Journal of Machine Learning Research 18 (2018) 1-46 Submitted 1/17; Revised 5/18; Published 6/18

Concentration inequalities for empirical processes of linear time series

Likai Chen LKCHEN@GALTON.UCHICAGO.EDU
Wei Biao Wu WBWU@GALTON.UCHICAGO.EDU
Department of Statistics
The University of Chicago
Chicago, IL 60637, USA

Editor: Mehryar Mohri

Abstract

The paper considers suprema of empirical processes for linear time series indexed by functional classes. We derive an upper bound for the tail probability of the suprema under conditions on the size of the function class, the sample size, temporal dependence and the moment conditions of the underlying time series. Due to the dependence and heavy-tailness, our tail probability bound is substantially different from those classical exponential bounds obtained under the independence assumption in that it involves an extra polynomial decaying term. We allow both short- and long-range dependent processes. For empirical processes indexed by half intervals, our tail probability inequality is sharp up to a multiplicative constant.

Keywords: martingale decomposition, tail probability, heavy tail, MA(∞)

1. Introduction

Concentration inequalities for suprema of empirical processes play a fundamental role in statistical learning theory. They have been extensively studied in the literature; see for example Vapnik (1998), Ledoux (2001), Massart (2007), Boucheron et al. (2013) among others. To fix the idea, let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which a sequence of random variables (X_i) is defined, $\mathcal{A}$ be a set of real-valued measurable functions. For a function g, denote $S_n(g) = \sum_{i=1}^n g(X_i)$. We are interested in studying the tail probability

$$T(z) := \mathbb{P}(\Delta_n \geq z), \text{ where } \Delta_n = \sup_{g \in \mathcal{A}} |S_n(g) - \mathbb{E}S_n(g)|. \quad (1)$$

When $\mathcal{A}$ is uncountable, $\mathbb{P}$ is understood as the outer probability (van der Vaart (1998)). In the special case in which $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables and $\mathcal{A} = \{\mathbf{1}_{(-\infty,t]}, t \in \mathbb{R}\}$ is the set of indicator functions of half intervals, the Dvoretzky-Kiefer-Wolfowitz-Massart (Dvoretzky et al. (1956); Massart (1990)) theorem asserts that for all $z \geq 0$,

$$T(z) \leq 2e^{-2z^2/n}. \quad (2)$$

Talagrand (1994) obtained a concentration inequality with $\mathcal{A} = \{\mathbf{1}_A, A \in \mathcal{C}\}$, where $\mathcal{C}$ is a VC class; cf Vapnik and Chervonenkis (1971). For empirical processes of independent random variables, a substantial theory has been developed and various powerful techniques have been invented; see Talagrand (1995, 1996), Ledoux (1997), Massart (2000), Boucheron et al. (2003), Klein and Rio (2005) and the monograph Boucheron et al. (2013).

In this paper we shall consider tail probability inequalities for temporally dependent data which are commonly encountered in economics, engineering, finance, geography, physics and other fields. It is considerably more challenging to deal with dependent data. Previous results include uniform laws of large numbers and central limit theorems; see, for example, Adams and Nobel (2012), Levental (1988), Arcones and Yu (1994), Kontorovich and Brockwell (2014) and Yu (1994). Various uniform deviation results have been derived for mixing processes, Markov chains and their variants; see Marton (1996, 1998), Samson (2000), Kontorovich and Ramanan (2008), Adamczak (2008), Kontorovich and Weiss (2014), Kontorovich and Raginsky (2017), Kuznetsov and Mohri (2014, 2015) and Agarwal and Duchi (2013) among others. In many of the aforementioned papers, exponentially decaying tail bounds have been obtained which are similar to those obtained under independence.

Here we shall consider the widely used linear or moving average (MA) process

$$X_i = \sum_{k\geq 0} a_k \epsilon_{i-k}, \tag{3}$$

where innovations $\epsilon_i, i \in \mathbb{Z}$, are i.i.d random variables with mean 0 and $a_k, k \geq 0$, are real numbers such that X_i is a proper random variable. Assume that $\epsilon_i \in \mathcal{L}^q$, $q \geq 1$, namely $\mu_q := \|\epsilon_i\|_q = (\mathbb{E}|\epsilon_i|^q)^{1/q} < \infty$ and coefficients $a_k = O(k^{-\beta})$, $\beta > 1/q$. Namely there exists a constant $C > 0$ such that $|a_k| \leq Ck^{-\beta}$ holds for all large k. Then by Kolmogorov's three-series theorem (Chow and Teicher (1997)), the sum in (3) exists and X_i is well-defined. If $q \geq 2$ and $1/2 < \beta < 1$, then the auto-covariances of the process (X_i) may not be summable, suggesting that the process is long-memory or long-range dependent (LRD). When $\beta > 1$, the process is short-range dependent (SRD). The linear or MA(∞) process (3) is very widely used in practice and it includes many important time series models such as the autoregressive moving average (ARMA) process

$$(1 - \sum_{j=1}^{p} \theta_j B^j) X_i = X_i - \sum_{j=1}^{p} \theta_j X_{i-j} = \sum_{k=0}^{q} \phi_k \epsilon_{i-k}, \tag{4}$$

where θ_j and ϕ_k are real coefficients such that the roots to the equation $1 - \sum_{j=1}^{p} \theta_j u^j = 0$ are all outside the unit disk and B is the backshift operator, and the fractional autoregressive integrated moving average (FARIMA) (cf. Granger and Joyeux (1980); Hosking (1981))

$$(1 - B)^d (X_i - \sum_{j=1}^{p} \theta_j X_{i-j}) = \sum_{k=0}^{q} \phi_k \epsilon_{i-k}, \tag{5}$$

where the fractional integration index $d \in (0, 1/2)$. For (4), the corresponding coefficients $|a_i| = O(\rho^i)$ for some $\rho \in (0, 1)$. While for (5) under suitable causality and invertibility conditions the limit $\lim_{i\to\infty} i^{1-d} a_i = c \neq 0$ exists (Granger and Joyeux (1980); Hosking (1981)). Hence $a_i \sim ci^{-\beta}$ with $\beta = 1 - d$.

The primary goal of the paper is to establish a concentration inequality for $T(z)$ in (1) for the linear process (3). Our theory allows both short- and long-range dependence and heavy-tailed innovations. Heavy-tailed distributions have been substantially studied in the literature. For instance, Mandelbrot (1963) documented evidence of power-law behavior in asset prices. Rachev and Mittnik (2000) showed long memory and heavy tails in the high frequency asset return data. Recently researchers extended tail probability inequalities to independent heavy-tailed random variables. Lederer and van de Geer (2014) applied the truncation method to develop bounds for an envelope of functions with finite moment assumptions on the envelope. Based on the robust M-estimator introduced in Catoni (2012), Brownlees et al. (2015) proposed a risk minimization procedure using the generic chaining method. The case with both dependence and heavy tails is more challenging. Jiang (2009) introduced a triplex inequality to handle unbounded and dependent situations. Mohri and Rostamizadeh (2010) considered φ-mixing and β-mixing processes. It is generally not easy to verify that a process is strong mixing and computation of mixing coefficients can be very difficult. Some simple and widely used AR processes are not strong mixing (Andrews (1984)).

In the present paper, we propose a martingale approximation based method. An intuitive illustration is given in Section 6.2. Our tail probability bound is a combination of an exponential term and a polynomial term (cf. Theorems 4 and 8), whose order depends on both β and q, which quantify the dependence and the moment condition, respectively. Larger β or q implies thinner tails. Our tail inequality allows both short- and long- range dependent processes and can also be adapted to discontinuous function classes including empirical distribution functions, which is fundamental and is of independent interest. Our theorem implies that, if the innovation ϵ_0 has tail

$$\mathbb{P}(|\epsilon_0| \geq x) = O(\log^{-r_0}(x)x^{-q}), \quad \text{as } x \to \infty, \tag{6}$$

where $r_0 > 1$ and $q > 1$ signifies heaviness of the tail, namely there exists a constant $C > 0$ such that $\mathbb{P}(|\epsilon_0| \geq x) \leq C\log^{-r_0}(x)x^{-q}$ holds for all large x, and the coefficients

$$a_k = O(k^{-\beta}),\ \beta > 1 \text{ and } q\beta \geq 2, \tag{7}$$

where β quantifies the dependence with larger β implying weaker dependence, then for $z \geq \sqrt{n}\log(n)$, the tail probability

$$\mathbb{P}\Big(\sup_{t\in\mathbb{R}}\Big|\sum_{i=1}^{n}[\mathbf{1}_{X_i\leq t} - F(t)]\Big| > z\Big) \lesssim \frac{n}{z^{q\beta}\log^{r_0}(z)}, \tag{8}$$

where the constant in $\lesssim$ is independent of n and z, $F(t) = \mathbb{P}(X_i \leq t)$ is the cumulative distribution function (c.d.f.) for X_i. Note that the bound (8) involves both the dependence parameter β and the tail heaviness parameter q. In comparison with the sub-Gaussian bound $e^{-2z^2/n}$ in (2), the polynomial bound (8) is much larger. On the other hand, however, it turns out that the polynomial bound (8) is *sharp* and it is essentially not improvable. For example, let $F_\epsilon(t) = \mathbb{P}(\epsilon_0 \leq t)$ be the c.d.f. of ϵ_0, and assume that the innovation ϵ_i has a symmetric regularly varying tail: for some $r_0 > 1$,

$$F_\epsilon(-x) = 1 - F_\epsilon(x) \sim \log^{-r_0}(x)x^{-q} \text{ as } x \to \infty, \tag{9}$$

namely $\lim_{x\to\infty}(1 - F_\epsilon(x))\log^{r_0}(x)x^q = 1$, and that the coefficients

$$a_k = (k \vee 1)^{-\beta},\ \beta > 1. \tag{10}$$

Then by Theorem 14, when $n/\log^{\alpha_0}(n) \geq z \geq \sqrt{n}\log(n)$ for some $\alpha_0 > 0$, we can have the precise order of the tail probability

$$\mathbb{P}\Big(\sum_{i=1}^{n}[\mathbf{1}_{X_i \leq t} - F(t)] > z\Big) = C_1 \frac{n}{z^{q\beta}\log^{r_0}(z)}(1+o(1)), \quad n \to \infty,$$

and

$$\mathbb{P}\Big(\sum_{i=1}^{n}[\mathbf{1}_{X_i \leq t} - F(t)] < -z\Big) = C_2 \frac{n}{z^{q\beta}\log^{r_0}(z)}(1+o(1)), \quad n \to \infty,$$

where the constants C_1, C_2 are independent of z and n. Hence the bound in (8) is sharp up to a multiplicative constant.

On the technical side, to establish inequality (8) and more generally, a tail probability inequality for empirical processes indexed by function classes, we need to develop new approaches so that the two main challenges posed by dependence and heavy tails can be dealt with. Techniques developed for empirical processes with independent random variables are not directly applicable. Here, we apply the martingale approximation method, together with the Fuk-Nagaev inequalities for high-dimensional vectors recently obtained by Chernozhukov et al. (2017), projection techniques and martingale inequalities, so that an optimal bound can be derived. Intuitions are given in the proof of Theorem 4 in Section 6.2. As a result, we can allow short- and long-range dependent, and light- and heavy-tailed linear processes.

The remainder of the paper is organized as follows. Section 2 states the theoretical results: Subsections 2.1 and 2.2 show the tail probabilities for short- and long- range dependence situations respectively with heavy tailness, Subsection 2.3 presents results for light tailed innovations. In Section 3, we apply the concentration inequality to empirical distribution functions as an important special case. We also derive an exact order of decay speed under certain settings, which demonstrates the sharpness of our upper bound. Sections 4 and 5 present applications in kernel density estimation and empirical risk minimization, respectively. Detailed proofs are provided in Section 6.

We now introduce some notation. For a random variable X and $q > 0$, we write $X \in \mathcal{L}^q$ if $\|X\|_q := \mathbb{E}(|X|^q)^{1/q} < \infty$. Write $\|\cdot\| = \|\cdot\|_2$. For a function g, define $|g|_\infty := \sup_x |g(x)|$. Let $x \vee y = \max\{x, y\}$ and $x \wedge y = \min\{x, y\}$. For two sequences of positive numbers (a_n) and (b_n), write $a_n \lesssim b_n$ (resp. $a_n \ll b_n$, $a_n \asymp b_n$, $a_n \sim b_n$) if there exists a positive constant C such that $a_n/b_n \leq C$ for all large n (resp. $\lim_{n\to\infty} a_n/b_n = 0$, $1/C \leq a_n/b_n \leq C$ for all large n, $\lim_{n\to\infty} a_n/b_n = 1$). Denote by F_ϵ (resp. F) the c.d.f. of the innovation ϵ_i (resp. X_i) and by $f_\epsilon = F'_\epsilon$ (resp. $f = F'$) the probability density function (p.d.f.) of ϵ_i (resp. X_i).

2. Main results

Recall (3) for the MA(∞) process (X_i), where $\epsilon_j \in \mathcal{L}^q$, $j \in \mathbb{Z}$, are i.i.d. with c.d.f. F_ϵ and p.d.f. f_ϵ. Assume $a_0 \neq 0$ and without loss of generality, let $a_0 = 1$.

For a function class $\mathcal{A}$ of bounded functions, define the covering number

$$\mathcal{N}_{\mathcal{A}}(\delta) := \min\Big\{m : \text{ there exist } g_1, \ldots, g_m \in \mathcal{A} \text{ such that } \sup_{g\in\mathcal{A}} \min_{1\leq j\leq m} |g - g_j|_\infty \leq \delta\Big\}. \quad (11)$$

Let $H_{\mathcal{A}}(\delta) := \log(\mathcal{N}_{\mathcal{A}}(\delta))$ be the metric entropy.

Before stating the main theorems, we shall introduce some assumptions.

- (A) (Smoothness) For any $g \in \mathcal{A}$, g', g'' exist and $|g|, |g'|, |g''|$ are uniformly bounded, without loss of generality set the bound to be 1.
- (A') Functions in $\mathcal{A}$ are uniformly bounded in $|\cdot|_\infty$ with $\sup_{g\in\mathcal{A}} |g|_\infty \le 1$. Assume that $f'_\epsilon, f''_\epsilon$ exist and the integrals $\int_{-\infty}^{\infty} |f'_\epsilon(x)|\mathrm{d}x, \int_{-\infty}^{\infty} |f''_\epsilon(x)|\mathrm{d}x$ are bounded by 1.
- (B) (Algebraically Decaying Coefficients) For some $\gamma, \beta > 0$, $|a_k| \le \gamma k^{-\beta}$ holds for all $k \ge 1$.
- (B') (Exponentially Decaying Coefficients) For some $\gamma > 0, 0 < \rho < 1$, $|a_k| \le \gamma\rho^k$ holds for all $k \ge 1$.
- (D) (Exponential Class) For some constants $N, C, \theta > 0$, the covering number $\mathcal{N}_{\mathcal{A}}(\delta) \le N\exp(C\delta^{-\theta})$ holds for all $0 < \delta \le 1$.
- (D') (Algebraical Class) For some constants $N, \theta > 0$, the covering number $\mathcal{N}_{\mathcal{A}}(\delta) \le N\delta^{-\theta}$ holds for all $0 < \delta \le 1$.

Remark 1 *Assumption (A) requires that functions in $\mathcal{A}$ have up to second order derivatives. This is relaxed in (A'), where an extra differentiability condition of f_ϵ is imposed. It holds for many commonly used distributions such as Gaussian and t distributions.*

Remark 2 *Assumption (B) specifies the decay rate of the MA(∞) coefficients to be at most polynomial. The parameter β controls the dependence strength, with larger β implying weaker dependence. By Theorem 4(v) in Chen and Wu (2016), the AR(∞) process*

$$X_t = \sum_{i\ge 1} b_i X_{t-i} + \epsilon_t \tag{12}$$

with coefficients $|b_i| = O(i^{-\beta})$, $\beta > 1$, and $\sum_{i\ge1} |b_i| < 1$, can also be rewritten as an MA(∞) process with coefficients (a_i) decaying at the same polynomial rate. Assumption (B') allows ARMA processes (4).

Remark 3 *Assumptions (D) and (D′) quantify the magnitudes of the class $\mathcal{A}$. They are satisfied for many function classes; see van der Vaart and Wellner (1996) and Kosorok (2008). For example, the former holds for Hölder or Sobolev classes, while the latter holds for VC classes.*

In the MA(∞) model described in (3), the parameter β controls the dependence: if $\beta > 1$, the covariances $\mathrm{Cov}(X_i, X_0)$, $i \ge 1$, are absolutely summable and the process (X_i) is short-range dependent; if $1/2 < \beta < 1$, then the covariances may not be absolutely summable and the process exhibits long-range dependence. The two cases are dealt with in Subsections 2.1 and 2.2, respectively. Subsection 2.3 deals linear processes with sub-exponential innovations.

2.1 Short-range dependent linear processes

We first consider the short-range dependence case with $\beta > 1$ in model (3). Recall (1) for Δ_n. Assume throughout the paper that $n \geq 2$. Let $q' := q \wedge 2$ and

$$c(n,q) = \begin{cases} n^{1/q'}, & \text{if } q > 2 \text{ or } 1 < q < 2, \\ n^{1/2}\log^{1/2}(n), & \text{if } q = 2. \end{cases} \tag{13}$$

Theorems 4 and 7 concern algebraically and exponentially decaying coefficients, respectively. In the statements of our theorems we use the notation $C_{\alpha,\beta,\gamma,\ldots}$ to denote constants that only depend on subscripts $\alpha, \beta, \gamma, \ldots$. Since $|g|_\infty \leq 1$, we have $T(z) = 0$ if $z > n$ and thus assume throughout the paper that $z \leq n$.

Theorem 4 (Algebraically decaying coefficients) *Assume* (A) *and* (B), $\beta > 1, q > 1$ *and* $q\beta \geq 2$. *Then there exist positive constants* $C_q, C_{\beta,q,\gamma}$ *and* $C_{\beta,\gamma}$ *such that for all* $z > 0$,

$$\begin{aligned} &\mathbb{P}\Big(\Delta_n \geq C_q a_* \mu_q c(n,q) + z\Big) \\ \leq &C_{\beta,q,\gamma}\mu_q^q \frac{n}{z^{q\beta}} + 3\exp\Big(-\frac{z^2}{C_{\beta,\gamma}\mu_{q'}^{q'} n} + H_{\mathcal{A}}(z/(4n))\Big) + 2\exp\Big(-\frac{z^v}{8\mu_q^{v'}} + H_{\mathcal{A}}(z/(4n))\Big), \end{aligned} \tag{14}$$

where $\mu_q = (\mathbb{E}|\epsilon_i|^q)^{1/q}$, $a_* = \sum_{i=0}^\infty |a_i|$, *and*

$$v = v_{q,\beta} = (q'\beta - 1)(3q'\beta - 1)^{-1}, \quad v' = 2q'(3q'\beta - 1)^{-1}. \tag{15}$$

The specific values of the constants $C_q, C_{\beta,q,\gamma}$ and $C_{\beta,\gamma}$ will be given in Remark 25 (Section 6.2). The bound (14) is a combination of exponential and polynomial terms. For z relatively small, the exponential term contributes more, while for z relatively large, the polynomial term $n/z^{q\beta}$ dominates. Note that $0 < v < 1/3$. Comparing the last two terms in (14), if $n^{1/(2-v)} \lesssim z$, then the last term dominates, and vise versa.

In Theorem 4, under Assumption (A), the class $\mathcal{A}$ consists of differentiable functions. To incorporate non-continuous functions, we can impose Assumption (A'), which requires differentiability of f_ϵ; cf Proposition 5. Corollary 6 follows from Theorem 4 and Proposition 5.

Proposition 5 *Assume* (A') *and* (B), $\beta > 1, q > 1$ *and* $q\beta \geq 2$. *Then there exist positive constants* $C_q, C_{\beta,q,\gamma}$ *and* $C_{\beta,\gamma}$ *such that for all* $z > 0$,

$$\begin{aligned} &\mathbb{P}\Big(\Delta_n \geq C_q a_* \mu_q c(n,q) + z\Big) \\ \leq &C_{\beta,q,\gamma}\mu_q^q \frac{n}{z^{q\beta}} + 5\exp\Big(-\frac{z^2}{C_{\beta,\gamma}(\mu_{q'}^{q'} \vee 1) n} + H_{\mathcal{A}}(z/(4n))\Big) + 2\exp\Big(-\frac{z^v}{8\mu_q^{v'}} + H_{\mathcal{A}}(z/(4n))\Big), \end{aligned}$$

where $c(n,q)$ *is defined in* (13) *and* v, v' *are defined in* (15).

Corollary 6 *Assume* (A) *(or* (A')*) and* (B). *Let* $\beta > 1, q > 1$ *and* $q\beta \geq 2$. *Define* $c(n,q)$ *and* v *as in* (13) *and* (15), *respectively. If either (i) Assumption* (D) *holds,* $\alpha = \max\{\theta/(\theta+$

2), $(\theta - v)/(\theta + v)\}/2$, and $z \geq cn^{1/2+\alpha}$ for a sufficiently large c; or (ii) for some $N, \theta > 0$, Assumption (D') holds and $z \geq cn^{1/2}\log^{1/2}(n)$ for a sufficiently large c, then we have

$$\mathbb{P}\Big(\Delta_n \geq C_q a_* \mu_q c(n,q) + z\Big) \leq C\mu_q^q \frac{n}{z^{q\beta}}, \tag{16}$$

where the constant C only depends on $\beta, q, \gamma, \theta, c$ and N.

Observe that in (16), when $q > 2$, the term $C_q a_* \mu_q c(n,q) + z$ can actually be replaced by z by choosing a larger constant C at the right hand side of (16), since $z \geq cn^{1/2+\alpha}$ or $z \geq cn^{1/2}\log^{1/2}(n)$ for a sufficiently large c, under (i) or (ii), respectively. The tail bound depends on both the dependence parameter β and the moment q.

If the coefficients (a_k) decay exponentially (cf Assumption (B')), then the process is very weakly dependent. It turns out that the polynomial term can be removed and an exponential upper bound can be derived; cf Theorem 7. Note that the bound in Theorem 7 explicitly involves ρ, with larger ρ indicating stronger dependence. We emphasize that the constants C_q, $C_{q,\gamma}$ and $C'_{q,\gamma}$ in (17) does not depend on ρ and they are given in Remark 26 (Section 6.3). Concentration inequality of this form is useful in situations in which one needs to deal with the dependence on ρ.

Theorem 7 (Exponentially decaying coefficients) *Assume that the coefficients (a_k) of (X_i) defined in (3) satisfy (B') and $\mu_q = \|\epsilon_i\|_q < \infty$, $q > 1$. Let $\mathcal{A} = \{g : \mathbb{R} \mapsto \mathbb{R}, |g|_\infty \leq 1, |g'|_\infty \leq 1\}$. Then*

$$\mathbb{P}(\Delta_n \geq C_q \mu_q c^*(n,\rho,q) + z) \leq C_{q,\gamma} \frac{\exp\{-q(1-\rho)n\}\mu_q^q}{z^q(1-\rho)^{q+q/q'}} + \exp\Big\{-C'_{q,\gamma}\frac{z^2(1-\rho)^2}{n(\mu_q^q \vee 1)}\Big\}, \tag{17}$$

where $q' = \min\{q, 2\}$,

$$c^*(n,\rho,q) = \begin{cases} n^{1/q'}(1-\rho)^{-1-1/q'}, & \text{if } q \neq 2, \\ \sqrt{n}(1-\rho)^{-3/2}\log(n(1-\rho)^{-1}), & \text{if } q = 2. \end{cases}$$

2.2 Long-range dependent linear processes

The phenomenon of long-range dependence has been observed in various fields including economics, finance, hydrology, geophysics etc; see, for example, Beran (1994), Baillie (1996). This subsection considers $1/2 < \beta < 1$, the long-range dependence case in model (3). Weak convergence for empirical processes for long-memory time series was studied by Ho and Hsing (1996) and Wu (2003) among others. Under suitable conditions on the class $\mathcal{A}$, by Corollary 1 in Wu (2003), one has $\mathbb{E}(\Delta_n^2) \lesssim n^{3-2\beta}$, which by Markov's inequality implies

$$\mathbb{P}(\Delta_n \geq z) \leq \frac{\mathbb{E}(\Delta_n^2)}{z^2} \lesssim \frac{n^{3-2\beta}}{z^2}.$$

Here we shall derive a much sharper and more general bound; cf Theorem 8, which allows strong dependence with non-summable algebraically decaying coefficients since $\beta < 1$. In comparison the coefficients (a_k) in Theorem 4 are summable, since $\beta > 1$, and the process is weakly dependent. Proposition 9 is an analogous version of Proposition 5 which allows discontinuous functions. Corollary 10 provides an explicit upper bound under certain conditions on the bracketing numbers and it follows from Theorem 8 and Proposition 9.

Theorem 8 *Assume (A) and (B), $q > 2$, $1/2 < \beta < 1$. Then there exist positive constants $C'_{\beta,q,\gamma}$, $C_{\beta,q,\gamma}$ and $C_{\beta,\gamma}$ such that for all $z > 0$,*

$$\begin{aligned}&\mathbb{P}\Big(\Delta_n \geq C'_{\beta,q,\gamma}\mu_q n^{3/2-\beta} + z\Big)\\ \leq& C_{\beta,q,\gamma}(\mu_q^{2q} \vee \mu_q^q)\frac{n^{1+(1-\beta)q}}{z^q}\Big(1 + \frac{[H_{\mathcal{A}}(z/4n) + \log(n)]^q}{\tilde{c}^q(n,\beta)}\Big) + 3\exp\Big(-\frac{z^2}{C_{\beta,\gamma}n^{3-2\beta}\mu_2^2} + H_{\mathcal{A}}(z/(4n))\Big),\end{aligned} \tag{18}$$

where

$$\tilde{c}(n,\beta) = \begin{cases} n^{1/4-|3/4-\beta|} & \text{if } \beta \neq 3/4, \\ n^{1/4}/\log(n) & \text{if } \beta = 3/4. \end{cases} \tag{19}$$

Values of constants $C'_{\beta,q,\gamma}$, $C_{\beta,q,\gamma}$ and $C_{\beta,\gamma}$ in Theorem 8 are given in Remark 29 (Section 6.4). In comparison with the bound $nz^{-q\beta}$ in the short-range dependence case Theorem 4, the bound $n^{1+(1-\beta)q}z^{-q}$ in (18) of Theorem 8 is larger since $nz^{-q\beta} \leq n^{1+(1-\beta)q}z^{-q}$ and $n \geq z$.

Proposition 9 *Assume (A′) and (B), $q > 2$, $1/2 < \beta < 1$. Recall (19) for $\tilde{c}(n,q)$. Then there exist positive constants $C'_{\beta,q,\gamma}$, $C_{\beta,q,\gamma}$ and $C_{\beta,\gamma}$ such that for all $z > 0$,*

$$\begin{aligned}\mathbb{P}\Big(\Delta_n \geq C'_{\beta,q,\gamma}\mu_q n^{3/2-\beta} + z\Big) \quad\leq\quad & C_{\beta,q,\gamma}(\mu_q^{2q} \vee \mu_q^q)\frac{n^{1+(1-\beta)q}}{z^q}\Big(1 + \frac{[H_{\mathcal{A}}(z/4n) + \log(n)]^q}{\tilde{c}^q(n,\beta)}\Big)\\ &+5\exp\Big(-\frac{z^2}{C_{\beta,\gamma}n^{3-2\beta}(\mu_2^2 \vee 1)} + H_{\mathcal{A}}(z/(4n))\Big).\end{aligned}$$

Corollary 10 *Assume (A) (or (A′)) and (B). Let $q > 2$, $1/2 < \beta < 1$. If either (i) for some $N, \theta > 0$, Assumption (D) holds and $z \geq cn^{3/2-\beta+\alpha}$ for $\alpha = (\beta - 1/2)\theta/(\theta + 2)$ and a sufficiently large c or (ii) for some $N, \theta > 0$, Assumption (D′) holds and $z \geq cn^{3/2-\beta}\log^{1/2}(n)$ for a sufficiently large c. Then there exists a constant $C'_{q,\beta,\gamma}$ such that*

$$\mathbb{P}\Big(\Delta_n \geq C'_{q,\beta,\gamma}\mu_q n^{3/2-\beta} + z\Big) \lesssim \frac{n^{1+(1-\beta)q}}{z^q}(\mu_q^{2q} \vee \mu_q^q)\Big(1 + \frac{t_n^q}{\tilde{c}^q(n,\beta)}\Big), \tag{20}$$

where $t_n = n^{\theta(\beta-1/2-\alpha)}$ and $\log(n)$ for (i) and (ii) respectively, and the constant in $\lesssim$ only depends on $q, \beta, \gamma, \theta, c$ and N.

2.3 Linear processes with sub-exponential innovations

In this subsection, we shall consider concentration inequalities for linear processes with innovations having very light tails. In particular, we assume that innovations ϵ_i have sub-exponential tails. In this case for both short- and long-range dependent processes we have exponentially decaying tail probabilities, with different norming sequences.

Theorem 11 *Let $\mathcal{G} = \{g : |g|_\infty \leq 1, |g'|_\infty \leq 1\}$. Assume (B) and there exist constants $c_0 > 0, f_* > 0$ such that $|f_\epsilon|_\infty \leq f_*$, where f_ϵ is the p.d.f of ϵ_0, and $\mu_e := \mathbb{E}(e^{c_0|\epsilon_0|}) < \infty$. Then there exist constants C_1, C_2, C_3 and C_4 such that* *(a) for SRD case ($\beta > 1$), we have for all $z > 0$,*

$$\mathbb{P}(\Delta_n \geq C_1\sqrt{n} + z) \leq 2e^{-C_2 z^2/n},$$

(b) for LRD case ($1/2 < \beta < 1$), we have for all $z > 0$,

$$\mathbb{P}(\Delta_n \geq C_3 n^{3/2-\beta} + z) \leq 2e^{-C_4 z^2/n^{3-2\beta}}.$$

Here the constants C_1 and C_3 only depend on $f_*, \beta, \gamma, c_0, \mu_e$, constants C_2, C_4 only depend on $\beta, \gamma, c_0, \mu, \mu_e$ and their values are given in Remark 30 (Section 6.5). Note that Theorem 11(a) implies $\mathbb{P}(\Delta \geq z) \leq 2e^{-C_5 z^2/n}$ for all $z > 0$, where constant C_5 depends on $f_*, \beta, \gamma, c_0, \mu$ and μ_e. A similar claim can be made for case (b).

In comparison with the results in Theorem 4 and Theorem 8, due to the light tails of the innovations, we do not encounter the polynomial terms $n/z^{q\beta}$ or $n^{3-2\beta}/z^{q\beta}$ here.

3. Empirical distribution functions

In this section we shall consider the important class of indicators indexed by half intervals. Let

$$S_n(t) = n[\hat{F}_n(t) - F(t)] = \sum_{i=1}^{n}[\mathbf{1}_{X_i \leq t} - F(t)]. \tag{21}$$

In Massart (1990)'s result (2), X_i are i.i.d. In Theorem 12, we present a concentration inequality for dependent and possibly heavy-tailed random variables, which has a very different upper bound that involves a polynomial decaying tail. Theorem 14 provides a lower bound for the deviation with regularly varying innovations. That lower bound assures the sharpness of Theorem 12: the polynomial decaying tail is unavoidable. Recall F_ϵ is the c.d.f. of ϵ_0 and f_ϵ its p.d.f. The values of constants in Theorem 12 are given in Remark 31 (Section 6.6). Following assumption states the boundedness of $|f_\epsilon|_\infty$ and $|f'_\epsilon|_\infty$.

(A_1) Let $f_* := \max(1, |f_\epsilon|_\infty, |f'_\epsilon|_\infty)$. Assume $f_* < \infty$.

Theorem 12 *Assume (A_1) and (B). Recall $c(n,q)$ and v, v' in (13) and (15) respectively.*

(i). Let $\beta > 1, q > 1$ (SRD case) and $q\beta \geq 2$. Then there exist constants C_0, C_1, C_2, C_3 such that

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in\mathbb{R}}|S_n(t)|/f_* > C_0 a_* \mu_q c(n,q) + z\right) \\
\leq& C_1 \mu_q^q \frac{n}{z^{q\beta}} + 4\exp\Big\{-C_2\frac{z^2}{n(\mu_{q'}^{q'} \vee 1)} + C_3\log(n\mu_q)\Big\} \\
&+ 2\exp\Big\{-\frac{z^v}{2^{3+2v}\mu_q^{v'}} + C_3\log(n\mu_q)\Big\},
\end{aligned}$$

In particular, if $z \geq cn^{1/2}\log^{1/2}(n)$, where c is a sufficiently large constant, then the above upper bound becomes $2C_1\mu_q^q n/z^{q\beta}$.

(ii). If $1/2 < \beta < 1$ (LRD case) and $q > 2$, then there exist constants C_0', C_1', C_2', C_3' such that

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}} |S_n(t)|/f_* > C_0'\mu_q n^{3/2-\beta} + z\right)$$
$$\leq C_1'(\mu_q^{2q} \vee \mu_q^q)\frac{n^{1+(1-\beta)q}}{z^q} + 4\exp\Big\{ - C_2'\frac{z^2}{n^{3-2\beta}(\mu_2^2 \vee 1)} + C_3'\log(n\mu_q)\Big\},$$

If $z \geq cn^{3/2-\beta}\log^{1/2}(n)$ for a sufficiently large c, then the above upper bound becomes $2C_1'(\mu_q^{2q} \vee \mu_q^q)n^{1+(1-\beta)q}/z^q$.

In (i) the constant C_0 only depends on q, C_1, C_3 only depend on β, q, γ and C_2 only depends on β, γ; In (ii) the constants C_0', C_1', C_3' only depend on β, q, γ and C_2' only depends on β, γ, their specific values can be found in Remark 31 (Section 6.6).

Under certain forms of tail probability of the innovations, we can have a more refined result.

Proposition 13 *Assume (A_1), (B), $\beta > 1$ and $q > 2$. Assume for any $x > 1$, $\mathbb{P}(|\epsilon_0| > x) \leq L\log^{-r_0}(x)x^{-q}$, with some constants $r_0 > 1, L > 0$. If $z \geq c\sqrt{n}\log^{\alpha}(n)$, $\alpha > 1/2$, then*

$$\mathbb{P}\left(\sup_{t\in\mathbb{R}} |S_n(t)|/f_* > z\right) \lesssim \frac{\mu_q^q n}{z^{q\beta}\log^{r_0}(z)},$$

where the constant in $\lesssim$ only depends on $\beta, q, \gamma, r_0, L, c$ and α.

To appreciate the sharpness of the upper bound in Proposition 13, we derive an exact decay rate when $a_k = (k \vee 1)^{-\beta}$ and ϵ_0 is symmetric with a regularly varying tail.

Theorem 14 *Assume (A_1), (B) with coefficients $a_k = (k \vee 1)^{-\beta}$, $k \geq 0$, and that ϵ_0 is symmetric with tail distribution*

$$\mathbb{P}(\epsilon_0 \geq x) \sim \log^{-r_0}(x)x^{-q}, \text{ as } x \to \infty, \tag{22}$$

where $r_0 > 1$ is a constant. Let $\beta > 1$, $q > 2$ and $\alpha > 1/2$. Then there exists a constant $\Gamma > 0$ such that for all z with $\sqrt{n}\log^{\alpha}(n) \leq z \leq n/\log^{\Gamma}(n)$,

$$\mathbb{P}\left(S_n(t) > z\right) = (1 + o(1))C_1\frac{n}{\log^{r_0}(z)z^{q\beta}}, \tag{23}$$

and

$$\mathbb{P}\left(S_n(t) < -z\right) = (1 + o(1))C_2\frac{n}{\log^{r_0}(z)z^{q\beta}}, \tag{24}$$

where the constants C_1, C_2 only depend on q, β, r_0, t and F.

Values of C_1 and C_2 are given in Lemma 34, and the constant Γ can be found in Remark 35 (Section 6.7). The asymptotic expressions (23) and (24) in Theorem 14 precisely depict the magnitude of the tail probability $\mathbb{P}(S_n(t) > z)$ and $\mathbb{P}(S_n(t) < -z)$. It asserts that the upper bound order in Proposition 13 is optimal within the range $\sqrt{n}\log^{\alpha}(n) \leq z \leq n/\log^{\Gamma}(n)$. Thus the polynomial $n/z^{q\beta}$ in Theorems 4 and 12 is sharp up to a multiplicative logarithmic term.

4. Kernel density estimation

Let (X_i) be a stationary sequence satisfying (3) with the marginal p.d.f. f. Given the observations $X_1, \dots, X_n$, the kernel density estimator of f is

$$\hat{f}_n(x) = \frac{1}{n}\sum_{j=1}^{n} K_b(x - X_j), \quad K_b(\cdot) = b^{-1}K(\cdot/b),$$

where the bandwidth $b = b_n$ satisfies the natural condition $b_n \to 0$ and $nb_n \to \infty$. Wu and Mielniczuk (2002) established an asymptotic distribution theory for $A_n(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x))$ for both short- and long-range dependent processes, where A_n is a proper norming sequence. In this section we shall derive a bound for the tail probability

$$\mathbb{P}\Big(\sup_{x\in\mathbb{R}} n|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \geq z\Big).$$

Such a bound is useful for constructing non-asymptotic confidence bounds. Giné and Guillou (2002) and Giné and Nickl (2010) considered the latter problem for i.i.d. data. Einmahl and Mason (2005) derived uniform in bandwidth consistency result for kernel-type function estimators. Hang et al. (2016) studied consistency properties for observations generated by certain dynamical systems under mixing conditions. Rinaldo et al. (2012), Chen et al. (2016) and Arias-Castro et al. (2016) applied such bounds in clustering problem. Liu et al. (2011) and Lafferty et al. (2012) used it in forest density estimation. Here, we shall provide a polynomial decay bound for linear time series.

Corollary 15 *Assume (B), the kernel K is symmetric with support $[-1,1]$, $\max(|K|_\infty, |K'|_\infty) \leq K_*$ and $\max(1, |f_\epsilon|_\infty, |f'_\epsilon|_\infty, |f''_\epsilon|_\infty) \leq f_*$ for some constants $K_*, f_* > 0$.*

(a) In the SRD case with $\beta > 1, q > 1, q\beta \geq 2$, if $z \geq c(n/b_n)^{1/2}\log^{1/2}(n)$ for a sufficiently large c, then

$$\mathbb{P}\Big(\sup_{x\in\mathbb{R}} n|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \geq \max(f_*, K_*)z\Big) \lesssim \mu_q^q n / z^{q\beta}, \tag{25}$$

where the constant in $\lesssim$ only depends on β, q, γ and c.

(b) In the LRD case with $1/2 < \beta < 1, q > 2$, if $z \geq c\max\{n^{3/2-\beta}, (n/b_n)^{1/2}\}\log^{1/2}(n)$ holds for a sufficiently large c, then

$$\mathbb{P}\Big(\sup_{x\in\mathbb{R}} n|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \geq \max(f_*, K_*)z\Big) \lesssim (\mu_q^{2q} \vee \mu_q^q)\frac{n^{1+(1-\beta)q}}{z^q}, \tag{26}$$

where the constant in $\lesssim$ only depends on β, q, γ and c.

5. Empirical risk minimization

Empirical risk minimization is of fundamental importance in the statistical learning theory and it is studied in various contexts including classification, regression and clustering among others. To fix the notation, let (X, Y) be a random vector taking values in the space $\mathcal{X} \times \mathcal{Y}$

and $\mathcal{H}$ be a class of measurable functions $h : \mathcal{X} \to \mathcal{Y}$. For a function $h \in \mathcal{H}$, define the risk $R(h) = \mathbb{E}[L(X, Y, h(X))]$, where L is a loss function. Let $h^* = \text{argmin}_{h\in\mathcal{H}} R(h)$. Based on the observations $(X_1, Y_1), \ldots, (X_n, Y_n)$ which are identically distributed as (X, Y), consider the empirical risk minimizer

$$\hat{h} = \text{argmin}_{h\in\mathcal{H}} R_n(h), \text{ where } R_n(h) = n^{-1} \sum_{i=1}^{n} L(X_i, Y_i, h(X_i)) \tag{27}$$

is the empirical risk. Since $R_n(h^*) \geq R_n(\hat{h})$, it follows (cf. Devroye et al. (1996)) that

$$0 \leq R(\hat{h}) - R(h^*) \leq 2\Psi_n, \text{ where } \Psi_n = \sup_{h\in\mathcal{H}} |R_n(h) - R(h)|. \tag{28}$$

A primary goal in statistical learning theory is to bound the uniform deviation Ψ_n. The latter problem has been widely studied when (X_i, Y_i) are assumed to be i.i.d.; see, for example, Caponnetto and Rakhlin (2006), Vapnik (1998, 2000) and Gottlieb et al. (2017). In recent years various dependent processes have been considered; see Modha and Masry (1996), Guo and Shi (2011), Zou and Li (2007), Zou et al. (2009), Alquier and Wintenberger (2012), Mohri and Rostamizadeh (2010), Steinwart and Christmann (2009), Hang and Steinwart (2014, 2016), Shalizi and Kontorovich (2013) among others.

Here we shall provide an upper bound for Ψ_n with (X_i) being the MA(∞) process (3) and the regression model

$$Y_i = H_0(X_i, \eta_i),$$

where $\eta_i, i \in \mathbb{Z}$, are i.i.d. random errors independent of (ϵ_i) and H_0 is an unknown measurable function. Denote $\mathcal{A} = \{g(x, y) = L(x, y, h(x)), h \in \mathcal{H}\}$ and

$$\mathcal{N}_{\mathcal{A}}(\delta) = \min\{m : \text{there exist } g_1, \ldots, g_m \in \mathcal{A}, \text{such that } \sup_{g\in\mathcal{A}} \min_{1\leq j\leq m} |g - g_j|_\infty \leq \delta\},$$

where $|g|_\infty = \sup_{x,y} |g(x, y)|$. Assume that the loss function L take values in $[0, 1]$. Here for the sake of presentational clarity we do not seek the fullest generality but as an illustration on how to apply our main results. Recall that f_ϵ is the density function of ϵ_i.

Corollary 16 *Assume (B), the density $f_\epsilon \in \mathcal{C}^2(\mathbb{R})$ with $f_* := \max(\int_{-\infty}^{\infty} |f'_\epsilon(x)|\mathrm{d}x, \int_{-\infty}^{\infty} |f''_\epsilon(x)|\mathrm{d}x, 1)$. Under conditions (i) or (ii) in Corollary 6 on the function class $\mathcal{H}$, $q, \beta > 1$ and $q\beta \geq 2$ (resp. conditions (i) or (ii) in Corollary 10 on $\mathcal{H}$, $q > 2$ and $1/2 < \beta < 1$), we have* (16) *(resp.* (20)*) holds with Δ_n therein replaced by $n\Psi_n/f_*$.*

Remark 17 In literature, many concentration inequalities for time series are derived under various mixing conditions (see, for example, Mohri and Rostamizadeh (2010)). Since mixing and our model (3) cover different ranges of processes, our results are not directly comparable with theirs. Here we consider an example in which our result and Corollary 21 in Mohri and Rostamizadeh (2010) can be compared. Let $X_i = \sum_{k\geq 0} a_k \epsilon_{i-k}$, where ϵ_t are i.i.d. with finite qth moment, $q > 2$ and $a_0 = 1$, $a_k \asymp k^{-\alpha}$, $\alpha > 2 + 1/q$. Assume the p.d.f. of ϵ_i satisfies $\int_{x\in\mathbb{R}} |f'_\epsilon(x)|\mathrm{d}x < \infty$ and $\int_{x\in\mathbb{R}} |f''_\epsilon(x)|\mathrm{d}x < \infty$. By Theorem 2.1 in Pham and Tran (1985), X_i is β-mixing and its β-mixing coefficient $\beta(k) = O(k^{1-(\alpha-1)q/(1+q)})$.

Assume functions $h \in \mathcal{H}$ are bounded and the function class $\mathcal{H}$ satisfies condition (D'). Also assume that a $\hat{\beta}$-stable algorithm yields an estimate $\hat{h}_S$ with $\hat{\beta} = O(n^{-1})$ where the definition for $\hat{\beta}$-stable can be found in Definition 4 of Mohri and Rostamizadeh (2010).

Let $K = 1/4 - (q+1)/(2(\alpha-1)q)$. By Corollary 21 in Mohri and Rostamizadeh (2010), there exists a constant $C > 0$ such that for $\delta > n^{-K}$,

$$\mathbb{P}(n|R_n(\hat{h}) - R(\hat{h})| \geq Cz_\delta) \leq \delta, \text{ where } z_\delta = n^{1-K}(\log(\delta - n^{-K})^{-1})^{1/2}. \tag{29}$$

By our Corollary 17,

$$\mathbb{P}(\sup_{h\in\mathcal{H}} n|R_n(h) - R(h)| \geq Cz_\delta) \lesssim \frac{n}{z_\delta^{q\alpha}}. \tag{30}$$

Note that, if $\delta > n^{-K}$, $nz_\delta^{-q\alpha} = O(n^{1-(1-K)q\alpha})$, which is of order $o(n^{-K})$ since $1-(1-K)q\alpha < -K$. To give a numeric example, let $\alpha = 4$, $q = 4$. Then $K = 1/24$, $1-(1-K)q\alpha = -43/3$. So (30) gives a much smaller upper bound $O(n^{-43/3})$, while (29) leads to the bound $O(n^{-1/24})$. The latter phenomenon could be explained by the sharpness of our upper bounds.

6. Proofs

In this section we shall provide proofs for results stated in the previous sections. We shall first introduce some notation. For $k \geq 1$ define the functions

$$g_k(x) := \mathbb{E}\Big[g(\sum_{i=0}^{k-1} a_i\epsilon_{-i} + x)\Big], \quad g_\infty(x) := \mathbb{E}[g(X_0 + x)]. \tag{31}$$

Since $a_0 = 1$, $g_1(x) = \mathbb{E}g(\epsilon_0 + x)$. Write $g_0(\cdot) = g(\cdot)$. Define projection operator $P_k,\ k \in \mathbb{Z}$, by $P_k\cdot = \mathbb{E}(\cdot|\mathcal{F}_k) - \mathbb{E}(\cdot|\mathcal{F}_{k-1})$, where $\mathcal{F}_i = (\epsilon_i, \epsilon_{i-1}, \ldots)$, that is $P_k f = \mathbb{E}(f(X)|\mathcal{F}_k) - \mathbb{E}(f(X)|\mathcal{F}_{k-1})$. For $j \leq i$, let

$$X_{i,j} = \sum_{k\geq 0} a_{i-j+k}\epsilon_{j-k}$$

be the truncated process. Then $X_{i,j} = \mathbb{E}(X_i|\mathcal{F}_j)$ and $g_{i-j}(X_{i,j}) = \mathbb{E}(g(X_i)|\mathcal{F}_j)$.

Let $\lfloor x \rfloor = \max\{i \in \mathbb{Z}, i \leq x\}$ and $\lceil x \rceil = \min\{i \in \mathbb{Z}, i \geq x\}$. Recall $\mu_q = (\mathbb{E}|\epsilon_0|^q)^{1/q}$ and let $\mu = \mu_1$.

In Section 6.1 we shall first present some inequalities and lemmas that will be extensively used. Theorem 4 and Proposition 5 (resp. Theorem 8 and Proposition 9) are proved in Section 6.2 (resp. Section 6.4). Theorem 7 (resp. Theorem 11, Theorem 14) is shown in Section 6.3 (resp. Section 6.5, Section 6.7). Section 6.6 gives proofs of Theorem 12 and Proposition 13. Proofs of Corollaries 15 and 16 are provided in Section 6.8.

6.1 Some useful lemmas

Lemma 18 is a maximal form of Freedman's martingale inequality (cf Freedman (1975)) and it is a simple modified version of Lemma 1 in Haeusler (1984). Lemma 19 is Burkholder's martingale inequality for moments (Burkholder (1988)). Lemma 20 is a Fuk-Nagaev inequality for high dimensional vectors (Chernozhukov et al. (2017)).

Lemma 18 *Let $\mathcal{A}$ be an index set with $|\mathcal{A}| < \infty$. For each $a \in \mathcal{A}$, let $\{\xi_{a,i}\}_{i=1}^n$ be a martingale difference sequence with respect to the filtration $\{\mathcal{F}_i\}_{i=1}^n$. Let $M_a = \sum_{i=1}^n \xi_{a,i}$ and $V_a = \sum_{i=1}^n \mathbb{E}[\xi_{a,i}^2|\mathcal{F}_{i-1}]$. Then for all $z, u, v > 0$*

$$\mathbb{P}\Big(\max_{a\in\mathcal{A}} |M_a| \geq z\Big) \leq \sum_{i=1}^n \mathbb{P}\Big(\max_{a\in\mathcal{A}} |\xi_{a,i}| \geq u\Big) + 2\mathbb{P}\Big(\max_{a\in\mathcal{A}} V_a \geq v\Big) + 2|\mathcal{A}|e^{-z^2/(2zu+2v)}.$$

Lemma 19 (Burkholder (1988), Rio (2009)) *Let $q > 1$, $q' = \min\{q, 2\}$. Let $M_T = \sum_{t=1}^T \xi_t$, where $\xi_t \in \mathcal{L}^q$ are martingale differences. Then*

$$\|M_T\|_q^{q'} \leq K_q^{q'} \sum_{t=1}^T \|\xi_t\|_q^{q'}, \text{ where } K_q = \max((q-1)^{-1}, \sqrt{q-1}).$$

Lemma 20 (A Fuk-Nagaev type inequality) *Let $X_1, \ldots, X_n$ be independent mean 0 random vectors in $\mathbb{R}^p$ and $\sigma^2 = \max_{1\leq j\leq p} \sum_{i=1}^n \mathbb{E}(X_{i,j}^2)$. Then for every $s > 1$ and $t > 0$,*

$$\mathbb{P}\Big(\max_{1\leq j\leq p} |\sum_{i=1}^n X_{i,j}| \geq 2\mathbb{E}(\max_{1\leq j\leq p} |\sum_{i=1}^n X_{i,j}|) + t\Big) \leq e^{-t^2/(3\sigma^2)} + \frac{K_s}{t^s} \sum_{i=1}^n \mathbb{E}(\max_{1\leq j\leq p} |X_{i,j}|^s),$$

where K_s is a constant depending only on s.

Lemma 21 *Assume that function g has second order derivative and $|g|, |g'|, |g''|$ are all bounded by $M < \infty$. Then $g_k, k \geq 1$, and g_∞ also have second order derivatives and $|g_k|, |g_k'|, |g_k''|$, $|g_\infty|, |g_\infty'|, |g_\infty''|$ are all bounded by M, where g_k and g_∞ are defined in* (31).

Proof Since $|g'|$ is bounded by M, by the dominated convergence theorem,

$$\lim_{\delta\to 0} \mathbb{E}\Big(\frac{g(\sum_{i=0}^{k-1} a_i\epsilon_{-i} + x + \delta) - g(\sum_{i=0}^{k-1} a_i\epsilon_{-i} + x)}{\delta}\Big) = \mathbb{E}\Big(g'\big(\sum_{i=0}^{k-1} a_i\epsilon_{-i} + x\big)\Big).$$

Since $g_k(x) = \mathbb{E}g(\sum_{i=0}^{k-1} a_i\epsilon_{-i} + x)$, $g_k'(x)$ exists and equals to $\mathbb{E}(g'(\sum_{i=0}^{k-1} a_i\epsilon_{-i} + x))$ with $|g_k'| \leq M$. Similarly g_k'' exists and $|g_k''|_\infty \leq M$. Note that $g_\infty(x) = \mathbb{E}g(\sum_{i=0}^{\infty} a_i\epsilon_{-i} + x)$. Hence same arguments lead to the existence of g_∞' and g_∞'', and they are also bounded in absolute value by M. ■

Lemma 22 *Let $\lambda > 0$, $\beta > 1$ and $G(y) = \sum_{k=0}^\infty \min\{\lambda, (k \vee 1)^{-\beta} y\}$, $y > 0$. Then for all $y > 0$, $G(y) \leq K_{\beta,\lambda} \min\{y, y^{1/\beta}\}$, where $K_{\beta,\lambda} = \max\{(\beta-1)^{-1}, \lambda\} + 2$.*

Proof Clearly $G(y) \leq \sum_{k=0}^\infty (k \vee 1)^{-\beta} y \leq (2 + (\beta-1)^{-1})y$. If $y \geq 1$, we have $y^{1/\beta} \leq y$ and

$$\begin{aligned} G(y) &\leq \sum_{k=0}^{\lceil y^{1/\beta}\rceil} \lambda + \sum_{k=\lceil y^{1/\beta}\rceil+1}^{\infty} k^{-\beta} y \leq \lambda(y^{1/\beta} + 2) + (\beta-1)^{-1} y^{(1-\beta)/\beta} y \\ &\leq \max\{(\lambda+2), (\beta-1)^{-1}\} y^{1/\beta}. \end{aligned}$$

So the lemma follows by considering two cases $0 < y < 1$ and $y \geq 1$ separately. ■

6.2 Proof of Theorem 4 and Proposition 5

The proof of Theorem 4 is quite involved. Here we shall first provide intuitive ideas of our martingale approximation approach. Recall the projection operator $P_k \cdot = \mathbb{E}(\cdot|\mathcal{F}_k) - \mathbb{E}(\cdot|\mathcal{F}_{k-1})$ and (31) for g_k and g_∞. Then $P_k g(X_i) = 0$ if $k > i$. Note that $\phi_j(g) := P_j S_n(g)$, $j = \ldots, n-1, n$, are martingale differences. Since $g_{i-j}(X_{i,j}) = \mathbb{E}(g(X_i)|\mathcal{F}_j)$, $j \le i$,

$$S_n(g) - \mathbb{E}S_n(g) = \sum_{j \le n} \phi_j(g), \text{ where } \phi_j(g) = \sum_{i=1\vee j}^{n} (g_{i-j}(X_{i,j}) - g_{i-j+1}(X_{i,j-1})). \tag{32}$$

Let $\epsilon_i, \epsilon'_j, \epsilon''_k, i, j, k \in \mathbb{Z}$ be i.i.d. Since $g_{i-j+1}(x) = \mathbb{E}(g_{i-j}(x + a_{i-j}\epsilon_j))$, $g_{i-j}(x + a_{i-j}\epsilon_j) - g_{i-j+1}(x) = \mathbb{E}(\int_{a_{i-j}\epsilon'_j}^{a_{i-j}\epsilon_j} g'_{i-j}(x+t)\mathrm{d}t|\mathcal{F}_j)$. Note that $X_{i,j} - X_{i,j-1} = a_{i-j}\epsilon_j$. Then

$$g_{i-j}(X_{i,j}) - g_{i-j+1}(X_{i,j-1}) = \mathbb{E}\Big(\int_{a_{i-j}\epsilon'_j}^{a_{i-j}\epsilon_j} g'_{i-j}(x + X_{i,j-1})\mathrm{d}x|\mathcal{F}_j\Big). \tag{33}$$

Let $X''_{i,j} = \sum_{k\ge 0} a_{i-j+k}\epsilon''_{j-k}$. Then $g_\infty(x) = \mathbb{E}(g_{i-j}(X''_{i,j} + x)) = \mathbb{E}(g_{i-j}(X''_{i,j} + x)|\mathcal{F}_j)$ and

$$g_\infty(a_{i-j}\epsilon_j) - \mathbb{E}g_\infty(a_{i-j}\epsilon_j) = \mathbb{E}\Big(\int_{a_{i-j}\epsilon'_j}^{a_{i-j}\epsilon_j} g'_{i-j}(x + X''_{i,j})\mathrm{d}x|\mathcal{F}_j\Big). \tag{34}$$

Since $\|X_{i,j}\|_q \to 0$ as $j \to \infty$, intuitively we have $g'_{i-j}(x+X_{i,j-1}) \approx g'_{i-j}(x) \approx g'_{i-j}(x+X''_{i,j})$. These relations (33) and (34) motivate us to approximate $S_n(g) - \mathbb{E}S_n(g)$ by

$$T_n(g) = \sum_{j\le n} \tilde{\phi}_j(g), \text{ where } \tilde{\phi}_j(g) = \sum_{i=1\vee j}^{n} \big(g_\infty(a_{i-j}\epsilon_j) - \mathbb{E}g_\infty(a_{i-j}\epsilon_j)\big). \tag{35}$$

Note that $\tilde{\phi}_j(g), j \le n$, are independent random variables. Hence we can apply corresponding inequalities. In Lemma 23 a Fuk-Nagaev type inequality for $T_n(g)$ is derived. Lemma 24 concerns the closeness of $S_n(g) - \mathbb{E}S_n(g)$ and $T_n(g)$. Similar arguments are also applied in the proofs of other theorems in the paper.

Proof We now proceed with the formal argument. By (11), there exists a set A_n such that for any $g \in \mathcal{A}$, $\min_{h\in A_n} |h - g|_\infty \le z/(4n)$ and $|A_n| = \mathcal{N}_{\mathcal{A}}(z/(4n))$. Then

$$\sup_{g\in\mathcal{A}} \Big| \sum_{i=1}^{n} \big[(g - \tau_n(g))(X_i) - \mathbb{E}(g - \tau_n(g))(X_i)\big] \Big| \le z/2,$$

where $\tau_n(g) := \mathrm{argmin}_{h\in A_n}|h - g|_\infty$. Hence $\Delta_n \le z/2 + \max_{g\in A_n} |S_n(g) - \mathbb{E}S_n(g)|$ and

$$\Delta_n \le \frac{z}{2} + \max_{g\in A_n}|S_n(g) - \mathbb{E}S_n(g) - T_n(g)| + \max_{g\in A_n}|T_n(g)| =: \frac{z}{2} + \Omega_n + U_n. \tag{36}$$

For $U_n = \max_{g\in A_n} |T_n(g)|$, by Lemma 23, we have

$$\mathbb{P}\Big(U_n \ge C_q a_* \mu_q c(n, q) + \frac{z}{4}\Big) \le \exp\Big(-\frac{z^2}{C_{\beta,\gamma}\mu_{q'}^{q'} n}\Big) + C_{\beta,q,\gamma,1}\mu_q^q \frac{n}{z^{q\beta}}. \tag{37}$$

For the difference term $\Omega_n = \max_{g\in A_n} |S_n(g) - \mathbb{E}S_n(g) - T_n(g)|$, by Lemma 24,

$$\mathbb{P}(\Omega_n \geq \frac{z}{4}) \leq C_{\beta,q,\gamma,2}\mu_q^q \frac{n}{z^{q\beta}} + 2|A_n|\exp\Big(-\frac{z^2}{C_{\beta,\gamma}\mu_{q'}^{q'} n}\Big) + 2|A_n|\exp\Big(-\frac{z^v}{8\mu_q^{v'}}\Big), \tag{38}$$

where $C_{\beta,q,\gamma,1}$ and $C_{\beta,q,\gamma,2}$ are constants only depending on β, q, γ and $C_{\beta,q,\gamma} = C_{\beta,q,\gamma,1} + C_{\beta,q,\gamma,2}$. Combining (36), (37) and (38), we complete the proof. ■

Lemma 23 *Recall the definitions of $\tilde{\phi}_j(g)$ and $T_n(g)$ in (32) and (35) respectively. Under assumptions of Theorem 4, we have (37).*

Proof Recall $U_n = \max_{g\in A_n} |T_n(g)|$. The proof contains two parts:

(i). Apply the Fuk-Nagaev type inequality (Lemma 20) to bound $\mathbb{P}(U_n - 2\mathbb{E}U_n \geq z/4)$.

(ii). Show that $2\mathbb{E}U_n \leq C_q a_* \mu_q c(n, q)$.

Part (i): For $g \in A_n$, since $|g|, |g'|$ are bounded by 1, by Lemma 21, $|g_\infty|$ and $|g'_\infty|$ are also bounded by 1. Then

$$|\tilde{\phi}_j(g)| = \Big|\sum_{i=1\vee j}^{n} \mathbb{E}\Big(\int_{a_{i-j}\epsilon'_j}^{a_{i-j}\epsilon_j} g'_\infty(x)\mathrm{d}x|\mathcal{F}_j\Big)\Big| \leq \sum_{i=1\vee j}^{n} \min\big\{|a_{i-j}|(|\epsilon_j| + \mu),\, 2\big\}. \tag{39}$$

Therefore for $j < -n$ and any $g \in A_n$, by (39),

$$|\tilde{\phi}_j(g)| \leq \min\{\gamma n(-j)^{-\beta}(|\epsilon_j| + \mu), 2n\}, \tag{40}$$

for $-n \leq j \leq n$ and any $g \in A_n$, by Lemma 22 and (39),

$$|\tilde{\phi}_j(g)| \leq \gamma K_{\beta,2/\gamma}(|\epsilon_j| + \mu)^{1/\beta}. \tag{41}$$

Denote $V = \max_{g\in A_n} \sum_{j\leq n} \mathbb{E}\tilde{\phi}_j^2(g)$. Hence by (40) and (41),

$$\begin{aligned} V &\leq \sum_{j<-n} (\gamma n(-j)^{-\beta})^{q'} \mathbb{E}(|\epsilon_0| + \mu)^{q'}(2n)^{2-q'} + (\gamma K_{\beta,2/\gamma})^2 \sum_{-n\leq j\leq n} \mathbb{E}[(|\epsilon_j| + \mu)^{2/\beta}] \\ &\leq \Big(\frac{4\gamma^2}{\beta - 1} + 2^{1+2/\beta}(\gamma K_{\beta,2/\gamma})^2\Big) n\mu_{q'}^{q'}. \end{aligned} \tag{42}$$

By (40),

$$\sum_{j<-n} \mathbb{E}\big(\max_{g\in A_n} |\tilde{\phi}_j|^{q\beta}\big) \leq \sum_{j<-n} (2n)^{q\beta-q}(\gamma n(-j)^{-\beta})^q \mathbb{E}[(|\epsilon_j| + \mu)^q] \leq \frac{2^{q\beta}\gamma^q}{q\beta - 1} n\mu_q^q. \tag{43}$$

By (41),

$$\sum_{-n\leq j\leq n} \mathbb{E}\big(\max_{g\in A_n} |\tilde{\phi}_j|^{q\beta}\big) \leq 2n(\gamma K_{\beta,2/\gamma})^{q\beta} \mathbb{E}[(|\epsilon_j| + \mu)^q] \leq 2^{q+1}(\gamma K_{\beta,2/\gamma})^{q\beta} n\mu_q^q. \tag{44}$$

Inserting the bounds (42), (43) and (44) into Lemma 20, we obtain

$$\begin{aligned}\mathbb{P}(U_n - 2\mathbb{E}U_n \geq z/4) \leq & e^{-z^2/(48V)} + \frac{4^{q\beta} K_{q\beta}}{z^{q\beta}} \sum_{j\leq n} \mathbb{E}\big(\max_{g\in A_n} |\tilde{\phi}_j|^{q\beta}\big) \\ \leq & \exp\Big(-\frac{z^2}{C_{\beta,\gamma}\mu_{q'}^{q'} n}\Big) + C_{\beta,q,\gamma,1}\mu_q^q \frac{n}{z^{q\beta}}, \end{aligned} \tag{45}$$

where $C_{\beta,\gamma} = 48(4\gamma^2/(\beta-1) + 2^{1+2/\beta}(\gamma K_{\beta,2/\gamma})^2)$ and $C_{\beta,q,\gamma,1} = 4^{q\beta} K_{q\beta}(2^{q\beta}\gamma^q/(q\beta-1) + 2^{q+1}(\gamma K_{\beta,2/\gamma})^{q\beta})$.

Part (ii): Recall $a_* = \sum_{k=0}^{\infty} |a_k|$. Note that $T_n(g)$ can be rewritten as

$$\begin{aligned} T_n(g) = \sum_{j\leq n} \tilde{\phi}_j(g) &= \sum_{k\geq 0}\sum_{i=1}^{n}\{g_\infty(a_k\epsilon_{i-k}) - \mathbb{E}g_\infty(a_k\epsilon_{i-k})\} \\ &= \sum_{k\geq 0}\int_{-\infty}^{\infty}\sum_{i=1}^{n}\big(\mathbf{1}_{a_k\epsilon_{i-k}\geq x} - \mathbb{P}(a_k\epsilon_{i-k}\geq x)\big) g'_\infty(x)\mathrm{d}x. \end{aligned}$$

Let $W_n(x) = \sum_{i=1}^{n}\big(\mathbf{1}_{\epsilon_i\geq x} - \mathbb{P}(\epsilon_i \geq x)\big)$. By Lemma 21, $|g'_\infty(x)| \leq 1$. Then

$$\begin{aligned} \mathbb{E}\Big[\max_{g\in A_n}\big|T_n(g)\big|\Big] &\leq \sum_{k\geq 0}\int_{-\infty}^{\infty}\mathbb{E}\Big|\sum_{i=1}^{n}\big(\mathbf{1}_{a_k\epsilon_{i-k}\geq x} - \mathbb{P}(a_k\epsilon_{i-k}\geq x)\big)\Big|\mathrm{d}x \\ &= \sum_{k\geq 0}\int_{-\infty}^{\infty}\mathbb{E}|W_n(x/a_k)|\mathrm{d}x = a_*\int_{-\infty}^{\infty}\mathbb{E}|W_n(y)|\mathrm{d}y, \end{aligned} \tag{46}$$

where the last equality is obtained by change of variables $y = x/a_k$ and $a_* = \sum_{k=0}^{\infty}|a_k|$. Let $T_F(x) = \mathbb{P}(|\epsilon_0| \geq |x|)$. Note that $\mathbb{E}|\mathbf{1}_{\epsilon_i\geq x} - \mathbb{P}(\epsilon_i \geq x)| = 2F_\epsilon(x)(1-F_\epsilon(x)) \leq 2T_F(x)$, and $\mathbb{E}(\mathbf{1}_{\epsilon_i\geq x} - \mathbb{P}(\epsilon_i\geq x))^2 = F_\epsilon(x)(1-F_\epsilon(x)) \leq T_F(x)$. Hence

$$\mathbb{E}|W_n(x)| \leq \min\{\|W_n(x)\|,\, 2nT_F(x)\} \leq \min\{\sqrt{n}T_F(x)^{1/2},\, 2nT_F(x)\}. \tag{47}$$

We have different bounds for (46) when $q > 2$, $1 < q < 2$ and $q = 2$. By Markov's inequality,

$$T_F(x) \leq \min\{|x|^{-q}\mu_q^q, 1\}. \tag{48}$$

When $q > 2$, we have

$$\int_{-\infty}^{\infty} T_F(x)^{1/2}\mathrm{d}y \leq 2\Big(\int_0^{\mu_q} 1\mathrm{d}x + \int_{\mu_q}^{\infty}|x|^{-q/2}\mu_q^{q/2}\mathrm{d}x\Big) = q/(q/2-1)\mu_q.$$

Inserting above into (46) and (47), we obtain

$$\mathbb{E}U_n \leq a_*\int_{-\infty}^{\infty}\mathbb{E}|W_n(x)|\mathrm{d}x \leq q/(q/2-1)a_*\mu_q\sqrt{n}. \tag{49}$$

When $\underline{1 < q < 2}$, by (47) and (48),

$$\begin{aligned}\int_{-\infty}^{\infty} \mathbb{E}|W_n(x)|\mathrm{d}x &\leq 2\Big(\int_0^{n^{1/q}\mu_q} \sqrt{n}x^{-q/2}\mu_q^{q/2}\mathrm{d}x + \int_{n^{1/q}\mu_q}^{\infty} 2nx^{-q}\mu_q^q\mathrm{d}x\Big) \\ &\leq 4(1/(2-q)+1/(q-1))\mu_q n^{1/q}.\end{aligned}$$

When $\underline{q=2}$, $\mathrm{I}_1 := \int_{|x|\leq\mu_2} \sqrt{n}T_F(x)^{1/2}\mathrm{d}x \leq 2\mu_2\sqrt{n}$. By (48), $\mathrm{I}_2 := \int_{|x|>n\mu_2} 2nT_F(x)\mathrm{d}x \leq 4\int_{n\mu_2}^{\infty} n\mu_2^2x^{-2}\mathrm{d}x = 4\mu_2$. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\mathrm{I}_3^2 &:= \left[\int_{\mu_2<|x|\leq n\mu_2} \sqrt{n}T_F(x)^{1/2}\mathrm{d}x\right]^2 \leq 4n\int_{\mu_2}^{n\mu_2} xT_F(x)\mathrm{d}x \int_{\mu_2}^{n\mu_2} x^{-1}\mathrm{d}x \\ &\leq 4n\int_0^{\infty} x\mathbb{P}(|\epsilon_0|\geq x)\mathrm{d}x(\log n) = 2\mathbb{E}(\epsilon_0^2)n\log(n) = 2\mu_2^2 n\log(n).\end{aligned}$$

Then by (47), $\int_{-\infty}^{\infty}\mathbb{E}|W_n(x)|\mathrm{d}x \leq \mathrm{I}_1+\mathrm{I}_2+\mathrm{I}_3 \leq 2\mu_2\sqrt{n}+4\mu_2+\mu_2(2n\log n)^{1/2}$. Combining the three cases $q>2$, $1<q<2$ and $q=2$, by (46), we have $\mathbb{E}U_n \leq c_q a_* \mu_q c(n,q)$. where $c_q = \max\{q/(q/2-1), 4(1/(2-q)+1/(q-1)), 6+\sqrt{2}\}$. ∎

Lemma 24 *Recall the definitions of* $\phi_j(g), \tilde{\phi}_j(g)$ *and* $T_n(g)$ *in* (32) *and* (35). *Under conditions of Theorem 4, we have* (38).

Proof Since $S_n(g) - \mathbb{E}S_n(g) - T_n(g)$ is the sum of martingale differences $\phi_j(g) - \tilde{\phi}_j(g)$, $j \leq n$, we can apply Lemma 18 to bound the tail probability. To this end, we shall:

(i). Derive the upper bound for $\mathrm{I}_1 = \sum_{j\leq n}\mathbb{P}(\max_{g\in A_n}|\phi_j(g)-\tilde{\phi}_j(g)|\geq u)$.

(ii). Bound the term $\mathrm{I}_2 = \max_{g\in A_n}\sum_{j\leq n}\mathbb{E}[(\phi_j(g)-\tilde{\phi}_j(g))^2|\mathcal{F}_{j-1}]$.

First we derive some inequalities that will be used for I_1 and I_2. Let $\epsilon_i, \epsilon_j', \epsilon_k'', i,j,k\in\mathbb{Z}$, be i.i.d. and $X_{i,j}'' = \sum_{k\geq 0} a_{i-j+k}\epsilon_{j-k}''$. Write $\phi_j(g) - \tilde{\phi}_j(g) = \sum_{i=1\vee j}^n d_{i,j}(g)$, where

$$d_{i,j}(g) = g_{i-j}(X_{i,j}) - g_{i-j+1}(X_{i,j-1}) - g_{\infty}(a_{i-j}\epsilon_j) + \mathbb{E}g_{\infty}(a_{i-j}\epsilon_j) \tag{D_1}$$

$$= \mathbb{E}\Big[\int_{a_{i-j}\epsilon_j'}^{a_{i-j}\epsilon_j} \big(g_{i-j}'(x+X_{i,j-1}) - g_{i-j}'(x+X_{i,j}'')\big)\mathrm{d}x\big|\mathcal{F}_j\Big] \tag{D_2}$$

$$= \mathbb{E}\Big[\int_{X_{i,j}''}^{X_{i,j-1}} \big(g_{i-j}'(x+a_{i-j}\epsilon_j) - g_{i-j}'(x+a_{i-j}\epsilon_j')\big)\mathrm{d}x\big|\mathcal{F}_j\Big] \tag{D_3}$$

$$= \mathbb{E}\Big[\int_{a_{i-j}\epsilon_j'}^{a_{i-j}\epsilon_j}\int_{X_{i,j}''}^{X_{i,j-1}} g_{i-j}''(x+y)\mathrm{d}y\mathrm{d}x\big|\mathcal{F}_j\Big]. \tag{D_4}$$

By Lemma 21, $|g_j|, |g_j{}'|$ and $|g_j{}''|$ are bounded by 1. Hence by (D_1)-(D_4), we have

$$\begin{aligned}&\max_{g\in A_n}|d_{i,j}(g)| \\ &\leq \min\Big\{4, 2|a_{i-j}|(|\epsilon_j|+\mu), 2(|X_{i,j-1}|+\mathbb{E}|X_{i,j}|), |a_{i-j}|(|\epsilon_j|+\mu)(|X_{i,j-1}|+\mathbb{E}|X_{i,j}|)\Big\} \\ &= \min\Big\{|a_{i-j}|(|\epsilon_j|+\mu), 2\Big\}\min\Big\{(|X_{i,j-1}|+\mathbb{E}|X_{i,j}|), 2\Big\}.\end{aligned} \tag{50}$$

Part (i): Recall $q' = \min(q,2)$. For $i > j$, by Lemma 19,

$$\|X_{i,j-1}\|_q^{q'} \le K_q^{q'} \sum_{k\ge 1} (|a_{i-j+k}| \|\epsilon_{j-k}\|_q)^{q'} \le \big(K_q^{q'} \gamma^{q'} (\beta q' - 1)^{-1}\big)(i-j)^{-q'\beta+1} \mu_q^{q'}. \quad (51)$$

Let $r = (q'\beta - 1)/(2q')$, by Markov's inequality,

$$\mathrm{I}_1 \le \sum_{-n\le j\le n} u^{-q(\beta+r)} \mathbb{E}[\max_{g\in A_n} |\phi_j(g) - \tilde{\phi}_j(g)|^{q(\beta+r)}] + \sum_{j<-n} u^{-q} \mathbb{E}[\max_{g\in A_n} |\phi_j(g) - \tilde{\phi}_j(g)|^q]. \quad (52)$$

We shall consider the two cases $-n \le j \le n$ and $j < -n$ separately. For $-n \le j \le n$, by (50) and since ϵ_j and $X_{i,j-1}$ are independent,

$$\begin{aligned}
&\|\max_{g\in A_n} |\phi_j(g) - \tilde{\phi}_j(g)|\|_{q(\beta+r)} \\
&\le \sum_{i=j\vee 1}^{n} \big\|\min\{|a_{i-j}|(|\epsilon_j| + \mu), 2\}\big\|_{q(\beta+r)} \big\|\min\{|X_{i,j-1}| + \mathbb{E}|X_{i,j}|, 2\}\big\|_{q(\beta+r)} \\
&\le \sum_{i=j\vee 1}^{n} \Big(|a_{i-j}|^q \mathbb{E}(|\epsilon_j| + \mu)^q 2^{q(\beta+r)-q}\Big)^{1/q(\beta+r)} \Big(\mathbb{E}(|X_{i,j-1}| + \mathbb{E}|X_{i,j}|)^q 2^{q(\beta+r)-q}\Big)^{1/q(\beta+r)}.
\end{aligned}$$

By (51) and $2\beta q' - 1 > (\beta + r)q'$, above inequality is further bounded by

$$\|\max_{g\in A_n} |\phi_j(g) - \tilde{\phi}_j(g)|\|_{q(\beta+r)} \le c_1 \sum_{i=j\vee 1}^{n} ((i-j)\vee 1)^{\frac{-2\beta q'+1}{(\beta+r)q'}} \mu_q^{2/(\beta+r)} \le c_2 \mu_q^{2/(\beta+r)}, \quad (53)$$

where $c_1 = (K_q \gamma (\beta q' - 1)^{-1} 2^{\beta+r})^{1/(\beta+r)}$ and $c_2 = 4(2\beta q' - 1)(\beta q' - 1)^{-1} c_1$.

For $j < -n$, again by (50) and the independence between ϵ_j and $X_{i,j-1}$,

$$\begin{aligned}
\|\max_{g\in A_n} |\phi_j(g) - \tilde{\phi}_j(g)|\|_q &\le \sum_{i=1}^{n} |a_{i-j}| \||\epsilon_j| + \mu\|_q \||X_{i,j-1}| + \mathbb{E}|X_{i,j}|\|_q \\
&\le \big(4\gamma(\beta q' - 1)^{1/q'}\big) n (-j)^{-\frac{2\beta q' - 1}{q'}} \mu_q^2, \quad (54)
\end{aligned}$$

where the last inequality is due to (51).

Applying (53) and (54) to (52), we have

$$\mathrm{I}_1 \le c_3 \mu_q^{2q} n u^{-\beta(q+r)}, \text{ where } c_3 = 2c_2^{q(\beta+r)} + (4\gamma(\beta q' - 1)^{1/q'})^q.$$

Part (ii): We shall bound $\max_{g\in A_n} |\phi_j(g) - \tilde{\phi}_j(g)|$ for $-n \le j \le n$ and $j < -n$ seperately. For $-n \le j \le n$, by (50) and Lemma 22,

$$\max_{g\in A_n} |\phi_j(g) - \tilde{\phi}_j(g)| \le \sum_{i=1\vee j}^{n} \min\{|a_{i-j}|(|\epsilon_j| + \mu), 2\} \le \gamma K_{\beta,\gamma/2} (|\epsilon_j| + \mu)^{1/\beta},$$

Since ϵ_j is independent of $\mathcal{F}_{j-1}$, we have

$$\begin{aligned}\mathrm{I}_{21} := &\sum_{-n\le j\le n} \mathbb{E}\big[\max_{g\in A_n}(\phi_j(g)-\tilde\phi_j(g))^2|\mathcal{F}_{j-1}\big]\\ \le &\sum_{-n\le j\le n}(\gamma K_{\beta,\gamma/2})^2\mathbb{E}[(|\epsilon_j|+\mu)^{2/\beta}] \le \big(2^{1+2/\beta}(\gamma K_{\beta,\gamma/2})^2\big)n\mu_{q'}^{q'}.\end{aligned}$$

For $j<-n$, by Lemma 22,

$$\max_{g\in A_n}|\phi_j(g)-\tilde\phi_j(g)| \le n\min\{\gamma(-j)^{-\beta}(|\epsilon_j|+\mu),2\}.$$

Since ϵ_j is independent of $\mathcal{F}_{j-1}$, we have

$$\begin{aligned}\mathrm{I}_{22} := &\sum_{j<-n}\mathbb{E}\big[\max_{g\in A_n}(\phi_j(g)-\tilde\phi_j(g))^2|\mathcal{F}_{j-1}\big]\\ \le n^2 &\sum_{j<-n}2^{2-q'}\gamma^{q'}(-j)^{-q'\beta}\mathbb{E}[(|\epsilon_j|+\mu)^{q'}] \le \big(4\gamma^2/(\beta-1)\big)n\mu_{q'}^{q'},\end{aligned}$$

Hence we have $\mathrm{I}_2=\mathrm{I}_{21}+\mathrm{I}_{22}\le c_4 n\mu_{q'}^{q'}$, where $c_4=2^{1+2/\beta}(\gamma K_{\beta,\gamma/2})^2+4\gamma^{q'}/(q'\beta-1)$.

Inserting the bounds for I_1 and I_2 into Lemma 18 leads to

$$\mathbb{P}(\Omega_n\ge z/4)\le c_3 n\mu_q^{2q}u^{-q(\beta+r)}+2|A_n|\exp\Big(-\frac{z^2}{32c_4\mu_{q'}^{q'}n}\Big)+2|A_n|\exp\Big(-\frac{z^2}{8zu}\Big). \tag{55}$$

Take $u=z^{\beta/(\beta+r)}\mu_q^{1/(\beta+r)}$ and we complete the proof. ■

Remark 25 *Let $K_{q\beta}$ (resp. K_q and $K_{\beta,2/\gamma}$) be the constant in Lemma 20 (resp. Lemma 19 and Lemma 22). With a careful check of the proofs of Theorem 4, Lemmas 23 and 24, we can choose constants in Theorem 4 as follows:*

- $C_q=2\max\{q/(q/2-1),4(1/(2-q)+1/(q-1)),6+\sqrt{2}\}$.
- $C_{\beta,q,\gamma}=C_{\beta,q,\gamma,1}+C_{\beta,q,\gamma,2}$, *where* $C_{\beta,q,\gamma,1}=4^{q\beta}K_{q\beta}(2^{q\beta}\gamma^q/(q\beta-1)+2^{q+1}(\gamma K_{\beta,2/\gamma})^{q\beta})$ *and* $C_{\beta,q,\gamma,2}=2c_2^{q(\beta+r)}+(4\gamma(\beta q'-1)^{1/q'})^q$ *with* $r=(q'\beta-1)/(2q')$, $c_1=(K_q\gamma(\beta q'-1)^{-1}2^{\beta+r})^{1/(\beta+r)}$ *and* $c_2=4(2\beta q'-1)(\beta q'-1)^{-1}c_1$.
- $C_{\beta,\gamma}=48\big(4\gamma^2/(\beta-1)+2^{1+2/\beta}(\gamma K_{\beta,2/\gamma})^2\big)$.

Proof [Proof of Proposition 5] Construct A_n as in the proof of Theorem 4. Recall (31) for the function g_k. Note that $g_1(X_{i,i-1})=\mathbb{E}[g(X_i)|\mathcal{F}_{i-1}]$. By (36), we have

$$\begin{aligned}\mathbb{P}(|\Delta_n|\ge a+z)\le&\mathbb{P}\Big(\max_{g\in A_n}|S_n(g)-\mathbb{E}S_n(g)|\ge a+z/2\Big)\\ \le&\mathbb{P}\Big(\max_{g\in A_n}\Big|\sum_{i=1}^n\big(g_1(X_{i,i-1})-\mathbb{E}g_1(X_{i,i-1})\big)\Big|\ge a+z/4\Big)\\ &+\sum_{g\in A_n}\mathbb{P}\Big(\Big|\sum_{i=1}^n(g(X_i)-\mathbb{E}[g(X_i)|\mathcal{F}_{i-1}])\Big|\ge z/4\Big)=:\mathrm{I}_1+\mathrm{I}_2,\end{aligned}$$

where $a = C_q a_* \mu_q c(n,q)$.

Since $|g| \le 1$ and $g(X_i) - \mathbb{E}[g(X_i)|\mathcal{F}_{i-1}]$, $1 \le i \le n$, are martingale differences, by Azuma's inequality, $\mathrm{I}_2 \le 2|A_n|\exp\{-z^2/(64n)\}$. For I_1, notice

$$g_1(x) = \int_{-\infty}^{\infty} g(x+y) f_\epsilon(y)\mathrm{d}y = \int_{-\infty}^{\infty} g(y) f_\epsilon(y-x)\mathrm{d}y.$$

By Assumption (A'), $\sup_{g\in\mathcal{A}} |g_1|_\infty$, $\sup_{g\in\mathcal{A}} |g_1'|_\infty$ and $\sup_{g\in\mathcal{A}} |g_1''|_\infty$ are all bounded by 1. Thus in the I_1 part, the function g_1 satisfies Assumption (A) and can be dealt with by Theorem 4. Combining I_1 and I_2, we complete the proof. ■

6.3 Proof of Theorem 7

Proof [Proof of Theorem 7] Recall the projection operator $P_k \cdot = \mathbb{E}(\cdot|\mathcal{F}_k) - \mathbb{E}(\cdot|\mathcal{F}_{k-1})$. Let $D_k = P_k \Delta_n$, $k \le n$. Then $\Delta_n - \mathbb{E}\Delta_n = \sum_{k\le n} D_k$ and

$$\mathbb{P}(\Delta_n - \mathbb{E}\Delta_n \ge z) \le \mathbb{P}\Big(\sum_{k\le -n} D_k \ge z/2\Big) + \mathbb{P}\Big(\sum_{-n<k\le n} D_k \ge z/2\Big) =: \mathrm{I}_1 + \mathrm{I}_2. \tag{56}$$

Then the theorem follows from the following three claims which will be proved in the sequel:

(i). $\mathrm{I}_1 \le C_{q,\gamma} e^{-qn(1-\rho)} \mu_q^q (z^q (1-\rho)^{q+q/q'})^{-1}$.

(ii). $\mathrm{I}_2 \le \exp\{-C'_{q,\gamma} z^2 (1-\rho)^2 ((\mu_q^q \vee 1)n)^{-1}\}$.

(iii). $\mathbb{E}\Delta_n \le C_q \mu_q c^*(n,\rho,q)$.

To prove (i) and (ii), we need to apply coupling. Let $\epsilon_i, \epsilon'_j, i, j \in \mathbb{Z}$, be i.i.d. For a random variable $Z = H(\varepsilon)$, where H is a measurable function and $\varepsilon = (\epsilon_i)_{i\in\mathbb{Z}}$, we define the coupled version $Z_{\{j\}} = H(\varepsilon'_{\{j\}})$, where $\varepsilon'_{\{j\}} = (\dots, \epsilon_{j-1}, \epsilon'_j, \epsilon_{j+1}, \dots)$. We shall now derive an upper bound for $|D_k|$. Since $|g|, |g'|$ are bounded by 1, for any $k \le i$,

$$\mathbb{E}\big(\sup_{g\in\mathcal{A}} |g(X_i) - g(X_{i,\{k\}})| \big| \mathcal{F}_k\big) \le \mathbb{E}\big(|X_i - X_{i,\{k\}}| \big| \mathcal{F}_k\big) \le |a_{i-k}|(|\epsilon_k| + \mu). \tag{57}$$

Note $\mathbb{E}(\Delta_n|\mathcal{F}_{k-1}) = \mathbb{E}(\Delta_{n,\{k\}}|\mathcal{F}_k)$, thus $D_k = \mathbb{E}(\Delta_n - \Delta_{n,\{k\}}|\mathcal{F}_k)$ and by (57),

$$|D_k| \le \mathbb{E}\bigg(\sup_{g\in\mathcal{A}} \Big|\sum_{i=1}^n [g(X_i) - g(X_{i,\{k\}})]\Big| \bigg| \mathcal{F}_k\bigg) \le \sum_{i=1\vee k}^n \min\{|a_{i-k}|(|\epsilon_k| + \mu), 2\}. \tag{58}$$

Part (i): Since D_k are martingale differences, by Lemma 19,

$$\mathrm{I}_1 \le (z/2)^{-q} \Big\|\sum_{k\le -n} D_k\Big\|_q^q \le K_q^q (z/2)^{-q} \Big(\sum_{k\le -n} \|D_k\|_q^{q'}\Big)^{q/q'}. \tag{59}$$

Since (58) implies $|D_k| \le \gamma\rho^{-k}(1-\rho)^{-1}(|\epsilon_k| + \mu)$ for any $k \le -n$, we further obtain from (59) and the elementary inequality $\log(\rho^{-1}) \ge 1 - \rho$ that

$$\mathrm{I}_1 \le (4K_q\gamma)^q \frac{\rho^{nq}\mu_q^q}{z^q (1-\rho)^q (1-\rho^{q'})^{q/q'}} \le (4K_q\gamma)^q \frac{e^{-nq(1-\rho)}\mu_q^q}{z^q (1-\rho)^{q+q/q'}}. \tag{60}$$

Part (ii): Note for any $y \geq 1$, since $\log(\rho^{-1}) \geq 1-\rho$,

$$\begin{aligned}\sum_{i\geq 0}\min(\rho^i y, 1) &\leq \sum_{i\geq -\log_\rho y}\rho^i y + (-\log_\rho y)\\ &\leq (1-\rho)^{-1} - \log_\rho y \leq (1-\rho)^{-1}[1+\log(y)]. \end{aligned}\tag{61}$$

Hence for $k > -n$, by (58) and (61),

$$|D_k| \leq (2\vee\gamma)(1-\rho)^{-1}\big[1+\log(|\epsilon_k|+\mu)\mathbf{1}_{\{|\epsilon_k|+\mu\geq 1\}}\big]. \tag{62}$$

Let $h^* := (2\vee\gamma)^{-1}(1-\rho)q$. Since $\epsilon_k \in \mathcal{L}^q$, for any $0 < h \leq h^*$, $\mathbb{E}(e^{D_k h}) < \infty$. Note $\mathbb{E}(D_k|\mathcal{F}_{k-1}) = 0$, then

$$\begin{aligned}\mathbb{E}(e^{D_k h}|\mathcal{F}_{k-1}) &= 1 + \mathbb{E}(e^{D_k h} - D_k h - 1|\mathcal{F}_{k-1})\\ &\leq 1 + \mathbb{E}\Big[\frac{e^{|D_k|h} - |D_k|h - 1}{h^2(1-\rho)^{-2}}\Big|\mathcal{F}_{k-1}\Big]h^2(1-\rho)^{-2}\end{aligned}\tag{63}$$

in view of $e^x - x \leq e^{|x|} - |x|$ for any x. Note that for any fixed $x > 0$, $(e^{tx} - tx - 1)/t^2$ is increasing on $t \in (0,\infty)$. Applying the upper bound of D_k in (62), we have

$$\begin{aligned}\mathbb{E}\Big[\frac{e^{|D_k|h} - |D_k|h - 1}{h^2(1-\rho)^{-2}}\Big|\mathcal{F}_{k-1}\Big] &\leq \mathbb{E}\Big[\frac{e^{|D_k|h^*} - |D_k|h^* - 1}{h^{*2}(1-\rho)^{-2}}\Big|\mathcal{F}_{k-1}\Big]\\ &\leq \mathbb{E}\Big[\frac{e^{q[1+\log(|\epsilon_k|+\mu)1_{\{|\epsilon_k|+\mu>1\}}]}}{h^{*2}(1-\rho)^{-2}}\Big|\mathcal{F}_{k-1}\Big] \leq c_1\mu_q^q,\end{aligned}\tag{64}$$

where $c_1 = 2^q e^q (2\vee\gamma)^2 q^{-2}$. Hence for any $h \leq h^*$,

$$\mathbb{E}(e^{D_k h}|\mathcal{F}_{k-1}) \leq 1 + c_1\mu_q^q h^2(1-\rho)^{-2}. \tag{65}$$

By Markov's inequality we have $\mathrm{I}_2 \leq e^{-zh/2}\mathbb{E}[\exp(\sum_{-n<k\leq n} D_k h)]$. Then by recursively applying (65), let $h = z(1-\rho)^2[8c_1(\mu_q^q\vee 1)n]^{-1} \leq h^*$, we further obtain

$$\begin{aligned}\mathrm{I}_2 \leq& e^{-zh/2}\mathbb{E}\Big(e^{\sum_{k=-n+1}^{n-1} D_k h}\mathbb{E}(e^{D_n h}|\mathcal{F}_{n-1})\Big) \leq e^{-zh/2}(1 + c_1\mu_q^q h^2/(1-\rho)^2)^{2n}\\ \leq& \exp\big(-zh/2 + 2nc_1\mu_q^q h^2/(1-\rho)^2\big) \leq \exp\Big(-\frac{z^2(1-\rho)^2}{32c_1(\mu_q^q\vee 1)n}\Big),\end{aligned}\tag{66}$$

where the third inequality is due to $1 + x \leq e^x$ for $x > 0$.

Part (iii): Note

$$\begin{aligned}g(X_i) - \mathbb{E}g(X_i) &= \sum_{j\geq 0}(g_j(X_{i,i-j}) - g_{j+1}(X_{i,i-j-1}))\\ &= \sum_{j\geq 0}\int_{-\infty}^{\infty} g_j'(x)\big(\mathbf{1}_{x\leq X_{i,i-j}} - \mathbb{E}(\mathbf{1}_{x\leq X_{i,i-j}}|\mathcal{F}_{i-j-1})\big)\mathrm{d}x.\end{aligned}$$

By above inequality and that $|g_j'|$ are bounded by 1,

$$\mathbb{E}(\Delta_n) \le \sum_{j\ge 0} \int_{-\infty}^{\infty} \mathbb{E}\Big| \sum_{i=1}^{n} \big(\mathbf{1}_{x\le X_{i,i-j}} - \mathbb{E}(\mathbf{1}_{x\le X_{i,i-j}}|\mathcal{F}_{i-j-1})\big)\Big| \mathrm{d}x. \tag{67}$$

Let $H_j(x) = \mathbb{P}(|X_{0,-j}| \ge |x|)$. Since for any fixed j, $\mathbf{1}_{x\le X_{i,i-j}} - \mathbb{E}(\mathbf{1}_{x\le X_{i,i-j}}|\mathcal{F}_{i-j-1})$, $i = 1,\ldots,n$, are martingale differences, by the same arguments as for (47), we have

$$\mathbb{E}\Big| \sum_{i=1}^{n} \big(\mathbf{1}_{x\le X_{i,i-j}} - \mathbb{E}(\mathbf{1}_{x\le X_{i,i-j}}|\mathcal{F}_{i-j-1})\big)\Big| \le \min\{\sqrt{n}H_j(x)^{1/2}, nH_j(x)\}. \tag{68}$$

For any $1 < r \le q$ and $r' = \min\{r,2\}$, by Lemma 19,

$$H_j(x) \le \frac{\|X_{0,-j}\|_r^r}{|x|^r} \le \frac{K_r^r}{|x|^r}\Big(\sum_{k\ge j}|a_k|^{r'}\mu_r^{r'}\Big)^{r/r'} \le \frac{K_r^r}{|x|^r}\rho^{jr}(1-\rho)^{-r/r'}\mu_r^r. \tag{69}$$

We need to deal with the three cases separately: $q > 2$, $1 < q < 2$ and $q = 2$.
Case $q > 2$: Let $r = 3/2$. By (67) and (68),

$$\mathbb{E}(\Delta_n) \le \sum_{j\ge 0}\sqrt{n}\int_{-\infty}^{\infty} H_j(x)^{1/2}\mathrm{d}x.$$

Since $1-\rho^x \ge 1-\rho$ for $x \ge 1$ and $1-\rho^x \ge 1-\rho^{1/2} \ge (1-\rho)/2$ for $1/2 \le x < 1$, by (69),

$$\begin{aligned}\mathbb{E}(\Delta_n) &\le (K_r \vee K_q)^{q/2}\sum_{j\ge 0}\sqrt{n}\Big(\int_{|x|>\mu_q}\rho^{jq/2}(1-\rho)^{-q/4}|x|^{-q/2}\mu_q^{q/2}\mathrm{d}x \\ &\quad + \int_{|x|\le\mu_q}\rho^{jr/2}(1-\rho)^{-1/2}|x|^{-r/2}\mu_r^{r/2}\mathrm{d}x\Big) \\ &\le (K_r \vee K_q)^{q/2}(2/(q-2)+8)\sqrt{n}(1-\rho)^{-3/2}\mu_q.\end{aligned}$$

Case $1 < q < 2$: By (67) and (68), for $a = n^{1/q}(1-\rho)^{-1/q}K_q\mu_q$,

$$\mathbb{E}(\Delta_n) \le \sum_{j\ge 0}\Big(\int_{|x|>a} nH_j(x)\mathrm{d}x + \int_{|x|\le a} n^{1/2}H_j(x)^{1/2}\mathrm{d}x\Big).$$

By (69), we further obtain

$$\begin{aligned}\mathbb{E}(\Delta_n) &\le K_q^q n\sum_{j\ge 0}\int_{|x|>a}\frac{\rho^{jq}\mu_q^q}{(1-\rho)|x|^q}\mathrm{d}x + K_q^{q/2}n^{1/2}\sum_{j\ge 0}\int_{|x|\le a}\frac{\rho^{qj/2}\mu_q^{q/2}}{(1-\rho)^{1/2}|x|^{q/2}}\mathrm{d}x \\ &\le (1/(q-1)+2/(2-q))K_q n^{1/q}(1-\rho)^{-1/q-1}\mu_q.\end{aligned}$$

Case $q = 2$: Take $a = n^{1/2}(1-\rho)^{-1/2}\mu_2$, $b = \mu_2$, then by (67) and (68),

$$\mathbb{E}(\Delta_n) \le \sum_{j\ge 0}\Big(\int_{|x|>a} nF_j(x)\mathrm{d}x + \int_{b<|x|\le a} n^{1/2}F_j(x)^{1/2}\mathrm{d}x + \int_{|x|\le b} n^{1/2}F_j(x)^{1/2}\mathrm{d}x\Big)$$

By (69), for $r = 3/2$,

$$\begin{aligned}\mathbb{E}(\Delta_n) \le{}& n\sum_{j\ge 0}\int_{|x|>a}\frac{\rho^{2j}\mu_2^2}{(1-\rho)|x|^2}\mathrm{d}x + n^{1/2}\sum_{j\ge 0}\int_{b<|x|\le a}\frac{\rho^j\mu_2}{(1-\rho)^{1/2}|x|}\mathrm{d}x\\ &+ n^{1/2}\sum_{j\ge 0}\int_{|x|\le b}\frac{\rho^{jr/2}\mu_2^{r/2}}{(1-\rho)^{1/2}|x|^{r/2}}\mathrm{d}x \le \frac{10\sqrt{n}\mu_2}{(1-\rho)^{3/2}}\log(n(1-\rho)^{-1}).\end{aligned}$$

■

Remark 26 *Let $K_{3/2}$ and K_q be the constants defined in Lemma 19. With a careful check of the proof of Theorem 7, we can choose constants in Theorem 7 as follows:*

- $C_q = \max\{(K_{3/2}\vee K_q)^{q/2}(2/(q-2)+8), (1/(q-1)+2/(2-q))K_q, 10\}$,
- $C_{q,\gamma} = (4K_q\gamma)^q$,
- $C'_{q,\gamma} = 2^{q+5}e^q(2\vee\gamma)^2q^{-2}$.

6.4 Proofs of Theorem 8 and Proposition 9

Proof [Proof of Theorem 8] The idea of proving Theorem 8 is similar to the proof of Theorem 4. Recall the definitions of $\phi_j(g)$, $\tilde{\phi}_j(g)$, $T_n(g)$ in (32), (35) and definitions of Ω_n, U_n in (36). Then the same argument as in Theorem 4 leads to

$$\begin{aligned}\mathbb{P}(\Delta_n \ge C'_{\beta,q,\gamma}\mu_q n^{3/2-\beta} + z) \le{}& \mathbb{P}\Big(\max_{g\in A_n}|S_n(g) - \mathbb{E}S_n(g)| \ge C'_{\beta,q,\gamma}\mu_q n^{3/2-\beta} + z/2\Big).\\ \le{}& \mathbb{P}\Big(U_n \ge C'_{\beta,q,\gamma}\mu_q n^{3/2-\beta} + z/4\Big) + \mathbb{P}(\Omega_n \ge z/4). \qquad (70)\end{aligned}$$

Again we shall use $T_n(g)$ to approximate $S_n(g) - \mathbb{E}S_n(g)$, and apply Fuk-Nagaev's inequality to deal with $T_n(g)$ part. By Lemma 27,

$$\mathbb{P}\Big(U_n \ge C'_{\beta,q,\gamma}\mu_q n^{3/2-\beta} + z/4\Big) \le C_{\beta,q,\gamma,1}\mu_q^q\frac{n^{1+(1-\beta)q}}{z^q} + \exp\Big(-\frac{z^2}{C_{\beta,\gamma}n^{3-2\beta}\mu_2^2}\Big), \qquad (71)$$

and by Lemma 28,

$$\mathbb{P}(\Omega_n \ge \frac{z}{4}) \le C_{\beta,q,\gamma,2}\mu_q^{2q}\frac{n^{1+(1-\beta)q}[\log|A_n| + \log(n)]^q}{\tilde{c}^q(n,\beta)z^q} + 2|A_n|\exp\Big(-\frac{z^2}{C_{\beta,\gamma}n^{3-2\beta}\mu_2^2}\Big). \qquad (72)$$

Combining (70), (71) and (72) with $C_{\beta,q,\gamma} = C_{\beta,q,\gamma,1} + C_{\beta,q,\gamma,2}$, the result follows. ■

Lemma 27 *Recall the definitions of $\phi_j(g)$, $T_n(g)$ in (32), (35) and U_n in (36). Under assumptions of Theorem 8, we have (71).*

Proof The proof is similar to the one of Lemma 23. We shall

(i). Bound the probability $\mathbb{P}(U_n - 2\mathbb{E}U_n \geq z/4)$.

(ii). Bound the expectation $\mathbb{E}U_n$.

Part (i): For $j < -n$, by (39),

$$|\tilde{\phi}_j(g)| \leq \sum_{i=1}^{n} |a_{i-j}|(|\epsilon_j| + \mu) \leq \gamma n(-j)^{-\beta}(|\epsilon_j| + \mu). \tag{73}$$

For $-n \leq j \leq n$, by (39),

$$|\tilde{\phi}_j(g)| \leq \sum_{i=1\vee j}^{n} |a_{i-j}|(|\epsilon_j| + \mu) \leq 2(1-\beta)^{-1}\gamma n^{1-\beta}(|\epsilon_j| + \mu). \tag{74}$$

Denote $V = \max_{g\in A_n}\sum_{j\leq n}\mathbb{E}\tilde{\phi}_j^2(g)$. Hence by (73) and (74),

$$V \leq \max_{g\in A_n}\sum_{j<-n}\mathbb{E}|\tilde{\phi}_j|^2 + \max_{g\in A_n}\sum_{j=-n}^{n}\mathbb{E}|\tilde{\phi}_j|^2 \leq c_1\mu_2^2 n^{3-2\beta}, \tag{75}$$

where $c_1 = 4\gamma^2((2\beta-1)^{-1} + 8(1-\beta)^{-2})$. Also by (73) and (74), we have

$$\sum_{j\leq n}\mathbb{E}\big(\max_{g\in A_n}|\tilde{\phi}_j|^q\big) \leq \sum_{j<-n}\mathbb{E}\big(\max_{g\in A_n}|\tilde{\phi}_j|^q\big) + \sum_{j=-n}^{n}\mathbb{E}\big(\max_{g\in A_n}|\tilde{\phi}_j|^q\big) \leq c_2 n^{1+(1-\beta)q}\mu_q^q, \tag{76}$$

where $c_2 = \gamma^q 2^q(2^{1+q}(1-\beta)^{-q} + (q\beta-1)^{-1})$.

Using the bounds (75) and (76) in the Fuk-Nagaev inequality Lemma 20, we obtain

$$\mathbb{P}\Big(U_n - 2\mathbb{E}U_n \geq z/4\Big) \leq \exp\Big(-\frac{z^2}{C_{\beta,\gamma}n^{3-2\beta}\mu_2^2}\Big) + C_{\beta,q,\gamma,1}\mu_q^q\frac{n^{1+(1-\beta)q}}{z^q}. \tag{77}$$

Part (ii): By Lemma 21 the derivatives $|g'_\infty(x)| \leq 1$, thus

$$\begin{aligned}
\max_{g\in A_n}|T_n(g)| &= \max_{g\in A_n}\Big|\int_{-\infty}^{\infty}\sum_{j\leq n}\sum_{i=1\vee j}^{n}\big(\mathbf{1}_{a_{i-j}\epsilon_j\geq x} - \mathbb{P}(a_{i-j}\epsilon_j \geq x)\big)g'_\infty(x)\mathrm{d}x\Big| \\
&\leq \int_{-\infty}^{\infty}\Big|\sum_{j\leq -n}\sum_{i=1\vee j}^{n}\big(\mathbf{1}_{a_{i-j}\epsilon_j\geq x} - \mathbb{P}(a_{i-j}\epsilon_j \geq x)\big)\Big|\mathrm{d}x \\
&+ \int_{-\infty}^{\infty}\Big|\sum_{-n\leq j\leq n}\sum_{i=1\vee j}^{n}\big(\mathbf{1}_{a_{i-j}\epsilon_j\geq x} - \mathbb{P}(a_{i-j}\epsilon_j \geq x)\big)\Big|\mathrm{d}x =: \mathrm{I}_1 + \mathrm{I}_2.
\end{aligned}$$

For I_1 : since ϵ_j are independent,

$$\begin{aligned}\mathbb{E}(\mathrm{I}_1) &\le \sum_{i=1}^{n} \int_{-\infty}^{\infty} \Big\| \sum_{j<-n} \big(\mathbf{1}_{a_{i-j}\epsilon_j \ge x} - \mathbb{P}(a_{i-j}\epsilon_j \ge x)\big) \Big\| \mathrm{d}x \\ &= \sum_{i=1}^{n} \int_{-\infty}^{\infty} \Big[\sum_{j<-n} (1 - F_\epsilon(x/a_{i-j}))F_\epsilon(x/a_{i-j}) \Big]^{1/2} \mathrm{d}x. \qquad (78)\end{aligned}$$

Denote $F^*(x) = \mathbb{P}(|\epsilon_0| \ge |x|)$, then

$$\max\big\{F_\epsilon(x) \wedge (1 - F_\epsilon(x)), F_\epsilon(-x) \wedge (1 - F_\epsilon(-x))\big\} \le F^*(x).$$

Since $F^*(x)$ decreases in $|x|$ and $|a_{i-j}| \le \gamma(-j)^\beta$, (78) can be further bounded by

$$\begin{aligned}\mathbb{E}(\mathrm{I}_1) &\le 2\sum_{i=1}^{n} \int_{0}^{\infty} \Big[\sum_{j<-n} F^*(x/a_k) \Big]^{1/2} \mathrm{d}x \\ &\le 2n \int_{0}^{\infty} \Big[\sum_{j<-n} F^*(x\gamma^{-1}(-j)^\beta) \Big]^{1/2} \mathrm{d}x \\ &\le 2n \int_{0}^{\infty} \Big[\int_{n}^{\infty} F^*(\gamma^{-1}xy^\beta)\mathrm{d}y \Big]^{1/2} \mathrm{d}x \\ &= 2n^{3/2-\beta}\gamma \int_{0}^{\infty} \Big[\int_{1}^{\infty} F^*(xy^\beta)\mathrm{d}y \Big]^{1/2} \mathrm{d}x, \qquad (79)\end{aligned}$$

where the last equality is due to a change of variables: $x \mapsto n^\beta x/\gamma$, $y \mapsto y/n$.

Let $r = 1 + 1/(2\beta)$. Then $1/\beta < r < 2$. Since $r < q$, we have $F^*(x) \le |x|^{-r}\mu_q^r$, $F^*(x) \le |x|^{-q}\mu_q^q$ and

$$\begin{aligned}\int_{0}^{\infty} \Big[\int_{1}^{\infty} F^*(xy^\beta)\mathrm{d}y \Big]^{1/2} \mathrm{d}x \;&\le\; \int_{0}^{\mu_q} \Big[\int_{1}^{\infty} x^{-r}y^{-r\beta}\mu_q^r \mathrm{d}y \Big]^{1/2} \mathrm{d}x \\ &+\; \int_{\mu_q}^{\infty} \Big[\int_{1}^{\infty} x^{-q}y^{-q\beta}\mu_q^q \mathrm{d}y \Big]^{1/2} \mathrm{d}x \le c_3\mu_q,\end{aligned}$$

where $c_3 = 2(2-r)^{-1}(r\beta - 1)^{-1} + 2(q-2)^{-1}(q\beta - 1)^{-1/2}$.

For I_2: Since ϵ_j are independent, we have

$$\begin{aligned}\mathbb{E}(\mathrm{I}_2) &= \mathbb{E}\int_{-\infty}^{\infty} \Big| \sum_{k=0}^{2n} \sum_{i=(k-n)\vee 1}^{n} \big(\mathbf{1}_{a_k\epsilon_{i-k} \ge x} - \mathbb{P}(a_k\epsilon_{i-k} \ge x)\big) \Big| \mathrm{d}x \\ &\le \sum_{k=0}^{2n} \int_{-\infty}^{\infty} \Big(\sum_{i=(k-n)\vee 1}^{n} (1 - F_\epsilon(x/a_k))F_\epsilon(x/a_k) \Big)^{1/2} \mathrm{d}x \\ &\le \sum_{k=0}^{2n} |a_k| \int_{-\infty}^{\infty} [nF^*(x)]^{1/2}\mathrm{d}x \le \gamma(1-\beta)^{-1}2^{1-\beta}n^{3/2-1} \int_{-\infty}^{\infty} F^*(x)^{1/2}\mathrm{d}x,\end{aligned}$$

where the second inequality is by a change of variable and the last inequality is by $|a_k| \le \gamma k^{-\beta}$. Note by definition of $F^*(x)$,

$$\int_{-\infty}^{\infty} F^*(x)^{1/2}\mathrm{d}x = 2\int_0^{\mu_q} 1\mathrm{d}x + 2\int_{\mu_q}^{\infty} F^*(x)^{1/2}\mathrm{d}x \le 2q/(q-2)\mu_q.$$

Combining I_1 and I_2, $\mathbb{E}U_n \le c_4\mu_q n^{3/2-\beta}$, where $c_4 = 2\gamma c_3 + 4\gamma(1-\beta)^{-1}q(q-2)^{-1}$. ■

Lemma 28 *Recall the definitions of* $\phi_j(g)$, $\tilde{\phi}_j(g)$, $T_n(g)$ *in* (32), (35) *and definition of* Ω_n *in* (36). *Under assumptions of Theorem 8, we have* (72).

Proof The argument is similar to the proof of Lemma 24, that is, we shall apply Lemma 18 to bound the tail probability. To this aim, we need to:

(i). Derive the upper bound for $\mathrm{I}_1 = \sum_{j\le n}\mathbb{P}(\max_{g\in A_n}|\phi_j(g) - \tilde{\phi}_j(g)| \ge u)$.

(ii). Bound the term $\mathrm{I}_2 = \max_{g\in A_n}\sum_{j\le n}\mathbb{E}[(\phi_j(g) - \tilde{\phi}_j(g))^2|\mathcal{F}_{j-1}]$.

Part (i): By (50), we have

$$\max_{g\in A_n}|\phi_j(g) - \tilde{\phi}_j(g)| \le \sum_{i=1\vee j}^{n}|a_{i-j}|(|\epsilon_j| + \mu)(|X_{i,j-1}| + \mathbb{E}|X_{i,j}|).$$

Since ϵ_j are independent of $X_{i,j-1}$, above together with (51) leads to

$$\|\max_{g\in A_n}|\phi_j(g) - \tilde{\phi}_j(g)|\|_q \le c_1'\sum_{i=1\vee j}^{n}(i-j)^{-2\beta+1/2}\mu_q^2 \le \begin{cases} c_1' n(-j)^{-2\beta+1/2}, & \text{if } j < -n,\\ 2c_1' h(n,\beta)\mu_q^2, & \text{if } -n \le j \le n,\end{cases}$$

where $c_1' = 4(2\beta-1)^{-1/2}\gamma^2 K_q$, $h(n,\beta) = \log(n)$ if $\beta = 3/4$; $h(n,\beta) = (4\beta-1)/(4\beta-3)$ if $\beta > 3/4$; $h(n,\beta) = 2(3-4\beta)^{-1}n^{3/2-2\beta}$ if $\beta < 3/4$. Therefore by Markov's inequality

$$\begin{aligned}\mathrm{I}_1 &\le u^{-q}\Big(\sum_{-n\le j\le n}\|\max_{g\in A_n}|\phi_j(g) - \tilde{\phi}_j(g)|\|_q^q + \sum_{j<-n}\|\max_{g\in A_n}|\phi_j(g) - \tilde{\phi}_j(g)|\|_q^q\Big)\\ &\lesssim u^{-q}n^{1+(1-\beta)q}\tilde{c}^{-q}(n,\beta)\mu_q^{2q},\end{aligned}$$

where the constant in $\lesssim$ only depends on β, q, γ.
Part (ii): By (50) we obtain

$$\max_{g\in A_n}|\phi_j(g) - \tilde{\phi}_j(g)| \le \sum_{i=1\vee j}^{n}2|a_{i-j}|(|\epsilon_j| + \mu) \le \begin{cases} 2\gamma n(-j)^{-\beta}(|\epsilon_j| + \mu), & \text{if } j < -n,\\ 2\gamma n^{1-\beta}(|\epsilon_j| + \mu), & \text{if } -n \le j \le n.\end{cases}$$

Since ϵ_j is independent of $\mathcal{F}_{j-1}$, $\mathbb{E}(|\epsilon_j|^2|\mathcal{F}_{j-1}) = \mu_2^2$. Hence

$$\begin{aligned}\mathrm{I}_2 \le& \sum_{-n\le j\le n}\mathbb{E}\big[\max_{g\in A_n}(\phi_j(g) - \tilde{\phi}_j(g))^2|\mathcal{F}_{j-1}\big] + \sum_{j<-n}\mathbb{E}\big[\max_{g\in A_n}(\phi_j(g) - \tilde{\phi}_j(g))^2|\mathcal{F}_{j-1}\big]\\ \le& 16\gamma^2\sum_{-n\le j\le n}n^{2(1-\beta)}\mu_2^2 + 16\gamma^2\sum_{j<-n}n^2(-j)^{-2\beta}\mu_2^2 \le c_2' n^{3-2\beta}\mu_2^2,\end{aligned}$$

where $c_2' = 32\gamma^2\beta(2\beta-1)^{-1}$.

Combining two parts and applying them to Lemma 18, we have

$$\mathbb{P}\Big(\Omega_n \geq \frac{z}{4}\Big) \lesssim \frac{n^{1+(1-\beta)q}\tilde{c}^{-q}(n,\beta)\mu_q^{2q}}{u^q} + 2|A_n|\exp\Big(-\frac{z^2}{C_{\beta,\gamma}n^{3-2\beta}\mu_2^2}\Big) + 2|A_n|\exp\Big(-\frac{z^2}{2zu}\Big),$$

where the constant in $\lesssim$ only depends on β, q, γ. Let $R_n = 2|A_n|\tilde{c}^q(n,\beta)z^q/n^{1+(1-\beta)q}$ and $u = z/(4\log(R_n))$. Notice $\log(R_n) \lesssim \log|A_n| + \log(n)$, where constant in $\lesssim$ only depends on β, q, γ. Then the desired result follows. ■

Remark 29 *With a careful check of the proofs of Theorem 8, Lemmas 27 and 28, we can choose constants in Theorem 8 as follows:*

- $C_{\beta,\gamma} = 64\max\{3\gamma^2((2\beta-1)^{-1} + 8(1-\beta)^{-2}), 16\gamma^2\beta(2\beta-1)^{-1}\}$.
- $C_{\beta,q,\gamma} = \max\{C_{\beta,q,\gamma,1}, C_{\beta,q,\gamma,2}\}$, *where* $C_{\beta,q,\gamma,1} = 4^q K_q' c_2$, *with* $c_2 = \gamma^q 2^q(2^{1+q}(1-\beta)^{-q} + (q\beta-1)^{-1})$ *and* $C_{\beta,q,\gamma,2} = 1 + 8{c_1'}^q(2\beta q - q/2 - 1)^{-1} + 2^{q+3}{c_1'}^q\max^q\{1, (4\beta-1)/(4\beta-3), 2(3-4\beta)^{-1}\}$, *with* $c_1' = 4(2\beta-1)^{-1/2}\gamma^2 K_q$.
- $C'_{\beta,q,\gamma} = 2\gamma c_3 + 4\gamma(1-\beta)^{-1}q(q-2)^{-1}$, *where* $c_3 = 2(2-r)^{-1}(r\beta-1)^{-1} + 2(q-2)^{-1}(q\beta-1)^{-1/2}$ *with* $r = 1 + 1/(2\beta)$.

Here K_q *(resp.* K_q' *) is the constant given in Lemma 19 (resp. Lemma 20).*

Proof [Proof of Proposition 9] Construct A_n as in the proof of Theorem 4. Recall (31) for the function g_k. Note that $g_1(X_{i,i-1}) = \mathbb{E}[g(X_i)|\mathcal{F}_{i-1}]$. By (36), we have

$$\begin{aligned}\mathbb{P}(|\Delta_n| \geq a + z) \leq & \mathbb{P}\Big(\max_{g\in A_n}|S_n(g) - \mathbb{E}S_n(g)| \geq a + z/2\Big) \\ \leq & \mathbb{P}\Big(\max_{g\in A_n}\Big|\sum_{i=1}^n\big(g_1(X_{i,i-1}) - \mathbb{E}g_1(X_{i,i-1})\big)\Big| \geq a + z/4\Big) \\ & + \sum_{g\in A_n}\mathbb{P}\Big(\Big|\sum_{i=1}^n(g(X_i) - \mathbb{E}[g(X_i)|\mathcal{F}_{i-1}])\Big| \geq z/4\Big) =: \mathrm{I}_1 + \mathrm{I}_2,\end{aligned}$$

where $a = C_q a_* \mu_q c(n,q)$ and $a = C'_{\beta,q,\gamma}\mu_q n^{3/2-\beta}$ for Propositions 5 and 9, respectively.

Since $|g| \leq 1$ and $g(X_i) - \mathbb{E}[g(X_i)|\mathcal{F}_{i-1}]$, $1 \leq i \leq n$, are martingale differences, by Azuma's inequality, $\mathrm{I}_2 \leq 2|A_n|\exp\{-z^2/(64n)\}$. For I_1, notice

$$g_1(x) = \int_{-\infty}^{\infty} g(x+y)f_\epsilon(y)\mathrm{d}y = \int_{-\infty}^{\infty} g(y)f_\epsilon(y-x)\mathrm{d}y.$$

By Assumption (A'), $\sup_{g\in\mathcal{A}}|g_1|_\infty$, $\sup_{g\in\mathcal{A}}|g_1'|_\infty$ and $\sup_{g\in\mathcal{A}}|g_1''|_\infty$ are all bounded by 1. Thus in the I_1 part, the function g_1 satisfies Assumption (A) and can be dealt with by Theorem 4 and Theorem 8 for Propositions 5 and 9 respectively. Combining I_1 and I_2, we complete the proof. ■

6.5 Proof of Theorem 11

Proof [Proof of Theorem 11] We shall apply the argument in the proof of Theorem 7. Recall (56) for D_k, I_1 and I_2. Case (a) follows from the following three claims:

(a.i) $\mathrm{I}_1 \le e^{-C_2 z^2/n}$, (a.ii) $\mathrm{I}_2 \le e^{-C_2 z^2/n}$, (a.iii) $\mathbb{E}\Delta_n \le C_1\sqrt{n}$,

while Case (b) follows from the following three:

(b.i) $\mathrm{I}_1 \le e^{-C_4 z^2 n^{-(3-2\beta)}}$, (b.ii) $\mathrm{I}_2 \le e^{-C_4 z^2 n^{-(3-2\beta)}}$, (b.iii) $\mathbb{E}\Delta_n \le C_3 n^{3/2-\beta}$.

Part (a.i) and (b.i): By (58), for $k \le -n$,

$$|D_k| \le \sum_{i=1\vee k}^{n} |a_{i-k}|(|\epsilon_k| + \mu) \le \gamma n(-k)^{-\beta}(|\epsilon_k| + \mu). \tag{80}$$

Let $h^* = c_0/(2\gamma)$. By the same argument in (63) and (64), for $0 < h \le h^*$,

$$\mathbb{E}(e^{D_k h}|\mathcal{F}_{k-1}) \le 1 + \mathbb{E}\Big[\frac{e^{|D_k|h^*} - |D_k|h^* - 1}{h^{*2}}\Big|\mathcal{F}_{k-1}\Big]h^2. \tag{81}$$

Denote $\theta = n(-k)^{-\beta}/2$. Note that for any fixed $x > 0$, $e^{tx} - tx - 1$ is increasing on $t \in (0, \infty)$. Applying the upper bound for $|D_k|$ in (80), we have

$$\begin{aligned}
&\mathbb{E}(e^{|D_k|h^*} - |D_k|h^* - 1|\mathcal{F}_{k-1}) \le \mathbb{E}\big[e^{c_0\theta(|\epsilon_k|+\mu)} - c_0\theta(|\epsilon_k| + \mu) - 1\big] \\
&\quad = \mathbb{E}\Big[\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}x}(e^{\theta x} - \theta x - 1)\cdot \mathbf{1}_{\{c_0(|\epsilon_k|+\mu)\ge x\}}\mathrm{d}x\Big] = \int_0^\infty (\theta e^{\theta x} - \theta)\mathbb{P}(c_0(|\epsilon_k| + \mu) \ge x)\mathrm{d}x,
\end{aligned}$$

where the last equality is by Fubini's theorem. Note that $\mathbb{P}(c_0(|\epsilon_k|+\mu) \ge x) \le c_1 e^{-x}$, where $c_1 = e^{c_0\mu}\mu_e$. Then we further have

$$\mathbb{E}(e^{|D_k|h^*} - |D_k|h^* - 1|\mathcal{F}_{k-1}) \le \int_0^\infty c_1 e^{-x}(\theta e^{\theta x} - \theta)\mathrm{d}x = c_1\theta^2/(1-\theta) \le 2c_1\theta^2. \tag{82}$$

where the last inequality is due to $\theta \le 1/2$. Hence by (81) and (82) we have for any $h \le h^*$,

$$\mathbb{E}(e^{D_k h}|\mathcal{F}_{k-1}) \le 1 + c_2 n^2(-k)^{-2\beta}h^2 \le e^{c_2 n^2(-k)^{-2\beta}h^2}. \tag{83}$$

where $c_2 = 2c_1\gamma^2/c_0^2$ and the last inequality is due to $1 + x \le e^x$. By Markov's inequality

$$\mathrm{I}_1 \le e^{-zh/2}\mathbb{E}\Big(e^{\sum_{k\le -n} D_k h}\Big) \le e^{-zh/2}\mathbb{E}\Big(e^{\sum_{k\le -n-1} D_k h}\mathbb{E}(e^{D_{-n}h}|\mathcal{F}_{-n-1})\Big).$$

Hence recursively applying (83), we obtain

$$\mathrm{I}_1 \le \exp\Big(-zh/2 + c_2 n^2\sum_{k\le -n}(-k)^{-2\beta}h^2\Big) \le \exp\Big(-zh/2 + c_2 c_3(2\beta-1)^{-1}n^{3-2\beta}h^2\Big),$$

where $c_3 = \max\{(2\beta-1)/(4c_2h^*), 1\}$. Take $h = (2\beta-1)(4c_2c_3)^{-1}z/n$ for (a.i) and $h = (2\beta-1)(4c_2c_3)^{-1}z/n^{3-2\beta}$ for (b.i) respectively, then $h \le h^*$ and we have $\mathrm{I}_1 \le e^{-C_{21}z^2/n}$ for (a.i) and $\mathrm{I}_1 \le e^{-C_{21}z^2/n^{3-2\beta}}$ for (b.i), where $C_{21} = (2\beta-1)/(16c_2c_3)$.

Part (a.ii): By (58), for $-n < k \le n$,

$$|D_k| \le \sum_{i=1\vee k}^{n} |a_{i-k}|(|\epsilon_k| + \mu) \le 2\beta(\beta-1)^{-1}\gamma(|\epsilon_k| + \mu). \tag{84}$$

Let $c_4 = 2\beta(\beta-1)^{-1}\gamma$ and $h^* = c_0 c_4^{-1}$. By the same argument as (83) in Part (a.i), we have for any $h \le h^*$,

$$\mathbb{E}(e^{D_k h}|\mathcal{F}_{k-1}) \le e^{c_5 h^2}, \tag{85}$$

where $c_5 = c_1 c_4^2/(2c_0^2)$. Similarly to Part (a.i), by Markov's inquality and recursively applying (85),

$$\mathrm{I}_2 \le e^{-zh/2}\mathbb{E}\Big(e^{\sum_{-n<k\le n} D_k h}\Big) \le \exp\Big(-zh/2 + 2c_5c_6nh^2\Big),$$

where $c_6 = \max\{c_4/(8c_0c_5), 1\}$. Let $h = (8c_5c_6)^{-1}z/n$. Then $h \le h^*$ and we have $\mathrm{I}_2 \le e^{-C_{22}z^2/n}$, where $C_{22} = (32c_5c_6)^{-1}$.
Part (b.ii): By (58), for $-n < k \le n$,

$$|D_k| \le (1-\beta)^{-1}\gamma n^{1-\beta}(|\epsilon_k| + \mu). \tag{86}$$

Let $c_7 = (1-\beta)^{-1}\gamma$, $h^* = c_0 c_7^{-1} n^{-1+\beta}$ and $c_8 = 2\max\{c_7/(8c_0), c_1c_7^2/(2c_0^2)\}$. By the same argument as in Part (a.ii) with the bound in (84) replaced by (86), we have for any $h \le h^*$,

$$\mathrm{I}_2 \le e^{-zh/2}\mathbb{E}\Big(e^{\sum_{-n<k\le n} D_k h}\Big) \le \exp\Big(-zh/2 + c_8 n^{3-2\beta}h^2\Big).$$

Take $h = (4c_8)^{-1}zn^{-(3-2\beta)}$, then $h \le h^*$ and we have $\mathrm{I}_2 \le e^{-C_{42}z^2n^{-(3-2\beta)}}$.

Part (a.iii) and (b.iii): Applying Theorem 1 in Wu (2003) with $p = 0$, $k = 1$ and $\gamma = 2$ therein, we have $\mathbb{E}(\Delta_n) = C_1 n^{1/2}$ (resp. $\mathbb{E}(\Delta_n) = C_3 n^{3/2-\beta}$) for SRD (resp. LRD) processes, where the constants C_1 and C_3 only depend on $\beta, \gamma, f_*, \mu_e, c_0$. ■

Remark 30 *Based on the proof of Theorem 11, the constants can take the following values:* $C_2 = \max(C_{21}, C_{22})$, $C_4 = \max(C_{21}, C_{42})$, *where* $C_{21} = (2\beta-1)/(16c_2c_3)$, $C_{22} = (32c_5c_6)^{-1}$ *and* $C_{42} = (16c_8)^{-1}$, *with* $c_1 = e^{c_0\mu}\mu_e$, $c_2 = 2c_1\gamma^2/c_0^2$, $c_3 = \max\{(2\beta-1)\gamma/(2c_0c_2), 1\}$, $c_4 = 2\beta(\beta-1)^{-1}\gamma$, $c_5 = c_1c_4^2/(2c_0^2)$, $c_6 = \max\{c_4/(8c_0c_5), 1\}$, $c_7 = (1-\beta)^{-1}\gamma$ *and* $c_8 = 2\max\{c_7/(8c_0), c_1c_7^2/(2c_0^2)\}$. *Constants* C_1 *and* C_3 *only depend on* $\beta, \gamma, f_*, \mu_e, c_0$.

6.6 Proofs of Theorem 12 and Proposition 13

Proof [Proof of Theorem 12] Since F_ϵ is the c.d.f of ϵ_i and $a_0 = 1$, $\mathbb{E}(\mathbf{1}_{X_i\le t}|\mathcal{F}_{i-1}) = F_\epsilon(t - X_{i,i-1})$. The summation $S_n(t)$ can be decomposed into two parts:

$$S_n(t) = \sum_{i=1}^{n}[\mathbf{1}_{X_i\le t} - \mathbb{E}(\mathbf{1}_{X_i\le t}|\mathcal{F}_{i-1})] + \sum_{i=1}^{n}[F_\epsilon(t - X_{i,i-1}) - F(t)] =: Q_n(t) + R_n(t). \tag{87}$$

Note that summands of $Q_n(t)$ are martingale differences. We shall derive bounds for (i). $\mathbb{P}(\sup_{t\in\mathbb{R}} |Q_n(t)|/f_* \geq z/2)$.

(ii). $\mathbb{P}\big(\sup_{t\in\mathbb{R}} |R_n(t)|/f_* \geq C_0 a_* \mu_q c(n,q) + z/2\big)$, for SRD case;
$\mathbb{P}(\sup_{t\in\mathbb{R}} |R_n(t)|/f_* \geq C_0' \mu_q n^{3/2-\beta} + z/2)$, for LRD case.

We shall apply Azuma's inequality on $Q_n(t)$ since it is the sum of martingale differences. For $R_n(t)$, since F_ϵ is smooth, we apply Theorems 4 and 8 for SRD and LRD cases, respectively. Part (i): Let $M = 2\mu_q n^{2\beta}$, $H(t) = \sum_{i=1}^n \mathbf{1}_{X_i \leq t}$ and $\tilde{H}(t) = \sum_{i=1}^n F_\epsilon(t - X_{i,i-1})$. Then $Q_n(t) = H(t) - \tilde{H}(t)$ and

$$\mathbb{P}\Big(\sup_{t\in\mathbb{R}} |Q_n(t)|/f_* \geq z/2\Big) \leq \mathrm{I}_1 + \mathrm{I}_2, \text{ where}$$

$$\mathrm{I}_1 = \mathbb{P}\Big(\sup_{t\in\mathbb{R}} |H(t) - \tilde{H}(t)|/f_* \geq z/2,\ \max_{i\leq n} |X_{i,i-1}| \leq M\Big),\ \mathrm{I}_2 = \sum_{i=1}^n \mathbb{P}\Big(|X_{i,i-1}| \geq M\Big).$$

For I_1, let $t_k = -2M + \delta k,\ k = 0, \ldots, \lceil 4M/\delta \rceil$, where $\delta = z/(4n)$. Since $|F_\epsilon'| \leq f_*$, under this construction, $|\tilde{H}(t_k) - \tilde{H}(t_{k+1})|/f_* \leq z/4$.

Moreover, since $n\mathbb{P}(|\epsilon_0| \geq M) \leq n^{1-2q\beta} \leq z/4$, $\tilde{H}(t_0)$ and $1 - \tilde{H}(t_{\lceil 4M/\delta \rceil})$ are less than $z/4$ on the set $\{\max_{i\leq n} |X_{i,i-1}| \leq M\}$.

Since $H(t)$ and $\tilde{H}(t)$ are both non-decreasing, for $s_1 \leq s_2$ and $t \in [s_1, s_2]$, we have

$$|H(t) - \tilde{H}(t)| \leq |\tilde{H}(s_1) - \tilde{H}(s_2)| + \max\big\{|H(s_1) - \tilde{H}(s_1)|, |H(s_2) - \tilde{H}(s_2)|\big\}.$$

Consequently,

$$\mathrm{I}_1 \leq \sum_{k=0}^{\lceil 4M/\delta \rceil} \mathbb{P}\left(|H(t_k) - \tilde{H}(t_k)|/f_* > z/4\right). \tag{88}$$

For any $t \in \mathbb{R}$, since the martingale differences $\mathbf{1}_{X_i \leq t} - \mathbb{E}(\mathbf{1}_{X_i \leq t} | \mathcal{F}_{i-1}), i = 1, \ldots, n$, are bounded in absolute value by 1, by Azuma's inequality,

$$\mathbb{P}\left(|H(t) - \tilde{H}(t)| > z\right) \leq 2\exp(-z^2/2n). \tag{89}$$

With (88) and (89), we obtain

$$\mathrm{I}_1 \leq (64\mu_q) z^{-1} n^{2\beta+1} \exp(-z^2/(32n)).$$

For I_2, by (51) and Markov's inequality,

$$\mathrm{I}_2 \leq M^{-q} \sum_{i=1}^n \|X_{i,i-1}\|_q^q \leq c_1 n^{1-2q\beta}, \tag{90}$$

where $c_1 = (K_q \gamma/2)^q (\beta q' - 1)^{-q/q'}$. Combining I_1 and I_2 we complete the proof for this part.

Part (ii): Let $M = c_2\mu_q n^{2\beta}$, where $c_2 = 2K_q\gamma q'\beta$, then for any $\tau > 0$,

$$\begin{aligned}
&\mathbb{P}(\sup_{t\in\mathbb{R}} |R_n(t)|/f_* \geq z/2+\tau) \\
\leq &\mathbb{P}\Big(\sup_{|t|\leq 2M} |R_n(t)|/f_* \geq z/4+\tau\Big) + \sum_{i=1}^{n} \mathbb{P}\big(|X_{i,i-1}| \geq M\big) \\
+&\mathbb{P}\Big(\sup_{|t|\geq 2M} |R_n(t)|/f_* \geq z/4,\ \max_{i\leq n} |X_{i,i-1}| \leq M\Big) =: \mathrm{I}_1' + \mathrm{I}_2' + \mathrm{I}_3'. \qquad (91)
\end{aligned}$$

For I_1', let

$$A_n = \{-2M + \delta k | \delta = z/(8n), k = 0, 1, \ldots, \lceil 4M/\delta \rceil\}. \qquad (92)$$

Then $\sup_{|t|\leq 2M} \min_{s\in A_n}(|F_\epsilon(t-\cdot) - F_\epsilon(s-\cdot)|_\infty + |F(t)-F(s)|)/f_* \leq z/(4n)$, and the cardinal number $|A_n| \leq (32c_2\mu_q)n^{2\beta+1}/z$. Hence under short- (resp. long-) range dependence, take $\tau = C_q a_* \mu_q c(n,q)$ (resp. $\tau = C'_{\beta,q\gamma}\mu_q n^{3/2-\beta}$), then I_1' can be bounded by Theorem 4 (resp. Theorem 8), that is, for SRD case,

$$\mathrm{I}_1' \leq 2^{2q\beta} C_{\beta,q,\gamma}\mu_q^q \frac{n}{z^{q\beta}} + 3\exp\Big(-\frac{z^2}{16C_{\beta,\gamma}\mu_{q'}^{q'} n} + \log(|A_n|)\Big) + 2\exp\Big(-\frac{z^v}{2^{3+2v}\mu_q^{v'}} + \log(|A_n|)\Big),$$

and for LRD case,

$$\mathrm{I}_1' \leq 2^{2q} C_{\beta,q,\gamma}(\mu_q^{2q} \vee \mu_q^q) \frac{n^{1+(1-\beta)q}}{z^q}\Big(1 + \frac{[\log(|A_n|) + \log(n)]^q}{\tilde{c}^q(n,\beta)}\Big) + 3\exp\Big(-\frac{z^2}{16C_{\beta,\gamma} n^{3-2\beta}\mu_2^2} + \log(|A_n|)\Big),$$

where $C_{\beta,q,\gamma}$ and $C_{\beta,\gamma}$ take the same values as in Theorems 4 and 8, respectively.

For I_2', by (90) we have $\mathrm{I}_2' \leq n^{1-2q\beta}$.

For I_3', if $|X_{i,i-1}| \leq M$ and $t \leq -2M$, then $F_\epsilon(t - X_{i,i-1}) \leq F_\epsilon(-M) \leq M^{-q}\mu_q^q$ and $F(t) \leq (2M)^{-q}\mathbb{E}|X_0|^q$. By a similar argument for $t \geq 2M$, we obtain $R_n(x) < z/4$ for $|X_{i,i-1}| \leq M$ and $|t| \geq 2M$, that is $\mathrm{I}_3' = 0$. ■

Remark 31 *Recall Lemma 19 for K_q. We can choose constants in Theorem 12 as follows:*

SRD $C_0 = C_q$, $C_1 = (K_q\gamma/2)^q(\beta q' - 1)^{-q/q'} + 1 + 2^{q\beta}C_{\beta,q,\gamma}$, $C_2 = (16C_{\beta,\gamma} \vee 32)^{-1}$, $C_3 = 64K_q\gamma q'\beta(2\beta+1)$, *where* $C_{\beta,q,\gamma}$, $C_{\beta,\gamma}$ *and* C_q *take same values as those in Theorem 4.*

LRD $C_0' = C'_{\beta,q,\gamma}$, $C_1' = (K_q\gamma/2)^q(\beta q' - 1)^{-q/q'} + 1 + 2^{2q}C_{\beta,q,\gamma}c_0$, $C_2' = (16C_{\beta,\gamma} \vee 32)^{-1}$, $C_3' = 64K_q\gamma q'\beta(2\beta+1)$, *where* $C'_{\beta,q,\gamma}$, $C_{\beta,q,\gamma}$, $C_{\beta,\gamma}$ *take same values as those in Theorem 8,* $c_0 = 1 + \max_{n\geq 1}\log^q(c_0' n^{2\beta+1})\tilde{c}^{-q}(n,\beta)$, *with* $c_0' = 64K_q\gamma q'\beta$. *Since* $\tilde{c}(n,\beta) = n^\alpha$ *some* $\alpha > 0$ *and* $\log(n)/n^\alpha \to 0$, c_0' *is a finite constant.*

The following lemma is a variant of the Fuk-Nagaev inequality which will be used in the proof of Proposition 13.

Lemma 32 *Let $X_i = (X_{i1}, \ldots, X_{ip})^\top$, $i \in \mathbb{Z}$, be independent mean 0 random vectors in $\mathbb{R}^p$ and $S_{nj} = \sum_{i \le n} X_{ij}$. Assume there exist constants $s, r, c > 0$ such that*

$$\sum_{i \le n} \mathbb{P}\big(\max_{1\le j\le p} |X_{ij}| \ge y\big) \le cn/(y^s \log^r(y)), \qquad \text{for all } y > e.$$

Let $\sigma_n^2 = \max_{1\le j\le p} \sum_{i\le n} \mathbb{E}(X_{ij}^2)$. Then for any $z \ge c' n^{1/2}$, where $c' > 0$,

$$\mathbb{P}\Big(\max_{1\le j\le p} |S_{nj}| \ge 2\mathbb{E}\big[\max_{1\le j\le p} |S_{nj}|\big] + z\Big) \le C_1 e^{-z^2/(3\sigma_n^2)} + C_2 n/(z^s \log^r(z)),$$

where C_1, C_2 are positive constants that only depend on c, c', s and r.

Proof We shall apply the argument in Theorem 3.1 of Einmahl and Li (2008) with $(B, \|\cdot\|) = (\mathbb{R}^p, |\cdot|_\infty)$, $\eta = \delta = 1$ and $\beta(y) = \beta_{sr}(y) = M/(y^s \log^r(y))$. Notice Λ_n^2 in Theorem 3.1 of Einmahl and Li (2008) is bounded by σ^2 in our settings (cf. proof of Lemma A.2 in Chernozhukov et al. (2017)). ■

Proof [Proof of Proposition 13] Recall the proof of Theorem 12 for I_1, I_2, I_1'-I_3' and A_n in (92). For $z \ge c\sqrt{n}\log^\alpha(n)$, where $\alpha > 1/2$, all terms except I_1' are of order $o\big(nz^{-q\beta}\log^{-r_0}(z)\big)$. Hence we only need to show that $\mathrm{I}_1' \lesssim z^{-q\beta}\log^{-r_0}(z) n\mu_q^q$ for $\tau = C_q a_* \mu_q \sqrt{n}$. Let

$$\phi_j(t) = \sum_{i=1}^n P_j F_\epsilon(t - X_{i,i-1}) = \sum_{i=1\vee(j+1)}^n \big(F_{i-j}(t - X_{i,j}) - F_{i-j+1}(t - X_{i,j-1})\big),$$

Then $R_n(t) = \sum_{j \le n-1} \phi_j(t)$. Define

$$\tilde{\phi}_j(t) = \sum_{i=1\vee(j+1)}^n \big(F(t - a_{i-j}\epsilon_j) - \mathbb{E}F(t - a_{i-j}\epsilon_j)\big) \text{ and } \tilde{R}_n(t) = \sum_{j\le n-1} \tilde{\phi}_j(t).$$

By the same argument for I_1' in the proof of Theorem 12, we have

$$\mathrm{I}_1' \le \mathbb{P}\Big(\max_{t\in A_n} |R_n(t)|/f_* \ge z/4 + C_q a_* \mu_q \sqrt{n}\Big).$$

The idea is similar to the proof of Theorem 4, that is, we shall show:

(i). $R_n(t)$ can be approximated by $\tilde{R}_n(t)$, specifically,

$$\mathbb{P}\Big(\max_{t\in A_n} |R_n(t) - \tilde{R}_n(t)|/f_* \ge z/8\Big) = o\big(nz^{-q\beta}\log^{-r_0}(z)\big). \tag{93}$$

(ii). The tail probability of $\tilde{R}_n(t)$,

$$\mathbb{P}\big(\max_{t\in A_n} |\tilde{R}_n(t)|/f_* \ge C_q a_* \mu_q \sqrt{n} + z/8\big) \lesssim \frac{\mu_q^q n}{z^{q\beta}\log^{r_0}(z)}.$$

Part (i): Note that $\log(|A_n|) \lesssim \log(n)$ (actually $\log(|A_n|) \asymp \log(n)$) and that F_ϵ/f_*, f_ϵ/f_*, f'_ϵ/f_* are all bounded in absolute value by 1. By the same argument as in the proof of Lemma 24, using $u = z/\log^{3/2}(z)$ in inequality (55), we obtain (93).

Part (ii): Denote $V = \max_{t\in A_n} \sum_{j\le n-1} \mathbb{E}[\tilde{\phi}_j^2(t)]$, then by (39), we have $V \lesssim n$, where the constant in $\lesssim$ only depends on β, q, γ. For $j \le -n$, by (40) and $f' \le f_*$, we have

$$\sum_{j\le -n} \mathbb{P}(\max_{t\in A_n} |\tilde{\phi}_j(t)|/f_* \ge z) \le \sum_{j\le -n} \mathbb{P}(\gamma n(-j)^{-\beta}(|\epsilon_j| + \mu) \ge z) \lesssim \mu_q^q n/(z^{q\beta}\log^{r_0}(z)),$$

where the constant in $\lesssim$ only depends on β, q, γ, r_0, L. For $-n < j \le n$, by (41),

$$\sum_{-n<j\le n} \mathbb{P}(\max_{t\in A_n} |\tilde{\phi}_j(t)|/f_* \ge z) \le \sum_{-n<j\le n} \mathbb{P}(c_\beta(|\epsilon_j| + \mu)^{1/\beta} \ge z) \lesssim \mu_q^q n/(z^{q\beta}\log^{r_0}(z)),$$

where the constant in $\lesssim$ only depends on β, q, γ, r_0, L. By Lemma 32 we have

$$\mathbb{P}\Big(\max_{t\in A_n} \big|\tilde{R}_n(t)\big|/f_* - 2\mathbb{E}\Big[\max_{t\in A_n} |\tilde{R}_n(t)|\Big]/f_* \ge z\Big) \lesssim e^{-z^2/(3V)} + \frac{\mu_q^q n}{z^{q\beta}\log^{r_0}(z)} \lesssim \frac{\mu_q^q n}{z^{q\beta}\log^{r_0}(z)}.$$

By (49), we have $\mathbb{E}[\max_{t\in A_n} |\tilde{R}_n(t)|] \le C_q a_* \mu_q \sqrt{n}$, which implies the desired result. ■

6.7 Proof of Theorem 14

Proof [Proof of Theorem 14] Define

$$W_n(t) = \sum_{j=0}^{n-1} \varphi_j(t), \text{ where } \varphi_j(t) = \sum_{k\ge 1} \big[F(t - k^{-\beta}\epsilon_j) - \mathbb{E}F(t - k^{-\beta}\epsilon_j)\big]. \tag{94}$$

We claim that:

(i). $S_n(t)$ can be approximated by $W_n(t)$, specifically for $\theta_0 = (2\alpha - 1)/4$,

$$\mathbb{P}\Big(|S_n(t) - W_n(t)| \ge z/\log^{\theta_0}(z)\Big) = o\big(nz^{-q\beta}/\log^{r_0}(z)\big). \tag{95}$$

(ii). The tail distribution of $\varphi_0(t)$ satisfies

$$\mathbb{P}\left(\varphi_0(t) > z\right) \sim \frac{C_1}{z^{q\beta}\log^{r_0}(z)} \text{ and } \mathbb{P}\left(\varphi_0(t) < -z\right) \sim \frac{C_2}{z^{q\beta}\log^{r_0}(z)}, \tag{96}$$

where C_1, C_2 only depend on q, β, r_0, t, F.

Proofs of (95) and (96) will be given in Lemmas 33 and 34, respectively. By (95),

$$\begin{aligned}\mathbb{P}(S_n(t) > z) \ge& \mathbb{P}(W_n(t) \ge z + z/\log^{\theta_0}(z)) - \mathbb{P}(|S_n(t) - W_n(t)| \ge z/\log^{\theta_0}(z)),\\ =& \mathbb{P}(W_n(t) \ge z + z/\log^{\theta_0}(z)) + o\big(nz^{-q\beta}/\log^{r_0}(z)\big),\end{aligned} \tag{97}$$

and similarly

$$\mathbb{P}(S_n(t) > z) \leq \mathbb{P}(W_n(t) \geq z - z/\log^{\theta_0}(z)) + o\big(nz^{-q\beta}/\log^{r_0}(z)\big). \tag{98}$$

Since φ_j has a regularly varying tail (96), by Theorem 1.9 in Nagaev (1979),

$$\sup_{w \geq \sqrt{n}\log^{\alpha}(n)} \Big|\frac{\mathbb{P}(W_n(t) \geq w)}{n\mathbb{P}(\varphi_0(t) \geq w)} - 1\Big| \to 0, \text{ as } n \to \infty.$$

Hence we have $\mathbb{P}(S_n(t) \geq z) \sim C_1 nz^{-q\beta}\log^{-r_0}(z)$ by (96), (97) and (98) in view of

$$\mathbb{P}(W_n(t) \geq z + z/\log^{\theta_0}(z)) \sim C_1 \frac{n}{z^{q\beta}\log^{r_0}(z)} \sim \mathbb{P}(W_n(t) \geq z - z/\log^{\theta_0}(z)).$$

Similarly we can derive $\mathbb{P}(S_n(t) \leq -z) \sim C_2 nz^{-q\beta}\log^{-r_0}(z)$. ■

Lemma 33 *Recall definitions of $S_n(t)$, $W_n(t)$ in* (94). *Under assumptions of Theorem 14, we have for $\theta_0 = (2\alpha - 1)/4$,* (95) *holds.*

Proof Recall (87) for $Q_n(t)$ and $R_n(t)$. Let

$$\tilde{W}_n(t) = \sum_{j \leq n-1} \sum_{i=1\vee(j+1)}^{n} [F(t - (i-j)^{-\beta}\epsilon_j) - \mathbb{E}F(t - (i-j)^{-\beta}\epsilon_j)].$$

Then

$$\begin{aligned}
&\mathbb{P}\big(|S_n(t) - W_n(t)| \geq z\log^{-\theta_0}(z)\big) \\
\leq &\mathbb{P}\big(|Q_n(t)| \geq z\log^{-\theta_0}(z)/3\big) + \mathbb{P}\big(|R_n(t) - \tilde{W}_n(t)| \geq z\log^{-\theta_0}(z)/3\big) \\
+ &\mathbb{P}\big(|W_n(t) - \tilde{W}_n(t)| \geq z\log^{-\theta_0}(z)/3\big) =: \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3.
\end{aligned}$$

Part I_1: Since $Q_n(t)$ is the summation of martingale differences bounded in absolute value by 1, Azuma's inequality leads to

$$\mathbb{P}\big(|Q_n(t)| \geq z/\log^{\theta_0}(z)\big) \leq 2\exp\Big\{-\frac{z^2}{2n\log^{2\theta_0}(z)}\Big\} = o\big(nz^{-q\beta}/\log^{r_0}(z)\big). \tag{99}$$

Part I_2: Note that

$$R_n(t) = \sum_{j \leq n-1} \sum_{i=1\vee(j+1)}^{n} \big(F_{i-j}(t - X_{i,j}) - F_{i-j+1}(t - X_{i,j-1})\big).$$

Take $F_\epsilon(t - \cdot)$ as $g(\cdot)$ in Lemma 24, then $g_\infty(\cdot) = F(t - \cdot)$. By Lemma 24, but in inequality (55), take $u = z/\log^{\theta_0+2}(z)$ instead, we obtain

$$\mathbb{P}\big(|R_n(t) - \tilde{W}_n(t)| \geq z/\log^{\theta_0}(z)\big) = o\big(nz^{-q\beta}\log^{-r_0}(z)\big).$$

Part I_3: Since $\tilde{W}_n(t)$ can be rewritten as

$$\tilde{W}_n(t) = \sum_{j \le n-1} \sum_{k=1\vee(1-j)}^{n-j} [F(t - k^{-\beta}\epsilon_j) - \mathbb{E}F(t - k^{-\beta}\epsilon_j)].$$

Notice that

$$\begin{aligned} &\tilde{W}_n(t) - W_n(t) \\ =& \sum_{j \le -1} \sum_{k=1-j}^{n-j} [F(t - k^{-\beta}\epsilon_j) - \mathbb{E}F(t - k^{-\beta}\epsilon_j)] - \sum_{j=0}^{n-1} \sum_{k \ge n-j+1} [F(t - k^{-\beta}\epsilon_j) - \mathbb{E}F(t - k^{-\beta}\epsilon_j)]. \end{aligned}$$

For $j < 0$, let $\phi_j = \sum_{k=1-j}^{n-j}[F(t - k^{-\beta}\epsilon_j) - \mathbb{E}F(t - k^{-\beta}\epsilon_j)]$, then by Lemma 22,

$$\begin{aligned} |\phi_j| &\le \sum_{k=1-j}^{n-j} \min\left\{f_* k^{-\beta}(|\epsilon_j| + \mu), 1\right\} \\ &\le f_* \min\Big\{2\beta(\beta-1)^{-1}(|\epsilon_j| + \mu)^{1/\beta}, n(-j)^{-\beta}(|\epsilon_j| + \mu)\Big\}. \end{aligned} \tag{100}$$

Denote $V = \sum_{j \le -1} \mathbb{E}(|\phi_j|^2)$. By Corollary 1.8 in Nagaev (1979), for $x = \lfloor n/\log^{\Gamma_0}(n) \rfloor$ with $\Gamma_0 = r_0 + \theta_0 q\beta + 1/2$,

$$\begin{aligned} \mathbb{P}\Big(\big|\sum_{j \le -1} \phi_j\big| \ge z/\log^{\theta_0}(z)\Big) \;\lesssim\; & \sum_{j=-x}^{-1} \frac{\log^{\theta_0 q}(z)}{z^q} \mathbb{E}(|\phi_j|^q) + \sum_{j<-x} \frac{\log^{\theta_0 q\beta}(z)}{z^{q\beta}} \mathbb{E}(|\phi_j|^{q\beta}) \\ & + \exp\Big(-\frac{z^2}{\log^{2\theta_0}(z)V}\Big) = \mathrm{II}_1 + \mathrm{II}_2 + \mathrm{II}_3, \end{aligned}$$

where the constant in $\lesssim$ only depends on q and β.

For II_1, by (100),

$$\mathrm{II}_1 \lesssim \frac{\log^{\theta_0 q\beta}(z)}{z^{q\beta}} x\mu_q^q = \frac{n}{z^{q\beta}\log^{r_0}(z)} \frac{x}{n} [\log(z)]^{\theta_0 q\beta + r_0} \mu_q^q = o\big(nz^{-q\beta}/\log^{r_0}(z)\big),$$

where the constant in $\lesssim$ only depends on μ_q, f_*, q and β.

For II_2, by (100),

$$\begin{aligned} \mathrm{II}_2 &\lesssim \frac{\log^{\theta_0 q}(z)}{z^q} \sum_{j<-x} n^q(-j)^{q\beta}\mu_q^q \\ &\lesssim \frac{n}{z^{q\beta}\log^{r_0}(z)} \frac{[\log(z)]^{\theta_0 q + r_0} z^{q(\beta-1)} n^{q-1}}{x^{q\beta-1}} \mu_q^q = o\big(nz^{-q\beta}/\log^{r_0}(z)\big), \end{aligned}$$

where the constants in $\lesssim$ only depend on μ_q, f_*, q and β.

For II_3, by (100),

$$V \lesssim \sum_{j<-n} (n(-j)^{-\beta})^{q'} n^{2-q'} \mu_q^{q'} + \sum_{-n \le j \le n} \mu_q^q \lesssim n\mu_q^q,$$

where the constants in $\lesssim$ only depends on f_*, q and β.

Combining II_1-II_3, we have $\mathbb{P}(|\sum_{j\le -1}\phi_j| \ge z/\log^{\theta_0}(z)) = o(nz^{-q\beta}/\log^{r_0}(z))$. A similar argument will lead to the same bound for $j\ge 0$ part, thus

$$\mathbb{P}\big(|\tilde{W}_n(t) - W_n(t)| \ge z/\log^{\theta_0}(z)\big) = o\big(nz^{-q\beta}/\log^{r_0}(z)\big).$$

Thus the lemma follows from I_1-I_3. ■

Lemma 34 *Recall* (94) *for* $\varphi_0(t)$*. Under conditions of Theorem 14, we have*

$$\mathbb{P}\left(\varphi_0(t) > z\right) \sim \frac{C_1}{z^{q\beta}\log^{r_0}(z)} \quad \textit{and} \quad \mathbb{P}\left(\varphi_0(t) < -z\right) \sim \frac{C_2}{z^{q\beta}\log^{r_0}(z)}, \;\; \textit{as } z\to\infty,$$

where $C_1 = L_1^{q\beta}(t)\beta^{-r_0}$, $C_2 = L_2^{q\beta}(t)\beta^{-r_0}$, *and*

$$L_1(t) = \int_0^\infty \frac{F(t+u)-F(t)}{\beta u^{1+1/\beta}}\mathrm{d}u, \quad L_2(t) = \int_0^\infty \frac{F(t)-F(t-u)}{\beta u^{1+1/\beta}}\mathrm{d}u. \tag{101}$$

Proof Since $f_\epsilon \le 1$, by Lemma 21, f is bounded by 1. Let

$$\tilde{\varphi}_0(t) = \int_0^\infty \big[F(t-s^{-\beta}\epsilon_0) - F(t)\big]\mathrm{d}s. \tag{102}$$

Since $|F(t-s^{-\beta}\epsilon_0) - F(t)| \le \min\{1, s^{-\beta}|\epsilon_0|\}$, we have

$$|\tilde{\varphi}_0(t)| \le 1 + \int_1^\infty s^{-\beta}|\epsilon_0|\mathrm{d}s \le 1 + (\beta-1)^{-1}|\epsilon_0|.$$

Thus $\tilde{\varphi}_0(t)$ is well defined. Note that the lemma follows from the following two claims:

(i). $|\varphi_0(t) - \tilde{\varphi}_0(t)| \le f_*\mu\beta/(\beta-1) + 1$, which is bounded.

(ii). $\mathbb{P}(\tilde{\varphi}_0(t) > z) \sim C_1\log^{-r_0}(z)z^{-q\beta}$ and $\mathbb{P}(\tilde{\varphi}_0(t) < -z) \sim C_2\log^{-r_0}(z)z^{-q\beta}$ as $z\to\infty$.

Part (i): Since F is non-decreasing, for any $s\in[k-1,k]$,

$$|F(t-s^{-\beta}\epsilon_0) - F(t-k^{-\beta}\epsilon_0)| \le \mathrm{sign}(\epsilon_0)\Big\{F(t-k^{-\beta}\epsilon_0) - F(t-(k-1)^{-\beta}\epsilon_0)\Big\}.$$

Since $F(-\infty) = 0$ and $F(\infty) = 1$,

$$\mathrm{I}_1 := \sum_{k=1}^\infty \int_{k-1}^k |F(t-s^{-\beta}\epsilon_0) - F(t-k^{-\beta}\epsilon_0)|\mathrm{d}s \le 1.$$

Since f is bounded, we have

$$\mathrm{I}_2 := \sum_{k=1}^\infty |F(t) - \mathbb{E}F(t-k^{-\beta}\epsilon_0)| \le f_*\sum_{k=1}^\infty k^{-\beta}\mu \le f_*\mu\beta/(\beta-1).$$

Thus $|\varphi_0(t) - \tilde{\varphi}_0(t)| \le \mathrm{I}_1 + \mathrm{I}_2 \le f_*\mu\beta/(\beta-1) + 1$, a finite constant.

Part (ii): Let $u > 0$. Then $0 \le F(t+u) - F(t) \le \min\{f_* u, 1\}$. Hence $L_1(t)$ is bounded by $\int_0^\infty \min\{f_* u, 1\}/(\beta u^{1+1/\beta})\mathrm{d}u \le f_*\beta/(\beta-1)$. Similarly $L_2(t) \le f_*\beta/(\beta-1)$. Note that

$$\int_0^\infty \big[F(t - s^{-\beta}y) - F(t)\big]\mathrm{d}s = \begin{cases} L_1(t)|y|^{1/\beta}, & \text{if } y < 0, \\ -L_2(t)|y|^{1/\beta}, & \text{if } y > 0. \end{cases}$$

Since ϵ_0 is symmetric, by (22) and the definition of $\tilde{\varphi}_0(t)$ in (102), (ii) follows. ∎

Remark 35 *Values of C_1 and C_2 are given in Lemma 34. A careful check of the proof of Theorem 14 suggests that the constant Γ can be chosen as $\Gamma = [\theta_0 q + r_0 + (q\beta - 1)\Gamma']/(q\beta - q)$, where $\theta_0 = (2\alpha - 1)/4$ and $\Gamma' = r_0 + \theta_0 q\beta + 1$.*

6.8 Proof of Corollaries 15 and 16

Proof [Proof of Corollary 15] We shall first deal with the SRD case. Recall $\mathcal{F}_j = (\epsilon_j, \epsilon_{j-1}, \ldots)$. Write

$$M_n(x) = \sum_{j=1}^n \Big(K_b(x - X_j) - \mathbb{E}[K_b(x - X_j)|\mathcal{F}_{j-1}]\Big),$$
$$R_n(x) = \sum_{j=1}^n \Big(\mathbb{E}[K_b(x - X_j)|\mathcal{F}_{j-1}] - \mathbb{E}[K_b(x - X_j)]\Big) = n(\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)) - M_n(x).$$

Note that $M_n(x)$ is a martingale w.r.t. filter $\sigma(\mathcal{F}_n)$. Let $\tau_n = n^\beta$ and $l_* = K_* \vee f_*$. Then

$$\begin{aligned} \mathrm{I} := {} & \mathbb{P}\Big(\sup_{x\in\mathbb{R}} n|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \ge l_* z\Big) \\ \le {} & \sum_{j=1}^n \mathbb{P}(|X_j| \ge \tau_n) + \mathbb{P}\Big(\sup_{x\in\mathbb{R}} n|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \ge l_* z, \max_{1\le j\le n}|X_j| < \tau_n\Big) \\ =: {} & \mathrm{I}_1 + \mathrm{I}_1'. \end{aligned} \tag{103}$$

Since K has support $[-1,1]$, $K_b(x - X_j) = 0$ when $|X_j| < \tau_n$ and $|x| > \tau_n + b_n$. Hence if $\max_{j\le n}|X_j| < \tau_n$ and $|x| > \tau_n + b_n$, we have $\hat{f}_n(x) = 0$. Note that $\sup_{|x|\le\tau_n+b_n}|M_n(x)| + \sup_{|x|\le\tau_n+b_n}|R_n(x)| \ge \sup_{|x|\le\tau_n+b_n} n|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)|$, we have

$$\begin{aligned} \mathrm{I}_1' \le {} & \mathbb{P}\Big(\sup_{|x|>\tau_n+b_n} n|\mathbb{E}K_b(x - X_1)| \ge l_* z/4\Big) + \mathbb{P}\Big(\sup_{|x|\le\tau_n+b_n}|M_n(x)| \ge l_* z/2\Big) \\ & + \mathbb{P}\Big(\sup_{|x|\le\tau_n+b_n}|R_n(x)| \ge l_* z/4\Big) =: \mathrm{I}_2 + \mathrm{I}_3 + \mathrm{I}_4. \end{aligned} \tag{104}$$

Hence by (103) and (104), we have $\mathrm{I} \le \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 + \mathrm{I}_4$. For I_1-I_3 we shall bound them through some basic inequalities, for I_4, we will apply Corollary 6.

For I_1: By Lemma 19, $\mathbb{E}(|X_0|^q) \le K_q^q(\sum_{j\ge 0}|a_j|^{q'})^{q/q'}\mu_q^q$. Hence by Markov's inequality $\mathrm{I}_1 \le n\tau_n^{-q}\mathbb{E}(|X_0|^q) \lesssim nz^{-q\beta}\mu_q^q$, where the constant in $\lesssim$ only depends on q, β and γ.

For I_2: Since $|K|_\infty$ is bounded by K_* with support $[-1,1]$, we have $|\mathbb{E}K_b(x-X_1)| \le K_* b_n^{-1}\mathbb{P}(|X_1 - x| \le b_n)$. When $|x| > \tau_n + b_n$, $\mathbb{P}(|X_1 - x| \le b_n) \le \mathbb{P}(|X_1| \ge \tau_n)$. Hence

$$n|\mathbb{E}K_b(x-X_1)| \le nK_* b_n^{-1}\mathbb{P}(|X_1| \ge \tau_n) \le K_* b_n^{-1} n^{1-q\beta}\mathbb{E}|X_0|^q = o(K_* z),$$

in view of $z \ge c(n/b_n)^{1/2}\log^{1/2}(n)$ and $nb_n \to \infty$. Thus $\mathrm{I}_2 = 0$ for all large n.

For I_3: Let $A_n = \{-(\tau_n + b_n) + \delta_n k, k = 0, 1, \ldots, \lfloor 2(\tau_n + b_n)/\delta_n + 1\rfloor\}$, where $\delta_n = zb_n^2/(8n)$. Then

$$\sup_{|x|\le \tau_n + b_n}\min_{y\in A_n}|M_n(x) - M_n(y)| \le K_* z/4,$$

and $\mathrm{I}_3 \le \sum_{x\in A_n}\mathbb{P}(|M_n(x)| \ge l_* z/4)$. Since $|K|_\infty \le K_*$ and $|f_\epsilon|_\infty \le f_*$, for $X_{j,j-1} = \sum_{k\ge 1}a_k\epsilon_{j-k}$,

$$\begin{aligned}\mathbb{E}[K_b^2(x-X_j)|\mathcal{F}_{j-1}]| &= \int_{-\infty}^{\infty} b_n^{-2}K^2(\frac{x - X_{j,j-1} - u}{b_n})f_\epsilon(u)\mathrm{d}u \\ &= \int_{-\infty}^{\infty} b_n^{-1}K^2(y)f_\epsilon(x - b_n y - X_{j,j-1})\mathrm{d}y \\ &\le K_* f_* b_n^{-1}\int_{-\infty}^{\infty}K(y)\mathrm{d}y = K_* f_* b_n^{-1}.\end{aligned}$$

Therefore for $\xi_j(x) = K_b(x - X_j) - \mathbb{E}[K_b(x-X_j)|\mathcal{F}_{j-1}]$,

$$V(x) := \sum_{j=1}^n \mathbb{E}(\xi_j(x)^2|\mathcal{F}_{j-1}) \le nK_* f_* b_n^{-1}.$$

Note $|K_b| \le K_*/b_n$, therefore by Freedman's inequality (Lemma 18), we have

$$\begin{aligned}\mathrm{I}_3 &\le \sum_{x\in A_n}\mathbb{P}\big(|M_n(x)| \ge l_* z/4\big) \le 2\sum_{x\in A_n}\exp(-\frac{z^2}{2zb_n^{-1} + 2nb_n^{-1}}) \\ &\le \frac{32n(\tau_n + b_n)}{zb_n^2}\exp(-\frac{z^2 b_n}{4n}).\end{aligned}$$

Since $z \ge c(n/b_n)^{1/2}\log^{1/2}(n)$, for c sufficiently large $\mathrm{I}_3 = o(n/z^{q\beta})$.

For I_4: Since $\mathbb{E}[K_{b_n}(x-X_j)|\mathcal{F}_{j-1}] = \int_{\mathbb{R}}K(u)f_\epsilon(x - b_n u - X_{j,j-1})\mathrm{d}u$, we have $R_n(x) = \sum_{j=1}^n N_n(x, X_{j,j-1})$, where

$$N_n(x,y) = \int_{-\infty}^{\infty}K(u)[f_\epsilon(x - b_n u - y) - f(x - b_n u)]\mathrm{d}u. \tag{105}$$

Let function class $\mathcal{A}_n = \{N_n(x,\cdot), |x| \le \tau_n + b_n\}$, then for any function in $\mathcal{A}_n$, its up to second order derivatives are bounded by f_* and $\mathcal{N}_{\mathcal{A}_n}(f_* z/n) \le 4n(\tau_n + b_n)/z$. Therefore by Corollary 6, we have $\mathrm{I}_4 \lesssim \mu_q^q nz^{-q\beta}$, where the constant in $\lesssim$ only depends on β, q and γ. Thus (25) follows from I_1-I_4.

For the LRD case, define $M_n(x)$ and $R_n(x)$ as in the SRD case and let $\tau_n = z$. Again we have

$$\mathbb{P}\Big(\sup_{x\in\mathbb{R}} n|\hat{f}_n(x) - \mathbb{E}\hat{f}_n(x)| \geq l_* z\Big) \leq \sum_{j=1}^n \mathbb{P}(|X_j| \geq z)$$
$$+ \mathbb{P}\Big(\sup_{|x|>z+b_n} n|\mathbb{E}K_b(x - X_1)| \geq l_* z/4\Big) + \mathbb{P}\Big(\sup_{|x|\leq z+b_n} |M_n(x)| \geq l_* z/2\Big)$$
$$+ \mathbb{P}\Big(\sup_{|x|\leq z+b_n} |R_n(x)| \geq l_* z/4\Big) =: \mathrm{I}_1 + \mathrm{I}_2 + \mathrm{I}_3 + \mathrm{I}_4.$$

Using same argument as for SRD case with τ_n replaced by z, we obtain $\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3 \lesssim nz^{-q}\mu_q^q$, where the constants in $\lesssim$ only depend on q, β and γ. For I_4, we still have (105). Let $\mathcal{A}_n = \{N_n(x, \cdot), |x| \leq z + b_n\}$. Then $\mathcal{N}_{\mathcal{A}_n}(f_* z/n) \leq 4n(z + b_n)/z$. Therefore by Corollary 10, we have $\mathrm{I}_3 \lesssim (\mu_q^{2q} \vee \mu_q^q) n^{3/2-\beta} z^{-q}$, where the constant in $\lesssim$ only depends on β, q and γ. ■

Proof [Proof of Corollary 16] Let $\mathcal{G}_i = (\epsilon_i, \epsilon_{i-1}, \ldots; \eta_i, \eta_{i-1}, \ldots)$ and $X_{i,i-1} = \sum_{j=1}^\infty a_j \epsilon_{i-j}$. Then we have $\mathbb{E}[L(X_i, Y_i, h(X_i))|\mathcal{G}_{i-1}] = Q_h(X_{i,i-1})$, where

$$Q_h(w) = \int_{-\infty}^{\infty} \mathbb{E}[L(u, H_0(u, \eta_i), h(u))] f_\epsilon(u - w) \mathrm{d}u. \tag{106}$$

Let $J_h(x) = Q_h(x) - \mathbb{E}[L(X_i, Y_i, h(X_i))]$. Write

$$n(R_n(h) - R(h)) = \sum_{i=1}^n \Big[L(X_i, Y_i, h(X_i)) - Q_h(X_{i,i-1})\Big] + \sum_{i=1}^n J_h(X_{i,i-1}) =: \mathrm{I}_1(h) + \mathrm{I}_2(h).$$

For $h, g \in \mathcal{H}$, let $D(h, g) = \sup_{x,y\in\mathbb{R}} |L(x, y, h(x)) - L(x, y, g(x))|$. Let $\mathcal{H}_n$ be the subset of $\mathcal{H}$ such that $\sup_{h_1\in\mathcal{H}} \inf_{h_2\in\mathcal{H}_n} D(h_1, h_2) \leq z/(4n)$ and $|\mathcal{H}_n| \leq \mathcal{N}_{\mathcal{A}}(z/(4n))$. Then for $\tau > 0$,

$$\mathbb{P}(n\Psi_n/f_* \geq z + \tau) \leq \mathbb{P}\Big(\max_{h\in\mathcal{H}_n} n|R_n(h) - R(h)|/f_* \geq z/2 + \tau\Big)$$
$$\leq \sum_{h\in\mathcal{H}_n} \mathbb{P}(|\mathrm{I}_1(h)|/f^* \geq z/4) + \mathbb{P}\Big(\max_{h\in\mathcal{H}_n} |\mathrm{I}_2(h)|/f_* \geq z/4 + \tau\Big).$$

Since $0 \leq L \leq 1$, $f^* \geq 1$ and the summands of $\mathrm{I}_1(h)$ are bounded martingale differences with respect to $\mathcal{G}_i$, by Azuma's inequality, we have $\sum_{h\in\mathcal{H}_n} \mathbb{P}(|\mathrm{I}_1(h)| \geq z) \leq 2|\mathcal{H}_n| e^{-z^2/(32n)}$. Since both $\int_{-\infty}^\infty |f'_\epsilon(x)| \mathrm{d}x$ and $\int_{-\infty}^\infty |f''_\epsilon(x)| \mathrm{d}x$ are bounded by f_*, by (106), for $h \in \mathcal{H}$, Q_h, Q'_h and Q''_h exist and are uniformly bounded by f_* in absolute value. Thus (16) (resp. (20)) follows from applying Corollary 6 to Q_h/f_* with $\tau = C_q a_* \mu_q c(n, q)$ (resp. Corollary 10 with $\tau = C_{\beta,q,\gamma} \mu_q n^{3/2-\beta}$). ■

Acknowledgments

We thank the reviewers and the editor for very helpful suggestions, which substantially improve the paper.

References

Radosław Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008. ISSN 1083-6489.

Terrence M. Adams and Andrew B. Nobel. Uniform approximation of Vapnik-Chervonenkis classes. *Bernoulli*, 18(4):1310–1319, 2012. ISSN 1350-7265.

Alekh Agarwal and John C Duchi. The generalization ability of online algorithms for dependent data. *IEEE Transactions on Information Theory*, 59(1):573–587, 2013.

Pierre Alquier and Olivier Wintenberger. Fast rates in learning with dependent observations. *JMLR: Workshop and Conference Proceedings 1-15*, 2012.

Donald W. K. Andrews. Nonstrong mixing autoregressive processes. *J. Appl. Probab.*, 21 (4):930–934, 1984. ISSN 0021-9002.

M. A. Arcones and B. Yu. Central limit theorems for empirical and U-processes of stationary mixing sequences. *J. Theoret. Probab.*, 7(1):47–71, 1994. ISSN 0894-9840.

Ery Arias-Castro, David Mason, and Bruno Pelletier. On the estimation of the gradient lines of a density and the consistency of the mean-shift algorithm. *J. Mach. Learn. Res.*, 17:Paper No. 43, 28, 2016. ISSN 1532-4435.

Richard T. Baillie. Long memory processes and fractional integration in econometrics. *J. Econometrics*, 73(1):5–59, 1996. ISSN 0304-4076.

Jan Beran. *Statistics for long-memory processes*, volume 61 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1994. ISBN 0-412-04901-5.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003. ISSN 0091-1798.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. ISBN 978-0-19-953525-5. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.

Christian Brownlees, Emilien Joly, and Gábor Lugosi. Empirical risk minimization for heavy-tailed losses. *Ann. Statist.*, 43(6):2507–2536, 2015. ISSN 0090-5364.

Donald L. Burkholder. Sharp inequalities for martingales and stochastic integrals. *Astérisque*, (157-158):75–94, 1988. ISSN 0303-1179. Colloque Paul Lévy sur les Processus Stochastiques (Palaiseau, 1987).

Andrea Caponnetto and Alexander Rakhlin. Stability properties of empirical risk minimization over Donsker classes. *J. Mach. Learn. Res.*, 7:2565–2583, 2006. ISSN 1532-4435.

Olivier Catoni. Challenging the empirical mean and empirical variance: a deviation study. *Ann. Inst. Henri Poincaré Probab. Stat.*, 48(4):1148–1185, 2012. ISSN 0246-0203.

Likai Chen and Wei Biao Wu. Stability and asymptotics for autoregressive processes. *Electron. J. Stat.*, 10(2):3723–3751, 2016. ISSN 1935-7524.

Yen-Chi Chen, Christopher R. Genovese, and Larry Wasserman. A comprehensive approach to mode clustering. *Electron. J. Stat.*, 10(1):210–241, 2016. ISSN 1935-7524.

Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Testing many moment inequalities. *arXiv preprint arXiv:1312.7614, Review of Economic Studies, revise and resubmit*, 2017.

Yuan Shih Chow and Henry Teicher. *Probability theory*. Springer Texts in Statistics. Springer-Verlag, New York, third edition, 1997. ISBN 0-387-98228-0. Independence, interchangeability, martingales.

Luc Devroye, László Györfi, and Gábor Lugosi. *A probabilistic theory of pattern recognition*, volume 31 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1996. ISBN 0-387-94618-7.

A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.*, 27: 642–669, 1956. ISSN 0003-4851.

Uwe Einmahl and Deli Li. Characterization of LIL behavior in Banach space. *Trans. Amer. Math. Soc.*, 360(12):6677–6693, 2008. ISSN 0002-9947.

Uwe Einmahl and David M. Mason. Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.*, 33(3):1380–1403, 2005. ISSN 0090-5364.

David A. Freedman. On tail probabilities for martingales. *Ann. Probability*, 3:100–118, 1975.

Evarist Giné and Armelle Guillou. Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):907–921, 2002. ISSN 0246-0203. En l'honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.

Evarist Giné and Richard Nickl. Confidence bands in density estimation. *Ann. Statist.*, 38 (2):1122–1170, 2010. ISSN 0090-5364.

Lee-Ad Gottlieb, Aryeh Kontorovich, and Robert Krauthgamer. Efficient regression in metric spaces via approximate lipschitz extension. *IEEE Transactions on Information Theory*, 2017.

Clive WJ Granger and Roselyne Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of time series analysis*, 1(1):15–29, 1980.

Zheng-Chu Guo and Lei Shi. Classification with non-i.i.d. sampling. *Math. Comput. Modelling*, 54(5-6):1347–1364, 2011. ISSN 0895-7177.

Erich Haeusler. An exact rate of convergence in the functional central limit theorem for special martingale difference arrays. *Z. Wahrsch. Verw. Gebiete*, 65(4):523–534, 1984. ISSN 0044-3719.

Hanyuan Hang and Ingo Steinwart. Fast learning from α-mixing observations. *Journal of Multivariate Analysis*, 127:184–199, 2014.

Hanyuan Hang and Ingo Steinwart. A bernstein-type inequality for some mixing processes and dynamical systems with an application to learning. *Annals of Statistics, To Appear*, 2016.

Hanyuan Hang, Ingo Steinwart, Yunlong Feng, and Johan AK Suykens. Kernel density estimation for dynamical systems. *arXiv preprint arXiv:1607.03792*, 2016.

Hwai-Chung Ho and Tailen Hsing. On the asymptotic expansion of the empirical process of long-memory moving averages. *Ann. Statist.*, 24(3):992–1024, 1996. ISSN 0090-5364.

Jonathan RM Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

Wenxin Jiang. On uniform deviations of general empirical risks with unboundedness, dependence, and high dimensionality. *J. Mach. Learn. Res.*, 10:977–996, 2009. ISSN 1532-4435.

T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005. ISSN 0091-1798.

A Kontorovich and A Brockwell. A strong law of large numbers for strongly mixing processes. *Communications in Statistics-Theory and Methods*, 43(18):3777–3796, 2014.

Aryeh Kontorovich and Maxim Raginsky. Concentration of measure without independence: a unified approach via the martingale method. In *Convexity and Concentration*, pages 183–210. Springer, 2017.

Aryeh Kontorovich and Roi Weiss. Uniform Chernoff and Dvoretzky-Kiefer-Wolfowitz-type inequalities for Markov chains and related processes. *J. Appl. Probab.*, 51(4):1100–1113, 2014. ISSN 0021-9002. doi: 10.1239/jap/1421763330. URL `http://dx.doi.org/10.1239/jap/1421763330`.

Leonid Kontorovich and Kavita Ramanan. Concentration inequalities for dependent random variables via the martingale method. *Ann. Probab.*, 36(6):2126–2158, 2008. ISSN 0091-1798.

Michael R. Kosorok. *Introduction to empirical processes and semiparametric inference*. Springer Series in Statistics. Springer, New York, 2008. ISBN 978-0-387-74977-8. doi: 10.1007/978-0-387-74978-5. URL `http://dx.doi.org/10.1007/978-0-387-74978-5`.

Vitaly Kuznetsov and Mehryar Mohri. Generalization bounds for time series prediction with non-stationary processes. In *International Conference on Algorithmic Learning Theory*, pages 260–274. Springer, 2014.

Vitaly Kuznetsov and Mehryar Mohri. Learning theory and algorithms for forecasting non-stationary time series. In *Advances in neural information processing systems*, pages 541–549, 2015.

John Lafferty, Han Liu, and Larry Wasserman. Sparse nonparametric graphical models. *Statist. Sci.*, 27(4):519–537, 2012. ISSN 0883-4237.

Johannes Lederer and Sara van de Geer. New concentration inequalities for suprema of empirical processes. *Bernoulli*, 20(4):2020–2038, 2014. ISSN 1350-7265.

Michel Ledoux. On talagrand's deviation inequalities for product measures. *ESAIM: Probability and statistics*, 1:63–87, 1997.

Michel Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. ISBN 0-8218-2864-9.

Shlomo Levental. Uniform limit theorems for Harris recurrent Markov chains. *Probab. Theory Related Fields*, 80(1):101–118, 1988. ISSN 0178-8051.

Han Liu, Min Xu, Haijie Gu, Anupam Gupta, John Lafferty, and Larry Wasserman. Forest density estimation. *J. Mach. Learn. Res.*, 12:907–951, 2011. ISSN 1532-4435.

Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996. ISSN 1016-443X.

Katalin Marton. Measure concentration for a class of random processes. *Probab. Theory Related Fields*, 110(3):427–439, 1998. ISSN 0178-8051.

P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, 18(3):1269–1283, 1990. ISSN 0091-1798.

Pascal Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000. ISSN 0091-1798.

Pascal Massart. *Concentration inequalities and model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007. ISBN 978-3-540-48497-4; 3-540-48497-3. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003, With a foreword by Jean Picard.

Dharmendra S. Modha and Elias Masry. Minimum complexity regression estimation with weakly dependent observations. *IEEE Trans. Inform. Theory*, 42(6, part 2):2133–2145, 1996. ISSN 0018-9448.

Mehryar Mohri and Afshin Rostamizadeh. Stability bounds for stationary ϕ-mixing and β-mixing processes. *J. Mach. Learn. Res.*, 11:789–814, 2010. ISSN 1532-4435.

S. V. Nagaev. Large deviations of sums of independent random variables. *Ann. Probab.*, 7 (5):745–789, 1979. ISSN 0091-1798.

Tuan D. Pham and Lanh T. Tran. Some mixing properties of time series models. *Stochastic Process. Appl.*, 19(2):297–303, 1985. ISSN 0304-4149. doi: 10.1016/0304-4149(85) 90031-6. URL `http://dx.doi.org/10.1016/0304-4149(85)90031-6`.

Svetlozar Rachev and Stefan Mittnik. *Stable Paretian models in finance.* John Willey & Sons, New York, 2000.

Alessandro Rinaldo, Aarti Singh, Rebecca Nugent, and Larry Wasserman. Stability of density-based clustering. *J. Mach. Learn. Res.*, 13:905–948, 2012. ISSN 1532-4435.

Emmanuel Rio. Moment inequalities for sums of dependent random variables under projective conditions. *J. Theoret. Probab.*, 22(1):146–163, 2009. ISSN 0894-9840.

Paul-Marie Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.*, 28(1):416–461, 2000. ISSN 0091-1798.

Cosma Shalizi and Aryeh Kontorovich. Predictive pac learning and process decompositions. In *Advances in neural information processing systems*, pages 1619–1627, 2013.

Ingo Steinwart and Andreas Christmann. Fast learning from non-iid observations. In *Advances in Neural Information Processing Systems*, pages 1768–1776, 2009.

M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Ann. Probab.*, 22(1): 28–76, 1994. ISSN 0091-1798.

Michel Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995. ISSN 0073-8301.

Michel Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126 (3):505–563, 1996. ISSN 0020-9910.

A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998. ISBN 0-521-49603-9; 0-521-78450-6.

Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3. With applications to statistics.

Vladimir N. Vapnik. *Statistical learning theory.* Adaptive and Learning Systems for Signal Processing, Communications, and Control. John Wiley & Sons, Inc., New York, 1998. ISBN 0-471-03003-1. A Wiley-Interscience Publication.

Vladimir N. Vapnik. *The nature of statistical learning theory.* Statistics for Engineering and Information Science. Springer-Verlag, New York, second edition, 2000. ISBN 0-387-98780-0.

VN Vapnik and A Ya Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264, 1971.

Wei Biao Wu. Empirical processes of long-memory sequences. *Bernoulli*, 9(5):809–831, 2003. ISSN 1350-7265.

Wei Biao Wu and Jan Mielniczuk. Kernel density estimation for linear processes. *Ann. Statist.*, 30(5):1441–1459, 2002. ISSN 0090-5364.

Bin Yu. Rates of convergence for empirical processes of stationary mixing sequences. *The Annals of Probability*, pages 94–116, 1994.

Bin Zou and Luoqing Li. The performance bounds of learning machines based on exponentially strongly mixing sequences. *Comput. Math. Appl.*, 53(7):1050–1058, 2007. ISSN 0898-1221.

Bin Zou, Luoqing Li, and Zongben Xu. The generalization performance of erm algorithm with strongly mixing observations. *Machine learning*, 75(3):275–295, 2009.