Econometric Theory, **42**, 2026, 787–836.
doi:10.1017/S0266466625100054

HIGHER-ORDER APPROXIMATION FOR UNCERTAINTY QUANTIFICATION IN TIME-SERIES ANALYSIS

ANNIKA BETKEN
University of Twente

MARIE-CHRISTINE DÜKER
FAU Erlangen-Nürnberg

For time series with high temporal correlation, the empirical process converges rather slowly to its limiting distribution. Many statistics in change-point analysis, goodness-of-fit testing, and uncertainty quantification admit a representation as functionals of the empirical process and therefore inherit its slow convergence. As a result, inference based on the asymptotic distribution of those quantities is significantly affected by relatively small sample sizes. We assess the quality of higher-order approximations (HOAs) of the empirical process by deriving the asymptotic distribution of the corresponding error terms. Based on the limiting distribution of the higher-order terms, we propose a novel approach to calculate confidence intervals for statistical quantities such as the median. In a simulation study, we compare coverage rates and lengths of these confidence intervals with those based on the asymptotic distribution of the empirical process and highlight some benefits of HOAs of the empirical process.

1. INTRODUCTION

Let $X_n, n = 1, \dots, N$, be a time-series stemming from a stationary stochastic process $X_n, n \in \mathbb{N}$, with marginal distribution function F, such that $F(x) = \mathrm{P}(X_n \leqslant x)$ for all $n \in \mathbb{N}$. We study the empirical distribution function $F_N(x) := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{X_n \leqslant x\}}$. The rate of convergence, i.e., the increase of the sequence a_N, $N \in \mathbb{N}$, which ensures weak convergence of the empirical process

$$a_N^{-1} N(F_N(x) - F(x)) \tag{1.1}$$

We thank the editor, co-editor, and referees for their careful reading of the manuscript and their thoughtful comments, which led to a significant improvement of the article. Annika Betken gratefully acknowledges financial support from the Dutch Research Council (NWO) through VENI grant 212.164. Marie-Christine Düker gratefully acknowledges financial support from the National Science Foundation under grants 1934985, 1940124, 1940276, and 2114143. This research was conducted with support from the Cornell University Center for Advanced Computing, which receives funding from Cornell University, the National Science Foundation, and members of its Partner Program. Address correspondence to Annika Betken, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands, a.betken@utwente.nl. Annika Betken, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Enschede, The Netherlands, e-mail: a.betken@utwente.nl

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

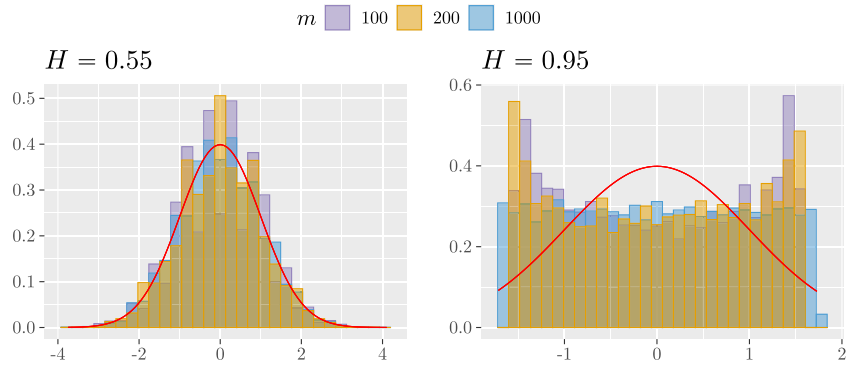


FIGURE 1. The empirical distribution of the centered and standardized empirical distribution $F_N(x)$ evaluated at zero ($x = 0$) under the assumption of Gaussian long-range dependent data with different Hurst parameters and for sample sizes $m = 100, 200, 1000$. The red line depicts the standard Gaussian density function.

to a nondegenerate limit, crucially depends on the behavior of the process' autocorrelation function $\gamma(k) := \mathrm{Cov}(X_1, X_{k+1})$. For short-range dependent time series, i.e., for stochastic processes with summable autocorrelations, $a_N = \sqrt{N}$. In contrast, for long-range dependent time series, i.e., for $\gamma(k) = k^{-(2-2H)}L(k)$ with L a slowly varying function and $H \in (1/2, 1)$ the so-called Hurst parameter, we have $a_N = N^H L^{\frac{1}{2}}(N)$. In fact, under long-range dependence (LRD), the distribution of the empirical process converges much slower to its limit than under short-range dependence.

To illustrate the slow convergence of the empirical process under strong temporal correlation, we would like to draw the reader's attention to Figure 1. The figure depicts the asymptotic behavior of the centered and standardized empirical distribution $F_N(x)$ evaluated at $x = 0$ for different sample sizes. The underlying process is assumed to be fractional Gaussian noise with Hurst parameter H. In this case, the sequence a_N, $N \in \mathbb{N}$, in (1.1) can be explicitly calculated as $a_N = \varphi(0)N^H$ with φ denoting the standard Gaussian density. The quantity $\frac{1}{\varphi(0)}N^{1-H}(F_N(0) - F(0))$ is computed independently 10,000 times for different sample sizes N and the resulting values are summarized in the histograms in Figure 1. Due to Dehling and Taqqu (1989, Thm. 1.1), the quantity is expected to converge to a standard Gaussian random variable. Therefore, the histograms in Figure 1 are expected to approach the standard Gaussian density function (depicted in red). The convergence rate depends on the value of the Hurst parameter in that a small value ($H = 0.55$) results in relatively fast convergence while a large value ($H = 0.95$), and implied stronger temporal correlation, results in much slower convergence. This phenomenon is specific to long-range dependent time series and the focus of this work.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

The empirical process serves as a powerful tool for characterizing the asymptotic behavior of a variety of test statistics used in change-point analysis and goodness-of-fit testing (Wilcoxon, Kolmogorov–Smirnov, and Cramèr–von Mises statistics); see Beran (1992); Betken (2016, 2017); Dehling et al. (2013); Tewes (2018). When testing the hypothesis of stationarity against the alternative hypothesis of a structural change in a time series, the phenomenon illustrated in Figure 1 results in a high number of false positives; see Dehling et al. (2013).

Against this background, the contribution of this is twofold: On the one hand, we address a statistical issue arising in the calculation of confidence intervals under strong temporal correlation. On the other hand, we push forward the theoretical investigation of the empirical process by proving a novel limit theorem. More precisely:

- We study the construction of confidence intervals for the marginal distribution of stationary time-series data and confidence intervals for its quantiles in long-range dependent time series. We propose a novel approach to calculate confidence intervals based on a higher-order approximation (HOA) of the empirical distribution function. Under LRD, an asymptotic expansion of the empirical distribution that is similar in spirit to a Taylor expansion can be derived. This expansion can be used to obtain HOAs of certain statistical functionals of the empirical process.
- We establish the theoretical validity of our method for statistics that can be considered as functionals of the empirical process. For statistical applications beyond the construction of confidence intervals, e.g., change-point and goodness-of-fit tests, uniform convergence of the one-parameter empirical process (1.1) does not suffice in order to derive limit distributions of corresponding statistics. These typically require consideration of the two-parameter (or sequential) empirical process

  $$a_N^{-1} \lfloor Nt \rfloor (F_{\lfloor Nt \rfloor}(x) - F(x)), \ t \in [0,1], \ x \in \mathbb{R}.$$

  We derive the asymptotic distribution of HOAs of the sequential empirical process by proposing a new chaining technique.

Constructing confidence intervals for unknown quantities in time series is a problem of substantial interest in statistics. In the statistical literature, the main focus has been on approximating the limiting distribution through finite sample procedures like subsampling and bootstrapping; see Bühlmann (2002); Huang and Shao (2016); Kim et al. (2015); Nordman et al. (2013); Shao (2010). From an entirely theoretical perspective, Youndjé and Vieu (2006) investigate consistency properties of kernel-type estimators of quantiles under LRD. The interest in confidence intervals is also due to their relevance for uncertainty quantification in other sciences where they are used in a variety of fields, including climate science, economics, finance, industrial engineering, and machine learning; see Massah and Kantz (2016); Fang et al. (2018); Hoga (2019); Purwanto and Sudargini (2021).

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Empirical process theory became one of the major themes in the historical progress of nonparametric statistics; see Donsker (1952); Dudley (1978); Doukhan and Surgailis (1998); Shorack and Wellner (2009); Wellner and van der Vaart (2013). The applications are manifold, especially since many statistics have a representation as functionals of the empirical process, such that statistical inference can be based on the properties of the empirical process itself. In the empirical sciences, confidence intervals for unknown parameters or critical values for hypothesis tests are derived from the distributional properties of the empirical process.

For stationary Gaussian processes, Koul and Surgailis (2002) derived the asymptotic distribution of higher-order terms of the empirical process. We extend their results substantially by considering the sequential empirical process and by allowing the underlying time series to be driven by subordinated Gaussian processes. Subordination extends the model's flexibility by allowing for a large class of marginal distributions. Furthermore, we are the first to propose a utilization of HOAs of the empirical process for the calculation of confidence intervals which are robust to high temporal correlation in time-series data.

Although long-range dependent processes are a popular modeling tool in a variety of domains (Rust et al. 2010; Weron 2002), the construction of confidence intervals under has not gotten much attention. We provide an empirical study comparing confidence intervals derived from the asymptotic distribution of the empirical process to confidence intervals based on HOAs of the empirical process.

For the population mean, Hall et al. (1998) propose a sampling window method to set confidence intervals under LRD. Nordman et al. (2007) consider the empirical likelihood for confidence intervals. For mean functions, Bagchi et al. (2016) study a monotone function plus noise model with potential in the noise term and derive confidence intervals for the monotone functions. In contrast, we deal with a different, rank-based class of statistics.

The literature review, as well as our motivation illustrated in Figure 1, show the strong influence of high temporal correlation on the performance of statistics derived from the empirical process. In this article, we aim to address this issue by introducing a procedure based on HOAs of the empirical process to construct confidence intervals for statistics of long-range dependent time series robust to high temporal correlation. Our theoretical contribution is of independent interest and potentially has further applications in change-point analysis and goodness-of-fit testing. Furthermore, a reduction to a limit theorem in the short-range dependent regime allows an application of established resampling procedures such as the moving block bootstrap which has been proved to be invalid under LRD; see Lahiri (1993).

The rest of th is organized as follows: In Section 2, we introduce the considered setting in all details. Section 3 motivates the consideration of HOAs of the empirical process. Section 4 focuses on theoretical contributions which manifest the formal validity of the proposed method. In Section 5, we discuss how to calculate confidence intervals based on the asymptotic distribution of the empirical

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

process and propose an alternative approach based on HOAs. The numerical study in Section 6 provides a comparison between the two methods. We conclude with Section 7. Proofs of the theoretical results can be found in Appendices A–D.

2. PRELIMINARIES

While Section 1 provides insight into the motivation for considering HOAs of statistics, we introduce here model assumptions which allow for this type of approximations (Section 2.1) and give some technical details necessary for our analysis (Section 2.2).

2.1. Setting

For future reference, we subsume assumptions on the data-generating process under the following model specification.

Model 2.1. Let X_n, $n \in \mathbb{N}$, be a subordinated Gaussian process, i.e., $X_n = G(\xi_n)$ for some measurable function $G : \mathbb{R} \to \mathbb{R}$ and with ξ_n, $n \in \mathbb{N}$, denoting a (standardized) long-range dependent Gaussian process, i.e., $\mathrm{E}(\xi_n) = 0$, $\mathrm{Var}(\xi_n) = 1$, and

$$\gamma(k) = \mathrm{Cov}(\xi_1, \xi_{k+1}) = \mathrm{E}(\xi_1 \xi_{k+1}) = k^{-D} L(k), \tag{2.1}$$

where $D \in (0, 1)$ (the so-called LRD parameter) and L a slowly varying function.

Relation (2.1) corresponds to one of multiple different ways to define LRD. A more general definition characterizes long-range dependent time series by the nonsummability of the absolute values of its autocovariance function; see Pipiras and Taqqu (2017, eqn. (2.1.6)). In fact, (2.1) implies that the series of the autocovariances diverges. We refer to Pipiras and Taqqu (2017, Chap. 2.1) for a detailed representation of different ways to define and their relations to each other.

For any particular distribution function F, an appropriate choice of the transformation G yields subordinated Gaussian processes with marginal distribution F. Moreover, there exist algorithms for generating Gaussian processes that, after suitable transformation, yield subordinated Gaussian processes with marginal distribution F and a predefined covariance structure; see Pipiras and Taqqu (2017).

The following example presents a process which satisfies Model 2.1.

Example 2.2 (Pipiras and Taqqu (2017, Def. 2.8.3)). Let $B_H(t), t \in \mathbb{R}$, be a fractional Brownian motion. Then, the process $\xi_H(k), k \in \mathbb{Z}$, defined by

$$\xi_H(k) := B_H(k+1) - B_H(k)$$

is called fractional Gaussian noise with Hurst parameter H.

2.2. Gaussian Subordination

In the study of functionals of Gaussian processes, Hermite polynomials play a fundamental role. In particular, they form a basis for the space of finite-variance

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

functions of Gaussian random variables. Since they are an inevitable tool in our analysis, we provide a short review.

Let $L^2(\mathbb{R}, \varphi(x)dx)$ be the space of functions which are square-integrable with respect to the Gaussian measure (here denoted by $\varphi(x)dx$). For $g \in L^2(\mathbb{R}, \varphi(x)dx)$ and ξ_n, $n \in \mathbb{N}$, we call the sequence $X_n = g(\xi_n)$, $n \in \mathbb{N}$, a subordinated Gaussian sequence.

A collection of orthogonal elements in $L^2(\mathbb{R}, \varphi(x)dx)$ is given by the sequence of Hermite polynomials; see Pipiras and Taqqu (2017, Prop. 5.1.3).

DEFINITION 2.3. *For $n \geqslant 0$, the Hermite polynomial of order n is defined by*

$$H_n(x) = (-1)^n \mathrm{e}^{\frac{1}{2}x^2} \frac{d^n}{dx^n} \mathrm{e}^{-\frac{1}{2}x^2}, \ x \in \mathbb{R}.$$

The Hermite polynomials form an orthogonal basis of $L^2(\mathbb{R}, \varphi(x)dx)$. As a result, every $g \in L^2(\mathbb{R}, \varphi(x)dx)$ has an expansion in Hermite polynomials, i.e., for $g \in L^2(\mathbb{R}, \varphi(x)dx)$ and ξ standard normally distributed, we have

$$g(\xi) = \sum_{r=0}^{\infty} \frac{J_r(g)}{r!} H_r(\xi), \quad J_r(g) = \mathrm{E}\, g(\xi) H_r(\xi), \tag{2.2}$$

where $J_r(g), r \geqslant 0$, are the so-called *Hermite coefficients*.

Given the Hermite expansion (2.2), it is possible to characterize the dependence structure of subordinated Gaussian time series $g(\xi_n)$, $n \in \mathbb{N}$. In fact, it holds that

$$\mathrm{Cov}(g(\xi_1), g(\xi_{k+1})) = \sum_{r=1}^{\infty} \frac{J_r^2(g)}{r!} \gamma^r(k), \tag{2.3}$$

where γ denotes the autocovariance function of ξ_n, $n \in \mathbb{N}$; see Pipiras and Taqqu (2017, Prop. 5.1.4). Under the assumption that, as k tends to ∞, $\gamma(k)$ converges to 0 with a certain rate, the asymptotically dominating term in the series (2.3) is the summand corresponding to the smallest integer r for which the Hermite coefficient $J_r(g)$ is nonzero. This index, which decisively depends on g, is called *Hermite rank*.

DEFINITION 2.4 (Pipiras and Taqqu (2017, Def. 5.2.1)). *Let $g \in L^2(\mathbb{R}, \varphi(x)dx)$ with $\mathrm{E}\, g(\xi) = 0$ for standard normally distributed X and let $J_r(g)$, $r \geqslant 0$, be the Hermite coefficients in the Hermite expansion of g. The smallest index $k \geqslant 1$ for which $J_k(g) \neq 0$ is called the Hermite rank of g, i.e.,*

$$r := \min\{k \geqslant 1 : J_k(g) \neq 0\}.$$

3. HOA

We utilize our model assumptions and give details on a characterization of the empirical process as a sum of first- and higher-order terms.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Given time-series data $X_1, \ldots, X_N$ stemming from a subordinated Gaussian process X_n, $n \in \mathbb{N}$, according to Model 2.1 and with marginal distribution function F, we are interested in characterizing HOAs of the sequential empirical process

$$e_N(t,x) := \sum_{n=1}^{\lfloor Nt \rfloor} \left(\mathbb{1}_{\{X_n \leqslant x\}} - F(x) \right), \; t \in [0,1], \; x \in \mathbb{R}. \tag{3.1}$$

HOAs can be derived through the Hermite expansion

$$\mathbb{1}_{\{X_n \leqslant x\}} - F(x) = \sum_{l=r}^{\infty} \frac{\widetilde{c}_l(x)}{l!} H_l(\xi_n),$$

where $\widetilde{c}_l(x) = \mathrm{E}\left(\mathbb{1}_{\{G(\xi_0) \leqslant x\}} H_l(\xi_0) \right)$ and where r denotes the corresponding Hermite rank

$$r := \min_{x \in \mathbb{R}} r(x) \quad \text{with} \quad r(x) := \min\{q \geqslant 1 | \widetilde{c}_q(x) \neq 0\}.$$

Dehling and Taqqu (1989) show that the first summand of this expansion determines the asymptotic distribution of the empirical process through the reduction principle

$$\frac{1}{d_{N,r}} \sum_{n=1}^{N} \left(\mathbb{1}_{\{X_n \leqslant x\}} - F(x) \right) = \frac{\widetilde{c}_r(x)}{r!} \frac{1}{d_{N,r}} \sum_{n=1}^{N} H_r(\xi_n) + o_P(1), \tag{3.2}$$

where $d_{N,r}^2 = \mathrm{Var}\left(\sum_{n=1}^{N} H_r(\xi_n) \right)$.

In order to study higher-order terms, we utilize the following observation:

$$\sum_{n \in \mathbb{N}} |\mathrm{Cov}(H_l(\xi_1), H_l(\xi_{n+1})| = l! \sum_{n \in \mathbb{N}} |\gamma(n)|^l \begin{cases} = \infty, & lD < 1, \\ < \infty, & lD > 1; \end{cases} \tag{3.3}$$

see Pipiras and Taqqu (2017, eqn. (5.1.1)) for the first equality in (3.3). Then, distinguishing the two cases in (3.3), the last relation is a consequence of (2.1) and the assumption that $\frac{1}{D} \notin \mathbb{N}$.

The convergence behavior of the partial sums of autocovariances provides another way of distinguishing short- and long-range dependence. While convergence is associated with short-range dependence, divergence indicates LRD.

As a result, the sequence $H_l(\xi_n)$, $n \in \mathbb{N}$, can be considered as long-range dependent when $lD < 1$, while short-range dependent when $lD > 1$. Moreover, the following holds

$$\begin{aligned} &\frac{\widetilde{c}_l(x)}{l!} \sum_{n=1}^{N} H_l(\xi_n) = \mathcal{O}_P(N^{-\frac{Dl}{2}+1} L^{\frac{l}{2}}(N)) \text{ for } l < \frac{1}{D}, \text{while} \\ &\frac{\widetilde{c}_l(x)}{l!} \sum_{n=1}^{N} H_l(\xi_n) = \mathcal{O}_P(\sqrt{N}) \text{ for } l > \frac{1}{D}, \end{aligned} \tag{3.4}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

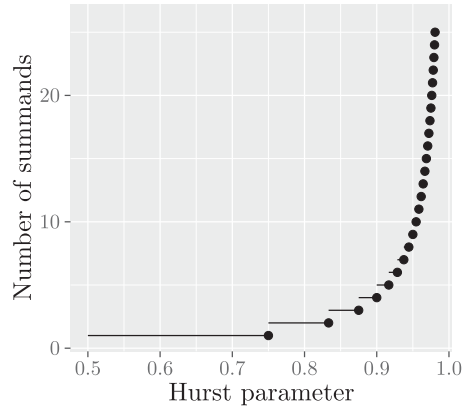


FIGURE 2. Number of summands in the "lower-order term" given that the Hermite rank $r = 1$.

where we refer to Beran et al. (2016, Eqn. (4.20)) for the first relation in (3.4). Note also that the memory parameter D corresponds to the Hurst parameter H through the relation $H = 1 - \frac{D}{2}$.

Motivated by the behavior of the series over the autocovariances in (3.3) and the different convergence rates in (3.4), we consider the separation

$$\sum_{l=r}^{\infty} \frac{\widetilde{c}_l(x)}{l!} H_l(\xi_n) = L_n(x) + S_n(x)$$

with

$$L_n(x) = \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{\widetilde{c}_l(x)}{l!} H_l(\xi_n) \quad \text{and} \quad S_n(x) = \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{\widetilde{c}_l(x)}{l!} H_l(\xi_n), \tag{3.5}$$

where for some $x \in \mathbb{R}$, $\lfloor x \rfloor$ and $\lceil x \rceil$ map x to the greatest integer less than or equal and the smallest integer greater than or equal to x. Based on (3.3), the series over the autocovariances of $L_n(x)$ diverges, while $S_n(x)$ has an absolutely summable autocovariance function. We refer to $L_n(x)$ in (3.5) as "lower-order term" and to $S_n(x)$ as "higher-order term."

For the empirical process (3.1), HOAs result from

$$\frac{1}{N} e_N(t,x) = \frac{1}{N} \sum_{n=1}^{\lfloor Nt \rfloor} L_n(x) + \frac{1}{N} \sum_{n=1}^{\lfloor Nt \rfloor} S_n(x).$$

Based on the previous considerations, the two summands are expected to converge at different rates. For our purpose, we aim at proving the convergence of $\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} S_n(x)$ parameterized in t and x.

To illustrate the observations made in this section, Figure 2 depicts the Hurst parameter $H \in (\frac{1}{2}, 1)$ and the corresponding number of summands $\lfloor \frac{1}{D} \rfloor$,

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

$D = 2 - 2H$, which contribute to the long-range dependent part $L_n(x)$ in (3.5). Note that the number of summands contributing to the lower-order term increases exponentially with the value of the Hurst parameter, while the interval length, i.e., the length of the subintervals of $(\frac{1}{2}, 1)$ which correspond to a certain number of summands, decreases.

4. MAIN RESULT

In this section, we present our main technical contributions. Our main result is stated in Section 4.1, followed by a layout of the proof ideas in Section 4.2.

4.1. Statement

We establish a limit theorem for the higher-order term in the decomposition of the sequential empirical process in two parameters. For this, recall that

$$\frac{\lfloor Nt \rfloor}{\sqrt{N}} \left(F_{\lfloor Nt \rfloor}(x) - F(x) \right) = \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} N^{-\frac{lD}{2}} L^{\frac{1}{2}}(N) Z_N^{(l)}(t,x) + \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} S_n(x)$$

with $Z_N^{(l)}(t,x) = N^{\frac{lD}{2}-1} L^{-\frac{l}{2}}(N) \sum_{n=1}^{\lfloor Nt \rfloor} \frac{\widetilde{c}_l(x)}{l!} H_l(\xi_n)$. According to Pipiras and Taqqu (2017, Thm. 5.3.1), if suitably standardized, each of the first $\lfloor \frac{1}{D} \rfloor - r + 1$ summands converges to a Hermite process of order l. More precisely, it holds that

$$Z_N^{(l)}(t,x) = N^{\frac{lD}{2}-1} L^{-\frac{1}{2}}(N) \sum_{n=1}^{\lfloor Nt \rfloor} \frac{\widetilde{c}_l(x)}{l!} H_l(\xi_n) \xrightarrow{\mathcal{D}} \frac{\widetilde{c}_l(x)}{l!} \beta_{l,H} Z_H^{(l)}(t)$$

in $D([-\infty, \infty] \times [0,1])$, where $\beta_{l,H}$ is a constant and $Z_H^{(l)}(t)$, $t \in [0,1]$, a Hermite process of order l with self-similarity parameter $H = 1 - \frac{lD}{2}$. The limit of the higher-order term $\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} S_n(x)$ is characterized by the following theorem.

THEOREM 4.1. *Suppose $X_n, n \in \mathbb{N}$, satisfies Model 2.1 and X_n has a strictly monotone, continuous distribution function F and $\frac{1}{D} \notin \mathbb{N}$. Then, as $N \to \infty$,*

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} S_n(x) \xrightarrow{\mathcal{D}} S(x,t)$$

in $D([-\infty, \infty] \times [0,1])$, where $S(x,t)$ is a mean zero Gaussian process with cross-covariances

$$\operatorname{Cov}(S(x,t), S(y,u)) = \min\{t,u\} \sum_{n \in \mathbb{Z}} \operatorname{Cov}(S_0(x), S_n(y)). \tag{4.1}$$

The proof of Theorem 4.1 can be found in Appendix A.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Remark 4.2. Note that we exclude the case $\frac{1}{D} \in \mathbb{N}$. That excludes in particular the case $D = 1$, that is when the underlying time series is short-range dependent. Therefore, our result as it is stated cannot recover existing results for short-range dependent time series. Under short-range dependence, the empirical process is known to converge to the so-called Kiefer–Müller process; see Müller (1970); Kiefer (1972).

4.2. Proof

While the detailed proof of Theorem 4.1 is given in Appendices A–C, we aim here to provide a roadmap of our proofs and to emphasize some of the main technical challenges.

For a proof of convergence in distribution as stated in Theorem 4.1, convergence of the finite-dimensional distributions and tightness are being established; see Sections A.1 and A.2, respectively. While proving convergence of the finite-dimensional distributions can be considered straightforward, the main technical challenges arise in the proof of tightness. These challenges are the subject of this section.

The partial sum of the higher-order terms in (3.5)

$$\widetilde{m}_N(x,t) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} S_n(x) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{\widetilde{c}_l(x)}{l!} H_l(\xi_n) \tag{4.2}$$

is a stochastic process in two parameters, which is one reason why proving tightness becomes particularly challenging. Another challenge results from the structure of the higher-order terms. In contrast to the empirical process, the higher-order terms are no longer bounded. While the transformed variables $H_l(\xi_n), n \in \mathbb{N}$, for $l \geqslant \lceil \frac{1}{D} \rceil$ are short-range dependent, the underlying process $\xi_n, n \in \mathbb{N}$, is still long-range dependent with nonsummable autocovariance function. The dependence on the memory parameter D appears in the summation determining the number of summands going into the higher-order terms.

The articles Dehling and Taqqu (1989); Koul and Surgailis (2002), and El Ktaibi and Ivanoff (2016) are closest to our work. In the following layout of our proof, we emphasize how our results differ from these works.

1. It is necessary to prove tightness in two parameters, more precisely, in the space $D([-\infty,\infty] \times [0,1])$. Furthermore, we allow the underlying process to be subordinated Gaussian. This makes our proofs decisively different from the proofs established in Koul and Surgailis (2002), who only consider (4.2) for fixed t and did not allow for subordinated transformations of the underlying Gaussian process.
2. The first step of our proof is to reduce tightness in $D([-\infty,\infty] \times [0,1])$ to proving tightness in $D([0,1] \times [0,1])$. The corresponding object in $D([0,1] \times [0,1])$ can be written as

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

$$m_N(x,t) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{c_l(x)}{l!} H_l(\xi_n) \quad \text{with} \quad c_l(x) = \mathrm{E}\left(\mathbb{1}_{\{F(G(\xi_0)) \leqslant x\}} H_l(\xi_0)\right). \tag{4.3}$$

3. We use a tightness criterion introduced in Ivanoff (1980) and later utilized in El Ktaibi and Ivanoff (2016) to prove tightness of the sequential empirical process under short-range dependence. El Ktaibi and Ivanoff (2016) take advantage of the boundedness of the empirical process. Those techniques fail for (4.2) since the higher-order terms of the empirical process can no longer be represented as an indicator function.
4. In the main part of our proof, we reduce the tightness criterion in El Ktaibi and Ivanoff (2016) to bounding the probability

$$\mathrm{P}\left(\bar{m}_{N,b}(y,x) > \lambda\right) \quad \text{with} \quad \bar{m}_{N,b}(x,y) := \sup_{t \in [0,b]} |m_N(y,t) - m_N(x,t)|$$

for some $b > 0$ and $x, y \in [0,1]$ with $m_N(y,t)$ as in (4.3). Typically, such bounds are derived through chaining techniques. Dehling and Taqqu (1989) establish a corresponding argument for proving tightness of the empirical process of long-range dependent observations. For this, they take advantage of the reduction principle as stated in (3.2). The reduction principle reduces the problem to proving convergence of the partial sums of the dominating Hermite polynomial. Since none of the summands of the infinite series (4.2) is asymptotically negligible, the chaining technique of Dehling and Taqqu (1989) does not apply to the considered situation. Betken et al. (2023) establish a chaining technique for proving tightness of the tail empirical process of long memory stochastic volatility (LMSV) time series. The major difference to our argument results from a martingale structure of the tail empirical process of LMSV time series. This allows to apply Freedman's inequality, i.e., a Bernstein-type inequality for martingale difference sequences which, as well, does not apply to the situation in this article.
5. A crucial part of the proof and second main technical contribution is to find a bound of the form

$$\mathrm{P}\left(\bar{m}_{N,b}(y,x) > \lambda\right) \leqslant C_{1,\gamma} \frac{1}{\lambda^4} b^{2-\theta} \frac{1}{N^\theta} (y-x) + C_{2,\gamma} \frac{1}{\lambda^4} b^2 (y-x)^{\frac{3}{2}}$$

for some $\theta > 0$, any $b > 0$ and all $x, y \in [0,1]$. The result is formally stated in Lemma B.1. Our proof consists of two major parts. The first one is to extend Billingsley (1968, Thm. 12.2) which provides a probabilistic bound for maxima over partial sums. In Lemma C.1, we provide a similar result, allowing the bound to take a more general form. The second main part of the proof is to verify the assumptions of Lemma C.1. Both, Lemma C.1 and Billingsley (1968, Thm. 12.2) apply under very general assumptions in that both do not impose any assumptions on the dependence structure of the underlying process.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

However, both are based on a probabilistic bound on the distances between partial sums. Given strong temporal dependence, as in our setting, verifying this condition becomes particularly challenging.

5. CONFIDENCE INTERVALS

In this section, we focus on how to utilize HOAs of the empirical process for the construction of confidence intervals. To begin with, we determine confidence intervals for values of the marginal distribution F of a time series $X_1, \dots, X_N$ following Model 2.1. The confidence intervals are based on the empirical analog $F_N(x) := \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{X_n \leqslant x\}}$ of $F(x)$; see Section 5.1. Following this, we derive confidence intervals for quantiles of the marginal distribution; see Section 5.2. Section 5.3 provides a comparative discussion.

First and foremost, we are interested in how well these confidence intervals approximate optimal confidence intervals. For this, note that the goodness of confidence intervals can be assessed on the basis of the following two criteria.

C1: A high coverage probability, i.e., the probability that the true value of the estimated quantity lies in the considered confidence interval should be high.
C2: A short length of the confidence interval.

Based on these criteria, we aim to compare the confidence intervals derived from HOAs of the empirical process to confidence intervals that result from the asymptotic distribution of $F_N(x)$. Therefore, we will first rephrase how to compute the asymptotic confidence intervals and then move on to introducing our approach to derive confidence intervals.

For ease of computations, we base all analysis on the assumption that we are given a subordinated Gaussian time series $X_n = G(\xi_n), n = 1, \dots, N$, resulting from a strictly monotone function G. In this case, the Hermite rank r equals 1 and the Hermite coefficients $\widetilde{c}_l(x)$ can be determined analytically. In particular, it holds that

$$\widetilde{c}_l(x) = \begin{cases} -H_{l-1}(G^{-1}(x))\varphi\left(G^{-1}(x)\right) & \text{if } G \text{ is increasing,} \\ H_{l-1}(G^{-1}(x))\varphi\left(G^{-1}(x)\right) & \text{if } G \text{ is decreasing;} \end{cases} \tag{5.1}$$

see Lemma D.2.

5.1. Confidence Intervals for the Marginal Distribution

Asymptotic Confidence Intervals: For a construction of confidence intervals based on the asymptotic distribution of the empirical process, note that

$$\frac{N}{d_N}\left(F_N(x) - F(x)\right) \xrightarrow{\mathcal{D}} \widetilde{c}_1(x) Z, \tag{5.2}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

where Z is a standard normally distributed random variable, $\widetilde{c}_1(x) = \mathrm{E}\left(\mathbb{1}_{\{G(\xi_0)\leqslant x\}}\xi_0\right)$ and $d_N^2 := \mathrm{Var}\left(\sum_{i=1}^N \xi_i\right) \sim N^{2H}L(N)$; see Dehling and Taqqu (1989). Due to the fact that convergence in (5.2) holds in $D[-\infty,\infty]$, we have

$$1-\alpha = \mathrm{P}\left(|\widetilde{c}_1(x)|^{-1}\frac{N}{d_N}\left(F_N(x)-F(x)\right) \in \left(z_{\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)\right) + o(1),$$

where $z_\alpha := \Phi^{-1}(\alpha)$ and Φ denotes the standard normal distribution function. Therefore, an approximate $1-\alpha$ confidence interval for $F(x)$ based on the asymptotic distribution of the empirical process is given by

$$\left(F_N(x) - \frac{d_N}{N}|\widetilde{c}_1(x)|z_{1-\frac{\alpha}{2}}, F_N(x) - \frac{d_N}{N}|\widetilde{c}_1(x)|z_{\frac{\alpha}{2}}\right). \tag{5.3}$$

Referring back to Example 2.2, the following example establishes these confidence intervals (5.3) for fractional Gaussian noise.

Example 5.1. For fractional Gaussian noise time series with Hurst parameter H, $d_N \sim N^H$ and $|\widetilde{c}_1(x)| = \varphi(x)$, such that the interval in (5.3) equals

$$\left(F_N(x) - N^{H-1}\varphi(x)z_{1-\frac{\alpha}{2}}, F_N(x) - N^{H-1}\varphi(x)z_{\frac{\alpha}{2}}\right).$$

Confidence Intervals Based on HOAs: For a construction of confidence intervals based on the HOA, note that according to Theorem 4.1

$$\sqrt{N}\left(F_N(x)-F(x)\right) - \frac{1}{\sqrt{N}}\sum_{n=1}^N L_n(x) \xrightarrow{\mathcal{D}} Z(x),$$

where $Z(x)$ is normally distributed with mean zero and variance $\sigma^2(x) := \sum_{n\in\mathbb{Z}}\mathrm{Cov}(S_0(x), S_n(x))$ and convergence holds in $D[-\infty,\infty]$. As a result, we have

$$1-\alpha = \mathrm{P}\left((\sigma(x))^{-1}\left(\sqrt{N}\left(F_N(x)-F(x)\right) - \frac{1}{\sqrt{N}}\sum_{n=1}^N L_n(x)\right) \in \left(z_{\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)\right) + o(1)$$

with L_n as in (3.5). Therefore, an approximate $1-\alpha$ confidence interval for $F(x)$ based on HOAs of the empirical process is given by

$$\left(F_N(x) - \frac{1}{N}\sum_{n=1}^N L_n(x) - \frac{\sigma(x)}{\sqrt{N}}z_{1-\frac{\alpha}{2}}, F_N(x) - \frac{1}{N}\sum_{n=1}^N L_n(x) - \frac{\sigma(x)}{\sqrt{N}}z_{\frac{\alpha}{2}}\right). \tag{5.4}$$

5.2. Confidence Intervals for Quantiles

In this section, we establish confidence intervals for quantiles of the marginal distribution of long-range dependent time series. Initially, we describe the

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

construction of confidence intervals for quantiles based on the convergence of the empirical process. Subsequently, we discuss the construction of confidence intervals for quantiles based on HOAs of the empirical process.

Asymptotic Confidence Intervals: The asymptotic distribution of empirical quantiles can be derived from the asymptotic behavior of the empirical process (5.2) and an application of the delta method. In fact, Hössjer and Mielniczuk (1995) showed that for a functional $\phi : (D[-\infty,\infty], \|\cdot\|_\infty) \to \mathbb{R}$, Hadamard-differentiable at F,

$$\frac{N}{d_N}(\phi(F_N)-\phi(F)) \xrightarrow{\mathcal{D}} Z\phi'(F)(\phi(F))\,\widetilde{c}_1(\phi(F)), \tag{5.5}$$

where Z is a standard normally distributed random variable, $\widetilde{c}_1(x) = \mathrm{E}\left(\mathbb{1}_{\{G(\xi_0)\leqslant x\}}\xi_0\right)$, $\phi'(F)$ the derivative in F and $d_N^2 := \mathrm{Var}\left(\sum_{i=1}^N \xi_i\right) \sim N^{2H}L(N)$; see Hössjer and Mielniczuk (1995, Thm. 1). Since our goal is to establish confidence intervals for quantiles $q_p = \inf\{x \mid F(x) \geqslant p\}$, we consider $\phi : (D[-\infty,\infty], \|\cdot\|_\infty) \to \mathbb{R}$, $\phi(F) = F^{-1}(p)$. Given that $r = 1$, (5.5) corresponds to

$$\frac{N}{d_N}\left(F_N^{-1}(p) - F^{-1}(p)\right) \xrightarrow{\mathcal{D}} -Z\frac{1}{F'(F^{-1}(p))}\widetilde{c}_1(F^{-1}(p)).$$

As a result, we have

$$1-\alpha = \mathrm{P}\left(-F'(F^{-1}(p))|\widetilde{c}_1(F^{-1}(p))|^{-1}\frac{N}{d_N}\left(F_N^{-1}(p)-F^{-1}(p)\right) \in \left(z_{\frac{\alpha}{2}}, z_{1-\frac{\alpha}{2}}\right)\right) + o(1).$$

Therefore, an approximate $1-\alpha$ confidence interval for $F^{-1}(p)$ based on the asymptotic distribution of the empirical process is given by

$$\left(F_N^{-1}(p) - \frac{d_N}{N}\frac{1}{F'(F^{-1}(p))}|\widetilde{c}_1(F^{-1}(p))|z_{\frac{\alpha}{2}}, F_N^{-1}(p) - \frac{d_N}{N}\frac{1}{F'(F^{-1}(p))}|\widetilde{c}_1(F^{-1}(p))|z_{1-\frac{\alpha}{2}}\right). \tag{5.6}$$

Example 5.2. For fractional Gaussian noise time series with Hurst parameter H, $d_N \sim N^H$ and $|\widetilde{c}_1(x)| = \varphi(x)$, such that the interval in (5.6) equals

$$\left(F_N^{-1}(p) + N^{H-1}z_{\frac{\alpha}{2}}, F_N^{-1}(p) + N^{H-1}z_{1-\frac{\alpha}{2}}\right).$$

Confidence Intervals Based on HOAs: We propose an alternative way to derive confidence intervals for the quantiles of the marginal distribution of long-range dependent time series based on HOAs of the empirical process. Recall that quantiles can be written as a functional of the distribution F as well as their estimated counterparts. Based on Taylor approximation of the functional ϕ, we can then write

$$\frac{N}{d_N}(\phi(F_N)-\phi(F)) = \phi'_F\left(\frac{N}{d_N}(F_N - F)\right) + o_P(1); \tag{5.7}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

see Hössjer and Mielniczuk (1995) and also Van der Vaart (2000, Thm. 20.8). The right-hand side can be further simplified by

$$\phi'_F\left(\frac{N}{d_N}(F_N - F)\right) = \frac{N}{d_N}\frac{p - F_N(F^{-1}(p))}{F'(F^{-1}(p))} = -\frac{N}{d_N}\frac{(F_N - F)(F^{-1}(p))}{F'(F^{-1}(p))}; \tag{5.8}$$

see Van der Vaart (2000, p. 294). Under the assumption that the underlying time series has Gaussian marginals and for $p = \frac{1}{2}$ (such that $F^{-1}(p)$ corresponds to the median) we get $\phi'_F(\frac{N}{d_N}(F_N - F)) = -\frac{N}{d_N}\frac{(F_N - F)(0)}{\varphi(0)}$. Then, an approximate $1-\alpha$ confidence interval of $F^{-1}(p)$ can be written as

$$\begin{aligned}\Bigg(\phi(F_N) + \frac{1}{\varphi(F^{-1}(p))}\Big(\frac{1}{N}\sum_{n=1}^{N} L_n(F^{-1}(p)) + \frac{1}{\sqrt{N}}\sigma(F^{-1}(p))z_{1-\frac{\alpha}{2}}\Big),\\ \phi(F_N) + \frac{1}{\varphi(F^{-1}(p))}\Big(\frac{1}{N}\sum_{n=1}^{N} L_n(F^{-1}(p)) + \frac{1}{\sqrt{N}}\sigma(F^{-1}(p))z_{\frac{\alpha}{2}}\Big)\Bigg);\end{aligned} \tag{5.9}$$

see Lemma D.1 and its proof for more details on the calculations.

5.3. Discussion

The confidence intervals established in Sections 5.1 and 5.2 all depend on the subordinating function G as well as the Hurst parameter H, quantities that are unknown in practice. Additionally, confidence intervals based on the asymptotic distribution of the empirical process (such as (5.3) and (5.6)) depend on the slowly varying function L through d_N. By definition, d_N^2 corresponds to the long-run variance of a long-range dependent Gaussian process. Due to the fact that the data is assumed to be subordinated to this process, d_N cannot be estimated straightforwardly, i.e., by a long-run variance estimator applied to the observed data. On the other hand, an estimation can be based on the asymptotic relation $d_N \sim N^H L^{\frac{1}{2}}(N)$. For this, it has to be taken into account that H characterizes the autocovariances of the Gaussian process (not the observed subordinated process). Only for a Hermite rank of the subordinating function G that equals 1, the Hurst parameter of the Gaussian process and that of the subordinated Gaussian process coincide, such that H can be estimated by established methods (such as R/S-estimation or local Whittle estimation); see also Section 6.4. Nonetheless, estimation of d_N also requires an approximation of the slowly varying function L. Unfortunately, we are not aware of any estimation procedure meeting this task. In particular cases, e.g., when the data stems from fractional Gaussian noise, L corresponds to a multiplicative constant depending on the parameter H only; see Examples 5.1 and 5.2. In these cases, the estimation can solely be based on estimation of H, but presupposes knowledge of the subordinating function G. In contrast to confidence intervals based on the asymptotic distribution of the empirical process, confidence intervals based on HOAs of the empirical process (such as (5.4) and (5.9)) do not depend on L. For an empirical comparison of

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

the two procedures for confidence interval construction in Section 6, we assume knowledge of the slowly varying function L. This knowledge can be exploited for confidence interval construction based on the asymptotic distribution of the empirical process, but is not needed for the proposed method of confidence interval construction based on HOAs of the empirical process. Moreover, we would like to point out that (5.9) only depends on the Hurst parameter H through L_n; see (5.4). In particular, the ceiling function applied to H determines the number of summands included in the construction of confidence intervals based on HOAs. Accordingly, these are less sensitive to small errors in the estimation of H than confidence intervals based on the asymptotic distribution. Given that inference on long-range dependent time series relies on how well the corresponding Hurst parameter is estimated, we expect confidence intervals based on HOAs to be more robust to misspecification of H.

Section 5.2 focuses on deriving confidence intervals for quantiles of the marginal distribution based on an HOA of the empirical process. The proposed procedure takes advantage of the Taylor expansion (5.7) of a general functional ϕ. Due to the generality of the results, we believe that similar results can be achieved for other estimators that have a representation as functionals of the empirical process, such as Huber's estimator and M-estimators.

6. NUMERICAL STUDIES

For our numerical studies, we consider the procedures proposed in Section 5. We compare the coverage rate as well as the length of asymptotic confidence intervals with those based on HOAs. To assess the performance of the proposed procedures, we assume that the underlying time series follows Model 2.1 with $G = id$, i.e., the time series is assumed to be long-range dependent with Gaussian marginals. In particular, we assume that $G = id$ is known although in practice G needs to be estimated. Estimation of G can, for example, be based on the relation $X \stackrel{\mathcal{D}}{=} F^{-1}(\Phi(\xi))$ (resulting from $X = G(\xi)$ for a standard normally distributed random variable ξ), where Φ denotes the standard normal distribution function and F the marginal distribution of X. Accordingly, G could be estimated by $\hat{F}^{-1} \circ \Phi$, where $\hat{F}^{-1}$ corresponds to the generalized inverse of the empirical distribution function of the observed data. Note that, nonetheless, estimation of G will add uncertainty to both procedures, such that for the purpose of comparison we refrain from estimation of G.

In the following, we focus on confidence intervals for the marginal distribution and confidence intervals for the median (Sections 6.2 and 6.3). Section 6.1 discusses the estimation of the long-run variance and the Hurst parameter.

6.1. Estimation of Long-Run Variance and Hurst Parameter

In order to compute the confidence intervals discussed in Section 5, we need to estimate the long-run variance. Furthermore, we provide simulation results under

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

the assumption that the Hurst parameter H is known and under the assumption that H is unknown.

The long-run variance $\sigma^2(x) := \sum_{n\in\mathbb{Z}} \mathrm{Cov}(S_0(x), S_n(x))$ cannot be computed analytically. In order to make our results applicable, we therefore need to estimate $\sigma^2(x)$. We use the kernel smoothing long-run variance estimator

$$\widehat{\sigma}^2(x) = \sum_{j=-(N-1)}^{N-1} K\left(\frac{j}{b_N}\right)\widehat{\gamma}_N(j),$$

where $K(x) = (1-|x|)\mathbb{1}_{\{|x|\leq 1\}}$ is the Bartlett kernel function, b_N denotes a bandwidth parameter, and $\widehat{\gamma}_N(j)$ is the sample autocovariance at lag j. For our simulation study, we use the command `hurstexp` in the `R` package `cointReg`. To determine the bandwidth, we use the command `getBandwidth`. For an estimation of the Hurst parameter H, we used the R/S procedure following the description in Weron (2002, Sect. 2.1). The estimator is implemented by `getLongRunVar` in the `R` package `pracma`.

6.2. Confidence Intervals for the Marginal Distribution

We construct confidence intervals for the marginal distribution F based on the asymptotic distribution and based on HOAs of the empirical process of long-range dependent time series. For a visual comparison of the two different methods, see Figures 3–5. To numerically assess the quality of the computed intervals, we report their coverage rate and width evaluated at different x. In our simulation study, we consider different scenarios ranging from small to large sample sizes ($N = 200$ and $N = 1000$) as well as from small to large Hurst parameters ($H = 0.55$ and $H = 0.95$). Pointing toward Figures 3 and 4, which are based on sample sizes $N = 200$ and $N = 1000$, we see only a slight improvement of the interval length for the asymptotic confidence intervals. The mild improvement emphasizes how the asymptotic confidence intervals are impacted by the slow convergence rate of the empirical process under LRD. That said, we fix the sample size to $N = 200$ and compare Figures 3 and 4. Focusing on $x = 0$, one can observe that for larger Hurst parameters, the width increases significantly for the asymptotic method. Naturally, the increase in width results in a higher coverage rate. Confidence intervals based on the proposed HOA method, however, are robust with respect to the value of the Hurst parameter and outperform the traditional construction of confidence intervals with respect to the coverage rate; see Figures 3 and 4.

Most notably, Figures 3 and 4 reveal that asymptotic confidence intervals may have lower coverage rates than those based on HOAs, while confidence intervals based on HOAs are shorter. This phenomenon results from the fact that the centers of the confidence intervals differ, i.e., the smaller confidence interval is not necessarily contained in the larger one. In particular, it therefore happens that the asymptotic confidence interval is larger, but nonetheless does not cover $F(x)$.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

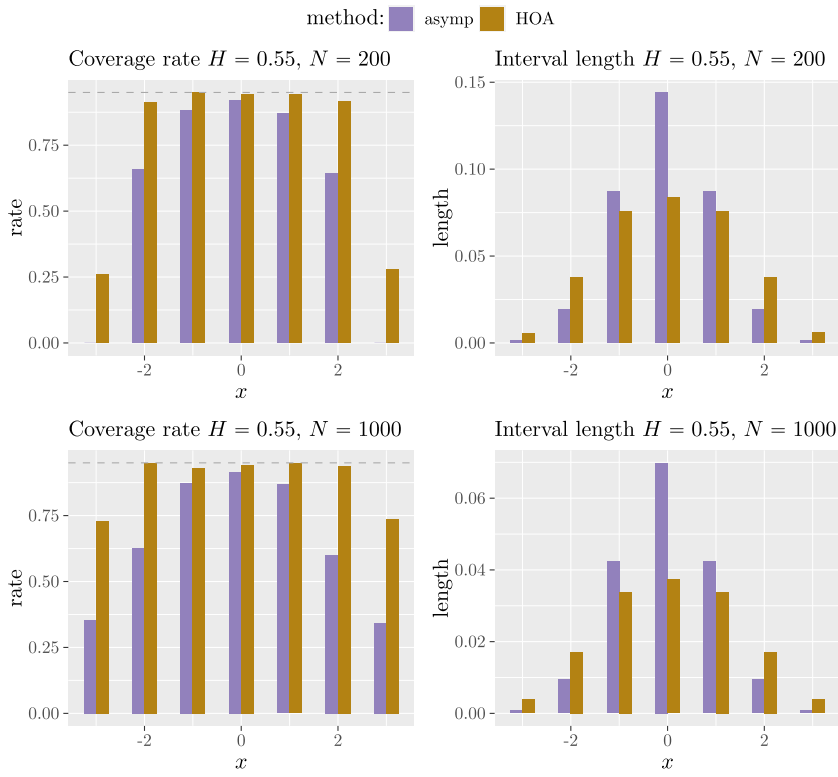


FIGURE 3. The coverage rate and length of confidence intervals for the marginal distribution $F(x)$ evaluated at different x. The two displayed methods to calculate the confidence intervals are based on the asymptotic distribution (asymp) and our HOA. The simulations are based on 2000 repetitions for Gaussian time series of length $N = 200$ (first row) and $N = 1000$ (second row) with Hurst parameter $H = 0.55$. The dashed gray line depicts the significance level of 95%.

6.3. Confidence Intervals for the Median

In this section, we consider confidence intervals for the median based on long-range dependent time series characterized by different Hurst parameters. Again, we consider different scenarios ranging from small to large sample sizes ($N = 200$ and $N = 1000$) as well as from small to large Hurst parameters (from $H = 0.55$ to $H = 0.95$), and we assess the quality of the confidence intervals through interval length and coverage rate. Pointing toward Figure 6, which is based on sample sizes $N = 200$ and $N = 1000$, we see only a slight improvement of the interval length for the asymptotic confidence intervals. This emphasizes how the asymptotic confidence intervals are impacted by the slow convergence rate of the empirical process under LRD. Therefore, instead of considering the impact of the sample size, we focus on how varying the Hurst parameter influences the coverage rate and interval length. In this regard, Figure 6 clearly

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

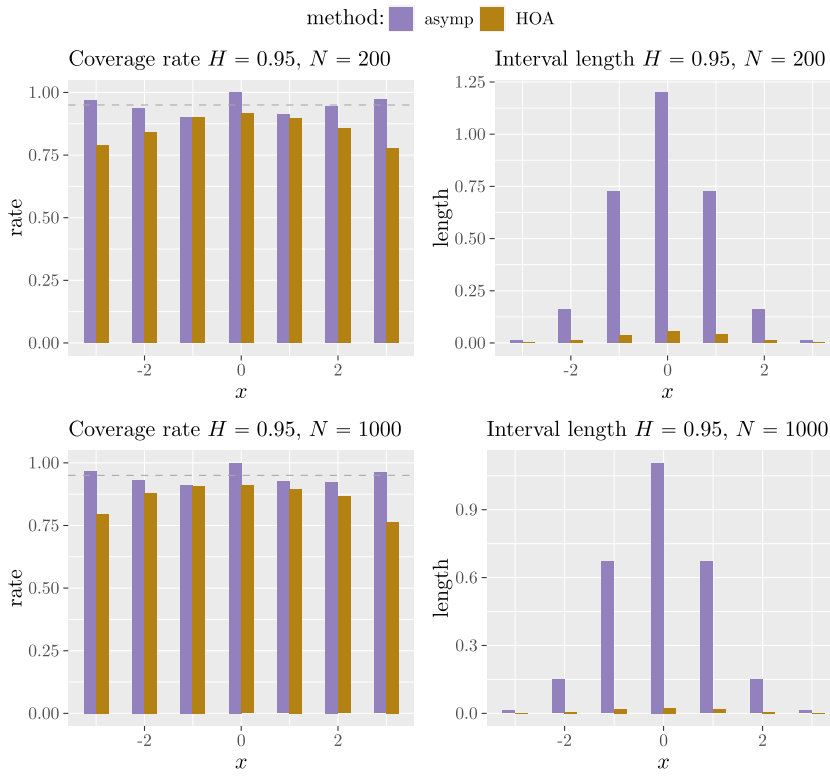


FIGURE 4. The coverage rate and length of confidence intervals for the marginal distribution $F(x)$ evaluated at different x. The two displayed methods to calculate the confidence intervals are based on the asymptotic distribution (asymp) and our HOA. The simulations are based on 2000 repetitions for Gaussian time series of length $N = 200$ (first row) and $N = 1000$ (second row) with Hurst parameter $H = 0.95$. The dashed gray line depicts the significance level of 95%.

demonstrates that the length of a confidence interval constructed on the basis of the asymptotic distribution of the empirical process increases almost exponentially with increasing value of H. This may be attributed to the exponential increase of the number of summands needed to calculate the lower-order terms of the empirical process; see Figure 2.

In contrast to basing confidence intervals on the asymptotic distribution of the empirical process, Figure 6 illustrates robustness of the confidence interval lengths to different values of the Hurst parameter if the construction of confidence intervals is based on HOAs of the empirical process. For Hurst parameters bigger than $H = 0.9$ a significant drop of the coverage rate can be observed. We attribute this observation to the fact that the stronger the dependence in a time series the higher the number of observations needed to reflect this dependence. When adjusting confidence intervals by the true number of summands in the lower-order term this finite-sample phenomenon is not accounted for resulting in lower coverage rates.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

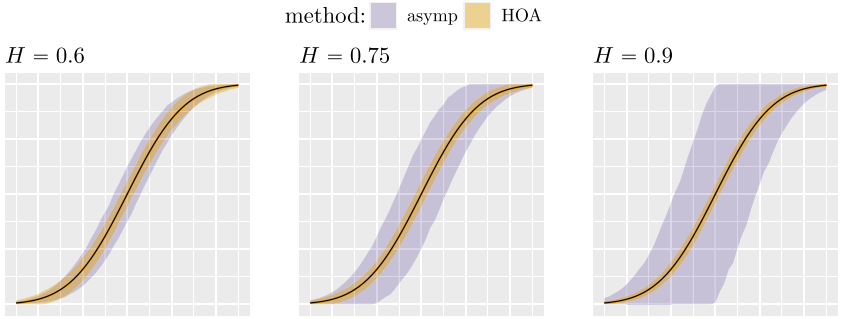


FIGURE 5. Confidence intervals for the marginal distribution F. The two displayed methods to calculate the confidence intervals are based on the asymptotic distribution (asymp) and our HOA. The simulations are based on 1000 repetitions for Gaussian time series of length $N = 100$ with Hurst parameters $H = 0.6, H = 0.75,$ and $H = 0.9$.

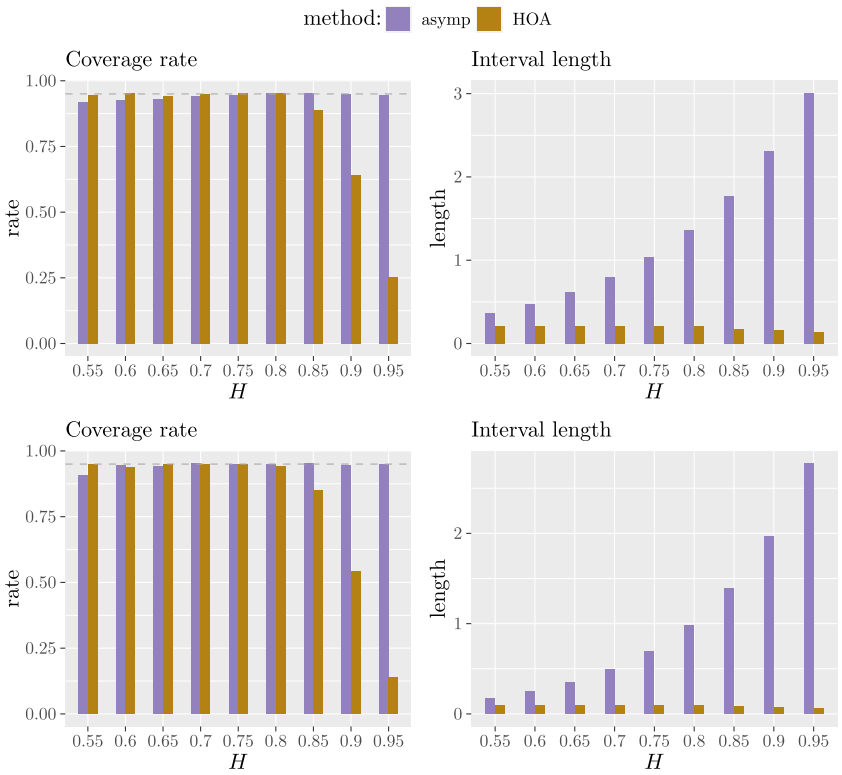


FIGURE 6. Coverage rate and interval length of confidence intervals for the median $F^{-1}(1/2)$ based on long-range dependent time series characterized by different Hurst parameters. For this, the distribution of the median is approximated by the asymptotic distribution (asymp) and our HOA of the empirical process. Simulations are based on 2000 repetitions for Gaussian time series of length $N = 200$ (first row) and $N = 1000$ (second row). The dashed gray line depicts the significance level of 95%.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

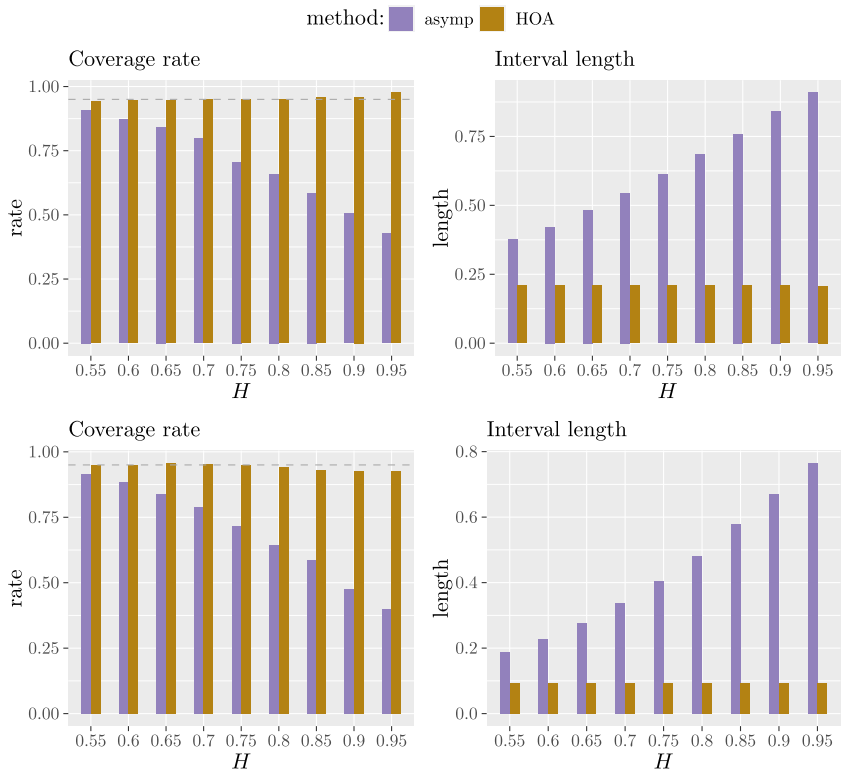


FIGURE 7. Coverage rate and interval length of confidence intervals for the median $F^{-1}(1/2)$ based on long-range dependent time series characterized by different Hurst parameters. For this, the distribution of the median is approximated by the asymptotic distribution (asymp) and our HOA of the empirical process. The Hurst parameter is replaced by its R/S-estimator. Simulations are based on 2000 repetitions for Gaussian time series of length $N = 200$ (first row) and $N = 1000$ (second row). The dashed gray line depicts the significance level of 95%.

6.4. Confidence Intervals Based on an Estimated Hurst Parameter

To make the construction of confidence intervals based on HOAs of the empirical process feasible for practical purposes, we need to consider the case where the Hurst parameter is unknown. As discussed in Section 6.1, we base estimation of the Hurst parameter on the so-called R/S-method. In this section, we focus on studying confidence intervals for the median. As done in Figure 6, the median was considered for a range of different Hurst parameters. We therefore use the median to illustrate how an estimated Hurst parameter changes the empirical coverage rates and lengths; see Figure 7.

Next, we compare the numerical results based on estimation of the Hurst parameter (Figure 7) with the numerical results that are based on the assumption that the

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Hurst parameter is known (Figure 6). It is notable that the lengths of confidence intervals resulting from approximation of the empirical process by its asymptotic distribution tend to be shorter when the Hurst parameter is estimated while their coverage rate is lower. Although the coverage rate of confidence intervals that are based on HOAs of the empirical process and estimation of the Hurst parameter declines for Hurst parameters larger than 0.9, this effect is not as pronounced as for the simulations that assumed knowledge of the Hurst parameter. We attribute this phenomenon to the fact that R/S estimation tends to underestimate the Hurst parameter and that the higher the value of the Hurst parameter, the bigger the estimation bias; see Taqqu et al. (1995). An underestimation of the Hurst parameter results in a smaller number of lower-order terms entering the approximation of confidence intervals. We conjecture that basing the approximation of confidence intervals on a lower number of summands than suggested by our theory for long-range dependent time series compensates for very strong dependence in time series not being reflected in the finite-sample behavior of time series with relatively low sample size. We call this phenomenon "benign underestimation of LRD". Our conjecture is supported by the fact that an increasing number of observations results in a drop of coverage rate based on the estimated Hurst parameter; see Figure 7.

Note that similar to the confidence intervals for the marginal distribution, we can observe that the asymptotic confidence intervals result in a smaller coverage rate than the ones based on HOAs while the latter are shorter. In particular, one can see that the larger the Hurst parameter, the smaller the coverage rate of the asymptotic confidence intervals while the intervals based on HOAs maintain a constant coverage rate and interval length.

7. CONCLUSION AND DISCUSSION

In this work, we study HOAs of the empirical process as an approach to improving statistical inference for long-range dependent time series. More precisely, we study confidence intervals for values of the empirical process and for quantiles of the marginal distribution of stationary time series that are based on an approximation of the empirical process through higher-order terms in its Hermite expansion. For statistics that can be expressed as partial sums of subordinated Gaussian processes, the Hermite expansion corresponds to an L^2-expansion of the subordinating function in orthonormal polynomials. The inclusion of higher-order terms in this expansion for the construction of confidence intervals results in narrower and more accurate confidence bands, especially when compared to those derived from first-order asymptotic theory. Most notably, this approach differs from Gram–Charlier and Edgeworth expansions, which aim at improving approximations of the cumulative distribution function through incorporation of higher-order information. The latter techniques incorporate correction terms relating to skewness, kurtosis, higher-order cumulants, or moments and, to the best of our knowledge, have so far been analyzed against the background of short-range dependent time series and

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

under the assumption of existing higher-order moments. Nonetheless, just as in the Edgeworth expansion framework, the number of terms included in the Hermite expansion of the considered statistic improves upon approximation quality from a theoretical perspective. In practice, however, the theoretical improvement does not show due to the finite-sample behavior of statistics not adequately reflecting in time series. Interestingly, our numerical results suggest that this mismatch between theory and practice may be mitigated by a phenomenon we term *benign underestimation of LRD*; see Section 6.4. This effect appears to stabilize inference procedures in practice and presents another intriguing direction for future research.

The main theoretical contribution of this article is a proof for the convergence of higher-order terms in the Hermite expansion of the sequential empirical process for long-range dependent time series. This result is of general interest for empirical process theory and paves the way for novel approaches with respect to statistical inference for long-range dependent time series. First numerical approaches using the established theory illustrate an alternative way of constructing confidence intervals based on long-range dependent observations.

In comparison to the construction of confidence intervals based on the asymptotic distribution of the empirical process, the proposed procedure improves the quality of confidence intervals for the empirical process and quantiles of the marginal distribution. Generally speaking, our results provide sufficient theoretical groundwork for the use of HOAs for statistical inference on long-range dependent time series. We conjecture that analogous theory would establish HOAs for sequential partial sum processes of subordinated Gaussian sequences. Such results would lay the foundation for improving upon statistical inference in change-point analysis for long-range dependent time series. When testing stationarity against the alternative hypothesis of a structural change in a time series by means of the Wilcoxon test, the phenomenon illustrated in Figure 1 results in a high number of false positives; see Dehling et al. (2013). While our proposed second-order approximation is expected to resolve an inflated test size, a potential drawback could be a lower size being accompanied by a loss in test power. This is an expected trade-off since critical values derived from HOAs would rely heavily on the lower-order term that also drives most of the behavior of the corresponding test statistic. In addition to applications in change-point analysis, we envision that the established theory for the two-parameter empirical process applies to goodness-of-fit testing based on Kolmogorov–Smirnov and Cramér–von Mises statistics. We leave both change-point analysis and goodness-of-fit testing based on HOA of the empirical process as challenges for future research.

A. PROOF OF THEOREM 4.1

In order to prove Theorem 4.1, we first investigate the convergence of the finite-dimensional distributions and then tightness in $D([-\infty,\infty]\times[0,1])$; see Sections A.1 and A.2, respectively.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

For the proof, we will make use of the following notation

$$\widetilde{m}_N(x,t) := \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} S_n(x) \tag{A.1}$$

with $S_n(x)$ as in (3.5).

A.1. Convergence of the Finite-Dimensional Distributions

We need to show convergence of the finite-dimensional distributions, i.e.,

$$\widetilde{m}_N(x,t) \overset{f.d.d.}{\to} S(x,t),$$

where $\{S(x,t)\}$ is the limiting process with cross-covariances given in (4.1). For this, it suffices to show that for all $q_1, q_2 \in \mathbb{N}$, and $(x_i, t_j) \in [-\infty, \infty] \times [0,1]$, $i = 1, \dots, q_1$; $j = 1, \dots, q_2$,

$$\left(\widetilde{m}_N(x_i,t_j)\right)_{i=1,\dots,q_1;\, j=1,\dots,q_2} \xrightarrow{\mathcal{D}} \left(S(x_i,t_j)\right)_{i=1,\dots,q_1;\, j=1,\dots,q_2}.$$

Recall from (3.5) and (A.1) that

$$\widetilde{m}_N(x_i,t_j) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt_j \rfloor} \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{\widetilde{c}_l(x_i)}{l!} H_l(\xi_n) \tag{A.2}$$

and set

$$\left(\widetilde{m}_N(x_i,t_j)\right)_{\substack{i=1,\dots,q_1 \\ j=1,\dots,q_2}} = \left(\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt_j \rfloor} G_i(\xi_n) \right)_{\substack{i=1,\dots,q_1 \\ j=1,\dots,q_2}} \quad \text{with } G_i(\cdot) = \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{\widetilde{c}_l(x_i)}{l!} H_l(\cdot), \tag{A.3}$$

such that we have a $q_1 \times q_2$-dimensional matrix of normalized partial sums of subordinated Gaussian sequences. In particular, different indices i correspond to different functions G_i.

Given (2.1) and since the summation in (A.3) starts with $l = \lceil \frac{1}{D} \rceil$, all $G_i(\xi_n)$ are short-range dependent in the sense that their autocovariances are absolutely summable as shown in (3.3). Furthermore, let r_i denote the Hermite rank of G_i. Due to (A.3), we have $\lceil \frac{1}{D} \rceil \leqslant r_i$ such that $\frac{1}{D} < \lceil \frac{1}{D} \rceil \leqslant r_i$ since $\frac{1}{D} \notin \mathbb{N}$.

Then, by Bai and Taqqu (2013, Thm. 3), we have

$$\left(\frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt_j \rfloor} G_i(\xi_n) \right)_{\substack{i=1,\dots,q_1 \\ j=1,\dots,q_2}} \xrightarrow{\mathcal{D}} (\mathcal{G}_1, \dots, \mathcal{G}_{q_1})',$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

where $\mathcal{G}_i, i = 1, \ldots, q_1$, are q_2-dimensional Gaussian vectors $\mathcal{G}_i = (\mathcal{G}_i(t_1), \ldots, \mathcal{G}_i(t_{q_2}))'$ with

$$\operatorname{Cov}(\mathcal{G}_{i_1}(t_{j_1}), \mathcal{G}_{i_2}(t_{j_2})) = \lim_{N\to\infty} \frac{1}{N} \sum_{n_1=1}^{\lfloor Nt_{j_1}\rfloor} \sum_{n_2=1}^{\lfloor Nt_{j_2}\rfloor} \sum_{l_1,l_2=\lceil \frac{1}{D}\rceil}^{\infty} \frac{\widetilde{c}_{l_1}(x_{i_1})\widetilde{c}_{l_2}(x_{i_2})}{l_1!\, l_2!} \operatorname{E}(H_{l_1}(\xi_{n_1})H_{l_2}(\xi_{n_2}))$$

$$= \min(t_{j_1}, t_{j_2}) \sum_{l=\lceil \frac{1}{D}\rceil}^{\infty} \frac{\widetilde{c}_l(x_{i_1})\widetilde{c}_l(x_{i_2})}{l!} \sum_{n=-\infty}^{\infty} \gamma^l(n), \tag{A.4}$$

where (A.4) follows by Bai and Taqqu (2013, eqn. 11) and since, for $l \geqslant \lceil \frac{1}{D}\rceil$ and due to (3.3),

$$\sum_{n\in\mathbb{Z}} |\gamma(n)|^l < \infty.$$

A.2. Tightness

Since the object of interest $\widetilde{m}_N(x,t)$ in (A.1) is a process in two parameters, proving tightness becomes particularly challenging. We will first give a tightness criterion in $D([-\infty,\infty]\times[0,1])$ and then argue that it suffices to prove tightness in $D([0,1]\times[0,1])$.

In order to prove tightness of $\widetilde{m}_N(x,t)$ in $D([-\infty,\infty]\times[0,1])$, we validate the following tightness criterion: for all $\varepsilon > 0$

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \operatorname{P}\left(\sup_{\substack{|x_2-x_1|<\delta \\ x_1,x_2\in\mathbb{R}}} \sup_{\substack{|t_2-t_1|<\delta \\ 0\leqslant t_1,t_2\leqslant 1}} |\widetilde{m}_N(x_2,t_2) - \widetilde{m}_N(x_1,t_1)| > \varepsilon \right) = 0;$$

see El Ktaibi and Ivanoff (2016, Formula (26)). In a more general setting, the criterion was introduced in Ivanoff (1980). We further write

$$\widetilde{m}_N(x_2,t_2) - \widetilde{m}_N(x_1,t_1) = \widetilde{m}_N(x_2,t_2) - \widetilde{m}_N(x_1,t_2) + \widetilde{m}_N(x_1,t_2) - \widetilde{m}_N(x_1,t_1).$$

Then, it suffices to show

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \operatorname{P}\left(\sup_{\substack{|x_2-x_1|<\delta \\ x_1,x_2\in\mathbb{R}}} \sup_{t\in[0,1]} |\widetilde{m}_N(x_2,t) - \widetilde{m}_N(x_1,t)| > \varepsilon \right) = 0, \tag{A.5}$$

$$\lim_{\delta\to 0} \limsup_{N\to\infty} \operatorname{P}\left(\sup_{x\in\mathbb{R}} \sup_{\substack{|t_2-t_1|<\delta \\ 0\leqslant t_1,t_2\leqslant 1}} |\widetilde{m}_N(x,t_2) - \widetilde{m}_N(x,t_1)| > \varepsilon \right) = 0. \tag{A.6}$$

For (A.5), note that due to continuity of F^{-1} (following from strict monotonicity and continuity of F) for every $\delta > 0$, there exists a $\widetilde{\delta} > 0$ such that $|x_2 - x_1| < \widetilde{\delta}$ implies $|F^{-1}(x_1) - F^{-1}(x_2)| < \delta$. It then follows that

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

$$\begin{aligned}
&\sup_{\substack{|x_2-x_1|<\widetilde{\delta}\\ x_1,x_2\in[-\infty,\infty]}} \sup_{t\in[0,1]} |\widetilde{m}_N(x_2,t)-\widetilde{m}_N(x_1,t)|\\
&\quad= \sup_{\substack{|x_2-x_1|<\widetilde{\delta}\\ x_1,x_2\in[-\infty,\infty]}} \sup_{t\in[0,1]} \left|\widetilde{m}_N\left(F(F^{-1}(x_2)),t\right)-\widetilde{m}_N\left(F(F^{-1}(x_1)),t\right)\right|\\
&\quad\leqslant \sup_{\substack{|x_2-x_1|<\delta\\ x_1,x_2\in[0,1]}} \sup_{t\in[0,1]} |\widetilde{m}_N(F(x_2),t)-\widetilde{m}_N(F(x_1),t)|\\
&\quad= \sup_{\substack{|x_2-x_1|<\delta\\ x_1,x_2\in[0,1]}} \sup_{t\in[0,1]} |m_N(x_2,t)-m_N(x_1,t)|
\end{aligned}$$

and accordingly

$$\begin{aligned}
&\mathrm{P}\left(\sup_{\substack{|x_2-x_1|<\widetilde{\delta}\\ x_1,x_2\in[-\infty,\infty]}} \sup_{t\in[0,1]} |\widetilde{m}_N(x_2,t)-\widetilde{m}_N(x_1,t)|>\varepsilon\right)\\
&\quad\leqslant \mathrm{P}\left(\sup_{\substack{|x_2-x_1|<\delta\\ x_1,x_2\in[0,1]}} \sup_{t\in[0,1]} |m_N(x_2,t)-m_N(x_1,t)|>\varepsilon\right).
\end{aligned}$$

For (A.6), note that due to $F:[-\infty,\infty]\longrightarrow[0,1]$ being a bijective function and due to $m_N(x,t)=\widetilde{m}_N(F^{-1}(x),t)$ with $m_N(x,t)$ as in (4.3),

$$\begin{aligned}
&\lim_{\delta\to 0}\limsup_{N\to\infty}\mathrm{P}\left(\sup_{x\in[-\infty,\infty]} \sup_{\substack{|t_2-t_1|<\delta\\ 0\leqslant t_1,t_2\leqslant 1}} |\widetilde{m}_N(x,t_2)-\widetilde{m}_N(x,t_1)|>\varepsilon\right)\\
&\quad=\lim_{\delta\to 0}\limsup_{N\to\infty}\mathrm{P}\left(\sup_{x\in[0,1]} \sup_{\substack{|t_2-t_1|<\delta\\ 0\leqslant t_1,t_2\leqslant 1}} \left|\widetilde{m}_N\left(F^{-1}(x),t_2\right)-\widetilde{m}_N\left(F^{-1}(x),t_1\right)\right|>\varepsilon\right)\\
&\quad=\lim_{\delta\to 0}\limsup_{N\to\infty}\mathrm{P}\left(\sup_{x\in[0,1]} \sup_{\substack{|t_2-t_1|<\delta\\ 0\leqslant t_1,t_2\leqslant 1}} |m_N(x,t_2)-m_N(x,t_1)|>\varepsilon\right).
\end{aligned}$$

It follows that the criteria (A.5) and (A.6) can be reformulated as

$$\lim_{\delta\to 0}\limsup_{N\to\infty}\mathrm{P}\left(\sup_{\substack{|x_2-x_1|<\delta\\ 0\leqslant x_1,x_2\leqslant 1}} \sup_{t\in[0,1]} |m_N(x_2,t)-m_N(x_1,t)|>\varepsilon\right)=0, \tag{A.7}$$

$$\lim_{\delta\to 0}\limsup_{N\to\infty}\mathrm{P}\left(\sup_{x\in[0,1]} \sup_{\substack{|t_2-t_1|<\delta\\ 0\leqslant t_1,t_2\leqslant 1}} |m_N(x,t_2)-m_N(x,t_1)|>\varepsilon\right)=0. \tag{A.8}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

We consider (A.7) and (A.8) separately. Both proofs are based on chaining techniques following the ideas in Dehling and Taqqu (1989, p. 1778) and Betken et al. (2023, Sect. 5.1.4).

A.2.1. *Proof of (A.7).* In order to prove (A.7), we apply a chaining technique. For this, we define the intervals

$$I_{1,p} := [2p\delta, 2(p+1)\delta] \quad \text{and} \quad I_{2,p} := [(2p+1)\delta, (2(p+1)+1)\delta]$$

for $p = 0, \dots, L_\delta := \lceil \frac{1}{2\delta} - \frac{3}{2} \rceil$. Then, the expression inside P in (A.7) can be bounded as

$$\begin{aligned} &\sup_{\substack{|x_2 - x_1| < \delta \\ 0 \leqslant x_1, x_2 \leqslant 1}} \sup_{t \in [0,1]} |m_N(x_2, t) - m_N(x_1, t)| \\ &\quad \leqslant \max_{0 \leqslant p \leqslant L_\delta} \sup_{x_1, x_2 \in I_{1,p}} \sup_{t \in [0,1]} |m_N(x_2, t) - m_N(x_1, t)| \\ &\qquad + \max_{0 \leqslant p \leqslant L_\delta} \sup_{x_1, x_2 \in I_{2,p}} \sup_{t \in [0,1]} |m_N(x_2, t) - m_N(x_1, t)|. \end{aligned} \tag{A.9}$$

In the following, we consider only the first summand in (A.9), since for the second summand analogous considerations hold. For this reason, it remains to show that

$$\lim_{\delta \to 0} \limsup_{N \to \infty} \mathrm{P}\left(\max_{0 \leqslant p \leqslant L_\delta} \sup_{x_1, x_2 \in I_{1,p}} \sup_{t \in [0,1]} |m_N(x_2, t) - m_N(x_1, t)| > \varepsilon \right) = 0.$$

For this, it suffices to show that

$$\lim_{\delta \to 0} \limsup_{N \to \infty} \frac{1}{\delta} \max_{0 \leqslant p \leqslant L_\delta} \mathrm{P}\left(\sup_{x_1, x_2 \in I_{1,p}} \sup_{t \in [0,1]} |m_N(x_2, t) - m_N(x_1, t)| > \varepsilon \right) = 0.$$

We write $I_{1,p} = [a_p, a_{p+1}]$, i.e., $a_p := 2p\delta$ and $a_{p+1} := 2(p+1)\delta$. Note that

$$\sup_{x_1, x_2 \in I_{1,p}} \sup_{t \in [0,1]} |m_N(x_2, t) - m_N(x_1, t)| \leqslant 2 \sup_{x \in [0, 2\delta]} \sup_{t \in [0,1]} \left| m_N(a_p, t) - m_N(a_p + x, t) \right|. \tag{A.10}$$

Define refining partitions $x_i(k)$ for $k = 0, \dots, K_N$ with $K_N \to \infty$, for $N \to \infty$, and

$$x_i(k) := a_p + \frac{i}{2^k} 2\delta, \quad i = 0, \dots, 2^k, \tag{A.11}$$

and choose $i_k(x)$ such that

$$a_p + x \in \left(x_{i_k(x)}(k), x_{i_k(x)+1}(k) \right].$$

We write

$$\bar{m}_{N,b}(x, y) := \sup_{t \in [0,b]} |m_N(y, t) - m_N(x, t)|, \quad \bar{m}_N(x, y) := \bar{m}_{N,1}(x, y). \tag{A.12}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Then, with help of the introduced partition (A.11), (A.10) can be bounded as

$$\sup_{t\in[0,1]}\left|m_N(a_p,t)-m_N(a_p+x,t)\right| \leqslant \sum_{k=1}^{K_N}\bar{m}_N(x_{i_k(x)}(k),x_{i_{k-1}(x)}(k-1))+\bar{m}_N(x_{i_{K_N}(x)}(K_N),a_p+x). \tag{A.13}$$

Consequently, (A.13) can be used to infer (A.14) below

$$\begin{aligned}&\mathrm{P}\left(\sup_{x\in[0,2\delta]}\sup_{t\in[0,1]}\left|m_N(a_p,t)-m_N(a_p+x,t)\right|>\varepsilon\right)\\ &\leqslant\sum_{k=1}^{K_N}\mathrm{P}\left(\sup_{x\in[0,2\delta]}\bar{m}_N(x_{i_k(x)}(k),x_{i_{k-1}(x)}(k-1))>\frac{\varepsilon}{(k+3)^2}\right)\\ &\quad+\mathrm{P}\left(\sup_{x\in[0,2\delta]}\bar{m}_N(x_{i_{K_N}(x)}(K_N),a_p+x)>\varepsilon-\sum_{k=0}^{\infty}\frac{\varepsilon}{(k+3)^2}\right) \qquad \text{(A.14)}\\ &\leqslant\sum_{k=1}^{K_N}\sum_{i=0}^{2^k-1}\mathrm{P}\left(\bar{m}_N(x_{i+1}(k),x_i(k))>\frac{\varepsilon}{(k+3)^2}\right)\\ &\quad+\mathrm{P}\left(\sup_{x\in[0,2\delta]}\bar{m}_N(x_{i_{K_N}(x)}(K_N),a_p+x)>\frac{\varepsilon}{2}\right), \qquad \text{(A.15)}\end{aligned}$$

since $\sum_{k=0}^{\infty}\frac{\varepsilon}{(k+3)^2}\leqslant\frac{\varepsilon}{2}$.

Throughout all following arguments, C is a generic constant that can change upon each appearance. We consider the two probabilities in (A.15) separately. The first one can be dealt with as follows:

$$\begin{aligned}&\sum_{k=1}^{K_N}\sum_{i=0}^{2^k-1}\mathrm{P}\left(\bar{m}_N(x_{i+1}(k),x_i(k))>\frac{\varepsilon}{(k+3)^2}\right)\\ &\leqslant C\sum_{k=1}^{K_N}\sum_{i=0}^{2^k-1}\frac{(k+3)^8}{\varepsilon^4}\left(\frac{1}{N^\theta}\left(x_{i+1}(k)-x_i(k)\right)+\left(x_{i+1}(k)-x_i(k)\right)^{\frac{3}{2}}\right) \qquad \text{(A.16)}\\ &=C\sum_{k=1}^{K_N}\sum_{i=0}^{2^k-1}\frac{(k+3)^8}{\varepsilon^4}\left(\frac{1}{N^\theta}\frac{2\delta}{2^k}+\left(\frac{2\delta}{2^k}\right)^{\frac{3}{2}}\right) \qquad \text{(A.17)}\\ &\leqslant C\delta\sum_{k=1}^{K_N}\frac{(k+3)^8}{\varepsilon^4}\frac{1}{N^\theta}+C\delta^{\frac{3}{2}}\sum_{k=1}^{K_N}\frac{(k+3)^8}{\varepsilon^4}\left(\frac{1}{2^k}\right)^{\frac{1}{2}}\\ &\leqslant C\delta^{\frac{3}{2}} \qquad \text{(A.18)}\end{aligned}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

for sufficiently large N, where (A.16) follows from Lemma B.1 with $b = 1$ and (A.17) is a consequence of the choice of our partition in (A.11). The last inequality (A.18) is then satisfied for large enough N since $\sum_{k=1}^{\infty} \frac{(k+3)^8}{\varepsilon^4} \left(\frac{1}{2^k}\right)^{\frac{1}{2}} < \infty$ by the ratio test for the convergence of series and $\sum_{k=1}^{K_N} \frac{(k+3)^8}{\varepsilon^4} \frac{1}{N^\theta} \sim K_N^9 \frac{1}{N^\theta} \to 0$ choosing K_N such that $K_N^9 = o(N^\theta)$.

Now, we consider the second summand in (A.15). Choosing K_N such that $K_N^9 = o(N^\theta)$, but $\frac{K_N}{\log_2(N)} \to \infty$, we get

$$\lim_{\delta \to 0} \limsup_{N \to \infty} \mathrm{P}\left(\sup_{x \in [0, 2\delta]} \bar{m}_N(x_{i_{K_N}(x)}(K_N), a_p + x) > \frac{\varepsilon}{2} \right) \leqslant \lim_{\delta \to 0} \limsup_{N \to \infty} \frac{C}{\varepsilon^4} \max\left\{ \frac{1}{N^{\frac{1}{2}}}, \frac{N}{2^{K_N \frac{1}{2}}} \right\} = 0$$

for all $N \geqslant N_\varepsilon$ by applying Lemma B.2 below with $a = 2\delta$, $b = 1$, and $c = a_p$.

Proof of (A.8). In order to prove (A.8), we first split the interval over t_1, t_2 in (A.8) into subintervals. This allows to bound the quantity of interest in terms of a supremum over a single parameter t in a specific interval. We then apply a similar chaining technique as in the proof of (A.7). Note that here the chaining is applied to $x \in [0, 1]$.

To deal with the supremum over t_1, t_2 in (A.8), define

$$I_{1,p} := [2p\delta, 2(p+1)\delta] \quad \text{and} \quad I_{2,p} := [(2p+1)\delta, (2(p+1)+1)\delta]$$

for $p = 0, \ldots, L_\delta := \lceil \frac{1}{2\delta} - \frac{3}{2} \rceil$. We first note that the expression in P in (A.8) can be bounded through

$$\begin{aligned} \sup_{0 \leqslant x \leqslant 1} \sup_{\substack{|t_2 - t_1| < \delta \\ 0 \leqslant t_1, t_2 \leqslant 1}} |m_N(x, t_2) - m_N(x, t_1)| \leqslant & \sup_{0 \leqslant x \leqslant 1} \max_{0 \leqslant p \leqslant L_\delta} \sup_{t_1, t_2 \in I_{1,p}} |m_N(x, t_2) - m_N(x, t_1)| \\ & + \sup_{0 \leqslant x \leqslant 1} \max_{0 \leqslant p \leqslant L_\delta} \sup_{t_1, t_2 \in I_{2,p}} |m_N(x, t_2) - m_N(x, t_1)|. \end{aligned} \tag{A.19}$$

In the following, we consider only the first summand in (A.19), since for the second summand analogous considerations hold. For this reason, it remains to show that

$$\lim_{\delta \to 0} \limsup_{N \to \infty} \mathrm{P}\left(\sup_{0 \leqslant x \leqslant 1} \max_{0 \leqslant p \leqslant L_\delta} \sup_{t_1, t_2 \in I_{1,p}} |m_N(x, t_2) - m_N(x, t_1)| > \varepsilon \right) = 0.$$

We write $I_{1,p} = [a_p, a_{p+1}]$, i.e., $a_p := 2p\delta$ and $a_{p+1} := 2(p+1)\delta$. Note that

$$\begin{aligned} \sup_{t_1, t_2 \in I_{1,p}} |m_N(x, t_2) - m_N(x, t_1)| \leqslant & \sup_{t_2 \in I_{1,p}} \left| m_N(x, t_2) - m_N\left(x, a_p\right) \right| \\ & + \sup_{t_1 \in I_{1,p}} \left| m_N\left(x, a_p\right) - m_N(x, t_1) \right| \\ \leqslant & 2 \sup_{t \in [0, 2\delta]} \left| m_N\left(x, a_p\right) - m_N\left(x, a_p + t\right) \right|. \end{aligned} \tag{A.20}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

For the supremum over $x \in [0,1]$, we apply a similar chaining technique as in the proof of (A.7). Define refining partitions $x_i(k)$ for $k = 0, \dots, K_N$ with $K_N \to \infty$, for $N \to \infty$, and

$$x_i(k) = \frac{i}{2^k},\ i = 0, \dots, 2^k, \tag{A.21}$$

and choose $i_k(x)$ such that

$$x \in \left(x_{i_k(x)}(k), x_{i_k(x)+1}(k)\right].$$

Moreover, define

$$\begin{aligned} m_N(x,t,p) &:= m_N\left(x, a_p\right) - m_N\left(x, a_p + t\right) \\ m_N(x,y,t,p) &:= m_N(y,t,p) - m_N(x,t,p). \end{aligned} \tag{A.22}$$

Then, continuing with (A.20), it follows that

$$\begin{aligned} &\max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} \left| m_N\left(x, a_p\right) - m_N\left(x, a_p + t\right) \right| \\ &\quad = \max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} |m_N(x,t,p)| \\ &\quad = \max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} |m_N(0,t,p) - m_N(x,t,p)|, \end{aligned}$$

since $c_l(0) = \mathrm{E}\left(\mathbb{1}_{\{F(X_0) \leqslant 0\}} H_l(\xi_0)\right) = 0$. We have

$$\begin{aligned} &\max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} |m_N(0,t,p) - m_N(x,t,p)| \\ &\quad \leqslant \sum_{k=1}^{K_N} \max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} \left| m_N(x_{i_k(x)}(k), x_{i_{k-1}(x)}(k-1), t, p) \right| \\ &\qquad + \max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} \left| m_N(x_{i_{K_N}(x)}(K_N), x, t, p) \right| \\ &\quad =: \sum_{k=1}^{K_N} \bar{m}_N(x_{i_k(x)}(k), x_{i_{k-1}(x)}(k-1)) + \bar{m}_N(x_{i_{K_N}(x)}(K_N), x), \end{aligned} \tag{A.23}$$

where $\bar{m}_N(x,y) := \max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} |m_N(x,y,t,p)|$. Consequently, (A.23) can be used to infer (A.24) below

$$\begin{aligned} &\mathrm{P}\left(\sup_{x \in [0,1]} \max_{0 \leqslant p \leqslant L_\delta} \sup_{t \in [0,2\delta]} |m_N(0,t,p) - m_N(x,t,p)| > \varepsilon \right) \\ &\quad \leqslant \sum_{k=1}^{K_N} \mathrm{P}\left(\sup_{x \in [0,1]} \bar{m}_N(x_{i_k(x)}(k), x_{i_{k-1}(x)}(k-1)) > \frac{\varepsilon}{(k+3)^2} \right) \\ &\qquad + \mathrm{P}\left(\sup_{x \in [0,1]} \bar{m}_N(x_{i_{K_N}(x)}(K_N), x) > \varepsilon - \sum_{k=0}^{\infty} \frac{\varepsilon}{(k+3)^2} \right) \end{aligned} \tag{A.24}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

$$\leqslant \sum_{k=1}^{K_N}\sum_{i=0}^{2^k} \mathrm{P}\left(\bar{m}_N(x_{i+1}(k),x_i(k)) > \frac{\varepsilon}{(k+3)^2}\right) + \mathrm{P}\left(\sup_{x\in[0,1]} \bar{m}_N(x_{i_{K_N}(x)}(K_N),x) > \frac{\varepsilon}{2}\right), \tag{A.25}$$

since $\sum_{k=0}^{\infty} \frac{\varepsilon}{(k+3)^2} \leqslant \frac{\varepsilon}{2}$. We consider the two summands in (A.25) separately. For the first summand in (A.25) we need some preliminary results. Note that for any $\eta > 0$,

$$\mathrm{P}\left(\bar{m}_N(x_{i+1}(k),x_i(k)) > \eta\right) = \mathrm{P}\left(\max_{0\leqslant p\leqslant L_\delta} \sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),x_i(k),t,p)\right| > \eta\right) \leqslant \sum_{p=0}^{L_\delta} \mathrm{P}\left(\sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),x_i(k),t,p)\right| > \eta\right). \tag{A.26}$$

Due to stationarity, it follows that

$$\mathrm{P}\left(\sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),x_i(k),t,p)\right| > \eta\right) = \mathrm{P}\left(\sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),x_i(k),t,0)\right| > \eta\right). \tag{A.27}$$

Combining (A.26) and (A.27), we get

$$\mathrm{P}\left(\bar{m}_N(x_{i+1}(k),x_i(k)) > \eta\right) \leqslant \frac{1}{\delta}\mathrm{P}\left(\sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),x_i(k),t,0)\right| > \eta\right). \tag{A.28}$$

Due to the notation in (A.22), we have

$$\begin{aligned} &m_N(x_{i+1}(k),x_i(k),t,0)\\ &\quad = m_N(x_i(k),t,0) - m_N(x_{i+1}(k),t,0)\\ &\quad = m_N\left(x_i(k),0\right) - m_N\left(x_i(k),t\right) - \left(m_N(x_{i+1}(k),0) - m_N(x_{i+1}(k),t)\right) \end{aligned}$$

such that

$$\sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),x_i(k),t,0)\right| \leqslant 2 \sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),t) - m_N\left(x_i(k),t\right)\right|.$$

We can then bound the first summand in (A.25), with further explanations given below, as follows:

$$\sum_{k=1}^{K_N}\sum_{i=0}^{2^k} \mathrm{P}\left(\bar{m}_N(x_{i+1}(k),x_i(k)) > \frac{\varepsilon}{(k+3)^2}\right) \leqslant \sum_{k=1}^{K_N}\sum_{i=0}^{2^k} \frac{1}{\delta} P\left(2 \sup_{t\in[0,2\delta]} \left|m_N(x_{i+1}(k),t) - m_N\left(x_i(k),t\right)\right| > \frac{\varepsilon}{(k+3)^2}\right) \tag{A.29}$$

$$\leqslant C\frac{1}{\delta}\sum_{k=1}^{K_N}\sum_{i=0}^{2^k} \frac{(k+3)^8}{\varepsilon^4}\left(\delta^{2-\theta}\frac{1}{N^\theta}\left(x_{i+1}(k)-x_i(k)\right) + \delta^2\left(x_{i+1}(k)-x_i(k)\right)^{\frac{3}{2}}\right) \tag{A.30}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

$$\leqslant \frac{1}{\delta} C \sum_{k=1}^{K_N} \sum_{i=0}^{2^k} \frac{(k+3)^8}{\varepsilon^4} \left(\delta^{2-\theta} \frac{1}{N^\theta} \frac{1}{2^k} + \delta^2 \left(\frac{1}{2^k} \right)^{\frac{3}{2}} \right) \tag{A.31}$$

$$\leqslant C \frac{1}{\delta} \delta^{2-\theta} \sum_{k=1}^{K_N} \frac{(k+3)^8}{\varepsilon^4} \frac{1}{N^\theta} + C \frac{1}{\delta} \delta^2 \sum_{k=1}^{K_N} \frac{(k+3)^8}{\varepsilon^4} \left(\frac{1}{2^k} \right)^{\frac{1}{2}}$$

$$\leqslant C \max\{\delta^{1-\theta}, \delta\} \tag{A.32}$$

for sufficiently large N, where (A.29) is due to (A.28) and (A.30) follows from Lemma B.1 with $b = \delta$, (A.31) is a consequence of the choice of our partition in (A.21). The last inequality (A.32) is then satisfied for large enough N since $\sum_{k=1}^{\infty} \frac{(k+3)^8}{\varepsilon^4} \left(\frac{1}{2^k} \right)^{\frac{1}{2}} < \infty$ by the ratio test for the convergence of series and $\sum_{k=1}^{K_N} \frac{(k+3)^8}{\varepsilon^4} \frac{1}{N^\theta} \sim K_N^9 \frac{1}{N^\theta} \to 0$, as $N \to \infty$, choosing K_N such that $K_N^9 = o\left(N^\theta\right)$.

Now, we consider the second summand in (A.25).

$$\begin{aligned}
&\mathrm{P}\left(\sup_{x\in[0,1]} \bar{m}_N(x_{i_{K_N}(x)}(K_N), x) > \frac{\varepsilon}{2} \right) \\
&= \mathrm{P}\left(\sup_{x\in[0,1]} \max_{0\leqslant p\leqslant L_\delta} \sup_{t\in[0,2\delta]} \left| m_N(x_{i_{K_N}(x)}(K_N), x, t, p) \right| > \frac{\varepsilon}{2} \right) \\
&\leqslant \frac{1}{\delta} \mathrm{P}\left(\sup_{x\in[0,1]} \sup_{t\in[0,2\delta]} \left| m_N(x_{i_{K_N}(x)}(K_N), x, t, 0) \right| > \frac{\varepsilon}{2} \right) \\
&\leqslant \frac{1}{\delta} \mathrm{P}\left(2 \sup_{x\in[0,1]} \sup_{t\in[0,2\delta]} \left| m_N(x_{i_{K_N}(x)}(K_N), t) - m_N(x, t) \right| > \frac{\varepsilon}{2} \right).
\end{aligned}$$

Choosing K_N such that $K_N^9 = o\left(N^\theta\right)$, but $\frac{K_N}{\log_2(N)} \to \infty$, we get

$$\frac{1}{\delta} \mathrm{P}\left(2 \sup_{x\in[0,1]} \bar{m}_N(x_{i_{K_N}(x)}(K_N), x) > \frac{\varepsilon}{2} \right) \leqslant \frac{C}{\varepsilon^4} \delta^{\frac{1}{2}} \max\left\{ \frac{1}{N^{\frac{1}{2}}}, \frac{N}{2^{K_N \frac{1}{2}}} \right\} \to 0 \quad \text{as } N \to \infty,$$

for all $N \geqslant N_\varepsilon$ by applying Lemma B.2 below with $a = 1$, $b = \delta$, and $c = 0$. □

B. TECHNICAL RESULTS AND THEIR PROOFS

In this section, we provide some technical results and their proofs.

LEMMA B.1. *Let $\bar{m}_{N,b}$ be as in (A.12). Then, there are constants $C_1, C_2 > 0$ and a $\theta \in (0, \frac{1}{2}]$ such that for any $\lambda > 0$,*

$$\mathrm{P}\left(\bar{m}_{N,b}(x, y) > \lambda \right) \leqslant C_1 \frac{1}{\lambda^4} b^{2-\theta} \frac{1}{N^\theta} (y - x) + C_2 \frac{1}{\lambda^4} b^2 (y - x)^{\frac{3}{2}} \tag{B.1}$$

for any $b > 0$ and all $x, y \in [0, 1]$ with $y > x$.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Proof. In order to bound the probability in (B.1), we use arguments from Billingsley (1968). For this, we express $\bar{m}_{N,b}$ in (A.12) as

$$\bar{m}_{N,b}(x,y) = \sup_{t\in[0,b]} \frac{1}{\sqrt{N}} \left| \sum_{n=1}^{\lfloor Nt \rfloor} \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{c_l(y)-c_l(x)}{l!} H_l(\xi_n) \right|$$
$$= \max_{1\leqslant k\leqslant \lfloor Nb \rfloor} \frac{1}{\sqrt{N}} \left| \sum_{n=1}^{k} \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{c_l(y)-c_l(x)}{l!} H_l(\xi_n) \right| =: \max_{1\leqslant k\leqslant \lfloor Nb \rfloor} |\mathrm{s}_k| .$$

Note that

$$\mathrm{s}_j - \mathrm{s}_i = \frac{1}{\sqrt{N}} \sum_{n=i+1}^{j} \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{c_l(y)-c_l(x)}{l!} H_l(\xi_n)$$

and define

$$h_{x,y}(\xi_n) := \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{c_l(y)-c_l(x)}{l!} H_l(\xi_n) = \mathbb{1}_{\{x<F(G(\xi_n))\leqslant y\}} - (y-x) - \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{c_l(y)-c_l(x)}{l!} H_l(\xi_n).$$

Then,

$$\mathrm{E}\left|\mathrm{s}_j - \mathrm{s}_i\right|^4 = \mathrm{E}\left| \frac{1}{\sqrt{N}} \sum_{n=i+1}^{j} h_{x,y}(\xi_n) \right|^4 \tag{B.2}$$
$$= \frac{1}{N^2} \mathrm{E}\left(\Sigma_1 + 4\Sigma_{21} + 3\Sigma_{22} + 6\Sigma_3 + \Sigma_4\right)$$

with

$$\Sigma_1 := \sum_{n=i+1}^{j} h_{x,y}^4(\xi_n),$$
$$\Sigma_{21} := {\sum}' h_{x,y}^3(\xi_{n_1}) h_{x,y}(\xi_{n_2}),\ \Sigma_{22} := {\sum}' h_{x,y}^2(\xi_{n_1}) h_{x,y}^2(\xi_{n_2}),$$
$$\Sigma_3 := {\sum}' h_{x,y}^2(\xi_{n_1}) h_{x,y}(\xi_{n_2}) h_{x,y}(\xi_{n_3}),$$
$$\Sigma_4 := {\sum}' h_{x,y}(\xi_{n_1}) h_{x,y}(\xi_{n_2}) h_{x,y}(\xi_{n_3}) h_{x,y}(\xi_{n_4}),$$

where $\sum'$ extends over all different indices $i+1 \leqslant n_1, \ldots, n_p \leqslant j$, $n_r \neq n_s$, $r \neq s$, $p = 1, \ldots, 4$.

Note that for any even integer $p \geqslant 2$, there is a constant $C > 0$ such that

$$\mathrm{E}\left(h_{x,y}^p(\xi_0)\right) \leqslant C\,(y-x) \tag{B.3}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

since

$$\mathrm{E}\left(h_{x,y}^{p}(\xi_0)\right) \leqslant C\left(\mathrm{E}\,\mathbb{1}_{\{x<F(G(\xi_0))\leqslant y\}} + (y-x)^p + \mathrm{E}\left(\left|\sum_{l=r}^{\lfloor \frac{1}{D}\rfloor}\left(\frac{c_l(y)-c_l(x)}{l!}\right)H_l(\xi_0)\right|^p\right)\right) \tag{B.4}$$

$$\leqslant C\left(\mathrm{E}\,\mathbb{1}_{\{x<F(G(\xi_0))\leqslant y\}} + (y-x)^p + \sum_{l=r}^{\lfloor \frac{1}{D}\rfloor}\left(\frac{c_l(y)-c_l(x)}{l!}\right)^p \mathrm{E}\left(|H_l(\xi_0)|^p\right)\right) \tag{B.5}$$

$$\leqslant C\left(\mathrm{E}\,\mathbb{1}_{\{x<F(G(\xi_0))\leqslant y\}} + (y-x)^p + \sum_{l=r}^{\lfloor \frac{1}{D}\rfloor}\frac{|c_l(y)-c_l(x)|^p}{(l!)^{\frac{p}{2}}}(p-1)^{\frac{lp}{2}}\right) \tag{B.6}$$

$$\leqslant C\,(y-x), \tag{B.7}$$

where C is a generic constant that can change upon each appearance. Inequalities (B.4) and (B.5) follow from

$$\left(\sum_{k=1}^{n}|x_k|\right)^p \leqslant n^{p-1}\sum_{k=1}^{n}|x_k|^p$$

which holds for any $p \geqslant 1$ and is a direct consequence of Hölder's inequality. Inequality (B.6) follows by Nelson's inequality; see Nourdin and Rosiński (2014, Lem. 2.1). Finally, (B.7) is a consequence of applying the Cauchy–Schwarz inequality

$$(c_l(y)-c_l(x))^2 = \mathrm{E}^2\left(\mathbb{1}_{\{x<F(G(\xi_0))\leqslant y\}}H_l(\xi_0)\right) \leqslant (y-x)\,\mathrm{E}\left(H_l^2(\xi_0)\right) = (y-x)l! \tag{B.8}$$

and by noticing that $(y-x)^{\frac{p}{2}} \leqslant y-x$ for $p \geqslant 2$ and $x, y \in (0,1)$.

We now consider the summands on the right-hand side of formula (B.2) separately. Starting with Σ_1, note that (B.3) gives

$$\mathrm{E}\,|\Sigma_1| \leqslant C(j-i)(y-x).$$

In order to estimate the remaining quantities, we make use of Taqqu (1977, Lem. 4.5). This together with (B.3) above, immediately yields

$$\mathrm{E}\,|\Sigma_{21}| \leqslant C(j-i)^{\frac{3}{2}}\left(\mathrm{E}\left(h_{x,y}^2(\xi_0)\right)\right)^{\frac{1}{2}}\left(\mathrm{E}\left(h_{x,y}^6(\xi_0)\right)\right)^{\frac{1}{2}} \leqslant C(j-i)^{\frac{3}{2}}(y-x),$$

$$\mathrm{E}\,|\Sigma_3| \leqslant C(j-i)^2\,\mathrm{E}\left(h_{x,y}^2(\xi_0)\right)\left(\mathrm{E}\left(h_{x,y}^4(\xi_0)\right)\right)^{\frac{1}{2}} \leqslant C(j-i)^2(y-x)^{\frac{3}{2}},$$

$$\mathrm{E}\,|\Sigma_4| \leqslant C(j-i)^2\,\mathrm{E}^2\left(h_{x,y}^2(\xi_0)\right) \leqslant C(j-i)^2(y-x)^2.$$

It remains to find an upper bound for $\mathrm{E}\,\Sigma_{22}$. For this, define

$$L_{x,y}(\xi_n) := \sum_{l=r}^{\lfloor \frac{1}{D}\rfloor}\frac{c_l(y)-c_l(x)}{l!}H_l(\xi_n) \quad \text{and} \quad \widetilde{h}_{x,y}(z) = \mathbb{1}_{\{x<F(G(z))\leqslant y\}} - (y-x).$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

It then holds that

$$\begin{aligned}\mathrm{E}\left(h_{x,y}^2(\xi_{n_1})h_{x,y}^2(\xi_{n_2})\right) &= \mathrm{E}\left((\widetilde{h}_{x,y}(\xi_{n_1})-L_{x,y}(\xi_{n_1}))^2(\widetilde{h}_{x,y}(\xi_{n_2})-L_{x,y}(\xi_{n_2}))^2\right)\\ &\leqslant 4\,\mathrm{E}\left((\widetilde{h}_{x,y}^2(\xi_{n_1})+L_{x,y}^2(\xi_{n_1}))(\widetilde{h}_{x,y}^2(\xi_{n_2})+L_{x,y}^2(\xi_{n_2}))\right)\\ &= 4\Big(\mathrm{E}(\widetilde{h}_{x,y}^2(\xi_{n_1})\widetilde{h}_{x,y}^2(\xi_{n_2}))+\mathrm{E}(\widetilde{h}_{x,y}^2(\xi_{n_1})L_{x,y}^2(\xi_{n_2}))\\ &\quad+\mathrm{E}(L_{x,y}^2(\xi_{n_1})\widetilde{h}_{x,y}^2(\xi_{n_2}))+\mathrm{E}(L_{x,y}^2(\xi_{n_1})L_{x,y}^2(\xi_{n_2}))\Big). \qquad \textbf{(B.9)}\end{aligned}$$

Before we consider the four summands in (B.9) separately, we make the following observation. With arguments given below,

$$\begin{aligned}\mathrm{E}(L_{x,y}^4(\xi_n)) &= \mathrm{E}\left(\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{c_l(y)-c_l(x)}{l!}H_l(\xi_n)\right)^4\\ &\leqslant C\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\left(\frac{c_l(y)-c_l(x)}{l!}\right)^4\mathrm{E}\,H_l^4(\xi_n) \qquad \textbf{(B.10)}\\ &\leqslant C\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{(c_l(y)-c_l(x))^4}{l!^2}3^{2l}, \qquad \textbf{(B.11)}\end{aligned}$$

where (B.10) follows by Hölder's inequality and (B.11) by Nelson's inequality; see Nourdin and Rosiński (2014) Lemma 2.1. Then, combining (B.11) with (B.8),

$$\mathrm{E}(L_{x,y}^4(\xi_n)) \leqslant C(y-x)^2. \qquad \textbf{(B.12)}$$

In the following, we consider the summands in (B.9) separately. For the last one, the Cauchy–Schwarz inequality and (B.12) yield

$$\mathrm{E}(L_{x,y}^2(\xi_{n_1})L_{x,y}^2(\xi_{n_2})) \leqslant (\mathrm{E}(L_{x,y}^4(\xi_{n_1})))^{\frac{1}{2}}(\mathrm{E}(L_{x,y}^4(\xi_{n_2})))^{\frac{1}{2}} \leqslant C(y-x)^2. \qquad \textbf{(B.13)}$$

Since $\mathrm{E}(\widetilde{h}_{x,y}^2(\xi_{n_1})L_{x,y}^2(\xi_{n_2}))$ and $\mathrm{E}(L_{x,y}^2(\xi_{n_1})\widetilde{h}_{x,y}^2(\xi_{n_2}))$ in (B.9) can be treated analogously, we only consider $\mathrm{E}(\widetilde{h}_{x,y}^2(\xi_{n_1})L_{x,y}^2(\xi_{n_2}))$. Given the definition of $\widetilde{h}_{x,y}^2$ in (??) and with further explanations provided below, we get

$$\begin{aligned}\mathrm{E}(\widetilde{h}_{x,y}^2(\xi_{n_1})L_{x,y}^2(\xi_{n_2})) &= \mathrm{E}((\mathbb{1}_{\{x<F(G(\xi_{n_1}))\leqslant y\}}-(y-x))^2L_{x,y}^2(\xi_{n_2}))\\ &= \mathrm{E}(\mathbb{1}_{\{x<F(G(\xi_{n_1}))\leqslant y\}}L_{x,y}^2(\xi_{n_2}))+(y-x)^2\,\mathrm{E}(L_{x,y}^2(\xi_{n_2}))\\ &\quad+2(y-x)\,\mathrm{E}(\mathbb{1}_{\{x<F(G(\xi_{n_1}))\leqslant y\}}L_{x,y}^2(\xi_{n_2})) \qquad \textbf{(B.14)}\\ &\leqslant C(y-x)^{\frac{3}{2}}+C(y-x)^3+C(y-x)^{\frac{5}{2}}\\ &\leqslant C(y-x)^{\frac{3}{2}}. \qquad \textbf{(B.15)}\end{aligned}$$

For the first and third summand in (B.14), the Cauchy–Schwarz inequality and (B.12) yield

$$\mathrm{E}(\mathbb{1}_{\{x<F(G(\xi_{n_1}))\leqslant y\}}L_{x,y}^2(\xi_{n_2})) \leqslant \left(\mathrm{E}\left(\mathbb{1}_{\{x<F(G(\xi_{n_1}))\leqslant y\}}\right)\right)^{\frac{1}{2}}\left(\mathrm{E}(L_{x,y}^4(\xi_{n_2}))\right)^{\frac{1}{2}} \leqslant C(y-x)^{\frac{3}{2}}.$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Moreover, we have, by orthogonality of the Hermite polynomials, that the second summand in (B.14) can be bounded as

$$\mathrm{E}(L^2_{x,y}(\xi_{n_2})) = \mathrm{E}\left(\sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{c_l(y)-c_l(x)}{l!} H_l(\xi_{n_2})\right)^2 = \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{(c_l(y)-c_l(x))^2}{(l!)^2} \mathrm{E} H_l^2(\xi_{n_2}) \leqslant C(y-x).$$

The first summand in (B.9) requires some more calculations, leading to

$$\begin{aligned}
&\mathrm{E}\left(\widetilde{h}^2_{x,y}(\xi_{n_1})\widetilde{h}^2_{x,y}(\xi_{n_2})\right) \\
&\quad = \mathrm{E}\left[\left(\mathbb{1}_{\{x<F(X_{n_1})\leqslant y\}} - (y-x)\right)^2 \left(\mathbb{1}_{\{x<F(X_{n_2})\leqslant y\}} - (y-x)\right)^2\right] \\
&\quad = \mathrm{E}\Big[\left(\mathbb{1}_{\{x<F(X_{n_1})\leqslant y\}} + (y-x)^2 - 2(y-x)\mathbb{1}_{\{x<F(X_{n_1})\leqslant y\}}\right) \\
&\qquad \times \left(\mathbb{1}_{\{x<F(X_{n_2})\leqslant y\}} + (y-x)^2 - 2(y-x)\mathbb{1}_{\{x<F(X_{n_2})\leqslant y\}}\right)\Big] \\
&\quad = \mathrm{E}\left(\mathbb{1}_{\{x<F(X_{n_1})\leqslant y\}}\mathbb{1}_{\{x<F(X_{n_2})\leqslant y\}}\right)(1-2(y-x))^2 + 2(y-x)^3 - 3(y-x)^4 \\
&\quad = \mathrm{E}\left(\widetilde{h}_{x,y}(\xi_{n_1})\widetilde{h}_{x,y}(\xi_{n_2})\right)(1-2(y-x))^2 + (y-x)^2(1-(y-x))^2 \\
&\quad \leqslant \left|\mathrm{E}\left(\widetilde{h}_{x,y}(\xi_{n_1})\widetilde{h}_{x,y}(\xi_{n_2})\right)\right| + (y-x)^2.
\end{aligned}$$

By orthogonality of the Hermite expansion

$$\begin{aligned}
\mathrm{E}\left(\widetilde{h}_{x,y}(\xi_{n_1})\widetilde{h}_{x,y}(\xi_{n_2})\right) &= \mathrm{E}(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2})) + \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \left(\frac{c_l(y)-c_l(x)}{l!}\right)^2 \mathrm{E}\left(H_l(\xi_{n_1})H_l(\xi_{n_2})\right) \\
&= \mathrm{E}(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2})) + \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{(c_l(y)-c_l(x))^2}{l!}\gamma^l(n_1-n_2) \\
&\leqslant \mathrm{E}(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2})) + (y-x)\sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \gamma^l(n_1-n_2) \\
&\leqslant \mathrm{E}(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2})) + (y-x)C\gamma^r(n_1-n_2).
\end{aligned}$$

Then,

$$\left|\mathrm{E}\left(\widetilde{h}_{x,y}(\xi_{n_1})\widetilde{h}_{x,y}(\xi_{n_2})\right)\right| \leqslant \left|\mathrm{E}(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2}))\right| + C(y-x)\gamma^r(n_1-n_2),$$

such that

$$\mathrm{E}\left(\widetilde{h}^2_{x,y}(\xi_{n_1})\widetilde{h}^2_{x,y}(\xi_{n_2})\right) \leqslant \left|\mathrm{E}(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2}))\right| + C(y-x)\gamma^r(n_1-n_2). \tag{B.16}$$

Combining (B.9), (B.13), (B.15), and (B.16) finally gives

$$\begin{aligned}
\mathrm{E}\left(h^2_{x,y}(\xi_{n_1})h^2_{x,y}(\xi_{n_2})\right) &\leqslant C\left|\mathrm{E}h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2})\right| \\
&\quad + C(y-x)^2 + C(y-x)^{\frac{3}{2}} + C(y-x)\gamma^r(n_1-n_2).
\end{aligned}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Taqqu (1977, Lem. 4.5) yields

$$\sum\nolimits' \left|\mathrm{E}\left(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2})\right)\right| \leqslant C(j-i)(y-x).$$

Since $\gamma(k) = k^{-D}L(k)$,

$$\begin{aligned}\sum_{i+1\leqslant n_1\neq n_2\leqslant j} \gamma^r(n_1-n_2) &= \sum_{n_1=1}^{j-i}\sum_{n_2=1}^{j-i}\gamma^r(n_1-n_2) - (j-i)\gamma^r(0)\\ &= 2\sum_{k=1}^{j-i-1}(j-i-k)\gamma^r(k)\\ &\sim \frac{2}{1-rD}(j-i)^{2-rD}L^r(j-i);\end{aligned}$$

see Pipiras and Taqqu (2017, Prop. 2.2.1). Therefore, for any $\epsilon > 0$, there exists an $N_0 \in \mathbb{N}$ such that

$$\sum_{i+1\leqslant n_1\neq n_2\leqslant j} \gamma^r(n_1-n_2) \leqslant (1+\epsilon)\frac{2}{1-rD}(j-i)^{2-rD}L^r(j-i)$$

for $j-i \geqslant N_0$. As a consequence thereof, there exists a constant $C > 0$ such that

$$\sum_{i+1\leqslant n_1\neq n_2\leqslant j} \gamma^r(n_1-n_2) \leqslant C(j-i)^{2-rD}L^r(j-i)$$

for all $i,j \in \mathbb{N}$.

Due to slow variation of L for any $\eta > 0$, there exists a $C > 0$, such that

$$\begin{aligned}\mathrm{E}\,|\Sigma_{22}| &\leqslant C\sum\nolimits' \left|\mathrm{E}\left(h_{x,y}(\xi_{n_1})h_{x,y}(\xi_{n_2})\right)\right| + C(j-i)^2(y-x)^2\\ &\quad + C(j-i)^2(y-x)^{\frac{3}{2}} + C(y-x)(j-i)^{2-rD+\eta}\\ &\leqslant C(j-i)^2(y-x)^{\frac{3}{2}} + C(j-i)(y-x) + C(j-i)^{2-rD+\eta}(y-x).\end{aligned}$$

Finally, we can use the bounds on $\mathrm{E}\,\Sigma_1$ to $\mathrm{E}\,\Sigma_4$ to continue bounding (B.2) as follows:

$$\begin{aligned}\mathrm{E}\left|s_j - s_i\right|^4 &= \frac{1}{N^2}\mathrm{E}\,(\Sigma_1 + 4\Sigma_{21} + 3\Sigma_{22} + 6\Sigma_3 + \Sigma_4)\\ &\leqslant C\frac{1}{N^2}\Big((j-i)(y-x) + (j-i)^{\frac{3}{2}}(y-x)\\ &\quad + (j-i)(y-x) + (j-i)^{2-rD+\eta}(y-x) + (j-i)^2(y-x)^{\frac{3}{2}} + (j-i)^2(y-x)^2\Big)\\ &\leqslant C\frac{1}{N^2}((j-i)^{\frac{3}{2}}(y-x) + (j-i)^{2-rD+\eta}(y-x) + (j-i)^2(y-x)^{\frac{3}{2}})\\ &\leqslant C\frac{1}{N^2}((j-i)^{2-\theta}(y-x) + (j-i)^2(y-x)^{\frac{3}{2}})\end{aligned}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

with $\theta = \min\{\frac{1}{2}, rD - \eta\}$. We obtain,

$$\begin{aligned} \mathrm{E}\left|\mathrm{s}_j - \mathrm{s}_i\right|^4 &\leqslant C\frac{1}{N^\theta}\left(\frac{j-i}{N}\right)^{2-\theta}(y-x) + C\left(\frac{j-i}{N}\right)^2 (y-x)^{\frac{3}{2}} \\ &\leqslant C\left((y-x)^{\frac{1}{2-\theta}}\sum_{q=i+1}^{j}\frac{1}{N^{\frac{2}{2-\theta}}}\right)^{2-\theta} + C\left((y-x)^{\frac{3}{4}}\sum_{q=i+1}^{j}\frac{1}{N}\right)^2. \end{aligned} \tag{B.17}$$

Applying Markov's inequality to the probability $\mathrm{P}(|\mathrm{s}_j - \mathrm{s}_i| > \lambda) = \mathrm{P}(|\mathrm{s}_j - \mathrm{s}_i|^4 > \lambda^4)$ and using the bound (B.17), the conditions of Lemma C.1 are satisfied with $\gamma = 4$, $\alpha_1 = 2-\theta, \alpha_2 = 2$, $v_1 = (y-x)^{\frac{1}{2-\theta}}\frac{1}{N^{\frac{2}{2-\theta}}}$, and $v_2 = (y-x)^{\frac{3}{4}}\frac{1}{N}$. Then, applying Lemma C.1 yields

$$\begin{aligned} \mathrm{P}\left(\max_{1\leqslant k\leqslant \lfloor Nb\rfloor}|\mathrm{s}_k| > \lambda\right) &\leqslant C_{\alpha_1,\gamma}\frac{1}{\lambda^4}\left((y-x)^{\frac{1}{2-\theta}}\sum_{n=1}^{\lfloor Nb\rfloor}\frac{1}{N^{\frac{2}{2-\theta}}}\right)^{2-\theta} \\ &\quad + C_{\alpha_2,\gamma}\frac{1}{\lambda^4}\left((y-x)^{\frac{3}{4}}\sum_{n=1}^{\lfloor Nb\rfloor}\frac{1}{N}\right)^2 \\ &\leqslant C_{\alpha_1,\gamma}\frac{1}{N^\theta}\frac{1}{\lambda^4}b^{2-\theta}(y-x) + C_{\alpha_2,\gamma}\frac{1}{\lambda^4}b^2(y-x)^{\frac{3}{2}}. \end{aligned}$$

□

The subsequent lemmas are used in Appendix A and are all formulated in terms of a generic sequence of refining partitions which covers the two sequences of partitions (A.11) and (A.21) in Appendix A. For $k = 0, \ldots, K_N$ define refining partitions

$$x_i(k) := \widetilde{a}_p + \frac{i}{2^k}2\delta, \quad i = 0, \ldots, 2^k, \tag{B.18}$$

of the interval $[\widetilde{a}_p, \widetilde{a}_p + a]$ and for $x \in [0, a]$ choose $i_k(x)$ such that

$$\widetilde{a}_p + x \in \left(x_{i_k(x)}(k), x_{i_k(x)+1}(k)\right].$$

Note that for the partitions defined in (A.11) and (A.21), we consider $\widetilde{a}_p = a_p$, $a = 2\delta$ and $\widetilde{a}_p = 0$, $a = 1$, respectively. All following lemmas in this section refer to these partitions.

LEMMA B.2. *Let m_N be defined as in (4.3) and let $\sqrt{N}/2^{K_N} \to 0$. Then, for all $\lambda, b \in (0, 1]$ there is an N_λ and a constant $C > 0$ such that*

$$\mathrm{P}\left(\sup_{x\in[0,a]}\sup_{t\in[0,b]}\left|m_N(x_{i_{K_N}(x)}(K_N), t) - m_N(x+c, t)\right| > \lambda\right) \leqslant \frac{C}{\lambda^4}b^{\frac{3}{2}}\max\left\{\frac{1}{N^{\frac{1}{2}}}, \frac{N}{2^{K_N\frac{1}{2}}}\right\}$$

for all $N \geqslant N_\lambda$ and for all $c \geqslant 0$ such that $x + c \leqslant x_{i_{K_N}(x)+1}(K_N)$.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Proof. Note first that for all x, y,

$$
\begin{aligned}
|m_N(y,t) - m_N(x,t)| &= \left| \sum_{l=\lceil \frac{1}{D} \rceil}^{\infty} \frac{c_l(y) - c_l(x)}{l!} \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} H_l(\xi_n) \right| \\
&= \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \mathbb{1}_{\{x < F(G(\xi_n)) \leqslant y\}} - \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{c_l(y) - c_l(x)}{l!} \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} H_l(\xi_n) \right| \\
&\leqslant \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \mathbb{1}_{\{x < F(G(\xi_n)) \leqslant y\}} + \left| \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{c_l(y) - c_l(x)}{l!} \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} H_l(\xi_n) \right|.
\end{aligned}
\tag{B.19}
$$

Then,

$$
\begin{aligned}
&\mathrm{P}\left(\sup_{x \in [0,a]} \sup_{t \in [0,b]} \left| m_N(x+c,t) - m_N(x_{i_{K_N}(x)}(K_N), t) \right| > \lambda \right) \\
&\quad \leqslant \mathrm{P}\left(\sup_{x \in [0,a]} \sup_{t \in [0,b]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \mathbb{1}_{\{x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x+c\}} \right| > \frac{\lambda}{2} \right) \\
&\qquad + \mathrm{P}\left(\sup_{x \in [0,a]} \sup_{t \in [0,b]} \left| \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{c_l(x+c) - c_l(x_{i_{K_N}(x)}(K_N))}{l!} \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} H_l(\xi_n) \right| > \frac{\lambda}{2} \right)
\end{aligned}
\tag{B.20}
$$

$$
\leqslant \mathrm{P}\left(\sup_{x \in [0,a]} \sup_{t \in [0,b]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \mathbb{1}_{\{x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x+c\}} \right| > \frac{\lambda}{2} \right) + \frac{4}{D^2} \frac{Nb^2}{2^{K_N} \lambda^2},
\tag{B.21}
$$

where (B.20) follows by (B.19) and (B.21) is a consequence of applying Lemma B.3 below. It remains to bound the probability in (B.21). Therefore, we write

$$
\begin{aligned}
&\mathrm{P}\left(\sup_{x \in [0,a]} \sup_{t \in [0,b]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \mathbb{1}_{\left\{x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x+c\right\}} \right| > \frac{\lambda}{2} \right) \\
&\quad \leqslant \mathrm{P}\left(\sup_{x \in [0,a]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nb \rfloor} \mathbb{1}_{\left\{x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x+c\right\}} \right| > \frac{\lambda}{2} \right) \\
&\quad \leqslant \mathrm{P}\left(\sup_{x \in [0,a]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nb \rfloor} \mathbb{1}_{\left\{x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x_{i_{K_N}(x)+1}(K_N)\right\}} \right| > \frac{\lambda}{2} \right)
\end{aligned}
$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

$$\leqslant \mathrm{P}\Bigg(\sup_{x\in[0,a]} \Bigg| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nb \rfloor} \Bigg(\mathbb{1}_{\left\{ x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x_{i_{K_N}(x)+1}(K_N) \right\}} - \left(x_{i_{K_N}(x)+1}(K_N) - x_{i_{K_N}(x)}(K_N) \right) \Bigg) \Bigg| + \sup_{x\in[0,a]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nb \rfloor} \left(x_{i_{K_N}(x)+1}(K_N) - x_{i_{K_N}(x)}(K_N) \right) \right| > \frac{\lambda}{2} \Bigg). \tag{B.22}$$

The second summand in (B.22) is deterministic and can be bounded by

$$\sup_{x\in[0,a]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nb \rfloor} \left(x_{i_{K_N}(x)+1}(K_N) - x_{i_{K_N}(x)}(K_N) \right) \right| \leqslant \frac{\sqrt{N}}{2^{K_N}} b \leqslant \frac{\sqrt{N}}{2^{K_N}}.$$

Then, choose $N_\lambda \geqslant 1$ such that $\frac{\sqrt{N}}{2^{K_N}} < \frac{\lambda}{4}$ for all $N \geqslant N_\lambda$. As a result, we get

$$\begin{aligned}
&\mathrm{P}\left(\sup_{x\in[0,a]} \sup_{t\in[0,b]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \mathbb{1}_{\left\{ x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x+c \right\}} \right| > \frac{\lambda}{2} \right) \\
&\leqslant \mathrm{P}\Bigg(\sup_{x\in[0,a]} \Bigg| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nb \rfloor} \Bigg(\mathbb{1}_{\left\{ x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x_{i_{K_N}(x)+1}(K_N) \right\}} - \left(x_{i_{K_N}(x)+1}(K_N) - x_{i_{K_N}(x)}(K_N) \right) \Bigg) \Bigg| > \frac{\lambda}{2} - \frac{\sqrt{N}}{2^{K_N}} \Bigg) \\
&\leqslant \mathrm{P}\Bigg(\sup_{x\in[0,a]} \Bigg| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nb \rfloor} \Bigg(\mathbb{1}_{\left\{ x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x_{i_{K_N}(x)+1}(K_N) \right\}} - \left(x_{i_{K_N}(x)+1}(K_N) - x_{i_{K_N}(x)}(K_N) \right) \Bigg) \Bigg| > \frac{\lambda}{4} \Bigg)
\end{aligned} \tag{B.23}$$

for all $N \geqslant N_\lambda$. Lemma B.4 gives an upper bound on the probability in (B.23):

$$\mathrm{P}\left(\sup_{x\in[0,a]} \sup_{t\in[0,b]} \left| \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} \mathbb{1}_{\left\{ x_{i_{K_N}(x)}(K_N) < F(G(\xi_n)) \leqslant x+c \right\}} \right| > \frac{\lambda}{2} \right) \leqslant \frac{C}{\lambda^4} b^{\frac{3}{2}} \max\left\{ \frac{1}{N^{\frac{1}{2}}}, \frac{N}{2^{K_N \frac{1}{2}}} \right\}.$$

□

LEMMA B.3. *Let $c_l(\cdot)$ be defined as in (4.3). For all $\lambda > 0$, it holds that*

$$\mathrm{P}\left(\sup_{x\in[0,a]} \sup_{t\in[0,b]} \left| \sum_{l=r}^{\lfloor \frac{1}{D} \rfloor} \frac{c_l(x+c) - c_l(x_{i_{K_N}(x)}(K_N))}{l!} \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nt \rfloor} H_l(\xi_n) \right| > \lambda \right) \leqslant \frac{1}{D^2} \frac{Nb^2}{2^{K_N} \lambda^2}$$

for $a, b > 0$ and $c \geqslant 0$, such that $x + c \leqslant x_{i_{K_N}(x)+1}(K_N)$.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Proof. In order to bound the probability of interest, we use (B.8) in (B.24) below

$$\begin{aligned}
&\mathrm{P}\left(\sup_{x\in[0,a]}\sup_{t\in[0,b]}\left|\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{c_l(x+c)-c_l(x_{i_{K_N}(x)}(K_N))}{l!}\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nt\rfloor}H_l(\xi_n)\right|>\lambda\right)\\
&\leqslant \mathrm{P}\left(\sup_{x\in[0,a]}\sup_{t\in[0,b]}\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{(x+c-x_{i_{K_N}(x)}(K_N))^{\frac{1}{2}}}{l!}\sqrt{l!}\left|\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nt\rfloor}H_l(\xi_n)\right|>\lambda\right) \qquad \textbf{(B.24)}\\
&\leqslant \mathrm{P}\left(\sup_{x\in[0,a]}\sup_{t\in[0,b]}\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{(x_{i_{K_N}(x)+1}(K_N)-x_{i_{K_N}(x)}(K_N))^{\frac{1}{2}}}{\sqrt{l!}}\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nt\rfloor}|H_l(\xi_n)|>\lambda\right)\\
&\leqslant \mathrm{P}\left(\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{1}{\sqrt{2^{K_N}l!}}\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nb\rfloor}|H_l(\xi_n)|>\lambda\right)\\
&\leqslant \frac{1}{2^{K_N}}\mathrm{E}\left(\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{1}{\sqrt{l!}}\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nb\rfloor}|H_l(\xi_n)|\right)^2\left(\frac{1}{\lambda}\right)^2 \qquad \textbf{(B.25)}\\
&\leqslant \frac{1}{2^{K_N}}\sum_{l_1,l_2=r}^{\lfloor\frac{1}{D}\rfloor}\frac{1}{\sqrt{l_1!\,l_2!}}\frac{1}{N}\sum_{n_1,n_2=1}^{\lfloor Nb\rfloor}\mathrm{E}\left(|H_{l_1}(\xi_{n_1})||H_{l_2}(\xi_{n_2})|\right)\left(\frac{1}{\lambda}\right)^2\\
&\leqslant \frac{1}{2^{K_N}}\sum_{l_1,l_2=r}^{\lfloor\frac{1}{D}\rfloor}\frac{1}{N}\sum_{n_1,n_2=1}^{\lfloor Nb\rfloor}\left(\frac{1}{\lambda}\right)^2 \qquad \textbf{(B.26)}\\
&\leqslant \frac{1}{D^2}\frac{Nb^2}{2^{K_N}\lambda^2},
\end{aligned}$$

where (B.25) follows by Markov's inequality. We then used the Cauchy–Schwarz inequality to get (B.26). □

LEMMA B.4. *Let F denote the marginal distribution function of $X_n, n\in\mathbb{N}$. Then, there is a constant $C>0$ such that*

$$\begin{aligned}
&\mathrm{P}\left(\sup_{x\in[0,a]}\left|\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nb\rfloor}\left(\mathbb{1}_{\left\{x_{i_{K_N}(x)}(K_N)<F(X_n)\leqslant x_{i_{K_N}(x)+1}(K_N)\right\}}\right.\right.\right.\\
&\qquad\left.\left.\left.-\left(x_{i_{K_N}(x)+1}(K_N)-x_{i_{K_N}(x)}(K_N)\right)\right)\right|>\lambda\right)\\
&\leqslant \frac{C}{\lambda^4}b^{\frac{3}{2}}\max\left\{\frac{1}{N^{\frac{1}{2}}},\frac{N}{2^{K_N\frac{1}{2}}}\right\}
\end{aligned}$$

for $a>0$ and $b,\lambda\in(0,1]$.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Proof. With further explanations given below, we can infer the following bounds:

$$\begin{aligned}
&\mathrm{P}\left(\sup_{x\in[0,a]}\left|\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nb\rfloor}\left(\mathbb{1}_{\left\{x_{i_{K_N}(x)}(K_N)<F(X_n)\leqslant x_{i_{K_N}(x)+1}(K_N)\right\}}\right.\right.\right.\\
&\qquad\left.\left.\left.-\left(x_{i_{K_N}(x)+1}(K_N)-x_{i_{K_N}(x)}(K_N)\right)\right)\right|>\lambda\right)\\
&\leqslant \mathrm{P}\left(\sup_{x\in[0,a]}\left|\sum_{l=r}^{\infty}\frac{c_l(x_{i_{K_N}(x)+1}(K_N))-c_l(x_{i_{K_N}(x)}(K_N))}{l!}\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nb\rfloor}H_l(\xi_n)\right|>\lambda\right)\\
&\leqslant \sum_{i=0}^{2^{K_N}}\mathrm{P}\left(\left|m_N\left(x_{i+1}(K_N),b\right)-m_N\left(x_i(K_N),b\right)\right|>\frac{\lambda}{2}\right)\\
&\quad+\mathrm{P}\left(\sup_{x\in[0,a]}\left|\sum_{l=r}^{\lfloor\frac{1}{D}\rfloor}\frac{c_l(x_{i_{K_N}(x)+1}(K_N))-c_l(x_{i_{K_N}(x)}(K_N))}{l!}\frac{1}{\sqrt{N}}\sum_{n=1}^{\lfloor Nb\rfloor}H_l(\xi_n)\right|>\frac{\lambda}{2}\right) \qquad \textbf{(B.27)}\\
&\leqslant \frac{16C_1}{\lambda^4}b^{\frac{3}{2}}\sum_{i=0}^{2^{K_N}}\frac{1}{N^{\frac{1}{2}}}\left(x_{i+1}(K_N)-x_i(K_N)\right)+\frac{16C_2}{\lambda^4}b^2\sum_{i=0}^{2^{K_N}}\left(x_{i+1}(K_N)-x_i(K_N)\right)^{\frac{3}{2}}+\frac{4}{D^2}\frac{Nb^2}{2^{K_N}\lambda^2} \qquad \textbf{(B.28)}\\
&\leqslant \frac{16C_1}{\lambda^4}b^{\frac{3}{2}}\sum_{i=0}^{2^{K_N}}\left(\frac{1}{2^{K_N}}\right)\frac{1}{N^{\frac{1}{2}}}+\frac{16C_2}{\lambda^4}b^2\sum_{i=0}^{2^{K_N}}\left(\frac{1}{2^{K_N}}\right)^{\frac{3}{2}}+\frac{4}{D^2}\frac{Nb^2}{2^{K_N}\lambda^2}\\
&\leqslant \frac{C}{\lambda^4}b^{\frac{3}{2}}\max\left\{\frac{1}{N^{\frac{1}{2}}},\frac{N}{2^{K_N\frac{1}{2}}}\right\},
\end{aligned}$$

where we used the representation (4.3) in the first summand of (B.27). The first probability in (B.28) can be bounded by Lemma B.1 and the second one by Lemma B.3. We deduce the last inequality by using that $b,\lambda\in(0,1]$. □

C. A COMPLEMENTARY RESULT AND ITS PROOF

In order to prove Lemma B.1, we use a slightly modified version of Billingsley (1968, Thm. 12.2). We recall some notation from Billingsley (1968, Ch. 12). Let $\xi_1,\dots,\xi_N$ be independent or identically distributed random variables and $\mathrm{s}_k=\sum_{j=1}^{k}\xi_j$ with $\mathrm{s}_0=0$ and set

$$M_N=\max_{0\leqslant k\leqslant N}|\,\mathrm{s}_k\,|.$$

Lemma C.1. *Suppose there are $\gamma>0$, $\alpha_1,\alpha_2>1$, $v_1,v_2>0$ and a positive sequence $(u_\ell)_{1\leqslant\ell\leqslant N}$, such that for all $\lambda>0$,*

$$\mathrm{P}\left(|\mathrm{s}_j-\mathrm{s}_i|>\lambda\right)\leqslant\frac{1}{\lambda^\gamma}\left(\left(v_1\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_1}+\left(v_2\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_2}\right),\quad 0\leqslant i\leqslant j\leqslant N. \qquad \textbf{(C.1)}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Then, there are constants $C_{\alpha_1,\gamma}, C_{\alpha_2,\gamma} > 0$ only depending on α_i, $i = 1,2$ and γ, such that

$$\mathrm{P}(M_N > \lambda) \leqslant \frac{C_{\alpha_1,\gamma}}{\lambda^\gamma}\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \frac{C_{\alpha_2,\gamma}}{\lambda^\gamma}\left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}.$$

Remark C.2. Note that Lemma C.1 reduces to Billingsley (1968, Thm. 12.2) by setting $v_2 = 0$. In particular, our statement remains true if either v_1 or v_2 are zero. Our proof below reveals that the generalization only works when the two summands in (C.1) depend on the same sequence $(u_\ell)_{1\leqslant\ell\leqslant N}$.

Proof. The proof follows the proofs of Billingsley (1968, Thms. 12.1 and 12.2) and requires only slight modifications of the arguments. First, note that

$$\mathrm{P}(M_N > \lambda) \leqslant \mathrm{P}(M'_N > \lambda/2) + \mathrm{P}(\mathrm{s}_N > \lambda/2) \tag{C.2}$$

with $M'_N = \max_{0\leqslant i\leqslant N}\min\{|\,\mathrm{s}_i\,|, |\,\mathrm{s}_N - \mathrm{s}_i\,|\}$. We consider the two probabilities in (C.2) separately. Using assumption (C.1) with $j = N$ and $i = 0$, the second probability can be bounded as

$$\mathrm{P}(\mathrm{s}_N > \lambda/2) \leqslant \frac{2^\gamma}{\lambda^\gamma}\left(\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right).$$

We prove a bound for the first probability in (C.2) via induction over N. Our induction hypothesis is, for $\mu = \lambda/2$,

$$\mathrm{P}(M'_N > \mu) \leqslant \frac{C_{\alpha_1,\gamma}}{\mu^\gamma}\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \frac{C_{\alpha_2,\gamma}}{\mu^\gamma}\left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}. \tag{C.3}$$

Beginning the induction with the base case $N = 2$, we get

$$\mathrm{P}(M'_2 > \mu) = \mathrm{P}(\min\{|\,\mathrm{s}_1\,|, |\,\mathrm{s}_2 - \mathrm{s}_1\,|\} > \mu) \leqslant \frac{1}{\mu^\gamma}\left(\left(v_1\sum_{\ell=1}^{2} u_\ell\right)^{\alpha_1} + \left(v_2\sum_{\ell=1}^{2} u_\ell\right)^{\alpha_2}\right) \tag{C.4}$$

by applying Lemma C.3 with $i = 0, j = 1, k = 2$ in (C.4).

For the inductive step, we assume that the induction hypothesis (C.3) is satisfied for all integers smaller and equal to $N-1$ and move toward N during the inductive step. Note that by Billingsley (1968, eqn. (12.29)), there is an h such that

$$\sum_{\ell=1}^{h-1} u_\ell \leqslant \frac{1}{2}\sum_{\ell=1}^{N} u_\ell \leqslant \sum_{\ell=1}^{h} u_\ell, \tag{C.5}$$

where the sum on the left is zero if $h = 1$. By algebraic computations, one can infer

$$\sum_{\ell=h+1}^{N} u_\ell \leqslant \frac{1}{2}\sum_{\ell=1}^{N} u_\ell \leqslant \sum_{\ell=h}^{N} u_\ell. \tag{C.6}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

To see this, note that

$$\begin{aligned}\frac{1}{2}\sum_{\ell=1}^{N} u_\ell \leqslant \sum_{\ell=1}^{h} u_\ell &\Rightarrow \frac{1}{2}\sum_{\ell=1}^{N} u_\ell + \sum_{\ell=h+1}^{N} u_\ell \leqslant \sum_{\ell=1}^{h} u_\ell + \sum_{\ell=h+1}^{N} u_\ell \\ &\Rightarrow \sum_{\ell=h+1}^{N} u_\ell \leqslant \sum_{\ell=1}^{N} u_\ell - \frac{1}{2}\sum_{\ell=1}^{N} u_\ell \\ &\Rightarrow \sum_{\ell=h+1}^{N} u_\ell \leqslant \frac{1}{2}\sum_{\ell=1}^{N} u_\ell. \end{aligned} \tag{C.7}$$

Using the first inequality of (C.5), we get

$$\begin{aligned}\sum_{\ell=1}^{h-1} u_\ell \leqslant \frac{1}{2}\sum_{\ell=1}^{N} u_\ell &\Rightarrow \sum_{\ell=1}^{h-1} u_\ell + \sum_{\ell=h}^{N} u_\ell \leqslant \frac{1}{2}\sum_{\ell=1}^{N} u_\ell + \sum_{\ell=h}^{N} u_\ell \\ &\Rightarrow \sum_{\ell=1}^{N} u_\ell - \frac{1}{2}\sum_{\ell=1}^{N} u_\ell \leqslant \sum_{\ell=h}^{N} u_\ell \\ &\Rightarrow \frac{1}{2}\sum_{\ell=1}^{N} u_\ell \leqslant \sum_{\ell=h}^{N} u_\ell. \end{aligned} \tag{C.8}$$

Combining (C.7) and (C.8), we get the desired result.

By Billingsley (1968, eqn. (12.36)),

$$M_N' \leqslant \max\{U_1 + D_1, U_2 + D_2\}$$

and therefore

$$\mathrm{P}(M_N' > \mu) \leqslant \mathrm{P}(U_1 + D_1 > \mu) + \mathrm{P}(U_2 + D_2 > \mu) \tag{C.9}$$

with

$$U_1 = \max_{0 \leqslant i \leqslant h-1} \min\{|\mathrm{s}_i|, |\mathrm{s}_{h-1} - \mathrm{s}_i|\}, \quad U_2 = \max_{h \leqslant i \leqslant N} \min\{|\mathrm{s}_j - \mathrm{s}_h|, |\mathrm{s}_N - \mathrm{s}_j|\}, \tag{C.10}$$

$$D_1 = \min\{|\mathrm{s}_{h-1}|, |\mathrm{s}_N - \mathrm{s}_{h-1}|\}, \quad D_2 = \min\{|\mathrm{s}_h|, |\mathrm{s}_N - \mathrm{s}_h|\}. \tag{C.11}$$

The tail probabilities of the random variables (C.10) and (C.11) can be bounded by using the inductive hypothesis (C.3) and Lemma C.3, respectively. Exemplarily, we consider U_1 and D_1. For U_1, we get the following bounds:

$$\mathrm{P}(U_1 > \mu) \leqslant \frac{C_{\alpha_1,\gamma}}{\mu^\gamma}\left(v_1 \sum_{\ell=1}^{h-1} u_\ell\right)^{\alpha_1} + \frac{C_{\alpha_2,\gamma}}{\mu^\gamma}\left(v_2 \sum_{\ell=1}^{h-1} u_\ell\right)^{\alpha_2} \tag{C.12}$$

$$\leqslant \frac{C_{\alpha_1,\gamma}}{\mu^\gamma}\frac{1}{2^{\alpha_1}}\left(v_1 \sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \frac{C_{\alpha_2,\gamma}}{\mu^\gamma}\frac{1}{2^{\alpha_2}}\left(v_2 \sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2} \tag{C.13}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

by applying the inductive hypothesis (C.3) in (C.12) and the inequality (C.5) in (C.13). The tail probability of U_2 can be dealt with analogously by applying (C.6). For D_1, we get

$$\mathrm{P}(D_1 > \mu) \leqslant \frac{1}{\mu^\gamma}\left(\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right) \tag{C.14}$$

by Lemma C.3 with $i = 0, j = h-1,$ and $k = N$. The tail probability of D_2 can be handled analogously by applying Lemma C.3 with $i = 0, j = h$ and $k = N$.

We now continue with bounding (C.9) with focus on the first summand since the second summand can be bounded by analogous arguments. With explanations given below, for some positive μ_0, μ_1 with $\mu_0 + \mu_1 = \mu$,

$$\begin{aligned}
&\mathrm{P}(U_1 + D_1 > \mu)\\
&\quad\leqslant \mathrm{P}(U_1 > \mu_0) + \mathrm{P}(D_1 > \mu_1)\\
&\quad\leqslant \frac{C_{\alpha_1,\gamma}}{\mu_0^\gamma}\frac{1}{2^{\alpha_1}}\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \frac{C_{\alpha_2,\gamma}}{\mu_0^\gamma}\frac{1}{2^{\alpha_2}}\left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2} + \frac{1}{\mu_1^\gamma}\left(\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right) \qquad \textbf{(C.15)}\\
&\quad= \frac{1}{\mu^\gamma}\left(\left(\frac{C_{\alpha_1,\gamma}}{2^{\alpha_1}}\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \frac{C_{\alpha_2,\gamma}}{2^{\alpha_2}}\left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right)^{\delta} + \left(\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right)^{\delta}\right)^{\frac{1}{\delta}} \qquad \textbf{(C.16)}\\
&\quad\leqslant \frac{1}{\mu^\gamma}\left(\left(C_{\alpha_1,\gamma}\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + C_{\alpha_2,\gamma}\left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right)^{\delta}\left[\left(\frac{1}{2^{\alpha_1}} + \frac{1}{2^{\alpha_2}}\right)^{\delta} + \left(\frac{1}{C_{\alpha_1,\gamma}} + \frac{1}{C_{\alpha_2,\gamma}}\right)^{\delta}\right]\right)^{\frac{1}{\delta}} \qquad \textbf{(C.17)}\\
&\quad\leqslant \frac{1}{\mu^\gamma}\left(C_{\alpha_1,\gamma}\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + C_{\alpha_2,\gamma}\left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right), \qquad \textbf{(C.18)}
\end{aligned}$$

where (C.15) follows by (C.13) and (C.14). For (C.16), we recall Billingsley (1968, eqn. (12.39)). It states that for positive numbers A, B, λ,

$$\min_{\substack{\lambda_0,\lambda_1>0\\ \lambda_0+\lambda_1=\lambda}}\left(\frac{A}{\lambda_0^\gamma} + \frac{B}{\lambda_1^\gamma}\right) = \frac{1}{\lambda^\gamma}(A^\delta + B^\delta)^{\frac{1}{\delta}}$$

with $\delta = \frac{1}{\gamma+1}$. The inequality (C.17) follows since

$$\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + \left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2} \leqslant \left(C_{\alpha_1,\gamma}\left(v_1\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_1} + C_{\alpha_2,\gamma}\left(v_2\sum_{\ell=1}^{N} u_\ell\right)^{\alpha_2}\right)\left(\frac{1}{C_{\alpha_1,\gamma}} + \frac{1}{C_{\alpha_2,\gamma}}\right).$$

Finally, we get (C.18) by choosing the constants $C_{\alpha_1,\gamma}, C_{\alpha_2,\gamma}$ large enough to get

$$\left[\left(\frac{1}{2^{\alpha_1}} + \frac{1}{2^{\alpha_2}}\right)^{\delta} + \left(\frac{1}{C_{\alpha_1,\gamma}} + \frac{1}{C_{\alpha_2,\gamma}}\right)^{\delta}\right]^{\frac{1}{\delta}} \leqslant 1 \tag{C.19}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

which is possible since (C.19) is equivalent to

$$\left(\frac{1}{2^{\alpha_1}}+\frac{1}{2^{\alpha_2}}\right)^{\delta}+\left(\frac{1}{C_{\alpha_1,\gamma}}+\frac{1}{C_{\alpha_2,\gamma}}\right)^{\delta}\leqslant 1$$

and due to our assumption that $\alpha_1,\alpha_2>1$. □

LEMMA C.3. *Suppose there are $\gamma>0$, $\alpha_1,\alpha_2>1$, (C.1) is satisfied with $v_1,v_2>0$ and a positive sequence $(u_\ell)_{1\leqslant\ell\leqslant N}$. Then, for all $\lambda>0$,*

$$\mathrm{P}\left(|s_j-s_i|>\lambda,\,|s_k-s_j|>\lambda\right)\leqslant\frac{1}{\lambda^{\gamma}}\left(\left(v_1\sum_{\ell=i+1}^{k}u_\ell\right)^{\alpha_1}+\left(v_2\sum_{\ell=i+1}^{k}u_\ell\right)^{\alpha_2}\right),\quad 0\leqslant i\leqslant j\leqslant k\leqslant N.$$

Proof. We follow the arguments in the proof of Billingsley (1968, Thm. 12.1). That is,

$$\begin{aligned}
&\mathrm{P}\left(|s_j-s_i|>\lambda,\,|s_k-s_j|>\lambda\right)\\
&\quad\leqslant \mathrm{P}^{\frac{1}{2}}\left(|s_j-s_i|>\lambda\right)\mathrm{P}^{\frac{1}{2}}\left(|s_k-s_j|>\lambda\right)\\
&\quad\leqslant\frac{1}{\lambda^{\gamma}}\left(\left(v_1\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_1}+\left(v_2\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_2}\right)^{\frac{1}{2}}\left(\left(v_1\sum_{\ell=j+1}^{k}u_\ell\right)^{\alpha_1}+\left(v_2\sum_{\ell=j+1}^{k}u_\ell\right)^{\alpha_2}\right)^{\frac{1}{2}} \qquad \textbf{(C.20)}\\
&\quad=\frac{1}{\lambda^{\gamma}}\left(\left(v_1\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_1}\left(v_1\sum_{\ell=j+1}^{k}u_\ell\right)^{\alpha_1}+\left(v_2\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_2}\left(v_1\sum_{\ell=j+1}^{k}u_\ell\right)^{\alpha_1}\right.\\
&\qquad\left.+\left(v_1\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_1}\left(v_2\sum_{\ell=j+1}^{k}u_\ell\right)^{\alpha_2}+\left(v_2\sum_{\ell=i+1}^{j}u_\ell\right)^{\alpha_2}\left(v_2\sum_{\ell=j+1}^{k}u_\ell\right)^{\alpha_2}\right)^{\frac{1}{2}}\\
&\quad\leqslant\frac{1}{\lambda^{\gamma}}\left(\left(v_1\sum_{\ell=i+1}^{k}u_\ell\right)^{2\alpha_1}+2\left(v_1\sum_{\ell=i+1}^{k}u_\ell\right)^{\alpha_1}\left(v_2\sum_{\ell=i+1}^{k}u_\ell\right)^{\alpha_2}+\left(v_2\sum_{\ell=i+1}^{k}u_\ell\right)^{2\alpha_2}\right)^{\frac{1}{2}} \qquad \textbf{(C.21)}\\
&\quad\leqslant\frac{1}{\lambda^{\gamma}}\left(\left(v_1\sum_{\ell=i+1}^{k}u_\ell\right)^{\alpha_1}+\left(v_2\sum_{\ell=i+1}^{k}u_\ell\right)^{\alpha_2}\right), \qquad \textbf{(C.22)}
\end{aligned}$$

where (C.20) is due to (C.1) and (C.21) follows since $xy\leqslant(x+y)^2$ for $x,y>0$ and $\sum_{\ell=j+1}^{k}u_\ell\leqslant\sum_{\ell=i+1}^{k}u_\ell$. □

D. ADDITIONAL RESULTS AND THEIR PROOFS

For shortness' sake, we set $\bar{L}_N(x)=\frac{1}{N}\sum_{n=1}^{N}L_n(x)$ in this section.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

LEMMA D.1. *An approximate $1-\alpha$ confidence interval of $F^{-1}(p)$ can be written as*

$$\left(\phi(F_N)+\frac{1}{\varphi(F^{-1}(p))}\left(\bar{L}_N(F^{-1}(p))+\frac{1}{\sqrt{N}}\sigma(F^{-1}(p))z_{1-\frac{\alpha}{2}}\right)\right.$$
$$\left.,\phi(F_N)+\frac{1}{\varphi(F^{-1}(p))}\left(\bar{L}_N(F^{-1}(p))+\frac{1}{\sqrt{N}}\sigma(F^{-1}(p))z_{\frac{\alpha}{2}}\right)\right).$$

Proof. Note that

$$\begin{aligned}\mathrm{P}\left(c_1\leqslant\frac{N}{d_N}(\phi(F_N)-\phi(F))\leqslant c_2\right)&=\mathrm{P}\left(\frac{d_N}{N}c_1\leqslant\phi(F_N)-\phi(F)\leqslant\frac{d_N}{N}c_2\right)\\&=\mathrm{P}\left(\phi(F_N)-\frac{d_N}{N}c_2\leqslant\phi(F)\leqslant\phi(F_N)-\frac{d_N}{N}c_1\right).\end{aligned}\tag{D.1}$$

In order to find an approximate $1-\alpha$ confidence interval, one has to determine the critical values c_1,c_2 in (D.1). Instead of utilizing the asymptotic distribution of the empirical process, we consider the asymptotic behavior of the HOA of the empirical process.

Then, with explanations given below,

$$\begin{aligned}&\mathrm{P}\left(c_1\leqslant\frac{N}{d_N}(\phi(F_N)-\phi(F))\leqslant c_2\right)\\&\quad=\mathrm{P}\left(c_1\leqslant\phi'_F\left(\frac{N}{d_N}(F_N-F)\right)\leqslant c_2\right)+o(1)\qquad\text{(D.2)}\\&\quad=\mathrm{P}\left(\frac{d_N}{N}c_1\leqslant\frac{-1}{\varphi(F^{-1}(p))}(F_N(F^{-1}(p))-F(F^{-1}(p)))\leqslant\frac{d_N}{N}c_2\right)+o(1)\qquad\text{(D.3)}\\&=\mathrm{P}\left(-\varphi(F^{-1}(p))\frac{d_N}{N}c_2\leqslant F_N(F^{-1}(p))-F(F^{-1}(p))\leqslant-\varphi(F^{-1}(p))\frac{d_N}{N}c_1\right)+o(1)\\&\quad=\mathrm{P}\Big(-\varphi(F^{-1}(p))\frac{d_N}{\sqrt{N}}c_2-\sqrt{N}\bar{L}_N(F^{-1}(p))\\&\qquad\leqslant\sqrt{N}S_N(F^{-1}(p))\leqslant-\varphi(F^{-1}(p))\frac{d_N}{\sqrt{N}}c_1-\sqrt{N}\bar{L}_N(F^{-1}(p))\Big)+o(1)\\&\quad=\mathrm{P}\left(\sigma(F^{-1}(p))z_{1-\frac{\alpha}{2}}\leqslant S(F^{-1}(p),1)\leqslant\sigma(F^{-1}(p))z_{\frac{\alpha}{2}}\right)+o(1),\qquad\text{(D.4)}\end{aligned}$$

where (D.2) is due to the Taylor approximation (5.7), (D.3) follows by the relation (5.8) and (D.4) is due to the asymptotic result in Theorem 4.1 where $S(x,1)$ is a mean zero Gaussian process with cross-covariances $\sigma^2(x)=\sum_{n\in\mathbb{Z}}\mathrm{Cov}(S_0(x),S_n(x))$ given in (4.1).

Based on the last approximation (D.4), we can infer the following relation between c_1,c_2 and the quantiles of the normal distribution

$$\begin{aligned}\sigma(F^{-1}(p))z_{1-\frac{\alpha}{2}}&=-\varphi(F^{-1}(p))\frac{d_N}{\sqrt{N}}c_1-\sqrt{N}\bar{L}_N(F^{-1}(p)),\\\sigma(F^{-1}(p))z_{\frac{\alpha}{2}}&=-\varphi(F^{-1}(p))\frac{d_N}{\sqrt{N}}c_2-\sqrt{N}\bar{L}_N(F^{-1}(p)).\end{aligned}\tag{D.5}$$

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Combining (D.1) and (D.5), we can then infer the statement of the lemma since for example

$$\begin{aligned}
&\phi(F_N) - \frac{d_N}{N} c_1 \\
&\quad = \phi(F_N) - \frac{1}{\varphi(F^{-1}(p))} \frac{d_N}{N} \frac{\sqrt{N}}{d_N} \Big(- \sqrt{N} \bar{L}_N(F^{-1}(p)) - \sigma(F^{-1}(p)) z_{1-\frac{\alpha}{2}} \Big) \\
&\quad = \phi(F_N) + \frac{1}{\varphi(F^{-1}(p))} \Big(\bar{L}_N(F^{-1}(p)) + \frac{1}{\sqrt{N}} \sigma(F^{-1}(p)) z_{1-\frac{\alpha}{2}} \Big).
\end{aligned}$$

□

For the following lemma, recall from Section 3 that $\widetilde{c}_l(x) = \mathrm{E}\left(\mathbb{1}_{\{G(\xi_0) \leqslant x\}} H_l(\xi_0)\right)$.

LEMMA D.2. *Suppose $G : \mathbb{R} \to \mathbb{R}$ is a monotonically increasing (decreasing), bijective function. Then,*

$$\widetilde{c}_l(x) = \begin{cases} -H_{l-1}(G^{-1}(x)) \varphi\left(G^{-1}(x)\right) & \text{if } G \text{ is increasing,} \\ H_{l-1}(G^{-1}(x)) \varphi\left(G^{-1}(x)\right) & \text{if } G \text{ is decreasing.} \end{cases}$$

Proof. For a strictly monotonically increasing, bijective function G, it holds that

$$\begin{aligned}
\widetilde{c}_l(x) &= \mathrm{E}\left(\mathbb{1}_{\{G(\xi_0) \leqslant x\}} H_l(\xi_0)\right) \\
&= \int_{\mathbb{R}} \mathbb{1}_{\{G(y) \leqslant x\}} H_l(y) \varphi(y) dy \\
&= \int_{\mathbb{R}} \mathbb{1}_{\{y \leqslant G^{-1}(x)\}} H_l(y) \varphi(y) dy \\
&= \int_{-\infty}^{G^{-1}(x)} H_l(y) \varphi(y) dy = -H_{l-1}(G^{-1}(x)) \varphi(G^{-1}(x)),
\end{aligned}$$

where the last equality follows from the definition of the Hermite polynomial H_l as

$$H_l(x) = (-1)^l \frac{1}{\varphi(x)} \frac{\partial^l}{\partial x^l} \varphi(x);$$

see Pipiras and Taqqu (2017, Formula (4.1.1)). Analogously, it follows that for a strictly monotonically decreasing, bijective function G, it holds that

$$\begin{aligned}
\widetilde{c}_l(x) &= \mathrm{E}\left(\mathbb{1}_{\{G(\xi_0) \leqslant x\}} H_l(\xi_0)\right) \\
&= \int_{\mathbb{R}} \mathbb{1}_{\{G(y) \leqslant x\}} H_l(y) \varphi(y) dy \\
&= \int_{\mathbb{R}} \mathbb{1}_{\{y \geqslant G^{-1}(x)\}} H_l(y) \varphi(y) dy \\
&= \int_{G^{-1}(x)}^{\infty} H_l(y) \varphi(y) dy = H_{l-1}(G^{-1}(x)) \varphi(G^{-1}(x)).
\end{aligned}$$

□

REFERENCES

Bagchi, P., Banerjee, M., & Stoev, S. (2016). Inference for monotone functions under short- and long-range dependence: Confidence intervals and new universal limits. *Journal of the American Statistical Association*, 111(516), 1634–1647.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Bai, S., & Taqqu, M. S. (2013). Multivariate limit theorems in the context of long-range dependence. *Journal of Time Series Analysis*, 34(6), 717–743.

Beran, J. (1992). Statistical methods for data with long-range dependence. *Statistical Science*, 7(4), 404–416.

Beran, J., Feng, Y., Ghosh, S., & Kulik, R. (2016). *Long-Memory Processes*. Springer.

Betken, A. (2016). Testing for change-points in long-range dependent time series by means of a self-normalized Wilcoxon test. *Journal of Time Series Analysis*, 37(6), 785–809.

Betken, A. (2017). Change point estimation based on Wilcoxon tests in the presence of long-range dependence. *Electronic Journal of Statistics*, 11(2), 3633–3672.

Betken, A., Giraudo, D., & Kulik, R. (2023). Change-point tests for the tail parameter of long memory stochastic volatility time series. *Statistica Sinica*, 33(3), 2017–2039.

Billingsley, P. (1968). *Convergence of probability measures*, Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics. Wiley.

Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science*, 17(1), 52–72.

Dehling, H., Rooch, A., & Taqqu, M. (2013). Non-parametric change-point tests for long-range dependent data. *Scandinavian Journal of Statistics*, 40(1), 153–173.

Dehling, H., & Taqqu, M. (1989). The empirical process of some long-range dependent sequences with an application to U-Statistics. *The Annals of Statistics*, 17(4), 1767–1783.

Donsker, M. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *The Annals of Mathematical Statistics*, 2, 277–281.

Doukhan, P., & Surgailis, D. (1998). Functional central limit theorem for the empirical process of short memory linear processes. *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, 326(1), 87–92.

Dudley, R. (1978). Central limit theorems for empirical measures. *The Annals of Probability*, 6(6), 899–929.

El Ktaibi, F., & Ivanoff, B. G. (2016). Change-point detection in the marginal distribution of a linear process. *Electronic Journal of Statistics*, 10(2), 3945–3985.

Fang, Y., Xu, J., & Yang, L. (2018). Online bootstrap confidence intervals for the stochastic gradient descent estimator. *The Journal of Machine Learning Research*, 19(1), 3053–3073.

Hall, P., Jing, B.-Y., & Lahiri, S. (1998). On the sampling window method for long-range dependent data. *Statistica Sinica*, 8, 1189–1204.

Hoga, Y. (2019). Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics*, 37(4), 613–624.

Hössjer, O., & Mielniczuk, J. (1995). Delta method for long-range dependent observations. *Journal of Nonparametric Statistics*, 5(1), 75–82.

Huang, Y., & Shao, X. (2016). Coverage bound for fixed-b subsampling and generalized subsampling for time series. *Statistica Sinica*, 26(4), 1499–1524.

Ivanoff, B. G. (1980). The function space $D([0,\infty)^q, E)$. *Canadian Journal of Statistics*, 8(2), 179–191.

Kiefer, J. (1972). Skorohod embedding of multivariate RV's, and the sample DF. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 24(1), 1–35.

Kim, S., Zhao, Z., & Shao, X. (2015). Nonparametric functional central limit theorem for time series regression with application to self-normalized confidence interval. *Journal of Multivariate Analysis*, 133, 277–290.

Koul, H., & Surgailis, D. (2002). Asymptotic expansion of the empirical process of long memory moving averages. In H. Dehling, T. Mikosch, & M. Sørensen (Eds.), *Empirical process techniques for dependent data* (213–239). Springer.

Lahiri, S. N. (1993). On the moving block bootstrap under long range dependence. *Statistics & Probability Letters*, 18(5), 405–413.

Massah, M., & Kantz, H. (2016). Confidence intervals for time averages in the presence of long-range correlations, a case study on earth surface temperature anomalies. *Geophysical Research Letters*, 43(17), 9243–9249.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press

Müller, D. W. (1970). On Glivenko-Cantelli convergence. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 16(3), 195–210.

Nordman, D., Bunzel, H., & Lahiri, S. (2013). A nonstandard empirical likelihood for time series. *The Annals of Statistics*, 41(6), 3050–3073.

Nordman, D., Sibbertsen, P., & Lahiri, S. N. (2007). Empirical likelihood confidence intervals for the mean of a long-range dependent process. *Journal of Time Series Analysis*, 28(4), 576–599.

Nourdin, I., & Rosiński, J. (2014). Asymptotic independence of multiple Wiener–Itô integrals and the resulting limit laws. *The Annals of Probability*, 42(2), 497–526.

Pipiras, V., & Taqqu, M. (2017). *Long-Range Dependence and Self-Similarity*, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Purwanto, A., & Sudargini, Y. (2021). Partial least squares structural squation modeling (pls-sem) analysis for social and management research: A literature review. *Journal of Industrial Engineering & Management Research*, 2(4), 114–123.

Rust, H. W., Kallache, M., Schellnhuber, H. J., & Kropp, J. P. (2010). Confidence intervals for flood return level estimates assuming long-range dependence. In J. Kropp & H.-J. Schellnhuber (Eds.), *Extremis: Disruptive events and trends in climate and hydrology* (60–88). Springer.

Shao, X. (2010). A self-normalized approach to confidence interval construction in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3), 343–366.

Shorack, G., & Wellner, J. (2009). *Empirical processes with applications to statistics*. SIAM.

Taqqu, M. S. (1977). Law of the iterated logarithm for sums of non-linear functions of Gaussian variables that exhibit a long range dependence. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 40(3), 203–238.

Taqqu, M. S., Teverovsky, V., & Willinger, W. (1995). Estimators for long-range dependence: An empirical study. *Fractals*, 3(4), 785–798.

Tewes, J. (2018). Block bootstrap for the empirical process of long-range dependent data. *Journal of Time Series Analysis*, 39(1), 28–53.

Van der Vaart, A. (2000). *Asymptotic statistics* (vol. 3). Cambridge University Press.

Wellner, J. & van der Vaart, A. (2013). *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media.

Weron, R. (2002). Estimating long-range dependence: finite sample properties and confidence intervals. *Physica A: Statistical Mechanics and its Applications*, 312(1), 285–299.

Youndjé, E., & Vieu, P. (2006). A note on quantile estimation for long-range dependent stochastic processes. *Statistics & Probability Letters*, 76(2), 109–116.

https://doi.org/10.1017/S0266466625100054 Published online by Cambridge University Press