 4.2.1 (Encoder-decoder assumption)**.** *We assume that*

$$f_0 = f_{dec} \circ f_{enc} \tag{4.2.1}$$

for $f_{enc} : \mathbb{R}^{dr} \to \mathbb{R}^{\tilde{d}}$ with $\tilde{d} \in \{1, ..., d\}$, and $f_{dec} : \mathbb{R}^{\tilde{d}} \to \mathbb{R}^d$ only depending on a maximum of $t_{dec} \in \{1, ..., \tilde{d}\}$ arguments in each component. Furthermore,

$$f_{enc} = g_{enc,1} \circ g_{enc,0} \tag{4.2.2}$$

where $g_{enc,0} : \mathbb{R}^{dr} \to \mathbb{R}^D$, $D \in \mathbb{N}$, only depends on a maximum of $t_{enc,0} \in \{1, ..., dr\}$ arguments in each component and $g_{enc,1} : \mathbb{R}^D \to \mathbb{R}^{\tilde{d}}$ only depends on a maximum of $t_{enc,1} \in \{1, ..., D\}$ arguments in each component.

The structure of f_0 (which has not to be a neural network itself) is depicted in Figure 4.1. Condition (4.2.1) means that f_0 decomposes into a function f_{enc}, which reduces the dimension from dr to $\tilde{d} \in \{1, ..., d\}$ (the "encoder"), and $f_{dec} : \mathbb{R}^{\tilde{d}} \to \mathbb{R}^d$ which expands the dimension to d (the "decoder"). For $r = 1$, such structures typically arise when information has to be compressed into a vector $\mathbb{R}^{\tilde{d}}$ (with the encoder) but also should be restorable close to its original information content (with the decoder).

The domain of f_{enc} is d-dimensional, with a possibly large d. Therefore, a structural constraint in the form of (4.2.2) is one possibility to control the convergence rate of the corresponding network estimator. A typical example we have in mind are additive models of the following form, where $g_{enc,1}$ is basically chosen as a summation function.

Example 4.2.2 (Additive models)**.**

(1) Reduction to one dimension: Suppose that $f_0 = f_{dec} \circ f_{enc}$ where $f_{dec} : \mathbb{R} \to \mathbb{R}^d$ and

$$f_{enc}(x) = \sum_{j=1}^{d} g_j(x_j)$$

for functions $g_j : \mathbb{R} \to \mathbb{R}$. Then, Assumption 4.2.1 is fulfilled with $t_{enc,0} = \tilde{d} = t_{dec} = 1$, $t_{enc,1} = D = d$.

(2) Reduction to $\tilde{d}$ dimensions: Suppose that $f_0 = f_{dec} \circ f_{enc}$ where $f_{dec} : \mathbb{R}^{\tilde{d}} \to \mathbb{R}^d$ and $f_{enc} = (f_{enc,k})_{k=1,...,\tilde{d}}$ with

$$f_{enc,k}(x) = \sum_{i_1,...,i_{t_{enc,0}}=1}^{d} g^{(k)}_{i_1,...,i_{t_{enc}}}(x_{i_1}, ..., x_{i_{t_{enc}}})$$

for functions $g_{i_1,...,i_{t_{enc,0}}} : \mathbb{R}^{t_{enc,0}} \to \mathbb{R}$. Then, Assumption 4.2.1 is fulfilled with the given $t_{enc,0}, \tilde{d} = t_{dec}$ and $t_{enc,1} = D = d^{t_{enc,0}}$.

4.2.2 Neural networks and the estimator

We now present the network estimator, formally. To do so, we use the formulation from Schmidt-Hieber [2017]. Let $\sigma(x) := \max\{x, 0\}$ be the ReLU (rectified linear unit) activation function. For a vector $v = (v_1, ..., v_r) \in \mathbb{R}^r$ and its transpose v' define

$$\sigma_v : \mathbb{R}^r \to \mathbb{R}^r, \quad \sigma_v(x) := (\sigma(x_1 - v_1), ..., \sigma(x_r - v_r))'.$$

Let $L \in \mathbb{N}_0$ and $p = (p_0, ..., p_{L+1}) \in \mathbb{N}^{L+2}$. A neural network with network architecture (L, p) and ReLU activation function is a function of the form

$$f : \mathbb{R}^{p_0} \to \mathbb{R}^{p_{L+1}}, \quad f(x) = W^{(L)}\sigma_{v^{(L)}}W^{(L-1)}\sigma_{v^{(L-1)}} \dots W^{(1)}\sigma_{v^{(1)}}W^{(0)}x, \tag{4.2.3}$$

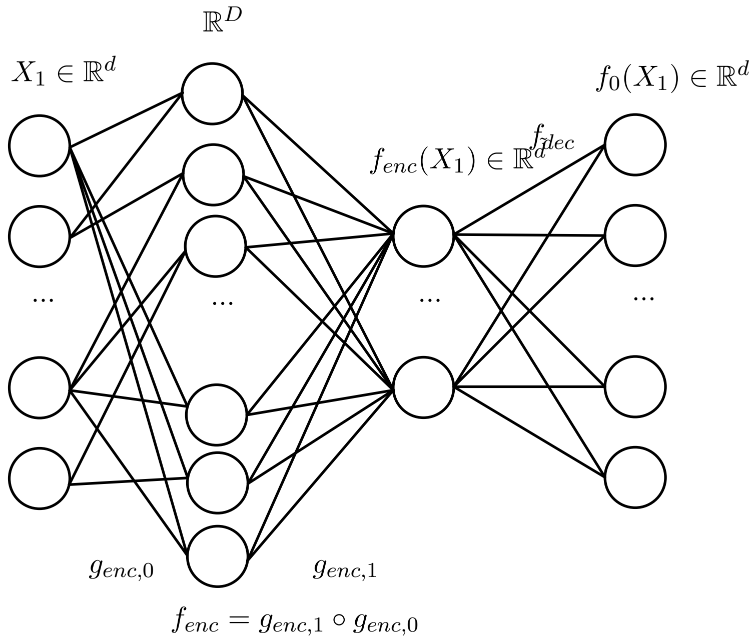

Figure 4.1: Graphical representation of the encoder-decoder assumption on f_0 in the special case $r = 1$.

where $W^{(i)} \in \mathbb{R}^{p_i \times p_{i+1}}$ are weight matrices and $v^{(i)} \in \mathbb{R}^{p_i}$ are bias vectors. We see that $L \in \mathbb{N}_0$ describes the number of hidden layers and $p = (p_0, ..., p_{L+1}) \in \mathbb{N}^{L+2}$ is the number of hidden units for each layer. For $L_1 \in \{1, ..., L\}$, let

$$\mathcal{F}_{ed}(L, L_1, p) \quad := \quad \big\{f : \mathbb{R}^{p_0} \to \mathbb{R}^{p_{L+1}} \text{ is of the form (4.2.3)} : \\ \max_{k=0,...,L} |W^{(j)}|_\infty \vee |v^{(j)}|_\infty \leq 1, p_{L_1} = \tilde{d}\big\},$$

be a class of networks, where the L_1-th hidden layer is $\tilde{d}$-dimensional. In our setting, since we aim to approximate f_0, $p_0 = dr$ and $p_{L+1} = d$ are predetermined.

In practice, a neural network $\hat{f} \in \mathcal{F}_{ed}(L, L_1, p)$ obtained by minimizing $\hat{R}_n(f)$ from (4.1.3) via a stochastic gradient descent method contains weight matrices and bias vectors in which many entries are not relevant for the evaluation $\hat{f}(x)$ of $x \in [0, 1]^d$. This behavior can be explained by the random initialization of the weight matrices and large step sizes of the gradient method. In fact, by employing dropout techniques during the learning process or imposing some additional penalties we can force $W^{(j)}, v^{(j)}$, $j = 0, ..., L$, to be sparse. For a compact overview on a simple stochastic gradient descent method we refer to Richter [2019]. A conceptual explanation of dropout can be found in, for example, Murphy [2022]. To indicate this type of sparsity in the model class, we introduce for $s \in \mathbb{N}$ and $F > 0$,

$$\mathcal{F}(L, L_1, p, s, F) := \Big\{f \in \mathcal{F}_{ed}(L, L_1, p) : \sum_{j=0}^{L} |W^{(j)}|_0 + |v^{(j)}|_0 \leq s, \|f\|_\infty \leq F\Big\}$$

and define the final neural network estimator via

$$\hat{f}^{net} \in \underset{f \in \mathcal{F}(L, L_1, p, s, F)}{\arg\min} \hat{R}_n(f). \tag{4.2.4}$$

In particular, the resulting network $\hat{f}^{net}$ (with estimated weight matrices $\hat{W}^{(j)}$ and bias vectors $\hat{v}^{(j)}$) can provide an estimator of the encoder function f_{enc} by only using its representation up to the L_1-th layer, that is,

$$\hat{f}^{net}_{enc}(x) := \hat{W}^{(L_1)} \sigma_{\hat{v}^{(L_1)}} \hat{W}^{(L_1-1)} \sigma_{\hat{v}^{(L_1-1)}} \dots \hat{W}^{(1)} \sigma_{\hat{v}^{(1)}} W^{(0)} x.$$

Another typical observation made is that fitted neural networks $\hat{f}^{net}$ tend to be rather smooth functions. This can be enforced by adding a gradient penalty in the learning procedure (common, for instance, in the training of WGANs (for Wasserstein generative adversarial networks), where a restricted Lipschitz constant is part of the optimization functional, cf. Gulrajani et al. [2017]). We see in Section 4.1 that we also formally need a bound on the Lipschitz constant when quantifying dependence with the functional dependence measure. We therefore introduce a second neural network estimator based on the function class

$$\mathcal{F}(L, L_1, p, s, F, \mathrm{Lip}) := \big\{ f \in \mathcal{F}(L, L_1, p, s, F) : \|f\|_{Lip} \le \mathrm{Lip} \big\}$$

where $\|f\|_{Lip} := \sup_{x \in \mathbb{R}^d} \frac{|f(x) - f(x')|_\infty}{|x - x'|_\infty}$. This estimator becomes

$$\hat{f}^{net,lip} \in \underset{f \in \mathcal{F}(L, L_1, p, s, F, \mathrm{Lip})}{\arg\min} \hat{R}_n(f). \tag{4.2.5}$$

In the following results, we will assume that Lip is constant in n. From Theorem 4.6.20 in Section 4.6, Subsection 4.6.7, we see that in principle, there exist approximating neural networks which provide the best approximation rate and have Lipschitz constant uniformly bounded in n. In practice, we may force a bounded Lipschitz constant by using an additional gradient penalty (cf. Gulrajani et al. [2017]).

4.2.3 Smoothness assumptions

To state convergence rates of $\hat{f}$, we have to quantify smoothness assumptions of the underlying true function f_0 and its components $g_{enc,1}, g_{enc,0}$ and f_{dec}. We measure smoothness with the well-known Hölder balls. A function has Hölder smoothness index β if all partial derivatives up to order $\lfloor \beta \rfloor := \max\{k \in \mathbb{N}_0 : k < \beta\}$ exist, are bounded and the partial derivatives of order $\lfloor \beta \rfloor$ are $\beta - \lfloor \beta \rfloor$ Hölder continuous. The ball of β-Hölder functions with radius $K > 0$ and domain of definition $P \subset \mathbb{R}^r$ is defined as

$$\begin{aligned} C^\beta(P, K) \quad := \quad & \Big\{ f : P \to \mathbb{R} : \\ & \sum_{\alpha : |\alpha| < \beta} \|\partial^\alpha f\|_\infty + \sum_{\alpha : |\alpha| = \lfloor \beta \rfloor} \sup_{\substack{x, y \in P \\ x \neq y}} \frac{|\partial^\alpha f(x) - \partial^\alpha f(y)|}{|x - y|_\infty^{\beta - \lfloor \beta \rfloor}} \le K \Big\} \end{aligned}$$

where $\alpha = (\alpha_1, ..., \alpha_r) \in \mathbb{N}_0^r$ is a multi-index and $\partial^\alpha := \partial^{\alpha_1} ... \partial^{\alpha_r}$, $|\alpha| := \alpha_1 + ... + \alpha_r$.

We now impose the following assumptions.

Assumption 4.2.3 (Smoothness assumption). *Suppose that for some constants $K \geq 1$ and $\beta_{dec}, \beta_{enc,1}, \beta_{enc,0} \geq 1$,*

- $g_{enc,0} \in C^{\beta_{enc,0}}([0,1]^{dr}, K)$ *and* $g_{enc,0}([0,1]^{dr}) \subset [0,1]^D$,
- $g_{enc,1} \in C^{\beta_{enc,1}}([0,1]^D, K)$ *and* $g_{enc,1}([0,1]^D) \subset [0,1]^{\tilde{d}}$,
- $f_{dec} \in C^{\beta_{dec}}([a_{enc,1}, b_{enc,1}]^{\tilde{d}}, K)$.

The restriction to the unit intervals for the domain and image is only done for the sake of simplicity in our presentation and can be easily enlarged to compact sets by rescaling.

4.2.4 Network conditions

For the following theorems we need assumptions on the network class itself. These assumptions are mainly adapted from [Schmidt-Hieber, 2017, Theorem 1] and are necessary to control the approximation error of the class $\mathcal{F}(L, L_1, p, s, F)$ as well as the size $\mathbb{H}(\delta, \mathcal{F}(L, L_1, p, s, F), \|\cdot\|_\infty)$ of the corresponding covering numbers. The parameter N therein is a parameter in the final theorems.

Assumption 4.2.4. *Fix $N \in \{1, ..., n\}$. The parameters L, L_1, p, s, F of $\mathcal{F}(L, L_1, p, s, F)$ are chosen such that*

(i) $K \leq F$,

(ii) $\{\log_2(4(t_{enc,0} \vee \beta_{enc,0})) + \log_2(4(t_{enc,1} \vee \beta_{enc,1}))\} \log_2(n) \leq L_1$ *and* $L_1 + \log_2(4(t_{dec} \vee \beta_{dec})) \log_2(n) \leq L \lesssim \log_2(n)$,

(iii) $N \lesssim \min_{i \in \{1,...,L\} \setminus \{L_1\}} \{p_i\}$,

(iv) $N \log_2(n) \asymp s$.

We now give a small discussion on the conditions. As we will see below, the optimal N is roughly of the size n^a, where a depends on smoothness properties of the underlying function f_0. Assumption (i) encodes the necessary fact that the network class has to include networks which have a supremum norm larger than the true function f_0. The second condtion (ii) is a condition on the layer size. It should be chosen of order $L \asymp \log_2(n)$. In fact, the upper bound on L is not necessary but produces the best convergence rates (cf. the proof of Theorem 4.2.5 or Theorem 4.2.6, respectively). Condition (iii) poses a lower bound on the size of the hidden layers (i.e. number of hidden units) in the network. From a practical point of view, it seems rather unusual to impose such a large dimension $\geq n^a$ to *all* the hidden layers. This is due to the approximation technique used and surely can be improved. The last condition (iv) is about the number of nonzero parameters $s \asymp N \log_2(n)$ which for instance could be enforced by computational methods during the learning process.

4.2.5 Theoretical results

During this section, let $\mathcal{W} : \mathbb{R}^{dr} \to [0,1]$ be an arbitrary (measurable) weight function with $\text{supp}(\mathcal{W}) \subset [0,1]^{dr}$. The weight function appears in the optimization functional (4.1.3) and the corresponding prediction error (4.1.2).

Theorem 4.2.5 (Mixing). *Suppose that Assumption 4.1.1, 4.2.1, 4.2.3 and 4.1.2 are fulfilled. If Assumption 4.2.4 is satisfied for some $N \in \{1,...,n\}$, then*

$$\mathbb{E}R(\hat{f}^{net}) - R(f_0) \lesssim \Lambda^{mix}\Big(\frac{N\log(n)^3}{n}\Big) + N^{-2A}$$

where $A := \min\big\{\frac{\beta_{dec}}{t_{dec}}, \frac{\beta_{enc,0}}{t_{enc,0}}, \frac{\beta_{enc,1}}{t_{enc,1}}\big\}$.

To formulate an analogous result for the functional dependence measure, we have to assume that the weight function in (4.1.3) is Lipschitz continuous in the sense that for some $\varsigma > 0$,

$$|\mathcal{W}(x) - \mathcal{W}(x')| \leq \frac{1}{\varsigma} \cdot |x - x'|_\infty.$$

Theorem 4.2.6 (Functional dependence). *Suppose that Assumption 4.1.1, 4.2.1, 4.2.3 are fulfilled. Let Assumption 4.1.5 hold true for $L_{\mathcal{G}} = 2dr\big(\frac{2}{\varsigma} + \frac{(\text{Lip}+K)}{F}\big)$. Then there exists some constant $\mathbb{C}_L > 0$ independent of n such that if Assumption 4.2.4 is satisfied for some $N \in \{1,...,n\}$ and $\text{Lip} \geq \mathbb{C}_L$,*

$$\mathbb{E}R(\hat{f}^{net,lip}) - R(f_0) \lesssim \Lambda^{dep}\Big(\frac{N\log(n)^3}{n}\Big) + N^{-2A}$$

where $A := \min\big\{\frac{\beta_{dec}}{t_{dec}}, \frac{\beta_{enc,0}}{t_{enc,0}}, \frac{\beta_{enc,1}}{t_{enc,1}}\big\}$.

A specific expression for $\mathbb{C}_L$ becomes available later on but due to its complicated form we reduce it to its formal existence.

Remark 4.2.7. Note that in the case of independent observations X_i, we can choose $\Lambda^{mix}(x) = \Lambda^{dep}(x) = x$ in Theorem 4.2.5 and 4.2.6 which yields then the same result as in [Schmidt-Hieber, 2017, Theorem 1].

To get a glimpse on the convergence rates which can be achieved, we formulate the following two corollaries of Theorem 4.2.5. Due to the similar form, an analogous result is available also in the case of the functional dependence measure. The first corollary is a simple consequence of Lemma 4.1.4 and Theorem 4.2.5 in the case of polynomial decaying dependence.

Corollary 4.2.8 (Mixing and polynomial decay). *Suppose that Assumption 4.1.1, 4.2.1 and 4.2.3 are fulfilled and that X_i is absolutely regular mixing with coefficients satisfying $\sum_{k=0}^\infty k^{\alpha-1}\beta^{mix}(k) < \infty$ for some $\alpha > 1$. Let*

$$A = \min\Big\{\frac{\beta_{dec}}{t_{dec}}, \frac{\beta_{enc,0}}{t_{enc,0}}, \frac{\beta_{enc,1}}{t_{enc,1}}\Big\}.$$

If Assumption 4.2.4 is satisfied with

$$N = \left\lceil n^{\frac{\frac{\alpha}{\alpha+1}}{2A+\frac{\alpha}{\alpha+1}}} \right\rceil,$$

then

$$\mathbb{E}R(\hat{f}^{net}) - R(f_0) \lesssim n^{-\frac{2A\cdot\frac{\alpha}{\alpha+1}}{2A+\frac{\alpha}{\alpha+1}}} \log(n)^{\frac{3\alpha}{\alpha+1}}.$$

We now investigate this rate for a specific model from Example 4.2.2, (2) with only one lag $r = 1$. Suppose that $t_{enc,0} = \tilde{d}$ and

$$f_{dec}, \quad g_{i_1,\dots,i_{\tilde{d}}} \in C^{\beta}([0,1]^{\tilde{d}}, K)$$

for some $\beta > 0$. This means that the encoder function produces a compressed result of $\tilde{d} \leq d$ components, where each of the $\tilde{d}$ components is constructed as follows: For each possibility to choose $\tilde{d}$ from d arguments, a different function can be used to process the given values. These results are all summed up. Since the summation is infinitely often differentiable with bounded derivatives, the situation in Corollary 4.2.8 becomes

$$A = \min\left\{\frac{\beta}{\tilde{d}}, \frac{\infty}{d^{\tilde{d}}}, \frac{\beta}{\tilde{d}}\right\} = \frac{\beta}{\tilde{d}},$$

which yields the following result.

Corollary 4.2.9. *Suppose that Assumption 4.1.1 holds true and that X_i is mixing with coefficients satisfying $\sum_{k=0}^{\infty} k^{\alpha-1}\beta^{mix}(k) < \infty$ for some $\alpha > 1$. Let Assumption 4.2.4 be satisfied with*

$$N = \left\lceil n^{\tilde{d}\cdot\frac{\frac{\alpha}{\alpha+1}}{2\beta+\tilde{d}\cdot\frac{\alpha}{\alpha+1}}} \right\rceil.$$

Then,

$$\mathbb{E}R(\hat{f}^{net}) - R(f_0) \lesssim n^{-\frac{2\beta\cdot\frac{\alpha}{\alpha+1}}{2\beta+\tilde{d}\cdot\frac{\alpha}{\alpha+1}}} \log(n)^{\frac{3\alpha}{\alpha+1}}.$$

In contrast to the rate of a (naive) standard nonparametric estimator which suffers from the curse of the dimension d, we are thus able to formulate structural conditions on the evolution of the time series to obtain much faster rates which only depend on the compressed dimension $\tilde{d} \in \{1, \dots, d\}$. Of course, the list in Example 4.2.2 is not exhaustive and much more models are suitable for our theory.

4.3 Simulations

In this section, we discuss the behavior of the estimator $\hat{f}$ from (4.2.4). Note that we use the approximation obtained via a stochastic gradient descent method and based on an unrestricted neural network function class. During the presentation, v' denotes the transpose of a vector or matrix v.

4.3.1 Simulated data

We first consider a low-dimensional example given by

$$X_i = f_0(X_{i-1}) + \varepsilon_i$$

where $\varepsilon_i \sim \mathcal{N}(0, 0.5 I_{5\times 5})$, $I_{5\times 5}$ denoting the 5-dimensional identity matrix, and

$$f_0 : \mathbb{R}^5 \to \mathbb{R}^5, \quad f_0(x) = va'x \tag{4.3.1}$$

for $a = (0.5, 0.6, 0.2, 0.3, 0.5)' \in \mathbb{R}^5$ and $v = (0.4, 0.6, 0.5, -0.2, 0.5)' \in \mathbb{R}^5$, that is,

$$X_i = v \cdot \sum_{j=1}^{5} a_j x_j.$$

We generate $n = 1000$ observations $X_1, ..., X_n$ following the above recursion and use $n_{test} - n = 1000$ further realizations of the time series to quantify the true prediction error. For the fitting process, we use an encoder-decoder network of the form

$$p = (5, 20, 10, 1, 10, 20, 5), \qquad L = 5,$$

that is, the network encodes the given information to one dimension and afterwards restores the value again to 5 dimensions. The network is learned by a standard stochastic gradient descent method (cf. Richter [2019]) with learning rate $\gamma = 0.003$ for the first 30 epochs and $\gamma = 0.0002$ afterwards. Furthermore, we use a penalty weight of $\lambda = 0.00001$ and the ReLU activation function. We can deduce from Figure 4.2 that the neural network can easily learn the underlying function f already after approximately 40 epochs. The process is stopped if the (empirical) error

$$D(f) := \frac{1}{d}\mathbb{E}[|f(\mathbb{X}_r) - f_0(\mathbb{X}_r)|_2^2], \qquad \hat{D}_n(f) := \frac{1}{n}\sum_{i=r+1}^{n} \frac{1}{d}|f(\mathbb{X}_{i-1}) - f_0(\mathbb{X}_{i-1})|_2^2$$

does not significantly fluctuate for a certain period of epochs (manual stopping criterion). We surmise that for low dimensional data the testing error (the error calculated on the test data) can be seen on par with the training error (the error calculated on the training data), converging rapidly towards the optimal prediction error $\frac{1}{5}\mathbb{E}[|\varepsilon_1|^2] = 0.5^2 = 0.25$.

We now turn to an example of higher dimension. We take the same model but consider

$$f_0 : \mathbb{R}^{30} \to \mathbb{R}^{30}, \quad f_0(x) = va'x, \tag{4.3.2}$$

where we define the vector $s = (0.05, -0.05, ..., 0.05, -0.05)' \in \mathbb{R}^{24}$ that alternates between the values 0.05 and -0.05 and put

$$\begin{aligned} a &= \begin{pmatrix} 0.3 & 0.6 & 0.5 & s' & 0 & -1 & 0.4 \\ 0.5 & -0.6 & 0.2 & s' & 0.4 & 0.9 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 30}, \\ v &= \begin{pmatrix} 0.4 & 0.4 & \dots & 0.4 & 0.4 \\ 0.5 & -0.3 & \dots & 0.5 & -0.3 \end{pmatrix}^t \in \mathbb{R}^{30\times 2}. \end{aligned}$$

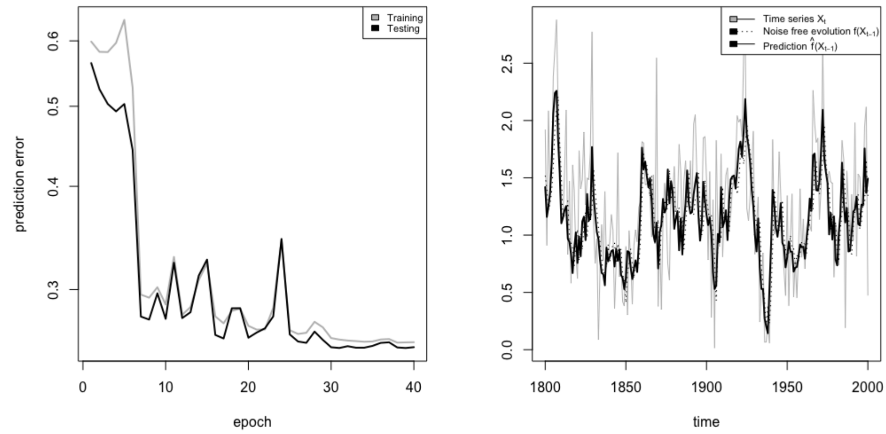

Figure 4.2: We depict the learning process under model (4.3.1). After 40 epochs the neural network learned the underlying function f_0 provided by a noisy version of the data. We can clearly see that the neural network is able to predict the noise free evolution of the times series.

for alternating values 0.5 and -0.3 in the second row of v. The network architecture is adjusted to

$$p = (30, 60, 30, 2, 30, 60, 30), \qquad L = 5,$$

using again a stochastic gradient descent method with learning rate $\gamma = 0.0002$ for the first 50 epochs and $\gamma = 0.00002$ afterwards. Furthermore, we set $\lambda = 0.00001$ and employ the ReLU activation function. Although the network is dealing with an input and output of dimension 30, Figure 4.3 shows that a good prediction already can be realized and most of the information can be preserved despite the data passing a layer of only two dimensions.

4.3.2 Real weather data

For our simulation study we consider weather data of $d = 32$ German cities provided by the Deutscher Wetterdienst (DWD, German Meteorological Service). Note that the cities chosen are spread throughout Germany which can be seen in Figure 4.7. The data we are interested in is the daily mean of temperature and can be found on the DWD's webpage under *Germany's historical data*, https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/daily/kl/historical/. In total we observe 4779 temperature values for each city over the period of 2006/07/01 to 2019/08/01. A subset of $n_{\text{train}} = 4415$ values serves as training data for the network and represents the data from 2006/07/01 to 2018/07/31. We validate our prediction on the year 2018/08/01 to 2019/07/31 which contains $n - n_{\text{train}} = 354$ values. For fitting, we

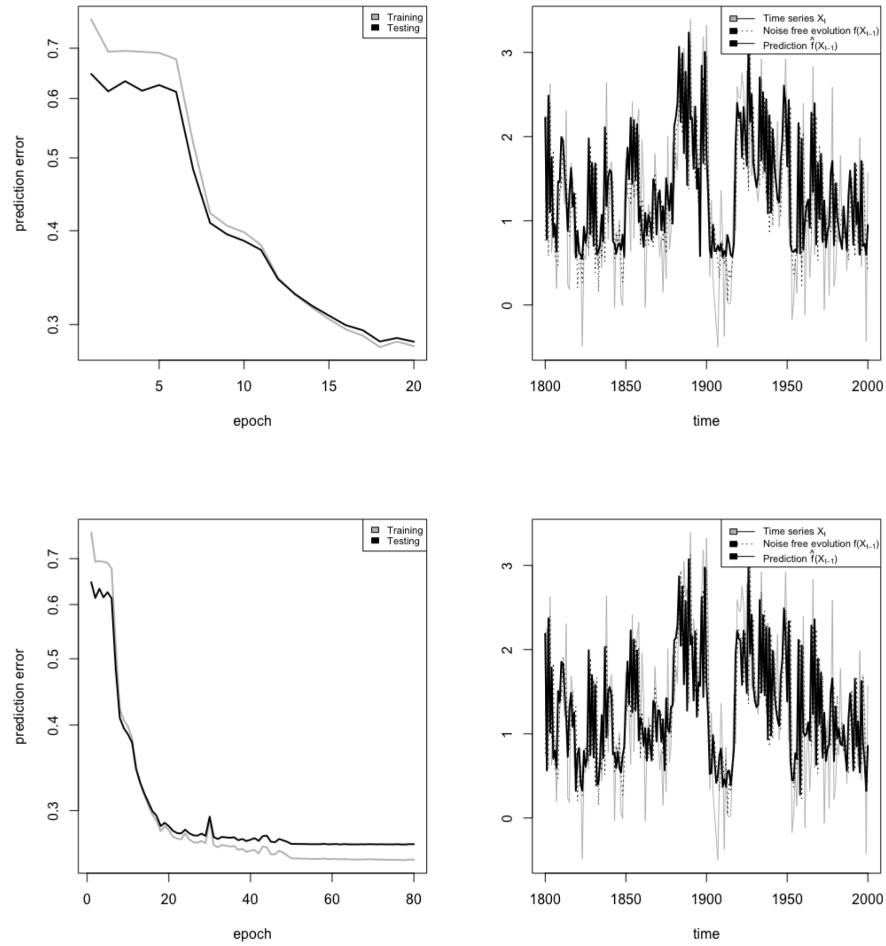

Figure 4.3: The underlying model here is given by (4.3.2). After 20 epochs the neural network learned the overall behavior of the function f_0. The network has still the potential to improve for the lower peaks. After about 40 epochs the learning process can be seen as completed. Due to overfitting, the testing error now begins to slowly increase.

use a network with architecture

$$\mathcal{F}(5, (rd, rd, 24, m, 24, d, d))$$

where $r \in \{1, 2, 3, 5\}$ and $m \in \{4, 6, 8, 10\}$, apply a stochastic gradient descent method for learning the approximation $\hat{f}^{\approx}$ of $\hat{f}$ over 150 epochs. The learning rate is chosen to be $\gamma = 0.000002$ until epoch 45 and $\gamma = 0.0000002$ thereafter. We let the simulation run 5 times over every step r for each network described by m.

In Figure 4.6 we summarize the prediction errors of $D(\hat{f}^{\approx})$ obtained during the testing process. The smallest error can be found for $r = 2$ (that is, using X_{i-1}, X_{i-2} for predicting X_i) with a "bottleneck" layer L_1 of $m = 10$ hidden units. However, it is also possible to take a layer with $m \in \{6, 8\}$ hidden units and still obtain a comparable result. Thus, we surmise that according to our model, when considering the errors, the weather should be predicted based on the two previous days. Taking one previous day or more than three days before the date of interest does not seem to yield a good prediction.

In comparison, the naive prediction method $\hat{f}^{\text{naive}}$ of taking the temperature value of the current day as it is to forecast the next day's value yields an error of approximately $D(\hat{f}^{\text{naive}}) \approx 4.99$. Choosing an AR(1) model, the usual predictor $\hat{f}^{\text{AR}(1)}$ yields an error of $D(\hat{f}^{\text{AR}(1)}) \approx 4.29$, which performs betters than the naive approach and comes close to that of our encoder-decoder network. We therefore see that employing encoder-decoder neural networks produces more accurate predictions.

For $r = 2, m = 6$, we depict the development of the training and testing error for the network $\mathcal{F}(5, (2d, 2d, 24, 6, 24, d, d))$ in Figure 4.4. After 45 epochs the testing error already drops down to a magnitude of 4 which means that we anticipate a deviation (i.e. $\sqrt{D(\cdot)}$) of 2 Kelvin for the prediction itself. The fitting process is displayed for the city of Mannheim in Figure 4.5.

Additionally, the 1-step predictor can be used to forecast k-steps ahead in time by applying the learned neural network k-times, accordingly. In our example, we applied this to the next week's temperature, i.e. $k = 7$. The chosen predictor with architecture $\mathcal{F}(5, (d, d, 24, 6, 24, d, d))$ yields a deviation of around 4.44 Kelvin.

4.4 Approximation error

We consider a network $\tilde{f}_0$ approximating the true regression function f_0. The network $\tilde{f}_0$ is assumed to have the form

$$\tilde{f}_0 = \tilde{f}_{dec} \circ \tilde{f}_{enc} : \mathbb{R}^{dr} \to \mathbb{R}^d \tag{4.4.1}$$

where $\tilde{f}_{enc} : \mathbb{R}^{dr} \to \mathbb{R}^{\tilde{d}}$ and $\tilde{f}_{dec} : \mathbb{R}^{\tilde{d}} \to \mathbb{R}^d$, $\tilde{d} \in \{1, ..., d\}$, and $\tilde{f}_{enc}$ has the additional network structure

$$\tilde{f}_{enc} = \tilde{g}_{enc,1} \circ \tilde{g}_{enc,0}$$

where $\tilde{g}_{enc,0} : \mathbb{R}^{dr} \to \mathbb{R}^D$, $D \in \mathbb{N}$, depends on at most $t_{enc,0} \in \{1, ..., dr\}$ arguments in each component, and $\tilde{g}_{enc,1} : \mathbb{R}^D \to \mathbb{R}^{\tilde{d}}$ depends on at most $t_{enc,1} \in \{1, ..., D\}$ arguments

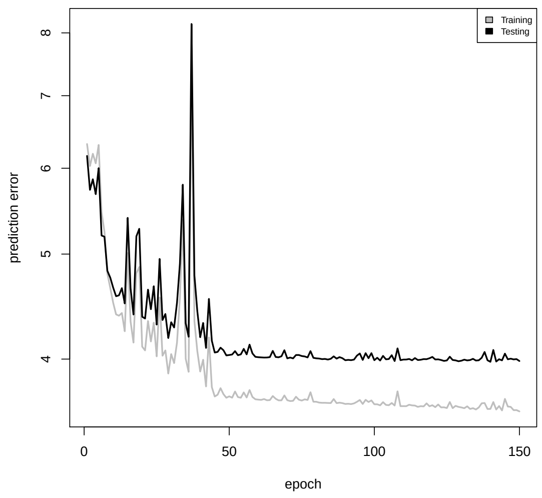

Figure 4.4: Depicted is the training and testing error in the learning process of the network $\mathcal{F}(5,(2d,2d,24,6,24,d,d))$ applied to the weather data. We clearly see that consistently, as expected, the testing error is higher than the training error. At an early stage the network already learns basic properties of the evolution scheme of the time series because the testing error rapidly drops. After 45 epochs the error is in the range of 4.

in each component. We denote by

$$\mathcal{F}(L,p,s) \quad := \quad \Big\{ f:\mathbb{R}^{p_0} \to \mathbb{R}^{p_{L+1}} \text{ is of the form (4.2.3)} : \\ \max_{k=0,\dots,L} |W^{(j)}|_\infty \vee |v^{(j)}|_\infty \le 1, \sum_{j=0}^{L} |W^{(j)}|_0 + |v^{(j)}|_0 \le s \Big\}$$

the set of all networks with constraints on $W^{(j)}$, $v^{(j)}$ and sparsity level s. We explicitly do not ask for the presence of an encoder-decoder structure or an intermediate hidden layer at position L_1.

Now, let $\boldsymbol{t} := (t_{dec}, t_{enc,0}, t_{enc,1})$ and $\boldsymbol{\beta} := (\beta_{dec}, \beta_{enc,0}, \beta_{enc,1})$ where $t_{dec} \in \{1,...,\tilde{d}\}$, $t_{enc,1} \in \{1,...,D\}$ and $t_{enc,0} \in \{1,...,dr\}$.

Theorem 4.4.1. *Consider the d-dimensional time series that follows the recursion relation (4.1.1) and Assumption 4.2.1, 4.2.3. Let $N \in \{1,...,n\}$. Suppose that the parameters of $\mathcal{F}(L,L_1,p,s,F,\mathrm{Lip})$ satisfy*

(i) $K \le F$,

(ii) $\sum_{i\in\{enc,0;enc,1\}} \log_2(4(t_i \vee \beta_i)) \log_2(n) \le L_1$ *and* $L_1 + \log_2(4(t_{dec} \vee \beta_{dec})) \log_2(n) \le L$,

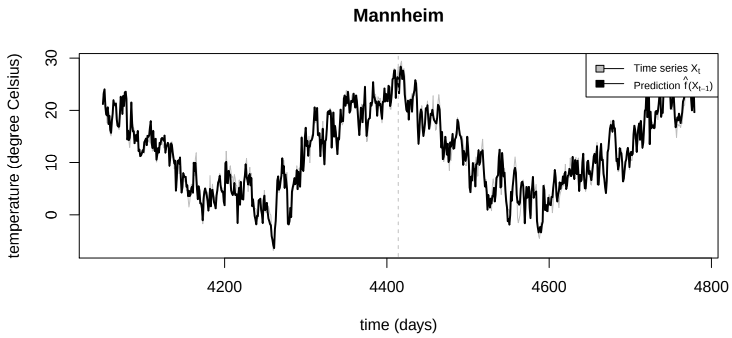

Figure 4.5: The graphic shows the daily mean temperature data from 2017/08/01 to 2019/07/31 measured in Mannheim. Note that we continuously count the days from 1 to 4779 beginning on 2006/07/01 (day 1). The training process ends on 2018/31/07 (day 4414), indicated by the gray vertical dashed line in the middle. Beginning on 2018/08/01 we see the values predicted $\hat{f}^{\approx}(X_{i-1}, X_{i-2})$ by the learned neural network on top of the actual data observed.

(iii) $N \lesssim \min_i\{p_i\}$,

(iv) $N \log_2(n) \lesssim s$,

(v) Lip $\gtrsim 1$.

Then,

$$\inf_{f^* \in \mathcal{F}(L, L_1, p, s, F, \mathrm{Lip})} \|f^* - f_0\|_\infty^2 \leq C \max_{k \in \{dec; enc, 0; enc, 1\}} \left\{ \frac{N}{n} + N^{-\frac{2\beta_k}{t_k}} \right\}$$

for a large enough constant C that only depends on $\tilde{d}, d, \boldsymbol{t}, \boldsymbol{\beta}$.

The proof can be found in Section 4.6, Subsection 4.4.

4.5 Concluding remarks

In this chapter, we proposed a method to forecast high-dimensional time series with encoder-decoder neural networks and quantified their prediction abilities theoretically with a convergence rate. A key step was to provide oracle-type inequalities for minimizers of the empirical prediction error under mixing or functional dependence. Besides the fact that the corresponding neural network is required to have a similar encoder-decoder structure to avoid overfitting, we also formulated appropriate conditions on the network parameters, such as bounds for the number of layers or active parameters. The encoder-decoder structure we used is fundamental to possibly circumvent the curse of dimension. The conditions imposed are similar to those in Schmidt-Hieber [2017] since we have used the same approximation results.

$r = 1$		prediction error upon validation				
layer m	4	4.81	4.77	4.63	4.68	4.86
	6	4.41	4.65	4.75	4.83	4.64
	8	4.41	4.49	4.42	4.47	4.45
	10	4.41	4.55	4.44	4.45	4.45
$r = 2$						
layer m	4	4.63	4.22	4.41	4.19	4.03
	6	3.98	4.10	4.21	4.16	4.22
	8	3.98	3.95	4.23	4.05	4.01
	10	4.11	4.09	3.93	3.93	4.02
$r = 3$						
layer m	4	4.30	4.79	4.11	4.72	4.10
	6	4.27	4.46	4.18	4.04	4.18
	8	4.36	4.29	4.08	4.24	4.28
	10	4.27	4.12	4.08	4.15	4.28
$r = 5$						
layer m	4	4.28	4.75	4.83	4.48	4.85
	6	4.10	4.27	4.28	4.34	4.71
	8	4.81	4.09	4.06	4.42	4.45
	10	4.24	4.47	4.28	4.37	4.36

Figure 4.6: The testing errors obtained during the simulation. For each of the 4 distinct network architectures and each of r-step predictions we ran the simulation 5 times.

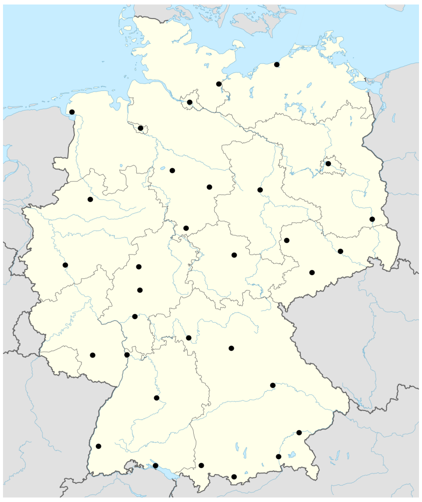

Figure 4.7: We collected weather data from the cities of Berlin, Braunschweig, Bremen, Chemnitz, Cottbus, Dresden, Erfurt, Frankfurt, Freiburg, Garmisch-Patenkirchen, Göttingen, Münster, Hamburg, Hannover, Kaiserslautern, Kempten, Köln, Konstanz, Leipzig, Lübeck, Magdeburg, Cölbe, Mühldorf, München, Nürnberg, Regensburg, Rosenheim, Rostock, Stuttgart, Würzburg, Emden and Mannheim.

Our theory can be seen as an extension of the upper bounds found in Schmidt-Hieber [2017] to dependent observations with high-dimensional outputs.

We also studied the performance of our neural network estimators with simulated data and saw that the estimators could detect and adapt to a specific encoder-decoder structure of the true evolution function quite successfully. We applied our procedure to real temperature data and showed that without too much tuning we were able to outperform a naive forecast.

4.6 Lemmata and Proofs of Chapter 4

4.6.1 A variance bound under mixing for Subsection 4.1.1

Recall the definition of the β-mixing coefficients $\beta^{mix}(k)$, $k \in \mathbb{N}_0$ from (4.1.4). In this section, we use the abbreviation $\beta(\cdot) = \beta^{mix}(\cdot)$.

We now introduce the $\|\cdot\|_{2,\beta}$-norm which originally was defined in Doukhan et al. [1995].

Define $\beta(t) = \beta(\lfloor t \rfloor)$ for $t \geq 1$ and $\beta(t) = 1$, otherwise. For some cadlag function $g : I \to \mathbb{R}$ defined on a domain $I \subset \mathbb{R}$, the cadlag inverse is defined as

$$g^{-1}(u) := \inf\{s \in I : f(s) \leq u\},$$

which we especially use for $\beta^{-1}(u)$. For any measurable $h : \mathbb{R} \to \mathbb{R}$, let $Q_h(u)$ denote the quantile function of $h(X_1)$, that is, $Q_h(u)$ is the cadlag inverse of $t \mapsto \mathbb{P}(h(X_1) \leq t)$. Let

$$\|h\|_{2,\beta} := \Big(\int_0^1 \beta^{-1}(u) Q_h(u)^2 du \Big)^{1/2}.$$

This norm can be used to upper bound the variance of a sum $\sum_{i=1}^k h(X_i)$. Furthermore, it is possible to upper bound $\|h\|_{2,\beta}$ in terms of $\|h\|_1 = \mathbb{E}|h(X_1)|$ and $\|h\|_\infty$ which we will need in the proofs to relate the variance of the empirical risk with the risk itself. Let

$$\Phi := \{\phi : [0,\infty) \to [0,\infty) \quad | \quad \phi \text{ increasing, convex, differentiable,} \\ \phi(0) = 0 \text{ and } \lim_{x\to\infty} \frac{\phi(x)}{x} = \infty\}.$$

For $\phi \in \Phi$, let $\phi^*(y) := \sup_{y>0}\{xy - \phi(x)\}$ be the convex dual function. Define the Orlicz norm associated to $\phi(x^2)$ via

$$\|h\|_{\phi,2} := \inf\{c > 0 : \mathbb{E}\phi\Big((\frac{|h(X_1)|}{c})^2 \Big) \leq 1\}.$$

The following two results are from [Doukhan et al., 1995, Proposition 1 and Lemma 2].

Lemma 4.6.1 (Variance bounds and bound of $\|\cdot\|_{2,\beta}$-norm)**.** *For $k \in \mathbb{N}$,*

$$\mathrm{Var}\Big(\sum_{i=1}^k h(X_i) \Big) \leq 4k\|h\|_{2,\beta}^2.$$

For $\phi \in \Phi$ assume that $\int_0^1 \phi^*(\beta^{-1}(u)) < \infty$. Then,

$$\|h\|_{2,\beta} \leq C_\beta \cdot \|h\|_{\phi,2}, \qquad C_\beta := \Big(1 + \int_0^1 \phi^*(\beta^{-1}(u)) du\Big)^{1/2}.$$

If $\|h\|_\infty \leq 1$, then

$$\|h\|_{\phi,2} \leq C_\beta \varphi(\|h\|_1) \tag{4.6.1}$$

where $\varphi(x) := \phi^{-1}(x^{-1})^{-1/2}$.

Only the last statement needs to be proven and is postponed to Subsection 4.6.4. The main goal of this section is to prove Theorem 4.6.4, which is Theorem 4.1.3. To do so, we use techniques and decomposition ideas from Dedecker and Louhichi [2002], Rio [1995] and Liebscher [1996]. We begin by establishing maximal inequalities under mixing. The proofs can be found here in Section 4.6, as well.

4.6.2 Maximal inequalities under mixing

Let $\mathcal{G} \subset \{g : \mathbb{R}^{dr} \to \mathbb{R} \text{ measurable}\}$ be a finite class of functions and

$$S_n(g) := \sum_{i=r+1}^{n} \{g(\mathbb{X}_{i-1}) - \mathbb{E}g(\mathbb{X}_{i-1})\}.$$

In the following, let $H = 1 \vee \log |\mathcal{G}|$. Recall

$$q^*(x) = q^{*,mix}(x) = \min\{q \in \mathbb{N} : \beta^{mix}(q) \le qx\}.$$

Lemma 4.6.2 (Maximal inequalities for mixing sequences). *Suppose that* $\sup_{g\in\mathcal{G}} \|g\|_\infty \le 1$ *and that there exists* $\nu(g) > 0$ *such that* $\sup_{g\in\mathcal{G}} \|\frac{g}{\nu(g)}\|_{2,\beta} \le 1$. *Then, there exists another process* $S_n^\circ(g)$ *and some universal constant* $c > 0$ *such that*

$$\mathbb{E}\sup_{g\in\mathcal{G}} |S_n(g) - S_n^\circ(g)| \le cnr \cdot q^*(\frac{H}{n})\frac{H}{n}. \tag{4.6.2}$$

Furthermore, with $N(g) := q^*(\frac{H}{n})\sqrt{\frac{H}{n}} \vee \nu(g)$,

(i)

$$\mathbb{E}\sup_{g\in\mathcal{G}} |\frac{S_n^\circ(g)}{N(g)}| \le c\sqrt{nrH}, \tag{4.6.3}$$

(ii)

$$\mathbb{E}\big[\sup_{g\in\mathcal{G}} |\frac{S_n^\circ(g)}{N(g)}|^2\big] \le cnr^2H. \tag{4.6.4}$$

Now, for $\mathcal{G} \subset \{g : \mathbb{R}^{dr} \to \mathbb{R}^d \text{ measurable}\}$, define

$$M_n(g) := \sum_{i=1}^{n} \frac{1}{d}\langle \varepsilon_i, g(\mathbb{X}_{i-1})\rangle.$$

Lemma 4.6.3 (Maximal inequalities for mixing martingale sequences). *Let Assumption 4.1.1 hold true. Furthermore, assume that* X_i *is* β*-mixing and* $\beta^{mix}(\cdot)$ *is submultiplicative in the sense of (4.1.5). Suppose that elements* $g \in \mathcal{G}$ *fulfill* $\sup_{g\in\mathcal{G}} \|g\|_\infty \le 1$ *and that there exists* $\nu(g) > 0$ *such that* $\sup_{g\in\mathcal{G}} \|\frac{|g(\mathbb{X}_r)|_2}{\sqrt{d}\nu(g)}\|_2 \le 1$. *Then, there exists another process* $M_n^\circ(g)$ *and some universal constant* $c > 0$ *such that*

$$\mathbb{E}\sup_{g\in\mathcal{G}} |M_n(g) - M_n^\circ(g)| \le crC_\varepsilon C_{\beta,sub} n \cdot q^*(\frac{H}{n})\frac{H}{n}. \tag{4.6.5}$$

Furthermore, with $N(g) := q^*(\frac{H}{n})\sqrt{\frac{H}{n}} \vee \nu(g)$,

(i)
$$\mathbb{E}\sup_{g\in\mathcal{G}}|\frac{M_n^\circ(g)}{N(g)}| \le cC_\varepsilon\sqrt{nH}, \tag{4.6.6}$$

(ii)
$$\mathbb{E}\big[\sup_{g\in\mathcal{G}}|\frac{M_n^\circ(g)}{N(g)}|^2\big] \le cC_\varepsilon^2 nH. \tag{4.6.7}$$

4.6.3 Oracle inequalities under mixing

Define as a short hand

$$\begin{aligned} D(f) &:= \frac{1}{d}\mathbb{E}[|f(\mathbb{X}_r) - f_0(\mathbb{X}_r)|_2^2\mathcal{W}(\mathbb{X}_r)], \\ \hat{D}_n(f) &:= \frac{1}{n}\sum_{i=r+1}^n \frac{1}{d}|f(\mathbb{X}_{i-1}) - f_0(\mathbb{X}_{i-1})|_2^2\mathcal{W}(\mathbb{X}_{i-1}) \end{aligned}$$

where $\mathcal{W}:\mathbb{R}^{dr}\to[0,1]$ denotes an arbitrary weight function. In this section, we show an oracle-type inequality for minimum empirical risk estimators

$$\hat{f} \in \arg\min_{f\in\mathcal{F}} \hat{R}_n(f), \qquad \hat{R}_n(f) = \frac{1}{n}\sum_{i=r+1}^n \frac{1}{d}|X_i - f(\mathbb{X}_{i-1})|_2^2\mathcal{W}(\mathbb{X}_{i-1})$$

for any weight function $\mathcal{W}:\mathbb{R}^{dr}\to[0,1]$. The function classes considered are of the form

$$\mathcal{F} \subset \{f = (f_j)_{j=1,\dots,d} : \mathbb{R}^{dr} \to \mathbb{R}^d \text{ measurable}\}$$

and have to satisfy $\sup_{f\in\mathcal{F}}\sup_{j\in\{1,\dots,d\}}\sup_{x\in\mathrm{supp}(\mathcal{W})}|f_j(x)| \le F$. For the proof of the following theorem we require Lemma 4.6.5 and Lemma 4.6.6 which, among others, are shown below.

Theorem 4.6.4. *Let Assumption 4.1.1 and 4.1.2 hold true. Suppose that each function $f = (f_j)_{j=1,\dots,d} \in \mathcal{F}$ satisfies $\sup_{j=1,\dots d}\sup_{x\in supp(\mathcal{W})}|f_j(x)| \le F$. Let $\delta\in(0,1)$ and $\mathbb{H} = \log\mathbb{N}(\delta,\mathcal{F},\|\cdot\|_\infty)$. Then, for any $\eta>0$ there exists a constant $\mathbb{C} = \mathbb{C}(\eta, c_0, r, C_{\beta,sub}, C_\varepsilon, F)$ such that*
$$\mathbb{E}D(\hat{f}) \le (1+\eta)^2\inf_{f\in\mathcal{F}} D(f) + \mathbb{C}\cdot\big\{\Lambda(\frac{H}{n}) + \delta\big\}.$$

Proof of Theorem 4.6.4. Note that $\mathbb{E}\hat{D}_n(f) = D(f)$ and

$$\mathbb{E}\hat{D}_n(\hat{f}) = \mathbb{E}\Big[\frac{1}{n}\sum_{i=r+1}^n \frac{1}{d}|\hat{f}(\mathbb{X}_{i-1}) - f_0(\mathbb{X}_{i-1})|_2^2\mathcal{W}(\mathbb{X}_{i-1})\Big].$$

By the model equation (4.1.1) for any $f \in \mathcal{F}$,

$$\begin{aligned}
\hat{R}_n(f) &:= \frac{1}{n}\sum_{i=r+1}^{n}\frac{1}{d}|X_i - f(\mathbb{X}_{i-1})|_2^2\mathcal{W}(X_{i-1}) \\
&= \frac{1}{n}\sum_{i=r+1}^{n}\frac{1}{d}|\varepsilon_i + (f_0(\mathbb{X}_{i-1}) - f(\mathbb{X}_{i-1}))|_2^2\mathcal{W}(\mathbb{X}_{i-1}) \\
&= \frac{1}{n}\sum_{i=r+1}^{n}\frac{1}{d}|\varepsilon_i|_2^2\mathcal{W}(\mathbb{X}_{i-1}) + \frac{1}{n}\sum_{i=r+1}^{n}\frac{1}{d}|f_0(\mathbb{X}_{i-1}) - f(\mathbb{X}_{i-1})|_2^2\mathcal{W}(\mathbb{X}_{i-1}) \\
&\qquad +\frac{2}{n}\sum_{i=r+1}^{n}\frac{1}{d}\langle\varepsilon_i, f_0(\mathbb{X}_{i-1}) - f(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1}) \\
&= \frac{1}{n}\sum_{i=r+1}^{n}\frac{1}{d}|\varepsilon_i|_2^2\mathcal{W}(\mathbb{X}_{i-1}) \\
&\qquad +\frac{2}{n}\sum_{i=r+1}^{n}\frac{1}{d}\langle\varepsilon_i, f_0(\mathbb{X}_{i-1}) - f(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1}) + \hat{D}_n(f).
\end{aligned}$$

Since $\hat{f} = \arg\min_{f\in\mathcal{F}}\hat{R}_n(f)$, we have for all $f \in \mathcal{F}$,

$$\begin{aligned}
\hat{D}_n(\hat{f}) &= \underbrace{\hat{R}_n(\hat{f})}_{\leq \hat{R}_n(f)} - \frac{1}{n}\sum_{i=r+1}^{n}\frac{1}{d}|\varepsilon_i|_2^2\mathcal{W}(\mathbb{X}_{i-1}) \\
&\qquad -\frac{2}{n}\sum_{i=r+1}^{n}\frac{1}{d}\langle\varepsilon_i, f_0(\mathbb{X}_{i-1}) - \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1}) \\
&\leq \hat{D}_n(f) + \frac{2}{n}\sum_{i=r+1}^{n}\frac{1}{d}\langle\varepsilon_i, f_0(\mathbb{X}_{i-1}) - f(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1}) \\
&\qquad -\frac{2}{n}\sum_{i=r+1}^{n}\frac{1}{d}\langle\varepsilon_i, f_0(\mathbb{X}_{i-1}) - \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1}).
\end{aligned}$$

Since $\mathbb{E}[\varepsilon_i|\mathcal{A}_{i-1}] = \mathbb{E}\varepsilon_i = 0$ for $\mathcal{A}_{i-1} = \sigma(\varepsilon_{i-1}, \varepsilon_{i-2}, ...)$,

$$\mathbb{E}\hat{D}_n(\hat{f}) \leq \underbrace{\mathbb{E}\hat{D}_n(f)}_{=D(f)} + \mathbb{E}\Big[\frac{2}{n}\sum_{i=r+1}^{n}\langle\varepsilon_i, \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1})\Big],$$

that is,

$$\mathbb{E}\hat{D}_n(\hat{f}) \leq \inf_{f\in\mathcal{F}} D(f) + 2\mathbb{E}\Big[\frac{1}{n}\sum_{i=r+1}^{n}\langle\varepsilon_i, \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1})\Big]. \tag{4.6.8}$$

Let $\eta > 0$. Define

$$\begin{aligned} R_{1,n} &:= (1+\eta)crF^2 q^*(\frac{H}{n})\frac{H}{n} + \frac{\eta F^2}{2}(\varphi^{-1})^*\Big(2\frac{1+\eta}{\eta F^2}rC_\beta\sqrt{\frac{H}{n}}\Big), \\ R_{1,\delta} &:= crF^2C_\beta\sqrt{\frac{H}{n}}\varphi(2F^{-2}\delta^2), \\ R_{2,n} &:= cC_\varepsilon C_{\beta,sub}rFq^*(\frac{H}{n})\frac{H}{n}, \\ R_{2,\delta} &:= C_\varepsilon\delta + cC_\varepsilon C_{\beta,sub}rF\sqrt{\frac{H}{n}}\delta. \end{aligned}$$

By Lemma 4.6.5, (4.6.8) and Lemma 4.6.6,

$$\begin{aligned} \mathbb{E}D(\hat{f}) &\leq (1+\eta)\mathbb{E}\hat{D}_n(\hat{f}) + R_{1,n} + (1+\eta)R_{1,\delta} \\ &\leq (1+\eta)\Big\{\inf_{f\in\mathcal{F}} D(f) + 2\mathbb{E}\Big[\frac{1}{n}\sum_{i=r+1}^n \langle\varepsilon_i, \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1})\Big] + R_{1,\delta}\Big\} + R_{1,n} \\ &\leq (1+\eta)\Big\{\inf_{f\in\mathcal{F}} D(f) + 2cC_\varepsilon C_{\beta,sub}rF\sqrt{\frac{H}{n}}\mathbb{E}[D(\hat{f})]^{1/2} + R_{2,n} + R_{2,\delta} + R_{1,\delta}\Big\} \\ &\quad +R_{1,n}. \end{aligned}$$

Due to $2ab \leq a^2+b^2$ with $a := (1+\eta)cC_\varepsilon C_{\beta,sub}rF\sqrt{\frac{H}{n}}(\frac{1+\eta}{\eta})^{1/2}$, $b := (\frac{\eta}{1+\eta})^{1/2}\mathbb{E}[D(\hat{f})]^{1/2}$, we obtain

$$\begin{aligned} \mathbb{E}D(\hat{f}) &\leq (1+\eta)\inf_{f\in\mathcal{F}} D(f) + \frac{(1+\eta)^3}{\eta}(cC_\varepsilon C_{\beta,sub}rF)^2\frac{H}{n} + \frac{\eta}{1+\eta}\mathbb{E}[D(\hat{f})] \\ &\quad +(1+\eta)(R_{2,n} + R_{2,\delta} + R_{1,\delta}) + R_{1,n}. \end{aligned}$$

This implies

$$\begin{aligned} \mathbb{E}D(\hat{f}) &\leq (1+\eta)^2\inf_{f\in\mathcal{F}} D(f) + (1+\eta)R_{1,n} \\ &\quad +(1+\eta)^2(R_{2,n} + R_{2,\delta} + R_{1,\delta}) \\ &\quad +\frac{(1+\eta)^4}{\eta}(cC_\varepsilon C_{\beta,sub}rF)^2\frac{H}{n}. \end{aligned} \tag{4.6.9}$$

Using Young's inequality applied to φ^{-1} (φ^{-1} is convex) and Lemma 4.6.7, we obtain

$$R_{1,\delta} \leq crF^2C_\beta(\varphi^{-1})^*\big(\sqrt{\frac{H}{n}}\big) + 2crC_\beta\delta^2 \leq crF^2C_\beta(4c_0)^2\Lambda(\frac{H}{n}) + 2crC_\beta\delta^2.$$

By Lemma 4.6.8, $R_{2,n} \leq 2cC_\varepsilon C_{\beta,sub}F\Lambda(\frac{H}{n})$, and

$$R_{1,n} \leq (1+\eta)crF^2\Lambda(\frac{H}{n}) + \frac{\eta F^2}{2}\Big(2\frac{1+\eta}{\eta F^2}rC_\beta\Big)^2(4c_0)^2\Lambda(\frac{H}{n}).$$

Furthermore,

$$R_{2,\delta} \le C_\varepsilon \delta + cC_\varepsilon C_{\beta,sub} rF\delta^2 + cC_\varepsilon C_{\beta,sub} rF\frac{H}{n}.$$

Insertion of these results into (4.6.9) yields

$$\begin{aligned}
\mathbb{E}D(\hat{f}) \quad \le \quad & (1+\eta)^2 \inf_{f\in\mathcal{F}} D(f) \\
& +\Lambda(\frac{H}{n}) \cdot \Big\{(1+\eta)^2 cF^2 + 32(\frac{(1+\eta)^3}{\eta F^2})C_\beta^2 r^2 c_0^2 \\
& \qquad +2(1+\eta)^2 crC_\varepsilon C_{\beta,sub}F + 16(1+\eta)^2 c_0^2 crC_\varepsilon C_{\beta,sub}F^2\Big\} \\
& +\delta^2 \cdot (1+\eta)^2 cC_\varepsilon C_{\beta,sub} rF + C_\varepsilon \delta \cdot (1+\eta)^2 \\
& +\frac{(1+\eta)^4}{\eta}(cC_\varepsilon C_{\beta,sub} rF)^2 \frac{H}{n}
\end{aligned}$$

□

Lemma 4.6.5. *Suppose that Assumption 4.1.1, 4.1.2 hold true. Assume that each* $f \in \mathcal{F}$ *satisfies* $\sup_{x\in supp(\mathcal{W})} |f(x)|_\infty \le F$. *Let* $\mathbb{H} = \log \mathbb{N}(\delta, \mathcal{F}, \|\cdot\|_\infty)$. *Then, there exists an universal constant* $c > 0$ *such that for any* $\delta \in (0,1), \eta > 0$,

$$\begin{aligned}
\mathbb{E}D(\hat{f}) \quad \le \quad & (1+\eta)\mathbb{E}\hat{D}_n(\hat{f}) + \Big\{(1+\eta)crF^2 q^*(\frac{H}{n})\frac{H}{n} + \frac{\eta F^2}{2}(\varphi^{-1})^*\Big(2\frac{1+\eta}{\eta F^2} rC_\beta \sqrt{\frac{H}{n}}\Big)\Big\} \\
& +(1+\eta)crF^2 C_\beta \sqrt{\frac{H}{n}}\varphi(2F^{-2}\delta^2).
\end{aligned}$$

Proof of Lemma 4.6.5. Let $(f_j)_{j=1,\dots,\mathcal{N}_n}$ be a δ-covering of $\mathcal{F}$, where $\mathcal{N}_n := \mathbb{N}(\delta, \mathcal{F}, \|\cdot\|_\infty)$. Let $j^* \in \{1, ..., \mathcal{N}_n\}$ be such that $\|\hat{f} - f_{j^*}\|_\infty \le \delta$ for all $k = 1, ..., d$. Without loss of generality, assume that $\delta \le F$.

Let $(X_i')_{i\in\mathbb{Z}}$ be an independent copy of the original time series $(X_i)_{i\in\mathbb{Z}}$. Then, the process $(\mathbb{X}_{i-1}, \mathbb{X}'_{i-1})$ is still β-mixing with the coefficients $2\tilde{\beta}(q) = 2\beta(q-r)$. We have

$$\begin{aligned}
& \big|\mathbb{E}D(\hat{f}) - \mathbb{E}\hat{D}_n(\hat{f})| \\
= & \Big|\mathbb{E}\Big[\frac{1}{nd}\sum_{i=r+1}^{n} |\hat{f}(\mathbb{X}'_{i-1}) - f_0(\mathbb{X}'_{i-1})|_2^2 \mathcal{W}(\mathbb{X}'_{i-1}) \\
& \qquad - \frac{1}{nd}\sum_{i=r+1}^{n} |\hat{f}(\mathbb{X}_{i-1}) - f_0(\mathbb{X}_{i-1})|_2^2 \mathcal{W}(\mathbb{X}_{i-1})\Big]\Big| \\
\leq & \Big|\mathbb{E}\Big[\frac{1}{nd}\sum_{i=r+1}^{n} |f_{j^*}(\mathbb{X}'_{i-1}) - f_0(\mathbb{X}'_{i-1})|_2^2 \mathcal{W}(\mathbb{X}'_{i-1}) \\
& \qquad - \frac{1}{nd}\sum_{i=r+1}^{n} |f_{j^*}(\mathbb{X}_{i-1}) - f_0(\mathbb{X}_{i-1})|_2^2 \mathcal{W}(\mathbb{X}_{i-1})\Big]\Big| + 10\delta F \\
\leq & \mathbb{E}\Big|\frac{1}{n}\sum_{i=r+1}^{n} g_{j^*}(\mathbb{X}_{i-1}, \mathbb{X}'_{i-1})\Big| + 10\delta F \\
= & \frac{F^2}{n}\mathbb{E}|S_n(g_{j^*})| + 10\delta F, \qquad (4.6.10)
\end{aligned}$$

where we used $a = \hat{f}(\mathbb{X}_{i-1}) - f_{j^*}(\mathbb{X}_{i-1}), b = f_{j^*}(\mathbb{X}_{i-1}) - f_0(\mathbb{X}_{i-1}) \in \mathbb{R}^d$,

$$\big||a+b|_2^2 - |b|_2^2\big| = |a|_2^2 + 2|\langle a, b\rangle| \leq |a|_2^2 + 2|a|_2|b|_2 \leq d\delta^2 + 4d\delta F \leq 5d\delta F,$$

defined for $x, x' \in \mathbb{R}^d$,

$$g_j(x, x') := \frac{1}{dF^2}|f_j(x') - f_0(x')|_2^2 \mathcal{W}(x') - \frac{1}{dF^2}|f_j(x) - f_0(x)|_2^2 \mathcal{W}(x),$$

and $S_n(\cdot)$ is from Lemma 4.6.2. By the same Lemma 4.6.2, there exists another process $S_n^\circ(\cdot)$ and some universal constant $c > 0$ such that

$$\mathbb{E}|S_n(g_{j^*}) - S_n^\circ(g_{j^*})| \leq \mathbb{E}\sup_{g\in\mathcal{G}} |S_n(g) - S_n^\circ(g)| \leq crq^*(\frac{H}{n})\frac{H}{n}. \qquad (4.6.11)$$

Note that $\|g\|_{2,\tilde{\beta}} \leq r\,\|g\|_{2,\beta}$. Put

$$N(g) := q^*(\frac{H}{n})\sqrt{\frac{H}{n}} \vee 2\|g\|_{2,\beta}.$$

We use the Cauchy-Schwarz inequality and again Lemma 4.6.2, yielding a universal constant $c > 0$ such that

$$\begin{aligned}
\mathbb{E}|S_n^\circ(g_{j^*})| &= \mathbb{E}\Big|\frac{S_n^\circ(g_{j^*})}{N(g_{j^*})} \cdot N(g_{j^*})\Big| \\
&\leq \Big\|\frac{S_n^\circ(g_{j^*})}{N(g_{j^*})}\Big\|_2 \mathbb{E}[\|g\|_{2,\beta}^2\big|_{g=g_{j^*}}]^{1/2} + \mathbb{E}\Big|\frac{S_n^\circ(g_{j^*})}{N(g_{j^*})}\Big| \cdot q^*(\frac{H}{n})\sqrt{\frac{H}{n}} \\
&\leq c\Big[r\sqrt{nH}\mathbb{E}[\|g\|_{2,\beta}^2\big|_{g=g_{j^*}}]^{1/2} + \sqrt{r}q^*(\frac{H}{n})H\Big]. \qquad (4.6.12)
\end{aligned}$$

Insertion of (4.6.11) and (4.6.12) into (4.6.10) delivers

$$|\mathbb{E}D(\hat{f}) - \mathbb{E}\hat{D}_n(\hat{f})| \le crF^2\Big[q^*(\frac{H}{n})\frac{H}{n} + \sqrt{\frac{H}{n}}\mathbb{E}[\|g\|_{2,\beta}^2\big|_{g=g_{j^*}}]^{1/2}\Big]. \quad (4.6.13)$$

By Lemma 4.6.1,

$$\|g\|_{2,\beta} \le C_\beta\|g\|_{\phi,2} \le C_\beta\varphi(\|g\|_1),$$

where $\varphi(x)^2 = \phi^{-1}(x^{-1})^{-1}$ is concave. Thus by Jensen's inequality and due to the fact that φ is concave (therefore subadditive),

$$\begin{aligned}\mathbb{E}[\|g\|_{2,\beta}^2\big|_{g=g_{j^*}}]^{1/2} &\le C_\beta\mathbb{E}[\varphi(\|g\|_1\big|_{g=g_{j^*}})^2]^{1/2} \le C_\beta\varphi(\mathbb{E}[\|g\|_1\big|_{g=g_{j^*}}])\\ &\le C_\beta\varphi(F^{-2}\mathbb{E}D(f_{j^*}))\\ &\le C_\beta[\varphi(2F^{-2}\delta^2) + \varphi(2F^{-2}\mathbb{E}D(\hat{f}))].\end{aligned}$$

Insertion into (4.6.13) yields

$$\big|\mathbb{E}D(\hat{f}) - \mathbb{E}\hat{D}_n(\hat{f})| \le cF^2r\Big[q^*(\frac{H}{n})\frac{H}{n} + C_\beta\sqrt{\frac{H}{n}}[\varphi(2F^{-2}\delta^2) + \varphi(2F^{-2}\mathbb{E}D(\hat{f}))]\Big].$$

By Lemma 4.6.10,

$$\begin{aligned}\mathbb{E}D(\hat{f}) \;\le\; &(1+\eta)\Big[\mathbb{E}\hat{D}_n(\hat{f}) + crF^2q^*(\frac{H}{n})\frac{H}{n} + crF^2C_\beta\sqrt{\frac{H}{n}}\varphi(2F^{-2}\delta^2)\Big]\\ &+\frac{\eta F^2}{2}(\varphi^{-1})^*\Big(2\frac{1+\eta}{\eta F^2}rC_\beta\sqrt{\frac{H}{n}}\Big).\end{aligned}$$

□

Lemma 4.6.6. *Suppose that Assumption 4.1.1, 4.1.2 hold true. Let each $f \in \mathcal{F}$ satisfy $\sup_{x\in supp(\mathcal{W})}|f(x)|_\infty \le F$. Then there exists an universal constant $c > 0$ such that for any $\delta \in (0,1)$,*

$$\begin{aligned}&\Big|\mathbb{E}\Big[\frac{1}{nd}\sum_{i=r+1}^n\langle\varepsilon_i, \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1})\Big]\Big|\\ \le\;& C_\varepsilon\delta + cC_\varepsilon C_{\beta,sub}rF\Big[q^*(\frac{H}{n})\frac{H}{n} + \sqrt{\frac{H}{n}}(\mathbb{E}[D(\hat{f})]^{1/2} + \delta)\Big].\end{aligned}$$

Proof of Lemma 4.6.6. Let $(f_j)_{j=1,\dots,\mathcal{N}_n}$ denote a δ-covering of $\mathcal{F}$ w.r.t. $\|\cdot\|_\infty$. Let $j^* \in \{1,...,\mathcal{N}_n\}$ be such that $\|\hat{f} - f_{j^*}\|_\infty \le \delta$. Let $\mathbb{H} = \mathbb{H}(\delta) = \log\mathbb{N}(\delta,\mathcal{F},\|\cdot\|_\infty)$. Since ε_i is independent of $\mathbb{X}_{i-1}$ and $\mathbb{E}\varepsilon_i = 0$, we have

$$\Big|\mathbb{E}\Big[\frac{1}{nd}\sum_{i=r+1}^n\langle\varepsilon_i, \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1})\Big]\Big| \le \delta\cdot\underbrace{\frac{1}{nd}\sum_{i=r+1}^n\mathbb{E}|\varepsilon_i|_1}_{\le\frac{1}{d}\sum_{k=1}^d\mathbb{E}|\varepsilon_{1k}|\le C_\varepsilon} + \frac{F}{n}|\mathbb{E}M_n(g_{j^*})| \quad (4.6.14)$$

where $g_j(x) = \frac{1}{F}(f_j(x) - f_0(x))\mathcal{W}(x)$ and M_n is from Lemma 4.6.3. By the same Lemma 4.6.3, there exists a process $M_n^\circ(\cdot)$ and some universal constant $c > 0$ such that

$$\frac{1}{n}|\mathbb{E}\{M_n(g_{j^*}) - M_n^\circ(g_{j^*})\}| \leq \frac{1}{n}\mathbb{E}\sup_{g\in\mathcal{G}}|M_n(g) - M_n^\circ(g)| \leq crC_\varepsilon C_{\beta,sub}q^*(\frac{H}{n})\frac{H}{n}. \quad (4.6.15)$$

Define $N(g) := \|\frac{1}{\sqrt{d}}|g(\mathbb{X}_r)|_2\|_2 \vee q^*(\frac{H}{n})\sqrt{\frac{H}{n}}$. Note that

$$\begin{aligned}
\mathbb{E}|M_n^\circ(g_{j^*})| &= \mathbb{E}\Big|\frac{M_n^\circ(g_{j^*})}{N(g_{j^*})}\cdot N(f_{j^*})\Big| \\
&\leq \Big\|\frac{M_n^\circ(g_{j^*})}{N(g_{j^*})}\Big\|_2 \underbrace{\mathbb{E}[\|\frac{1}{\sqrt{d}}|g(\mathbb{X}_r)|_2\|_2^2\big|_{g=g_{j^*}}]^{1/2}}_{=F^{-1}\mathbb{E}[D(f_{j^*})]^{1/2}} + \mathbb{E}\Big|\frac{M_n^\circ(g_{j^*})}{N(g_{j^*})}\Big|\cdot q^*(\frac{H}{n})\sqrt{\frac{H}{n}} \\
&\leq \mathbb{E}\Big[\sup_{j=1,\ldots,\mathcal{N}_n}\Big|\frac{M_n^\circ(g_j)}{N(g_j)}\Big|^2\Big]^{1/2}\cdot F^{-1}(\mathbb{E}[D(\hat{f})]^{1/2} + \delta) \\
&\qquad + \mathbb{E}\Big[\sup_{j=1,\ldots,\mathcal{N}_n}\Big|\frac{M_n^\circ(g_j)}{N(g_j)}\Big|\Big]\cdot q^*(\frac{H}{n})\sqrt{\frac{H}{n}}.
\end{aligned}$$

By Lemma 4.6.3, there exists some universal constant $c > 0$ such that

$$\mathbb{E}|M_n^\circ(g_{j^*})| \leq cC_\varepsilon\Big[\sqrt{nH}\cdot F^{-1}(\mathbb{E}[D(\hat{f})]^{1/2} + \delta) + q^*(\frac{H}{n})H\Big]. \quad (4.6.16)$$

Insertion of (4.6.15) and (4.6.16) into (4.6.14) yields

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\frac{1}{nd}\sum_{i=1}^n\langle\varepsilon_i, \hat{f}(X_{i-1})\rangle\mathcal{W}(X_{i-1})\Big| \\
\leq\ & C_\varepsilon\delta + 2cC_\varepsilon\Big[FrC_{\beta,sub}q^*(\frac{H}{n})\frac{H}{n} + \sqrt{\frac{H}{n}}(\mathbb{E}[D(\hat{f})]^{1/2} + \delta)\Big].
\end{aligned}$$

□

4.6.4 Auxiliary results for mixing time series

Variance bound for mixing

Proofs of Lemma 4.6.1. We only have to prove the last inequality. Note that for any $c > 0$, due to monotonicity of $x \mapsto \frac{\phi(x)}{x}$,

$$\mathbb{E}\phi\Big((\frac{|h(X_1)|}{c})^2\Big) = \mathbb{E}\Big[\frac{\phi\Big((\frac{|h(X_1)|}{c})^2\Big)}{|h(X_1)|}\cdot|h(X_1)|\Big] \leq \frac{\phi\Big((\frac{\|h\|_\infty}{c})^2\Big)}{\|h\|_\infty}\cdot\|h\|_1.$$

This upper bound attains the value 1 for

$$c = \|h\|_\infty\cdot\phi^{-1}(\frac{\|h\|_\infty}{\|h\|_1})^{-1/2},$$

which shows

$$\|h\|_{\phi,2} \le \|h\|_\infty \cdot \phi^{-1}(\frac{\|h\|_\infty}{\|h\|_1})^{-1/2}. \tag{4.6.17}$$

The result (4.6.1) now follows from $\|h\|_\infty \le 1$. □

During the proofs for the oracle inequalities under mixing, there will occur two quantities:

$$(\varphi^{-1})^*(\sqrt{x}) \qquad \text{and} \qquad q^*(x)x, \tag{4.6.18}$$

where $q^*(x) = q^{*,mix}(x) = \min\{q \in \mathbb{N} : \beta^{mix}(q) \le qx\}$. For x, we have to plug in a specific rate of the form $\frac{H}{n}$. It is therefore of interest to upper bound both quantities in (4.6.18) by one common quantity.

Recall the definitions $\psi(x) := \phi^*(x)x$ and $\Lambda(x) := \lceil \psi^{-1}(x^{-1}) \rceil x$ from (4.1.6).

In Lemma 4.6.7, we show that $(\varphi^{-1})^*(\sqrt{x})$ is upper bounded by a constant times $\Lambda(x)$. Lemma 4.6.8 shows that $q^*(x)x$ is upper bounded by a constant times $\Lambda(x)$. Thus, $\Lambda(x)$ serves as a common upper bound for both quantities in (4.6.18).

Unification theory for (4.6.18)

Lemma 4.6.7. *Let Assumption 4.1.2 hold true. Then $\varphi(x) = \phi^{-1}(\frac{1}{x})^{-1/2}$ and $\psi(x) = \phi^*(x)x$ satisfy:*

(i) for any $C \ge 1$, $(\varphi^{-1})^*(Cx) \le C^2(\varphi^{-1})^*(x)$,

(ii) φ^2, φ are concave and thus subadditive,

(iii) $(\varphi^{-1})^*(\sqrt{x}) \le (4c_0)^2\psi^{-1}(\frac{1}{x})x \le (4c_0)^2\Lambda(x)$.

Proof of Lemma 4.6.7. (i) Since $y \mapsto \frac{\phi(y)}{y}$ is increasing, $\frac{\phi(y)}{y} \le \frac{\phi(C^2y)}{C^2y}$. Thus, $\frac{1}{\phi(y)} \ge \frac{C^2}{\phi(C^2y)}$. We obtain

$$\begin{aligned}
(\varphi^{-1})^*(Cx) &= \sup_{z>0}\{Cxz - \phi(\frac{1}{z^2})^{-1}\} \\
&= C^2 \sup_{z>0}\{x\frac{z}{C} - \frac{\frac{1}{C^2}}{\phi(\frac{1}{z^2})}\} \\
&\le C^2 \sup_{z>0}\{x\frac{z}{C} - \frac{1}{\phi(\frac{C^2}{z^2})}\} \\
&\overset{u:=\frac{z}{C}}{=} C^2 \sup_{u>0}\{xu - \phi(\frac{1}{u^2})^{-1}\} = C^2(\varphi^{-1})^*(x).
\end{aligned}$$

(ii) We have

$$(\varphi^2)^{-1}(y) = \phi(\frac{1}{y})^{-1}. \tag{4.6.19}$$

By assumption, $y \mapsto \frac{y}{\phi(y)}$ is convex. Since $x \mapsto f(\frac{1}{x})$ is convex on $(0,\infty)$ if and only if $x \cdot f(x)$ is convex on $(0,\infty)$ (cf. [Baricz, 2012, page 1]), we obtain that

$y \mapsto \phi(\frac{1}{y})^{-1}$ is convex. By (4.6.19), its inverse φ^2 is concave. By concavity of $\sqrt{\cdot}$, φ is concave. Since concavity implies subadditivity, the claim follows.

(iii) *First Claim:* If $f : [0, \infty) \to \mathbb{R}$ is a convex function with $f(0) = 0$ and $w_0 \in [0, \infty)$ is such that $xw_0 - f(w_0) \leq 0$, then

$$f^*(x) \leq w_0 f'(w_0).$$

Proof: $F : [0, \infty) \to \mathbb{R}$, $F(w) = xw - f(w)$ is concave with $F(0) = 0$, $F(w_0) \leq 0$. Thus, F attains its global maximum in $[0, w_0]$. Since F is concave, the tangent $t(w) := F'(w_0)(w - w_0) + F(w_0)$ at w_0 satisfies

$$\begin{aligned} f^*(x) &= \sup_{w>0} F(w) \leq \sup_{w>0} t(w) = t(0) = -F'(w_0)w_0 + F(w_0) \\ &= -(x - f'(w_0))w_0 + (xw_0 - f(w_0)) \\ &= f'(w_0)w_0 - f(w_0) \leq f'(w_0)w_0. \end{aligned}$$

This proves the claim.

Second Claim: If $f : [0, \infty) \to \mathbb{R}$ is convex with $f(0) = 0$, then $f(x) \leq f'(x)x$ for all $x > 0$.

Proof: Since f is convex, the tangent $t(y) = f'(x)(y - x) + f(x)$ at x satisfies $t(0) \leq f(0) = 0$, which gives the result.

Let $y_0(x) := (\phi')^{-1}(c_0 x)$. Then,

$$x = \frac{1}{c_0}\phi'(y_0(x)) \leq \frac{\phi(y_0(x))}{y_0(x)}.$$

Application of the first claim to ϕ^* and $y_0(x)$ yields

$$\psi(x) = \phi^*(x)x \leq xy_0(x) = c_0 x^2 (\phi')^{-1}(c_0 x) =: g(x).$$

Since $g, \psi : [0, \infty) \to [0, \infty)$ are strictly increasing, we conclude that for any $y \in [0, \infty)$,

$$g^{-1}(y) = \psi^{-1}(\psi(g^{-1}(y))) \leq \psi^{-1}(g(g^{-1}(y))) = \psi^{-1}(y).$$

Especially, we obtain for any $x > 0$,

$$\psi^{-1}(\frac{1}{x})x \geq g^{-1}(\frac{1}{x})x. \tag{4.6.20}$$

As in (ii), we obtain that $y \mapsto \phi(\frac{1}{y})^{-1}$ is convex. Additionally, it is increasing, whence $y \mapsto \phi(\frac{1}{y^2})^{-1}$ is convex. Moreover, for all $z > 0$,

$$z \cdot \partial_z(\phi(\frac{1}{z^2})^{-1}) = \frac{2\phi'(\frac{1}{z^2})\frac{1}{z^2}}{\phi(\frac{1}{z^2})^2} \leq 2c_0\phi(\frac{1}{z^2})^{-1}. \tag{4.6.21}$$

Choose $z_0(x) > 0$ such that

$$\sqrt{x}z_0(x) - \phi(\frac{1}{z_0(x)^2})^{-1} = 0. \tag{4.6.22}$$

We obtain from the first claim and (4.6.21) that

$$\begin{aligned}(\varphi^{-1})^*(\sqrt{x}) &= \sup_{y>0}\{xy - \phi(\frac{1}{y^2})^{-1}\} \le z_0(x)\partial_z(\phi(\frac{1}{z^2})^{-1})\big|_{z=z_0(x)} \\ &\le 2c_0\phi(\frac{1}{z_0(x)^2})^{-1} \\ &= 2c_0 \cdot \sqrt{x}z_0(x). \end{aligned} \tag{4.6.23}$$

By the second claim, we obtain

$$\sqrt{x} = \frac{\phi(\frac{1}{z_0(x)^2})^{-1}}{z_0(x)} \le \partial_z(\phi(\frac{1}{z^2})^{-1})\big|_{z=z_0(x)} = \frac{2\phi'(\frac{1}{z_0(x)^2})}{\phi(\frac{1}{z_0(x)^2})^2 z_0(x)^3}. \tag{4.6.24}$$

By (4.6.22) and (4.6.24),

$$\frac{z_0(x)}{\sqrt{x}} = \sqrt{x}z_0(x)^3\phi(\frac{1}{z_0(x)^2})^2 \le 2\phi'(\frac{1}{z_0(x)^2}).$$

Thus,

$$g(\frac{1}{c_0}\frac{z_0(x)}{2\sqrt{x}}) = \frac{1}{c_0}(\frac{z_0(x)}{2\sqrt{x}})^2 \cdot (\phi')^{-1}\Big(\frac{z_0(x)}{2\sqrt{x}}\Big) \le \frac{1}{c_0}(\frac{z_0(x)}{2\sqrt{x}})^2 \cdot \frac{1}{z_0(x)^2} = \frac{1}{4c_0}\frac{1}{x}.$$

Since g is increasing,

$$g^{-1}(\frac{1}{4c_0}\frac{1}{x}) \ge \frac{1}{c_0}\frac{z_0(x)}{2\sqrt{x}}, \quad \text{and thus} \quad g^{-1}(\frac{1}{4c_0}\frac{1}{x})x \ge \frac{1}{2c_0}\sqrt{x}z_0(x).$$

By (4.6.20) and (4.6.23), g is increasing. Due to the fact that $c_0 \ge 1$ (see the second claim), we have

$$\begin{aligned}(\varphi^{-1})^*(\sqrt{x}) &\le 2c_0 \cdot \sqrt{x}z_0(x) \\ &\le (4c_0)^2 g^{-1}(\frac{1}{4c_0}\frac{1}{x})x \le (4c_0)^2 g^{-1}(\frac{1}{x})x \le (4c_0)^2\psi^{-1}(\frac{1}{x})x.\end{aligned}$$

□

Lemma 4.6.8. *Let Assumption 4.1.2 hold true. Then,*

$$q^*(x)x \le 2C\Lambda(x)$$

where $C \le \sum_{k=0}^{\infty}\{\phi^*(k+1) - \phi^*(k)\}\beta(k)$.

Proof of Lemma 4.6.8. Since $\beta^{-1}(u) = \sum_{i=0}^{\infty} \mathbb{1}_{\{u<\beta(i)\}}$ and $\beta(0) = 1$, we have

$$\begin{aligned}
\int_0^1 \phi^*(\beta^{-1}(u))du &= \sum_{i=0}^{\infty} \int_{\beta(i+1)}^{\beta(i)} \phi^*(i+1)du = \sum_{i=0}^{\infty} (\beta(i) - \beta(i+1))\phi^*(i+1) \\
&= \sum_{i=0}^{\infty} \sum_{k=0}^{i} \{\phi^*(k+1) - \phi^*(k)\}(\beta(i) - \beta(i+1)) \\
&= \sum_{k=0}^{\infty} \{\phi^*(k+1) - \phi^*(k)\} \cdot \sum_{i=k}^{\infty} (\beta(i) - \beta(i+1)) \\
&= \sum_{k=0}^{\infty} \{\phi^*(k+1) - \phi^*(k)\}\beta(k) < \infty. \qquad (4.6.25)
\end{aligned}$$

Let Z be a nonnegative $\mathbb{N}_0$-valued random variable with $\mathbb{P}(Z \geq k) = \beta(k)$. Then, $\mathbb{P}(Z = k) = \beta(k) - \beta(k+1)$, so (4.6.25) shows that $C := \mathbb{E}\phi^*(Z) \leq \mathbb{E}\phi^*(Z+1) < \infty$. Markov's inequality implies

$$\beta(k) = \mathbb{P}(Z \geq k) \leq \frac{\mathbb{E}\phi^*(Z)}{\phi^*(k)},$$

that is, $\beta(k)\phi^*(k) \leq C$. We then obtain

$$\begin{aligned}
q^*(x) &= \min\{q \in \mathbb{N} : \frac{\beta(q)}{q} \leq x\} \leq \min\{q \in \mathbb{N} : \frac{C}{\phi^*(q)q} \leq x\} \\
&= \min\{q \in \mathbb{N} : Cx^{-1} \leq \phi^*(q)q\} \leq \lceil \psi^{-1}(Cx^{-1}) \rceil,
\end{aligned}$$

whence

$$q^*(x)x \leq \lceil \psi^{-1}(Cx^{-1}) \rceil x. \qquad (4.6.26)$$

Since ϕ^* is increasing, $\psi(Cx) = Cx\phi^*(Cx) \geq Cx\phi^*(x) = C\psi(x)$. This implies for any $y > 0$ and $z := \psi^{-1}(y)$,

$$\psi^{-1}(Cy) = \psi^{-1}(C\psi(z)) \leq \psi^{-1}(\psi(Cz)) \leq Cz = C\psi^{-1}(y).$$

From (4.6.26) we obtain

$$q^*(x)x \leq \lceil C\psi^{-1}(x^{-1}) \rceil x \leq \lceil C \lceil \psi^{-1}(x^{-1}) \rceil \rceil x \leq 2C \lceil \psi^{-1}(x^{-1}) \rceil x.$$

□

The following proof shows the announced special forms for $\Lambda = \Lambda^{mix}$ in Lemma 4.1.4.

Proof of Lemma 4.1.4. (i) Let $\phi(x) = x^{\frac{\alpha}{\alpha-1}}$ with $\alpha > 1$. Then obviously, Assumption 4.1.2 (i), (ii) are fulfilled for $c_0 = \frac{\alpha}{\alpha-1}$. Furthermore,

$$\phi^*(x) = \sup_{y>0}\{xy - \phi(y)\} = C_\alpha x^\alpha, \qquad C_\alpha := (1 - \frac{1}{\alpha})^\alpha \cdot \frac{1}{\alpha - 1},$$

which implies $\phi^*(k+1) - \phi^*(k) = O(k^{\alpha-1})$ and thus proves $\sum_{k=0}^\infty(\phi^*(k+1) - \phi^*(k))\beta(k) = O(\sum_{k=0}^\infty k^{\alpha-1}\beta(k)) < \infty$.

In this case, we have $\psi(x) = \phi^*(x)x = C_\alpha x^{\alpha+1}$, $\psi^{-1}(x) = (C_\alpha^{-1}x)^{\frac{1}{\alpha+1}}$ and

$$\Lambda(x) = \lceil \psi^{-1}(x^{-1})\rceil x = \lceil C_\alpha^{-\frac{1}{\alpha+1}} x^{-\frac{1}{\alpha+1}}\rceil x.$$

For $x > 1$, the above quantity is bounded by $2C_\alpha^{-\frac{1}{\alpha+1}}x$; for $x < 1$, it is bounded by $2C_\alpha^{-\frac{1}{\alpha+1}}x^{-\frac{1}{\alpha}}x$. This yields the result with $c_\alpha = 2C_\alpha^{-\frac{1}{\alpha+1}}$.

(ii) Let $a := \frac{\rho+1}{2\rho} > 1$. Then $a\rho < 1$. Define $\phi(x) = x\frac{\log(x+1)}{\log(a)}$. Obviously, Assumption 4.1.2 (i), (ii) are fulfilled with $c_0 = 2$. Furthermore, by the first claim in the proof of Lemma 4.6.7 applied to $w_0 := a^x - 1$,

$$\phi^*(x) \le w_0 \cdot \phi'(w_0) \le 2\phi(w_0) = 2x\{a^x - 1\}.$$

On the other hand, for $x \ge 1$, $0 \le w_1 := a^{x-1} - 1$, thus

$$\phi^*(x) \ge xw_1 - \phi(w_1) = a^{x-1} - 1. \tag{4.6.27}$$

We obtain

$$\begin{aligned}\sum_{k=1}^\infty(\phi^*(k+1) - \phi^*(k))\beta(k) &\le \sum_{k=1}^\infty \big(2(k+1)(a^{k+1}-1) - a^{k-1} + 1)\kappa\rho^k \\ &= O(\sum_{k=1}^\infty k(\rho a)^k) < \infty.\end{aligned}$$

To upper bound the rate function, we use (4.6.27) to obtain

$$\psi(x) = \phi^*(x)x \ge x(a^{x-1} - 1) =: g(x)$$

where $g : [1,\infty) \to [0,\infty)$ is bijective. Thus, for any $y \ge 0$ we obtain

$$\psi^{-1}(y) = \psi^{-1}(g(g^{-1}(y))) \le \psi^{-1}(\psi(g^{-1}(y))) = g^{-1}(y).$$

We conclude that $\Lambda(x) \le \lceil g^{-1}(\frac{1}{x})\rceil x$. Here,

$$g(\frac{\log(2a(y\vee e))}{\log(a)}) = \frac{\log(2a(y\vee e))}{\log(a)}(\frac{2a(y\vee e)}{a} - 1) \ge 2(y\vee e) - 1 \ge y \vee e.$$

Thus for $y \ge e$,

$$g^{-1}(y) \le \frac{\log(2ay)}{\log(a)} \le 2 + \frac{\log(y)}{\log(a)}.$$

We obtain for $x \le e^{-1}$,

$$\Lambda(x) \le 2\Big(2 + \frac{\log(x^{-1})}{\log(a)}\Big)x. \tag{4.6.28}$$

Note that for $y \leq e$, $c := 1 + \frac{e}{a-1}$ satisfies $\phi^*(c) = \sup_{y>0}\{cy - \phi(y)\} \overset{y=a-1}{=} (a-1)[c - \frac{\log((a-1)+1)}{\log(a)}] = e$ and $\psi(c) = \phi^*(c)c \geq ec \geq e \geq y$. Therefore,

$$\psi^{-1}(y) \leq c.$$

So, for $x \geq \frac{1}{e}$,

$$\Lambda(x) = \lceil \psi^{-1}(x^{-1}) \rceil x \leq 2cx. \tag{4.6.29}$$

A combination of (4.6.28) and (4.6.29) gives the result.

□

Proofs of maximal inequalities under mixing

In this subsection, we prove auxiliary maximal inequalities under mixing. To do so, we use techniques and decomposition ideas from Dedecker and Louhichi [2002], Rio [1995] and Liebscher [1996].

Proof of Lemma 4.6.2. During the proof, let $q \in \{1, ..., n\}$ be arbitrary. Later we will choose $q = q^*(\frac{H}{n}) \leq n$. Note that

$$\begin{aligned} \beta^{\mathbb{X}}(k) &:= \beta(\sigma(\mathbb{X}_{i-1} : i \leq 0), \sigma(\mathbb{X}_{i-1} : i \geq k)) \\ &= \beta(\sigma(X_{i-1} : i \leq 0), \sigma(X_{i-1} : i \geq k - r + 1)) = \beta^X(k - r + 1) \leq \beta^X(k - r). \end{aligned}$$

We define $\tilde{\beta}(k) := \beta^X((k - r) \vee 0)$. Now, following Dedecker and Louhichi [2002], there exist random variables $\mathbb{X}_{i-1}^{\circ}$ with the following properties:

- for all $i \geq 0$, $U_{i-1}^{\circ} = (\mathbb{X}_{(i-1)q+1}^{\circ}, ..., \mathbb{X}_{(i-1)q+q}^{\circ})$ and $U_{i-1} = (\mathbb{X}_{(i-1)q+1}, ..., \mathbb{X}_{(i-1)q+q})$ have the same distribution,
- $(U_{2(i-1)}^{\circ})_{i\geq 1}$ and $(U_{2i}^{\circ})_{i\geq 1}$ are i.i.d.,
- for all $i \geq 1$, $\mathbb{P}(\mathbb{X}_{i-1} \neq \mathbb{X}_{i-1}^{\circ}) \leq \mathbb{P}(U_{i-1} \neq U_{i-1}^{\circ}) \leq \tilde{\beta}(q)$.

Put

$$S_n^{\circ}(g) := \sum_{i=r+1}^{n} \{g(\mathbb{X}_{i-1}^{\circ}) - \mathbb{E}g(\mathbb{X}_{i-1}^{\circ})\}.$$

Then,

$$|S_n(g) - S_n^{\circ}(g)| \leq 2\|g\|_{\infty} \Big| \sum_{i=r+1}^{n} (\mathbb{1}_{\{\mathbb{X}_{i-1} \neq \mathbb{X}_{i-1}^{\circ}\}} + \mathbb{P}(\mathbb{X}_{i-1} \neq \mathbb{X}_{i-1}^{\circ})) \Big|. \tag{4.6.30}$$

We now proceed with the proof of the announced inequalities. First, we have

$$\mathbb{E} \sup_{g \in \mathcal{G}} |S_n(g) - S_n^{\circ}(g)| \leq 4 \sum_{i=r+1}^{n} \mathbb{P}(\mathbb{X}_{i-1} \neq \mathbb{X}_{i-1}^{\circ}) = 4n\tilde{\beta}(q).$$

Let $\tilde{q}^*(x) = \min\{q \in \mathbb{N} : \tilde{\beta}(q) \le qx\}$. For $q = \tilde{q}^*(\frac{H}{n}) \le n$, we obtain

$$\mathbb{E} \sup_{g\in\mathcal{G}} |S_n(g) - S_n^\circ(g)| \le 4n\tilde{\beta}(q^*(\frac{H}{n})) \le 4n\tilde{q}^*(\frac{H}{n})\frac{H}{n}.$$

Now, let $\tilde{q} := q^*(x) + r$. Then,

$$\beta(\tilde{q} - r) = \beta(q^*(x)) \le q^*(x)x = (\tilde{q} - r)x \le \tilde{q}x.$$

This yields $\tilde{q}^*(x) \le \tilde{q} = q^*(x) + r \le rq^*(x)$. Finally,

$$\mathbb{E} \sup_{g\in\mathcal{G}} |S_n(g) - S_n^\circ(g)| \le 4n\tilde{q}^*(\frac{H}{n})\frac{H}{n} \le 4nrq^*(\frac{H}{n})\frac{H}{n},$$

which proves (4.6.2).

We now show (4.6.3) and (4.6.4). We have

$$S_n^\circ(g) = \sum_{k=r+1, k \text{ even}}^{\lfloor\frac{n}{q}\rfloor+1} Y_k^\circ(g) + \sum_{k=r+1, k \text{ odd}}^{\lfloor\frac{n}{q}\rfloor+1} Y_k^\circ(g)$$

where

$$Y_k^\circ(g) := \sum_{i=(k-1)q+1}^{kq\wedge n} \{g(\mathbb{X}_{i-1}^\circ) - \mathbb{E}g(\mathbb{X}_{i-1}^\circ)\}.$$

Furthermore, $(Y_k)_{k \text{ even}}$ and $(Y_k)_{k \text{ odd}}$ are independent with

$$\|\frac{Y_k^\circ(g)}{N(g)}\|_\infty \le 2qN(g)^{-1}, \qquad \|\frac{Y_k^\circ(g)}{N(g)}\|_2^2 \le \frac{1}{N(g)^2}\mathrm{Var}(\sum_{i=(k-1)q+1}^{kq\wedge n} g(\mathbb{X}_{i-1})) \le 4q\frac{\|g\|_{2,\tilde{\beta}}^2}{N(g)^2}.$$

Next,

$$\begin{aligned}
\|g\|_{2,\tilde{\beta}} &= \int_0^1 \tilde{\beta}^{-1}(u)Q_g(u)du = \sum_{i=1}^\infty \int_{\tilde{\beta}(i+1)}^{\tilde{\beta}(i)} iQ_g(u)du \\
&= \sum_{i=1}^\infty \int_{\beta((i-r+1)\vee 0)}^{\beta((i-r)\vee 0))} iQ_g(u)du \\
&= \sum_{i=r+1}^\infty \int_{\beta((i-r+1)\vee 0)}^{\beta((i-r)\vee 0))} iQ_g(u)du \\
&= \sum_{j=1}^\infty \int_{\beta(j+1)}^{\beta(j)} (j+r)Q_g(u)du \\
&\le r\sum_{j=1}^\infty \int_{\beta(j+1)}^{\beta(j)} jQ_g(u)du \\
&= r\,\|g\|_{2,\beta}.
\end{aligned}$$

Hence,

$$\|\frac{Y_k^\circ(g)}{N(g)}\|_2^2 \le 4q\frac{\|g\|_{2,\tilde\beta}^2}{N(g)^2} \le 4qr\frac{\|g\|_{2,\beta}^2}{N(g)^2} \le 4qr.$$

We obtain by Bernstein's inequality,

$$\begin{aligned}
&\mathbb{P}\Big(|\frac{S_n^\circ(g)}{N(g)}| > x\Big)\\
\le\ & \mathbb{P}\Big(|\sum_{k=r+1,k \text{ even}}^{\lfloor \frac{n}{q}\rfloor+1}\frac{Y_k^\circ(g)}{N(g)}| > \frac{x}{2}\Big) + \mathbb{P}\Big(|\sum_{k=r+1,k \text{ odd}}^{\lfloor \frac{n}{q}\rfloor+1}\frac{Y_k^\circ(g)}{N(g)}| > \frac{x}{2}\Big)\\
\le\ & 4\exp\Big(-\frac{1}{2}\frac{(x/2)^2}{4nr+2qN(g)^{-1}x/2}\Big)\\
\le\ & 4\exp\Big(-\frac{1}{32}\frac{x^2}{nr+qN(g)^{-1}x}\Big).
\end{aligned}\tag{4.6.31}$$

(i) Using standard arguments (cf. van der Vaart [1998], Lemma 19.35), we obtain from (4.6.31) that there exists a universal constant $c > 0$ such that

$$\mathbb{E}\sup_{g\in\mathcal{G}}|\frac{S_n^\circ(g)}{N(g)}| \le c\big[\sqrt{nrH} + qN(g)^{-1}H\big].$$

For $q = q^*(\frac{H}{n}) \le n$, we obtain

$$\mathbb{E}\sup_{g\in\mathcal{G}}|\frac{S_n^\circ(g)}{N(g)}| \le c\big[\sqrt{nrH} + q^*(\frac{H}{n})N(g)^{-1}H\big] \le 2c\sqrt{nrH}.$$

This shows (4.6.3).

(ii) Here, we use

$$\mathbb{E}\big[\sup_g|\frac{S_n^\circ(g)}{N(g)}|^2\big] = \int_0^\infty \mathbb{P}\Big(\sup_g|\frac{S_n^\circ}{N(g)}| > \sqrt{t}\Big)dt.$$

Put $a := q^*(\frac{H}{n})\sqrt{\frac{H}{n}}$. Choose $G := 64\frac{nr}{q}a$. Then for $t \ge G^2$, $\frac{q}{a}\sqrt{t} \ge nr$. With (4.6.31) and $\int_{b^2}^\infty \exp(-b_2\sqrt{t})dt = \int_b^\infty 2s\exp(-b_2 s)ds = 2(b_2 b+1)b_2^{-2}\exp(-b_2 b)$ we obtain

$$\begin{aligned}
&\int_0^\infty \mathbb{P}\Big(\sup_g |\frac{S_n^\circ(g)}{N(g)}| > \sqrt{t}\Big)dt \\
= \;& G^2 + \int_{G^2}^\infty \mathbb{P}\Big(\sup_g |\frac{S_n^\circ(g)}{N(g)}| > \sqrt{t}\Big)dt \\
\le \;& G^2 + 4|\mathcal{G}| \int_{G^2}^\infty \exp\Big(-\frac{1}{32}\frac{t}{nr + qN(g)^{-1}\sqrt{t}}\Big)dt \\
\le \;& G^2 + 4|\mathcal{G}| \int_{G^2}^\infty \exp\Big(-\frac{1}{32}\frac{t}{nr + qa^{-1}\sqrt{t}}\Big)dt \\
\le \;& G^2 + 4|\mathcal{G}| \int_{G^2}^\infty \exp(-\frac{1}{64}\frac{\sqrt{t}}{qa^{-1}})dt \\
\le \;& G^2 + 8|\mathcal{G}|\Big(\frac{Ga}{q}+1\Big)(64qa^{-1})^2\exp(-\frac{1}{64}\frac{Ga}{q}) \\
\le \;& 2^{15}\Big[\Big(\frac{nar}{q}\Big)^2 + |\mathcal{G}|\cdot\exp\Big(-nr(\frac{a}{q})^2\Big)\cdot(nr+(qa^{-1})^2)\Big]. \qquad (4.6.32)
\end{aligned}$$

We receive

$$\mathbb{E}\big[\sup_g |\frac{S_n^\circ(g)}{N(g)}|^2\big] \;\le\; 2^{16}\Big[\Big(\frac{nar}{q}\Big)^2 + |\mathcal{G}|\cdot\exp\Big(-nr(\frac{a}{q})^2\Big)\cdot(nr+(qa^{-1})^2)\Big].$$

With $q = q^*(\frac{H}{n})$, the latter is upper bounded by

$$2^{16}\Big[nr^2H + nr + \frac{n}{H}\Big] = 2^{18}nr^2H,$$

proving (4.6.4).

□

Proof of Lemma 4.6.3. Note that $(\varepsilon_i, \mathbb{X}_{i-1})$ is still β-mixing with coefficients $\tilde{\beta}(k) := \beta(k-r)$. This is due to the following argument: The model equation yields $X_i = f_0(\mathbb{X}_{i-1}) + \varepsilon_i$, that is, $\varepsilon_i = X_i - f_0(\mathbb{X}_{i-1})$. Thus, the generated sigma fields fulfill

$$\sigma((\varepsilon_i, \mathbb{X}_{i-1}) : i \le 0) = \sigma(X_i : i \le 0)$$

and

$$\sigma((\varepsilon_i, \mathbb{X}_{i-1}) : i \ge k) = \sigma(X_{i-1} : i \ge k-r+1) = \sigma(X_i : i \ge k-r).$$

Similar to the proof of Lemma 4.6.2, for each $q \in \{1, ..., n\}$ we can construct coupled versions $(\varepsilon_i^\circ, \mathbb{X}_{i-1}^\circ)$ of $(\varepsilon_i, \mathbb{X}_{i-1})$ and define

$$M_n^\circ(g) := \sum_{i=1}^n \frac{1}{d}\langle \varepsilon_i^\circ, g(\mathbb{X}_{i-1}^\circ)\rangle.$$

We will apply the following theory to $q = q^*(\frac{H}{n})^2$. Since $\sum_{q\in\mathbb{N}}\beta(q) < \infty$, $q^*(\frac{H}{n}) \le \sqrt{\frac{n}{H}}$ and hence $q = q^*(\frac{H}{n})^2 \le n$.

Now we have

$$\begin{aligned}|M_n(g) - M_n^\circ(g)| &\le \sum_{i=1}^n \frac{1}{d}\big(|\varepsilon_i^\circ|_2|g(\mathbb{X}_{i-1}^\circ)|_2 + |\varepsilon_i|_2|g(\mathbb{X}_{i-1})|_2\big)\mathbb{1}_{\{(\varepsilon_i,\mathbb{X}_{i-1})\neq(\varepsilon_i^\circ,\mathbb{X}_{i-1}^\circ)\}}\\ &\le 2\sum_{i=1}^n \frac{1}{\sqrt{d}}(|\varepsilon_i|_2 + |\varepsilon_i^\circ|_2)\mathbb{1}_{\{(\varepsilon_i,\mathbb{X}_{i-1})\neq(\varepsilon_i^\circ,\mathbb{X}_{i-1}^\circ)\}}. \qquad (4.6.33)\end{aligned}$$

With (4.6.33) and the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}[\sup_{g\in\mathcal{G}}|M_n(g) - M_n^\circ(g)|] \le 2n\frac{1}{\sqrt{d}}\||\varepsilon_1|_2\|_2\|\mathbb{1}_{\{(\varepsilon_i,\mathbb{X}_{i-1})\neq(\varepsilon_i^\circ,\mathbb{X}_{i-1}^\circ)\}}\|_2 \le 4nC_\varepsilon\tilde{\beta}(q)^{1/2}. \qquad (4.6.34)$$

With $q = q^*(\frac{H}{n})^2$ and (4.6.34) we derive

$$\begin{aligned}\mathbb{E}[\sup_{g\in\mathcal{G}}|M_n(g) - M_n^\circ(g)|] &\le 4nC_\varepsilon\tilde{\beta}\Big(q^*(\frac{H}{n})^2\Big)^{1/2} \le 4nC_\varepsilon C_{\beta,sub}(\tilde{\beta}(q^*(\frac{H}{n}))^2)^{1/2}\\ &\le 4nC_\varepsilon C_{\beta,sub}\tilde{q}^*(\frac{H}{n})\frac{H}{n}\end{aligned}$$

where $\tilde{q}^*(x) = \min\{q\in\mathbb{N} : \tilde{\beta}(q) \le qx\}$. By a similar argument as discussed in Lemma 4.6.2,

$$\mathbb{E}[\sup_{g\in\mathcal{G}}|M_n(g) - M_n^\circ(g)|] \le 4nrC_\varepsilon C_{\beta,sub}q^*(\frac{H}{n})\frac{H}{n},$$

which shows (4.6.5).

We now show (4.6.6) and (4.6.7). We first decompose

$$M_n^\circ(g) = \sum_{k=r+1,\ k \text{ even}}^{\lfloor\frac{n}{q}\rfloor+1} Y_k^\circ(g) + \sum_{k=r+1,\ k \text{ odd}}^{\lfloor\frac{n}{q}\rfloor+1} Y_k^\circ(g) \qquad (4.6.35)$$

where

$$Y_k^\circ(g) := \sum_{i=(k-1)q+1}^{kq\wedge n}\frac{1}{d}\langle\varepsilon_i^\circ, g(\mathbb{X}_{i-1}^\circ)\rangle$$

are independent. Since $(\langle\varepsilon_i, g(\mathbb{X}_{i-1})\rangle)_i$ is a martingale, by [Rio, 2009, Theorem 2.1] we have

$$\begin{aligned}\|Y_k^\circ(g)\|_m &\le (m-1)^{1/2}\Big(\sum_{i=(k-1)q+1}^{kq\wedge n}\|\frac{1}{d}\langle\varepsilon_i^\circ, g(\mathbb{X}_{i-1}^\circ)\rangle\|_m^2\Big)^{1/2}\\ &\le (m-1)^{1/2}q^{1/2}\frac{1}{d}\|\langle\varepsilon_1^\circ, g(\mathbb{X}_r^\circ)\rangle\|_m. \qquad (4.6.36)\end{aligned}$$

Due to independence we get

$$\Big(\frac{1}{d}\|\langle\varepsilon_1^\circ, g(\mathbb{X}_r^\circ)\rangle\|_m\Big)^m \le \Big\|\frac{|\varepsilon_1|_2}{\sqrt{d}}\Big\|_m^m \cdot \frac{1}{d}\mathbb{E}[|g(\mathbb{X}_r)|_2^2]\cdot\|g\|_\infty^{m-2}. \tag{4.6.37}$$

Furthermore,

$$\begin{aligned}\Big\|\frac{|\varepsilon_1|_2}{\sqrt{d}}\Big\|_m^m &\le \mathbb{E}\Big[\Big(\frac{1}{d}\sum_{j=1}^d \varepsilon_{1j}^2\Big)^{m/2}\Big] = \Big\|\frac{1}{d}\sum_{j=1}^d \varepsilon_{1j}^2\Big\|_{m/2}^{m/2} \le \Big(\frac{1}{d}\sum_{j=1}^d \|\varepsilon_{1j}^2\|_{m/2}\Big)^{m/2}\\ &\le \|\varepsilon_{11}\|_m^m \le C_\varepsilon^m m^{m/2}. \end{aligned}\tag{4.6.38}$$

Insertion of (4.6.38) into (4.6.37) and afterwards into (4.6.36) yields with $a := q^*(\frac{H}{n})\sqrt{\frac{H}{n}}$,

$$\begin{aligned}\|\frac{Y_k^\circ(g)}{N(g)}\|_m^m &\le (m-1)^{m/2}m^{m/2}\cdot(C_\varepsilon a^{-1}q^{1/2})^{m-2}\cdot qC_\varepsilon^2\mathbb{E}[\frac{\frac{1}{d}|g(\mathbb{X}_r)|_2^2}{\nu(g)^2}]\\ &\le \frac{m!}{2}\cdot 2e^2qC_\varepsilon^2(C_\varepsilon e\cdot a^{-1}q^{1/2})^{m-2}.\end{aligned}\tag{4.6.39}$$

By Bernstein's inequality for independent variables, we conclude from (4.6.39) that

$$\mathbb{P}\Big(\Big|\frac{1}{N(g)}\sum_{k=r+1,\ k\text{ even}}^{\lfloor\frac{n}{q}\rfloor+1} Y_k^\circ(g)\Big| > x\Big) \le 2\exp\Big(-\frac{1}{2}\frac{x^2}{\frac{n}{q}\cdot 2e^2C_\varepsilon^2 q + eC_\varepsilon q^{1/2}a^{-1}x}\Big).$$

Insertion into (4.6.35) yields

$$\begin{aligned}\mathbb{P}(|\frac{M_n^\circ(g)}{N(g)}| > x) &\le 4\exp\Big(-\frac{1}{2}\frac{(x/2)^2}{2e^2C_\varepsilon^2 n + eC_\varepsilon q^{1/2}a^{-1}(x/2)}\Big)\\ &\le 4\exp\Big(-\frac{1}{8}\frac{x^2}{2e^2C_\varepsilon^2 n + eC_\varepsilon q^{1/2}a^{-1}x}\Big).\end{aligned}\tag{4.6.40}$$

(i) Standard arguments (cf. [van der Vaart, 1998, Lemma 19.35]) applied to (4.6.40) state that there exists some universal constant $c > 0$ such that

$$\mathbb{E}\sup_{g\in\mathcal{G}}|\frac{M_n^\circ(g)}{N(g)}| \le cC_\varepsilon\Big[\sqrt{nH} + q^{1/2}a^{-1}H\Big]. \tag{4.6.41}$$

With $q = q^*(\frac{H}{n})^2$, we obtain

$$\mathbb{E}[\sup_{g\in\mathcal{G}}|M_n^\circ(g)|] \le cC_\varepsilon\Big[\sqrt{nH} + q^*(\frac{H}{n})a^{-1}H\Big] \le 2cC_\varepsilon\sqrt{nH},$$

which shows (4.6.6).

(ii) Here, we use

$$\mathbb{E}\big[\sup_g |\frac{M_n^\circ(g)}{N(g)}|^2\big] = \int_0^\infty \mathbb{P}\Big(\sup_g |\frac{M_n^\circ(g)}{N(g)}| > \sqrt{t}\Big)dt.$$

Choose $G := 16eC_\varepsilon \frac{na}{q^{1/2}}$. Then for $t \geq G^2$, $q^{1/2}a^{-1}\sqrt{t} \geq 2e^2C_\varepsilon^2 n$. With (4.6.40) and $\int_{b^2}^\infty \exp(-b_2\sqrt{t})dt = \int_b^\infty 2s\exp(-b_2 s)ds = 2(b_2 b+1)b_2^{-2}\exp(-b_2 b)$,

$$\begin{aligned}
& \int_0^\infty \mathbb{P}\Big(\sup_g |\frac{M_n^\circ(g)}{N(g)}| > \sqrt{t}\Big)dt \\
& G^2 + \int_{G^2}^\infty \mathbb{P}\Big(\sup_g |\frac{M_n^\circ(g)}{N(g)}| > \sqrt{t}\Big)dt \\
\leq\; & G^2 + 4|\mathcal{G}| \int_{G^2}^\infty \exp\Big(-\frac{1}{8}\frac{t}{2e^2C_\varepsilon^2 n + eC_\varepsilon q^{1/2}a^{-1}\sqrt{t}}\Big)dt \\
\leq\; & G^2 + 4|\mathcal{G}| \int_{G^2} \exp(-\frac{1}{16}\frac{\sqrt{t}}{eC_\varepsilon q^{1/2}a^{-1}})dt \\
\leq\; & G^2 + 8|\mathcal{G}|\Big(\frac{Ga}{eC_\varepsilon q^{1/2}} + 1\Big)(16eC_\varepsilon q^{1/2}a^{-1})^2 \exp(-\frac{1}{8}\frac{Ga}{eC_\varepsilon q^{1/2}}) \\
\leq\; & 2^{10}e^2C_\varepsilon^2\Big[\Big(\frac{na}{q^{1/2}}\Big)^2 + |\mathcal{G}|\cdot\exp\Big(-n(\frac{a}{q^{1/2}})^2\Big)\cdot(n + (q^{1/2}a^{-1})^2)\Big] \quad (4.6.42)
\end{aligned}$$

We obtain

$$\begin{aligned}
& \mathbb{E}\big[\sup_g |M_n^\circ(g)|^2\big] \\
\leq\; & 2^{10}e^2C_\varepsilon^2\Big[\Big(\frac{na}{q^{1/2}}\Big)^2 + |\mathcal{G}|\cdot\exp\Big(-n(\frac{a}{q^{1/2}})^2\Big)\cdot(n + (q^{1/2}a^{-1})^2)\Big].
\end{aligned}$$

For $q = q^*(\frac{H}{n})^2$ this implies

$$\begin{aligned}
\mathbb{E}\big[\sup_g |\frac{M_n^\circ(g)}{N(g)}|^2\big] \;\leq\; & 2^{10}e^2C_\varepsilon^2\Big[\Big(\frac{na}{q^*(\frac{H}{n})}\Big)^2 \\
& + |\mathcal{G}|\cdot\exp\Big(-n(\frac{a}{q^*(\frac{H}{n})})^2\Big)\cdot(n + (q^*(\frac{H}{n})a^{-1})^2)\Big] \\
\leq\; & 2^{10}e^2C_\varepsilon^2\Big[nH + n + \frac{n}{H}\Big] \leq 2^{11}e^2C_\varepsilon^2 nH,
\end{aligned}$$

proving (4.6.7).

□

Lemma 4.6.9. *Let $\tilde{\beta}(k) := \beta((k-r)\vee 0)$. Suppose that $\beta^{mix}(\cdot)$ is submultiplicative in the sense of (4.1.5). Then for any $q_1, q_2, r \in \mathbb{N}$ there exists $C_{\beta,sub}$, such that*

$$\tilde{\beta}(q_1q_2) \leq C_{\beta,sub}\tilde{\beta}(q_1)\tilde{\beta}(q_2).$$

Proof of Lemma 4.6.9. By case distinction it is elementary to prove

$$((q_1 - r) \vee 0)((q_2 - r) \vee 0) \leq (q_1 q_2 - r) \vee 0.$$

Since β is decreasing, we directly have

$$\begin{aligned}\tilde{\beta}(q_1 q_2) &= \beta((q_1 q_2 - r) \vee 0)\\ &\leq \beta(((q_1 - r) \vee 0)((q_2 - r) \vee 0))\\ &\leq C_{\beta,sub}\beta((q_1 - r) \vee 0)\beta((q_2 - r) \vee 0) = C_{\beta,sub}\tilde{\beta}(q_1)\tilde{\beta}(q_2)\end{aligned}$$

□

Auxiliary results for oracle inequalities under mixing

The following lemma is applied to $\varphi(x) = \phi^{-1}(\frac{1}{x})^{-1/2}$ in the proof of Theorem 4.6.4.

Lemma 4.6.10. *Let $r_1, r_2, b, P > 0$ and φ some concave function with $\varphi(0) = 0$. If $a \geq 0$ satisfies $|a - b| \leq r_1\varphi(r_2 a) + P$, then for any $\eta > 0$,*

$$a \leq \frac{\eta}{r_2}(\varphi^{-1})^*(\frac{1+\eta}{\eta} r_1 r_2) + (1+\eta)(b+P).$$

Proof of Lemma 4.6.10. The mapping $g(x) := \frac{\eta}{1+\eta}\frac{1}{r_2}\varphi^{-1}(x)$ is convex. By Young's inequality and denoting by g^* the convex conjugate of g,

$$r_1\varphi(r_2 a) \leq g^*(r_1) + g(\varphi(r_2 a)) = \frac{\eta}{(1+\eta)r_2}(\varphi^{-1})^*(\frac{1+\eta}{\eta} r_1 r_2) + \frac{\eta}{1+\eta}a.$$

We therefore have

$$a \leq |a-b| + b \leq r_1\varphi(a) + (b+P) \leq \frac{\eta}{(1+\eta)r_2}(\varphi^{-1})^*(\frac{1+\eta}{\eta} r_1 r_2) + \frac{\eta}{1+\eta}a + (b+P).$$

Rearranging the terms leads to

$$a \leq \frac{\eta}{r_2}(\varphi^{-1})^*(\frac{1+\eta}{\eta} r_1 r_2) + (1+\eta)(b+P).$$

□

4.6.5 Results for the functional dependence measure

Recall the definition of the functional dependence measure coefficients $\delta_q^X(k)$, $k \in \mathbb{N}_0$ from (4.1.8).

During the proofs for the oracle inequalities under functional dependence, there will occur two quantities:

$$(\tilde{V}^{-1})^*(\sqrt{x}) \qquad \text{and} \qquad q^*(\sqrt{x})x, \tag{4.6.43}$$

where

$$q^*(x) = q^{*,dep}(x) = \min\{q \in \mathbb{N} : \beta^{dep}(q) \leq qx\} \tag{4.6.44}$$

and

$$\beta^{dep}(q) = \sum_{j=q}^{\infty} \Delta(j). \tag{4.6.45}$$

Here, $\Delta(k)$, $k \in \mathbb{N}_0$, is an upper bound chosen dependent on the function class of interest and specified below. For x, we have to plug in a specific rate of the form $\frac{H}{n}$. It is therefore of interest to upper bound both quantities in (4.6.43) by one common quantity. We saw this approach already when we discussed mixing sequences.

Recall the definitions

$$\tilde{V}(x) = x^{1/2} + \sum_{j=0}^{\infty} \min\{x^{1/2}, \Delta(j)\}$$

and $\Lambda(x) = \sqrt{x}\bar{y}(x)$ as well as $\bar{y}(x)$ from (4.1.11) and (4.1.12).

In Lemma 4.6.11, we show that both terms in (4.6.43) are bounded by a constant times $\Lambda(x)$. Thus, $\Lambda(x)$ serves as a common upper bound for both quantities in (4.6.43). Besides that, we show in Lemma 4.6.13 that $\tilde{V}$ is a concave function which is needed to obtain meaningful upper bounds in Theorem 4.6.16, which was originally stated as Theorem 4.1.6.

Unification theory for (4.6.43)

Lemma 4.6.11. *Let $\tilde{V}'$ denote the left derivative of $\tilde{V}$.*

(i) Let $h, \delta \geq 0$. Then $\tilde{V}(h) \leq \delta$ implies $\sqrt{h} \leq r(\delta)$.

(ii) If there exists $C > 0$ such that for all $q \in \mathbb{N}$, $\beta^{dep}(q) \leq Cq \cdot \Delta(q)$, then

$$\inf_{x\in[0,\infty)} \frac{\tilde{V}'(x)x}{\tilde{V}(x)} \geq \frac{1}{2(1+C)}.$$

(iii) Under the assumptions of (ii),

$$(\tilde{V}^{-1})^*(\sqrt{x}) \leq 2(1+C) \cdot \Lambda(x), \qquad q^{*,dep}(\sqrt{x})x \leq \Lambda(x). \tag{4.6.46}$$

Proof of Lemma 4.6.11. (i) It can be shown as in the proof of Lemma 2.8.8 that for any $h \in [0, \infty)$,

$$\tilde{V}(h) = \sqrt{h} \cdot a^* + \beta^{dep}(a^*) \tag{4.6.47}$$

with some $a^* \in \mathbb{N}$ dependent on h.

Let $\delta > 0$. If $\tilde{V}(h) \leq \delta$, then $\beta^{dep}(a^*) \leq \delta$, that is, $\frac{\beta^{dep}(a^*)}{a^*} \leq \frac{\delta}{a^*}$. By definition of $q^{*,dep}$, $q^{*,dep}(\frac{\delta}{a^*}) \leq a^*$. This implies $q^{*,dep}(\frac{\delta}{a^*})\frac{\delta}{a^*} \leq \delta$ and thus by definition of $r(\cdot)$ and (4.6.47),

$$r(\delta) \geq \frac{\delta}{a^*} \geq \sqrt{h}.$$

(ii) Let $x \in [0,\infty)$. If $\Delta(N) < \sqrt{x} < \Delta(N-1)$ for some $N \in \mathbb{N}$, then

$$\tilde{V}(x) = x^{1/2} + \sum_{j=0}^{N-1} x^{1/2} + \sum_{j=N}^{\infty} \Delta(j) = (N+1)\sqrt{x} + \beta^{dep}(N),$$

and thus $x \cdot \partial_x \tilde{V}(x) = (N+1) \cdot \frac{1}{2} x^{1/2}$. By assumption, $\beta^{dep}(N) \leq N\Delta(N)$ and $\sqrt{x} > \Delta(N)$, implying

$$\frac{x \cdot \partial_x \tilde{V}(x)}{\tilde{V}(x)} \geq \frac{1}{2} \cdot \frac{(N+1)\sqrt{x}}{(N+1)\sqrt{x} + CN\Delta(N)} \geq \frac{1}{2} \frac{(N+1)\Delta(N)}{(N+1)\Delta(N) + CN\Delta(N)} \geq \frac{1}{2(1+C)}. \tag{4.6.48}$$

Writing the left derivative as a limit of $\partial_x \tilde{V}(x)$, the result follows.

(iii) Fix $x \geq 0$. Define $y_0(x) \in [0,\infty)$ as a solution of

$$\tilde{V}(\sqrt{x} y_0(x)) = y_0(x). \tag{4.6.49}$$

Since $z \mapsto \frac{\tilde{V}(z)}{z}$ is decreasing, $y_0(x) \leq \bar{y}(x)$. It is therefore enough to show that the quantities in (4.6.46) are bounded by multiples of $\sqrt{x} y_0(x) \leq \Lambda(x)$.

Let $\tilde{W}$ denote the right derivative of $\tilde{V}^{-1}$. By the First Claim in the proof of Lemma 4.6.7(iii) (which also holds true for left or right derivatives), we obtain

$$(\tilde{V}^{-1})^*(\sqrt{x}) \leq \tilde{W}(y_0) \cdot y_0. \tag{4.6.50}$$

For any $y \geq 0$ with $z = \tilde{V}^{-1}(y)$,

$$\frac{\tilde{W}(y) y}{\tilde{V}^{-1}(y)} \leq \frac{\tilde{V}(z)}{z \tilde{V}'(z)} \leq 2(1+C).$$

Insertion into (4.6.50) and using the definition of y_0 yields

$$(\tilde{V}^{-1})^*(\sqrt{x}) \leq \tilde{W}(y_0) \cdot y_0 \leq \tilde{V}^{-1}(y_0) = 2(1+C) y_0 \sqrt{x}.$$

By (4.6.49) and (i), we have

$$\sqrt{x} y_0(x) \leq r(y_0(x))^2.$$

Together with $r(\delta) \leq \delta$ for any $\delta \geq 0$, this implies

$$\sqrt{x} = \frac{\sqrt{x} y_0}{y_0} \leq \frac{r(y_0)^2}{y_0} \leq r(y_0).$$

Since $z \mapsto q^{*,dep}(z) z$ is increasing,

$$q^{*,dep}(\sqrt{x}) x \leq q^{*,dep}(r(y_0)) r(y_0) \sqrt{x} \leq y_0 \sqrt{x}.$$

□

Lemma 4.6.12. *For $c > 0$, $(\tilde{V}^{-1})^*(cx) \leq c^2(\tilde{V}^{-1})^*(x)$.*

Proof of Lemma 4.6.12. We have

$$\begin{aligned}
(\tilde{V}^{-1})^*(cx) &= \sup_{y>0}\{cxy - \tilde{V}^{-1}(y)\} \\
&= \sup_{z>0}\{cx\tilde{V}(z) - z\} \\
&= \sup_{w>0}\{cx\tilde{V}(c^2w) - c^2w\} \\
&\leq \sup_{w>0}\{c^2x\tilde{V}(w) - c^2w\} \\
&\leq c^2\sup_{w>0}\{x\tilde{V}(w) - w\} = c^2(\tilde{V}^{-1})^*(x)
\end{aligned}$$

due to

$$\tilde{V}(c'a) = \sum_{j=0}^{\infty}\min\{c'a, \Delta(j)\} \leq \sqrt{c'}\tilde{V}(a).$$

for $c' > 0$. □

The following lemma shows that $\tilde{V}(\cdot)$ defined in (4.1.10) is concave. This property is needed in the next section in order to get a good upper bound in the maximal inequalities.

Lemma 4.6.13. *Let $(a_k)_{k\in\mathbb{N}}$ be a decreasing nonnegative sequence of real numbers for which $\sum_{k=0}^{\infty} a_k < \infty$. Then,*

$$y \mapsto v(y) := \Big(\sum_{k=1}^{N}\min\{\sqrt{y}, a_k\}\Big)^2, \qquad N \in \mathbb{N}\cup\{\infty\},$$

is a concave map.

Proof of Lemma 4.6.13. It is obvious that v is concave on $y \in (a_j^2, a_{j-1}^2]$ because v can be represented as a sum of concave functions, namely

$$v(y) = \Big((j-1)\sqrt{y} + \sum_{k=j}^{N} a_k\Big)^2 = (j-1)^2y + 2(j-1)\sum_{k=j}^{N} a_k\sqrt{y} + (\sum_{k=j}^{N} a_k)^2.$$

We investigate the slope's behavior on the interval's open boundary. The derivative's left limit at a_j yields

$$\begin{aligned}
\lim_{y\to a_j^2, y<a_j^2}\partial_y v(y) = \lim_{y\to a_j^2, y<a_j^2}\frac{j}{\sqrt{y}}\Big(j\sqrt{y} + \sum_{k=j+1}^{N} a_k\Big) &= \frac{j}{a_j}\Big(ja_j + \sum_{k=j+1}^{N} a_k\Big) \\
&= \frac{j}{a_j}\Big((j-1)a_j + \sum_{k=j}^{N} a_k\Big).
\end{aligned}$$

On the other hand, the right limit is given by

$$\lim_{y\to a_j^2, y>a_j^2} \partial_y v(y) = \lim_{y\to a_j^2, y<a_j^2} \frac{j-1}{\sqrt{y}}\Big((j-1)\sqrt{y} + \sum_{k=j}^{N} a_k\Big) = \frac{j-1}{a_j}\Big((j-1)a_j + \sum_{k=j}^{N} a_k\Big).$$

Hence,

$$\lim_{y\to a_j^2, y<a_j^2} \partial_y f(y) \geq \lim_{y\to a_j^2, y>a_j^2} \partial_y f(y).$$

Since f is concave on intervals of the form $(a_j^2, a_{j-1}^2]$, the just proven inequality for the derivative implies that f has a representation as $f(y) = \int_0^y \tilde{v}(x)dx$. Now let $\lambda \in (0,1)$. Since

$$\begin{aligned}\tilde{v}(x+\lambda(y-x)) = \int_0^{x+\lambda(y-x)} \tilde{v}(z)dz &= \int_0^x \tilde{v}(z)dz + \int_x^{x+\lambda(y-x)} \tilde{v}(z)dz \\ &= \int_0^x \tilde{v}(z)dz + \int_0^{\lambda(y-x)} \tilde{v}(x+u)du.\end{aligned}$$

and

$$\begin{aligned}\tilde{v}(x) + \lambda(\tilde{v}(y) - \tilde{v}(x)) &= \int_0^x \tilde{v}(z)dz + \lambda\int_x^y \tilde{v}(z)dz \\ &= \int_0^x \tilde{v}(z)dz + \lambda\int_0^{x-y} \tilde{v}(x+u)du\end{aligned}$$

for $x, y > 0$, we conclude that v is concave.

The result for $N = \infty$ can be obtained since the limit of concave functions is concave. □

Here, we prove the upper bounds from Lemma 4.1.8 for $\Lambda(\cdot)$ which arise in the special cases of polynomial and exponential decay.

Proof of Lemma 4.1.8. (i) In Lemma 2.8.12 it was shown that

$$\tilde{V}(z) \leq C_{\kappa,\alpha} \cdot \max\{z^{\frac{1}{2}\frac{\alpha-1}{\alpha}}, z^{\frac{1}{2}}\},$$

where $C_{\kappa,\alpha} > 0$ is some constant only depending on κ, α. Fix $x \in [0, \infty)$. With $\bar{y}(x) = c\max\{x^{\frac{1}{2}\frac{\alpha-1}{\alpha+1}}, x^{\frac{1}{2}}\}$, we have

$$\sqrt{x}\bar{y}(x) = c\max\{x^{\frac{\alpha}{\alpha+1}}, x\}$$

and by case distinction $x > 1$, $x \leq 1$,

$$\begin{aligned}\tilde{V}(\sqrt{x}\bar{y}(x)) &\leq C_{\kappa,\alpha} \cdot \max\{c^{\frac{1}{2}\frac{\alpha-1}{\alpha}}, c^{\frac{1}{2}}\} \cdot \max\{x^{\frac{1}{2}\frac{\alpha-1}{\alpha+1}}, x^{\frac{1}{2}\frac{\alpha-1}{\alpha}}, x^{\frac{1}{2}\frac{\alpha}{\alpha+1}}, x^{\frac{1}{2}}\} \\ &\leq C_{\kappa,\alpha}\max\{c^{\frac{1}{2}\frac{\alpha-1}{\alpha}}, c^{\frac{1}{2}}\} \cdot \max\{x^{\frac{1}{2}\frac{\alpha-1}{\alpha+1}}, x^{\frac{1}{2}}\} = C_{\kappa,\alpha}\max\{c^{-\frac{1}{2}\frac{\alpha+1}{\alpha}}, c^{-\frac{1}{2}}\}\bar{y}(x).\end{aligned}$$

So, choosing $c = \max\{C_{\kappa,\alpha}^{\frac{2\alpha}{\alpha+1}}, C_{\kappa,\alpha}^2\}$ yields the result.

(ii) Lemma 2.8.12 yields

$$\tilde{V}(z) \leq C_{\kappa,\rho} \cdot z^{\frac{1}{2}} \log(z^{-1} \vee 1),$$

where $C_{\kappa,\rho} > 1$ is some constant only depending on κ, ρ. Fix $x \in [0, \infty)$. With $\bar{y}(x) = cx^{\frac{1}{2}} \log(x^{-1} \vee 1)^2$, we have

$$\sqrt{x}\bar{y}(x) = cx \log(x^{-1} \vee 1)^2$$

and by case distinction $x > 1$, $x \leq 1$,

$$\begin{aligned}
\tilde{V}(\sqrt{x}\bar{y}(x)) &\leq c^{\frac{1}{2}} \log(c^{-1} \vee e) C_{\kappa,\rho} \cdot \max\{x^{\frac{1}{2}} \log(x^{-1} \vee 1)(\log(x^{-1} \vee 1) - 2\log(\log(x^{-1} \vee 1) \vee 1)), \\
&\qquad x^{\frac{1}{2}} \log(x^{-1} \vee 1), x^{\frac{1}{2}}\} \\
&\leq \frac{1}{4} c^{\frac{1}{2}} \log(c^{-1} \vee e) C_{\kappa,\rho} \cdot \max\{x^{\frac{1}{2}} \log(x^{-1} \vee 1)^2, x^{\frac{1}{2}}\} \\
&= \frac{1}{4} c^{-\frac{1}{2}} \log(c^{-1} \vee e) C_{\kappa,\rho} \bar{y}(x).
\end{aligned}$$

So, choosing $c = 16 C_{\kappa,\rho}^2$ yields the result.

□

Maximal inequalities under functional dependence

The following empirical process results are based on the theory developed in Chapter 2.

Let $\mathcal{G} \subset \{g : \mathbb{R}^{dr} \to \mathbb{R} \text{ measurable}\}$ be a finite class of Lipschitz continuous functions in the sense that there exists $b \in (0, 1]$ such that for $x_j \in \mathbb{R}^d, x_j' \in \mathbb{R}^d$, $j = 1, ..., r$,

$$|g(x_1, ..., x_r) - g(x_1', ..., x_r')| \leq L_{\mathcal{G}} \cdot \max_{j=1,...,r} |x_j - x_j'|_\infty, \tag{4.6.51}$$

and for some $G > 0$,

$$\sup_{g \in \mathcal{G}} \|g\|_\infty \leq G. \tag{4.6.52}$$

For $\theta \in (0, 1]$,

$$\begin{aligned}
|g(x_1, ..., x_r) - g(x_1', ..., x_r')| &\leq \min\{2G, L_{\mathcal{G}} \max_{i=1,...,r} |x_i - x_i'|_\infty\} \\
&\leq (2G)^{1-\theta} L_{\mathcal{G}}^\theta (\max_{i=1,...,r} |x_i - x_i'|_\infty)^\theta.
\end{aligned}$$

Then, using the notation from (4.1.7) and (4.1.8),

$$\begin{aligned}
\delta_2^{g(\mathbb{X}_{\cdot-1})}(k) &= \|g(\mathbb{X}_{i-1}) - g(\mathbb{X}_{i-1}^{*(i-k)})\|_2 \\
&\leq (2G)^{1-\theta} L_{\mathcal{G}}^\theta \cdot \| \max_{j=1,...,r} |X_{i-j} - X_{i-j}^{*(i-k)}|_\infty^\theta \|_2 \\
&\leq dr(2G)^{1-\theta} L_{\mathcal{G}}^\theta \max_{j=\{1,...,r\}, l \in \{1,...,d\}} \|X_{i-j,l} - X_{i-j,l}^{*(i-k)}\|_{2\theta}^\theta \\
&\leq dr(2G)^{1-\theta} L_{\mathcal{G}}^\theta \cdot \sup_{l=1,...,r} \sup_{j=1,...,r} \delta_{2\theta}^{X_{\cdot,l}}(k-j)^\theta.
\end{aligned} \tag{4.6.53}$$

In the following, we suppose that $\Delta(k)$, $k \in \mathbb{N}_0$, is a decreasing sequence chosen such that

$$dr(2G)^{1-\theta}L_{\mathcal{G}}^{\theta} \cdot \sup_{l=1,...,r}\sup_{j=1,...,r} \delta_{2\theta}^{X_{\cdot,l}}(k-j)^{\theta} \le \Delta(k). \tag{4.6.54}$$

Recall the definition of $q^{*,dep}$ and β^{dep} from (4.6.44) and (4.6.45).

Put

$$S_n(g) := \sum_{i=r+1}^{n}\{g(\mathbb{X}_{i-1}) - \mathbb{E}g(\mathbb{X}_{i-1})\}.$$

To prove maximal inequalities for $S_n(g)$, we use the decomposition technique throughout Chapter 2, equation (2.8.23) therein, or Chapter 3. For $j \ge 1$ define

$$S_{n,j}(g) := \sum_{i=r+1}^{n} W_{i,j}(g), \qquad W_{i,j}(g) := \mathbb{E}[g(\mathbb{X}_{i-1})|\varepsilon_{i-j},...,\varepsilon_{i-1}].$$

Then,

$$S_n(g) = S_n(g) - S_{n,q}(g) + \sum_{l=1}^{L}(S_{n,\tau_l} - S_{n,\tau_{l-1}}) + S_{n,0}(g) \tag{4.6.55}$$

where $L = \lfloor\frac{\log(q)}{\log(2)}\rfloor$ and $\tau_l = 2^l$ $(l = 0,...,L-1)$, $\tau_L = q$ for arbitrary $q \in \{1,...,n\}$. Set

$$S_n^{\circ}(g) := S_{n,q}(g)$$

and

$$S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g) = \sum_{i=1}^{\lfloor\frac{n}{\tau_l}\rfloor+1} T_{i,l}(g), \qquad T_{i,l}(g) := \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n}\big[W_{k,\tau_l}(g) - W_{k,\tau_{l-1}}(g)\big].$$

Hence,

$$S_n^{\circ}(g) = \sum_{l=1}^{L}\Big[\sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1} T_{i,l}(g) + \sum_{\substack{i=1\\ i \text{ odd}}}^{\lfloor\frac{n}{\tau_l}\rfloor+1} T_{i,l}(g)\Big] + S_{n,0}(g).$$

Lemma 4.6.14 (Maximal inequalities under functional dependence)**.** *Assume that X_i is of the form (4.1.7). Suppose that $\mathcal{G}$ satisfies (4.6.51) and (4.6.52) with $G = 1$ and some $L_{\mathcal{G}} > 0$. Let $\theta \in (0,1]$. Then, for any decreasing sequence $\Delta(k)$, $k \in \mathbb{N}_0$, satisfying*

$$drL_{\mathcal{G}}^{\theta} \cdot \sup_{l=1,...,d}\sup_{j=1,...,r} \delta_{2\theta}^{X_{\cdot,l}}(k-j)^{\theta} \le \Delta(k),$$

there exists some universal constant $c > 0$ such that

$$\mathbb{E}\sup_{g\in\mathcal{G}}|S_n(g) - S_n^{\circ}(g)| \le cHq^{*,dep}(\sqrt{\frac{H}{n}}). \tag{4.6.56}$$

Furthermore, for any estimator $\hat{g} \in \mathcal{G}$ we have for some universal constant $c > 0$

$$\mathbb{E}|S_n^{\circ}(\hat{g})| \le c(\sqrt{nH}\tilde{V}(\mathbb{E}[\|g(\mathbb{X}_r)\|_1|_{g=\hat{g}}]) + qH). \tag{4.6.57}$$

Proof of Lemma 4.6.14. Let $q \in \{1, ..., n\}$ be arbitrary. Then, as in the proof of Theorem 2.4.1 (cf. the term A_1 accordingly), there exists a universal constant $c > 0$ such that

$$\mathbb{E} \sup_{g \in \mathcal{G}} |S_n(g) - S_n^{\circ}(g)| \leq c\sqrt{nH}\beta^{dep}(q).$$

If $q := q^{*,dep}(\sqrt{\frac{H}{n}})$,

$$\mathbb{E} \sup_{g \in \mathcal{G}} |S_n(g) - S_n^{\circ}(g)| \leq c\sqrt{nH}\beta(q^{*,dep}(\sqrt{\frac{H}{n}})) \leq c\sqrt{nH} q^{*,dep}(\sqrt{\frac{H}{n}})\sqrt{\frac{H}{n}} = cH q^{*,dep}(\sqrt{\frac{H}{n}}),$$

which proves (4.6.56).

We employ a similar strategy as in the proof of Lemma 4.6.2. Let $N_l(g) := \tau_l \sqrt{\frac{H}{n}} \vee V_l(g)$ for $V_l(g) := \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\|g\|_2, \Delta(\lfloor \frac{j}{2} \rfloor)\}$. We show the following two inequalties first, where c denotes some universal constant:

(i)

$$\mathbb{E} \sup_{g \in \mathcal{G}} |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}| \leq c\sqrt{nH}, \tag{4.6.58}$$

(ii)

$$\mathbb{E}\big[\sup_{g \in \mathcal{G}} |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}|^2\big] \leq cnH. \tag{4.6.59}$$

For $g \in \mathcal{G}$, we have

$$\Big|\frac{T_{i,l}(g)}{N_l(g)}\Big| \leq 2\tau_l \|g\|_\infty N_l(g)^{-1} \leq 2\tau_l \tau_l^{-1}\sqrt{\frac{n}{H}} = 2\sqrt{\frac{n}{H}}$$

and by the same calculation as in the proof of Theorem 2.4.1,

$$\frac{1}{\frac{n}{\tau_l}} \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \Big\|\frac{T_{i,l}(g)}{N_l(g)}\Big\|_2^2 \leq \frac{1}{N_l(g)^2}\Big(\sqrt{\tau_l} \sum_{j=\tau_{l-1}+1}^{\tau_l} \min\{\|g\|_2, \Delta(\lfloor \frac{j}{2} \rfloor)\}\Big)^2 \leq \tau_l.$$

By Bernstein's inequality we obtain

$$\begin{aligned}
&\mathbb{P}\Big(|\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}| > x\Big) \\
\leq\ & \mathbb{P}\Big(\Big|\sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{T_{i,l}(g)}{N_l(g)}\Big| > \frac{x}{2}\Big) + \mathbb{P}\Big(\Big|\sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{T_{i,l}(g)}{N_l(g)}\Big| > \frac{x}{2}\Big) \\
\leq\ & 4\exp\Big(-\frac{1}{2}\frac{(x/2)^2}{n + 2\sqrt{\frac{n}{H}}x/2}\Big) \\
\leq\ & 4\exp\Big(-\frac{1}{8}\frac{x^2}{n + \sqrt{\frac{n}{H}}x}\Big).
\end{aligned} \tag{4.6.60}$$

(i) Using standard arguments (cf. [van der Vaart, 1998, Lemma 19.35]), we derive from (4.6.60) that there exists a universal constant $c > 0$ such that

$$\mathbb{E} \sup_{g\in\mathcal{G}} |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}| \leq c\sqrt{nH}.$$

This shows (4.6.58).

(ii) Next, we use

$$\mathbb{E}\big[\sup_{g\in\mathcal{G}} |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}|^2\big] = \int_0^\infty \mathbb{P}\Big(\sup_{g\in\mathcal{G}} |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}| > \sqrt{t}\Big) dt.$$

Put $a := \sqrt{\frac{H}{n}}$ and choose $G := 16na$. Then for $t \geq G^2$, $a^{-1}\sqrt{t} \geq n$. With (4.6.60) and $\int_{b^2}^\infty \exp(-b_2\sqrt{t})dt = \int_b^\infty 2s\exp(-b_2 s)ds = 2(b_2 b + 1)b_2^{-2}\exp(-b_2 b)$, we obtain

$$\begin{aligned}
& \int_0^\infty \mathbb{P}\Big(\sup_g |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}| > \sqrt{t}\Big) dt \\
= \;& G^2 + \int_{G^2}^\infty \mathbb{P}\Big(\sup_g |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}| > \sqrt{t}\Big) dt \\
\leq \;& G^2 + 4|\mathcal{G}| \int_{G^2}^\infty \exp\Big(-\frac{1}{8}\frac{t}{n + \sqrt{\frac{n}{H}}\sqrt{t}}\Big) dt \\
\leq \;& G^2 + 4|\mathcal{G}| \int_{G^2}^\infty \exp\Big(-\frac{1}{8}\frac{t}{n + a^{-1}\sqrt{t}}\Big) dt \\
\leq \;& G^2 + 4|\mathcal{G}| \int_{G^2}^\infty \exp(-\frac{1}{16}\frac{\sqrt{t}}{a^{-1}}) dt \\
\leq \;& G^2 + 8|\mathcal{G}|\Big(Ga + 1\Big)(16a^{-1})^2 \exp(-\frac{1}{16}Ga) \\
\leq \;& 2^{11}\Big[\Big(na\Big)^2 + |\mathcal{G}| \cdot \exp\Big(-na^2\Big) \cdot (n + a^{-2})\Big]. \qquad (4.6.61)
\end{aligned}$$

We have

$$\mathbb{E}\big[\sup_g |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}|^2\big] \leq 2^{12}\Big[(na)^2 + |\mathcal{G}| \cdot \exp\big(-na^2\big) \cdot (n + a^{-2})\Big].$$

which can be upper bounded by

$$\mathbb{E}\big[\sup_g |\frac{S_{n,\tau_l}(g) - S_{n,\tau_{l-1}}(g)}{N_l(g)}|^2\big] \leq 2^{12}\Big[nH + n + \frac{n}{H}\Big] = 2^{14}nH,$$

proving (4.6.58).

Moving on to $\mathbb{E}|S_n^\circ(\hat{g})|$, the Cauchy-Schwarz inequality yields

$$\begin{aligned}\mathbb{E}|S_n^\circ(\hat{g})| &\leq \sum_{l=1}^{L}\Big[\Big\|\frac{S_{n,\tau_l}(\hat{g})-S_{n,\tau_{l-1}}(\hat{g})}{N_l(\hat{g})}\Big\|_2\mathbb{E}[V_l(\hat{g})^2]^{1/2} \\ &\quad +\mathbb{E}\Big|\frac{S_{n,\tau_l}(\hat{g})-S_{n,\tau_{l-1}}(\hat{g})}{N_l(\hat{g})}\Big|\cdot\tau_l\sqrt{\frac{H}{n}}\Big] \\ &\quad +\mathbb{E}[S_{n,0}(\hat{g})] \\ &\leq c(\sqrt{nH}\sum_{l=1}^{L}\mathbb{E}[V_l(\hat{g})^2]^{1/2}+qH)+\mathbb{E}[|S_{n,0}(\hat{g})|]. \qquad (4.6.62)\end{aligned}$$

The last summand can be discussed as follows. Let $N(g) := \sqrt{\frac{H}{n}}\vee\|g\|_2$,

$$\begin{aligned}\mathbb{E}[|S_{n,0}(\hat{g})|] &= \mathbb{E}[|\frac{S_{n,0}(\hat{g})}{N(\hat{g})}\cdot N(\hat{g})|] \\ &\leq \Big\|\frac{S_{n,0}(\hat{g})}{N(\hat{g})}\Big\|_2\mathbb{E}[\|g\|_2^2\,|_{g=\hat{g}}]^{1/2}+\mathbb{E}[|\frac{S_{n,0}(\hat{g})}{N(\hat{g})}|]\sqrt{\frac{H}{n}}. \qquad (4.6.63)\end{aligned}$$

Since $S_{n,0}(g)=\sum_{i=r+1}^{n}W_{i,0}(g)$ is a sum of independent variables with $|W_{i,0}(g)|\leq\|g\|_\infty\leq 1$ and $\|W_{i,0}(g)\|_2\leq 2\|g\|_2$, the Bernstein inequality yields

$$\mathbb{P}\Big(|\frac{S_{n,0}(g)}{N(g)}|>x\Big)\leq 2\exp\big(-\frac{1}{4}\frac{x^2}{n+N(g)^{-1}x/2}\big),$$

from which we derive

$$\mathbb{E}\big[\sup_{g\in\mathbb{G}}|\frac{S_{n,0}(g)}{N(g)}|\big]\leq c(\sqrt{nH}+N(g)^{-1}H)\leq c\sqrt{nH}$$

for some universal constant $c>0$. In analogy to the calculation of equation (4.6.61),

$$\mathbb{E}\Big[\sup_{g\in\mathcal{G}}\Big|\frac{S_{n,0}(g)}{N(g)}\Big|^2\Big]\leq 2^{10}nH.$$

Therefore, equation (4.6.63) can be bounded by

$$\mathbb{E}\sup_{g\in\mathcal{G}}|S_{n,0}(\hat{g})|\leq c(\sqrt{nH}\mathbb{E}[\|g(\mathbb{X}_r)\|_1\,|_{g=\hat{g}}]^{1/2}+H). \qquad (4.6.64)$$

Now, let us define $v_l(x):=\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\sqrt{x},\Delta(\lfloor\frac{j}{2}\rfloor)\}$. Then,

$$\begin{aligned}V_l(h)^2 &= \Big(\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|h(\mathbb{X}_r)\|_2,\Delta(\lfloor\frac{j}{2}\rfloor)\}\Big)^2 \\ &\leq \Big(\sum_{j=\tau_{l-1}+1}^{\tau_l}\min\{\|h(\mathbb{X}_r)\|_1^{1/2},\Delta(\lfloor\frac{j}{2}\rfloor)\}\Big)^2=v_l(\|h\|_1)^2.\end{aligned}$$

This implies the first bound of the following inequality. The second bound follows from Jensen's inequality while taking into account that v_l^2 is concave by Lemma 4.6.13:

$$\mathbb{E}[V_l(\hat{g})^2]^{1/2} \le \mathbb{E}[v_l(\|g(\mathbb{X}_r)\|_1)^2\big|_{g=\hat{g}}]^{1/2} \le v_l(\mathbb{E}[\|g(\mathbb{X}_r)\|_1\big|_{g=\hat{g}}]). \tag{4.6.65}$$

Inserting equations (4.6.65), (4.6.64) into (4.6.62) and applying Lemma 2.8.5 afterwards, which allows to replace $\Delta(\lfloor\frac{j}{2}\rfloor)$ in $v_l(\cdot)$ by $\Delta(j)$, gives

$$\begin{aligned}
\mathbb{E}\sup_{g\in\mathcal{G}}|S_n^\circ(\hat{g})| &\le c\Big(\sqrt{nH}(\sum_{l=1}^{L} v_l(\mathbb{E}[\|g(\mathbb{X}_r)\|_1\big|_{g=\hat{g}}]) + \mathbb{E}[\|g(\mathbb{X}_r)\|_1\,|_{g=\hat{g}}]^{1/2} + (q+1)H\Big) \\
&\le c2\Big(\sqrt{nH}(2\sum_{j=1}^{\infty}\min\{\mathbb{E}[\|g(\mathbb{X}_r)\|_1\big|_{g=\hat{g}}]^{1/2},\Delta(j)\} + \mathbb{E}[\|g(\mathbb{X}_r)\|_1\,|_{g=\hat{g}}]^{1/2}) + qH\Big) \\
&\le c(\sqrt{nH}\tilde{V}(\mathbb{E}[\|g(\mathbb{X}_r)\|_1\,|_{g=\hat{g}}]) + qH).
\end{aligned}$$

□

For $g:\mathbb{R}^{dr}\to\mathbb{R}^d$, let

$$M_n(g) := \sum_{i=r+1}^{n}\frac{1}{d}\langle\varepsilon_i, g(\mathbb{X}_{i-1})\rangle.$$

Lemma 4.6.15 (Maximal inequalities for martingale sequences under functional dependence)**.** *Assume that X_i is of the form (4.1.7) and that Assumption 4.1.1 holds true. Suppose that any component of $g\in\mathcal{G}$ satisfies (4.6.51) and (4.6.52) with $G=1$ and some $L_{\mathcal{G}}>0$. Let $\theta\in(0,1]$. Then, with any decreasing sequence $\Delta(k)$, $k\in\mathbb{N}_0$, satisfying*

$$drL_{\mathcal{G}}^{\theta}\cdot\sup_{l=1,\dots,r}\sup_{j=1,\dots,r}\delta_{2\theta}^{X_{\cdot,l}}(k-j)^{\theta}\le\Delta(k),$$

there exists another process $M_n^\circ(g)$ and some universal constant $c>0$ such that

$$\mathbb{E}\sup_{g\in\mathcal{G}}|M_n(g)-M_n^\circ(g)|\le cC_\varepsilon\sqrt{nH}\beta^{dep}(q). \tag{4.6.66}$$

Furthermore for an estimator $\hat{g}:\mathbb{R}^{dr}\to\mathbb{R}^d$,

$$\mathbb{E}[|M_n^\circ(\hat{g})|]\le cC_\varepsilon(\sqrt{nH}(\sqrt{\log(q)}+1)\mathbb{E}[\||g(\mathbb{X}_1)|_2\|_2\big|_{g=\hat{g}}]^{1/2}+q^{1/2}H) \tag{4.6.67}$$

Proof of Lemma 4.6.15. We use a similar decomposition as in the proof of Lemma 4.6.14. For $j\ge1$ define

$$M_{n,j}(g) := \sum_{i=r+1}^{n}\bar{W}_{i,j}(g),\qquad \bar{W}_{i,j}(g) := \mathbb{E}[\frac{1}{d}\langle\varepsilon_i, g(\mathbb{X}_{i-1})\rangle|\varepsilon_{i-j},\dots,\varepsilon_i] = \frac{1}{d}\langle\varepsilon_i, W_{i,j}(g)\rangle,$$

where $W_{i,j}(g) := \mathbb{E}[g(\mathbb{X}_{i-1})|\varepsilon_{i-j},\dots,\varepsilon_{i-1}]$. Let

$$M_n^\circ(g) := M_{n,q}(g).$$

Note that $(\langle \varepsilon_i, g(\mathbb{X}_{i-1})\rangle)_i$ is a martingale and for fixed j, the sequence

$$(E_{i,j}(g))_{g\in\mathcal{G}} \quad = \quad \big((\bar{W}_{i,j+1}(g) - \bar{W}_{i,j}(g))\big)_{g\in\mathcal{G}}$$

is a $|\mathcal{G}|$-dimensional martingale difference vector with respect to $\mathcal{A}^i := \sigma(\varepsilon_{i-j}, \varepsilon_{i-j+1}, ...)$. Since

$$\sup_{g\in\mathcal{G}} |E_{i,j}(g)| = \sup_{g\in\mathcal{G}} |\bar{W}_{i,j+1}(g) - \bar{W}_{i,j}(g)| \leq \frac{1}{d}|\varepsilon_i|_2 \cdot \sup_{g\in\mathcal{G}} |W_{i,j+1}(g) - W_{i,j}(g)|_2,$$

we have by (4.6.53) (which also holds with $\sup_g$ inside the $\|\cdot\|_2$-norm),

$$\big\| \sup_{g\in\mathcal{G}} |E_{i,j}(g)| \,\big\|_2 \quad \leq \quad \frac{1}{d}\sum_{k=1}^{d} \|\varepsilon_{ik}\|_2 \left\|\sup_{g\in\mathcal{G}} |g(\mathbb{X}_{i-1})_k - g(\mathbb{X}_{i-1}^{*(i-j)})_k|\right\|_2 \leq C_\varepsilon \Delta(j).$$

Therefore, in analogy to the proof of Theorem 2.4.1,

$$\mathbb{E}\sup_{g\in\mathcal{G}} |M_n(g) - M_n^\circ(g)| \leq \sum_{j=q}^{\infty} \left\|\sup_{g\in\mathcal{G}} \Big| \sum_{i=r+1}^{n} E_{i,j}(g)\Big|\right\|_2 \leq cC_\varepsilon \sqrt{nH}\beta^{dep}(q).$$

for some universal constant $c > 0$.

For $q \in \{1, ..., n\}$,

$$M_n^\circ(g) = \sum_{l=1}^{L} \Big[\sum_{\substack{i=1\\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l}\rfloor + 1} \bar{T}_{i,l}(g) + \sum_{\substack{i=1\\ i \text{ odd}}}^{\lfloor \frac{n}{\tau_l}\rfloor + 1} \bar{T}_{i,l}(g)\Big] + M_{n,0}(g)$$

where

$$\bar{T}_{i,l}(g) := \sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \big[\bar{W}_{k,\tau_l}(g) - \bar{W}_{k,\tau_{l-1}}(g)\big] = \frac{1}{d}\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \langle \varepsilon_k, W_{k,\tau_l}(g) - W_{k,\tau_{l-1}}(g)\rangle.$$

Next, let

$$N_l(g) = \max\{\tau_l^{1/2} \cdot \sqrt{\frac{H}{n}}, D_l(g)\}, \qquad D_l(g) := \mathbb{E}[\frac{1}{d}|W_{1,\tau_l}(g) - W_{1,\tau_{l-1}}(g)|_2^2]^{1/2}.$$

We show the following two inequalities first, where c denotes a universal constant:

(i)
$$\mathbb{E}\sup_{g\in\mathcal{G}} |\frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)}| \leq cC_\varepsilon \sqrt{nH}, \tag{4.6.68}$$

(ii)
$$\mathbb{E}\big[\sup_{g\in\mathcal{G}} |\frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)}|^2\big] \leq cC_\varepsilon^2 nH. \tag{4.6.69}$$

We have, similar to (4.6.39), by [Rio, 2009, Theorem 2.1] and Assumption 4.1.1,

$$\begin{aligned}
&\frac{1}{\lfloor \frac{n}{\tau_l} \rfloor} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \mathbb{E}[|\bar{T}_{i,l}(g)|^m] \\
\leq\ & (m-1)^{m/2} \frac{1}{\lfloor \frac{n}{\tau_l} \rfloor} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{d} \Big(\sum_{k=(i-1)\tau_l+1}^{(i\tau_l)\wedge n} \mathbb{E}[|\langle \varepsilon_k, W_{k,\tau_l}(g) - W_{k,\tau_{l-1}}(g) \rangle|^m]^2 \Big)^{1/2} \\
\leq\ & (m-1)^{m/2} \frac{1}{d} \tau_l^{m/2} \mathbb{E}[|\varepsilon_1|_2^m] \cdot \mathbb{E}[|W_{1,\tau_l}(g) - W_{1,\tau_{l-1}}(g)|_2^m] \\
\leq\ & (m-1)^{m/2} \tau_l^{m/2} C_\varepsilon^m \cdot \frac{1}{d} \mathbb{E}[|W_{1,\tau_l}(g) - W_{1,\tau_{l-1}}(g)|_2^2] \, \|g\|_\infty^{m-2} \\
\leq\ & \frac{m!}{2} \cdot 2e^2 C_\varepsilon^2 \tau_l \mathbb{E}[\frac{1}{d} |W_{1,\tau_l}(g) - W_{1,\tau_{l-1}}(g)|_2^2] \cdot (e C_\varepsilon \tau_l^{1/2})^{m-2}. \qquad (4.6.70)
\end{aligned}$$

With $\tilde{a} := \tau_l^{1/2} \cdot \sqrt{\frac{H}{n}}$,

$$\frac{1}{\lfloor \frac{n}{\tau_l} \rfloor} \sum_{i=1}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{1}{N_l(g)} \mathbb{E}[|T_{i,l}(g)|^m] \leq \frac{m!}{2} \cdot 2e^2 C_\varepsilon^2 \tau_l (\tilde{a}^{-1} e C_\varepsilon \tau_l^{1/2})^{m-2}$$

By Bernstein's inequality for independent variables, we conclude that

$$\begin{aligned}
&\mathbb{P}\Big(\Big| \frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)} \Big| > x \Big) \\
\leq\ & \mathbb{P}\Big(\Big| \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{T_{i,l}(g)}{N_l(g)} \Big| > \frac{x}{2} \Big) + \mathbb{P}\Big(\Big| \sum_{\substack{i=1 \\ i \text{ even}}}^{\lfloor \frac{n}{\tau_l} \rfloor + 1} \frac{T_{i,l}(g)}{N_l(g)} \Big| > \frac{x}{2} \Big) \\
\leq\ & 4 \exp\Big(-\frac{1}{8} \frac{x^2}{2(eC_\varepsilon)^2 n + e C_\varepsilon \sqrt{\frac{n}{H}} x} \Big). \qquad (4.6.71)
\end{aligned}$$

(i) Using standard arguments (cf. [van der Vaart, 1998, Lemma 19.35]), we derive from (4.6.60) that there exists a universal constant $c > 0$ such that

$$\mathbb{E} \sup_{g \in \mathcal{G}} \Big| \frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)} \Big| \leq c C_\varepsilon \sqrt{nH}.$$

This shows (4.6.68).

(ii) Next, we use

$$\mathbb{E}\Big[\sup_{g \in \mathcal{G}} \Big| \frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)} \Big|^2 \Big] = \int_0^\infty \mathbb{P}\Big(\sup_{g \in \mathcal{G}} \Big| \frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)} \Big| > \sqrt{t} \Big) dt.$$

Put $a := \sqrt{\frac{H}{n}}$. Choose $G := 16(eC_\varepsilon)na$. Then for $t \geq G^2$, $a^{-1}\sqrt{t} \geq n$. With (4.6.71) and $\int_{b^2}^\infty \exp(-b_2\sqrt{t})dt = \int_b^\infty 2s\exp(-b_2 s)ds = 2(b_2 b+1)b_2^{-2}\exp(-b_2 b)$, we obtain

$$\begin{aligned} & \int_0^\infty \mathbb{P}\Big(\sup_g |\frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)}| > \sqrt{t}\Big)dt \\ = \;& G^2 + \int_{G^2}^\infty \mathbb{P}\Big(\sup_g |\frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)}| > \sqrt{t}\Big)dt \\ \leq \;& G^2 + 4|\mathcal{G}|\int_{G^2}^\infty \exp\Big(-\frac{1}{8}\frac{x^2}{2(eC_\varepsilon)^2 n + eC_\varepsilon\sqrt{\frac{n}{H}}x}\Big)dt \\ \leq \;& G^2 + 4|\mathcal{G}|\int_{G^2}^\infty \exp\Big(-\frac{1}{8}\frac{t}{2(eC_\varepsilon)^2 n + eC_\varepsilon a^{-1}\sqrt{t}}\Big)dt \\ \leq \;& G^2 + 4|\mathcal{G}|\int_{G^2}^\infty \exp(-\frac{1}{16}\frac{\sqrt{t}}{eC_\varepsilon a^{-1}})dt \\ \leq \;& G^2 + 8|\mathcal{G}|\Big(\frac{Ga}{eC_\varepsilon}+1\Big)(16eC_\varepsilon a^{-1})^2\exp(-\frac{1}{16eC_\varepsilon}Ga) \\ \leq \;& 2^{11}(eC_\varepsilon)^2\Big[(na)^2 + |\mathcal{G}|\cdot\exp\big(-na^2\big)\cdot(n+a^{-2})\Big]. \end{aligned} \tag{4.6.72}$$

We conclude that

$$\mathbb{E}\big[\sup_g |\frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)}|^2\big] \;\leq\; 2^{12}(eC_\varepsilon)^2\Big[(na)^2 + |\mathcal{G}|\cdot\exp\big(-na^2\big)\cdot(n+a^{-2})\Big]$$

which can be upper bounded by

$$\mathbb{E}\big[\sup_g |\frac{M_{n,\tau_l}(g) - M_{n,\tau_{l-1}}(g)}{N_l(g)}|^2\big] \leq 2^{12}(eC_\varepsilon)^2\Big[nH + n + \frac{n}{H}\Big] \leq 2^{14}(eC_\varepsilon)^2 nH.$$

This shows (4.6.69).

Using (4.6.68) and (4.6.69), we can now upper bound $\mathbb{E}|M_n^\circ(\hat{g})|$. By the Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E}|M_n^\circ(\hat{g})| \;\leq\; & \sum_{l=1}^L \Big[\Big\|\frac{M_{n,\tau_l}(\hat{g}) - M_{n,\tau_{l-1}}(\hat{g})}{N_l(\hat{g})}\Big\|_2 \mathbb{E}[D_l(\hat{g})^2]^{1/2} \\ & +\mathbb{E}\Big|\frac{M_{n,\tau_l}(\hat{g}) - M_{n,\tau_{l-1}}(\hat{g})}{N_l(\hat{g})}\Big|\cdot\tau_l^{1/2}\sqrt{\frac{H}{n}}\,\Big] \\ & +\mathbb{E}[M_{n,0}(\hat{g})] \\ \leq\; & cC_\varepsilon(\sqrt{nH}\sum_{l=1}^L \mathbb{E}[D_l(\hat{g})^2]^{1/2} + q^{1/2}H) + \mathbb{E}[|M_{n,0}(\hat{g})|]. \end{aligned} \tag{4.6.73}$$

For $v \in \mathbb{R}^L$, $|v|_1 \le \sqrt{L}|v|_2$. Thus,

$$\begin{aligned}\sum_{l=1}^L \mathbb{E}[D_l(\hat{g})^2]^{1/2} &\le \sqrt{L}\cdot\Big(\sum_{l=1}^L \mathbb{E}[D_l(\hat{g})^2]\Big)^{1/2}\\ &= \sqrt{L}\mathbb{E}\Big[\Big(\frac{1}{d}\sum_{l=1}^L \mathbb{E}[|W_{1,\tau_l}(g)-W_{1,\tau_{l-1}}(g)|_2^2]\Big)\Big|_{g=\hat{g}}\Big]^{1/2}. \quad (4.6.74)\end{aligned}$$

Note that $(W_{1,\tau_l}(g) - W_{1,\tau_{l-1}}(g))_l$ is a martingale difference sequence with respect to $\tilde{\mathcal{A}}^l := \sigma(\varepsilon_{1-\tau_l}, ..., \varepsilon_1)$. We therefore have

$$\mathbb{E}[|W_{1,q}(g)-W_{1,0}(g)|_2^2] = \mathbb{E}\Big[\Big|\sum_{l=1}^L W_{1,\tau_l}(g)-W_{1,\tau_{l-1}}(g)\Big|_2^2\Big] = \sum_{l=1}^L \mathbb{E}[|W_{1,\tau_l}(g)-W_{1,\tau_{l-1}}(g)|_2^2].$$

Since the left hand side is bounded by $4\mathbb{E}[|g(\mathbb{X}_1)|_2^2]$ by the projection property of conditional expectations, insertion into (4.6.74) yields

$$\sum_{l=1}^L \mathbb{E}[D_l(\hat{g})]^{1/2} \le 4\sqrt{L}\cdot\mathbb{E}[\frac{1}{d}\||g(\mathbb{X}_1)|_2\|_2^2|_{g=\hat{g}}]^{1/2}. \quad (4.6.75)$$

The last summand in (4.6.73) can be similarly dealt with. With $N(g) := \sqrt{\frac{H}{n}} \vee \||g(\mathbb{X}_1)|_2\|_2$,

$$\begin{aligned}\mathbb{E}[\big|M_{n,0}(\hat{g})\big|] &= \mathbb{E}[\big|\frac{M_{n,0}(\hat{g})}{N(\hat{g})}\cdot N(\hat{g})\big|]\\ &\le \Big\|\frac{M_{n,0}(\hat{g})}{N(\hat{g})}\Big\|_2 \mathbb{E}[\||g(\mathbb{X}_1)|_2\|_2^2\,\big|_{g=\hat{g}}]^{1/2}\\ &\qquad + \mathbb{E}[\big|\frac{M_{n,0}(\hat{g})}{N(\hat{g})}\big|]\sqrt{\frac{H}{n}}. \quad (4.6.76)\end{aligned}$$

Since $M_{n,0}(g) = \sum_{i=r+1}^n \bar{W}_{i,0}(g)$ is a sum of independent variables, we can proceed as before in Lemma 4.6.14 and obtain the existence of universal constants $c > 0$ such that

$$\Big\|\frac{M_{n,0}(\hat{g})}{N(\hat{g})}\Big\|_2 \le cC_\varepsilon\sqrt{nH}, \qquad \mathbb{E}[\big|\frac{M_{n,0}(\hat{g})}{N(\hat{g})}\big|] \le cC_\varepsilon^2\sqrt{nH}.$$

Insertion into (4.6.76) yields

$$\mathbb{E}[\big|M_{n,0}(\hat{g})\big|] \le cC_\varepsilon\big\{\sqrt{nH}\cdot\mathbb{E}[\||g(\mathbb{X}_1)|_2\|_2^2\,\big|_{g=\hat{g}}]^{1/2} + H\big\}. \quad (4.6.77)$$

Insertion of (4.6.75) and (4.6.77) into (4.6.73) gives the result. □

Oracle inequalities under functional dependence

Let $\mathcal{F} \subset \{f : \mathbb{R}^{dr} \to \mathbb{R}^d \text{ measurable}\}$ such that any $f = (f_j)_{j=1,...,d} \in \mathcal{F}$ satisfies

$$\sup_{j\in\{1,...,d\}} |f_j(x) - f_j(x')| \le L_{\mathcal{F}} \cdot |x - x'|_\infty \tag{4.6.78}$$

and

$$\sup_{j\in\{1,...,d\}} \sup_{x\in \text{supp}(\mathcal{W})} |f_j(x)| \le F \tag{4.6.79}$$

where $\mathcal{W} : \mathbb{R}^{dr} \to [0, 1]$ is an arbitrary weight function depending on $\varsigma > 0$ with

$$|\mathcal{W}(x) - \mathcal{W}(x')| \le \frac{1}{\varsigma} \cdot |x - x'|_\infty.$$

Let

$$\hat{f} \in \arg\min_{f\in\mathcal{F}} \hat{R}_n(f).$$

The main result of this section is the following theorem, restated from Theorem 4.1.6. Here, $\mathbb{H}(\delta) = \log \mathbb{N}(\delta, \mathcal{F}, \|\cdot\|_\infty)$.

Theorem 4.6.16. *Suppose that X_i is of the form (4.1.7) and that Assumption 4.1.1 holds true. Assume that there exist $F > 0, L_{\mathcal{F}} > 0$ such that $\mathcal{F}$ satisfies (4.6.78) and (4.6.79). Furthermore, suppose that $f_0 : \mathbb{R}^{dr} \to \mathbb{R}^d$ from (4.1.1) is such that $|f_0(x) - f_0(x')|_\infty \le K|x - x'|_\infty$ for some $K > 0$.*

Let Assumption 4.1.5 be fulfilled with $L_{\mathcal{G}} = 2dr\big(\frac{2}{\varsigma} + \frac{(L_{\mathcal{F}}+K)}{F}\big)$, $\delta \in (0, 1)$. Then, for any $\eta > 0$ there exists a constant $\mathbb{C} = \mathbb{C}(\eta, C_\varepsilon, F)$ such that

$$\mathbb{E}D(\hat{f}) \le (1 + \eta)^2 \inf_{f\in\mathcal{F}} D(f) + \mathbb{C} \cdot \big\{\Lambda(\frac{\mathbb{H}(\delta)}{n}) + \delta\big\}.$$

Proof of Theorem 4.6.16. Let $\eta > 0$. We follow the proof of Theorem 4.6.4. Define

$$\begin{aligned}
R_{1,n} &:= (1 + \eta)cF^2 q^{*,dep}(\sqrt{\frac{H}{n}})\frac{H}{n} + \frac{\eta F^2}{2}(\tilde{V}^{-1})^*\Big(2\frac{1 + \eta}{\eta}\sqrt{\frac{H}{n}}\Big),\\
R_{1,\delta} &:= cF^2\sqrt{\frac{H}{n}}\tilde{V}(2F^{-2}\delta^2),\\
R_{2,n} &:= 2cC_\varepsilon F q^{*,dep}(\frac{H}{n})\frac{H}{n},\\
R_{2,\delta} &:= C_\varepsilon\delta + 2cC_\varepsilon F\sqrt{\frac{H}{n}}\delta.
\end{aligned}$$

Then, as in the mixing case, by Lemma 4.6.18, (4.6.8) and Lemma 4.6.17,

$$\begin{aligned}
\mathbb{E}D(\hat{f}) &\le (1 + \eta)\mathbb{E}\hat{D}_n(\hat{f}) + R_{1,n} + (1 + \eta)R_{1,\delta}\\
&\le (1 + \eta)\Big\{\inf_{f\in\mathcal{F}} D(f) + 2cC_\varepsilon F\sqrt{\frac{H}{n}}\sqrt{q^{*,dep}(\sqrt{\frac{H}{n}})}\mathbb{E}[D(\hat{f})]^{1/2} + R_{2,n} + R_{2,\delta} + R_{1,\delta}\Big\} + R_{1,n}.
\end{aligned}$$

Due to $2ab \leq a^2+b^2$ with $a := (1+\eta)cC_\varepsilon F\sqrt{\frac{H}{n}}\sqrt{q^{*,dep}(\sqrt{\frac{H}{n}})}(\frac{1+\eta}{\eta})^{1/2}$, $b := (\frac{\eta}{1+\eta})^{1/2}\mathbb{E}[D(\hat{f})]^{1/2}$, we obtain

$$\begin{aligned}\mathbb{E}D(\hat{f}) &\leq (1+\eta)\inf_{f\in\mathcal{F}} D(f) + \frac{(1+\eta)^3}{\eta}(cC_\varepsilon F)^2 q^*(\sqrt{\frac{H}{n}})\frac{H}{n} + \frac{\eta}{1+\eta}\mathbb{E}[D(\hat{f})]\\ &\quad +(1+\eta)(R_{2,n}+R_{2,\delta}+R_{1,\delta}) + R_{1,n}.\end{aligned}$$

This implies

$$\begin{aligned}\mathbb{E}D(\hat{f}) &\leq (1+\eta)^2\inf_{f\in\mathcal{F}} D(f) + (1+\eta)R_{1,n}\\ &\quad +(1+\eta)^2(R_{2,n}+R_{2,\delta}+R_{1,\delta})\\ &\quad +\frac{(1+\eta)^4}{\eta}(cC_\varepsilon F)^2 q^{*,dep}(\sqrt{\frac{H}{n}})\frac{H}{n}. \qquad (4.6.80)\end{aligned}$$

Using Young's inequality applied to $\tilde{V}^{-1}$ ($\tilde{V}^{-1}$ is convex) and Lemma 4.6.11, we obtain

$$R_{1,\delta} \leq cF^2(\tilde{V}^{-1})^*(\sqrt{\frac{H}{n}}) + 2c\delta^2 \leq cF^2\Lambda(\frac{H}{n}) + 2c\delta^2,$$

as well as $R_{2,n} \leq 4cF\Lambda(\frac{H}{n})$. Furthermore,

$$R_{1,n} \leq (1+\eta)cF^2\Lambda(\frac{H}{n}) + \frac{\eta F^2}{2}\Big(2\frac{1+\eta}{\eta}\Big)^2\Lambda(\frac{H}{n}).$$

and

$$R_{2,\delta} \leq C_\varepsilon(\delta + cF\delta^2 + cF\frac{H}{n}).$$

Insertion of these results into (4.6.80) yields the assertion. □

To prove Theorem 4.6.16, the following two lemmata are used.

Lemma 4.6.17. *Suppose that X_i is of the form (4.1.7) and that Assumption 4.1.1 holds true. Assume that there exist $F > 0, L_{\mathcal{F}} > 0$ such that $\mathcal{F}$ satisfies (4.6.78) and (4.6.79). Furthermore, suppose that $f_0 : \mathbb{R}^{dr} \to \mathbb{R}^d$ from (4.1.1) is such that $|f_0(x) - f_0(x')|_\infty \leq K|x-x'|_\infty$ for some $K > 0$.*

If additionally Assumption 4.1.5 is fulfilled with $L_{\mathcal{G}} = 2dr\big(\frac{K+L_{\mathcal{F}}}{F} + \frac{2}{\varsigma}\big)$, then

$$\begin{aligned}&\Big|\mathbb{E}\Big[\frac{1}{nd}\sum_{i=1}^n\langle\varepsilon_i, \hat{f}(\mathbb{X}_{i-1})\rangle\mathcal{W}(\mathbb{X}_{i-1})\Big]\Big|\\ &\leq C_\varepsilon\delta + 2cC_\varepsilon F\Big[q^{*,dep}(\sqrt{\frac{H}{n}})\frac{H}{n} + \sqrt{\frac{H}{n}}\Big(\sqrt{q^{*,dep}(\sqrt{\frac{H}{n}})}\mathbb{E}[D(\hat{f})]^{1/2} + \delta\Big)\Big].\end{aligned}$$

Proof of Lemma 4.6.17. As in the proof of Lemma 4.6.6, let $j^* \in \{1, ..., \mathcal{N}_n\}$ be such that $\|\hat{f} - f_{j^*}\|_\infty \le \delta$. Since ε_i is independent of $\mathbb{X}_{i-1}$ and $\mathbb{E}\varepsilon_i = 0$, we have

$$\Big|\mathbb{E}\Big[\frac{1}{nd}\sum_{i=1}^n \langle \varepsilon_i, \hat{f}(\mathbb{X}_{i-1})\rangle \mathcal{W}(\mathbb{X}_{i-1})\Big]\Big| \le \delta \cdot \underbrace{\frac{1}{nd}\sum_{i=1}^n \mathbb{E}|\varepsilon_i|_1}_{\le \frac{1}{d}\sum_{k=1}^d \mathbb{E}|\varepsilon_{1k}| \le C_\varepsilon} + \frac{F}{n}|\mathbb{E}M_n(g_{j^*})|, \tag{4.6.81}$$

where $g_j(x) := \frac{1}{F}(f_j(x) - f_0(x))\mathcal{W}(x)$ and $M_n(\cdot)$ is from Lemma 4.6.15. We choose $\mathcal{G} = \{g_j : j \in \{1, ..., \mathcal{N}_n\}\}$. Note that

$$\sup_{j=1,...,\mathcal{N}_n} \sup_{k=1,...,d} \|g_{jk}\|_\infty \le \frac{1}{F} \cdot F \cdot \|\mathcal{W}\|_\infty \le 1$$

and

$$\begin{aligned} |g_{jk}(x) - g_{jk}(x')| &\le \frac{1}{F}|f_{jk}(x) - f_{0k}(x) - f_{jk}(x') - f_{0k}(x')| \cdot \mathcal{W}(x) \\ &\quad + \frac{1}{F}|f_{jk}(x') - f_{0k}(x')| \cdot |\mathcal{W}(x) - \mathcal{W}(x')| \\ &\le \Big(\frac{K + L_{\mathcal{F}}}{F} + \frac{2}{\varsigma}\Big) \cdot |x - x'|_\infty. \end{aligned}$$

That is, $\mathcal{G}$ satisfies (4.6.51) and (4.6.52) with $G = 1$ and $L_{\mathcal{G}} = \frac{K+L_{\mathcal{F}}}{F} + \frac{2}{\varsigma}$. With the argument (4.6.53), we conclude that for any $l \in \{1, ..., d\}$,

$$\delta_2^{g_{jl}(\mathbb{X})}(k) \le 2drL_{\mathcal{G}} \cdot \sup_{j\in\{1,...,r\}} \sup_{j=1,...,r} \delta_{2\theta}^{\mathbb{X}_{\cdot},l}(k-j)^\theta \le \Delta(k).$$

By Lemma 4.6.15, (4.6.66) and (4.6.67) (taking $q := q^{*,dep}(\sqrt{\frac{H}{n}})$), we obtain

$$\begin{aligned} &\mathbb{E}|M_n(g_{j^*})| \\ \le\; & cC_\varepsilon\sqrt{nH}\beta^{dep}(q^{*,dep}(\sqrt{\frac{H}{n}})) + \mathbb{E}|M_n^\circ(g_{j^*})| \\ \le\; & cC_\varepsilon H q^{*,dep}(\sqrt{\frac{H}{n}}) + cC_\varepsilon\Big(\sqrt{nH}(\sqrt{\log(q^{*,dep}(\sqrt{\frac{H}{n}})) + 1}) \\ & \qquad \times \mathbb{E}\big[\||g(\mathbb{X}_r)|_2\|_2\big|_{g=g_{j^*}}\big]^{1/2} + q^{*,dep}(\sqrt{\frac{H}{n}})^{1/2} H\Big) \\ \le\; & 2cC_\varepsilon\Big(Hq^{*,dep}(\sqrt{\frac{H}{n}}) + \sqrt{nH}\sqrt{q^{*,dep}(\sqrt{\frac{H}{n}}) \cdot \mathbb{E}[D(f_{j^*})]^{1/2}}\Big). \end{aligned} \tag{4.6.82}$$

Since $\|\hat{f}_k - f_{j^*k}\|_\infty \le \delta$, $k = 1, ..., d$, we have

$$\mathbb{E}[D(f_{j^*})]^{1/2} \le \frac{1}{\sqrt{d}}\||\hat{f}(\mathbb{X}_r) - f_{j^*}(\mathbb{X}_r)|_2\mathcal{W}(\mathbb{X}_r)\|_2 + \mathbb{E}[D(\hat{f})]^{1/2} \le \delta + \mathbb{E}[D(\hat{f})]^{1/2}. \tag{4.6.83}$$

Insertion of (4.6.82) and (4.6.83) into (4.6.81) yields the result. □

Lemma 4.6.18. *Suppose that X_i is of the form (4.1.7) and that Assumption 4.1.1 holds true. Assume that there exist $F > 0, L_{\mathcal{F}} > 0$ such that $\mathcal{F}$ satisfies (4.6.78) and (4.6.79). Furthermore, suppose that $f_0 : \mathbb{R}^{dr} \to \mathbb{R}^d$ from (4.1.1) is such that $|f_0(x) - f_0(x')|_\infty \le K|x - x'|_\infty$ for some $K > 0$.*

If additionally Assumption 4.1.5 is satisfied with $L_{\mathcal{G}} = 2dr\big(\frac{2}{\varsigma} + \frac{(L_{\mathcal{F}}+K)}{F}\big)$, then there exists some universal constant $c > 0$ such that for every $\eta > 0$,

$$\begin{aligned}\mathbb{E}D(\hat{f}) \;&\le\; (1+\eta)\mathbb{E}\hat{D}_n(\hat{f}) \\ &\quad +\Big\{(1+\eta)cF^2 q^{*,dep}(\sqrt{\frac{H}{n}})\frac{H}{n} + c\frac{F^2}{2}\eta(\tilde{V}^{-1})^*\Big(2\frac{1+\eta}{\eta}\sqrt{\frac{H}{n}}\Big)\Big\} \\ &\quad +(1+\eta)cF^2\sqrt{\frac{H}{n}}\tilde{V}(2F^{-2}\delta^2).\end{aligned}$$

Proof of Lemma 4.6.18. The proof follows a similar structure to Lemma 4.6.5. Let $(f_j)_{j=1,\dots,\mathcal{N}_n}$ be a δ-covering of $\mathcal{F}$ w.r.t. $\|\cdot\|_\infty$, where $\mathcal{N}_n := \mathbb{N}(\delta, \mathcal{F}, \|\cdot\|_\infty)$. Let $j^* \in \{1, \dots, \mathcal{N}_n\}$ be such that $\|\hat{f} - f_{j^*}\|_\infty \le \delta$. Without loss of generality, assume that $\delta \le F$.

Let $(X_i')_{i\in\mathbb{Z}}$ be an independent copy of the original time series $(X_i)_{i\in\mathbb{Z}}$. We have

$$\big|\mathbb{E}D(\hat{f}) - \mathbb{E}\hat{D}_n(\hat{f})\big| \le \frac{2F^2}{n}\mathbb{E}|S_n(g_{j^*})| + 10\delta F, \tag{4.6.84}$$

where for $x, x' \in \mathbb{R}^{dr}$,

$$g_j(x, x') := \frac{1}{2dF^2}|f_j(x') - f_0(x')|_2^2 \mathcal{W}(x') - \frac{1}{dF^2}|f_j(x) - f_0(x)|_2^2 \mathcal{W}(x),$$

and $S_n(\cdot)$ is from Lemma 4.6.14 based on the process (X_i, X_i'), where X_i', $i \in \mathbb{Z}$, is an independent copy of X_i, $i \in \mathbb{Z}$.

Here, due to the assumption on f_0 and on $\mathcal{F}$,

$$\begin{aligned}|g_j(x,x') - g_j(y,y')| \;&\le\; \frac{1}{2dF^2}\big(|f_j(x') - f_0(x')|_2^2 - |f_j(y') - f_0(y')|_2^2\big)\mathcal{W}(x') \\ &\quad + \frac{1}{2dF^2}|f_j(y') - f_0(y')|_2^2 \cdot |\mathcal{W}(x') - \mathcal{W}(y')| \\ &\quad + \frac{1}{2dF^2}\big(|f_j(x) - f_0(x)|_2^2 - |f_j(y) - f_0(y)|_2^2\big)\mathcal{W}(x) \\ &\quad + \frac{1}{2dF^2}|f_j(y) - f_0(y)|_2^2 \cdot |\mathcal{W}(x) - \mathcal{W}(y)| \\ &\le\; \Big(\frac{2}{\varsigma} + \frac{(L_{\mathcal{F}}+K)}{F}\Big)(|x-y|_\infty + |x'-y'|_\infty),\end{aligned}$$

and

$$\|g_j\|_\infty \le \frac{2}{2dF^2} \cdot dF^2 \cdot \|\mathcal{W}\|_\infty \le 1.$$

Thus, $\mathcal{G} = \{g_j : j = 1, \dots, \mathcal{N}_n\}$ satisfies the conditions (4.6.51) and (4.6.52) with $G = 1$ and

$$L_{\mathcal{G}} = \Big(\frac{2}{\varsigma} + \frac{L_{\mathcal{F}}+K}{F}\Big).$$

Since $X_i', i \in \mathbb{Z}$, has the same distribution as $X_i, i \in \mathbb{Z}$, the argument (4.6.53) yields for $j \in \{1, ..., \mathcal{N}_n\}$ that

$$\delta_2^{g_j(\mathbb{X}_{\cdot -1}, \mathbb{X}'_{\cdot -1})}(k) \leq 2drL_{\mathcal{G}} \cdot \sup_{l=1,...,r} \sup_{j=1,...,r} \delta_{2\theta}^{X_{\cdot},l}(k-j)^{\theta} \leq \Delta(k).$$

Furthermore, there exists another process $S_n^{\circ}(\cdot)$ and some universal constant $c > 0$ such that

$$\mathbb{E}|S_n(g_{j^*}) - S_n^{\circ}(g_{j^*})| \leq \mathbb{E} \sup_{g \in \mathcal{G}} |S_n(g) - S_n^{\circ}(g)| \leq c\sqrt{nH}\beta^{dep}(q).$$

For $q = q^*(\sqrt{\frac{H}{n}})$,

$$\mathbb{E}|S_n(g_{j^*}) - S_n^{\circ}(g_{j^*})| \leq cHq^*(\sqrt{\frac{H}{n}}). \tag{4.6.85}$$

Insertion of (4.6.85) and (4.6.57) into (4.6.84) yields

$$|\mathbb{E}D(\hat{f}) - \mathbb{E}\hat{D}_n(\hat{f})| \leq 2cF^2 \Big[q^*(\sqrt{\frac{H}{n}})\frac{H}{n} + \sqrt{\frac{H}{n}}\tilde{V}(\mathbb{E}[\|g(\mathbb{X}_r)\|_1 \,|_{g=g_{j^*}}])\Big]. \tag{4.6.86}$$

Now, observe that

$$\tilde{V}(\mathbb{E}[\|g(\mathbb{X}_r)\|_1 \,|_{g=g_{j^*}}]) \leq \tilde{V}(F^{-2}\mathbb{E}D(f_{j^*})) \leq \tilde{V}(2F^{-2}\delta^2) + \tilde{V}(2F^{-2}\mathbb{E}D(\hat{f}))$$

which together with Lemma 4.6.10 delivers,

$$\begin{aligned}\mathbb{E}D(\hat{f}) \quad &\leq \quad (1+\eta)\Big[\mathbb{E}\hat{D}_n(\hat{f}) + 2cF^2 q^*(\sqrt{\frac{H}{n}})\frac{H}{n} + 2cF^2\tilde{V}(2F^{-2}\delta^2)\Big] \\ &\quad + c\eta F^2(\tilde{V}^{-1})^*\Big(2\frac{1+\eta}{\eta}\sqrt{\frac{H}{n}}\Big).\end{aligned}$$

□

4.6.6 Proof of Section 4.2

We will now prove the theoretical results from Subsection 4.2.5.

Proof of Theorem 4.2.5. Choose $\eta = 1$ and $\delta = n^{-1}$. Applying Theorem 4.1.3, Assumption 4.1.2 and 4.1.1 we have

$$\mathbb{E}R(\hat{f}^{net}) - R(f_0) \lesssim \inf_{f \in \mathcal{F}(L,L_1,p,s,F)} \{R(f) - R(f_0)\} + \big(\Lambda^{mix}(\frac{\mathbb{H}(n^{-1})}{n}) + n^{-1}\big). \tag{4.6.87}$$

By Theorem 4.4.1 and Assumption 4.2.1, 4.2.3 and 4.2.4,

$$\inf_{f \in \mathcal{F}(L,L_1,p,s,F)} \{R(f) - R(f_0)\} \leq \inf_{f \in \mathcal{F}(L,L_1,p,s,F)} \|f - f_0\|_{\infty}^2 \lesssim (\frac{N}{n})^2 + N^{-2A}. \tag{4.6.88}$$

Proposition 4.6.19 and Assumption 4.2.4 deliver

$$\begin{aligned}\mathbb{H}(\delta) &\leq (s+1)\log(2^{2L+5}\delta^{-1}(L+1)p_0^2 p_{L+1}^2 s^{2L}) \lesssim sL\log(s) \\ &\lesssim N\log_2(n)\cdot\log_2(n)\log(n) \lesssim N\log(n)^3. \end{aligned} \quad (4.6.89)$$

Insertion of (4.6.88) and (4.6.89) into (4.6.87) yields the result. □

Proof of Theorem 4.2.6. Choose $\eta = 1$ and $\delta = n^{-1}$. Applying Theorem 4.1.6, Assumption 4.1.2 and 4.1.1 we have

$$\mathbb{E}R(\hat{f}^{net,lip}) - R(f_0) \lesssim \inf_{f\in\mathcal{F}(L,L_1,p,s,F)}\{R(f)-R(f_0)\} + \left(\Lambda^{dep}(\frac{\mathbb{H}(n^{-1})}{n}) + n^{-1}\right). \quad (4.6.90)$$

By Theorem 4.4.1 and Assumption 4.2.1, 4.2.3 and 4.2.4,

$$\inf_{f\in\mathcal{F}(L,L_1,p,s,F,\mathrm{Lip})}\{R(f)-R(f_0)\} \leq \inf_{f\in\mathcal{F}(L,L_1,p,s,F,\mathrm{Lip})}\|f-f_0\|_\infty \lesssim \frac{N}{n} + N^{-2A}. \quad (4.6.91)$$

Proposition 4.6.19 and Assumption 4.2.4 deliver

$$\begin{aligned}\mathbb{H}(\delta) &\leq (s+1)\log(2^{2L+5}\delta^{-1}(L+1)p_0^2 p_{L+1}^2 s^{2L}) \lesssim sL\log(s) \\ &\lesssim NL\log_2(n)\log(n) \lesssim N\log(n)^3. \end{aligned} \quad (4.6.92)$$

Insertion of (4.6.91) and (4.6.92) into (4.6.90) yields the result. □

4.6.7 Approximation results

In this section we consider the approximation error as well as the size of the corresponding network class.

Proof of the approximation error, Section 4.4

Proof of Theorem 4.4.1. We follow the proof given by [Schmidt-Hieber, 2017, Theorem 1] and employ [Schmidt-Hieber, 2017, Theorem 5] (recited here as part of Theorem 4.6.20), adapting it to the "encoder-decoder" structure. Since C is not explicitly given, it is enough to prove the result for large enough n. Fix $N \in \mathbb{N}$ and choose $m = \lceil \log_2(n) \rceil$.

By Theorem 4.6.20, we find for arbitrarily chosen $N > 0$ functions

$$\tilde{g}_{enc,0} \in \mathcal{F}(L_{enc,0}+2, (dr, p_{enc,0}, D), D(s_{enc,0}+4))$$

and

$$\tilde{g}_{enc,1} \in \mathcal{F}(L_{enc,1}+2, (D, p_{enc,1}, \tilde{d}), \tilde{d}(s_{enc,1}+4))$$

where

$$\begin{aligned} L_{enc,i} &= 8 + (m+5)(1+\log_2(t_{enc,i}\vee\beta_{enc,i})), \\ p_{enc,0} &= D(6(t_{enc,0}+\lceil\beta_{enc,0}\rceil)N, \dots, 6(t_{enc,0}+\lceil\beta_{enc,0}\rceil)N) \in \mathbb{R}^{L_{enc,0}+2}, \\ p_{enc,1} &= \tilde{d}(6(t_{enc,1}+\lceil\beta_{enc,1}\rceil)N, \dots, 6(t_{enc,1}+\lceil\beta_{enc,1}\rceil)N) \in \mathbb{R}^{L_{enc,1}+2} \end{aligned}$$

and

$$s_{enc,i} \leq 141((t_{enc,i} + \beta_{enc,i} + 1)^{3+t_{enc,i}} N(m+6), \qquad i = 0, 1,$$

such that

$$\|(g_{enc,i})_j - (\tilde{g}_{enc,i})_j\|_\infty \leq (2K+1)(1 + t_{enc,i}^2 + \beta_{enc,i}^2) 6^{t_{enc,i}} N 2^{-m} + K 3^{\beta_{enc,i}} N^{\frac{\beta_{enc,i}}{t_{enc,i}}}$$

for $i = 0, 1$. The composed network $\tilde{f}_{enc} := \tilde{g}_{enc,1} \circ \tilde{g}_{enc,0}$ satisfies

$$\tilde{f}_{enc} \in \mathcal{F}(L_{enc,0} + L_{enc,1} + 5, (dr, p_{enc,0}, D, p_{enc,1}, \tilde{d}), \tilde{d}(D(s_{enc,0} + 4) + \tilde{d}(s_{enc,1} + 4)))$$

as well as $\tilde{f}_{enc} \in \mathcal{F}(L_1, \bar{p}, \bar{s})$ for

$$L_{enc,0} + L_{enc,1} + 5 \leq \sum_{i \in \{enc,0; enc,1\}} \log_2(4(t_i \vee \beta_i)) \log_2(n) \leq L_1$$

(the first inequality holds true for n large enough) and

$$\begin{aligned} \bar{p} &:= (\underbrace{dr, ..., dr}_{(\bar{k}+1) \text{ times}}, p_{enc,0}, D, p_{enc,1}, \tilde{d}) \\ \bar{s} &:= \tilde{d}(D(s_{enc,0} + 4) + \tilde{d}(s_{enc,1} + 4))) + \bar{k}dr \end{aligned}$$

where $\bar{k} := L_1 - (L_{enc,0} + L_{enc,1} + 5)$ (cf. [Schmidt-Hieber, 2017, Section 7.1]). Furthermore, by Theorem 4.6.20 there exists a network

$$\tilde{f}_{dec} \in \mathcal{F}(L_{dec} + 2, (\tilde{d}, p_{dec}, d), d(s_{dec} + 4))$$

where

$$\begin{aligned} L_{dec} &= 8 + (m+5)(1 + \log_2(t_{dec} \vee \beta_{dec})), \\ p_{dec} &= d(6(t_{dec} + \lceil \beta_{dec} \rceil)N, \ldots, 6(t_{dec} + \lceil \beta_{dec} \rceil)N) \in \mathbb{R}^{L_{dec}+2}, \\ s_{dec} &\leq 141((t_{dec} + \beta_{dec} + 1)^{3+t_{dec}} N(m+6), \end{aligned}$$

such that

$$\|(f_{dec})_j - (\tilde{f}_{dec})_j\|_\infty \leq (2K+1)(1 + t_{dec,i}^2 + \beta_{dec,i}^2) 6^{t_{dec,i}} N 2^{-m} + K 3^{\beta_{dec,i}} N^{\frac{\beta_{dec,i}}{t_{dec,i}}}$$

for $j = 1, ..., d$. We then obtain $\tilde{f}_0 = \tilde{f}_{dec} \circ \tilde{f}_{enc} \in \mathcal{F}(L', p', s')$ by composing the networks $\tilde{f}_{enc}$ and $\tilde{f}_{dec}$ (cf. [Schmidt-Hieber, 2017, Section 7.1]) with the values

$$\begin{aligned} L' &:= L_1 + L_{dec} + 1, \\ p' &:= (\bar{p}, p_{dec}, d), \\ s' &:= \bar{s} + d(s_{dec} + 4). \end{aligned}$$

The composition also satisfies $\tilde{f}_0 \in \mathcal{F}(L, p, s)$ for additional layers where

$$L' \leq L_1 + \log_2(4(t_{dec} \vee \beta_{dec})) \log_2(n) \leq L$$

(the first inequality holds true for n large enough) and s, p are set according to [Schmidt-Hieber, 2017, Section 7.1, equation (18)], i.e.

$$k = L - L', \qquad p = (\underbrace{dr, ..., dr}_{k \text{ times}}, p'), \qquad s = s' + kp'_0.$$

The conditions (ii) to (v) are automatically met. In analogy to [Schmidt-Hieber, 2017, Section 7.1, Lemma 3],

$$\|\tilde{f}_0 - f_0\|_\infty^2 \quad \leq \quad C \max_{k \in \{dec; enc,0; enc,1\}} \left\{ \frac{N}{n} + N^{-\frac{2\beta_k}{t_k}} \right\} \tag{4.6.93}$$

for a constant C that only depends on $\boldsymbol{t}, \boldsymbol{\beta}$. By Theorem 4.6.20, since $N2^{-m} \lesssim 1$, $\tilde{f}_0$ has Lipschitz constant

$$\|\tilde{f}_0\|_{\text{Lip}} \leq \|\tilde{f}_{dec}\|_{\text{Lip}} \cdot \|\tilde{g}_{enc,1}\|_{\text{Lip}} \cdot \|\tilde{g}_{enc,0}\|_{\text{Lip}} \leq C_2$$

for a constant C_2 only depending on $\boldsymbol{\beta}$, $\boldsymbol{t}$.

Up to now, $\tilde{f}_0$ is not bounded by a given F. For large enough n we are able to generate a sequence $(\tilde{f}_n)_{n \in \mathbb{N}}$ in $\mathcal{F}(L, L_1, p, s, \bar{F}, C_2)$ ($\bar{F}$ chosen arbitrarily large) satisfying equation (4.6.93). If we define $f_n^* := (\frac{\|f_0\|_\infty}{\|\tilde{f}_n\|_\infty} \wedge 1)\tilde{f}_n$,

$$\|f_n^*\|_\infty \leq \|f_0\|_\infty \leq \|f_{dec}\|_\infty \leq K \leq F$$

by assumption (i). Therefore, $f_n^* \in \mathcal{F}(L, L_1, p, s, F, C_2)$. Equation (4.6.93) also holds true for the class $\mathcal{F}(L, L_1, p, s, F, C_2)$ since $\|f_n^* - f_0\|_\infty \leq 2\|\tilde{f}_n - f_0\|_\infty$. This completes the proof. □

We cite [Schmidt-Hieber, 2017, Remark 1] in order to maintain a consistent reading flow and for the sake of completeness.

Proposition 4.6.19. *For the network $\mathcal{F}(L, L_1, p, s, \infty)$ we have the covering entropy bound*

$$\log \mathcal{N}(\delta, \mathcal{F}(L, L_1, p, s, \infty), \|\cdot\|_\infty) \leq (s+1) \log(2^{2L+5}\delta^{-1}(L+1)p_0^2 p_{L+1}^2 s^{2L}).$$

Approximation error and Lipschitz continuity of neural networks

The first part of the following theorem is taken from [Schmidt-Hieber, 2017, Theorem 5]. The second part (4.6.94) is proven below.

Theorem 4.6.20. *For any function $f \in C_t^\beta([0,1]^t, K)$ and any integers $m \geq 1$, $N \geq (\beta+1)^t \vee (K+1)e^t$, there exists a network*

$$\tilde{f} \in \mathcal{F}(L, (t, 6(t + \lceil \beta \rceil)N, ..., 6(t + \lceil \beta \rceil)N, 1), s, \infty)$$

with depth

$$L = 8 + (m+5)(1 + \lceil \log_2(t \vee \beta) \rceil)$$

and number of active parameters

$$s \leq 141(t+\beta+1)^{3+r} N(m+6)$$

such that

$$\|\tilde{f} - f\|_\infty \leq (2K+1)(1+t^2+\beta^2)6^t N 2^{-m} + K 3^\beta N^{-\frac{\beta}{t}}.$$

Furthermore, $\tilde{f}$ satisfies for any $x, y \in [0,1]^t$,

$$|\tilde{f}(x) - \tilde{f}(y)| \leq \mathrm{Lip}(N,m) \cdot |x-y|_\infty \tag{4.6.94}$$

where

$$\mathrm{Lip}(N,m) := 2\beta F(K+1)e^t(24t^6 2^t N 2^{-m} + 3t).$$

To prove (4.6.94), we first recap how $\tilde{f}$ is constructed in [Schmidt-Hieber, 2017, Theorem 5].

As in Schmidt-Hieber [2017], we define for $x, y \in [0,1]$, $m \in \mathbb{N}$,

$$\mathrm{mult}_m(x,y) := \Big(\sum_{k=1}^{m+1} \big\{R^k(\frac{x-y+1}{2}) - R^k(\frac{x+y}{2})\big\} + \frac{x+y}{2} - \frac{1}{4}\Big)_+$$

where

$$R^k := T^k \circ T^{k-1} \circ ... \circ T^1, \quad k \in \mathbb{N},$$

and

$$T^k(x) := \min\{\frac{x}{2}, 2^{1-2k} - \frac{x}{2}\}, \quad k \in \mathbb{N}.$$

Lemma 4.6.21. *For $x, y \in [0,1]$ where $mult_m$ is differentiable, it holds that*

$$\partial_1 mult_m(x,y) = y + res_1(x,y), \qquad \partial_2 mult_m(x,y) = x + res_2(x,y)$$

where $|res_i(x,y)| \leq 2^{-m-1}$, $i = 1, 2$. Furthermore,

$$|mult_m(x,y) - x \cdot y| \leq 2^{-m-1}(x+y) \leq 2^{-m}.$$

Proof of Lemma 4.6.21. A straightforward calculation yields

$$\partial_1 R^k(x) = \begin{cases} \frac{1}{2^k}, & x \in A_{k+}, \\ -\frac{1}{2^k}, & x \in [0,1] \setminus A_{k+} \end{cases} = \frac{1}{2^k}(2 \cdot \mathbb{1}_{A_{k+}}(x) - 1)$$

where

$$A_{k+} := \bigcup_{j=0}^{2^k-1} [\frac{j}{2^k}, \frac{j+1}{2^k}].$$

We conclude that

$$\begin{aligned} \partial_1 \mathrm{mult}_m(x,y) &= \sum_{k=1}^{m+1} \big\{\partial_1 R^k(\frac{x-y+1}{2}) \cdot \frac{1}{2} - \partial_1 R^k(\frac{x+y}{2}) \cdot \frac{1}{2}\big\} + \frac{1}{2} \\ &= \sum_{k=1}^{m+1} \frac{1}{2^k}\big\{\mathbb{1}_{A_{k+}}(\frac{x-y+1}{2}) - \mathbb{1}_{A_{k+}}(\frac{x+y}{2})\big\} + \frac{1}{2}. \end{aligned} \tag{4.6.95}$$

Suppose that the following binary representations are valid for $x, y \in [0,1]$:

$$\frac{x+y}{2} = \sum_{k=1}^{\infty} \frac{a_k}{2^k}, \qquad \frac{x-y+1}{2} = \sum_{k=1}^{\infty} \frac{b_k}{2^k}$$

where $a_k, b_k \in \{0,1\}$ $(k \in \mathbb{N})$. Then,

$$\mathbb{1}_{A_{k+}}(\frac{x-y+1}{2}) = 1 - b_k, \qquad \mathbb{1}_{A_{k+}}(\frac{x+y}{2}) = 1 - a_k.$$

Insertion into (4.6.95) yields

$$\partial_1 \mathrm{mult}_m(x,y) \;=\; \sum_{k=1}^{m+1} \frac{1}{2^k}\{a_k - b_k\} + \frac{1}{2} = \frac{x+y}{2} - \frac{x-y+1}{2} + \frac{1}{2} + \mathrm{res}(x,y) = y + \mathrm{res}(x,y)$$

where

$$\mathrm{res}_k(x,y) := \sum_{k=m+2}^{\infty} \frac{b_k}{2^k} - \sum_{k=m+2}^{\infty} \frac{a_k}{2^k}.$$

Due to $a_k, b_k \in \{0,1\}$ $(k \in \mathbb{N})$, we see that $|\mathrm{res}(x,y)| \leq 2^{-(m+1)}$. The proof for $\partial_2 \mathrm{mult}_m$ is similar.

The second statement follows by the first one using the fundamental theorem of analysis:

$$\begin{aligned} \big|\mathrm{mult}_m(x,y) - xy\big| &\leq x\int_0^1 |\partial_1 \mathrm{mult}_m(xt,yt) - yt|dt + y\int_0^1 |\partial_2 \mathrm{mult}_m(xt,yt) - xt|dt \\ &\leq 2^{-m-1}(x+y). \end{aligned}$$

□

As in Schmidt-Hieber [2017], define recursively for $x \in [0,1]$,

$$\mathbb{M}_m(x) := x,$$

for $x = (x_1, ..., x_{2^q}) \in [0,1]^{2^q}$, $q \in \mathbb{N}$,

$$\mathbb{M}_m(x) := \mathrm{mult}_m(\mathbb{M}_m(x_1, ..., x_{2^{q-1}}), \mathbb{M}_m(x_{2^{q-1}+1}, ..., x_{2^q})),$$

and for $x = (x_1, ..., x_t) \in [0,1]^t$, $q = \lceil \log(r) \rceil$,

$$\mathbb{M}_m(x) := \mathbb{M}_m(x_1, ..., x_t, \underbrace{1, ..., 1}_{(2^q - t)\text{ ones}}).$$

The first part of the following lemma is taken from [Schmidt-Hieber, 2017, Lemma A.3].

Lemma 4.6.22. *For $y \in [0,1]^t$ we have*

$$|\mathbb{M}_m(y_1, ..., y_t) - \prod_{k=1}^{r} y_k| \leq t^2 \cdot 2^{-m} \tag{4.6.96}$$

and for $j \in \{1, ..., t\}$, at the points y where $\mathbb{M}_m$ is differentiable,

$$|\partial_j \mathbb{M}_m(y_1, ..., y_t) - \prod_{k=1, k\neq j}^{t} y_k| \leq 2t^3 \cdot 2^{-m}. \tag{4.6.97}$$

Proof of Lemma 4.6.22. We only have to show (4.6.97). We restrict ourselves to $j = 1$ for simplicity. With some abuse of notation, overload $y := (y, 1, ..., 1)$ (where we added $2^q - t$ entries of 1). Then by Lemma 4.6.21, (4.6.96) and $|y_k| \leq 1$ ($k = 1, ..., 2^q$),

$$\begin{aligned}
&|\partial_{y_1} \mathbb{M}_m(y) - \prod_{k=2}^{t} y_k| \\
\leq\ & \sum_{i=1}^{q} \Big\{ \Big(\prod_{k=2^{q-i+1}+1}^{2^q} y_k \Big) \\
&\qquad \times \Big| \partial_1 \mathbb{M}_m(\mathbb{M}_m(y_1, ..., y_{2^{q-i}}), \mathbb{M}_m(y_{2^{q-i}+1}, ..., y_{2^{q-i+1}})) - \prod_{k=2^{q-i}+1}^{2^{q-i+1}} y_k \Big| \\
&\qquad \times \prod_{j=i+1}^{q} \partial_1 \mathbb{M}_m(\mathbb{M}_m(y_1, ..., y_{2^{q-j}}), \mathbb{M}_m(y_{2^{q-j}+1}, ..., y_{2^{q-j+1}})) \Big\} \\
\leq\ & \sum_{i=1}^{q} \Big\{ \Big(\prod_{k=2^{q-i+1}+1}^{2^q} y_k \Big) \cdot \Big(\Big|\mathbb{M}_m(y_{2^{q-i}+1}, ..., y_{2^{q-i+1}}) - \prod_{k=2^{q-i}+1}^{2^{q-i+1}} y_k \Big| + 2^{-m-1} \Big) \\
&\qquad \times \prod_{j=i+1}^{q} \Big(\mathbb{M}_m(y_{2^{q-j}+1}, ..., y_{2^{q-j+1}}) + 2^{-m-1} \Big) \Big\} \\
\leq\ & \sum_{i=1}^{q} \Big\{ \Big(\prod_{k=2^{q-i+1}+1}^{2^q} y_k \Big) \cdot 2 \cdot 4^{q-i} 2^{-m} \cdot \prod_{j=i+1}^{q} \Big(\prod_{k=2^{q-j}+1}^{2^{q-j+1}} y_k + 4^{q-j} 2^{-m} \Big) \Big\} \\
\leq\ & 2^{-m+1} \cdot \sum_{i=1}^{q} \Big\{ 4^{q-i} \cdot \prod_{j=i+1}^{q} (1 + 4^{q-j} 2^{-m}) \Big\} \\
\leq\ & 2^{-m} \sum_{i=1}^{q} 8^{q-i} \leq \frac{8}{7} t^3 2^{-m}.
\end{aligned}$$

□

Now we show (4.6.94). To do so, we derive the mathematical expression Q_3 used in [Schmidt-Hieber, 2017, Theorem 5] to describe $\tilde{f}$.

Let M be the largest integer such that $(M+1)^t \leq N$. Define the grid

$$D(M) := \{x_l := (\ell_j/M)_{j=1,\dots,t} : (\ell_1, ..., \ell_t) \in \{0,1,...,M\}^t\}.$$

For $x, y \in [0,1]$, put

$$I_y(x) := (\frac{1}{M} - |x-y|)_+,$$

and for $x, y \in [0,1]^t$,

$$\mathrm{Hat}_x(y) := \mathbb{M}_m(I_{x_1}(y_1), ..., I_{x_r}(y_r)).$$

For $a, x \in [0,1]^r$, let

$$P_a^\beta f(y) = \sum_{0 \leq |\alpha| < \beta} (\partial^\alpha f)(a) \cdot \frac{(y-a)^\alpha}{\alpha!} = \sum_{0 \leq |\gamma| < \beta} c_\gamma(a) \cdot y^\gamma$$

denote the multivariate Taylor polynomial of f with degree β at a. In the above formula, α and γ denote multi-indices.

Then (cf. Schmidt-Hieber [2017], (31)-(35) therein), $|c_\gamma| \leq \frac{K}{\gamma!}$ and $\sum_{\gamma \geq 0} |c_\gamma| \leq Ke^t \leq \frac{1}{2}B$, where $B := \lceil 2Ke^t \rceil$. Put

$$Q_1(y)_{x_l} := \frac{1}{B} \sum_{0 \leq |\gamma| < \beta} c_\gamma(x_l) \cdot \mathbb{M}_m(y_\gamma) + \frac{1}{2},$$

where $y_\gamma := (y_{\gamma_1}, y_{\gamma_2}, ..., y_{\gamma_t})$. Define

$$Q_2(y) := \sum_{x_l \in D(M)} \mathrm{mult}_m\big(Q_1(y)_{x_l}, \mathrm{Hat}_{x_l}(y)\big)$$

and

$$Q_3 := S \circ Q_2$$

where $S(x) := BM^t(x - \frac{1}{2M^t})$. Since $\tilde{f} = Q_3$, (4.6.94) follows from Lemma 4.6.23.

Lemma 4.6.23. *For $x, y \in [0,1]^t$,*

$$|Q_3(x) - Q_3(y)| \leq \beta F B(24t^6 2^t N 2^{-m} + 3t) \cdot |x-y|_\infty.$$

Proof of Lemma 4.6.23. Since Q_2 is piecewise linear, it is enough to consider its first derivative at the points where it is differentiable to derive its Lipschitz constant.

With $q = \lceil \log(t) \rceil$,

$$\begin{aligned} \partial_{y_1} \mathrm{Hat}_{x_l}(y) &= \partial_1 \mathbb{M}_m(\mathbb{M}_m(y_1, ..., y_{2^{q-1}}), \mathbb{M}_m(y_{2^{q-1}+1}, ..., y_t, 1, ..., 1)) \\ &\quad \times \prod_{i=2}^{q-1} \partial_1 \mathbb{M}_m(\mathbb{M}_m(y_1, ..., y_{2^{q-i}}), \mathbb{M}_m(y_{2^{q-i}+1}, ..., y_{2^{q-i+1}})) \\ &\quad \times \partial_1 \mathrm{mult}_m(y_1, y_2). \end{aligned}$$

By Lemma 4.6.22,

$$
\begin{aligned}
|Q_1(y)_{x_l} - \big(\frac{1}{B}P^{\beta}_{x_l}f(y) + \frac{1}{2}\big)| &\leq \frac{1}{B}\sum_{0\leq|\gamma|<\beta} |c_\gamma(x_l)|\cdot|\mathbb{M}_m(y_\gamma) - y^\gamma| \\
&\leq \frac{r^2 2^{-m}}{B}\cdot\sum_{0\leq|\gamma|<\beta}|c_\gamma| \leq \frac{1}{2}t^2 2^{-m}. \qquad (4.6.98)
\end{aligned}
$$

Furthermore,

$$
\partial_{y_1}Q_1(y)_{x_l} = \frac{1}{B}\sum_{0\leq|\gamma|<\beta,\gamma_1\geq 1} c_\gamma(x_l)\cdot\partial_{y_1}\mathbb{M}_m(y_\gamma) = \frac{1}{B}\sum_{0\leq|\gamma|<\beta} c_\gamma(x_l)\cdot\sum_{j=1}^{\gamma_1}\partial_j\mathbb{M}_m(y_\gamma).
$$

Thus by Lemma 4.6.22,

$$
\begin{aligned}
\big|\partial_{y_1}Q_1(y)_{x_l} - \frac{1}{B}\partial_{y_1}P^{\beta}_{x_l}f(y)\big| &\leq \frac{1}{B}\sum_{0\leq|\gamma|<\beta,\gamma_1\geq 1}|c_\gamma(x_l)|\cdot\sum_{j=1}^{\gamma_1}\big|\partial_j\mathbb{M}_m(y_\gamma) - y^{\gamma-(1,0,\dots,0)}\big| \\
&\leq \frac{1}{B}\sum_{0\leq|\gamma|<\beta}|c_\gamma(x_l)|\cdot 2\gamma_1 t^3 2^{-m} \leq \beta t^3 2^{-m}. \qquad (4.6.99)
\end{aligned}
$$

Finally, Lemma 4.6.22 yields

$$
\big|\mathrm{Hat}_{x_l}(y) - \prod_{k=1}^{t} I_{(x_l)_k}(y_k)\big| \leq t^2 2^{-m} \qquad (4.6.100)
$$

and for $j\in\{1,...,t\}$, since $\partial_{y_j}I_{(x_l)_j}(y_j)\in\{-1,+1\}$,

$$
\begin{aligned}
&\big|\partial_{y_j}\mathrm{Hat}_{x_l}(y) - \partial_{y_j}\prod_{k=1}^{t} I_{(x_l)_k}(y_k)\big| \\
=\ &\Big|\partial_j\mathbb{M}_m(I_{(x_l)_1}(y_1),...,I_{(x_l)_t}(y_t)) - \prod_{k=1,k\neq j}^{t} I_{(x_l)_k}(y_k)\Big|\cdot|\partial_{y_j}I_{(x_l)_j}(y_j)| \\
\leq\ & t^2 2^{-m}. \qquad (4.6.101)
\end{aligned}
$$

Note furthermore that

$$
|\partial_{y_j}\prod_{k=1}^{t} I_{(x_l)_k}(y_k)| \leq \prod_{k=1,k\neq j}^{t} I_{(x_l)_k}(y_k)\cdot|\partial_{y_j}I_{(x_l)_j}(y_j)| \leq 1
$$

and

$$
|\frac{1}{B}\partial_{y_1}P^{\beta}_{x_l}f(y)| \leq \frac{1}{B}\sum_{0\leq|\gamma|<\beta}|c_\gamma(x_l)|\cdot\gamma_1 y^{\gamma-(1,0,\dots,0)} \leq \frac{\beta}{2}.
$$

By Lemma 4.6.21, (4.6.99) and (4.6.101), it holds that

$$\begin{aligned}
&\Big|\partial_{y_1} Q_2(y) \\
&\qquad - \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \Big\{ \mathrm{Hat}_{x_l}(y) \cdot \partial_{y_1} Q_1(y)_{x_l} + Q_1(y)_{x_l} \cdot \partial_{y_1} \mathrm{Hat}_{x_l}(y) \Big\} \Big| \\
\le\ & \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \Big\{ \big| \partial_1 \mathrm{mult}_m\big(Q_1(y)_{x_l}, \mathrm{Hat}_{x_l}(y)\big) - \mathrm{Hat}_{x_l}(y) \cdot \partial_{y_1} Q_1(y)_{x_l} \big| \\
&\qquad \times |\partial_{y_1} Q_1(y)_{x_l}| \\
&\qquad + \big| \partial_2 \mathrm{mult}_m(Q_1(y)_{x_l}, \mathrm{Hat}_{x_l}(y)) - Q_1(y)_{x_l} \big| \cdot \big| \partial_{y_1} \mathrm{Hat}_{x_l}(y) \big| \Big\} \\
\le\ & 2^{-m-1} \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \big\{ |\partial_{y_1} Q_1(y)_{x_l}| + |\partial_{y_1} \mathrm{Hat}_{x_l}(y)| \big\} \\
\le\ & 2^{-m-1} \cdot 2^t \cdot \big\{ (\beta t^3 2^{-m} + \frac{\beta}{2}) + (t^2 2^{-m} + 1) \big\} \\
\le\ & 4 \beta t^3 2^t \cdot 2^{-m}.
\end{aligned} \tag{4.6.102}$$

In a similar manner, we obtain with (4.6.98), (4.6.99), (4.6.100) and (4.6.101) that

$$\begin{aligned}
&\Big| \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \mathrm{Hat}_{x_l}(y) \cdot \partial_{y_1} Q_1(y)_{x_l} \\
&\qquad - \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \Big(\prod_{k=1}^{t} I_{(x_l)_k}(y_k) \Big) \cdot \frac{1}{B} \partial_{y_1} P^{\beta}_{x_l} f(y) \Big| \\
\le\ & 2^t \cdot \big(t^2 2^{-m} \cdot (\beta t^3 2^{-m} + \frac{\beta}{2}) + 1 \cdot (\beta t^3 2^{-m}) \big) \\
\le\ & 4 \beta t^5 2^t \cdot 2^{-m}
\end{aligned} \tag{4.6.103}$$

and

$$\begin{aligned}
&\Big| \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} Q_1(y)_{x_l} \cdot \partial_{y_1} \mathrm{Hat}_{x_l}(y) \\
&\qquad - \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \Big\{ \big(\frac{1}{B} P^{\beta}_{x_l} f(y) + \frac{1}{2} \big) \cdot \Big(\prod_{k=2}^{t} I_{(x_l)_k}(y_k) \Big) \cdot \partial_{y_1} I_{(x_l)_1}(y_1) \Big\} \Big| \\
\le\ & 2^t \cdot \big(\frac{1}{2} t^2 2^{-m} \cdot (t^2 2^{-m} + 1) + 1 \cdot (t^2 2^{-m}) \big) \\
\le\ & 4 t^4 2^t \cdot 2^{-m}.
\end{aligned} \tag{4.6.104}$$

Now, we have

$$\begin{aligned} &\Big| \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \Big(\prod_{k=1}^{t} I_{(x_l)_k}(y_k) \Big) \cdot \frac{1}{B} \partial_{y_1} P_{x_l}^{\beta} f(y) \Big| \\ \le\ & \frac{\beta}{2} \cdot \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \Big(\prod_{k=1}^{t} I_{(x_l)_k}(y_k) \Big) \le \frac{\beta}{2} \cdot M^{-t}. \end{aligned} \quad (4.6.105)$$

Let $u \in D(M)$ be the grid point which satisfies $u_j \le y_j \le u_j + M^{-1}$, $j = 1, ..., t$.

Let $b(|\alpha|) := b$, if $|\alpha| = \beta - 1$ and $b(\alpha) = 1$, otherwise. For general $a, a' \in [0,1]^t$ with $|y - a|_\infty, |y - a'|_\infty \le M^{-1}$, $|a - a'|_\infty \le M^{-1}$, it holds true that

$$\begin{aligned} |P_a^\beta f(y) - P_{a'}^\beta f(y)| \le & \sum_{0 \le |\alpha| < \beta} \frac{1}{\alpha!} \cdot \{ \big| \partial^\alpha f(a) - \partial^\alpha f(a') \big| \cdot |(y-a)^\alpha| \\ & + |\partial^\alpha f(a')| \cdot \big| (y-a)^\alpha - (y-a')^\alpha \big| \} \\ \le\ & \sum_{0 \le |\alpha| < \beta} \frac{1}{\alpha!} \cdot \{ K |a - a'|_\infty^{b(|\alpha|)} M^{-|\alpha|} \\ & + (K+F) \sum_{j=1}^{t} \Big(\prod_{k=1}^{j-1} |y_k - a_k|^{\alpha_k} \Big) \cdot \Big(\prod_{k=j+1}^{t} |y_k - a'_k|^{\alpha_k} \Big) \cdot |(y_j - a_j)^{\alpha_j} - (y_j - a'_j)^{\alpha_j}| \} \\ \le\ & \sum_{0 \le |\alpha| < \beta} \frac{1}{\alpha!} \cdot \{ K |a - a'|_\infty^{b(|\alpha|)} M^{-|\alpha|} + \sum_{j=1}^{t} \alpha_j M^{-(\alpha_j - 1)} |a_j - a'_j| \} \\ \le\ & \sum_{0 \le |\alpha| < \beta} \frac{1}{\alpha!} \cdot \{ K |a - a'|_\infty^{b(|\alpha|)} M^{-|\alpha|} + (K+F) \beta M^{-1} \} \\ \le\ & (K + \beta K + \beta F) e^t M^{-1}. \end{aligned}$$

The last step is due to the fact that f is assumed to have at least Hölder exponent 1.

Using this result, we obtain

$$\begin{aligned} & \Big| \sum_{x_l \in D(M), |x_l - y|_\infty \le M^{-1}} \Big\{ \big(\frac{1}{B} P_{x_l}^\beta f(y) + \frac{1}{2} \big) \cdot \Big(\prod_{k=2}^{t} I_{(x_l)_k}(y_k) \Big) \cdot \partial_{y_1} I_{(x_l)_1}(y_1) \Big\} \Big| \\ \le\ & \sum_{(i_2,...,i_t) \in \{0,1\}^r} \Big(\prod_{k=2}^{t} I_{u_k + \frac{i_k}{M}}(y_k) \Big) \cdot \Big| \big(\frac{1}{B} P^\beta_{(u_1 + M^{-1}, u_2 + \frac{i_2}{M}, ..., u_t + \frac{i_t}{M})} f(y) + \frac{1}{2} \big) \cdot \partial_{y_1} I_{u_1 + M^{-1}}(y_1) \\ & + \big(\frac{1}{B} P^\beta_{(u_1, u_2 + \frac{i_2}{M}, ..., u_t + \frac{i_t}{M})} f(y) + \frac{1}{2} \big) \cdot \partial_{y_1} I_{u_1}(y_1) \Big| \\ \le\ & \frac{1}{B} \sum_{(i_2,...,i_t) \in \{0,1\}^t} \Big(\prod_{k=2}^{t} I_{u_k + \frac{i_k}{M}}(y_k) \Big) \cdot \big| P^\beta_{(u_1 + M^{-1}, u_2 + \frac{i_2}{M}, ..., u_t + \frac{i_t}{M})} f(y) - P^\beta_{(u_1, u_2 + \frac{i_2}{M}, ..., u_t + \frac{i_t}{M})} f(y) \big| \\ \le\ & \frac{(K + \beta K + \beta F) e^t}{B} M^{-(t-1)} \cdot M^{-1} = \frac{(K + \beta K + \beta F) e^t}{B} \cdot M^{-t}. \end{aligned} \quad (4.6.106)$$

Using the bounds (4.6.102), (4.6.103), (4.6.104), (4.6.105) and (4.6.106), we obtain with $K \geq 1$ that

$$|\partial_{y_1} Q_2(y)| \leq 24\beta F t^5 2^t 2^{-m} + 3\beta M^{-t}.$$

The proof for the other derivatives ∂_{y_j}, $j = 2, ..., t$, is completely similar. Thus, for $x, y \in [0,1]^t$,

$$|Q_2(y) - Q_2(x)| \leq \int_0^1 |\langle \partial Q_2(x + t(y-x)), y - x\rangle| dt \leq t \sup_y |\partial Q_2(y)|_\infty \cdot |y - x|_\infty.$$

We obtain

$$|Q_3(x) - Q_3(y)| \leq BM^t |Q_2(x) - Q_2(y)| \leq \beta F B(24 t^6 2^t M^t 2^{-m} + 3t) \cdot |x - y|_\infty.$$

□

Chapter 5

Multiplicative deconvolution in survival analysis under dependence

A popular branch of Statistics is represented by survival analysis. It has many applications in, for example, Economics and Biology. So far, there are a lot of applications regarding independent and identically distributed observations which we do not have access to most of the time. In this chapter we extent the theory to data with certain dependence structures and consider the estimation of survival functions from an inverse problem point of view.

5.1 Motivation

In survival analysis one primary interest is the survival function. As always, we most often do not have access to such in real life and therefore have to recover it, given our observations. We will be focusing on the following scenario. Let $S : \mathbb{R}_+ \to \mathbb{R}_+$ be the unknown survival function of a positive random variable X given identically distributed copies of $Y = XU$ where X and U are independent of each other. We assume that U has a known density $g : \mathbb{R}_+ \to \mathbb{R}_+$. In this setting, the density $f_Y : \mathbb{R}_+ \to \mathbb{R}_+$ of Y is given by

$$f_Y(y) = [f * g](y) := \int_0^\infty f(x)g(y/x)x^{-1}dx \quad \forall y \in \mathbb{R}_+$$

where, as usual, "$*$" denotes the operation of multiplicative convolution. The estimation of S using a sample $Y_1, \dots, Y_n$ from f_Y is thus a multiplicative deconvolution problem. Under this view point, we can build up our theory by means of Mellin transforms, whose concept is closely related to that of Fourier transforms.

Subsequently, we assume that the errors $U_1, \dots, U_n$ are independent and identically distributed (i.i.d.) and that $X_1, \dots, X_n$ are i.i.d. or a stationary process with the following, already established, dependence structures. We briefly revise them here in this context for completeness.

Absolutely regular mixing (β-mixing) First, let us consider two sigma fields $\mathcal{U}, \mathcal{V}$ over some probability space Ω and define the quantity

$$\beta^{mix}(\mathcal{U}, \mathcal{V}) := \frac{1}{2} \sup \sum_{(p,q)\in P\times Q} |\mathbb{P}(U_p \cap V_q) - \mathbb{P}(U_p)\mathbb{P}(V_q)|$$

where the supremum is taken over all finite partitions $(U_p)_{p\in P}$, $(V_q)_{q\in Q}$ of Ω such that $(U_p)_{p\in P} \subset \mathcal{U}$, $(V_q)_{q\in Q} \subseteq \mathcal{V}$. Now, let $(\beta(k))_{k\in\mathbb{N}_0}$ be a sequence of real-valued numbers defined by

$$\beta(k) := \beta(X_0, X_k) := \beta^{mix}(\sigma(X_0), \sigma(X_k)), \tag{5.1.1}$$

for σ-fields generated by X_0 and X_k, respectively. Then, the process is said to be β-mixing if for the corresponding coefficients $\beta(k) \to 0$ as $k \to \infty$. We refer to the previous chapters for more references on this subject.

Functional dependence measure We assume that the given process has a representation as a Bernoulli shift process. Let X_j, $j = 1, ..., n$, be a one dimensional process of the form

$$X_j = J_{j,n}(\mathcal{G}_j), \tag{5.1.2}$$

where $\mathcal{G}_j = (\varepsilon_j, \varepsilon_{j-1}, ...)$ for ε_j, $j \in \mathbb{Z}$, a sequence of i.i.d. random variables in $\mathbb{R}$, and some measurable function $J_{j,n} : \mathbb{R}^{\mathbb{N}_0} \to \mathbb{R}$, $j = 1, ..., n$, $n \in \mathbb{N}$. For a real-valued random variable W and some $\nu > 0$, we define $\|W\|_\nu := \mathbb{E}[|W|^\nu]^{1/\nu}$. Note that this notation should not be confused with a norm on function spaces; whenever we refer to the functional dependence measure, it should be clear that we apply a moments based norm. If ε_k^* is an independent copy of ε_k, independent of $\varepsilon_j, j \in \mathbb{Z}$, we define $\mathcal{G}_j^{*(j-k)} := (\varepsilon_j, ..., \varepsilon_{j-k+1}, \varepsilon_{j-k}^*, \varepsilon_{j-k-1}, ...)$ and $X_j^{*(j-k)} := J_{j,n}(\mathcal{G}_j^{*(j-k)})$. Then, the functional dependence measure of X_j is given by

$$\delta_\nu^X(k) = \left\| X_j - X_j^{*(j-k)} \right\|_\nu. \tag{5.1.3}$$

As before, references and properties already appeared in previous chapters.

5.2 Mellin transform

In this section we establish the key concept of Mellin transforms and collect its relevant properties for our theory. Let $\mathbb{L}^1(\Omega, \omega)$, $\mathbb{L}^2(\mathbb{R}_+, \omega)$ denote the space of either weighted absolutely integrable or weighted square-integrable functions on a space Ω or $\mathbb{R}_+$ (the positive real line), respectively. The spaces $\mathbb{L}^1(\mathbb{R})$ or $\mathbb{L}^2(\mathbb{R})$ consist of absolutely integrable or square-integrable functions on $\mathbb{R}$, respectively. For mathematically sound definitions we refer to Section 1.2.

Multiplicative Convolution In Chapter 1 we already mentioned that the density f_Y of Y_1 can be written as the multiplicative convolution of the densities f and g. We will now define this convolution in a more general setting. Let $c \in \mathbb{R}$. For two functions $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ we define the multiplicative convolution $h_1 * h_2$ of h_1 and h_2 by

$$(h_1 * h_2)(y) := \int_0^\infty h_1(y/x)h_2(x)x^{-1}dx, \quad y \in \mathbb{R}. \tag{5.2.1}$$

It is possible to show that the function $h_1 * h_2$ is well-defined, $h_1 * h_2 = h_2 * h_1$ and $h_1 * h_2 \in \mathbb{L}^1(\mathbb{R}, x^{c-1})$, compare Brenner Miguel [2021]. It is worth pointing out, that the definition of the multiplicative convolution in equation (5.2.1) is independent of the model parameter $c \in \mathbb{R}$. We also know that for densities h_1, h_2; $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^0)$. If additionally $h_1 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$, then $h_1 * h_2 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. We would like to mention that in this case the weight function is a polynomial, referred to by its own evaluation, that is, by abuse of notation $x^a : x \mapsto x^a$ for $a \in \mathbb{R}$. However, this is quite common in the literature.

Mellin transform properties We will now collect the main properties of the Mellin transform which will be used in the upcoming theory. Proof sketches of these properties can be found in Brenner Miguel [2021]. Let $h_1 \in \mathbb{L}^1(\mathbb{R}, x^{c-1})$. Then, we define the Mellin transform of h_1 at the development point $c \in \mathbb{R}$ as the function $\mathcal{M}_c[h] : \mathbb{R} \to \mathbb{C}$ with

$$\mathcal{M}_c[h_1](t) := \int_0^\infty x^{c-1+it}h_1(x)dx, \quad t \in \mathbb{R}. \tag{5.2.2}$$

A very important result with regards to the Mellin transform, which makes it so appealing for the use in multiplicative deconvolution, is the so-called convolution theorem: For $h_1, h_2 \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$,

$$\mathcal{M}_c[h_1 * h_2](t) = \mathcal{M}_c[h_1](t)\mathcal{M}_c[h_2](t), \quad t \in \mathbb{R}. \tag{5.2.3}$$

Additionally, for the estimation of the survival function the following property is used. Let $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ be a density and $S_h := \mathbb{R}_+ \to \mathbb{R}_+$, $y \mapsto \int_y^\infty h(x)dx$ its corresponding survival function. Then for any $c > 0$, $S_h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1})$ if and only if $h \in \mathbb{L}^1(\mathbb{R}_+, x^c)$. Furthermore, for any $t \in \mathbb{R}$,

$$\mathcal{M}_c[S_h](t) = (c + it)^{-1}\mathcal{M}_{c+1}[h](t).$$

Let us now define the Mellin transform of a square-integrable function, that is, for $h_1 \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$. Let $\varphi : \mathbb{R} \to \mathbb{R}_+$, $x \mapsto \exp(-2\pi x)$ and $\varphi^{-1} : \mathbb{R}_+ \to \mathbb{R}$ be its inverse. Then, as diffeomorphisms, φ, φ^{-1} map Lebesgue null sets to Lebesgue null sets. Thus, the isomorphism $\Phi_c : \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \to \mathbb{L}^2(\mathbb{R})$, $h \mapsto \varphi^c \cdot (h \circ \varphi)$ is well-defined. Moreover, let $\Phi_c^{-1} : \mathbb{L}^2(\mathbb{R}) \to \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ denote the inverse of Φ_c. Then, for $h \in \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ we define the Mellin transform of h developed in $c \in \mathbb{R}$ by

$$\mathcal{M}_c[h](t) := (2\pi)\mathcal{F}[\Phi_c[h]](t), \quad t \in \mathbb{R},$$

where $\mathcal{F} : \mathbb{L}^2(\mathbb{R}) \to \mathbb{L}^2(\mathbb{R})$, $H \mapsto (t \mapsto \mathcal{F}[H](t) := \lim_{k\to\infty} \int_{-k}^{k} \exp(-2\pi itx)H(x)dt)$ is the Fourier-Plancherel transform. Due to this definition several properties of Mellin transforms can be deduced from the well-known theory of Fourier transforms. In the case that $h \in \mathbb{L}^1(\mathbb{R}_+, x^{c-1}) \cap \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ we have

$$\mathcal{M}_c[h](t) = \int_0^\infty x^{c-1+it}h(x)dx, \quad t \in \mathbb{R}, \tag{5.2.4}$$

which coincides with the usual notion of Mellin transforms as considered in Paris and Kaminski [2001].

Now, due to the construction of the operator $\mathcal{M}_c : \mathbb{L}^2(\mathbb{R}_+, x^{2c-1}) \to \mathbb{L}^2(\mathbb{R})$ it can easily be seen that it is an isomorphism. We denote by $\mathcal{M}_c^{-1} : \mathbb{L}^2(\mathbb{R}) \to \mathbb{L}^2(\mathbb{R}_+, x^{2c-1})$ its inverse. If additionally to $H \in \mathbb{L}^2(\mathbb{R})$, $H \in \mathbb{L}^1(\mathbb{R})$, we can express the inverse Mellin transform explicitly by

$$\mathcal{M}_c^{-1}[H](x) = \frac{1}{2\pi}\int_{-\infty}^\infty x^{-c-it}H(t)dt, \quad x \in \mathbb{R}_+. \tag{5.2.5}$$

Furthermore, we can directly show that a Plancherel-type identity holds true for the Mellin transform: For all $h_1, h_2 \in \mathbb{L}(\mathbb{R}_+, x^{2c-1})$,

$$\langle h_1, h_2\rangle_{x^{2c-1}} = (2\pi)^{-1}\langle \mathcal{M}_c[h_1], \mathcal{M}_c[h_2]\rangle_{\mathbb{R}} \quad \text{whence} \quad \|h_1\|^2_{x^{2c-1}} = (2\pi)^{-1}\|\mathcal{M}_c[h]\|^2_{\mathbb{R}}. \tag{5.2.6}$$

Examples 5.2.1. Given below are the Mellin transforms of commonly used distribution families.

(i) *Beta distribution*: Let us consider the family $(g_b)_{b\in\mathbb{N}}$, $g_b(x) := \mathbb{1}_{(0,1)}(x)b(1-x)^{b-1}$ for $b \in \mathbb{N}$ and $x \in \mathbb{R}_+$. Obviously, we see that $\mathcal{M}_c[g_b]$ is well-defined for $c > 0$ and

$$\mathcal{M}_c[g_b](t) = \prod_{j=1}^{b} \frac{j}{c-1+j+it}, \quad t \in \mathbb{R}.$$

(ii) *Scaled* log-*gamma distribution*: Consider the family $(g_{\mu,a,\lambda})_{(\mu,a,\lambda)\in\mathbb{R}\times\mathbb{R}_+\times\mathbb{R}_+}$ where for $a, \lambda, x \in \mathbb{R}_+$ and $\mu \in \mathbb{R}$ we have $g_{\mu,a,\lambda}(x) = \frac{\exp(\lambda\mu)}{\Gamma(a)}x^{-\lambda-1}(\log(x)-\mu)^{a-1}\mathbb{1}_{(e^\mu,\infty)}(x)$. Then for $c < \lambda + 1$,

$$\mathcal{M}_c[g_{\mu,a,\lambda}](t) = \exp(\mu(c-1+it))(\lambda - c + 1 - it)^{-a}, \quad t \in \mathbb{R}.$$

If $a = 1$ then $g_{\mu,1,\lambda}$ is the density of a Pareto distribution with parameter e^μ and λ. If $\mu = 0$ we have that $g_{0,a,\lambda}$ is the density of a log-gamma distribution.

(iii) *Gamma distribution*: Consider the family $(g_d)_{d\in\mathbb{R}_+}, g_d(x) = \frac{x^{d-1}}{\Gamma(d)}\exp(-x)\mathbb{1}_{\mathbb{R}^+}(x)$ for $d, x \in \mathbb{R}_+$. Obviously, we see that $\mathcal{M}_c[g_d]$ is well-defined for $c > -d+1$ and

$$\mathcal{M}_c[g_d](t) = \frac{\Gamma(c+d-1+it)}{\Gamma(d)}, \quad t \in \mathbb{R}.$$

(iv) *Weibull distribution:* Consider the family $(g_m)_{m\in\mathbb{R}_+}$, $g_m(x) = mx^{m-1}\exp(-x^m)\mathbb{1}_{\mathbb{R}_+}(x)$ for $m, x \in \mathbb{R}_+$. Obviously, we see that $\mathcal{M}_c[g_m]$ is well-defined for $c > -m+1$ and

$$\mathcal{M}_c[g_m](t) = \frac{(c-1+it)}{m}\Gamma(\frac{c-1+it}{m}), \quad t \in \mathbb{R}.$$

(v) log-*normal distribution:* Consider the family $(g_{\mu,\lambda})_{(\mu,\lambda)\in\mathbb{R}\times\mathbb{R}_+}$ where $g_{\mu,\lambda}$ for $\lambda, x \in \mathbb{R}_+$ and $\mu \in \mathbb{R}$ is given by $g_{\mu,\lambda}(x) = \frac{1}{\sqrt{2\pi}\lambda x}\exp(-(\log(x)-\mu)^2/2\lambda^2)\mathbb{1}_{\mathbb{R}_+}(x)$. We see that $\mathcal{M}_c[g_{\mu,\lambda}]$ is well-defined for any $c \in \mathbb{R}$ and

$$\mathcal{M}_c[g_{\mu,\lambda}](t) = \exp(\mu(c-1+it))\exp\left(\frac{\lambda^2(c-1+it)^2}{2}\right), \quad t \in \mathbb{R}.$$

5.3 Minimax theory

In the upcoming theory, we need to ensure that the survival function S of the sample $X_1, \dots, X_n$ is square-integrable. Furthermore, in order to define the estimator, we also need the square-integrablility of the empirical survival function $\widehat{S}_X$ which is defined by

$$\widehat{S}_X(x) := n^{-1}\sum_{j=1}^{n}\mathbb{1}_{(0,X_i)}(x) \tag{5.3.1}$$

for any $x \in \mathbb{R}_+$. The following proposition shows that we can derive the square-integrablilty condition for both functions by moment conditions.

Proposition 5.3.1. *Let $\mathbb{E}(X^{1/2}) < \infty$. Then, $S \in \mathbb{L}^1(\mathbb{R}_+, x^{-1/2}) \cap \mathbb{L}^2(\mathbb{R}, x^0)$. If additionally $\mathbb{E}(X_1) < \infty$ then $\widehat{S}_X \in \mathbb{L}^1(\mathbb{R}_+, x^{-1/2}) \cap \mathbb{L}^2(\mathbb{R}, x^0)$ almost surely.*

The proof of Proposition 5.3.1 can be found in Section 5.7.2.

We now define an estimator of S based on the contaminated data $Y_1, \dots, Y_n$ and use the rich theory of Mellin transforms in Section 5.2.

Estimation strategy Let $c = 1/2$. So, the weighted $\mathbb{L}^2(\mathbb{R}, x^{2c-1})$-norm becomes the usual unweighted $\mathbb{L}^2$-norm. Assuming now that $\mathbb{E}(X_1^{1/2}) < \infty$ we have for $k \in \mathbb{R}_+$, $\mathcal{M}_{1/2}[S]\mathbb{1}_{[-k,k]} \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Thus

$$S_k(x) := \mathcal{M}_{1/2}^{-1}[\mathcal{M}_{1/2}[S]\mathbb{1}_{[-k,k]}](x) = \frac{1}{2\pi}\int_{-k}^{k} x^{-1/2-it}\mathcal{M}_{1/2}[S](t)dt, \quad x \in \mathbb{R}_+, \tag{5.3.2}$$

is an approximation of S in the $\mathbb{L}^2(\mathbb{R}_+, x^0)$ sense, that is, $\|S_k - S\|^2 \to 0$ for $k \to \infty$. Now, applying the property of the Mellin transform for survival functions, we know that $\mathcal{M}_{1/2}[S](t) = (1/2+it)^{-1}\mathcal{M}_{3/2}[f](t)$ for all $t \in \mathbb{R}$. Assuming that $\mathbb{E}(U^{1/2}) < \infty$, and thus $\mathbb{E}(Y^{1/2}) < \infty$, we get $f_Y, g \in \mathbb{L}^1(\mathbb{R}, x^{1/2})$. We can deduce from the convolution theorem that $\mathcal{M}_{3/2}[f_Y] = \mathcal{M}_{3/2}[f]\mathcal{M}_{3/2}[g]$. Under the mild assumption that for all $t \in \mathbb{R}$, $\mathcal{M}_{3/2}[g](t) \neq 0$, we can rewrite equation (5.3.2) as

$$S_k(x) = \frac{1}{2\pi} \int_{-k}^{k} x^{-1/2-it} \frac{\mathcal{M}_{3/2}[f_Y](t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)} dt. \tag{5.3.3}$$

To derive an estimator from equation (5.3.3) we use the empirical Mellin transform $\widehat{\mathcal{M}}(t) := n^{-1} \sum_{j=1}^{n} Y_j^{1/2+it}$ as an unbiased estimator of $\mathcal{M}_{3/2}[f_Y](t)$ for all $t \in \mathbb{R}$. Keeping in mind that $|\widehat{\mathcal{M}}(t)| \leq |\widehat{\mathcal{M}}(0)| < \infty$ almost surely, it is sufficient to assume that $\int_{-k}^{k} |(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2} dt < \infty$ for all $k \in \mathbb{R}_+$ to ensure the well-definition of the spectral cut-off estimator

$$\widehat{S}_k(x) := \frac{1}{2\pi} \int_{-k}^{k} x^{-1/2-it} \frac{\widehat{\mathcal{M}}(t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)} dt, \quad k, x \in \mathbb{R}_+. \tag{5.3.4}$$

Up to now, we had two minor conditions on the Mellin transform of the error density g which we want to collect in the following assumption:

$$\forall\, t \in \mathbb{R} : \mathcal{M}_{3/2}[g](t) \neq 0 \quad \text{and} \quad \forall\, k \in \mathbb{R}_+ : \int_{-k}^{k} |(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-1} dt < \infty. \tag{[\textbf{G0}]}$$

The following proposition shows that the proposed estimator is consistent for a suitable choice of a cut-off parameter and under certain assumptions on the dependence structure of $X_1, \ldots, X_n$. It is worth stressing out that for $t \in \mathbb{R}$ the estimator $(1/2 + it)^{-1}\widehat{\mathcal{M}}_X(t) := (1/2+it)^{-1} n^{-1} \sum_{j=1}^{n} X_j^{1/2+it}$ is an unbiased estimator of $\mathcal{M}_{1/2}[S](t)$. Furthermore, there is a special link between the empirical survival function and the estimator $(1/2+it)^{-1/2}\widehat{\mathcal{M}}(t)$. In fact, (5.3.1) ensures that almost surely $\widehat{S}_X \in \mathbb{L}^1(\mathbb{R}_0, x^{-1/2})$ which almost surely implies the existence of the Mellin transform of $\widehat{S}_X$. From that, it can easily be shown that

$$\mathcal{M}_{1/2}[\widehat{S}_X](t) = (1/2+it)^{-1}\widehat{\mathcal{M}}_X(t) \tag{5.3.5}$$

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

Theorem 5.3.2. *Assume that $\mathbb{E}(Y) < \infty$ and that **[G0]** holds true. Then for any $k \in \mathbb{R}_+$,*

$$\mathbb{E}(\|\widehat{S}_k - S\|^2) \leq \|S - S_k\|^2 + \mathbb{E}(Y_1)\frac{\Delta_g(k)}{n} + \frac{1}{2\pi}\int_{-k}^{k} \operatorname{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) dt,$$

where $\Delta_g(k) = (2\pi)^{-1} \int_{-k}^{k} |(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2} dt$.
If $(k_n)_{n\in\mathbb{N}}$ is chosen such that $k_n \to \infty$ for $n \to \infty$, $\frac{1}{2\pi}\int_{-k_n}^{k_n} \operatorname{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) dt \to 0$ and $\Delta_g(k_n) n^{-1} \to 0$, the consistency of $\widehat{S}_{k_n}$, that is,

$$\mathbb{E}(\|\widehat{S}_{k_n} - S\|^2) \to 0, \qquad n \to \infty$$

is obtained.

Corollary 5.3.3. *Under the assumptions of Theorem 5.3.2 we have*

(I) *for independent observations* $X_1, \dots, X_n$*,*

$$\mathbb{E}(\|\widehat{S}_k - S\|^2) \leq \|S - S_k\|^2 + \mathbb{E}(Y_1)\frac{\Delta_g(k)}{n} + \frac{\mathbb{E}(X_1)}{n};$$

(B) *for* β*-mixing observations* $X_1, \dots, X_n$ *under the assumption that* $\mathbb{E}(X_1 b(X_1)) < \infty$*,*

$$\mathbb{E}(\|\widehat{S}_k - S\|^2) \leq \|S - S_k\|^2 + \mathbb{E}(Y_1)\frac{\Delta_g(k)}{n} + \frac{c\mathbb{E}(X_1 b(X_1))}{n},$$

where $c > 0$ *is a positive numerical constant;*

(F) *and for Bernoulli shift processes (5.1.2) under the dependence measure (5.1.3) provided that* $\sum_{j=1}^{\infty} \delta_1^X(j)^{1/2} < \infty$*,*

$$\mathbb{E}(\|\widehat{S}_k - S\|^2) \leq \|S - S_k\|^2 + \mathbb{E}(Y_1)\frac{\Delta_g(k)}{n} + \frac{c\log(k)}{n}\Big(\sum_{j=1}^{\infty} \delta_1^X(j)^{1/2}\Big)^2$$

where $c > 0$ *is a numerical positive constant.*

As we can see, the first term, the so-called bias term, in the upper bound in Theorem 5.3.2 is monotonically decreasing in $k \in \mathbb{R}_+$ while the second and the last term are monotonically increasing in $k \in \mathbb{R}_+$. The second and the third term are a decomposition of $\mathbb{E}(\|\widehat{S}_k - S_k\|^2)$, the so-called variance term. While Corollary 5.3.3 indicates how the third term can be bounded, the general assumptions on the error densities do not allow us to determine the exact growth of the second term. For a more sophisticated analysis of the variance's growth, we need to consider more specific assumptions on the error density g. More precisely, the growth of Δ_g is determined by the decay of the Mellin transform of g. In this work, we mainly focus on the case of *smooth error densities*, that is, there exist $c, C, \gamma \in \mathbb{R}_+$ such that

$$c(1+t^2)^{-\gamma/2} \leq |\mathcal{M}_c[g](t)| \leq C(1+t^2)^{-\gamma/2}, \quad t \in \mathbb{R}. \qquad (\textbf{[G1]})$$

This assumption on the error density is typical in context of additive deconvolution problems (cf. Fan [1991]) and is also considered in the works of Belomestny and Goldenshluger [2020] and Brenner Miguel et al. [2020].

In the context of smooth error densities, we have the following with regards to **[G1]**.

Examples 5.3.4 (Examples 5.2.1, continued)**.** (i) *Beta distribution:* For $b \in \mathbb{N}$ and $t \in \mathbb{R}$ we have already seen that $\mathcal{M}_{3/2}[g_b](t) = \prod_{j=1}^{b} \frac{j}{1/2+j+it}$. Therefore, there exist $c_g, C_g > 0$ such that

$$c_g(1+t^2)^{-b/2} \leq |\mathcal{M}_{3/2}[g_b](t)| \leq C_g(1+t^2)^{-b/2}.$$

(ii) *Scaled* log-*gamma distribution:* For $\lambda > 1/2$, $a \in \mathbb{R}_+$ and $\mu, t \in \mathbb{R}$ we have already seen that $\mathcal{M}_{3/2}[g_{\mu,a,\lambda}](t) = \exp(\mu(1/2+it))(\lambda - 1/2 - it)^{-a}$. Therefore, there exist $c_g, C_g > 0$ such that

$$c_g(1+t^2)^{-a/2} \leq |\mathcal{M}_{3/2}[g_{\mu,a,\lambda}](t)| \leq C_g(1+t^2)^{-a/2}.$$

We would like to mention that for small values of γ in **[G1]** it is possible to choose k independently of the decay of the bias term, in a way that the risk in Theorem 5.3.2 is of order n^{-1}, respectively $\log(n)n^{-1}$. These cases are covered in the following paragraphs.

Parametric rate If $\gamma \leq 1/2$ in **[G1]**, the parameter $k \in \mathbb{R}_+$ can be chosen in a way that leads to a parametric rate up to a log-term. This choice can be done independently of the precise decay of the bias term $\|S - S_k\|^2$. We accomplish it by the naive bound

$$\|S - S_k\|^2 = \frac{1}{\pi}\int_k^\infty |\mathcal{M}_{1/2}[S](t)|^2 dt \leq k^{-1}\mathbb{E}(X_1^{1/2})^2$$

where we exploit $\mathcal{M}_{1/2}[S](t) = (1/2+it)^{-1}\mathcal{M}_{3/2}[f](t)$ and the bound $|\mathcal{M}_{3/2}[f](t)| \leq \mathbb{E}(X_1^{1/2})$. The different cases are collected in the following proposition whose proof is omitted.

Proposition 5.3.5. *Let $\mathbb{E}(Y) < \infty$ and **[G1]** hold true for $\gamma \leq 1/2$. Then,*

$$\Delta_g(k) \leq \begin{cases} C(g) & ,\gamma < 1/2; \\ C(g)\log(k) & ,\gamma = 1/2. \end{cases}$$

*For $\gamma = 1/2$, choosing $k = n$ leads in all three cases **(I)**,**(B)** and **(F)** to a parametric rate up to a* log*-term, that is,*

$$\mathbb{E}(\|\widehat{S}_n - S\|^2) \leq C(f,g)\frac{\log(n)}{n}$$

*where $C(f,g)$ is dependent on $\mathbb{E}(X_1)$, $\mathbb{E}(U_1)$, the constants in **[G1]** and the dependence structure.*
*If $\gamma < 1/2$ in the cases **(I)** and **(B)**, choosing $k = \infty$ leads to a parametric rate,*

$$\mathbb{E}(\|\widehat{S}_\infty - S\|^2) \leq \frac{C(f,g)}{n}$$

*where $C(f,g)$ depends on $\mathbb{E}(X_1)$, $\mathbb{E}(Y_1)$, the constants in **[G1]** and the dependence structure. If $\gamma < 1/2$ and we are in the case of **(F)**, the third summand in Corollary 5.3.3 dominates the second, which leads to no improvement in the rate.*

Nonparametric rate Now let us consider the case where $\gamma > 1/2$, which forces $\Delta_g(k)$ to be polynomial increasing. Under **[G1]** we see that $c_g k^{2\gamma-1} \leq \Delta_g(k) \leq C_g k^{2\gamma-1}$ for every $k \in \mathbb{R}_+$.

In order to control the bias-term we introduce regularity spaces characterized by the decay of the Mellin transform in analogy to the usually considered Sobolev spaces for common deconvolution problems. Let us for $s \in \mathbb{R}_+$ define the *Mellin-Sobolev space* by

$$\mathcal{W}^s_{1/2}(\mathbb{R}_+) := \{h \in \mathbb{L}^2(\mathbb{R}_+, x^0) : |h|_s^2 := \|(1+t)^s \mathcal{M}_{1/2}[h]\|^2_{\mathbb{R}} < \infty\} \tag{5.3.6}$$

and the corresponding ellipsoids with $L \in \mathbb{R}$ by $\mathcal{W}^s_{1/2}(L) := \{h \in \mathcal{W}^s_{1/2}(\mathbb{R}_+) : |h|_s^2 \leq L\}$. For $f \in \mathcal{W}^s_{1/2}(L)$ we deduce that $\|S - S_k\|^2 \leq Lk^{-2s}$. Setting

$$\begin{aligned}\mathbb{W}^s_{1/2}(L) := \{S \in \mathcal{W}^s_{1/2}(L) : S \text{ survival function},\\ \operatorname{Var}_S(\mathcal{M}_{1/2}[\widehat{S}_X](t)) \leq L(1+|t|)^{-1}n^{-1} \text{ for any } t \in \mathbb{R}\},\end{aligned}$$

the previous discussion leads to the following statement whose proof is omitted.

Proposition 5.3.6. *Let $\mathbb{E}(U) < \infty$. Then under the assumptions **[G0]** and **[G1]**,*

$$\sup_{S \in \mathbb{W}^s_{1/2}(L)} \mathbb{E}(\|S - \widehat{S}_{k_n}\|)^2 \leq C(L, g, s) n^{-2s/(2s+2\gamma-1)}$$

for the choice $k_n := n^{1/(2s+2\gamma-1)}$.

Again, let us consider the three different cases of dependence considered in Corollary 5.3.3. As a direct consequence of Proposition 5.3.6 and Corollary 5.3.3 we get the following corollary.

Corollary 5.3.7. *The assumption* $\operatorname{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) \leq L(1+|t|)^{-1}n^{-1}$ *for any $t \in \mathbb{R}$ in Proposition 5.3.6 can be replaced by*

(I) $\mathbb{E}(X_1) \leq L$

(B) $\mathbb{E}(X_1 b(X_1)) \leq L$

(F) $\sum_{k=1}^{\infty} \delta_1^X(k)^{1/2} \leq L^{1/2}$

in the three different dependence cases.

We will now show that the rates presented in Corollary 5.3.7 are optimal in the sense, that there exists no estimator based on the i.i.d. sample $Y_1, \dots, Y_n$ that can achieve uniformly over $\mathbb{W}^s_{1/2}(L)$ a better rate. This implies that the estimator $\widehat{S}_{k_n}$ presented in Proposition 5.3.6 is minimax-optimal.
For technical reason we need an additional assumption on the error density g. Let us assume that the support of g is bounded, that is, there exists a constant $d > 0$ such that for all $x \geq d$, $g(x) = 0$. For the sake of simplicity we assume that $d = 1$. Furthermore, let there be constants $c, C, \gamma \in \mathbb{R}_+$ such that

$$c(1+t^2)^{-\gamma/2} \leq |\mathcal{M}_{1/2}[g](t)| \leq C(1+t^2)^{-\gamma/2}, \quad t \in \mathbb{R}. \tag{[G2]}$$

With this additional assumption we can show the following theorem. Its proof can be found in Section 5.7.2.

Theorem 5.3.8. *Let $s, \gamma \in \mathbb{N}$ and assume that [G1] and [G2] hold true. Then there exist constants $c_g, L_{s,g} > 0$ such that for all $L \geq L_{s,g}$, $n \in \mathbb{N}$ and for any estimator $\widehat{S}$ of S based on an i.i.d. sample $Y_1, \dots, Y_n$,*

$$\sup_{S \in \mathbb{W}^s_{1/2}(L)} \mathbb{E}(\|\widehat{S} - S\|^2) \geq c_g n^{-2s/(2s+2\gamma-1)}.$$

For the multiplicative censoring model, that is, U is uniformly-distributed on $[0,1]$, the assumptions **[G1]** and **[G2]** hold true.
Nevertheless, the rate presented in Proposition 5.3.6 is very pessimistic, meaning that we can find examples of $f \in \mathbb{W}^s_{1/2}(L)$ where the bias decays faster than Lk^{-2s}. These examples are considered in the next section.

Faster rates Let us revisit the families (iii)-(v) in Example 5.2.1.

Examples 5.3.9 (Example 5.2.1, continued)**.** By application of the Stirling formula,

(i) the *gamma distribution* $f(x) = \frac{x^{d-1}}{\Gamma(d)} \exp(-x)\mathbb{1}_{\mathbb{R}_+}(x)$, $d, x \in \mathbb{R}_+$ delivers

$$|\mathcal{M}_{1/2}[S](t)| = |(1/2 + it)|^{-1} |\mathcal{M}_{3/2}[f](t)| \leq C_d |t|^{d-1} \exp(-\pi|t|/2), \text{ for } |t| \geq 2;$$

(ii) the *Weibull distribution* $f(x) = mx^{m-1} \exp(-x^m)\mathbb{1}_{\mathbb{R}_+}(x)$, $m, x \in \mathbb{R}_+$, delivers for $|t| \geq 2$,

$$|\mathcal{M}_{1/2}[S](t)| = |(1/2 + it)|^{-1} |\mathcal{M}_{3/2}[f](t)| \leq C_m |t|^{(1-m)/2m} \exp(-\pi|t|/(2m));$$

(iii) the log-*normal distribution* $f(x) = (2\pi\lambda^2 x^2)^{-1/2} \exp(-(\log(x) - \mu)/2\lambda^2)\mathbb{1}_{\mathbb{R}_+}(x)$, $\lambda, x \in \mathbb{R}_+$ and $\mu \in \mathbb{R}$ delivers

$$|\mathcal{M}_{1/2}[S](t)| = |(1/2 + it)|^{-1} |\mathcal{M}_{3/2}[f](t)| \leq C_{\mu,\lambda} |t|^{-1} \exp(-\lambda^2 t^2/2), \text{ for } |t| \geq 1.$$

In all three cases, we can bound the bias term by $\|S - S_k\| \leq C \exp(-\delta k^r)$ for some $\delta, r \in \mathbb{R}_+$ leading to a much sharper bound than Lk^{-2s}, although it is easy to verify that all three examples lie in $\mathbb{W}^s_{1/2}(L)$ for any $s \in \mathbb{R}_+$ and $L \in \mathbb{R}_+$ large enough. For example, in the case of **(I)**, the choice of $k = k_n = n^{1/(2s+2\gamma-1)}$, which was suggested in Proposition 5.3.6, can be improved for any choice of $s \in \mathbb{R}_+$. Setting $k_n = (\log(n)\delta^{-1})^{1/r}$ leads to

$$\mathbb{E}(\|\widehat{S}_{k_n} - S\|^2) \leq C(f, g) \frac{\log(n)^{(2\gamma-1)/r}}{n}$$

which results in a sharper rate then $n^{-2s/(2s+2\gamma-1)}$.
Furthermore, despite the fact that the choice of k_n in Proposition 5.3.6 does not depend on the explicit density $f \in \mathbb{W}^s_{1/2}(L)$, it is still dependent on the regularity parameter $s \in \mathbb{R}_+$ of the unknown density f. While it is tempting to set the regularity parameter $s \in \mathbb{R}_+$ to a fixed value and interpret this as an additional model assumption, the discussion above suggests that we might end up with worse rates. In the next section, we therefore present a data-driven method in order to choose the parameter $k = k_n \in \mathbb{R}_+$ based only on the sample $Y_1, \dots, Y_n$.

5.4 Data-driven method

We now present a data-driven method for selecting the above appearing tuning parameter $k \in \mathbb{R}_+$ based on a penalized contrast approach. We only consider the case where **[G1]** holds true for $\gamma > 1/2$. For $\gamma \leq 1/2$, we already have discussed a choice of the parameter $k \in \mathbb{R}_+$, that is independent of the density f and achieves an almost parametric rate. In the case $\gamma > 1/2$, the second summand in Theorem 5.3.2 dominates the third term. Thus, the growth of the variance term is determined by the growth of Δ_g. Our aim now is to define an estimator $\widehat{k}$ which mimics the behavior of

$$k :\in \arg\min\{\|S - S_k\|^2 + \mathbb{E}(Y)C_g(2\pi n)^{-1}k^{2\gamma-1} : k \in \mathcal{K}_n\}$$

for a suitable large set of parameters $\mathcal{K}_n \subset \mathbb{R}_+$. Considering the result of Proposition 5.3.6 and the fact that $\|S - S_k\|^2 \leq k^{-1}\mathbb{E}(X^{1/2})^2$, which we have seen in the paragraph about the parametric case, we can ensure that the set $\mathcal{K}_n := \{k \in \{1, \dots, n\} : \Delta_g(k) \leq n^{-1}\}$ is suitably large enough. Starting with the bias term we see that $\|S - S_k\|^2 = \|S\|^2 - \|S_k\|^2$ behaves like $-\|S_k\|^2$. Furthermore, for $k \in \mathcal{K}_n$ we define the penalty term $\mathrm{pen}(k) = \chi\sigma_Y\Delta_g(k)n^{-1}$, $\sigma_Y := \mathbb{E}(Y_1)$, which imitates the behavior of the variance term. Exchanging $-\|S_k\|^2$ and $\mathbb{E}(Y_1)$ with their empirical counterparts $-\|\widehat{S}_k\|^2$ and $\widehat{\sigma}_Y := n^{-1}\sum_{j=1}^n Y_j$ we define a fully data-driven model selection $\widehat{k}$ by

$$\widehat{k} \in \arg\min\{-\|\widehat{S}_k\|^2 + \widehat{\mathrm{pen}}(k) : k \in \mathcal{K}_n\} \quad \text{where} \quad \widehat{\mathrm{pen}}(k) := 2\chi\widehat{\sigma}_Y\Delta_g(k)n^{-1} \tag{5.4.1}$$

for $\chi > 0$. The following theorem shows that this procedure is adaptive up to a negligible term.

Theorem 5.4.1. *Let g satisfy* ***[G1]*** *with $\gamma > 1/2$ and $\|xg\|_\infty < \infty$. Assume further that $\mathbb{E}(Y_1^{5/2}) < \infty$. Then for $\chi > 96$,*

$$\mathbb{E}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6 \inf_{k \in \mathcal{K}_n} \left(\|S - S_k\|^2 + \mathrm{pen}(k)\right) + C(g,f)\left(n^{-1} + \mathrm{Var}(\widehat{\sigma}_X) + \int_{-n}^{n} \mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t))dt\right)$$

where $C(g,f) > 0$ is a constant depending on χ, the error density g, $\mathbb{E}(X_1^{5/2})$, $\sigma_X := \mathbb{E}(X_1)$ and $\widehat{\sigma}_X := n^{-1}\sum_{j=1}^n X_j$.

The proof of Theorem 5.4.1 is postponed to Section 5.7.3. The assumption $\|xg\|_\infty < \infty$ is rather weak because it is satisfied for a large range of densities considered. Note that by abuse of notation $x : x \mapsto x$ is the identity mapping.

Corollary 5.4.2. *Let the assumptions of Theorem 5.4.1 hold true. Then,*

(I) *in the presence of i.i.d. observations $X_1, ..., X_n$ and $\mathbb{E}[X_1^2] < \infty$,*

$$\mathbb{E}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6 \inf_{k \in \mathcal{K}_n} \left(\|S - S_k\|^2 + \mathrm{pen}(k)\right) + \frac{C(g,f)}{n}\left(1 + \mathbb{E}[X_1^2] + \mathbb{E}|X_1|\right);$$

(B) *for β-mixing observations $X_1, ..., X_n$ under the assumption that $\mathbb{E}[X_1^2 b(X_1)] < \infty$,*

$$\mathbb{E}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6 \inf_{k \in \mathcal{K}_n} \left(\|S - S_k\|^2 + \text{pen}(k)\right) + \frac{C(g,f)}{n}\left(1 + \mathbb{E}[X_1^2 b(X_1)] + \mathbb{E}[|X_1| b(X_1)]\right);$$

(F) *for Bernoulli-shift processes (5.1.2) under the dependence measure (5.1.3) provided that $\sum_{j=1}^n \delta_1^X(j)^{1/2} < \infty$ and $\sum_{j=1}^n \delta_2^X(j) < \infty$,*

$$\mathbb{E}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6 \inf_{k \in \mathcal{K}_n} \left(\|S - S_k\|^2 + \text{pen}(k)\right) + \frac{C(g,f)\log(n)}{n}\left(1 + \Big(\sum_{j=1}^n \delta_2^X(j)\Big)^2 + \Big(\sum_{j=1}^n \delta_1^X(j)^{1/2}\Big)^2\right).$$

Under the assumptions of Theorem 5.4.1 and if $S \in \mathbb{W}^s_{1/2}(L)$, we can ensure that $s > 1/2$. Then we have $k_n := \lfloor n^{1/(2s+2\gamma-1} \rfloor \in \mathcal{K}_n$ and thus

$$\mathbb{E}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq C(f,g) n^{-2s/(2s+2\gamma-1)}$$

for all three cases **(I)**, **(B)** and **(F)**.

5.5 Numerical studies

In this section, we use Monte-Carlo simulations to visualize the properties of the estimator $\widehat{S}_{\widehat{k}}$. We will first consider the case of independent observations **[I]** and then study the behavior of the estimator in presence of dependence.

5.5.1 Independent data

Let us illustrate the performance of the fully-data driven estimator $\widehat{S}_{\widehat{k}}$ defined in (5.3.4) and (5.4.1). To do so, we consider the following densities whose corresponding survival function will be estimated.

(i) *Gamma distribution:* $f_1(x) = \frac{0.5^4}{\Gamma(4)} x^3 \exp(-0.5x) \mathbb{1}_{\mathbb{R}_+}(x)$,

(ii) *Weibull distribution:* $f_2(x) = 2x \exp(-x^2) \mathbb{1}_{\mathbb{R}_+}(x)$,

(iii) *Beta distribution:* $f_3(x) = \frac{1}{560}(0.5x)^3(1 - 0.5x)^4 \mathbb{1}_{(0,2)}(x)$ and

(iv) log-*gamma distribution:* $f_4(x) = \frac{x^{-4}}{6} \log(x)^3 \mathbb{1}_{(1,\infty)}(x)$.

As for the distribution of the error density, we consider the following three cases

(a) *Uniform distribution:* $g_1(x) = \mathbb{1}_{[0,1]}(x)$,

(b) *Symmetric noise:* $g_2(x) = \mathbb{1}_{(1/2,3/2)}(x)$ and

(c) *Beta distribution:* $g_3(x) = 2(1-x)\mathbb{1}_{(0,1)}(x)$.

We see that g_1 and g_2 satisfy **[G1]** with the parameter $\gamma = 1$ and g_3 fulfills the conditions with $\gamma = 2$. Due to the fact that the true survival function satisfies $S(x) \in [0,1]$, $x \in \mathbb{R}$, we can improve the estimator $\widehat{S}_{\widehat{k}}$ by defining

$$\widetilde{S}_{\widehat{k}}(x) := \begin{cases} 0 & , \widehat{S}_{\widehat{k}}(x) \leq 0; \\ \widehat{S}_{\widehat{k}}(x) & , \widehat{S}_{\widehat{k}}(x) \in [0,1]; \\ 1 & , \widehat{S}_{\widehat{k}}(x) \geq 1. \end{cases}$$

The resulting estimator $\widetilde{S}_{\widehat{k}}$ has a smaller risk then $\widehat{S}_{\widehat{k}}$, since $\|\widetilde{S}_{\widehat{k}} - S\|^2 \leq \|\widehat{S}_{\widehat{k}} - S\|$. On the other hand, the estimator has the desired property that it is $[0,1]$-valued. Nevertheless, it is difficult ensure that $\widetilde{S}_{\widehat{k}}(0) = 1$ and that $\widetilde{S}_{\widehat{k}}$ is monotone decreasing. Although there are many procedures to guarantee the monotonicity of an estimator $\widehat{S}$ and the property $\widehat{S}(0) = 1$, the presented theoretical results of this work are not applicable to the modified estimators.
We will now use a Monte-Carlo simulation to visualise the properties of the estimator $\widetilde{S}_{\widehat{k}}$ and discuss whether the numerical simulated behavior of the estimator coincides with the theoretical predictions.
After that we will construct a survival function estimator $\widehat{S}$ based on $\widehat{S}_{\widehat{k}}$ which, in fact, is a survival function, keeping in mind that the theoretical results of this work do not apply for this estimator.

	$n = 500$	$n = 1000$	$n = 2000$
(i)	1.31	0.75	0.23
(ii)	0.19	0.09	0.04
(iii)	0.21	0.12	0.06
(iv)	1.10	0.34	0.20

Table 5.1: The entries showcase the MISE (scaled by a factor of 100) obtained by Monte-Carlo simulations each with 200 iterations. We take a look at different densities and varying sample sizes. The error density is chosen as (a) in each case.

In Table 5.1 we can see that for an increasing sample size, the variance of the estimator seems to decrease. Also, increasing the sample size allows for the estimator to choose bigger $\widehat{k}$ values, which on the other hand decreases the bias induced by the approximation step. Next, let us consider different error densities.

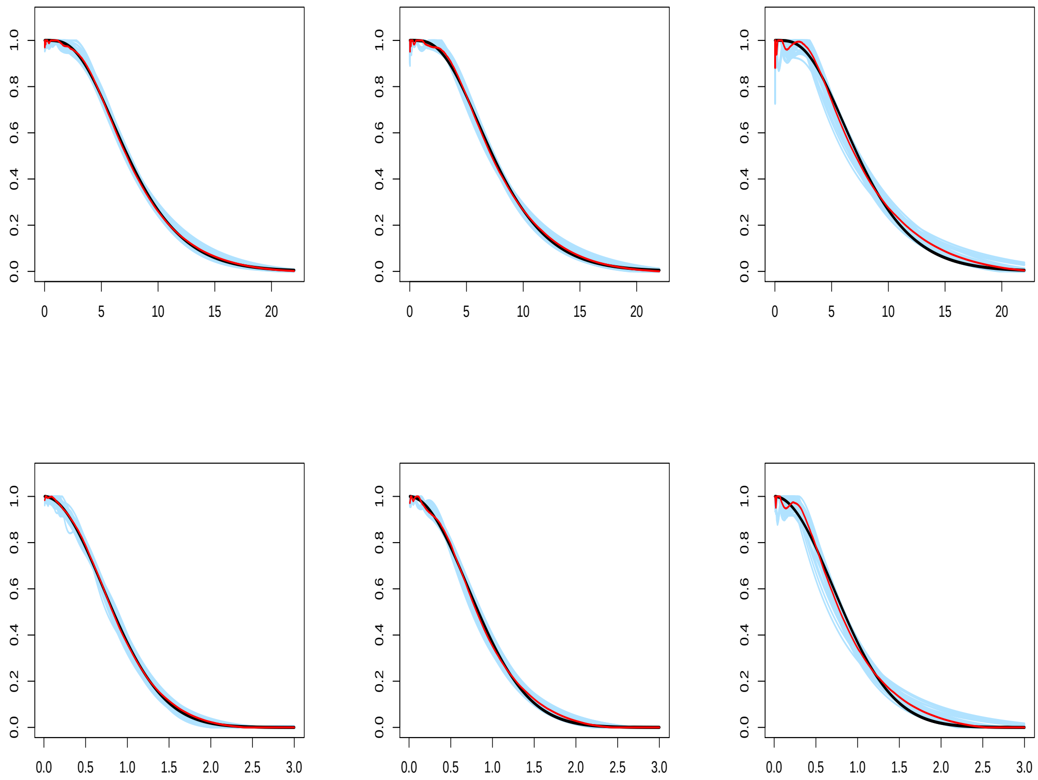

Figure 5.1: Considering the estimators $\widetilde{S}_{\widehat{k}}$, we depict 50 Monte-Carlo simulations with varying error density (a) (left), (b) (middle) and (c) (right) for (i) (top), (ii) (bottom) with $n = 1000$. The true survival function S is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

In Figure 5.1 we see that the reconstruction of the survival function with error density (a) and (b) is of the same complexity while the reconstruction with error density (c) seems to be more difficult. This behavior is predicted by the theoretical results because for **[G1]**, (a) and (b) share the same parameter $\gamma = 1$ while (c) has the parameter $\gamma = 2$.

Heuristic estimator We will now modify the estimator $\widehat{S}_{\widehat{k}}$ such that the resulting estimator $\widehat{S}$ is a survival function. To do so, we see that for any $k \in \mathbb{R}$ and $x \in \mathbb{R}_+$,

$$\widehat{S}_k(x) = \frac{1}{2\pi}\int_{-k}^{k} x^{-1/2-it}\frac{\widehat{\mathcal{M}}(t)}{\mathcal{M}_{3/2}[g](t)}(1/2+it)^{-1}dt = (\widehat{p}_k * g_u)(x),$$

where $g_u(x) = \mathbb{1}_{(0,1)}(x)$ is the density of the uniform distribution on $(0,1)$ and $\widehat{p}_k(x) = (2\pi)^{-1}\int_{-k}^{k} x^{-1/2-it}\frac{\widehat{\mathcal{M}}(t)}{\mathcal{M}_{3/2}[g](t)}dt$. Exploiting the definition of the multiplicative convolution we see that for any $x \in \mathbb{R}_+$,

$$\widehat{S}_k(x) = \int_x^\infty \widehat{p}_k(y)y^{-1}dy.$$

This motivates the following construction of a survival function estimator. First, exchanging $\widehat{p}_k$ with $(\widehat{p}_k(x))_+$ ensures the monotonicity and the positivity of our estimator. The final estimator is then defined as

$$\widehat{S}(x) := \widetilde{S}(x)/\widetilde{S}(0+), \quad \text{where } \widetilde{S}(x) := \int_x^\infty (\widehat{p}_{\widehat{k}}(y))_+ y^{-1}dy, \text{ for any } y \in \mathbb{R}_+.$$

where $0+$ denotes a positive real number very close to 0. Since our estimator is not defined in 0 we cannot normalise it with $\widetilde{S}(0)$.

Let us now illustrate the behavior of the heuristic estimator $\widehat{S}$ for an increasing number of observations compared to the estimator $\widetilde{S}_{\widehat{k}}$.

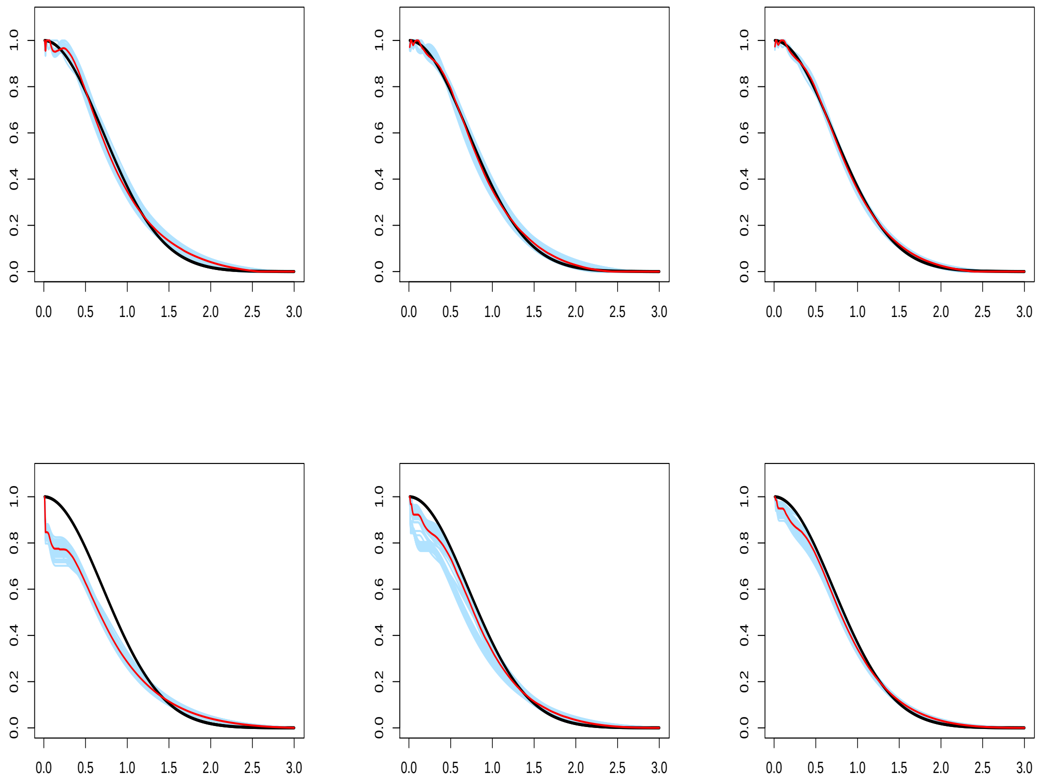

Figure 5.2: Considering the estimators $\widetilde{S}_{\widehat{k}}$ (top) and $\widehat{S}$ (bottom), we depict 50 Monte-Carlo simulations with varying sample sizes $n = 500$ (left), $n = 1000$ (middle) and $n = 2000$ (right) in the case (ii) with error density (b). The true survival function S is given by the black curve while the red curve is the point-wise empirical median of the 50 estimates.

Although the estimator $\widehat{S}$ is a survival function, the alteration of the estimator seems to introduce an additional bias. Based on the numerical study, this additional bias seems to decrease for n large enough. Nevertheless, this modification is purely heuristic.

Dependent data In this subsection we are going to take a look at an AR(1)-process of the following form and analyze its structure using the functional dependence measure. For $|\rho| < 1$ we define

$$X_n := \rho X_{n-1} + \varepsilon_n, \quad \text{where } \varepsilon_n \mid B_n \sim \Gamma_{(B_n, \lambda)} \text{ with } B_n \sim \mathrm{Bin}(m, 1 - |\rho|).$$

Here, we obviously refer to the gamma distribution by $\Gamma(\alpha, \lambda)$, $\alpha, \lambda > 0$, and the binomial distribution by $\mathrm{Bin}(m, p)$, $m \in \mathbb{N} \cup \{0\}$, $\rho \in (0, 1)$. It is apparent that the innovations constructed are i.i.d. and for this specific choice of ε_n, the marginals of the AR(1)-process follow a gamma distribution $\Gamma_{(m,\lambda)}$ as shown in Gaver and Lewis [1980]. The sequence $(X_n)_{n\in\mathbb{N}}$ then emits a representation as a Bernoulli shift, $X_n = \sum_{j=0}^{\infty} \rho^j \varepsilon_{n-j}$ (cf. Brockwell and Davis [2016] orChow and Teicher [1988]). In this case,

$$\delta_p^X(k) = |\rho^k| \sup_{j=1,\dots,n} \|\varepsilon_j - \varepsilon_j^{*(j-k)}\|_p \leq 2|\rho^k|\|\varepsilon_1\|_p.$$

Therefore, $\sum_{k=0}^{\infty} \delta_2^X(k) = \frac{1}{1-|\rho|} < \infty$ and $\sum_{k=0}^{\infty} \delta_1^X(k)^{1/2} = \frac{1}{1-|\rho|^{1/2}} < \infty$ as geometric series. Hence, the assumptions of Corollary 5.3.3 and 5.4.2 are satisfied. Specifically, the estimator's variance depends on the choice of ρ. The closer ρ is to one, the more dependent the data becomes whereas the variance increases. However, if ρ approaches zero, the data can be seen more independent, until it becomes i.i.d. This behavior can be seen in Table 5.2, as well.

		$n = 500$	$n = 1000$	$n = 2000$
$m = 1$	$\rho = 0.1$	0.15	0.11	0.05
	$\rho = 0.5$	0.29	0.16	0.07
	$\rho = 0.9$	1.19	0.63	0.35
$m = 4$	$\rho = 0.1$	0.69	0.39	0.13
	$\rho = 0.5$	0.93	0.52	0.18
	$\rho = 0.9$	2.87	1.57	0.70

Table 5.2: The entries showcase the MISE (scaled by a factor of 100) obtained by Monte-Carlo simulations each with 200 iterations. We take a look at different densities, three distinct sample sizes and varying ρ. The error density is chosen as (a) in each case.

5.6 Concluding remarks

We have developed a theory for the estimation of the survival function with multiplicative measurement error based not only on i.i.d. but also on dependent observations. As an inverse problem, employing the Mellin transform turned out to be a successful strategy. If the data is i.i.d. or β-mixing we obtained oracle rates of a similar magnitude. In case of Bernoulli shift processes under the functional dependence measure a slightly slower rate by an additional factor of $\log(n)$ was achieved. Under certain conditions almost parametric rates were recovered. The bias-variance trade-off was typically dealt with by a spectral cut-off for which a cut-off parameter was selected by a data-driven method. So far, we have established a minimax rate when our observations are independent and identically distributed; in the presence of dependence a completely new theory has yet to be provided. In a numerical study we have seen that the simulations support our theory, which is why would also expect a good performance in real life applications.

5.7 Lemmata and Proofs of Chapter 5

5.7.1 Useful inequalities

The following inequality is due to Talagrand [1996], the formulation of the first part can be found for example in Klein and Rio [2005].

Lemma 5.7.1. *(Talagrand's inequality) Let $Z_1, \dots, Z_n$ be independent $\mathcal{Z}$-valued random variables and let $\bar{\nu}_h = n^{-1} \sum_{i=1}^n [\nu_h(Z_i) - \mathbb{E}(\nu_h(Z_i))]$ for ν_h belonging to a countable class $\{\nu_h, h \in \mathcal{H}\}$ of measurable functions. Then,*

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}} |\bar{\nu}_h|^2 - 6\Psi^2\right)_+ \leq C\left[\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) + \frac{\psi^2}{n^2}\exp\left(\frac{-Kn\Psi}{\psi}\right)\right] \tag{5.7.1}$$

with numerical constants $K = (\sqrt{2}-1)/(21\sqrt{2})$ and $C > 0$ where

$$\sup_{h\in\mathcal{H}}\sup_{z\in\mathcal{Z}} |\nu_h(z)| \leq \psi, \qquad \mathbb{E}(\sup_{h\in\mathcal{H}} |\overline{\nu}_h|) \leq \Psi, \qquad \sup_{h\in\mathcal{H}} \frac{1}{n}\sum_{j=1}^n \mathrm{Var}(\nu_h(Z_i)) \leq \tau.$$

Remark 5.7.2. Keeping the bound (5.7.1) in mind, we can specify particular choices of K, e.g. $K \geq \frac{1}{100}$. As an immediate consequence we have

$$\mathbb{E}\left(\sup_{h\in\mathcal{H}} |\bar{\nu}_h|^2 - 6\Psi^2\right)_+ \leq C\left(\frac{\tau}{n}\exp\left(\frac{-n\Psi^2}{6\tau}\right) + \frac{\psi^2}{n^2}\exp\left(\frac{-n\Psi}{100\psi}\right)\right). \tag{5.7.2}$$

5.7.2 Proofs of Section 5.3

Proof of Lemma 5.3.1. Let us assume that $\mathbb{E}(X^{1/2}) < \infty$. Then, $f \in \mathbb{L}^1(\mathbb{R}, x^{1/2})$ and thus $S \in \mathbb{L}^1(\mathbb{R}, x^{-1/2})$ because

$$\int_0^\infty x^{-1/2} S(x)dx = \int_0^\infty \int_0^\infty x^{-1/2}\mathbb{1}_{(0,x)}(y)f(y)dydx = \int_0^\infty 2y^{1/2}f(y)dy < \infty.$$

By the generalized Minkowski inequality, cf. Tsybakov [2008], we have

$$\begin{aligned}\int_0^\infty S^2(x)dx &= \int_0^\infty \left(\int_0^\infty \mathbb{1}_{(0,x)}(y)f(x)dx\right)^2 dy \\ &\leq \left(\int_0^\infty \left(\int_0^\infty \mathbb{1}_{(0,x)}(y)f^2(x)dy\right)^{1/2} dx\right)^2 = \left(\int_0^\infty x^{1/2}f(x)dx\right)^2,\end{aligned}$$

which implies $S \in \mathbb{L}^2(\mathbb{R}_+, x^0)$. Next, we can easily see that $\mathbb{E}(\|\widehat{S}_X\|_{\mathbb{L}^1(\mathbb{R}_+,x^{-1/2})}) = \int_0^\infty x^{-1/2}S_X(x)dx < \infty$. Since $\mathbb{E}(X_1) < \infty$, we obtain

$$\mathbb{E}(\|\widehat{S}_X\|^2) \leq \int_0^\infty \mathbb{E}(\mathbb{1}_{(x,\infty)}(X_1))dx = \int_0^\infty S(x)dx = \mathbb{E}(X_1) < \infty.$$

Hence, $\widehat{S}_X \in \mathbb{L}^1(\mathbb{R}_+, x^{-1/2}) \cap \mathbb{L}^2(\mathbb{R}_+, x^0)$ almost surely. □

Proof of Theorem 5.3.2. Let $k \in \mathbb{R}_+$. Since $\mathcal{M}_{1/2}[S-S_k](t) = 0$ for $|t| \le k$ we get by application of the Plancherel identity, $\langle S-S_k, S_k-\widehat{S}_k\rangle = \frac{1}{2\pi}\int_{-k}^{k} \mathcal{M}_{1/2}[S-S_k](t)\mathcal{M}_{1/2}[S_k-\widehat{S}_k](-t)dt = 0$ and thus $\|S-\widehat{S}_k\|^2 = \|S-S_k\|^2 + \|S_k-\widehat{S}_k\|^2$. Again by application of the Plancherel identity, cf. equation (5.2.6), and the Fubini-Tonelli theorem,

$$\mathbb{E}(\|\widehat{S}_k - S_k\|^2) = \frac{1}{2\pi}\int_{-k}^{k} \frac{\mathrm{Var}(\widehat{\mathcal{M}}(t))}{|(1/2+it)\mathcal{M}_{3/2}[g](t)|^2}dt.$$

Now we will use the fact that $\mathrm{Var}(\widehat{\mathcal{M}}(t)) = \mathrm{Var}(\widehat{\mathcal{M}}(t) - \mathbb{E}_{|X}(\widehat{\mathcal{M}}(t))) + \mathrm{Var}(\mathbb{E}_{|X}(\widehat{\mathcal{M}}(t))$. For the first summand we see that by $\mathbb{E}_{|X}(Y_j^{1/2+it}) = \mathbb{E}(U_j^{1/2+it})X_j^{1/2+it}$,

$$\begin{aligned}
&\mathrm{Var}\left(\widehat{\mathcal{M}}(t) - \mathbb{E}_{|X}(\widehat{\mathcal{M}}(t))\right)\\
&= n^{-2}\sum_{j,j'=1}^{n} \mathbb{E}(X_j^{1/2+it}X_{j'}^{1/2-it}(U_j^{1/2+it} - \mathbb{E}(U_j^{1/2+it})(U_{j'}^{1/2-it} - \mathbb{E}(U_{j'}^{1/2-it}))\\
&= n^{-2}\sum_{j,j'=1}^{n} \mathbb{E}(X_j^{1/2+it}X_{j'}^{1/2-it})\mathbb{E}((U_j^{1/2+it} - \mathbb{E}(U_j^{1/2+it})(U_{j'}^{1/2-it} - \mathbb{E}(U_{j'}^{1/2-it}))\\
&= n^{-1}\mathbb{E}(X_1)\mathrm{Var}(U_1^{1/2+it}) \le n^{-1}\mathbb{E}(Y_1).
\end{aligned}$$

On the other hand, we have $\mathbb{E}_{|X}(\widehat{\mathcal{M}}(t)) = n^{-1}\sum_{j=1}^{n} X_j^{1/2+it}\mathbb{E}(U_j^{1/2+it})$ and thus

$$\mathrm{Var}(\mathbb{E}_{|X}(\widehat{\mathcal{M}}(t)) = |\mathbb{E}(U_1^{1/2+it})|^2\mathbb{E}(|n^{-1}\sum_{j=1}^{n} X_j^{1/2+it} - \mathcal{M}_{3/2}[f](t)|^2).$$

Due to these considerations and the fact that $\mathbb{E}(U_1^{1/2+it}) = \mathcal{M}_{3/2}[g](t)$ for all $t \in \mathbb{R}$, we have

$$\mathbb{E}(\|\widehat{S}_k - S_k\|^2) \le \mathbb{E}(Y_1)\frac{\Delta_g(k)}{n} + \frac{1}{2\pi}\int_{-k}^{k} \frac{\mathbb{E}(|\widehat{\mathcal{M}}_X(t) - \mathcal{M}_{3/2}[f](t)|^2)}{|1/2+it|^2}dt.$$

As for the last summand, for any $k \in \mathbb{R}_+$,

$$\frac{1}{2\pi}\int_{-k}^{k} \frac{\mathbb{E}(|\widehat{\mathcal{M}}_X(t) - \mathcal{M}_{3/2}[f](t)|^2)}{|1/2+it|^2}dt = \frac{1}{2\pi}\int_{-k}^{k} \mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t))dt$$

using equation (5.3.5). □

For the proof of Corollary 5.3.3 we will need the following results. The key statement regarding β-mixing processes is delivered by the variance bound in Asin and Johannes [2017] after Lemma 4.1 of the same work. Their approach is based on the original idea of [Viennet, 1997, Theorem 2.1].

Lemma 5.7.3. *Let $(Z_j)_{j\in\mathbb{Z}}$ be a strictly stationary process of real-valued random variables with common marginal distribution $\mathbb{P}$. There exists a sequence $(b_k)_{k\in\mathbb{N}}$ of measurable functions $b_k : \mathbb{R} \to [0,1]$ with $\mathbb{E}_{\mathbb{P}}[b_k(Z_0)] = \beta(Z_0, Z_k)$ such that for any measurable function h with $\mathbb{E}[|h(Z_0)|^2] < \infty$ and $b = \sum_{k=1}^{\infty}(k+1)^{p-2}b_k : \mathbb{R} \to [0,\infty]$, $p \geq 2$,*

$$\mathrm{Var}\Big(\sum_{j=1}^{n} h(Z_j)\Big) \leq 4n\mathbb{E}[|h(Z_0)|^2 b(Z_0)]$$

where $b_0 \equiv 1$.

As far as the functional dependence measure is concerned, we do not have access to the process (X_j) itself. In its given form $(X_j^{1/2+it})$ we make use of the two following statements.

Lemma 5.7.4. *The function $g : \mathbb{R}_+ \to \mathbb{R}$, $x \mapsto x^{1/2+it}$, $t \in \mathbb{R}$, is Hölder continuous with exponent $1/2$, i.e. $|g(x) - g(y)| \leq L(t) \cdot |x-y|^{1/2}$ where $L(t) = 1 + 4|t|^{1/2}$.*

Lemma 5.7.5. *Let $g : \mathbb{R}_+ \to \mathbb{R}$, $x \mapsto x^{1/2+it}$, $t \in \mathbb{R}$. In the case of **(F)** we have $\mathrm{Var}(\sum_{j=1}^{n} g(X_j))^{1/2} \leq L(t) \cdot n^{1/2} \sum_{k=1}^{\infty} \delta_1^X(k)^{1/2}$ where $L(t) = 1 + 4|t|^{1/2}$.*

Proof of Corollary 5.3.3. We consider the different dependence structures separately.

(I) The result for independent observations can directly be obtained since

$$\mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) = |1/2+it|^{-2}n^{-1}\mathrm{Var}(X_1^{1/2+it}) \leq |1/2+it|^{-2}n^{-1}\mathbb{E}(X_1).$$

On the other hand, as $k \to \infty$, $(2\pi)^{-1}\int_{-k}^{k} |1/2+it|^{-2}dt \to 1$.

(B) In the case of β-mixing we employ Lemma 5.7.3. We then have

$$\begin{aligned}\mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) &\leq 4|1/2+it|^{-2}n^{-2} \cdot n\mathbb{E}[|X_1^{1/2+it}|^2 b(X_1)] \\ &\leq 4|1/2+it|^{-2}n^{-1}\mathbb{E}[|X_1|b(X_1)].\end{aligned}$$

As before, as $k \to \infty$, $(2\pi)^{-1}\int_{-k}^{k} |1/2+it|^{-2}dt \to 1$.

(F) We now study the dependence measure. According to Lemma 5.7.5 we have

$$\mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t))dt \leq L^2(t)|1/2+it|^{-2} \cdot n^{-1}\Big(\sum_{k=1}^{\infty} \delta_1^X(k)^{1/2}\Big)^2$$

where $L(t) = 1 + 4|t|^{1/2}$. Therefore, $\int_{-k}^{k} L^2(t)(1/2+it)^{-2}dt \leq c\log(k)$ for a numerical constant $c > 0$.

□

Proof of Lemma 5.7.4. Without loss of generality let $x > y > 0$. By the elementary inequality $|x^{1/2} - y^{1/2}| \leq |x - y|^{1/2}$ we have

$$\begin{aligned} |g(x) - g(y)| = |x^{1/2+it} - y^{1/2+it}| &\leq |x^{it}| \cdot |x-y|^{1/2} + |y|^{1/2} \cdot |x^{it} - y^{it}| \\ &\leq |x-y|^{1/2} + |y|^{1/2} \cdot |x^{it} - y^{it}|. \end{aligned}$$

We then bound the second term. First we see that

$$|y|^{1/2} \cdot |x^{it} - y^{it}| \leq |y|^{1/2} \cdot (|\cos(t\log(x)) - \cos(t\log(y))| + |(\sin(t\log(y)) - \sin(t\log(y)))|).$$

Moreover,

$$\begin{aligned} |\cos(t\log(x)) - \cos(t\log(y))| &\leq |t| \cdot \log(x/y) = |t| \cdot \log\left(1 + \frac{x-y}{y}\right) \\ &\leq |t| \cdot \frac{x-y}{y} \end{aligned}$$

where we used $\log(1+z) \leq z$ for $z > 0$. At the same time $|\cos(t\log(x)) - \cos(t\log(y))| \leq 2$. Applying both bounds together and exploiting $\min\{1, z\} \leq z^s$ for $z \geq 0$, $s \in (0,1)$,

$$\begin{aligned} |y|^{1/2} \cdot |\cos(t\log(x)) - \cos(t\log(y))| &\leq 2|y|^{1/2} \min\left\{|t|\frac{x-y}{y}, 1\right\} \\ &\leq 2|t|^{1/2}|y|^{1/2}\left|\frac{x-y}{y}\right|^{1/2} \\ &= 2|t|^{1/2}| \cdot |x-y|^{1/2}. \end{aligned}$$

A similar argument applies for the sine terms, which delivers

$$|y|^{1/2} \cdot |x^{it} - y^{it}| \leq 4|t|^{1/2} \cdot |x-y|^{1/2}.$$

The case $y > x > 0$ follows analogously by interchanging the roles of x and y. Therefore,

$$|g(x) - g(y)| = |x^{1/2+it} - y^{1/2+it}| \leq (1 + 4|t|^{1/2})|x-y|^{1/2}.$$

□

Proof of Lemma 5.7.5. For a sequence $W_j := J_{j,n}(\mathcal{G}_i)$ with $\|W_1\|_1 < \infty$, let $P_{j-k}(W) := \mathbb{E}[W \mid \mathcal{G}_{j-k}] - \mathbb{E}[W \mid \mathcal{G}_{j-k-1}]$ denote its projection, $k \in \mathbb{N}_0$. Then, by the projection property of the conditional expectation and an elementary property of δ_2 (cf. Wu [2005], Theorem 1), we have

$$\begin{aligned} \mathrm{Var}(\sum_{j=1}^{n} g(X_j))^{1/2} &= \sum_{k=0}^{\infty} \Big\| \sum_{j=1}^{n} P_{j-k} g(X_j) \Big\|_2 \\ &\leq n^{1/2} \sum_{k=0}^{\infty} \delta_2^{g(X)}(k). \end{aligned}$$

By Lemma 5.7.4,

$$\begin{aligned}\delta_2^{g(X)}(k) &= \sup_{j=1,\dots,n} \|g(X_j) - g(X_j)^{*(j-k)}\|_2 \\ &\le L(t) \cdot \sup_{j=1,\dots,n} \|X_j - X_j^{*(j-k)}\|_1^{1/2} = L(t) \cdot \delta_1^X(k)^{1/2},\end{aligned}$$

which delivers our desired statement. □

Proof of Theorem 5.3.8. We first outline the main steps of the proof. We will construct a family of functions in $\mathbb{W}^s_{1/2}(L)$ by a perturbation of the survival function $S_o : \mathbb{R}_+ \to \mathbb{R}_+$ with small "bumps", such that their $\mathbb{L}^2$-distance and the Kullback-Leibler divergence of their induced distributions can be bounded from below and above, respectively. The claim follows by applying Theorem 2.5 in Tsybakov [2008]. We use the following construction, which we present first.

Denote by $C_c^\infty(\mathbb{R})$ the set of all smooth functions with compact support in $\mathbb{R}$ and let $\psi \in C_c^\infty(\mathbb{R})$ be a function with support in $[0,1]$ and $\int_0^1 \psi(x)dx = 0$. For each $K \in \mathbb{N}$ (to be selected below) and $k \in \{0,\dots,K\}$ we define "bump functions" $\psi_{k,K}(x) := \psi(xK - K - k)$, $x \in \mathbb{R}$. For $j \in \mathbb{N}_0$ we set the finite constant $C_{j,\infty} := \max\{\|\psi^{(l)}\|_\infty, l \in \{0,\dots,j\}\}$. Let us further define the operator $\mathcal{S} : C_c^\infty(\mathbb{R}) \to C_c^\infty(\mathbb{R})$ by $\mathcal{S}[f](x) = -xf^{(1)}(x)$ for all $x \in \mathbb{R}$ and define $\mathcal{S}^1 := \mathcal{S}$ and $\mathcal{S}^n := \mathcal{S} \circ \mathcal{S}^{n-1}$ for $n \in \mathbb{N}, n \ge 2$. Now, for $j \in \mathbb{N}$, we define the function $\psi_{k,K,j}(x) := \mathcal{S}^j[\psi_{k,K}](x) = (-1)^j \sum_{i=1}^j c_{i,j} x^i K^i \psi^{(i)}(xK - K - k)$ for $x \in \mathbb{R}_+$ and $c_{i,j} \ge 1$.

For a "bump amplitude" $\delta > 0, \gamma \in \mathbb{N}$ and a vector $\boldsymbol{\theta} = (\theta_1,\dots,\theta_K) \in \{0,1\}^K$ we define

$$S_{\boldsymbol{\theta}}(x) = S_o(x) + \delta K^{-s-\gamma+1} \sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K,\gamma-1}(x) \text{ where } S_o(x) := \exp(-x). \tag{5.7.3}$$

Taking the negative sign of the derivative of this function leads to the density

$$f_{\boldsymbol{\theta}}(x) = f_o(x) + \delta K^{-s-\gamma+1} \sum_{k=0}^{K-1} \theta_{k+1} \psi_{k,K,\gamma}(x) x^{-1} \text{ where } f_o(x) := \exp(-x). \tag{5.7.4}$$

Until now, we did not give a sufficient condition to ensure that our constructed functions $\{S_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \{0,1\}^K\}$ are survival functions. We do this by stating conditions such that the family $\{f_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \{0,1\}^K\}$ is a family of densities.

Lemma 5.7.6. *Let $0 < \delta < \delta_o(\psi,\gamma) := \exp(-2)2^{-\gamma}(C_{\gamma,\infty}c_\gamma)^{-1}$. Then for all $\boldsymbol{\theta} \in \{0,1\}^K$, $f_{\boldsymbol{\theta}}$ from (5.7.4) is a density.*

Furthermore, it is possible to show that these survival functions all lie inside the ellipsoids $\mathbb{W}^s_{1/2}(L)$ for L big enough. This is captured in the following lemma.

Lemma 5.7.7. *Let $s \in \mathbb{N}$. Then, there is $L_{s,\gamma,\delta} > 0$ such that S_o and any $S_{\boldsymbol{\theta}}$ as in* (5.7.3) *with $\boldsymbol{\theta} \in \{0,1\}^K$, $K \in \mathbb{N}$, belong to $\mathbb{W}^s_{1/2}(L_{s,\gamma,\delta})$.*

For sake of simplicity we denote for a function $\varphi \in \mathbb{L}^2(\mathbb{R}_+, x^0) \cap \mathbb{L}^1(\mathbb{R}_+, x^{-1/2})$ the multiplicative convolution with g by $\widetilde{\varphi} := \varphi * g$. Next, we see that for $y_2 \geq y_1 > 0$,

$$\widetilde{f}_o(y_1) = \int_0^\infty g(x)x^{-1}\exp(-y_1/x)dx \geq \int_0^\infty g(x)x^{-1}\exp(-y_2/x)dx = \widetilde{f}_o(y_2)$$

and thus $\widetilde{f}_o$ is monotonically decreasing. Additionally, we have that $\widetilde{f}_o(2) > 0$, since otherwise $g = 0$ almost everywhere. Exploiting *Varshamov-Gilbert's lemma* (cf. Tsybakov [2008]) in Lemma 5.7.8 we show that there exists $M \in \mathbb{N}$ with $M \geq 2^{K/8}$ and a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that for all $j, l \in \{0, \dots, M\}$, $j \neq l$, the $\mathbb{L}^2$-distance and the Kullback-Leibler divergence are bounded for $K \geq K_o(\gamma, \psi)$.

Lemma 5.7.8. *Let $K \geq \max\{K_o(\psi,\gamma), 8\}$. Then there exists a subset $\{\boldsymbol{\theta}^{(0)}, \dots, \boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0, \dots, 0)$ such that $M \geq 2^{K/8}$ and for all $j, l \in \{0, \dots, M\}$, $j \neq l$, $\|S_{\boldsymbol{\theta}^{(j)}} - S_{\boldsymbol{\theta}^{(l)}}\|^2 \geq \frac{\|\psi^{(\gamma-1)}\|^2\delta^2}{16}K^{-2s}$ and $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_1(g)\|\psi\|^2}{\widetilde{f}_o(2)\log(2)}\delta^2\log(M)K^{-2s-2\gamma+1}$ where $\mathrm{KL}(\cdot,\cdot)$ is the Kullback-Leibler divergence.*

Selecting $K = \lceil n^{1/(2s+2\gamma-1)} \rceil$ delivers

$$\frac{1}{M}\sum_{j=1}^M \mathrm{KL}((\widetilde{f}_{\boldsymbol{\theta}^{(j)}})^{\otimes n}, (\widetilde{f}_{\boldsymbol{\theta}^{(0)}})^{\otimes n}) = \frac{n}{M}\sum_{j=1}^M \mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq c_{\psi,\delta,g,\gamma,f_o}\log(M)$$

where $c_{\psi,\delta,g,\gamma,f_o} < 1/8$ for all $\delta \leq \delta_1(\psi, g, \gamma, f_o)$, $M \geq 2$ and for the choice of a large enough $n \geq n_{s,\gamma} := \max\{8^{2s+1}, K_o(\gamma,\psi)^{2s+2\gamma+1}\}$. Thereby, we can use [Tsybakov, 2008, Theorem 2.5], which in turn for any estimator S of $\mathbb{W}^s_{1/2}(L)$ implies

$$\sup_{S \in \mathbb{W}^s_{1/2}(L)} \mathbb{P}\big(\|\widehat{S} - S\|^2 \geq \tfrac{c_{\psi,\delta,\gamma}}{2} n^{-2s/(2s+2\gamma-1)}\big) \geq \tfrac{\sqrt{M}}{1+\sqrt{M}}\big(1 - 1/4 - \sqrt{\frac{1}{4\log(M)}}\big) \geq 0.07.$$

Note that the constant $c_{\psi,\delta,\gamma}$ only depends on ψ, γ and δ. Hence, it is independent of the parameters s and n. The claim of Theorem 5.3.8 follows by using Markov's inequality, which completes the proof. □

Proofs of the lemmata

Proof of Lemma 5.7.6. For any $h \in C_c^\infty(\mathbb{R})$, $\mathcal{S}[h] \in C_c^\infty(\mathbb{R})$ and thus $\mathcal{S}^j[h] \in C_c^\infty(\mathbb{R})$ for any $j \in \mathbb{N}$. Further, for $h \in C_c^\infty(\mathbb{R})$, $\int_{-\infty}^\infty h'(x)dx = 0$, which implies that for any $\delta > 0$ and $\boldsymbol{\theta} \in \{0,1\}^K$ we have $\int_0^\infty f_{\boldsymbol{\theta}}(x)dx = 1$.
Now due to the construction (5.7.4) of the functions $\psi_{k,K}$ we easily see that the function $\psi_{k,K}$ has support on $[1 + k/K, 1 + (k+1)/K]$, which leads to $\psi_{k,K}$ and $\psi_{l,K}$ having disjoint supports if $k \neq l$. Here, we want to emphasize that $\mathrm{supp}(\mathcal{S}[h]) \subseteq \mathrm{supp}(h)$ for all $h \in C_c^\infty(\mathbb{R})$, which implies that $\psi_{k,K,\gamma}$ and $\psi_{l,K,\gamma}$ have disjoint supports if $k \neq l$, too. For $x \in [1,2]^c$ we have $f_{\boldsymbol{\theta}}(x) = \exp(-x) \geq 0$. Now let us consider the case $x \in [1,2]$. Then there is $k_o \in \{0, \dots, K-1\}$ such that $x \in [1 + k_o/K, 1 + (k_o+1)/K]$ and hence

$$S_{\boldsymbol{\theta}}(x) = S_o(x) + \theta_{k_o+1}\delta K^{-s-\gamma+1}x^{-1}\psi_{k_o,K,\gamma}(x) \geq \exp(-2) - \delta 2^\gamma C_{\gamma,\infty}c_\gamma$$

since $\|\psi_{k,K,j}\|_\infty \leq 2^j C_{j,\infty} c_j K^j$ for any $k \in \{0,\dots,K-1\}$, $s \geq 1$ and $j \in \mathbb{N}$ where $c_j := \sum_{i=1}^{j} c_{i,j}$. Choosing $\delta \leq \delta_o(\psi,\gamma) = \exp(-2)2^{-\gamma}(C_{\gamma,\infty}c_\gamma)^{-1}$ ensures $f_{\boldsymbol{\theta}}(x) \geq 0$ for all $x \in \mathbb{R}_+$. $\square$

Proof of Lemma 5.7.7. Our proof starts with the observation that for all $t \in \mathbb{R}$ we have $\mathcal{M}_{1/2}[S_o](t) = \Gamma(1/2+it)$. Now, by applying the Stirling formula (cf. Belomestny and Goldenshluger [2020]) we get $|\Gamma(1/2+it)| \sim \exp(-\pi/2|t|)$, $|t| \geq 2$. Thus for every $s \in \mathbb{N}$ there exists L_s such that $|S_o|_s^2 \leq L$ for all $L \geq L_s$.
Next, we consider $|S_o - S_{\boldsymbol{\theta}}|_s$. Let us define $\Psi_K := \sum_{k=0}^{K-1} \theta_{k+1}\psi_{k,K}$ and $\Psi_{K,j} := \mathcal{S}^j[\Psi_K]$ for $j \in \mathbb{N}$. Then we have $|S_o - S_{\boldsymbol{\theta}}|_s^2 = \delta^2 K^{-2s-2\gamma+2}|\Psi_{K,\gamma-1}|_s^2$ where $|\,.\,|_s$ is defined in (5.3.6). Now since for any $j \in \mathbb{N}$, $\operatorname{supp}(\Psi_{K,j}) \subset [1,2]$ and $\|\Psi_{K,j}\|_\infty < \infty$, we have that the Mellin transform of $\Psi_{K,j}$ is defined for any $c \in (0,\infty)$. By a recursive application of the integration by parts we deduce that $|\mathcal{M}_{1/2}[\Psi_{K,s+\gamma-1}](t)|^2 = (1/4+t^2)^s|\mathcal{M}_{1/2}[\Psi_{k,\gamma-1}](t)|^2$, whence

$$|\Psi_{k,\gamma-1}|_s^2 \leq C_s \int_{-\infty}^{\infty} |\mathcal{M}_{1/2}[\Psi_{K,s+\gamma-1}](t)|^2 dt = C_s \int_0^\infty |\Psi_{K,s+\gamma-1}(x)|^2 dx$$

by the Parseval formula, see equation (5.2.6), where $C_s > 0$ is a positive constant. Since the $\psi_{k,K}$ have disjoint support for different values of k we reason that $|\Psi_{k,\gamma-1}|_s^2 \leq C_s \sum_{k=0}^{K-1} \theta_{k+1}^2 \int_0^\infty |\mathcal{S}^{\gamma-1+s}[\psi_{k,K}](x)|^2 dx$. Applying Jensen's inequality and considering the fact that $\operatorname{supp}(\psi_{k,K}) \subset [1,2]$, we obtain

$$\begin{aligned}|\Psi_{k,\gamma-1}|_s^2 &\leq C_{(\gamma,s)} \sum_{k=0}^{K-1}\sum_{j=1}^{\gamma+s-1} c_{j,\gamma-1+s}^2 \int_1^2 x^{2j}K^{2j}\psi^{(j)}(xK-K-k)^2 dx \\ &\leq C_{(\gamma,s)} K^{2(\gamma-1+s)} \sum_{k=0}^{K-1}\sum_{j=1}^{\gamma+s-1} c_{j,\gamma+s}^2 4^j C_{\psi,s,\gamma}^2 K^{-1} \leq C_{(\gamma,s)} K^{2(\gamma-1+s)}.\end{aligned}$$

Thus, $|S_o - S_{\boldsymbol{\theta}}|_s^2 \leq C_{(s,\gamma,\delta)}$ and $|S_{\boldsymbol{\theta}}|_s^2 \leq 2(|S_o - S_{\boldsymbol{\theta}}|_s^2 + |S_o|_s^2) \leq 2(C_{(s,\gamma,\delta)} + L_s) =: L_{s,\gamma,\delta,1}$. By Corollary 5.3.7 it is sufficient to show that $\int_0^\infty x f_{\boldsymbol{\theta}}(x)dx \leq L_{s,\gamma,\delta,2}$. In fact,

$$\int_0^\infty x f_{\boldsymbol{\theta}}(x)dx = 1 + \delta K^{-s-\gamma+1} \sum_{k=0}^{K-1} \int_{1+k/K}^{1+(k+1)/K} \psi_{k,K,\gamma}(x)dx \leq 1 + \delta C_\gamma$$

since $\|\psi_{k,K,\gamma}\|_\infty \leq 2^\gamma C_{\gamma,\infty} c_\gamma K^\gamma = C_\gamma K^\gamma$, cf. the proof of Lemma 5.7.6. The claim follows by choosing $L_{s,\gamma,\delta} = \max\{L_{s,\gamma,\delta,1}, L_{s,\gamma,\delta,2}\}$. $\square$

Proof of Lemma 5.7.8. Using the fact that the functions $(\psi_{k,K,\gamma})_{k\in\{0,\dots,K-1\}}$ with different index k have disjoint supports we get

$$\begin{aligned}\|S_{\boldsymbol{\theta}} - S_{\boldsymbol{\theta}'}\|^2 &= \delta^2 K^{-2(s+\gamma-1)} \|\sum_{k=0}^{K-1} (\theta_{k+1} - \theta'_{k+1})\psi_{k,K,\gamma-1}\|^2 \\ &= \delta^2 K^{-2(s+\gamma-1)} \rho(\boldsymbol{\theta},\boldsymbol{\theta}')\|\psi_{0,K,\gamma-1}\|^2\end{aligned}$$

with $\rho(\boldsymbol{\theta}, \boldsymbol{\theta}') := \sum_{j=0}^{K-1} \mathbb{1}_{\{\boldsymbol{\theta}_{j+1} \neq \boldsymbol{\theta}'_{j+1}\}}$, the *Hamming distance*. Now the first claim follows by showing that $\|\psi_{0,K,\gamma-1}\|^2 \geq \frac{K^{2\gamma-3}\|\psi^{(\gamma-1)}\|^2}{2}$ for K big enough. To do so, we observe

$$\|\psi_{0,K,\gamma-1}\|^2 = \sum_{i,j\in\{1,\dots,\gamma-1\}} c_{j,\gamma-1}c_{i,\gamma-1}\int_0^\infty x^{j+i+1}\psi_{0,K}^{(j)}(x)\psi_{0,K}^{(i)}(x)dx.$$

Defining $\Sigma := \|\psi_{0,K,\gamma-1}\|^2 - \int_0^\infty (x^{\gamma-1}\psi_{0,K}^{(\gamma-1)}(x))^2 dx$, we see

$$\|\psi_{0,K,\gamma-1}\|^2 = \Sigma + \int_0^\infty (x^{\gamma-1}\psi_{0,K}^{(\gamma-1)}(x))^2 dx \geq \Sigma + K^{2\gamma-3}\|\psi^{(\gamma-1)}\|^2 \geq \frac{K^{2\gamma-3}\|\psi^{(\gamma-1)}\|^2}{2} \tag{5.7.5}$$

as soon as $|\Sigma| \leq \frac{K^{2\gamma-3}\|\psi^{(\gamma-1)}\|^2}{2}$. This is obviously true because $K \geq K_o(\gamma,\psi)$ and thus $\|S_{\boldsymbol{\theta}} - S_{\boldsymbol{\theta}'}\|^2 \geq \frac{\delta^2\|\psi^{(\gamma-1)}\|^2}{2}K^{-2s-1}\rho(\boldsymbol{\theta},\boldsymbol{\theta}')$ for $K \geq K_o(\psi,\gamma)$.
Now we use the *Varshamov-Gilbert Lemma* (cf. Tsybakov [2008]) which states that for $K \geq 8$ there exists a subset $\{\boldsymbol{\theta}^{(0)},\dots,\boldsymbol{\theta}^{(M)}\}$ of $\{0,1\}^K$ with $\boldsymbol{\theta}^{(0)} = (0,\dots,0)$ such that $\rho(\boldsymbol{\theta}^{(j)},\boldsymbol{\theta}^{(k)}) \geq K/8$ for all $j,k \in \{0,\dots,M\}, j \neq k$ and $M \geq 2^{K/8}$. Therefore, $\|S_{\boldsymbol{\theta}^{(j)}} - S_{\boldsymbol{\theta}^{(l)}}\|_\omega^2 \geq \frac{\|\psi^{(\gamma-1)}\|^2\delta^2}{16}K^{-2s}$.
For the second part we have $f_o = f_{\boldsymbol{\theta}^{(0)}}$ and by using $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) \leq \chi^2(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) := \int_{\mathbb{R}_+}(\widetilde{f}_{\boldsymbol{\theta}}(x) - \widetilde{f}_o(x))^2/\widetilde{f}_o(x)dx$ it is sufficient to bound the χ^2-divergence. We notice that $\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o$ has support on $[0,2]$ since $f_{\boldsymbol{\theta}} - f_o$ has support on $[1,2]$ and g has support on $[0,1]$. For $y > 2$, $\widetilde{f}_{\boldsymbol{\theta}}(y) - \widetilde{f}_o(y) = \int_y^\infty (f_{\boldsymbol{\theta}} - f_o)(x)x^{-1}g(y/x)dx = 0$. Let $\Psi_{K,\gamma} := \sum_{k=0}^{K-1}\theta_{k+1}\psi_{k,K,\gamma} = \mathcal{S}^\gamma[\sum_{k=0}^{K-1}\theta_{k+1}\psi_{k,K}] =: \mathcal{S}^\gamma[\Psi_K]$. Now by using the compact support property and a single substitution we get

$$\chi^2(\widetilde{f}_{\boldsymbol{\theta}}, \widetilde{f}_o) \leq \widetilde{f}_o(2)^{-1}\|\widetilde{f}_{\boldsymbol{\theta}} - \widetilde{f}_o\|^2 = \widetilde{f}_o(2)^{-1}\delta^2K^{-2s-2\gamma+2}\|\widetilde{\omega^{-1}\Psi_{K,\gamma}}\|^2.$$

Let us now consider $\|\widetilde{\omega^{-1}\Psi_{K,\gamma}}\|^2$. In a first step we see by application of the Parseval equality that $\|\widetilde{\omega^{-1}\Psi_{K,\gamma}}\|^2 = \frac{1}{2\pi}\int_{-\infty}^\infty |\mathcal{M}_{1/2}[\widetilde{\omega^{-1}\Psi_{K,y}}](t)|^2dt$. Now for $t \in \mathbb{R}$, we see by using the multiplication theorem for Mellin transforms that $\mathcal{M}_{1/2}[\widetilde{\omega^{-1}\Psi_{K,\gamma}}](t) = \mathcal{M}_{1/2}[g](t)\cdot\mathcal{M}_{1/2}[\omega^{-1}\mathcal{S}^\gamma[\Psi_K]](t)$. Again we have $\mathcal{M}_{1/2}[\omega^{-1}S^\gamma[\Psi_K]](t) = (-1/2+it)^\gamma\mathcal{M}_{-1/2}[\Psi_K](t)$. Together with assumption **[G1']** we obtain

$$\|\widetilde{\omega^{-1}\Psi_{K,\gamma}}\|^2 \leq \frac{C_1(g)}{2\pi}\int_{-\infty}^\infty |\mathcal{M}_{-1/2}[\Psi_K](t)|^2dt = C_1(g)\|\omega^{-1}\Psi_K\|^2 \leq C_1(g)\|\psi\|^2.$$

Since $M \geq 2^K$ we have $\mathrm{KL}(\widetilde{f}_{\boldsymbol{\theta}^{(j)}}, \widetilde{f}_{\boldsymbol{\theta}^{(0)}}) \leq \frac{C_1(g)\|\psi\|^2}{\widetilde{f}_o(2)\log(2)}\delta^2\log(M)K^{-2s-2\gamma+1}$. □

5.7.3 Proofs of Section 5.4

Proof of Theorem 5.4.1. Let us define nested subspaces $(U_k)_{k\in\mathbb{R}_+}$ by $U_k := \{h \in \mathbb{L}^2(\mathbb{R}_+, x^0) : \forall |t| \geq k : \mathcal{M}_{1/2}[h](t) = 0\}$. For any $h \in U_k$ we consider the empirical contrast

$$\gamma_n(h) = \|h\|^2 - 2\frac{1}{2\pi}\int_{-\infty}^{\infty} \widehat{\mathcal{M}}(t)\frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}dt = \|h\|^2 - 2n^{-1}\sum_{j=1}^{n}\nu_h(Y_j)$$

with $\nu_h(Y_j) := \frac{1}{2\pi}\int_{-\infty}^{\infty} Y_j^{1/2+it}\frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}dt$. It can be seen easily that $\widehat{S}_k = \arg\min\{\gamma_n(h) : h \in U_k\}$ with $\gamma_n(\widehat{S}_k) = -\|\widehat{S}_k\|^2$. For $h \in U_k$ define the centered empirical process $\bar{\nu}_h := n^{-1}\sum_{j=1}^n \nu_h(Y_j) - \langle h, S\rangle$. Then we have for $h_1, h_2 \in U_k$,

$$\gamma_n(h_1) - \gamma_n(h_2) = \|h_1 - S\|^2 - \|h_2 - S\|^2 - 2\bar{\nu}_{h_1-h_2}. \tag{5.7.6}$$

Now since $\gamma_n(\widehat{S}_k) \leq \gamma_n(S_k)$ and by the definition of $\widehat{k}$, we have $\gamma_n(\widehat{S}_{\widehat{k}}) - \widehat{\mathrm{pen}}(\widehat{k}) \leq \gamma_n(\widehat{S}_k) - \widehat{\mathrm{pen}}(k) \leq \gamma_n(S_k) - \widehat{\mathrm{pen}}(k)$ for any $k \in \mathcal{K}_n$. Using (5.7.6),

$$\|S - \widehat{S}_{\widehat{k}}\|^2 \leq \|S - S_k\|^2 + 2\bar{\nu}_{\widehat{S}_{\widehat{k}}-S_k} + \widehat{\mathrm{pen}}(k) - \widehat{\mathrm{pen}}(\widehat{k}).$$

First we note that $U_{k_1} \subseteq U_{k_2}$ for $k_1 \leq k_2$. Let us now denote by $a \vee b := \max\{a, b\}$ and define for all $k \in \mathcal{K}_n$ the unit balls $B_k := \{h \in U_k : \|h\| \leq 1\}$. Next, we deduce from $2ab \leq a^2 + b^2$ that $2\bar{\nu}_{\widehat{S}_{\widehat{k}}-S_k} \leq 4^{-1}\|\widehat{S}_{\widehat{k}} - S_k\|^2 + 4\sup_{h\in B_{\widehat{k}\vee k}} \bar{\nu}_h^2$. Furthermore, we see that $4^{-1}\|\widehat{S}_{\widehat{k}} - S_k\|^2 \leq 2^{-1}(\|\widehat{S}_{\widehat{k}} - S\|^2 + \|S - S_k\|^2)$. Putting all the facts together and defining

$$p(\widehat{k} \vee k) := 24\sigma_Y \Delta_g(\widehat{k} \vee k)n^{-1}. \tag{5.7.7}$$

we have

$$\begin{aligned}\|S - \widehat{S}_{\widehat{k}}\|^2 \;&\leq\; 3\|S - S_k\|^2 \\ &\quad + 8\big(\sup_{h\in B_{\widehat{k}\vee k}} \bar{\nu}_h^2 - p(k \vee \widehat{k})\big)_+ + 8p(\widehat{k} \vee k) + 2\widehat{\mathrm{pen}}(k) - 2\widehat{\mathrm{pen}}(\widehat{k}).\end{aligned}$$

The decomposition $\bar{\nu}_h = \bar{\nu}_{h,in} + \bar{\nu}_{h,de}$ where

$$\bar{\nu}_{h,in} := n^{-1}\sum_{j=1}^{n}(\nu_h(Y_j) - \mathbb{E}_{|X}(\nu_h(Y_j))) \text{ and } \bar{\nu}_{h,de} = n^{-1}\sum_{j=1}^{n}\mathbb{E}_{|X}(\nu_h(Y_j))) - \mathbb{E}(\nu_h(Y_j))$$

implies the inequality

$$\begin{aligned}\|S - \widehat{S}_{\widehat{k}}\|^2 \leq& 3\|S - S_k\|^2 + 16\big(\sup_{h\in B_{\widehat{k}\vee k}} \bar{\nu}_{h,in}^2 - \frac{1}{2}p(k \vee \widehat{k})\big)_+ \\ &+ 16\sup_{h\in B_{\widehat{k}\vee k}} \bar{\nu}_{h,de}^2 + 8p(\widehat{k} \vee k) + 2\widehat{\mathrm{pen}}(k) - 2\widehat{\mathrm{pen}}(\widehat{k}).\end{aligned}$$

Assuming that $\chi \geq 96$, $4p(\widehat{k} \vee k) \leq \mathrm{pen}(k) + \mathrm{pen}(\widehat{k})$. Thus,

$$\begin{aligned}\|S - \widehat{S}_{\widehat{k}}\|^2 \leq & 6\big(\|S - S_k\|^2 + \mathrm{pen}(k)\big) + 16 \max_{k \in \mathcal{K}_n} \big(\sup_{h \in B_k} \bar{\nu}_{h,in}^2 - \frac{1}{2} p(k)\big)_+ \\ & + 16 \max_{k' \in \mathcal{K}_n} \sup_{h \in B_{k'}} \bar{\nu}_{h,de}^2 + 2(\widehat{\mathrm{pen}}(k) - 2\mathrm{pen}(k)) + 2(\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+.\end{aligned}$$

We will use the following Lemmata which will be proven afterwards.

Lemma 5.7.9. *Under the assumption of Theorem 5.4.1 we have*

$$\mathbb{E}(\max_{k \in \mathcal{K}_n} \sup_{h \in B_k} \bar{\nu}_{h,de}^2) \leq \frac{1}{2\pi} \int_{-n}^{n} \mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) dt$$

Lemma 5.7.10. *Under the assumption of Theorem 5.4.1 we have*

$$\mathbb{E}\big(\max_{k \in \mathcal{K}_n} \big(\sup_{h \in B_k} \bar{\nu}_{h,in}^2 - \frac{1}{2} p(k)\big)_+ \leq C(g) \big(\frac{\sigma_X}{n} + \frac{\mathbb{E}(X_1^{5/2})}{\sigma_X^{3/2} n} + \frac{\mathrm{Var}(\widehat{\sigma}_X)}{\sigma_X}\big)$$

Lemma 5.7.11. *Under the assumption of Theorem 5.4.1 we have*

$$\mathbb{E}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \leq 4\chi \frac{\mathbb{E}(Y_1^2)}{\sigma_Y n} + 4\chi \frac{\sigma_U}{\sigma_X} \mathrm{Var}(\widehat{\sigma}_X).$$

Applying the lemmata and using the fact that $\mathbb{E}(\widehat{\mathrm{pen}}(k)) = 2\mathrm{pen}(k)$,

$$\begin{aligned}\mathbb{E}(\|S - \widehat{S}_{\widehat{k}}\|^2) \leq 6\big(\|S - S_k\|^2 + \mathrm{pen}(k)\big) + \frac{C(g, \chi, f)}{n} + C(g, \chi) \frac{\mathrm{Var}(\widehat{\sigma}_X)}{\sigma_X} \\ + \frac{1}{2\pi} \int_{-n}^{n} \mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) dt.\end{aligned}$$

Since this inequality holds true for all $k \in \mathcal{K}_n$, it implies the claim. □

Proofs of the lemmata

Proof of Lemma 5.7.9. By applying the Cauchy-Schwarz inequality and the identity $\mathbb{E}_{|X}(Y_j^{1/2+it}) - \mathbb{E}(Y_j^{1/2+it}) = \mathcal{M}_{3/2}[g](t)(X_j^{1/2+it} - \mathbb{E}(X_1^{1/2+it})$, we get for any $k \in \mathcal{K}_n$ and any $h \in B_k$,

$$\begin{aligned}\frac{1}{n} \sum_{j=1}^{n} \mathbb{E}_{|X}(\nu_h(Y_j)) - \mathbb{E}(\nu_h(Y_j)) &= \frac{1}{2\pi n} \int_{-k}^{k} \frac{(\sum_{j=1}^n X_j^{1/2+it} - \mathbb{E}(X_j^{1/2+it}) \mathcal{M}_{1/2}[h](-t)}{(1/2 + it)} dt \\ &\leq \left(\frac{1}{2\pi} \int_{-k}^{k} \frac{|n^{-1} \sum_{j=1}^n X_j^{1/2+it} - \mathbb{E}(X_j^{1/2+it})|^2}{1/4 + t^2} dt \right)^{1/2} \|h\|_{\mathbb{L}^2(\mathbb{R}_+, x^0)}.\end{aligned}$$

Now since $\|h\|_{\mathbb{L}^2(\mathbb{R}_+, x^0)} \leq 1$, $\mathbb{E}(\max_{k \in \mathcal{K}_n} \sup_{h \in B_k} \bar{\nu}_{h,de}^2) \leq \frac{1}{2\pi} \int_{-n}^{n} \mathrm{Var}(\mathcal{M}_{1/2}[\widehat{S}_X](t)) dt$. □

Proof of Lemma 5.7.10. First let us define $\widetilde{p}(k) := 24\sigma_U \widehat{\sigma}_X \Delta_g(k) n^{-1}$. Then,

$$\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} \bar{\nu}^2_{h,in} - \frac{1}{2}p(k))_+ = \max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} \bar{\nu}^2_{h,in} - \frac{1}{2}\widetilde{p}(k))_+ + \frac{1}{2}\max_{k\in\mathcal{K}_n}(\widetilde{p}(k) - p(k))_+.$$

For the second summand we have

$$\mathbb{E}(\max_{k\in\mathcal{K}_n}(\widetilde{p}(k) - p(k))_+) \leq 24\sigma_U \mathbb{E}((\widehat{\sigma}_X - \sigma_X)_+).$$

Let us define $\Omega_X := \{|\widehat{\sigma}_X - \sigma_X| \leq \sigma_X/2\}$. Then on Ω_X we have $\widehat{\sigma}_X \leq 3\sigma_X/2$ and thus $\mathbb{E}((\widehat{\sigma}_X - \sigma_X)_+) = \mathbb{E}((\widehat{\sigma}_X - \sigma_X)_+ \mathbb{1}_{\Omega_X^c}) \leq 2\sigma_X^{-1}\mathrm{Var}(\widehat{\sigma}_X)$ by application of Cauchy-Schwarz and the Markov's inequality.

For the first summand we see

$$\mathbb{E}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} \bar{\nu}^2_{h,in} - \frac{1}{2}\widetilde{p}(k))_+) = \mathbb{E}(\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} \bar{\nu}^2_{h,in} - \frac{1}{2}\widetilde{p}(k))_+)).$$

Thus we start by considering the inner conditional expectation in order to bound the entire term. By the construction of $\bar{\nu}_{h,in}$, its summands conditioned on $\sigma(X_i, i \geq 0)$ are independent but not identically distributed. We therefore split the process again in the following way

$$\begin{aligned}\bar{\nu}_{h,1} :=& n^{-1}\sum_{j=1}^{n} \nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_j^{1/2}) - \mathbb{E}_{|X}(\nu_h(Y_1)\mathbb{1}_{(0,c_n)}(Y_1^{1/2}))\\ &\text{and } \bar{\nu}_{h,2} := n^{-1}\sum_{j=1}^{n} \nu_h(Y_j)\mathbb{1}_{(c_n,\infty)}(Y_j^{1/2}) - \mathbb{E}_{|X}(\nu_h(Y_1)\mathbb{1}_{(c_n,\infty)}(Y_1^{1/2}))\end{aligned}$$

to get

$$\begin{aligned}\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} |\bar{\nu}_{h,in}|^2 - \frac{1}{2}\widetilde{p}(k))_+) &\leq 2\mathbb{E}_{|X}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k} |\bar{\nu}_{h,1}|^2 - \frac{1}{4}\widetilde{p}(k))_+ + |\bar{\nu}_{h,2}|^2),\\ &:= M_1 + M_2,\end{aligned}$$

where we will now consider the two summands M_1, M_2 separately.
To bound the M_1 term we will make use of Talagrand's inequality (5.7.2) on the term $\mathbb{E}_{|X}(\sup_{t\in B_k} |\bar{\nu}_{h,1}|^2 - \frac{1}{4}\widetilde{p}(k))_+$. We have

$$M_1 \leq \sum_{k=1}^{K_n} \mathbb{E}_{|X}(\sup_{t\in B_k} |\bar{\nu}_{h,1}|^2 - \frac{1}{4}\widetilde{p}(k))_+,$$

which will be used to show the claim. We want to emphasize that we are able to apply the Talagrand's inequality on the sets B_k, since B_k has a dense countable subset and due to continuity arguments. Further, we see that the random variables $\nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_j^{1/2}) - \mathbb{E}_{|X}(\nu_h(Y_1)\mathbb{1}_{(0,c_n)}(Y_1^{1/2}))$, $j = 1, \dots, n$, are conditioned on $\sigma(X_i, i \geq 0)$, centered and independent but not identically distributed. In order to apply Talagrand's inequality, we need to find the constants Ψ, ψ, τ such that

$$\sup_{h\in B_k}\sup_{y>0}|\nu_h(y)\mathbb{1}_{(0,c_n)}(y^{1/2})|\leq\psi;\quad \mathbb{E}_{|X}(\sup_{h\in B_k}|\bar{\nu}_{h,1}|)\leq\Psi;$$

$$\sup_{h\in B_k}\frac{1}{n}\sum_{j=1}^{n}\mathrm{Var}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_j^{1/2}))\leq\tau.$$

We start by determining the constant Ψ^2.

Let us define $\widetilde{M}(t) := n^{-1}\sum_{j=1}^{n} Y_j^{1/2+it}\mathbb{1}_{(0,c_n)}(Y_j^{1/2}))$ as an unbiased estimator of $\mathcal{M}_{3/2}[f_Y\mathbb{1}_{(0,c_n)}](t)$ and

$$\widetilde{S}_k(x) := \frac{1}{2\pi}\int_{-k}^{k} x^{-1/2-it}\frac{\widetilde{\mathcal{M}}(t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}dt$$

where $n^{-1}\sum_{j=1}^{n}\nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_i) = \langle\widetilde{S}_k, h\rangle$. Thus, we have for any $h\in B_k$ that $\bar{\nu}_{h,1}^2 = \langle h, \widetilde{S}_k - \mathbb{E}_{|X}(\widetilde{S}_k)\rangle^2 \leq \|h\|^2\|\widetilde{S}_k - \mathbb{E}_{|X}(\widetilde{S}_k)\|^2$. Since $\|h\|\leq 1$, we get

$$\mathbb{E}_{|X}(\sup_{h\in B_k}\bar{\nu}_{h,1}^2)\leq\mathbb{E}_{|X}(\|\widetilde{S}_k-\mathbb{E}_{|X}(\widetilde{S}_k)\|^2) = \frac{1}{2\pi}\int_{-k}^{k}\frac{\mathbb{E}_{|X}(|\widetilde{\mathcal{M}}(t)-\mathbb{E}_{|X}(\widetilde{\mathcal{M}}(t))|^2)}{(1/4+t^2)|\mathcal{M}_{3/2}[g](t)|^2}dt.$$

Now since $Y_j^{1/2+it}\mathbb{1}_{(0,c_n)}(Y_j^{1/2}) - \mathbb{E}_{|X}(Y_j^{1/2+it}\mathbb{1}_{(0,c_n)}(Y_j^{1/2})$ are independent conditioned on $\sigma(X_i : i\geq 0)$ we obtain

$$\mathbb{E}_{|X}(|\widetilde{\mathcal{M}}(t)-\mathbb{E}_{|X}(\widetilde{\mathcal{M}}(t))|^2)\leq\frac{1}{n^2}\sum_{j=1}^{n}\mathbb{E}_{|X}(Y_j\mathbb{1}_{(0,c_n)}(Y_j^{1/2})) = \frac{\sigma_U}{n}\widehat{\sigma}_X,$$

which implies

$$\mathbb{E}(\sup_{h\in B_k}\bar{\nu}_{h,1}^2)\leq\sigma_U\widehat{\sigma}_X\frac{\Delta_g(k)}{n} =: \Psi^2.$$

Thus $6\Psi^2 = \frac{1}{4}\widetilde{p}(k)$.
We now deal with ψ. Let $y>0$ and $h\in B_k$. Then using the Cauchy-Schwarz inequality, $|\nu_h(y)\mathbb{1}_{(0,c_n)}(y)|^2 = (2\pi)^{-2}c_n^2|\int_{-k}^{k} y^{it}\frac{\mathcal{M}_{1/2}[h](-t)}{(1/2+it)\mathcal{M}_{3/2}[g](t)}dt|^2 \leq (2\pi)^{-1}c_n^2\int_{-k}^{k}|(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}dt\leq c_n^2\Delta_g(k) =: \psi^2$ since $|y^{it}|=1$ for all $t\in\mathbb{R}$.
Next, we consider τ. For $h\in B_k$ we can conclude

$$\begin{aligned}
&\mathrm{Var}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_j^{1/2}))\leq\mathbb{E}_{|X}(\nu_h(Y_j)^2)\\
&\quad=\frac{1}{4\pi^2}\int_{-k}^{k}\int_{-k}^{k}\frac{\mathbb{E}_{|X}(Y_j^{1+i(t_1-t_2)})}{(1/2+it_1)(1/2-it_2)}\frac{\mathcal{M}_{1/2}[h](-t_1)}{\mathcal{M}_{3/2}[g](t_1)}\frac{\mathcal{M}_{1/2}[h](t_2)}{\mathcal{M}_{3/2}[g](-t_2)}dt_1dt_2\\
&\quad=\frac{1}{4\pi^2}\int_{-k}^{k}\int_{-k}^{k}\frac{X_j^{1+i(t_1-t_2)}\mathbb{E}_g(U_1^{1+i(t_1-t_2)})}{(1/2+it_1)(1/2-it_2)}\frac{\mathcal{M}_{1/2}[h](-t_1)}{\mathcal{M}_{3/2}[g](t_1)}\frac{\mathcal{M}_{1/2}[h](t_2)}{\mathcal{M}_{3/2}[g](-t_2)}dt_1dt_2\\
&\quad=X_j\int_0^{\infty}g(u)u\left|\mathcal{M}_{1/2}^{-1}[\mathbb{1}_{[-k,k]}(t)\frac{\mathcal{M}_{1/2}[h](t)}{(1/2-it)\mathcal{M}_{3/2}[g](-t)}](u)\right|^2du.
\end{aligned}$$

Taking the supremum of $u \mapsto ug(u)$ and applying the Plancherel theorem delivers

$$\mathrm{Var}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_j^{1/2})) \leq X_j \|xg\|_\infty \frac{1}{2\pi}\int_{-k}^{k} \frac{|\mathcal{M}_{1/2}[h](t)|^2}{|(1/2+it)\mathcal{M}_{3/2}[g](t)|^2} dt.$$

Now since $\|h\|^2 \leq 1$ and for $G_k(t) := \mathbb{1}_{[-k,k]}(t)|(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2}$,

$$\sup_{h \in B_k} \frac{1}{n}\sum_{j=1}^{n} \mathrm{Var}_{|X}(\nu_h(Y_j)\mathbb{1}_{(0,c_n)}(Y_j^{1/2})) \leq \widehat{\sigma}_X \|G_k\|_\infty \|xg\|_\infty =: \tau.$$

Hence, we have $\frac{n\Psi^2}{6\tau} = \frac{\sigma_U \Delta_g(k)}{6\|xg\|_\infty \|G_k\|_\infty}$ and $\frac{n\Psi}{100\psi} = \frac{\sqrt{\sigma_U \widehat{\sigma}_X n}}{100 c_n}$. Choosing the sequence $c_n := \frac{\sqrt{\sigma_U \widehat{\sigma}_X n}}{a 100 \log(n)}$ gives $\frac{n\Psi}{100\psi} = a\log(n)$. We deduce

$$\begin{aligned}\mathbb{E}_{|X}\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - \frac{1}{4}\widetilde{p}(k)\big)_+ \leq \frac{C}{n}\bigg(&\widehat{\sigma}_X \|G_k\|_\infty \|xg\|_\infty \exp(-\frac{\pi\sigma_U \Delta_g(k)}{3\|xg\|_\infty \|G_k\|_\infty}) \\ &+ \frac{\sigma_U \widehat{\sigma}_X \Delta_g(k)}{\log(n)^2} n^{-a}\bigg).\end{aligned}$$

Under **[G1]** we have $C_g k^{2\gamma-1} \geq \Delta_g(k) \geq c_g k^{2\gamma-1}$ and for all $t \in \mathbb{R}$ it holds true that $c_g k^{2\gamma-2} \leq |G_k(t)| \leq C_g k^{2\gamma-2}$. Hence,

$$\sum_{k=1}^{K_n} \mathbb{E}_{|X}\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - \frac{1}{4}\widetilde{p}(k)\big)_+ \leq \frac{C(g)\widehat{\sigma}_X}{n}\left(\sum_{k=1}^{K_n} k^{2\gamma-1}\exp(-C(g)k) + \sum_{k=1}^{K_n} \frac{k^{2\gamma-1}}{\log(n)^2 n^a}\right)$$

where the first sum is bounded in $n \in \mathbb{N}$. The second sum can be bounded by the term $C(g) n^{\frac{2\gamma}{2\gamma-1}-a}/\log(n)^2$, which by choosing $a = \frac{2\gamma}{2\gamma-1}$ ensures the boundedness in $n \in \mathbb{N}$. Thus, we have

$$\sum_{k=1}^{K_n} \mathbb{E}_{|X}\big(\sup_{h\in B_k} \bar{\nu}_{h,1}^2 - \frac{1}{4}\widetilde{p}(k)\big)_+ \leq \frac{C(g)\widehat{\sigma}_X}{n}.$$

Now, we consider M_2. Let us define $\overline{S}_k := \widehat{S}_k - \widetilde{S}_k$. Then, from $\nu_{h,2} = \nu_{h,in} - \nu_{h,1}$ we deduce $\nu_{h,2}^2 = \langle \overline{S}_k - \mathbb{E}_{|X}(\overline{S}_k), h\rangle^2 \leq \|\overline{S_k} - \mathbb{E}_{|X}(\overline{S}_k)\|^2$ for any $h \in B_k$. Further, for any $k \in \mathcal{K}_n$, $\|\overline{S_k} - \mathbb{E}_{|X}(\overline{S}_k)\|^2 \leq \|\overline{S}_{K_n} - \mathbb{E}_{|X}(\overline{S}_{K_n})\|^2$ and

$$\begin{aligned}\mathbb{E}_{|X}(\|\overline{S}_{K_n} - \mathbb{E}_{|X}(\overline{S}_{K_n}))\|^2) &= \frac{1}{2\pi}\int_{-K_n}^{K_n} \mathrm{Var}_{|X}(\widehat{\mathcal{M}}(t) - \widetilde{\mathcal{M}}(t))|(1/2+it)\mathcal{M}_{3/2}[g](t)|^{-2} dt \\ &\leq \frac{1}{n^2}\sum_{j=1}^{n} \mathbb{E}_{|X}(Y_j \mathbb{1}_{(c_n,\infty)}(Y_j^{1/2}))\Delta_g(K_n).\end{aligned}$$

Let us define the event $\Xi_X := \{\widehat{\sigma}_X \geq \sigma_X/2\}$. Then, we have

$$\frac{1}{n^2}\sum_{j=1}^{n} \mathbb{E}_{|X}(Y_j \mathbb{1}_{(c_n,\infty)}(Y_j^{1/2}))\Delta_g(K_n)\mathbb{1}_{\Xi_X} \leq \frac{1}{n}\sum_{j=1}^{n} X_j^{1+p/2}\mathbb{E}(U_j^{1+p/2}) c_n^{-p}\mathbb{1}_{\Xi_X}$$

where on Ξ_X we can state that $c_n^{-p} = C(g)n^{-p/2}(\widehat{\sigma}_X)^{-p/2}\log(n)^p \le C(g)\sigma_X^{-p/2}n^{-p/2}\log(n)^p$. Choosing $p = 3$ leads to $\mathbb{E}_{|X}(\|\overline{S}_{K_n} - \mathbb{E}_{|X}(\overline{S}_{K_n}))\|^2)\mathbb{1}_{\Xi_X} \le \frac{C(g)\sigma_X^{-3/2}}{n}\mathbb{E}(U_1^{5/2})n^{-1}\sum_{j=1}^n X_j^{5/2}$. On the other hand,

$$\frac{1}{n^2}\sum_{j=1}^n \mathbb{E}_{|X}(Y_j\mathbb{1}_{(c_n,\infty)}(Y_j^{1/2}))\Delta_g(K_n)\mathbb{1}_{\Xi_X^c} \le \frac{\sigma_X}{2}\mathbb{1}_{\Xi_X^c} \le \frac{\sigma_X}{2}\mathbb{1}_{\Omega_X^c}.$$

These three bounds imply

$$\mathbb{E}(\max_{k\in\mathcal{K}_n}(\sup_{h\in B_k}\overline{\nu}_{h,in}^2 - \frac{1}{2}p(k))_+) \le C(g)\big(\frac{\sigma_X}{2n} + \frac{\mathbb{E}(X_1^{5/2})}{\sigma_X^{3/2}n} + \frac{2\mathrm{Var}(\widehat{\sigma}_X)}{\sigma_X}\big).$$

□

Proof of Lemma 5.7.11. First we see that

$$\mathbb{E}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) = 2\chi\mathbb{E}((\sigma_Y/2 - \widehat{\sigma}_Y)_+\Delta_g(\widehat{k})n^{-1}) \le 2\chi\mathbb{E}((\sigma_Y/2 - \widehat{\sigma}_Y)_+).$$

On $\Omega_Y := \{|\sigma_Y - \widehat{\sigma}_Y| \le \sigma_Y/2\}$ we have $\sigma_Y/2 - \widehat{\sigma}_Y \le 0$. Therefore,

$$\mathbb{E}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \le 2\chi\mathbb{E}((\sigma_Y/2 - \widehat{\sigma}_Y)_+\mathbb{1}_{\Omega^c}) \le 2\chi\sqrt{\mathrm{Var}_{f_Y}^n(\widehat{\sigma}_Y)\mathbb{P}_{f_Y}(\Omega^c)}$$

by applying the Cauchy-Schwarz inequality. Next, by Markov's inequality, $\mathbb{P}[|\widehat{\sigma}_Y - \sigma_Y| \ge \sigma_Y/2] \le 4\mathrm{Var}(\widehat{\sigma}_Y)\sigma_Y^{-2}$ which implies $\mathbb{E}((\mathrm{pen}(\widehat{k}) - \widehat{\mathrm{pen}}(\widehat{k}))_+) \le 4\chi\mathrm{Var}(\widehat{\sigma}_Y)\sigma_Y^{-1}$. In analogy to the proof of Theorem 5.3.2 we get

$$\mathrm{Var}(\widehat{\sigma}_Y) \le \frac{\mathbb{E}(Y_1^2)}{n} + \mathbb{E}(U_1)^2\mathrm{Var}(\widehat{\sigma}_X).$$

□

Proof of Corollary 5.4.2. We discuss each case separately. We already assessed the variance term in the integral in Corollary 5.3.3. It remains to upper bound the variance of $\widehat{\sigma}_X$.

(I) Trivially, $\mathrm{Var}(n^{-1}\sum_{j=1}^n X_j) \le n^{-1}\mathbb{E}(X_1^2)$.

(B) Exploiting Lemma 5.7.3 for the identity mapping $h = \mathrm{id}$, we have

$$\mathrm{Var}(n^{-1}\sum_{j=1}^n X_j) \le n^{-2}\cdot 4n\mathbb{E}(X_1^2 b(X_1)) = 4n^{-1}\mathbb{E}(X_1^2 b(X_1)).$$

(F) Setting the function g in Lemma 5.7.5 as the identity mapping $g := \mathrm{id}$, we simply have

$$\mathrm{Var}(\sum_{j=1}^n X_j)^{1/2} \le n^{1/2}\sum_{k=0}^\infty \delta_2^X(k).$$

Combined with the results of Corollary 5.3.3, we derive our statement. □

Chapter 6

Perspectives

Our work has led to some questions which might be addressed in a future work.

Empirical process theory

In principle, a similar empirical process theory for locally stationary processes can be established under mixing conditions such as absolute regularity (β-mixing). This would be a generalization of the results found in Rio [1995] and Dedecker and Louhichi [2002]. As we have seen in Section 2.2.2, such a theory would impose additional moment conditions on $f(Z_i, \frac{i}{n})$. Contrary to that, our framework only requires second moments of $f(Z_i, \frac{i}{n})$. However, the entropy integral in our case becomes enlarged by a factor which increases with stronger dependence.

Moreover, the derivation of a bound for mixing coefficients in nearly all models considered by the community requires that the innovation process is continuous, which may not be suitable in some examples. Especially for linear processes, the bounds are quite hard to obtain and do not seem to be optimal. Also, there exist no "invariance rules" which directly allow us to transfer the mixing properties of X_i to $f(Z_i, \frac{i}{n})$, which incorporates *infinitely* many lags of X_i. In this regard, our theory from Chapter 2 to 4 substantially generalize the existing theory even in the stationary case.

Many of the more elaborated dependence concepts developed in for example Dedecker and Prieur [2007], Borovkova et al. [2001], Durieu and Tusche [2014], Berkes et al. [2009] are restricted (at least in their original formulation) to the discussion of the empirical distribution function or connected one-dimensional indexed function classes. Here, results for a broader range of functions could be interesting.

Statistical learning, neural networks

A natural extension of our work would be the proof of lower bounds (and thus minimax optimality) under specific structural assumptions on the process or its recurrence relation. The conditions for dependent data have not been formulated yet, even for common nonparametric methods.

It would also be interesting to consider a general ARCH-type model,

$$X_i = f_0(\mathbb{X}_{i-1}) + \sigma(\mathbb{X}_{i-1})\varepsilon_i, \qquad i = r+1, ..., n,$$

with an additional matrix-valued function $\sigma : \mathbb{R}^{dr} \to \mathbb{R}^{d\times d}$. We conjecture that in such models, similar convergence rates could be obtained under appropriate structural assumptions on $\sigma(\cdot)$.

On another note, it should be possible to derive more precise results when approximating the estimators by, for example, a stochastic gradient descent algorithm. Moreover, similar to the theory of boosting, we could hope for explicit or adaptive stopping rules.

Finally in this field, there are many more architectures to choose from. However, little is known about them from a mathematical point of view. How can we build a theory based on a fully connected, recurrent or convolutional network?

Survival analysis

We have discussed the multiplicative deconvolution problem in the field of survival analysis, estimating the survival function under multiplicative noise. Although we restricted ourselves to sufficiently smooth error densities, we should, in principle, be able to generalize our theory for arbitrary error types.

In a similar context, we would also like to incorporate the case of an unknown error. Then, we would have to perform an additional estimation step for the underlying error density. In another setting, Meister [2009], Comte and Lacour [2010] and Johannes [2009] recovered the typical nonparametric rates, minimax optimality and included adaptivity in their theory.

So far, our proposed estimator delivers an estimator that *approximates* a survival function. However, the estimator itself is not required to satisfy the exact conditions of a survival function. We have provided a heuristic in which we took this fact into account; in numerical evaluations our approach seems to have introduced more bias. In this regard, further investigations are needed for a mathematical sound groundwork.

We hope to answer these questions soon enough and reserve them for future discussions or inspire other researchers to investigate them.

Bibliography

R. M. Dudley. *Uniform central limit theorems*, volume 142 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, New York, second edition, 2014. ISBN 978-0-521-73841-5; 978-0-521-49884-5.

Evarist Giné and Richard Nickl. *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York, 2016. ISBN 978-1-107-04316-9. doi: 10.1017/CBO9781107337862. URL https://doi.org/10.1017/CBO9781107337862.

Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3. doi: 10.1007/978-1-4757-2545-2. URL https://doi.org/10.1007/978-1-4757-2545-2. With applications to statistics.

A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. ISBN 0-521-49603-9; 0-521-78450-6. doi: 10.1017/CBO9780511802256. URL https://doi.org/10.1017/CBO9780511802256.

Monroe D. Donsker. Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, 23:277–281, 1952. ISSN 0003-4851.

R. M. Dudley. Weak convergences of probabilities on nonseparable metric spaces and empirical measures on Euclidean spaces. *Illinois J. Math.*, 10:109–126, 1966. ISSN 0019-2082. URL http://projecteuclid.org/euclid.ijm/1256055206.

R. M. Dudley. Central limit theorems for empirical measures. *Ann. Probab.*, 6(6): 899–929 (1979), 1978. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(197812)6:6<899:CLTFEM>2.0.CO;2-O&origin=MSN.

David Pollard. A central limit theorem for empirical processes. *J. Austral. Math. Soc. Ser. A*, 33(2):235–248, 1982. ISSN 0263-6115.

Mina Ossiander. A central limit theorem under metric entropy with L_2 bracketing. *Ann. Probab.*, 15(3):897–919, 1987. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(198707)15:3<897:ACLTUM>2.0.CO;2-Z&origin=MSN.

Shlomo Levental. Uniform limit theorems for Harris recurrent Markov chains. *Probab. Theory Related Fields*, 80(1):101–118, 1988. ISSN 0178-8051. doi: 10.1007/BF00348754. URL `https://doi.org/10.1007/BF00348754`.

Jorge D. Samur. A regularity condition and a limit theorem for Harris ergodic Markov chains. *Stochastic Process. Appl.*, 111(2):207–235, 2004. ISSN 0304-4149. doi: 10.1016/j.spa.2004.02.005. URL `https://doi.org/10.1016/j.spa.2004.02.005`.

Richard S. Ellis and Aaron D. Wyner. Uniform large deviation property of the empirical process of a Markov chain. *Ann. Probab.*, 17(3):1147–1151, 1989. ISSN 0091-1798. URL `http://links.jstor.org/sici?sici=0091-1798(198907)17:3<1147:ULDPOT>2.0.CO;2-R&origin=MSN`.

Radosł aw Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008. doi: 10.1214/EJP.v13-521. URL `https://doi.org/10.1214/EJP.v13-521`.

RafałKulik, Philippe Soulier, and Olivier Wintenberger. The tail empirical process of regularly varying functions of geometrically ergodic Markov chains. *Stochastic Process. Appl.*, 129(11):4209–4238, 2019. ISSN 0304-4149. doi: 10.1016/j.spa.2018.11.014. URL `https://doi.org/10.1016/j.spa.2018.11.014`.

Hans Arnfinn Karlsen and Dag Tjø stheim. Nonparametric estimation in null recurrent time series. *Ann. Statist.*, 29(2):372–416, 2001. ISSN 0090-5364. doi: 10.1214/aos/1009210546. URL `https://doi.org/10.1214/aos/1009210546`.

Herold Dehling, Olivier Durieu, and Dalibor Volny. New techniques for empirical processes of dependent data. *Stochastic Process. Appl.*, 119(10):3699–3718, 2009. ISSN 0304-4149. doi: 10.1016/j.spa.2009.07.003. URL `https://doi.org/10.1016/j.spa.2009.07.003`.

Paul Doukhan and Michael H Neumann. Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Processes and their Applications*, 117(7):878–903, 2007.

Olivier Durieu and Marco Tusche. An empirical process central limit theorem for multidimensional dependent data. *J. Theoret. Probab.*, 27(1):249–277, 2014. ISSN 0894-9840. doi: 10.1007/s10959-012-0450-3. URL `https://doi.org/10.1007/s10959-012-0450-3`.

István Berkes, Siegfried Hörmann, and Johannes Schauer. Asymptotic results for the empirical process of stationary sequences. *Stochastic Process. Appl.*, 119(4):1298–1324, 2009. ISSN 0304-4149. doi: 10.1016/j.spa.2008.06.010. URL `https://doi.org/10.1016/j.spa.2008.06.010`.

Svetlana Borovkova, Robert Burton, and Herold Dehling. Limit theorems for functionals of mixing processes with applications to U-statistics and dimension estimation.

Trans. Amer. Math. Soc., 353(11):4261–4318, 2001. ISSN 0002-9947. doi: 10.1090/S0002-9947-01-02819-7. URL https://doi.org/10.1090/S0002-9947-01-02819-7.

J. Dedecker. An empirical central limit theorem for intermittent maps. *Probab. Theory Related Fields*, 148(1-2):177–195, 2010. ISSN 0178-8051. doi: 10.1007/s00440-009-0227-5. URL https://doi.org/10.1007/s00440-009-0227-5.

Paul Doukhan. *Mixing*, volume 85 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1994. ISBN 0-387-94214-9. doi: 10.1007/978-1-4612-2642-0. URL https://doi.org/10.1007/978-1-4612-2642-0. Properties and examples.

Emmanuel Rio. The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann. Probab.*, 23(3):1188–1203, 07 1995. doi: 10.1214/aop/1176988179. URL https://doi.org/10.1214/aop/1176988179.

Eckhard Liebscher. Strong convergence of sums of [alpha]-mixing random variables with applications to density estimation. *Stochastic Processes and their Applications*, 65 (1):69–80, 1996. URL https://EconPapers.repec.org/RePEc:eee:spapps:v:65:y:1996:i:1:p:69-80.

M. A. Arcones and B. Yu. Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.*, 7(1):47–71, 1994. ISSN 0894-9840. doi: 10.1007/BF02213360. URL https://doi.org/10.1007/BF02213360.

Bin Yu. Rates of convergence for empirical processes of stationary mixing sequences. *Ann. Probab.*, 22(1):94–116, 1994. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(199401)22:1<94:ROCFEP>2.0.CO;2-K&origin=MSN.

P. Doukhan, P. Massart, and E. Rio. Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 31(2):393–427, 1995. ISSN 0246-0203.

Emmanuel Rio. Processus empiriques absolument réguliers et entropie universelle. *Probab. Theory Related Fields*, 111(4):585–608, 1998. ISSN 0178-8051. doi: 10.1007/s004400050179. URL https://doi.org/10.1007/s004400050179.

Jérôme Dedecker and Clémentine Prieur. An empirical central limit theorem for dependent sequences. *Stochastic Process. Appl.*, 117(1):121–142, 2007. ISSN 0304-4149. doi: 10.1016/j.spa.2006.06.003. URL https://doi.org/10.1016/j.spa.2006.06.003.

Emmanuel Rio. *Asymptotic theory of weakly dependent random processes*, volume 80 of *Probability Theory and Stochastic Modelling*. Springer, Berlin, 2017. ISBN 978-3-662-54322-1; 978-3-662-54323-8. doi: 10.1007/978-3-662-54323-8. URL https://doi.org/10.1007/978-3-662-54323-8. Translated from the 2000 French edition [MR2117923].

Jérôme Dedecker and Sana Louhichi. Maximal inequalities and empirical central limit theorems. In *Empirical process techniques for dependent data*, pages 137–159. Birkhäuser Boston, Boston, MA, 2002.

Donald W. K. Andrews and David Pollard. An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review / Revue Internationale de Statistique*, 62(1):119–132, 1994. ISSN 03067734, 17515823. URL `http://www.jstor.org/stable/1403549`.

Lothar Heinrich. Bounds for the absolute regularity coefficient of a stationary renewal process. *Yokohama Math. J.*, 40(1):25–33, 1992. ISSN 0044-0523.

Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability.* Cambridge University Press, Cambridge, second edition, 2009. ISBN 978-0-521-73182-9. doi: 10.1017/CBO9780511626630. URL `https://doi.org/10.1017/CBO9780511626630`. With a prologue by Peter W. Glynn.

Tuan D. Pham and Lanh T. Tran. Some mixing properties of time series models. *Stochastic Process. Appl.*, 19(2):297–303, 1985. ISSN 0304-4149. doi: 10.1016/0304-4149(85)90031-6. URL `https://doi.org/10.1016/0304-4149(85)90031-6`.

Abdelkader Mokkadem. Mixing properties of ARMA processes. *Stochastic Process. Appl.*, 29(2):309–315, 1988. ISSN 0304-4149. doi: 10.1016/0304-4149(88)90045-2. URL `https://doi.org/10.1016/0304-4149(88)90045-2`.

Christian Francq and Jean-Michel Zakoïan. Mixing properties of a general class of GARCH(1,1) models without moment assumptions on the observed process. *Econometric Theory*, 22(5):815–834, 2006. ISSN 0266-4666. doi: 10.1017/S0266466606060373. URL `https://doi.org/10.1017/S0266466606060373`.

Wei Biao Wu. Nonlinear system theory: another look at dependence. *Proc. Natl. Acad. Sci. USA*, 102(40):14150–14154, 2005. ISSN 0027-8424. doi: 10.1073/pnas.0506715102. URL `https://doi.org/10.1073/pnas.0506715102`.

Wei Biao Wu, Weidong Liu, and Han Xiao. Probability and moment inequalities under dependence. *Statist. Sinica*, 23(3):1257–1272, 2013. ISSN 1017-0405.

Danna Zhang and Wei Biao Wu. Gaussian approximation for high dimensional time series. *Ann. Statist.*, 45(5):1895–1919, 2017. ISSN 0090-5364. doi: 10.1214/16-AOS1512. URL `https://doi.org/10.1214/16-AOS1512`.

Wei Biao Wu. Empirical processes of stationary sequences. *Statistica Sinica*, 18(1): 313–333, 2008. ISSN 10170405, 19968507. URL `http://www.jstor.org/stable/24308259`.

Ulrike Mayer, Henryk Zähle, and Zhou Zhou. Functional weak limit theorem for a local empirical process of non-stationary time series and its application. *Bernoulli*, 26 (3):1891 – 1911, 2020. doi: 10.3150/19-BEJ1174. URL `https://doi.org/10.3150/19-BEJ1174`.

Rainer Dahlhaus, Stefan Richter, and Wei Biao Wu. Towards a general theory for nonlinear locally stationary processes. *Bernoulli*, 25(2):1013–1044, 2019. ISSN 1350-7265. doi: 10.3150/17-bej1011. URL `https://doi.org/10.3150/17-bej1011`.

Rainer Dahlhaus and Wolfgang Polonik. Empirical spectral processes for locally stationary time series. *Bernoulli*, 15(1):1–39, 2009. ISSN 1350-7265. doi: 10.3150/08-BEJ137. URL `https://doi.org/10.3150/08-BEJ137`.

Wei Biao Wu. Asymptotic theory for stationary processes. *Stat. Interface*, 4(2):207–226, 2011. ISSN 1938-7989. doi: 10.4310/SII.2011.v4.n2.a15. URL `https://doi.org/10.4310/SII.2011.v4.n2.a15`.

Johannes Schmidt-Hieber. Nonparametric regression using deep neural networks with relu activation function, 2017.

Zaiyong Tang and Paul A. Fishwick. Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing*, 5(4):374–385, 1993. doi: 10.1287/ijoc.5.4.374. URL `https://doi.org/10.1287/ijoc.5.4.374`.

Douglas Kline. *Methods for Multi-Step Time Series Forecasting with Neural Networks*, pages 226–250. IGI Global, 01 2004. doi: 10.4018/978-1-59140-176-6.ch012.

G. Peter Zhang. *Neural Networks for Time-Series Forecasting*, pages 461–477. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. ISBN 978-3-540-92910-9. doi: 10.1007/978-3-540-92910-9_14. URL `https://doi.org/10.1007/978-3-540-92910-9_14`.

Peter Bühlmann and Sara Van De Geer. *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media, 2011.

Gilles Blanchard, Olivier Bousquet, and Pascal Massart. Statistical performance of support vector machines. *Ann. Statist.*, 36(2):489–531, 2008. ISSN 0090-5364. doi: 10.1214/009053607000000839. URL `https://doi.org/10.1214/009053607000000839`.

Sergio Brenner Miguel, Fabienne Comte, and Jan Johannes. Spectral cut-off regularisation for density estimation under multiplicative measurement errors. *arXiv preprint arXiv:2009.10547*, 2020.

Sergio Brenner Miguel. Anisotropic spectral cut-off estimation under multiplicative measurement errors. *arXiv preprint arXiv:2107.02120*, 2021.

Denis Belomestny and Alexander Goldenshluger. Nonparametric density estimation from observations with multiplicative measurement errors. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 36–67. Institut Henri Poincaré, 2020.

Yehuda Vardi. Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika*, 76(4):751–761, 1989.

Y Vardi and Cun-Hui Zhang. Large sample study of empirical distributions in a random-multiplicative censoring model. *The Annals of Statistics*, pages 1022–1039, 1992.

Bert Van Es, Chris AJ Klaassen, and Karin Oudshoorn. Survival analysis under cross-sectional sampling: length bias and multiplicative censoring. *Journal of Statistical Planning and Inference*, 91(2):295–312, 2000.

Masoud Asgharian and David B Wolfson. Asymptotic behavior of the unconditional npmle of the length-biased survivor function from right censored prevalent cohort data. *The Annals of Statistics*, 33(5):2109–2131, 2005.

Kim E Andersen and Martin B Hansen. Multiplicative censoring: density estimation by a series expansion approach. *Journal of Statistical Planning and Inference*, 98(1-2): 137–155, 2001.

Elodie Brunel, Fabienne Comte, and Valentine Genon-Catalot. Nonparametric density and survival function estimation in the multiplicative censoring model. *Test*, 25(3): 570–590, 2016.

Fabienne Comte and Charlotte Dion. Nonparametric estimation in a multiplicative censoring model with symmetric noise. *Journal of Nonparametric Statistics*, 28(4): 768–801, 2016.

Denis Belomestny, Fabienne Comte, and Valentine Genon-Catalot. Nonparametric Laguerre estimation in the multiplicative censoring model. *Electronic Journal of Statistics*, 10(2):3114–3152, 2016.

Alexander Meister. Density deconvolution. In *Deconvolution Problems in Nonparametric Statistics*, pages 5–105. Springer, 2009.

Heinz W. Engl, Martin Hanke-Bourgeois, and Andreas Neubauer. *Regularization of inverse problems*. Kluwer Acad. Publ., 2000.

F. Comte, J. Dedecker, and M. L. Taupin. Adaptive density estimation for general ARCH models. *Econometric Theory*, 24(6):1628–1662, 2008. ISSN 0266-4666. doi: 10.1017/S026646660808064X. URL `https://doi.org/10.1017/S026646660808064X`.

Nathawut Phandoidaen and Stefan Richter. Empirical process theory for locally stationary processes. *Bernoulli*, 28(1):453 – 480, 2022. doi: 10.3150/21-BEJ1351. URL `https://doi.org/10.3150/21-BEJ1351`.

Vivek S. Borkar. White-noise representations in stochastic realization theory. *SIAM J. Control Optim.*, 31(5):1093–1102, 1993. ISSN 0363-0129. doi: 10.1137/0331050. URL `https://doi.org/10.1137/0331050`.

M. Rosenblatt. A comment on a conjecture of n. wiener. *Statistics & Probability Letters*, 79:347–348, 02 2009. doi: 10.1016/j.spl.2008.09.001.

David Pollard. *Convergence of stochastic processes*. Springer Science & Business Media, 2012.

Lionel Truquet. A perturbation analysis of Markov chains models with time-varying parameters. *Bernoulli*, 26(4):2876–2906, 2020. ISSN 1350-7265. doi: 10.3150/20-BEJ1210. URL `https://doi.org/10.3150/20-BEJ1210`.

Degui Li, Dag Tjø stheim, and Jiti Gao. Estimation in nonlinear regression with Harris recurrent Markov chains. *Ann. Statist.*, 44(5):1957–1987, 2016. ISSN 0090-5364. doi: 10.1214/15-AOS1379. URL `https://doi.org/10.1214/15-AOS1379`.

Bruce E. Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, 24(3):726–748, 2008. ISSN 0266-4666. doi: 10.1017/S0266466608080304. URL `https://doi.org/10.1017/S0266466608080304`.

Michael Vogt. Nonparametric regression for locally stationary time series. *Ann. Statist.*, 40(5):2601–2633, 2012. ISSN 0090-5364. doi: 10.1214/12-AOS1043. URL `https://doi.org/10.1214/12-AOS1043`.

Kenneth S Alexander. Probability inequalities for empirical processes and a law of the iterated logarithm. *The Annals of Probability*, pages 1041–1067, 1984.

Iosif Pinelis. Optimum bounds for the distributions of martingales in Banach spaces. *Ann. Probab.*, 22(4):1679–1706, 1994. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(199410)22:4<1679:OBFTDO>2.0.CO;2-2&origin=MSN.

David A. Freedman. On tail probabilities for martingales. *Ann. Probab.*, 3(1):100–118, 02 1975. doi: 10.1214/aop/1176996452. URL `https://doi.org/10.1214/aop/1176996452`.

Yoichi Nishiyama et al. Weak convergence of some classes of martingales with jumps. *The Annals of Probability*, 28(2):685–712, 2000.

Michael G. Akritas and Ingrid Van Keilegom. Non-parametric estimation of the residual distribution. *Scand. J. Statist.*, 28(3):549–567, 2001. ISSN 0303-6898. doi: 10.1111/1467-9469.00254. URL `https://doi.org/10.1111/1467-9469.00254`.

Emmanuel Rio. Inequalities and limit theorems for weakly dependent sequences. Lecture, September 2013. URL `https://cel.archives-ouvertes.fr/cel-00867106`.

Richard C. Bradley. Basic properties of strong mixing conditions. a survey and some open questions. *Probab. Surveys*, 2:107–144, 2005. doi: 10.1214/154957805100000104. URL `https://doi.org/10.1214/154957805100000104`.

Piotr Fryzlewicz and Suhasini Subba Rao. Mixing properties of arch and time-varying arch processes. *Bernoulli*, 17(1):320–346, 02 2011. doi: 10.3150/10-BEJ270. URL `https://doi.org/10.3150/10-BEJ270`.

Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

S. Richter. *Statistisches und maschinelles Lernen: Gängige Verfahren im Überblick*. Springer Berlin Heidelberg, 2019. ISBN 9783662593547. URL https://books.google.de/books?id=XSmsDwAAQBAJ.

Kevin P Murphy. *Probabilistic machine learning: an introduction*. MIT press, 2022.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Advances in neural information processing systems*, pages 5767–5777, 2017.

Árpád Baricz. Mills' ratio: reciprocal convexity and functional inequalities. *Acta Univ. Sapientiae Math.*, 4(1):26–35, 2012. ISSN 1844-6094.

Emmanuel Rio. Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.*, 22(1):146–163, 2009. ISSN 0894-9840. doi: 10.1007/s10959-008-0155-9. URL https://doi.org/10.1007/s10959-008-0155-9.

Richard B Paris and David Kaminski. *Asymptotics and mellin-barnes integrals*, volume 85. Cambridge University Press, 2001.

Jianqing Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, pages 1257–1272, 1991.

D. P. Gaver and P. A. W. Lewis. First-order autoregressive gamma sequences and point processes. *Adv. in Appl. Probab.*, 12(3):727–745, 1980. ISSN 0001-8678. doi: 10.2307/1426429. URL https://doi.org/10.2307/1426429.

Peter J. Brockwell and Richard A. Davis. *Introduction to time series and forecasting*. Springer Texts in Statistics. Springer, [Cham], third edition, 2016. ISBN 978-3-319-29852-8; 978-3-319-29854-2. doi: 10.1007/978-3-319-29854-2. URL https://doi.org/10.1007/978-3-319-29854-2.

Yuan Shih Chow and Henry Teicher. *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1988. ISBN 0-387-96695-1. doi: 10.1007/978-1-4684-0504-0. URL https://doi.org/10.1007/978-1-4684-0504-0. Independence, interchangeability, martingales.

M. Talagrand. New concentration inequalities in product spaces. *Inventiones mathematicae*, 126:505–563, 1996.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *The Annals of Probability*, 33(3):1060–1077, 2005.

Alexandre B. Tsybakov. *Introduction to nonparametric estimation*. Springer Publishing Company, Incorporated, 2008.

Nicolas Asin and Jan Johannes. Adaptive nonparametric estimation in the presence of dependence. *Journal of Nonparametric Statistics*, 29(4):694–730, 2017.

Gabrielle Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107(4):467–492, 1997. ISSN 0178-8051. doi: 10.1007/s004400050094. URL https://doi.org/10.1007/s004400050094.

Fabienne Comte and Claire Lacour. Pointwise deconvolution with unknown error distribution. *Comptes Rendus Mathematique*, 348(5):323–326, 2010. ISSN 1631-073X. doi: https://doi.org/10.1016/j.crma.2010.02.012. URL https://www.sciencedirect.com/science/article/pii/S1631073X10000579.

Jan Johannes. Deconvolution with unknown error distribution. *The Annals of Statistics*, 37(5A):2301 – 2323, 2009. doi: 10.1214/08-AOS652. URL https://doi.org/10.1214/08-AOS652.