Electron. Commun. Probab. **23** (2018), no. 38, 1–14.
https://doi.org/10.1214/18-ECP142
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Stein's method for nonconventional sums

Yeor Hafouta*

Abstract

We obtain almost optimal convergence rate in the central limit theorem for (appropriately normalized) “nonconventional" sums of the form $S_N = \sum_{n=1}^{N}(F(\xi_n, \xi_{2n}, ..., \xi_{\ell n}) - \bar{F})$. Here $\{\xi_n : n \geq 0\}$ is a sufficiently fast mixing vector process with some stationarity conditions, F is bounded Hölder continuous function and $\bar{F}$ is a certain centralizing constant. Extensions to more general functions F will be discusses, as well. Our approach here is based on the so called Stein's method, and the rates obtained in this paper significantly improve the rates in [7]. Our results hold true, for instance, when $\xi_n = (T^n f_i)_{i=1}^{\wp}$ where T is a topologically mixing subshift of finite type, a hyperbolic diffeomorphism or an expanding transformation taken with a Gibbs invariant measure, as well as in the case when $\{\xi_n : n \geq 0\}$ forms a stationary and exponentially fast ϕ-mixing sequence, which, for instance, holds true when $\xi_n = (f_i(\Upsilon_n))_{i=1}^{\wp}$ where Υ_n is a Markov chain satisfying the Doeblin condition considered as a stationary process with respect to its invariant measure.

Keywords: central limit theorem; Berry–Esseen theorem; mixing; nonconventional setup; Stein's method.
AMS MSC 2010: 60F05.
Submitted to ECP on April 6, 2017, final version accepted on June 5, 2018.

1 Introduction

Let Φ be the standard normal distribution function and let $X_1, X_2, X_3...$ be a sequence of independent and identically distributed random variables such that $\mathbb{E}X_1 = 0$ and $0 < \mathbb{E}X_1^2 = \sigma^2 < \infty$. The classical Berry-Esseen theorem provides a uniform approximation of the error term in the central limit theorem (CLT) for the sums $\hat{S}_n = \frac{1}{\sqrt{n}\sigma}\sum_{k=1}^{n} X_k$, stating that for any $n \in \mathbb{N}$,

$$\sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \leq \frac{C\mathbb{E}|X_1|^3}{\sqrt{n}} \tag{1.1}$$

where F_n is the distribution function of $\hat{S}_n$ (see Section 6 of Ch. III in [17]) and $C > 0$ is an absolute constant which by efforts of many researchers was optimized by now to a number a bit less than 1/2.

During the last 50 years there were several extensions of the CLT for sums of weakly dependent random variables and for martingales, including many estimates of error terms. Among the most used methods in the case of weak dependence are Gordin's

*Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel. E-mail: yeor.hafouta@mail.huji.ac.il

method for martingale approximation (see [5], [12] and [6]) and Stein's method (see [15]). While Stein's method can yield close to optimal convergence rate (see [15] and [13]), the martingale approximation method can not, since Berry-Esseen type estimates for martingales do not yield (in general) optimal convergence rates even for sums of independent random variables (see, for instance [6] and [1]).

Partially motivated by the research on nonconventional ergodic theorems (the term "nonconventional" comes from [4]), probabilistic limit theorems for sums of the form $S_N=\sum_{n=1}^{N}F(\xi_{q_1(n)},\xi_{q_2(n)},...,\xi_{q_\ell(n)})$ have become a well studied topic. Here $\{\xi_n,n\geq 0\}$ is a sufficiently fast mixing processes with some stationarity properties and F is a function satisfying some regularity conditions. The summands here are nonstationary and long range dependent which makes it difficult to apply standard methods. This line of research started in [9], where a functional CLT was proved for the normalized sums $N^{-\frac12}S_{[Nt]}$ using characteristic function estimates. In [11] a functional CLT was established for more general q_i's than in [9], where one of the main parts of the proof was showing that the martingale approximation approach is applicable. These results included the case when $q_i(n)=in$, which was the original motivation for the study of nonconventional averages (see [4]). In [7] the authors estimated the convergence rate of $\mathcal{Z}_N=N^{-\frac12}S_N$ in the Kolmogorov (uniform) metric towards its weak limit under the assumptions of [11]. The proof relied on Berry-Esseen type results for martingales, which led to estimates of order $N^{-\frac{1}{10}}\ln(N+1)$, which is far from optimal. In the special case when ξ_n's are independent the authors provided optimal rate of order $N^{-\frac12}$ relying on Stein's method for sums of locally dependent random variables (see [3]).

The goal of this paper is to show that Stein's method is applicable for nonconventional sums when ξ_n's are weakly dependent, and to significantly improve the rates obtained in [7]. We first consider the case when F is a bounded Hölder continuous function and $q_i(n)=in$ for any $1\leq i\leq \ell$ and $n\in\mathbb{N}$, and (in the self normalized case) provide almost optimal upper bound of the form

$$\sup_{x\in\mathbb{R}}|P(S_N\leq x\sqrt{\mathbb{E}S_N^2})-\Phi(x)|\leq CN^{-\frac12}\ln^2(N+1) \tag{1.2}$$

assuming that $D^2>0$, where $D^2=\lim_{N\to\infty}N^{-1}\mathbb{E}S_N^2$. We also obtain rates of the form

$$\sup_{x\in\mathbb{R}}|P(N^{-\frac12}S_N\leq x)-\Phi(xD^{-1})|\leq C_\epsilon N^{-\frac12+\epsilon} \tag{1.3}$$

where $\epsilon>0$ is an arbitrary positive constant and C_ϵ is a constant which in general depends on ϵ. When $\{\xi_n: n\geq 0\}$ forms a stationary and exponentially fast ϕ-mixing sequence then, in fact, we show that (1.2) and (1.3) hold true for any bounded function F which is not necessarily continuous. Convergence rates for more general functions and more general indexes $q_i(n)$'s will be discussed, as well.

As in [7], our results hold true when, for instance, $\xi_n=T^nf$ where $f=(f_1,...,f_d)$, T is a topologically mixing subshift of finite type, a hyperbolic diffeomorphism or an expanding transformation taken with a Gibbs invariant measure, as well as in the case when $\xi_n=f(\Upsilon_n)$, $f=(f_1,...,f_d)$ where Υ_n is a Markov chain satisfying the Doeblin condition considered as a stationary process with respect to its invariant measure. In fact, any stationary and exponentially fast ϕ-mixing sequence $\{\xi_n,\, n\geq 0\}$ can be considered. In the dynamical systems case each f_i should be either Hölder continuous or piecewise constant on elements of Markov partitions. As an application we can consider $\xi_n=((\xi_n)_1,...,(\xi_n)_\ell)$, $(\xi_n)_j=\mathbb{I}_{A_j}(T^nx)$ in the dynamical systems case and $(\xi_n)_j=\mathbb{I}_{A_j}(\Upsilon_n)$ in the Markov chain case, where $\mathbb{I}_A$ is the indicator of an appropriate set A. Let $F=F(x_1,...,x_\ell)$, $x_j=(x_j^{(1)},...,x_j^{(\ell)})$ be a bounded Hölder continuous function which identifies with the function $G(x_1,...,x_\ell)=x_1^{(1)}\cdot x_2^{(2)}\cdots x_\ell^{(\ell)}$ on the cube $([0,1]^\wp)^\ell$.

Let $N(n)$ be the number of l's between 0 and n for which $T^{q_j(l)}x \in A_j$ for $j = 0, 1, ..., \ell$ (or $\Upsilon_{q_j(l)} \in A_j$ in the Markov chains case), where we set $q_0 = 0$, namely the number of $\ell-$tuples of return times to A_j's (either by $T^{q_j(l)}$ or by $\Upsilon_{q_j(l)}$). Then our results yield a central limit theorem with almost optimal convergence rate for the numbers $N(n)$.

2 Preliminaries and main results

Our setup consists of a $\wp$-dimensional stochastic process $\{\xi_n, n \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, P)$ and a family of sub-$\sigma-$algebras $\mathcal{F}_{k,l}$, $-\infty \leq k \leq l \leq \infty$ such that $\mathcal{F}_{k,l} \subset \mathcal{F}_{k',l'} \subset \mathcal{F}$ if $k' \leq k$ and $l' \geq l$. We will impose restrictions of the mixing coefficients

$$\phi(n) = \sup\{\phi(\mathcal{F}_{-\infty,k}, \mathcal{F}_{k+n,\infty}) : k \in \mathbb{Z}\} \tag{2.1}$$

where we recall that for any two sub-$\sigma-$algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$,

$$\phi(\mathcal{G}, \mathcal{H}) = \sup\big\{\big|\frac{P(A \cap B)}{P(A)} - P(B)\big| : A \in \mathcal{G}, B \in \mathcal{H}, P(A) > 0\big\}.$$

In order to ensure some applications, in particular, to dynamical systems we will not assume that ξ_n is measurable with respect to $\mathcal{F}_{n,n}$ but instead impose conditions on the approximation rates

$$\beta_\infty(r) = \sup_{k\geq 0} \|\xi_k - \mathbb{E}[\xi_k|\mathcal{F}_{k-r,k+r}]\|_{L^\infty} \tag{2.2}$$

where $\|X\|_{L^\infty}$ denotes the essential supremum of the absolute value of a random variable X.

We do not require stationarity of the process $\{\xi_n, n \geq 0\}$, assuming only that the distribution of ξ_n does not depend on n and that the joint distribution of (ξ_n, ξ_m) depends only on $n - m$, which we write for further reference by

$$\xi_n \sim \mu \quad \text{and} \quad \big(\xi_n, \xi_m\big) \sim \mu_{m-n} \tag{2.3}$$

where $Y \sim \mu$ means that Y has μ for its distribution.

Let $F = F(x_1, ..., x_\ell) : (\mathbb{R}^\wp)^\ell \to \mathbb{R}$, $\ell \geq 1$ be a bounded Hölder function and let $M > 0$ and $\kappa \in (0, 1]$ be such that

$$|F(x)| \leq M \quad \text{and} \tag{2.4}$$

$$|F(x) - F(y)| \leq M \sum_{i=1}^{\ell} |x_i - y_i|^\kappa \tag{2.5}$$

for any $x = (x_1, ..., x_\ell)$ and $y = (y_1, ..., y_\ell)$ in $(\mathbb{R}^\wp)^\ell$. To simplify formulas we assume the centering condition

$$\bar{F} = \int F(x_1, x_2, ..., x_\ell)d\mu(x_1)d\mu(x_2)...d\mu(x_\ell) = 0 \tag{2.6}$$

which is not really a restriction since we can always replace F by $F - \bar{F}$. The main goal of this paper is to prove a central limit theorem with convergence rate for the normalized sums $(c_N)^{-1}S_N$, where

$$S_N = \sum_{n=1}^{N} F(\xi_n, \xi_{2n}, ..., \xi_{\ell n})$$

and either $c_N = N^{-\frac{1}{2}}$ or $c_N = \sqrt{\mathbb{E}S_N^2}$. Our results will rely on

Assumption 2.1. *There exist $d > 0$ and $c \in (0,1)$ such that*

$$\phi(n) + (\beta_\infty(n))^\kappa \le dc^n \tag{2.7}$$

for any $n \in \mathbb{N}$.

Before formulating our main results we will first state the following

Theorem 2.2. *Suppose that Assumption (2.1) is satisfied. Then the limit $D^2 = \lim_{N\to\infty} N^{-1}\mathbb{E}S_N^2$ exists and there exists $C_1 > 0$ which depends only on ℓ, c and d such that*

$$|\mathbb{E}S_N^2 - D^2 N| \le C_1 M^2 N^{\frac{1}{2}} \tag{2.8}$$

for any $N \in \mathbb{N}$. Moreover, $D^2 > 0$ if and only if there exists no stationary in the wide sense process $\{V_n : n \ge 1\}$ such that

$$F(\xi_n^{(1)}, \xi_{2n}^{(n)}, ..., \xi_{\ell n}^{(\ell)}) = V_{n+1} - V_n, \text{ P-almost surely}$$

for any $n \in \mathbb{N}$, where $\xi^{(i)}$, $i = 1, ..., \ell$ are independent copies of $\xi = \{\xi_n : n \ge 1\}$.

This theorem is a consequence of the arguments in [11], [10] and [7] and is formulated here for readers' convenience.

Next, recall that the Kolmogorov (uniform) metric is defined for each pairs of distributions $\mathcal{L}_1$ and $\mathcal{L}_2$ on $\mathbb{R}$ with distribution functions G_1 and G_2 by

$$d_K(\mathcal{L}_1, \mathcal{L}_2) = \sup_{x\in\mathbb{R}} |G_1(x) - G_2(x)|.$$

For any random variable X we denote its law by $\mathcal{L}(X)$. Our main result is the following

Theorem 2.3. *Suppose that Assumption (2.1) holds true and that $D^2 > 0$. Set $s_N = \sqrt{\mathbb{E}S_N^2}$ and $Z_N = (s_N)^{-1}S_N$ when $s_N > 0$, while when $s_N = 0$ we set $Z_N = N^{-\frac{1}{2}}S_N$. Let $\mathcal{N}(0,1)$ be the zero mean normal distribution with variance 1. Then there exists a constant $C > 0$ which depends only on ℓ, d and c such that*

$$d_K(\mathcal{L}(Z_N), \mathcal{N}(0,1)) \le C \max(1, \rho^3) N^{-\frac{1}{2}} \ln^2(N+1) \tag{2.9}$$

for any $N \in \mathbb{N}$, where $\rho = MD^{-1}$. Moreover, for any $\epsilon > 0$ there exists a constant $c_\epsilon > 0$ which depends only on ϵ, c, d and ℓ so that for any $N \ge 1$,

$$d_K(\mathcal{L}(N^{-\frac{1}{2}}S_N), \mathcal{N}(0, D^2)) \le c_\epsilon \max(1, \rho^3) N^{-\frac{1}{2}+\epsilon} \tag{2.10}$$

where $\mathcal{N}(0, D^2)$ is the zero mean normal distribution with variance D^2. When $\beta_\infty(r_0) = 0$ for some r_0 then (2.9) and (2.10) hold true with constants C and c_ϵ which depend also on r_0, assuming only that F is a bounded function satisfying (2.4).

Note that $\beta_\infty(0) = 0$ when $\mathcal{F}_{m,n} = \sigma\{\xi_{\max(0,m)}, ..., \xi_{\max(0,n)}\}$ and therefore when the processes $\{\xi_n : n \ge 0\}$ itself is exponentially fast ϕ-mixing (i.e. when (2.7) holds true with these $\sigma-$algebras) we obtain (2.9) for any bounded function F.

The outline of the proof is as follows. Relying on [13], Stein's method becomes effective for the sum S_N when $\{F_n : 1 \le n \le N\}$, $F_n = F(\xi_n, \xi_{2n}, ..., \xi_{\ell n})$ are locally weak dependent in the sense that there exist sets A_n and nonnegative integers $d_{n,m}$, $1 \le n, m \le N$ so that $n \in A_n$, $a_n = |A_n|$ and $b_n(k) = |\{1 \le m \le N : d_{n,m} = k\}|$, $k \ge 0$ are small relatively to N, F_n and $\{F_s : s \notin A_n\}$ are weakly dependent and the random vectors $\mathbf{F}_n = \{F_k : k \in A_n\}$ and $\mathbf{F}_m = \{F_s : s \in A_m\}$ are weakly dependent when $d_{n,m}$ is sufficiently large. We first reduce the problem of approximation of the left hand side of (2.9) to the case when $\xi = \{\xi_n : n \ge 0\}$ forms a sufficiently fast ϕ-mixing process. Then we consider the sets

$$A_n = A_{n,N,l} = \{1 \le m \le N : \min_{1\le i,j\le\ell} |in - jm| \le l\}$$

and the numbers $d_{n,m} = \min\{|ia - jb| : a \in A_n, b \in A_m, 1 \le i,j \le \ell\}$ and show that a_n and $b_n(k)$ defined above are of order l. In Section 3 we provide estimates which will show that the required type of the above weak dependence is satisfied, and then we take l of the form $l = A\ln(N+1)$ to complete the proof. In fact, existing estimates on the left hand side of (2.9) using Stein's method become effective only after using the expectation estimates obtained in Section 3 even for "conventional" sums of ϕ-mixing sequences (i.e. in the case $\ell = 1$), which is a particular case of our setup, and so, in particular, we show that Stein's method is effective for such sums and yields almost optimal convergence rate.

3 Auxiliary results

Lemma 3.1. *Let X and Y be two random variables defined on the same probability space. Let Z be a random variable with density ρ bounded from above by some constant $c > 0$. Then,*

$$d_K(\mathcal{L}(Y), \mathcal{L}(Z)) \le 3d_K(\mathcal{L}(X), \mathcal{L}(Z)) + 4c\|X - Y\|_{L^\infty} \quad \text{and for any } \ b \ge 1,$$
$$d_K(\mathcal{L}(Y), \mathcal{L}(Z)) \le 3d_K(\mathcal{L}(X), \mathcal{L}(Z)) + (1 + 4c)\|X - Y\|_{L^b}^{1-\frac{1}{b+1}}.$$

The second inequality is proved in Lemma 3.3 in [8], while the proof of the first inequality goes in the same way as the proof of that Lemma 3.3, taking in (3.2) from there $\delta = \|X - Y\|_{L^\infty}$.

Next, we recall that (see [2], Ch. 4) for any two sub-$\sigma-$algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$,

$$2\phi(\mathcal{G}, \mathcal{H}) = \sup\{\|\mathbb{E}[g|\mathcal{G}] - \mathbb{E}g\|_{L^\infty} : g \in L^\infty(\Omega, \mathcal{H}, P),\ \|g\|_{L^\infty} \le 1\}. \tag{3.1}$$

The following lemma does not seem to be new but for readers' convenience and completeness we will sketch its proof here.

Lemma 3.2. *Let $\mathcal{G}_1, \mathcal{G}_2 \subset \mathcal{F}$ be two sub-$\sigma-$algebras of $\mathcal{F}$ and for $i = 1, 2$ let V_i be a $\mathbb{R}^{d_i}$-valued random $\mathcal{G}_i$-measurable vector with distribution μ_i. Set $d = d_1 + d_2$, $\mu = \mu_1 \times \mu_2$, denote by κ the distribution of the random vector (V_1, V_2) and consider the measure $\nu = \frac{1}{2}(\kappa + \mu)$. Let $\mathcal{B}$ be the Borel $\sigma-$algebra on $\mathbb{R}^d$ and $H \in L^\infty(\mathbb{R}^d, \mathcal{B}, \nu)$. Then $\mathbb{E}[H(V_1, V_2)|\mathcal{G}_1]$ and $\mathbb{E}H(v, V_2)$ exist for μ_1-almost any $v \in \mathbb{R}^{d_1}$ and*

$$|\mathbb{E}[H(V_1, V_2)|\mathcal{G}_1] - h(V_1)| \le 2\|H\|_{L^\infty(\mathbb{R}^d, \mathcal{B}, \nu)}\phi(\mathcal{G}_1, \mathcal{G}_2),\ \ P - a.s. \tag{3.2}$$

where $h(v) = \mathbb{E}H(v, V_2)$ and a.s. stands for almost surely.

Proof. Clearly H is bounded μ and κ a.s.. Thus $\mathbb{E}[H(V_1, V_2)|\mathcal{G}_1]$ exists and existence of $\mathbb{E}H(v, V_2)$ (μ_1-a.s.) follows from the Fubini theorem. Relying on (3.1), inequality (3.2) follows easily for functions of the form $G(v_1, v_2) = \sum_i I(v_1 \in A_i)g_i(v_2)$ where $\{A_i\}$ is a measurable partition of the support of μ_1. Any uniformly continuous function H is a uniform limit of functions of the above form, which implies that (3.2) holds true for uniformly continuous functions. Finally, by Lusin's theorem (see [14]), any function $H \in L^\infty(R^d, \mathcal{B}, \nu)$ is an L^1 (and a.s.) limit of a sequence $\{H_n\}$ of continuous functions with compact support satisfying $\|H_n\|_{L^\infty(\mathbb{R}^d, \mathcal{B}, \nu)} \le \|H\|_{L^\infty(\mathbb{R}^d, \mathcal{B}, \nu)}$ and (3.2) follows for any $H \in L^\infty(\mathbb{R}^d, \mathcal{B}, \nu)$. $\square$

Corollary 3.3. *Let U_i be a d_i-dimensional random vector, $i = 1, ..., k$ defined on the probability space $(\Omega, \mathcal{F}, P)$ from Section 2. Suppose that each U_i is $\mathcal{F}_{m_i, n_i}$-measurable, where $n_{i-1} < m_i \le n_i < m_{i+1}$, $i = 1, ..., k$, $n_0 = -\infty$ and $m_{k+1} = \infty$. Let $\{\mathcal{C}_i : 1 \le i \le s\}$*

be a partition of $\{1,2,...,k\}$*. Denote by* μ_i *the distribution of the random vector* $U(\mathcal{C}_i)=\{U_j : j\in\mathcal{C}_i\}$*,* $i=1,...,s$*. Then, for any bounded Borel function* $H:\mathbb{R}^{d_1+d_2+...+d_k}\to\mathbb{R}$*,*

$$\Big|\mathbb{E}H(U_1,U_2,...,U_k)-\int H(u_1,u_2,...,u_k)d\mu_1(u^{(\mathcal{C}_1)})d\mu_2(u^{(\mathcal{C}_2)})...d\mu_s(u^{(\mathcal{C}_s)})\Big| \le 4\|H\|_\infty\sum_{i=2}^k\phi(m_i-n_{i-1}) \tag{3.3}$$

where $u^{(\mathcal{C}_i)}=\{u_j : j\in\mathcal{C}_i\}$*,* $i=1,...,s$ *and* $\|H\|_\infty$ *stands for the supremum of* $|H|$*. Namely, let* $U^{(i)}(\mathcal{C}_i)$ *be independent copies of the processes* $U(\mathcal{C}_i)$*,* $i=1,...,s$*. Then*

$$\Big|\mathbb{E}H(U_1,U_2,...,U_k)-\mathbb{E}H(U_1^{(j_1)},U_2^{(j_2)},...,U_k^{(j_k)})\Big|\le 4\|H\|_\infty\textstyle\sum_{i=2}^k\phi(m_i-n_{i-1})$$

where j_i *satisfies that* $i\in\mathcal{C}_{j_i}$*, for any* $1\le i\le k$*.*

Proof. Denote by ν_i the distribution of U_i, $i=1,..,k$. We first prove by induction on k that for any choice of H and U_i's with the required properties,

$$|\mathbb{E}H(U_1,U_2,...,U_k)-\int H(u_1,u_2,...,u_k)d\nu_1(u_1)d\nu_2(u_2)...d\nu_k(u_k)| \le 2\|H\|_\infty\sum_{i=2}^v\phi(m_i-n_{i-1}). \tag{3.4}$$

Indeed, suppose that $k=2$ and set $V_1=U_1$, $V_2=U_2$, $h(u_1)=E[H(u_1,U_2)]$, $\mathcal{G}_1=\mathcal{F}_{-\infty,n_1}$ and $\mathcal{G}_2=\mathcal{F}_{m_2,\infty}$. Taking expectation in (3.2) yields

$$|\mathbb{E}H(U_1,U_2)-\mathbb{E}h(U_1)|\le 2\|H\|_\infty\phi(m_2-n_1)$$

which means that (3.4) holds true when $k=2$. Now, suppose that (3.4) holds true for any $k\le j-1$, $U_1,...,U_k$ with the required properties and any bounded Borel function $H:\mathbb{R}^{e_1+..+e_{k-1}}\to\mathbb{R}$, where $e_1,...,e_{k-1}\in\mathbb{N}$. In order to deduce (3.4) for $k=j$, set $V_1=(U_1,...,U_{j-1})$, $V_2=U_j$, $h(v_1)=\mathbb{E}H(v_1,U_j)$, $v_1=(u_1,...,u_{j-1})$, $\mathcal{G}_1=\mathcal{F}_{-\infty,n_{j-1}}$ and $\mathcal{G}_2=\mathcal{F}_{m_j,\infty}$. Taking expectation in (3.2) yields

$$|\mathbb{E}H(U_1,U_2,...,U_j)-\mathbb{E}h(U_1,U_2,...,U_{j-1})|\le 2\|H\|_\infty\phi(m_j-n_{j-1}).$$

Applying the induction hypothesis with the function h completes the proof of (3.4), since $\|h\|_\infty\le\|H\|_\infty$. Next, we prove by induction on s that for any choice of k, H, U_i's with the required properties and $\mathcal{C}_1,...,\mathcal{C}_s$,

$$\begin{aligned}\Big|\textstyle\int H(u_1,u_2,...,u_k)d\mu_1(u^{(\mathcal{C}_1)})d\mu_2(u^{(\mathcal{C}_2)})...d\mu_s(u^{(\mathcal{C}_s)})-\\ \textstyle\int H(u_1,u_2,...,u_k)d\nu_1(u_1)d\nu_2(u_2)...d\nu_k(u_k)\Big|\le 2\|H\|_\infty\sum_{i=2}^k\phi(m_i-n_{i-1}).\end{aligned} \tag{3.5}$$

For $s=1$ this is just (3.4). Now suppose that (3.5) holds true for any $s\le j-1$, and any real valued bounded Borel function H defined on $\mathbb{R}^{d_1+...+d_k}$, where k and $d_1,...,d_k$ are some natural numbers. In order to prove (3.5) for $s=j$, set $u^{(I)}=(u^{(\mathcal{C}_1)},u^{(\mathcal{C}_2)},...,u^{(\mathcal{C}_{s-1})})$ and let the function I be defined by

$$I(u^{(I)})=\textstyle\int H(u_1,u_2,....,u_k)\prod_{j\in\mathcal{C}_s}d\nu_j(u_j). \tag{3.6}$$

Then

$$\textstyle\int H(u_1,u_2,...,u_k)d\nu_1(u_1)d\nu_2(u_2)...d\nu_k(u_k)=\int I(u^{(I)})\prod_{j\notin\mathcal{C}_s}d\nu_j(u_j). \tag{3.7}$$

Let the function J be defined by

$$J(u^{(I)}) = \int H(u_1, u_2, ..., u_k) d\mu_s(u^{(\mathcal{C}_s)}). \tag{3.8}$$

Then by (3.4), for any $u^{(\mathcal{C}_1)}, ..., u^{(\mathcal{C}_{s-1})}$,

$$|I(u^{(I)}) - J(u^{(I)})| \leq 2\|H\|_\infty \textstyle\sum_{i\in\mathcal{C}_s} \phi(m_i - n_{i-1}). \tag{3.9}$$

It is clear that $\|J\|_\infty \leq \|H\|_\infty$. Applying the induction hypothesis with the function J (considered as a function of the variable u) and taking into account (3.7) and (3.9) we obtain (3.5) with $s = j$. We have completed the induction. Inequality (3.3) follows by (3.4) and (3.5), and the proof of Corollary 3.3 is complete. □

Remark 3.4. In the notations of Corollary 3.3, let $Z_i, i = 1, ..., s$ be a bounded $\sigma\{U(\mathcal{C}_i)\}$-measurable random variable. Then each Z_i has the form $Z_i = H_i\big(U(\mathcal{C}_i)\big)$ for some function H_i which satisfies $\|H_i\|_\infty \leq \|Z_i\|_{L^\infty}$. By considering the function $H(u) = \prod_{i=1}^s H_i(u^{(\mathcal{C}_i)})$, we obtain from (3.3) that,

$$\big|\mathbb{E}[\textstyle\prod_{i=1}^s Z_i] - \prod_{i=1}^s \mathbb{E}Z_i\big| \leq 4\big(\prod_{i=1}^s \|Z_i\|_{L^\infty}\big) \sum_{j=2}^k \phi(m_j - n_{j-1}). \tag{3.10}$$

In general we can replace $\|H\|_\infty$ appearing in the right hand side of (3.3) by some essential supremum norm of H with respect to some measure which has a similar but more complicated form as κ defined in Lemma 3.2.

4 Nonconventional CLT with almost optimal convergence rate via Stein's method

First, the proof of Theorem 2.2 follows from arguments in [11], [10], and [7]. Indeed, relying on (2.25) in [11], the conditions of [7] and [10] hold true in our circumstances. Existence of D^2 follows from Theorem 2.2 in [11], inequality (2.8) follows from the arguments in [10] (first by considering the case when $M = 1$) and the condition for positivity follows from Theorem 2.3 in [7].

Before proving Theorem 2.3 we introduce the following notations. For any $a, b \in \mathbb{R}$ set

$$d_\ell(a, b) = \min_{1\leq i,j\leq \ell} |ia - jb|$$

and for any $A, B \subset \mathbb{R}$ set

$$dist(A, B) = \inf\{|a - b| : a \in A, b \in B\} \text{ and } d_\ell(A, B) = \inf\{d_\ell(a, b) : a \in A, b \in B\}.$$

Finally, for any $A_1, A_2, ..., A_L \subset \mathbb{R}$, we will write $A_1 < A_2 < ... < A_L$ if $a_1 < a_2 < ... < a_L$ for any $a_i \in A_i$, $i = 1, 2, ..., L$.

Proof of Theorem 2.3

Suppose that $D^2 > 0$. We consider first the self normalized case. Clearly, in the proof of (2.9) we can assume that $M = 1$. For any $N \geq 1$ set $s_N = \sqrt{\mathbb{E}(S_N)^2}$. Then by (2.8),

$$(s_N)^2 \geq D^2 N - N^{\frac{1}{2}} C_1. \tag{4.1}$$

Let N be so that $D^2 N > N^{\frac{1}{2}} C_1$. Then $s_N > 0$ and we set $Z_N = \frac{S_N}{s_N}$. Let l be of the form $l = 4A\ln(N + 1)$ where $A \geq 1$ is a positive constant considered here as a parameter which will be chosen later. Set $r = [\frac{l}{4}]$ and

$$S_{N,r} = \sum_{n=1}^{N} F(\xi_{n,r}, \xi_{2n,r}, ..., \xi_{\ell n,r})$$

where $\xi_{m,r}=\mathbb{E}[\xi_m|\mathcal{F}_{m-r,m+r}]$ for any $m\in\mathbb{N}$. Then by (2.5) and (2.7),

$$\|S_N-S_{N,r}\|_{L^\infty}\leq \ell N(\beta_\infty(r))^\kappa\leq d\ell c^{-1}Nc^{\frac{l}{4}}=d\ell c^{-1}Nc^{A\ln(N+1)}\leq c_0(N+1)^{1+A\ln c} \tag{4.2}$$

where $c_0=d\ell c^{-1}$ and we also used our assumption that $M=1$. Next, let $n>l$, consider the random vectors $U_i=\xi_{in,r}$ and set $m_i=in-r$ and $n_i=in+r$, $i=1,...,\ell$. Then each U_i is $\mathcal{F}_{in-r,in+r}$-measurable and $m_i-n_{i-1}=n-2r\geq l-2r\geq\frac{l}{2}$. Applying Corollary 3.3 with the sets $\mathcal{C}_i=\{in\},i=1,...,\ell$ we obtain

$$\begin{aligned}\big|\mathbb{E}F(\xi_{n,r},\xi_{2n,r}...,\xi_{\ell n,r})-\mathbb{E}F(\xi^{(1)}_{n,r},\xi^{(2)}_{2n,r},...,\xi^{(\ell)}_{\ell n,r})\big|&\\ \leq 4\ell\phi(\frac{l}{2})\leq 4\ell dc^{\frac{l}{2}}=c_1(N+1)^{2A\ln c}&\end{aligned}$$

where $c_1=4\ell d$ and $\xi^{(i)}_{in,r}$'s are independent copies of $\xi_{in,r}$'s. Considering the product measure of the laws of the vectors $(\xi_{in,r},\xi_{in}),i=1,...,\ell$, we can always assume that there (on a larger probability space) exist independent copies $\xi^{(i)}_{in}$ of the ξ_{in}'s such that $\|\xi^{(i)}_{in}-\xi^{(i)}_{in,r}\|_{L^\infty}\leq\beta_\infty(r)$ for any $i=1,2,...,\ell$. Thus by (2.5) and (2.7),

$$\begin{aligned}|\mathbb{E}F(\xi^{(1)}_{n,r},\xi^{(2)}_{2n,r},...,\xi^{(\ell)}_{\ell n,r})-\mathbb{E}F(\xi^{(1)}_{n},\xi^{(2)}_{2n},...,\xi^{(\ell)}_{\ell n})|&\\ \leq\ell\big(\beta_\infty(r)\big)^\kappa\leq\ell dc^{-1}c^{\frac{l}{4}}=c_0(N+1)^{A\ln c}&\end{aligned}$$

and notice that $\mathbb{E}F(\xi^{(1)}_{n},\xi^{(2)}_{2n},...,\xi^{(\ell)}_{\ell n})=\bar{F}=0$. We conclude from (2.4) and the above estimates that

$$\begin{aligned}|\mathbb{E}S_{N,r}|\leq|\mathbb{E}S_{l,r}|+N(4\ell dc^{\frac{l}{2}}+d\ell c^{-1}c^{\frac{l}{4}})&\\ \leq 2l+5N\ell dc^{-1}c^{\frac{l}{4}}\leq 8A\ln(N+1)+5c_0(N+1)^{1+A\ln c}.&\end{aligned} \tag{4.3}$$

We assume henceforth that $-A\ln c=A|\ln c|>2$ and set $\bar{S}_{N,r}=S_{N,r}-\mathbb{E}S_{N,r}$. For any two random variables X and Y defined on the same probability space we have $|\mathbb{E}X^2-\mathbb{E}Y^2|\leq\|X+Y\|_{L^2}\|X-Y\|_{L^2}$ and therefore by (4.2) and (2.8),

$$\begin{aligned}|\mathbb{E}S_N^2-\mathbb{E}S_{N,r}^2|\leq(2\|S_N\|_2+c_0(N+1)^{1+A\ln c})c_0(N+1)^{1+A\ln c}&\\ \leq 3c_0(2+c_0+C_1+D)(N+1)^{\frac{3}{2}+A\ln c}&\end{aligned}$$

where we also used that $A|\ln c|>1$. Next, by (4.3),

$$\begin{aligned}|\mathbb{E}S_{N,r}^2-\mathbb{E}\bar{S}_{N,r}^2|=|\mathbb{E}S_{N,r}^2-\mathrm{Var}S_{N,r}|=|\mathbb{E}S_{N,r}|^2&\\ \leq 32A^2\ln^2(N+1)+25c_0^2(N+1)^{2+2A\ln c}&\end{aligned}$$

and together with the previous inequality and our assumption that $A\ln c<-2$ we obtain that

$$|\mathbb{E}S_N^2-\mathbb{E}\bar{S}_{N,r}^2|\leq c_2\ln^2(N+1) \tag{4.4}$$

where $c_2=32A^2+25c_0^2+3c_0(2+c_0+C_1+D)$. Combining this wih (4.1), it follows that

$$\mathbb{E}\bar{S}_{N,r}^2\geq D^2N-N^{\frac{1}{2}}C_1-c_2\ln^2(N+1). \tag{4.5}$$

Let N be so large such that the right hand side in the previous inequality is positive. Then $\mathbb{E}\bar{S}_{N,r}^2>0$. Let $\bar{s}_{N,r}$ be its positive square root and set

$$\bar{Z}_{N,r}=\frac{\bar{S}_{N,r}}{\bar{s}_{N,r}}=\sum_{n=1}^{N}Y_n$$

where for each n,

$$Y_n = Y_{n,N,r} = \frac{F(\xi_{n,r}, \xi_{2n,r}, ..., \xi_{\ell n,r}) - \mathbb{E}F(\xi_{n,r}, \xi_{2n,r}, ..., \xi_{\ell n,r})}{\bar{s}_{N,r}}.$$

Observe now that

$$\|Z_N - \bar{Z}_{N,r}\|_{L^\infty} \le \|(s_N)^{-1}S_N - (\bar{s}_{N,r})^{-1}S_{N,r}\|_{L^\infty} + |(\bar{s}_{N,r})^{-1}\mathbb{E}S_{N,r}|$$
$$\le (s_N)^{-1}\|S_N - S_{N,r}\|_{L^\infty} + |(s_N)^{-1} - (\bar{s}_{N,r})^{-1}|\|S_{N,r}\|_{L^\infty} + |(\bar{s}_{N,r})^{-1}\mathbb{E}S_{N,r}|.$$

The inequality $|x^{-1} - y^{-1}| = |x^2 - y^2|(x^2y + y^2x)^{-1}$ holds true for any $x, y > 0$ yielding that

$$|(s_N)^{-1} - (\bar{s}_{N,r})^{-1}| \le \frac{c_2 \ln^2(N+1)}{(s_N + \bar{s}_{N,r})s_N\bar{s}_{N,r}} := e_1$$

where we used (4.4), and we conclude from (4.2), (4.3) and the above estimates that

$$\|Z_N - \bar{Z}_{N,r}\|_{L^\infty} \le (s_N)^{-1}c_0(N+1)^{1+A\ln c} + 2Ne_1 \\ +(\bar{s}_{N,r})^{-1}(8A\ln(N+1) + 5c_0(N+1)^{1+A\ln c}) \tag{4.6}$$

where we used that $\|S_{N,r}\|_{L^\infty} \le 2N$ (recall our assumption that $M = 1$). Next, using (4.1), (4.5) and that $\ln(N+1) \le N^{\frac{1}{2}}$ for any $N \ge 1$ we derive that $\min(s_N^2, \bar{s}_{N,r}^2) \ge \frac{1}{4}D^2N$ when $3N^{\frac{1}{2}}D^2 \ge 8(C_1 + c_2)$ and in this case

$$\|Z_N - \bar{Z}_{N,r}\|_{L^\infty} \le c_4 \max(D^{-1}, D^{-3})N^{-\frac{1}{2}}\ln^2(N+1) \tag{4.7}$$

where $c_4 = C_4(1 + c_0 + c_2 + A)$, $C_4 > 1$ is some absolute constant and we also used that $N^{1+A\ln c} < 1$.

Next, let N be sufficiently large so that $3N^{\frac{1}{2}}D^2 \ge 8(C_1 + c_2)$. Then by (2.4) and the above lower bound of $\bar{s}_{N,r}^2$,

$$\|Y_n\|_{L^\infty} \le 2(\bar{s}_{N,r})^{-1} \le 4D^{-1}N^{-\frac{1}{2}}. \tag{4.8}$$

For any $n = 1, 2, ..., N$ set

$$A_n = A_{n,l,N} = \{1 \le m \le N : \min_{1\le i,j\le \ell}|in - jm| \le l\} = \{1 \le m \le N : d_\ell(n,m) \le l\}$$

and for any $k \ge 0$ set

$$\mathcal{A}_n(k) = \{1 \le m \le N : d_\ell(A_n, A_m) = k\}.$$

We claim that there exist constants K_1 and K_2 which depend only on ℓ such that for any $n = 1, 2, ..., N$ and $k \ge 0$,

$$|A_n| \le K_1 l \ \text{ and } \ |\mathcal{A}_n(k)| \le K_2 l. \tag{4.9}$$

Indeed, since A_n is contained in a union of at most ℓ^2 intervals whose lengths do not exceed $2l+1$ we have $|A_n| \le \ell^2(2l+1)$ and since $l \ge 1$ we can take $K_1 = 3\ell^2$. To prove the second inequality, let n and m be such that $d_\ell(A_n, A_m) = k \ge 0$. Then there exist $1 \le i_s, j_s \le \ell$, $s = 1, 2, 3$ and $1 \le u, v \le N$ such that $|i_3u - j_3v| = k$, $|i_1n - j_1u| \le l$ and $|i_2m - j_2v| \le l$. When $j_3v - i_3u_3 = k$, we deduce from the last two inequalities that

$$\big|m - \frac{j_2i_3i_1}{i_2j_3j_1}n - \frac{j_2}{j_3i_2}k\big| \le l\big(\frac{1}{i_2} + \frac{j_2i_3}{i_2j_3j_1}\big)$$

and similar inequality holds when $j_3v - i_3u_3 = -k$. Thus, when n and k are fixed the set $\mathcal{A}_n(k)$ is contained in a union of $2\ell^6$ intervals whose lengths do not exceed $2(\ell^2+1)l$, and the choice of $K_2 = 4\ell^6 \cdot (\ell^2 + 2)$ is sufficient.

Now, set $\delta=\delta_{l,N}=\sum_{n=1}^N\sum_{m\in A_n}\mathbb{E}Y_nY_m$. Then

$$1=\mathrm{Var}\bar{Z}_{N,r}=\mathbb{E}\big(\sum_{n=1}^N Y_n\big)^2=\delta+\gamma \tag{4.10}$$

where $\gamma=\gamma_{l,N}=\sum_{n=1}^N\sum_{m\in\{1,...,N\}\setminus A_n}\mathbb{E}Y_nY_m$. Let $1\leq n,m\leq N$ be such that $m\notin A_n$. Consider the sets of indexes $\Gamma_k=\{jn:1\leq j\leq\ell\}$ where $k=n,m$ and set $\Gamma_{n,m}=\Gamma_n\cup\Gamma_m$. By the definition of the set A_n we have $dist(\Gamma_n,\Gamma_m)=d_\ell(n,m)>l$. Therefore, the set $\Gamma_{n,m}$ can be represented in the form

$$\Gamma_{n,m}=\bigcup_{t=1}^L B_t,\ B_1<B_2<...<B_L$$

where $L\leq 2\ell$, each B_t is either a subset of Γ_n or of Γ_m and $dist(B_t,B_{t-1})>l$. Set

$$U_t=\{\xi_{s,r}:s\in B_t\},\ t=1,...,L.$$

Since $r\leq\frac{l}{4}$, there exist numbers n_t and m_t, $t=1,...,L$ such that $n_{t-1}<m_t\leq n_t<m_{t+1}+\frac{l}{2}$, where $n_0=-\infty, m_{L+1}:=\infty$, and each U_t is measurable with respect to $\mathcal{F}_{m_t,n_t}$. Set $\mathcal{C}_1=\{1\leq t\leq L:B_t\subset\Gamma_n\}$ and $\mathcal{C}_2=\{1\leq t\leq L:B_t\subset\Gamma_m\}$. Then $\{\mathcal{C}_1,\mathcal{C}_2\}$ is a partition of $\{1,2,...,L\}$, Y_n is measurable with respect to $\sigma\{U_t:t\in\mathcal{C}_1\}$ and Y_m is measurable with respect to $\sigma\{U_t:t\in\mathcal{C}_2\}$. Therefore, by (3.10) and (4.8) and since $\mathbb{E}Y_n=0$,

$$|\mathbb{E}Y_nY_m|\leq 64\ell N^{-1}D^{-2}\phi(\frac{l}{2})\leq 64d\ell D^{-2}N^{-1}c^{\frac{l}{2}}\leq 64d\ell D^{-2}N^{2A\ln c-1}$$

implying that

$$|\gamma|=|\delta-1|=|\delta-\mathrm{Var}\bar{Z}_{N,r}|\leq 64d\ell D^{-2}N^{1+2A\ln c}. \tag{4.11}$$

We assume now, in addition to the previous restriction on N, that $64d\ell D^{-2}N^{-\frac{1}{2}}<\frac{1}{2}$. Then $\delta>\frac{1}{2}$ and so we can set $\sigma=\sqrt{\delta}$ and $W=\frac{\bar{Z}_{N,r}}{\sigma}$. Then $\sigma^2\geq\frac{1}{2}$ and using (4.11) we obtain

$$\|W-\bar{Z}_{N,r}\|_{L^\infty}\leq\|\bar{Z}_{N,r}\|_{L^\infty}|1-\frac{1}{\sigma}|\leq 4\|\bar{Z}_{N,r}\|_{L^\infty}|\delta-1|\leq 16D^{-3}N^{\frac{3}{2}+2A\ln c} \tag{4.12}$$

where we also used that $\bar{s}_{N,r}\geq\frac{1}{2}DN^{\frac{1}{2}}$. Since $A|\ln c|>1$ the above right hand side does not exceed $16D^{-3}N^{-\frac{1}{2}}$ which together with (4.7) and Lemma 3.1 yields that

$$d_K(\mathcal{L}(Z_N),\mathcal{N}(0,1))\leq 3d_K(\mathcal{L}(W),\mathcal{N}(0,1))+c_5\max(D^{-1},D^{-3})N^{-\frac{1}{2}}\ln^2(N+1) \tag{4.13}$$

where $c_5=16c_4$.

In order to approximate $d_K(\mathcal{L}(W),\mathcal{N}(0,1))$, set $X_n=\sigma^{-1}Y_n$, $n=1,2,...,N$. Then $W=\sum_{n=1}^N X_n$ and by (4.8) we have $\|X_n\|_{L^\infty}\leq R$, where $R=4N^{-\frac{1}{2}}D^{-1}\sigma^{-1}\leq 8N^{-\frac{1}{2}}D^{-1}$. Applying Theorem 2.1 in [13], using the equality (15) from there and taking into account (4.9) we obtain that

$$d_K(\mathcal{L}(W),\mathcal{N}(0,1))\leq R_1+R_2+R_3+K_1lR+2K_1^2l^2NR^3$$

where

$$R_1=4\|\textstyle\sum_{n=1}^N\sum_{m\in A_n}(X_nX_m-\mathbb{E}X_nX_m)\|_2,$$
$$R_2=\sqrt{2\pi}\textstyle\sum_{n=1}^N\mathbb{E}\big|\mathbb{E}[X_n|X_m:m\notin A_n]\big|,$$
$$R_3=2\|\textstyle\sum_{n=1}^N X_n\big(\sum_{m\in A_n}X_m\big)^2\|_2\big(\|W\|_2+5\big)$$

and $\|X\|_q^q = \mathbb{E}|X|^q = \|X\|_{L^q}^q$ for any random variable X. Now we estimate R_1, R_2 and R_3. Set $T_n = \sum_{m\in A_n}(X_nX_m - \mathbb{E}X_nX_m)$, $n = 1, ..., N$. Then

$$R_1^2 = 16\sum_{n_1,n_2=1}^{N}\mathbb{E}T_{n_1}T_{n_2}.$$

Let n_1 and n_2 be such that $d_\ell(A_{n_1}, A_{n_2}) = k > 2r$. Consider the sets $\Gamma_s = \{jm : m \in A_{n_s}, 1 \le j \le \ell\}$, $s = 1, 2$. Then $dist(\Gamma_1, \Gamma_2) = d_\ell(A_{n_1}, A_{n_2}) = k$. Set $\Gamma = \Gamma_1 \cup \Gamma_2$. Both Γ_i's are unions of at most ℓ^3 integer intervals (over the integers), and therefore there exist sets $B_1, B_2, ..., B_L$, $L \le 2\ell^3$ such that

$$\Gamma = \bigcup_{t=1}^{L} B_t, \;\; B_1 < B_2 < ... < B_L$$

where each B_t is either a subset of Γ_1 or a subset of Γ_2 and $dist(B_t, B_{t-1}) \ge k$, $t = 2, ..., L$. Set

$$U_t = \{\xi_{a,r} : a \in B_t\}, \;\; t = 1, ..., L.$$

Then there exist numbers $m_t, n_t, t = 1, 2, ..., L$ such that $n_{t-1} < m_t \le n_t \le m_{t+1} + k - 2r$, $n_0 = -\infty, m_{L+1} := \infty$ and each U_t is $\mathcal{F}_{m_t,n_t}$-measurable. Set $\mathcal{C}_s = \{1 \le t \le L : B_t \subset \Gamma_s\}$, $s = 1, 2$. Then $\{\mathcal{C}_1, \mathcal{C}_2\}$ is a partition of $\{1, 2, ..., L\}$ and T_{n_s}, $s = 1, 2$ is measurable with respect to $\sigma\{U_t : t \in \mathcal{C}_s\}$. Since $\|X_n\|_{L^\infty} \le R$ we have $\|T_n\|_{L^\infty} \le 2K_1lR^2$ (recall (4.9)) and thus by (3.10),

$$|\mathbb{E}T_{n_1}T_{n_2}| \le 16K_1^2l^2R^4L^2\phi(k-2r) \le 64\ell^6K_1^2l^2R^4dc^{k-2r}$$

where we used that $\mathbb{E}T_n = 0$. Given n_1 and $k > 2r$, the number of n_2's satisfying $d_\ell(A_{n_1}, A_{n_2}) = k$ is at most K_2l (recall (4.9)), while for any other n_2 and k we can use the trivial upper bound $|\mathbb{E}T_{n_1}T_{n_2}| \le \|T_{n_1}\|_{L^\infty}\|T_{n_2}\|_{L^\infty} \le 4K_1^2l^2R^4$. Therefore, by the definitions of R and r,

$$R_1^2 \le 64\ell^4K_1^2l^2R^4N\big(K_2ld\sum_{k=2r+1}^{N}c^{k-2r} + (2r+1)K_2l\big) \le C_0l^4N^{-1}D^{-4}$$

where C_0 is a constant which depends only on c and d and ℓ. In order to approximate R_2, let $1 \le n \le N$ and set $\mathscr{X}_n = \{X_m : m \notin A_n\}$. Then,

$$\|\mathbb{E}[X_n|\mathscr{X}_n]\|_1^2 \le \|\mathbb{E}[X_n|\mathscr{X}_n]\|_2^2 = \mathbb{E}X_n\mathbb{E}[X_n|\mathscr{X}_n]. \tag{4.14}$$

Consider the sets $\tau_1 = \{n, 2n, ..., \ell n\}$ and $\tau_2 = \{jm : m \notin A_n, 1 \le j \le \ell\}$. Then by the definition of A_n we have $dist(\tau_1, \tau_2) > l$. Thus, the union $\tau_1 \cup \tau_2$ can be written as a union of at most $2\ell + 1$ disjoint sets $B_1, B_2, ..., B_L$ such that $B_1 < B_2 < ... < B_L$, $dist(B_t, B_{t+1}) > l$ and each B_t is either a subset of τ_1 of a subset of τ_2. Consider the random vectors

$$U_t = \{\xi_{s,r} : s \in B_t\}, \; t = 1, ..., L$$

and the partition of $\{1, 2, ..., L\}$ into the sets $\{\mathcal{C}_1, \mathcal{C}_2\}$, where $\mathcal{C}_s = \{1 \le t \le L : B_t \subset \tau_s\}$, $s = 1, 2$. Then X_n is measurable with respect to $\sigma\{U_t : t \in \mathcal{C}_1\}$ and $\mathbb{E}[X_n|\mathscr{X}_n]$ is measurable with respect to $\sigma\{U_t : t \in \mathcal{C}_2\}$. Therefore by (3.10) and (2.7),

$$|\mathbb{E}[X_n\mathbb{E}[X_n|\mathscr{X}_n]]| \le 4(2\ell+1)R^2dc^{\frac{l}{2}} \tag{4.15}$$

where we used that $r \le \frac{l}{4}$, $\mathbb{E}X_n = 0$ and that $\|\mathbb{E}[X_n|\mathscr{X}_n]\|_{L^\infty} \le \|X_n\|_{L^\infty} \le R$. We conclude from (4.14) and (4.15) that there exists a constant C_0' which depends only on ℓ such that

$$R_2 \le C_0'N^{\frac{1}{2}}D^{-1}d^{\frac{1}{2}}c^{\frac{l}{4}}. \tag{4.16}$$

To estimate R_3, first observe that by the definition of W and by (4.10) we have $\|W\|_2^2 = \delta^{-1}\|\bar{Z}_{N,r}\|_2^2 = 1+\delta^{-1}\gamma$ and therefore $\|W\|_2 \le 2$, since $|\gamma| < \frac{1}{2}$ and $\delta > \frac{1}{2}$. The first factor in the definition of R_3 is clearly bounded from above by $2NK_1^2 l^2 R^3$ and we conclude that

$$R_3 \le C_4 l^2 D^{-3} N^{-\frac{1}{2}}$$

for some constant C_4 which depends only on ℓ. The estimate (2.9) in Theorem 2.3 follows now by taking any $A > \max(1, 2|\ln c|^{-1})$, using (4.13) and the above estimates of R_i's. Note that all the approximations in this section hold true only for N's satisfying $3N \ge 8D^{-2}(C_1+c_2)$ and $64d\ell D^{-2}N^{-\frac{1}{2}} < \frac{1}{2}$, but inequality (2.9) follows for all other N's from the basic estimate $d_K(\mathcal{L}(Z_N), \mathcal{N}(0,1)) \le 1$. We also remark that when $\beta_\infty(r_0) = 0$ for some r_0 then taking $r \ge r_0$ we get $S_{N,r} = S_N$ and so there is no need for (2.5) to hold true.

Now we derive (2.10) where again it is sufficient to consider the case when $M = 1$. Let $0 < \epsilon < \frac{1}{4}$. First for any $b > 1$,

$$\begin{aligned}\|D^{-1}N^{-\frac{1}{2}}S_N - Z_N\|_{L^b} &= \|S_N\|_{L^b}|N^{-\frac{1}{2}}D^{-1} - (s_N)^{-1}| \\ &= \|S_N\|_{L^b}\Big|\frac{\mathbb{E}S_N^2 - D^2N}{D^2Ns_N + D(s_N)^2N^{\frac{1}{2}}}\Big|\end{aligned}$$

where in the second equality we used that $|x^{-1}-y^{-1}| = |x^2-y^2|(xy^2+yx^2)^{-1}$ for any $x, y > 0$. By Lemma 5.2 in [7] for any $b > 1$ there exits a constant M_b which depends only on c, d, b and ℓ so that $\|S_N\|_{L^b} \le M_b N^{\frac{1}{2}}$. Using the previous estimates, for any N so that $3N^{\frac{1}{2}}D^2 \ge 8(C_1+c_2)$ and $64d\ell D^{-2}N^{-\frac{1}{2}} < \frac{1}{2}$ we have $s_N \ge \frac{1}{2}D$. Therefore,

$$\|D^{-1}N^{-\frac{1}{2}}S_N - Z_N\|_{L^b} \le 8D^{-3}C_1M_bN^{-\frac{1}{2}}$$

where and we also used (2.8). Applying the second statement of Lemma 3.1 with $b = \frac{1}{2\epsilon} - 1$ and using (2.9) completes the proof of (2.10). □

4.1 Extensions and remarks

Unbounded functions

Let $M, \iota > 0$, $\kappa \in (0,1]$ and $F : (\mathbb{R}^\wp)^\ell \to \mathbb{R}$ be a function satisfying

$$\begin{aligned} |F(x)| &\le M(1+\textstyle\sum_{i=1}^\ell |x_i|^\iota) \text{ and} \\ |F(x)-F(y)| &\le M(1+\textstyle\sum_{i=1}^\ell |x_i|^\iota + |y_i|^\iota)\sum_{i=1}^\ell |x_i-y_i|^\kappa \end{aligned} \tag{4.17}$$

for any $x = (x_1, ..., x_\ell)$ and $y = (y_1, ..., y_\ell)$ in $(\mathbb{R}^\wp)^\ell$. For any $R > 0$ set $F_R(x) = F(x)\mathbb{I}(|F(x)| \le R)$. Then, assuming that for some $p > \iota + 1$,

$$\gamma_p = \|\xi_1\|_{L^{\iota p}} < \infty$$

we can first approximate $F(\xi_n, \xi_{2n}, ..., \xi_{\ell n})$ by $F_R(\xi_{n,r}, \xi_{2n,r}..., \xi_{\ell n,r})$ in the L^p-norm and then use Lemma 3.3 3.1. Applying Theorem 2.3 with the function F_R and taking R with an appropriate dependence on N we obtain convergence rate of the form $CN^{-\frac{1}{2}+\varepsilon_p}$, where ε_p depends on p and satisfies $\lim_{p\to\infty}\varepsilon_p = 0$. In fact, similar type of rates can be obtained assuming only that $\phi(n)+\beta_q(n) \le dn^{-\theta}$ for some $q, d, \theta > 0$, where β_q is defined similarly to β_∞, but with the L^q norm.

Nonlinear indexes

Let $q_i, i = 1, ..., \ell$ be strictly increasing functions satisfying $q_i(\mathbb{N}) \subset \mathbb{N}$ which are ordered so that

$$q_1(n) < q_2(n) < ... < q_\ell(n) \text{ for any sufficiently large } n.$$

Consider the sums

$$S_N = \sum_{n=1}^{N} F(\xi_{q_1(n)}, \xi_{q_2(n)}, ..., \xi_{q_\ell(n)}).$$

The proof of Theorem 2.3 proceeds essentially in the same way when all q_i's are linear. For more general q_i's, set

$$A_n = \{1 \le m \le N : \min_{1\le i,j\le \ell} |q_i(n) - q_j(m)| \le l\}.$$

The proof of Theorem 2.3 will proceed similarly for the sums S_N if we show that the limit $D^2 = \lim_{N\to\infty} N^{-\frac{1}{2}} \mathbb{E} S_N^2$ exists, obtain convergence rate towards it and upper bounds similar to the ones in (4.9). Suppose that $q_1, ..., q_k$ are linear, for some $k < \ell$ and that $q_j, j \ge k$ are not. When all q_i's are polynomials, existence of D^2 is proved in [8]. Though the limit D^2 does not exist in general, if q_{j+1} grows faster then q_j for $j > k$ in the sense of (2.11) in [11], then existence of D^2 follows from Theorem 2.3 in [11]. Convergence rate towards D^2 when q_i's are polynomials can be obtained by proceeding similarly to the proof of Proposition 5.3 in [8]. If, instead, $q_{j+1}(n^\alpha) - q_j(n)$ converges to ∞ as $n \to \infty$ for some $0 < \alpha < 1$ and all $j \ge k$, then convergence rate towards D^2 with some dependence on α follows from the arguments in [11].

Each $q_i(n)$ grows at least as fast as linearly which implies that $|A_n|$ is of order l. When all q_i's are polynomials of the same degree then the limit $\lim_{n\to\infty} q_i^{-1}(q_j(n))/n$ exists for any $1 \le i, j \le \ell$ and therefore the proof of the second upper bound in (4.9) proceeds in a similar way but with $\tilde{d}_\ell(a,b) = \min_{1\le i,j\le \ell} |q_i(a) - q_j(b)|$ in place of $d_\ell(a,b)$. When q_i's do not necessarily have the same degree then beginning the summation in the definition of S_N from cN^γ for appropriate $\gamma < 1$ and $c > 0$, guarantees that $|q_i(n) - q_j(m)| > CN$ when $\deg q_i \ne \deg q_j$ and $cN^\gamma \le n, m \le N$. Similar to the latter inequality is satisfied when $\max(i,j) > k$ and q_s grows faster than q_{s-1} for $s = k+1, ..., \ell$ and so an appropriate version of (4.9) follows in this situation, as well.

Remark 4.1. Using arguments similar to the ones in [3], we can obtain (2.9) with $N^{-\frac{1}{2}} S_N$ in place of Z_N when F satisfies (4.17), and in particular when F is a bounded Hölder continuous function. In fact, Stein's method also yields (under appropriate conditions) the nonconventional functional CLT, which was proved in [11] using martingale approximation. These results require a relatively long presentation and their proofs are not short, so they will appear elsewhere.

References

[1] E. Bolthausen, *Exact convergence rates in some martingale central limit theorems*, Ann. Probab. 10 (1982), 672-688. MR-0659537

[2] R.C. Bradley, *Introduction to Strong Mixing Conditions*, Volume 1, Kendrick Press, Heber City, 2007. MR-2325295

[3] L.H.Y Chen and Q.M. Shao, *Normal approximation under local dependence*, Ann.Probab. 32 (2004), 1985-2028. MR-2073183

[4] H. Furstenberg, *Nonconventional ergodic averages*, Proc. Symp. Pure Math. 50 (1990), 43-56. MR-1067751

[5] M.I. Gordin, *On the central limit theorem for stationary processes*, Soviet Math. Dokl. 10 (1969), 1174-1176. MR-0251785

[6] P. Hall, C.C. Heyde, *Rates of convergence in the martingale central limit theorem*, Ann. Probab. 9 (1981) 395-404. MR-0614625

[7] Y. Hafouta and Yu. Kifer, *Berry-Esseen type estimates for nonconventional sums*, Stoch. Proc. Appl. 126 (2016), 2430-2464. MR-3505233

[8] Y. Hafouta and Yu. Kifer, *Nonconventional polynomial CLT*, Stochastics, 89 (2017), 550-591. MR-3590435

[9] Yu. Kifer, *Nonconventional limit theorems* , Probab. Th. Rel. Fields, 148 (2010), 71-106. MR-2653222

[10] Yu. Kifer, *Strong approximations for nonconventional sums and almost sure limit theorems*, Stochastic Process. Appl., 123 (2013), 2286-2302. MR-3038505

[11] Yu. Kifer and S.R.S Varadhan, *Nonconventional limit theorems in discrete and continuous time via martingales*, Ann. Probab., 42 (2014), 649-688. MR-3178470

[12] D.L. McLeish, *Invariance principles for dependent variables*, Z.Wahrscheinlichkeitstheorie und Verw. Gebiete 32 (1975), 165-178. MR-0388483

[13] Y. Rinott, *On normal approximation rates for certain sums of dependent random variables*, J. Comput. Appl. Math., 55 (1994), 135-143. MR-1327369

[14] W. Rudin *Real and Complex Analysis*, McGraw-Hill, New York, 1987. MR-0924157

[15] C. Stein, *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proc. Sixth Berkeley Symp. Math. Statist. Probab, 2 (1972), 583-602. Univ. California Press, Berkeley. MR-0402873

[16] C. Stein, *Approximation Computation of Expectations*, IMS, Hayward, CA (1986).

[17] N. Shiryaev, *Probability*, Springer-Verlag, Berlin, 1995. MR-0737192

Acknowledgments. This paper is a part of the author’s PhD thesis conducted at the Hebrew university of Jerusalem. I would like to thank my advisor Professor Yuri Kifer for suggesting to me the problem studied in this paper and for many helpful discussions.