Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2020, Vol. 56, No. 1, 428–448
https://doi.org/10.1214/19-AIHP967
© Association des Publications de l'Institut Henri Poincaré, 2020

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Nonconventional moderate deviations theorems and exponential concentration inequalities

Yeor Hafouta

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

Received 19 July 2018; revised 22 January 2019; accepted 8 February 2019

Abstract. We obtain moderate deviations theorems and exponential (Bernstein type) concentration inequalities for "nonconventional" sums of the form $S_N = \sum_{n=1}^{N}(F(\xi_{q_1(n)}, \xi_{q_2(n)}, \dots, \xi_{q_\ell(n)}) - \bar{F})$, where most of the time we consider $q_i(n) = in$, but our results also hold true for more general $q_i(n)$'s such as polynomials. Here ξ_n, $n \geq 0$ is a sufficiently fast mixing vector process with some stationarity conditions, F is a function satisfying certain regularity conditions and $\bar{F}$ is a certain centralizing constant. When $\xi_n, n \geq 0$ are independent and identically distributed a large deviations theorem was obtained in (*Probab. Theory Related Fields* **158** (2014) 197–224) and one of the purposes of this paper is to obtain related results in the (weakly) dependent case. Several normal approximation type results will also be derived. In particular, two more proofs of the nonconventional central limit theorem are given and a Rosenthal type inequality is obtained. 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 Gibbs–Markov map, a hyperbolic diffeomorphism, a Young tower or an expanding transformation taken with a Gibbs invariant measure, as well as in the case when ξ_n, $n \geq 0$ forms a stationary and (stretched) 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.

Résumé. Nous obtenons un théorème de déviations modérées et des inégalités de concentration exponentielles (du type de Bernstein) pour des sommes «non-conventionnelles» de la forme $S_N = \sum_{n=1}^{N}(F(\xi_{q_1(n)}, \xi_{q_2(n)}, \dots, \xi_{q_\ell(n)}) - \bar{F})$, où la plupart du temps nous considérons $q_i(n) = in$, mais nos résultats restent aussi vrais pour des $q_i(n)$ plus généraux tels que des polynômes. Ici, ξ_n, $n \geq 0$ est un processus vectoriel suffisamment mélangeant avec des conditions de stationnarité, F est une fonction satisfaisant certaines propriétés de régularité et $\bar{F}$ est une constante de centrage. Quand $\xi_n, n \geq 0$ sont indépendants et identiquement distribués, un principe de grande déviation a été obtenu dans (*Probab. Theory Related Fields* **158** (2014) 197–224) et un des objectifs de cet article est d'obtenir des résultats analogues dans le cas faiblement dépendant. Plusieurs résultats de type approximation normale sont aussi obtenus. En particulier, deux nouvelles preuves du théorème central limite non-conventionnel sont données et une inégalité de type Rosenthal est obtenue. Nos résultats sont vrais par exemple quand $\xi_n = (T^n f_i)_{i=1}^{\wp}$ où T est un sous-shift de type fini topologiquement mélangeant, une application Gibbs–Markov, un difféomorphisme hyperbolique, une tour de Young ou une transformation expansive pour une mesure invariante de Gibbs, tout comme dans le cas où $\xi_n, n \geq 0$ forme une suite stationnaire exponentiellement (ou streched exponentiellement) ϕ-mélangeante, ce qui, par exemple, et vrai lorsque $\xi_n = (f_i(\Upsilon_n))_{i=1}^{\wp}$ où Υ_n est une chaîne de Markov satisfaisant une condition de Doeblin, considérée comme un processus stationnaire par rapport à une mesure invariante.

MSC: Primary 60F10; secondary 60F05; 37D20; 37D25; 37A25

Keywords: Nonconventional setup; Mixing; Large deviations; Moderate deviations; Exponential concentration inequalities; The method of cumulants; Martingale approximation; Central limit theorem; Berry–Esseen theorem

1. Introduction

Partially motivated by the research on nonconventional ergodic theorems (the term "nonconventional" comes from [13]), probabilistic limit theorems for sums of the form $S_N = \sum_{n=1}^{N}(F(\xi_{q_1(n)}, \xi_{q_2(n)}, \dots, \xi_{q_\ell(n)}) - \bar{F})$ have become a well studied topic. Here ξ_n, $n \geq 0$ is a sufficiently fast mixing vector process with some stationarity properties, F is a function satisfying some regularity conditions and $\bar{F}$ is a certain centralizing constant. During the past decade many of the classical results such as the (functional) central limit theorem, Berry–Esseen type theorem, the local central limit theorem, Poissonian limit theorems and large deviations theorems were obtained for such sums (see [19,23,26,27] and references therein). One of the most interesting choices of q_i's is the situation when $q_i(n) = in$ for any $i = 1, 2, \dots, \ell$. This was the original motivation for the study of nonconventional sums and yields appropriate limit theorems for number of multiple recurrencies to a given set by ξ_k's at times forming arithmetic progressions of the type $n, 2n, \dots, \ell n$.

The large deviation priciple proved in [27] holds true in the case when $S_N = \sum_{n=1}^{N}(F(\xi_n, \xi_{2n}, \dots, \xi_{\ell n}) - \bar{F})$ only for independent and identically distributed ξ_n's, while when the $q_i(n)$'s satisfy certain (faster than linear) growth conditions the results from there hold true also for certain Markov chains and dynamical systems. The main goal of this paper is to obtain related results when the ξ_n's are weakly dependent and not necessarily generated by a Markov chain or a dynamical system. We will first obtain moderate deviation type theorems for such sums, namely, study the asymptotic behaviour as $N \to \infty$ of probabilities of the form

$$P\left(\frac{1}{N^{\zeta}} S_N \in \Gamma\right)$$

for arbitrary Borel measurable sets $\Gamma \subset \mathbb{R}$. Here $\frac{1}{2} < \zeta < 1$ depends on the amount of regularity of F and on the growth of $\mathbb{E}|\xi_1|^k$ as $k \to \infty$. Formally (see [9]), any choice of ζ is considered as large deviations type result, but under our conditions $\frac{1}{N} S_N$ will satisfy the law of large numbers (see [24]) and so we will use the standard informal convention of referring to the case when $\zeta = 1$ as the large deviations case, while the case when $0 < \zeta < 1$ will be referred to as the moderate deviations case, where in our situation it is natural to require that $\frac{1}{2} < \zeta$ since $N^{-\frac{1}{2}} S_N$ satisfies the central limit theorem (see [26] and [18]). Exponential concentration inequalities (i.e. estimates of $P(S_N \ge x)$, $x > 0$) and Gaussian type estimates of the moments of S_N will also be derived. All of the above results are obtained using the so-called method of cumulants (see [30]) and the local dependence structure of nonconventional sums introduced in [19]. The best exponential inequality obtained by this method yields estimates of the form

$$P(S_N \ge \varepsilon N) \le e^{-c(\varepsilon N)^{\frac{1}{2}}}, \quad \varepsilon > 0, N \ge c\varepsilon^{-\frac{5}{2}}$$

where $c > 0$ is some constant. Such estimates are not optimal since the power of N is $\frac{1}{2}$ and not 1. In the case when F is bounded we are able to improve these estimates. We first approximate S_N in the L^∞ norm by martingales with bounded differences and then apply the Hoeffding–Azuma inequality in order to obtain, in particular, estimates of the form

$$P(S_N \ge \varepsilon N) \le e^{-c(\varepsilon)N}, \quad \varepsilon > 0, N \ge 1$$

where $c(\varepsilon) > 0$ is some constant which depends on ε but not on N. In the case when either $\xi_n, n \ge 0$ forms a sufficiently fast ϕ-mixing process or it is generated by a topologically mixing subshift of finite type or a Young tower with exponential tails we can choose $c(\varepsilon) = c\varepsilon^2$ for some $c > 0$ which does not depend on ε and N. Note that all the results described above hold true also with $\bar{S}_N = S_N - \mathbb{E}S_N$ in place of S_N.

Our results hold true, for instance, when $\xi_n = T^n f$ where $f = (f_1, \dots, f_d)$, T is a topologically mixing subshift of finite type, a hyperbolic diffeomorphism (see [2]), a Young tower (see [31] and [32]) with sufficiently fast cecaying tails, a Gibbs–Markov map considered in [1] or an expanding transformation taken with a Gibbs invariant measure, as well as in the case when $\xi_n = f(\Upsilon_n)$, $f = (f_1, \dots, f_d)$ 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 $\{\xi_n\}$ 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 $\xi_n = ((\xi_n)_1, \dots, (\xi_n)_\ell)$, $(\xi_n)_j = \mathbb{1}_{A_j}(T^n x)$ in the dynamical systems case and $(\xi_n)_j = \mathbb{1}_{A_j}(\Upsilon_n)$ in the Markov chain case where $\mathbb{1}_A$ is the indicator of a set A. Let $F = F(x_1, \dots, x_\ell)$, $x_j = (x_j^{(1)}, \dots, x_j^{(\ell)})$ be a bounded Hölder continuous function which identifies with the function $G(x_1, \dots, x_\ell) = x_1^{(1)} \cdot x_2^{(2)} \cdots x_\ell^{(\ell)}$ 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 $j = 0, 1, \dots, \ell$ (or $\Upsilon_{q_j(l)} \in A_j$ in the Markov chains case), where we set $q_0 = 0$, namely the number of ℓ-tuples of return times to A_j's (either by $T^{q_j(l)}$ or by $\Upsilon_{q_j(l)}$). Then our results yield moderate deviation theorems and exponential concentration inequalities for the numbers $N(n)$. In fact, in this case, and more generally for product functions of the form $F(x_1, \dots, x_\ell) = \prod_{i=1}^{\ell} g_i(x_i)$, our results hold true for (stretched) exponentially fast mixing α-mixing processes. When f_i's and g_i's are Hölder continuous then our results also hold true for the (deterministic) distance expanding maps considered in [28], even though there are no underlying Markov partitions.

In general, the sum S_N is a nonlinear function of the random vector $\{\xi_1, \xi_2, \dots, \xi_{q_\ell(N)}\}$, and therefore our results can also be viewed as a part of the research on nonlinear large deviations theorems (see [4] and [5]). Moreover, in view of the large variety of dynamical systems that can be considered, our results can be viewed as a part of the research on concentration of measure for dynamical systems (see, for instance, [6]), as well.

2. Preliminaries and main results

Our setup consists of a $\wp$-dimensional stochastic process ξ_n, $n \geq 0$ on a probability space $(\Omega, \mathcal{F}, P)$ and a family of sub-σ-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 on 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-σ-algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$,

$$\phi(\mathcal{G}, \mathcal{H}) = \sup\left\{\left|\frac{P(A \cap B)}{P(A)} - P(B)\right| : A \in \mathcal{G}, B \in \mathcal{H}, P(A) > 0\right\}. \tag{2.2}$$

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 restrictions on the approximation rates

$$\beta_q(r) = \sup_{k \geq 0} \left\| \xi_k - \mathbb{E}[\xi_k | \mathcal{F}_{k-r,k+r}] \right\|_q \tag{2.3}$$

where $\|X\|_q := \|X\|_{L^q}$ for any $0 < q \leq \infty$ and a random variable X.

We do not require stationarity of the process ξ_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

$$\xi_n \sim \mu \quad \text{and} \quad (\xi_n, \xi_m) \sim \mu_{m-n} \tag{2.4}$$

where $Y \sim \mu$ means that Y has μ for its distribution. In fact, some of our results hold true assuming only that $\xi_n \sim \mu$ for any $n \geq 0$, and we will point out when the assumption about the distribution of (ξ_n, ξ_m) is not needed.

Let $F = F(x_1, \dots, x_\ell)$, $x_j \in \mathbb{R}^\wp$ be a function on $(\mathbb{R}^\wp)^\ell$ such that for some $K \geq 1$, an integer $\lambda \geq 0$, $\kappa \in (0, 1]$ and all $x_i, z_i \in \mathbb{R}^\wp$, $i = 1, \dots, \ell$, we have

$$\left|F(x) - F(z)\right| \leq K\left[1 + \sum_{i=1}^{\ell}\left(|x_i|^\lambda + |z_i|^\lambda\right)\right]\sum_{i=1}^{\ell}|x_j - z_j|^\kappa \tag{2.5}$$

and

$$\left|F(x)\right| \leq K\left[1 + \sum_{i=1}^{\ell}|x_i|^\lambda\right] \tag{2.6}$$

where $x = (x_1, \dots, x_\ell)$ and $z = (z_1, \dots, z_\ell)$. In fact, if ξ_n is measurable with respect to $\mathcal{F}_{n,n}$ then our results will follow with any Borel function F satisfying (2.6) without imposing (2.5), since the latter is needed only for approximation of ξ_n by conditional expectations $\mathbb{E}[\xi_n | \mathcal{F}_{n-r,n+r}]$ using (2.3). To simplify formulas we assume the centering condition

$$\bar{F} := \int F(x_1, \dots, x_\ell)\, d\mu(x_1) \cdots d\mu(x_\ell) = 0 \tag{2.7}$$

which is not really a restriction since we can always replace F by $F - \bar{F}$. Let $\ell \geq 1$ be an integer, set

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

and $\bar{S}_N = S_N - \mathbb{E}S_N$. All the results presented here hold true in the situation when $q_i(n)$'s are polynomials with positive leading coefficients taking integer values on the integers, while some of the results hold true even for more general $q_i(n)$'s. This "nonlinear indexation" case requires some preparation, and so, for the sake of readability, we will discuss it only in Section 5.

We will obtain our main results under either

Assumption 2.1. $\lambda = 0$ (i.e. F is a bounded Hölder function) and there exist $a, d, \eta > 0$ so that

$$\phi(n) + \beta_\kappa^\kappa(n) \le d e^{-an^\eta}$$

for any $n \ge 1$,

or

Assumption 2.2. $\lambda > 0$ and there exist $d, a, \eta, M, \zeta > 0$ so that

$$\phi(n) + \beta_\infty^\kappa(n) \le d e^{-an^\eta}$$

for any $n \ge 1$, and for any $k \in \mathbb{N}$,

$$\tau_k^k = \mathbb{E}|\xi_1|^k = \int |x|^k \, d\mu(x) \le M^k (k!)^\zeta. \tag{2.8}$$

Note that under either Assumption 2.1 or Assumption 2.2 there exists a constant a_0 so that $|\mathbb{E}S_N| \le a_0 K$ for any $N \ge 1$. In fact, this estimate holds true under weaker conditions, see the paragraph proceeding Theorem 2.7.

Our first result is the following

Theorem 2.3.

(i) *Suppose that Assumption 2.1 holds true and set $\gamma = \frac{1}{\eta}$. Then there exist constants $c_1, c_2 > 0$ which depend only on K, ℓ, d, a, η and κ so that for any $x > 0$,*

$$P(\bar{S}_N \ge x) \le \exp\Big(-\frac{x^2}{2(c_1 + c_2 x N^{-\frac{1}{2+4\gamma}})^{\frac{1+2\gamma}{1+\gamma}}}\Big). \tag{2.9}$$

(ii) *When Assumption 2.2 holds true then (2.9) hold true with $\gamma = \frac{1}{\eta} + \lambda\zeta$ in place of $\frac{1}{\eta}$ and constants c_1 and c_2 which depend only on $K, \ell, d, a, \eta, M, \zeta, \kappa, \lambda$ and τ_λ.*

The above theorem holds true also for certain nonlinear $q_i(n)$'s such as polynomials and functions with exponential growth, see Section 5. Note that when $\beta_q(r_0) = 0$ for some q and r_0 then Theorem 2.3 holds true for any Borel function F satisfying (2.6), namely, there is no need in (2.5) or of any other type of continuity.

Next, by taking $x = \varepsilon N$, $\varepsilon > 0$ in (2.9) (or in the corresponding estimate under Assumption 2.2) and using that $|\mathbb{E}S_N| \le a_0 K$ we obtain that

$$\max\big(P(\bar{S}_N \ge \varepsilon N), P(S_N \ge \varepsilon N)\big) \le e^{-c_7 (\varepsilon N)^{\frac{1}{1+\gamma}}}, \quad N \ge c_6 \varepsilon^{-2-\frac{1}{\gamma}} \tag{2.10}$$

where c_6 and c_7 are positive constants which do not depend on N and a, and γ equals either $\frac{1}{\eta}$ or $\frac{1}{\eta} + \lambda\zeta$, depending on the case. The power of N in (2.10) is not optimal since it is smaller than 1. In order to obtain more accurate estimates on the tail probabilities we also prove the following

Theorem 2.4. *Suppose that F is a bounded Hölder continuous function and that*

$$\varphi := \sum_{n=0}^{\infty} \phi(n) < \infty.$$

Fix some $N \ge 1$ and $r \ge 0$ and set $\delta_1 := K(\varphi + r + 1)$ and $\delta_2 = KN\beta_\infty^\kappa(r) + \delta_1$. Then there exists a constant $B > 0$ which depends only on ℓ so that for any $\lambda > 0$,

$$\mathbb{E}e^{\lambda S_N} \le e^{B\lambda^2 N \ell \delta_1 + B\lambda\delta_2}. \tag{2.11}$$

When $\beta_\infty(r_0) = 0$ for some $r_0 \ge 0$ then the above results hold true with $r = r_0$ for any bounded Borel function F, i.e. there is no need in any kind of continuity.

Theorem 2.4 holds true also when $q_i(n)$'s are polynomials with positive leading coefficients taking integer values on the integers, see Section 5. Note that the above theorem does not require that $(\xi_n, \xi_m) \sim \mu_{m-n}$ since it does not involve the limit $D^2 = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}S_N^2$ (which does not necessarily exist without this assumption about the distribution of (ξ_n, ξ_m)). Most of the rest of the results obtained in this paper require that D^2 exists, and so we have assumed that $(\xi_n, \xi_m) \sim \mu_{m-n}$ at the beginning for the sake of convenience.

Next, using the Chernoff bounding method, in Section 4 we derive from (2.11) that for any $t > 0$,

$$P(S_N \geq t + B\delta_2) \leq e^{-\frac{t^2}{4B^2N\ell\delta_1^2}}. \tag{2.12}$$

When $\beta_\infty(r_0) = 0$ for some $r_0 \geq 0$ then by taking $r = r_0$ the terms δ_1 and δ_2 are constants, and therefore we obtain optimal exponential concentration inequalities of the form

$$P(S_N \geq \varepsilon N) \leq e^{-c\varepsilon^2 N}, \quad N \geq \frac{2B\delta_2}{\varepsilon}$$

where $c = \frac{\delta_2}{16\ell\delta_1^2} > 0$ and $\varepsilon > 0$. When $\beta_\infty(r)$ convergence to 0 as $r \to \infty$ then for any $\varepsilon > 0$ we can take a sufficiently large $r_0 = r_0(\varepsilon)$ and obtain that there exists a constant $c(\varepsilon) > 0$ so that for any $N \geq 1$ and $t > 0$,

$$P(S_N \geq t + 0.5\varepsilon N) \leq e^{-c(\varepsilon)\frac{t^2}{N}}$$

and in particular

$$P(S_N \geq \varepsilon N) \leq e^{-c_1(\varepsilon)N} \tag{2.13}$$

for some constant $c_1(\varepsilon) > 0$ which depends on ε but not on N. When some rate of decay of $\beta_\infty^\kappa(r)$ to 0 is known we can find an explicit $c(\varepsilon)$. For instance, when $\beta_\infty^\kappa(r) \leq de^{-ur}$, $d, u > 0$ for any $r \geq 0$, we can take r_0 of the form $r_0 = -c\ln\varepsilon$ and then the above estimate will hold true with $c(\varepsilon)$ having the form $c(\varepsilon) = q_0|\ln\varepsilon|^{-1}$ for some constant q_0 which depends only on ℓ, d, u, κ and K.

Remark 2.5. Let $(\mathcal{X}, T)$ be a Young tower (see [31] and [32]) and μ be an appropriate absolutely continuous invariant measure. Consider the σ-algebras $\mathcal{F}_{n,m} = \bigvee_{k=n}^m T^{-k}\mathcal{M}$ which are generated by T and the partition $\mathcal{M}$ defining the separation time on the tower. Then (see [21]), the mixing coefficients $\phi(n)$ decay in the same speed as the tails of the tower. Let $h_1, \dots, h_\wp$ be real valued functions on $\mathcal{X}$ which are either constant on atoms of the partition or are Hölder continuous functions and let $\xi_n = (h_1 \circ T^n, \dots, h_\wp \circ T^n)$, $n \geq 1$. Then, (2.13) holds true (with an appropriate $c(\varepsilon)$'s) assuming that the tails converge sufficiently fast to 0. Note that when $h_1, \dots, h_\wp$ are Hölder continuous functions, then the centralized sum $\bar{S}_N$ can be written as a reverse martingale, and therefore (see [7]), in these circumstances we obtain optimal exponential concentration inequality of the form

$$P(\bar{S}_N \geq t) \leq e^{-c\frac{t^2}{N}}, \quad N \geq 1, t > 0$$

where c is some constant. Plugging in $t = \varepsilon N$, $\varepsilon > 0$ we derive that for any $N \geq 1$,

$$P(\bar{S}_N \geq \varepsilon N) \leq e^{-c\varepsilon^2 N},$$

namely we can take $c(\varepsilon)$ of the form $c(\varepsilon) = c\varepsilon^2$ when S_N is replaced with $\bar{S}_N$.

Recall now (see [9]) that a sequence of probability measures μ_N, $N \geq 1$ on a topological space $\mathcal{X}$ is said to satisfy the large deviation principle (LDP) with speed $s_N \nearrow \infty$ and good rate function $I(\cdot)$ if I is lower semicontinuous, the sets $I^{-1}[0, \alpha]$, $\alpha \geq 0$ are compact and for any Borel measurable set $\Gamma \subset \mathcal{X}$,

$$\liminf_{N\to\infty} \frac{1}{s_N} \ln \mu_N(\Gamma) \geq -\inf_{x\in\Gamma^o} I(x)$$

and

$$\limsup_{N\to\infty} \frac{1}{s_N} \ln \mu_N(\Gamma) \leq -\inf_{x\in\bar{\Gamma}} I(x)$$

where Γ^o denotes the interior of a set Γ and $\bar{\Gamma}$ denotes its closure. A sequence of random variables $W_N, N \ge 1$ is said to satisfy the LDP with speed s_N and good rate function $I(\cdot)$ if the sequence $\mathcal{L}(W_N)$, $N \ge 1$ of the laws of the W_N's satisfies the appropriate LDP. We also recall the following terminological convention. When W_N, $N \ge 1$ satisfies the law of large numbers and s_N grows slower than linear in N the appropriate LDP is usually called a moderate deviation principle (MDP) and the case when $s_N = N$ is referred to as the LDP.

We will also prove the following

Theorem 2.6.

(i) *Suppose that Assumption 2.1 holds true and set $\gamma = \frac{1}{\eta}$. Set $v_N = \sqrt{\mathrm{Var}(S_N)}$ and when $v_N > 0$ also set $Z_N = \frac{\bar{S}_N}{v_N}$. Let Φ be the standard normal distribution function. Then the limit $D^2 = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}S_N^2$ exists and when $D^2 > 0$ there exist constants $c_3, c_4, c_5 > 0$ which depend only on ℓ, K, κ, a, d and η so that for any $N \ge c_3$ we have $v_N > 0$ and for any $0 \le x < c_4 N^{\frac{1}{2+4\gamma}}$,*

$$\left|\ln \frac{P(Z_N \ge x)}{1-\Phi(x)}\right| \le c_5\big(1+x^3\big)N^{-\frac{1}{2+4\gamma}} \quad \text{and}$$
$$\left|\ln \frac{P(Z_N \le -x)}{\Phi(-x)}\right| \le c_5\big(1+x^3\big)N^{-\frac{1}{2+4\gamma}}. \tag{2.14}$$

Moreover, let a_N, $N \ge 1$ be a sequence of real numbers so that

$$\lim_{N\to\infty} a_N = \infty \quad \text{and} \quad \lim_{N\to\infty} a_N N^{-\frac{1}{2+4\gamma}} = 0.$$

Then the sequence $(DN^{\frac{1}{2}}a_N)^{-1}S_N$, $N \ge 1$ satisfies the MDP with the speed $s_N = a_N^2$ and the rate function $I(x) = \frac{x^2}{2}$.

(ii) *When Assumption 2.2 holds true all the results stated above hold true with $\gamma = \frac{1}{\eta} + \lambda\zeta$ in place of $\frac{1}{\eta}$ and constants c_1, c_2 and c_3 which depend only on K, ℓ, d, a, η, M, ζ, κ, λ and τ_λ.*

Theorem 2.6 also holds true when $q_i(n)$'s are polynomials, or functions with certain exponential growth, see Section 5. When $\beta_q(r_0) = 0$ for some q and r_0 then all the results stated in Theorem 2.6 hold true for any Borel function F satisfying (2.6). We also remark that (2.14) is obtained using Lemma 2.3 in [30]. This lemma yields certain estimates close to the ones in (2.14), but for larger domain of x's. For the sake of readability these results are not stated here.

Theorems 2.3, 2.4 and 2.6 will follow from the following general results. The first one is

Theorem 2.7. *Suppose that for some $b > 2$ and $m > 0$,*

$$\frac{1}{b} \ge \frac{\lambda}{m} + 1, \qquad \max(\tau_m, \tau_{\lambda b}) < \infty \tag{2.15}$$

and

$$\Theta(b,\kappa) := \sum_{n=0}^{\infty}(n+1)\phi^{1-\frac{1}{b}}(n) + \sum_{n=0}^{\infty}(n+1)\beta_\kappa^\kappa(n) < \infty.$$

Then the limit $D^2 = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}S_N^2$ exists and there exists $c_\ell > 0$ which depends only on ℓ so that

$$\big|\mathbb{E}S_N^2 - D^2 N\big| \le c_\ell C_0 N^{\frac{1}{2}} \tag{2.16}$$

for any $N \in \mathbb{N}$, where $C_0 = K^2(1+\gamma_m^\lambda)\Theta(b,\kappa)$. Moreover, $D^2 > 0$ if and only if there exists no stationary in the wide sense process $\{V_n : n \ge 1\}$ such that

$$F\big(\xi_n^{(1)}, \xi_{2n}^{(n)}, \dots, \xi_{\ell n}^{(\ell)}\big) = V_{n+1} - V_n, \quad P\text{-a.s.}$$

for any $n \in \mathbb{N}$, where $\xi^{(i)}$, $i = 1, \dots, \ell$ are independent copies of $\xi = \{\xi_n : n \ge 1\}$. When $\lambda = 0$ then the above results hold true without assuming (2.15) while when $\beta_\infty(r) = 0$ for some r they hold true for Borel measurable F's without assuming (2.5).

This theorem is a particular case of Theorem 1.3.4 in [19] and Theorem 2.2 in [15]. In fact, it is a consequence of the arguments in [25,26] and [17] and is formulated here for readers' convenience. We refer the readers to [16] for conditions in the special case when ξ_n, $n\geq 0$ forms a sufficiently fast mixing Markov chain. Remark that in the circumstance of Theorem 2.7 there exists a constant a_ℓ which depends only on ℓ so that $|\mathbb{E}S_N|\leq a_\ell K C_0$ for any $N\geq 1$. Indeed this is a consequence of (2.7) and Corollary 1.3.14 in [19]. Therefore, for any $N\geq 1$,

$$\left|\mathrm{Var}(S_N)-D^2N\right|\leq C_1N^{\frac{1}{2}} \tag{2.17}$$

for some constant C_1 which depends only on C_0, ℓ and K.

We recall next that the k-th cumulant of a random variable W with finite moments of all orders is given by

$$\Gamma_k(W)=\frac{1}{i^k}\frac{d^k}{dt^k}\left(\ln\mathbb{E}e^{itW}\right)|_{t=0}.$$

Note that $\Gamma_1(W)=\mathbb{E}W$, $\Gamma_2(W)=\mathrm{Var}(W)$ and that $\Gamma_k(aW)=a^k\Gamma_k(W)$ for any $a\in\mathbb{R}$ and $k\geq 1$.

Theorem 2.8. *Under Assumption 2.1, there exists a constant c_0 which depends only on K, ℓ, d, a, η and κ so that for any $k\geq 3$,*

$$\left|\Gamma_k(\bar{S}_N)\right|\leq N(k!)^{1+\gamma_1}(c_0)^{k-2}$$

where $\gamma_1=\frac{1}{\eta}$. When Assumption 2.2 holds true, there exists a constant c_0 which depends only on K, ℓ, d, a, η, M, ζ, κ and λ so that for any $k\geq 3$,

$$\left|\Gamma_k(\bar{S}_N)\right|\leq N(k!)^{1+\gamma_2}(c_0)^{k-2}$$

where $\gamma_2=\gamma_1+\lambda\zeta$.

Note that Theorem 2.8 holds true without assuming that $(\xi_n,\xi_m)\sim\mu_{m-n}$ since its proof does not require that the limit D^2 exists. When $(\xi_n,\xi_m)\sim\mu_{m-n}$ then $N^{-\frac{1}{2}}\bar{S}_N$ satisfies the CLT and so the term N on the above right hand sides should not be alarming since Theorem 2.8 implies that

$$\left|\Gamma_k\left(N^{-\frac{1}{2}}\bar{S}_N\right)\right|\leq (k!)^{1+\gamma}\left(N^{-\frac{1}{2}}c_0\right)^{k-2}$$

for any $k\geq 3$, where γ is either γ_1 or γ_2, depending on the case. After proving Theorem 2.8, the moderate deviations theorems and the (stretched) exponential concentration inequalities stated in Theorems 2.3 and 2.6 follow from the so called method of cumulants (see [30] and [10]).

Theorem 2.4 will follow from the following result together with the Hoeffding–Azuma inequality.

Theorem 2.9. *Suppose that F is a bounded Hölder function and that*

$$\varphi:=\sum_{n=0}^{\infty}\phi(n)<\infty.$$

Then there exists a constant $B>0$ which depends only on ℓ so that for any $N\geq 1$ and $r\geq 0$ there is a martingale $M_n^{(N,r)}$, $n\geq 1$ whose differences are bounded by $\delta_1':=BK(\varphi+r+1)$ and

$$\left\|S_N-M_{\ell N}^{(N,r)}\right\|_\infty\leq\delta_2':=BKN\beta_\infty^\kappa(r)+\delta_1'.$$

When $\beta_\infty(r_0)=0$ for some $r_0\geq 0$ then the above results hold true with $r=r_0$ for any bounded Borel function F.

2.1. *Product functions*

In the special case when F has the form

$$F(x_1,\dots,x_\ell)=\prod_{i=1}^{\ell}f_i(x_i) \tag{2.18}$$

the results stated in Theorems 2.3 (i) and Theorem 2.6 (i) hold true under weaker assumptions, as described in what follows.

Recall first that the α-mixing coefficients are given by

$$\alpha(n)=\sup\bigl\{\alpha(\mathcal{F}_{-\infty,k},\mathcal{F}_{k+n,\infty}):k\in\mathbb{Z}\bigr\} \tag{2.19}$$

where for any two sub-σ-algebras $\mathcal{G},\mathcal{H}\subset\mathcal{F}$,

$$\alpha(\mathcal{G},\mathcal{H})=\sup\bigl\{\bigl|P(A\cap B)-P(A)P(B)\bigr|:A\in\mathcal{G},B\in\mathcal{H}\bigr\}. \tag{2.20}$$

Then (see [3]) $\alpha(n)\le\frac{1}{2}\phi(n)$ for any $n\ge 0$, and so, assumptions involving $\alpha(n)$ are weaker than ones involving $\phi(n)$. We also recall that (see [11]) for any bounded functions $g_1,\dots,g_L$, numbers $m_1<n_1<m_2<n_2<\cdots<m_L<n_L$ and $\mathcal{F}_{m_i,n_i}$-measurable random vectors U_i, $i=1,2,\dots,L$,

$$\left|\mathbb{E}\prod_{i=1}^{L}g_i(U_i)-\prod_{i=1}^{L}\mathbb{E}g(U_i)\right|\le 8\left(\prod_{i=1}^{L}\sup|g_j|\right)\sum_{t=2}^{L}\alpha(m_t-n_{t-1}). \tag{2.21}$$

Relying on (2.21) we show in Section 3.4 that all the results stated in Theorems 2.3 (i) and Theorem 2.6 (i) hold true when f_i's are bounded. The situation of unbounded f_i's satisfying certain moment conditions is discussed there, as well.

Next, let $T:\Omega\to\Omega$ be a measurable and P-preserving map. We assume here that there exists a space $\mathcal{H}$ of real valued bounded functions on Ω, a norm $\|\cdot\|_{\mathcal{H}}$ on $\mathcal{H}$, a constant d and a sequence $c(m)$, $m\ge 1$, which converges to 0 as $m\to\infty$, so that for any $f,g\in\mathcal{H}$ and $n\ge 1$,

$$\mathrm{Cor}_P\bigl(g,f\circ T^n\bigr)\le d\|g\|_{\mathcal{H}}\sup|f|c(n). \tag{2.22}$$

Usually, in applications, Ω will be a topological space and $\mathcal{H}$ will be a space of Hölder continuous functions equipped with an appropriate norm. We also assume that the f_i's are members of $\mathcal{H}$. Obtaining the MDP and exponential concentration inequalities under condition (2.22) is important when either there are no underlying Markov partitions or there is no effective estimate on the diameter of such partitions (so it is impossible to approximate effectively Hölder continuous functions by functions which are constant on elements of such partitions). For instance, (2.22) holds true with $c(n)=e^{-an}$, $a>0$ in the (nonrandom) setup of [28], where T is a locally distance expanding map and $\mathcal{H}$ is a space of (locally) Hölder continuous functions, while there are no underlying Markov partitions. Let $n_1<n_L<\cdots<n_L$ and $g_1,\dots,g_L\in\mathcal{H}$. By writing

$$\prod_{i=1}^{L}g_i\circ T^{n_i}=\bigl(g_1\cdot G\circ T^{n_2-n_1}\bigr)\circ T^{n_1}$$

where $G=\prod_{i=2}^{L}g_i\circ T^{n_i-n_2}$ we obtain that

$$\left|\mathbb{E}_P\prod_{i=1}^{L}g_i\circ T^{n_i}-\prod_{i=1}^{L}\mathbb{E}_Pg_i\circ T^{n_i}\right|\le dM^L\sum_{t=2}^{L}c(n_t-n_{t-1}) \tag{2.23}$$

where $M=\max\{\sup|g_i|,\|g_i\|_{\mathcal{H}}:i=1,2,\dots,L\}$. Suppose next that

$$\sum_{n=1}^{\infty}nc(n)<\infty.$$

Using (2.23) in place of (2.21), we will prove in Section 3.4 that all the results stated in Theorem 2.3 (i), Theorem 2.6 (ii) and Theorem 2.7 hold true with $\beta_\kappa(n)\equiv 0$ and $c(n)$ in place of $\phi(n)$.

3. Nonconventional moderate deviations and exponential inequalities via the method of cumulants

3.1. *General estimates of cumulants*

Let V be a finite set and $\rho:V\times V\to[0,\infty)$ be so that $\rho(v,v)=0$ and $\rho(u,v)=\rho(v,u)$ for any $u,v\in V$. For any $A,B\subset V$ set

$$\rho(A,B)=\min\bigl\{\rho(a,b):a\in A,b\in B\bigr\}.$$

Let X_v, $v \in V$ be a collection of centered random variables with finite moments of all orders, and for each $v \in V$ and $t \in (0, \infty]$ let $\varrho_{v,t} \in (0, \infty]$ be so that $\|X_v\|_t \le \varrho_{v,t}$. Set $W = \sum_{v \in V} X_v$. The following result is (essentially) proved in [14] (see Theorem 1 there).

Theorem 3.1. *Let $0 < \delta \le \infty$. Suppose that for any $k \ge 1$, $b > 0$ and a finite collection A_j, $j \in \mathcal{J}$ of (nonempty) subsets of V so that $\min_{i \ne j} \rho(A_i, A_j) \ge b$ and $r := \sum_{j \in \mathcal{J}} |A_j| \le k$ we have*

$$\left| \mathbb{E} \prod_{j \in \mathcal{J}} \prod_{i \in A_j} X_i - \prod_{j \in \mathcal{J}} \mathbb{E} \prod_{j \in A_j} X_i \right| \le (r-1) \Big(\prod_{j \in \mathcal{J}} \prod_{i \in A_j} \varrho_{i,(1+\delta)k} \Big) \gamma_\delta(b,k) \tag{3.1}$$

where $\gamma_\delta(b,r)$ is some nonnegative number which depends only on δ, b and r, and $|\Delta|$ stands for the cardinality of a finite set Δ. Then for any $k \ge 2$ and $s > 0$,

$$\left|\Gamma_k(W)\right| \le k^k \big(2^k C(k) \big(L_s(k)\big)^{k-1} + R_s(\delta,k)\big)$$

where for any $0 < t \le \infty$,

$$L_s(t) = \sup\Big\{ \sum_{u \in V : \rho(u,v) \le s} \varrho_{u,t} : v \in V \Big\}, \qquad C(t) = \sum_{v \in V} \varrho_{v,t},$$

$$R_s(\delta,k) = \sum_{m \ge s+1} \big(L_m\big((1+\delta)k\big)\big)^{k-1} C\big((1+\delta)k\big) \lambda\big(\tilde{\gamma}_\delta(m,k), k\big),$$

$$\tilde{\gamma}_\delta(m,k) = \max\big\{\gamma_\delta(m,r)/r : 1 \le r \le k\big\}$$

$$\textit{and} \quad \lambda(\varepsilon,k) = k! \sum_{r=1}^{[\frac{k}{2}]} \frac{\varepsilon^r (3r+1)^{k-2r}}{r(k-2r)!}.$$

The difference in the formulations of Theorem 1 in [14] and Theorem 3.1 is that the result from [14] relies on a certain local mixing condition instead of (3.1). But in proof from there the author obtains (3.1) with $\varrho_{v,t} = \|X_v\|_t$ and appropriate $\gamma_\delta(b,k)$ relying on that mixing condition, and so Theorem 3.1 is proved exactly as in [14]. We reformulated this theorem in order to include the case when $\beta_q(r) \ne 0$ for any r and the second situation considered in Section 2.1.

Note that by Stirling's approximation there exists a constant $C > 0$ so that $k^k \le C e^k k!$ for any $k \ge 1$. Remark also that when condition (3.1) holds true only in the case when $|\mathcal{J}| = 2$, then using induction this implies that (3.1) holds true with $k\gamma_\delta(b,k)$ instead of $\gamma_\delta(b,k)$, for collections of more than two sets. Compare this with [12,22] and [8] in the case when $V = \{1, \ldots, n\}$ and $\rho(x,y) = |x-y|$.

Next, the following result is a consequence of Theorem 3.1.

Corollary 3.2. *Suppose, in addition to the assumptions of Theorem 3.1, that there exist $c_0 \ge 1$ and $u_0 \ge 0$ so that*

$$\left|\big\{u \in V : \rho(u,v) \le s\big\}\right| \le c_0 s^{u_0} \tag{3.2}$$

for any $v \in V$ and $s \ge 1$. Assume also that $\tilde{\gamma}_\delta(m,k) \le d e^{-am^\eta}$ for some $a, \eta > 0$, $d \ge 1$ and all $k, m \ge 1$. Then there exists a constant c which depends only on c_0, a, u_0 and η so that for any $k \ge 2$,

$$\left|\Gamma_k(W)\right| \le d^k |V| c^k (k!)^{1+\frac{u_0}{\eta}} \big(M_k^k + M_{(1+\delta)k}^k\big) \tag{3.3}$$

where for any $q > 0$,

$$M_q = \max\{\varrho_{v,q} : v \in V\} \quad \textit{and} \quad M_q^k = (M_q)^k.$$

When the X_v's are bounded and (3.1) holds true with $\delta = \infty$ we can always take $\varrho_{v,t} = \varrho_{v,\infty}$, $t > 0$ and then for any $k \ge 2$,

$$\left|\Gamma_k(W)\right| \le 2 d^k |V| M_\infty^k c^k (k!)^{1+\frac{u_0}{\eta}}. \tag{3.4}$$

When $\delta < \infty$ and there exist $\theta \geq 0$ and $M > 0$ so that

$$(\varrho_{v,k})^k \leq M^k (k!)^\theta \tag{3.5}$$

for any $v \in V$ and $k \geq 1$, then for any $k \geq 2$,

$$\left|\Gamma_k(W)\right| \leq 3C^{\frac{\theta}{1+\delta}} d^k |V| c^k (1+\delta)^k M^k (k!)^{1+\frac{u_0}{\eta}+\theta} \tag{3.6}$$

where C is some absolute constant.

The proof of this corollary is elementary but for readers' convenience we will give all the details.

Proof. Let $k \geq 2$ and $m \geq s \geq k^{\frac{1}{\eta}}$. Set $\varepsilon = \varepsilon_m = e^{-am^\eta}$. Then $\tilde{\gamma}_\delta(m,k) \leq d\varepsilon$ and so

$$\lambda\big(\tilde{\gamma}_\delta(m,k), k\big) \leq d^k k! 4^k \sum_{r=1}^{[\frac{k}{2}]} \frac{\varepsilon^r r^{k-2r-1}}{(k-2r)!} \leq d^k k! 4^k \sum_{r=1}^{[\frac{k}{2}]-1} \frac{\varepsilon^r r^{k-2r}}{(k-2r)!} + d^k k! 4^k \varepsilon^{[\frac{k}{2}]}.$$

Observe that $k!4^k\varepsilon^{[\frac{k}{2}]} \leq H\varepsilon$ for some constant H which depends only on a and η, where we used that $m^\eta \geq k$. Moreover, by Stirling's approximation there exists an absolute constant $C > 0$ so that for any $1 \leq r \leq [\frac{k}{2}] - 1$,

$$\frac{1}{(k-2r)!} \leq C \frac{e^{k-2r}}{(k-2r)^{k-2r}}.$$

Therefore,

$$\lambda\big(\tilde{\gamma}_\delta(m,k), k\big) \leq Ck!(4de)^k \sum_{r=1}^{[\frac{k}{2}]-1} \varepsilon^r \left(\frac{r}{k-2r}\right)^{k-2r} + d^k H\varepsilon. \tag{3.7}$$

Consider next the function $g_m = g_{m,k} : [1, \frac{k}{2} - 1] \to \mathbb{R}$ given by

$$g_m(r) = \varepsilon^r \left(\frac{r}{k-2r}\right)^{k-2r} = e^{r\ln\varepsilon - (k-2r)\ln(\frac{k}{r}-2)}.$$

Then,

$$g_m'(r) = \left(\ln\varepsilon + 2\ln\left(\frac{k}{r} - 2\right) + \frac{k}{r}\right) g_m(r).$$

If $g_m'(r_0) = 0$ for some $r_0 \in [1, \frac{k}{2} - 1]$ then

$$ak \leq am^\eta = -\ln\varepsilon = 2\ln\left(\frac{k}{r_0} - 2\right) + \frac{k}{r_0} \leq \frac{3k}{r_0}$$

and so $r_0 \leq \frac{3}{a} := q$. Hence,

$$\max_{r\in[1,\frac{k}{2}-1]} g_m(r) \leq \max\left(g_m(1), g_m\left(\frac{k}{2} - 1\right), \max_{w\in[1,q_k]} g_m(w)\right)$$

where $q_k = \min(\frac{k}{2} - 1, q)$ and we set $\max\varnothing = -\infty$. Observe now that

$$g_m(1) = \frac{\varepsilon}{(k-2)^{k-2}} \leq \frac{k^2\varepsilon}{k!} \leq 3^k (k!)^{-1}\varepsilon.$$

Since $m^\eta \geq k$ we also have

$$g_m\left(\frac{k}{2} - 1\right) \leq k^2 \varepsilon^{\frac{k}{2}-1} \leq \varepsilon k^2 e^{-ak(\frac{k}{2}-2)} \leq c_1 \varepsilon (k!)^{-1}$$

where c_1 is a constant which depends only on a. When $k \le 2(q+1)$ we can trivially write

$$\max_{w\in[1,q_k]} g_m(w) \le \varepsilon(\psi_0)^k (k!)^{-1}$$

for some constant ψ_0 which depends only on a. On the other hand, when $k > 2(q+1)$ then using that the function $x \to x^{-x}$ is strictly decreasing on $[1,\infty)$ and then Stirling's approximation we derive that

$$\max_{w\in[1,q_k]} g_m(w) = \max_{w\in[1,q]} g_m(w) \le \varepsilon(q+1)^k \big(k-[2q]-1\big)^{-(k-[2q]-1)} \le \varepsilon\psi^k (k!)^{-1}$$

where ψ is a constant which depends only on a, and we also used the inequality $k! \le (k-l)!k^l \le (k-l)!3^{kl}$, $1 \le l \le k$. We conclude from the above estimates that there exists a constant $R = R(a,\eta)$ which depends only on a and η so that for any $1 \le r \le \frac{k}{2} - 1$,

$$g_m(r) = \varepsilon^r \left(\frac{r}{k-2r}\right)^{k-2r} \le \varepsilon R^k (k!)^{-1}$$

which together with (3.7) yields

$$\lambda\big(\tilde{\gamma}_\delta(m,k),k\big) \le d^k R_0^k \varepsilon = d^k R_0^k e^{-am^\eta} \tag{3.8}$$

where $R_0 = R_0(a,\eta) \ge 1$ is another constant.

Next, using (3.2), (3.8) and the definitions of $C(t)$ and $L_s(t)$ we obtain that

$$R_s(\delta,k) \le d^k(1+H)R_0^k (M_{(1+\delta)k})^k |V| \sum_{m\ge s+1} m^{u_0(k-1)} e^{-am^\eta}$$

where $L_s(t)$, $C(t)$, $R_s(\delta,k)$ are defined in Theorem 3.1. Set $j_0 = j_0(k,\eta) = [\frac{(k-1)u_0+2}{\eta}] + 1$. Then

$$m^{u_0(k-1)} e^{-am^\eta} \le m^{u_0(k-1)} j_0!\big(am^\eta\big)^{-j_0} \le j_0! a^{-j_0} m^{-2}.$$

By Stirling's approximation there exists a constant Q which depends only on η and u_0 so that $j_0! \le Q^k (k!)^{\frac{u_0}{\eta}}$ and therefore,

$$\sum_{m\ge s+1} m^{u_0(k-1)} e^{-am^\eta} \le j_0! \sum_{m\ge s+1} \frac{1}{m^2} \le \frac{1}{s} j_0! \le \frac{1}{s}(Q_1)^k (k!)^{\frac{u_0}{\eta}}$$

where Q_1 is a constant which depends only on η, a and u_0. Taking $s = k^{\frac{1}{\eta}}$ the estimate (3.3) follows by Theorem 3.1, the definition of $L_s(m)$, Stirling's approximation and (3.2). By Stirling's approximation $((1+\delta)k)! \le C(k!)^{1+\delta}(1+\delta)^{(1+\delta)k}$ and (3.6) follows now by (3.3) and the inequality $(1+\delta)^{\frac{1}{1+\delta}} \le e$. $\square$

3.2. *Proof the Theorem* 2.8

Fix some $N \ge 1$ and set $V = V_N = \{1,2,\dots,N\}$. For any $n, m \in V$ set

$$\rho(n,m) = \rho_\ell(n,m) = \min_{1\le i,j\le \ell} |in - jm|.$$

Then for any $\Delta_1, \Delta_2 \subset V$,

$$\rho(\Delta_1,\Delta_2) = \inf\big\{|x-y| : x \in \mathcal{T}_1, y \in \mathcal{T}_2\big\} := \mathrm{dist}(\mathcal{T}_1,\mathcal{T}_2) \tag{3.9}$$

where $\mathcal{T}_i = \{jt : t \in \Delta_i, 1 \le j \le \ell\}$, $i = 1,2$. Moreover, for any $s \ge 1$ and $1 \le n \le N$,

$$A_s(n,N) := \big\{m \in V : \rho(m,n) \le s\big\} = \bigcup_{1\le i,j\le \ell} \left[\frac{in-s}{j}, \frac{in+s}{j}\right]$$

and so

$$\big|A_s(n,N)\big| \le 3\ell^2 s. \tag{3.10}$$

Therefore (3.2) holds true in our situation with $u_0 = 1$. For each $n \in V$ put $\Theta_n = (\xi_n, \xi_{2n}, \dots, \xi_{\ell n})$ and

$$X_n = F(\Theta_n) - \mathbb{E}F(\Theta_n).$$

Then $\bar{S}_N = \sum_{n \in V} X_n$. We will verify that the remaining assumptions of Corollary 3.2 hold true with the above X_n's. First, for each $r \ge 0$ and $n \ge 1$, set $\xi_{n,r} = \mathbb{E}[\xi_n|\mathcal{F}_{n-r,n+r}]$, $\Theta_{n,r} = (\xi_{n,r}, \xi_{2n,r}, \dots, \xi_{\ell n,r})$ and

$$X_{n,r} = F(\Theta_{n,r}) - \mathbb{E}F(\Theta_{n,r}).$$

Set $\rho_\infty = 2K(1+\ell)$ and $\varrho_t = 2K(1+\ell\tau_{\lambda t}^{\lambda})$, $0 < t < \infty$. When $\lambda = 0$ then by (2.6) for any $n \ge 1$ and $r \ge 0$,

$$\max\big(\|X_n\|_\infty, \|X_{n,r}\|_\infty\big) \le 2K(1+\ell) = \varrho_\infty \tag{3.11}$$

while when $\lambda > 0$ we derive similarly that for any $0 < t < \infty$, $n \ge 1$ and $r \ge 0$,

$$\max\big(\|X_n\|_t, \|X_{n,r}\|_t\big) \le 2K\big(1+\ell\tau_{\lambda t}^{\lambda}\big) = \varrho_t \tag{3.12}$$

where we also used the contraction of conditional expectations. Note that $\varrho_{t_1} \le \varrho_{t_2}$ whenever $0 < t_1 \le t_2 < \infty$. In our future applications of Corollary 3.2 we will always take $\varrho_{v,\infty} = \varrho_\infty$ and $\varrho_{v,t} = \varrho_t$ for $0 < t < \infty$.

Next, when (2.8) holds true and $\lambda > 0$ then by Stirling's approximation there exists an absolute constant $C > 1$ so that for any $k \ge 1$,

$$\tau_{\lambda k}^{\lambda k} = \mathbb{E}|\xi_1|^{k\lambda} \le M^{k\lambda}\big((k\lambda)!\big)^{\zeta} \le C^{\zeta(\lambda+1)}Q^k(k!)^{\lambda\zeta} \le \big(C^{\zeta(\lambda+1)}Q\big)^k(k!)^{\lambda\zeta} \tag{3.13}$$

where $Q = \lambda^{\zeta\lambda}M^{\lambda} \ge 1$. Therefore, the collection of numbers $\varrho_{v,k} = \varrho_k$ satisfies (3.5) with $4K\ell C^{\zeta(\lambda+1)}Q$ in place of M and with $\theta = \lambda\zeta$.

Now we will verify condition (3.1). We will need first the following general result. Let U_i, $i = 1, 2, \dots, L$ be d_i-dimensional random vectors defined on the probability space $(\Omega, \mathcal{F}, P)$ from Section 1, and $\{\mathcal{C}_j : 1 \le j \le s\}$ be a partition of $\{1, 2, \dots, L\}$. Consider the random vectors $U(\mathcal{C}_j) = \{U_i : i \in \mathcal{C}_j\}$, $j = 1, \dots, s$, and let

$$U^{(j)}(\mathcal{C}_i) = \big\{U_i^{(j)} : i \in \mathcal{C}_j\big\}, \quad j = 1, \dots, s$$

be independent copies of the $U(\mathcal{C}_j)$'s. For each $1 \le i \le L$ let $a_i \in \{1, \dots, s\}$ be the unique index such that $i \in \mathcal{C}_{a_i}$, and for any bounded Borel function $H : \mathbb{R}^{d_1+d_2+\dots+d_L} \to \mathbb{R}$ set

$$\mathcal{D}(H) = \big|\mathbb{E}H(U_1, U_2, \dots, U_L) - \mathbb{E}H\big(U_1^{(a_1)}, U_2^{(a_2)}, \dots, U_L^{(a_L)}\big)\big|. \tag{3.14}$$

The following result is proved in Corollary 1.3.11 in [19] (see also Corollary 3.3 in [15]),

Lemma 3.3. *Suppose that each U_i is $\mathcal{F}_{m_i,n_i}$-measurable, where $n_{i-1} < m_i \le n_i < m_{i+1}$, $i = 1, \dots, L$, $n_0 = -\infty$ and $m_{L+1} = \infty$. Then, for any bounded Borel function $H : \mathbb{R}^{d_1+d_2+\dots+d_L} \to \mathbb{R}$,*

$$\mathcal{D}(H) \le 4\sup|H| \sum_{i=2}^{L} \phi(m_i - n_{i-1}) \tag{3.15}$$

where $\sup|H|$ is the supremum of $|H|$. In particular, when $s = 2$ then

$$\alpha\big(\sigma\{U(\mathcal{C}_1)\}, \sigma\{U(\mathcal{C}_2)\}\big) \le 4\sum_{i=2}^{L} \phi(m_i - n_{i-1}) \tag{3.16}$$

where $\sigma\{X\}$ stands for the σ-algebra generated by a random variable X.

Next, in order to show that (3.1) holds true we first notice that for any set of pairs (a_i, b_i), $i = 1, 2, \dots, m$,

$$\prod_{i=1}^{m} a_i - \prod_{i=1}^{m} b_i = \sum_{i=1}^{m} \prod_{1 \le j < i} a_j (a_i - b_i) \prod_{i < j \le m} b_j. \tag{3.17}$$

Let $n_1, \dots, n_m \in V$ and $q \ge 0$. When $\lambda = 0$ using (3.17), (3.11) and (2.5) we obtain that for each $1 \le i \le m$,

$$\begin{aligned}\left|\mathbb{E}\prod_{i=1}^{m} X_{n_i} - \mathbb{E}\prod_{i=1}^{m} X_{n_i,q}\right| &\le m(\varrho_\infty)^{m-1}\max\big\{\mathbb{E}|X_{n_i} - X_{n_i,q}|, 1 \le i \le m\big\}\\ &\le m(\varrho_\infty)^m \ell \beta_\kappa^\kappa(q).\end{aligned} \tag{3.18}$$

When $\lambda > 0$ then by the contraction of conditional expectations for any $1 \le i \le m$,

$$\begin{aligned}\|X_{n_i} - X_{n_i,q}\|_m &\le K\left\|1 + \sum_{j=1}^{\ell}\big(|\xi_{jn_i}|^\lambda + |\xi_{jn_i,q}|^\lambda\big)\right\|_m \sum_{j=1}^{\ell}\big\||\xi_{jn_i} - \xi_{jn_i,q}|^\kappa\big\|_\infty\\ &\le K\ell\big(1 + 2\ell\tau_{m\lambda}^\lambda\big)\beta_\infty^\kappa(q) \le \varrho_m \ell \beta_\infty^\kappa(q),\end{aligned}$$

where ϱ_m is defined in (3.12). Therefore by (3.17), (3.12), (2.5) and the Hölder inequality,

$$\left|\mathbb{E}\prod_{i=1}^{m} X_{n_i} - \mathbb{E}\prod_{i=1}^{m} X_{n_i,q}\right| \le m(\varrho_m)^m \ell \beta_\infty^\kappa(q). \tag{3.19}$$

Now, let $k, b \ge 1$ and a finite collection A_j, $j \in \mathcal{J}$ of nonempty subsets of V be so that $r := \sum_{j \in \mathcal{J}} |A_j| \le k$ and $\rho(A_j, A_i) \ge b$ whenever $i \ne j$. Set $q_b = [\frac{b}{3}]$. When $\lambda = 0$ set $\delta = \infty$ and

$$\gamma(b, r) = \gamma_\infty(b, r) = 128\ell r\big(\phi(q_b) + \beta_\kappa^\kappa(q_b)\big)$$

while when $\lambda > 0$ set $\delta = 1$ and

$$\gamma(b, r) = \gamma_1(b, r) = 128\ell r\big(\phi^{\frac{1}{2}}(q_b) + \beta_\infty^\kappa(q_b)\big).$$

We claim that in both cases (3.1) holds true with $\varrho_{v,t} = \varrho_t$ defined in (3.12) and (3.11) and the above δ and $\gamma_\delta(b, r)$ (depending on the case). Indeed, when $\lambda = 0$ and $\delta = \infty$ set $\gamma'_\delta(b, r) = 32\ell r\phi(q_b)$, while when $\lambda > 0$ and $\delta = 1$ we set $\gamma'_\delta(b, r) = 32\ell r(\phi(q_b))^{\frac{\delta}{1+\delta}} = 32\ell r\sqrt{\phi(q_b)}$. In order to prove this claim we first assert that in both cases,

$$\left|\mathbb{E}\prod_{j \in \mathcal{J}}\prod_{i \in A_j} X_{i,q_b} - \prod_{j \in \mathcal{J}}\mathbb{E}\prod_{j \in A_j} X_{i,q_b}\right| \le (r - 1)\left(\prod_{j \in \mathcal{J}}\prod_{i \in A_j}\|X_{i,q_b}\|_{(1+\delta)k}\right)\gamma'_\delta(b, k). \tag{3.20}$$

It is clear that (3.1) with these A_j's, b and k follow from either (3.18) and (3.20) or (3.19) and (3.20), depending on the case, where when $r \ge 2$ we use that $r \le 2(r - 1)$. In order to obtain (3.20) we need first the following. Let $\Delta_1, \Delta_2 \subset \mathbb{N}$ be so that $\rho(\Delta_1, \Delta_2) \ge b$ and set $d_1 = |\Delta_1| + |\Delta_2|$ and $\mathcal{T}_i = \{jx : x \in \Delta_i, 1 \le j \le \ell\}$, $i = 1, 2$. Then by (3.9) we have $\mathrm{dist}(\mathcal{T}_1, \mathcal{T}_2) = \rho(\Delta_1, \Delta_2) \ge b$ and so we can write

$$\mathcal{T} := \mathcal{T}_1 \cup \mathcal{T}_2 = \bigcup_{i=1}^{L} C_i$$

where $L \le \ell d_1$, $c_i + b \le c_{i+1}$ for any $c_i \in C_i$ and $c_{i+1} \in C_{i+1}$, $i = 1, 2, \dots, L - 1$ and each one of the C_i's is either a subset of $\mathcal{T}_1$ or a subset of $\mathcal{T}_2$. Applying (3.16) with the random vectors $U_i = \{\xi_{j,q_b} : j \in C_i\}$, $i = 1, 2, \dots, L$ and the partition of $\{1, 2, \dots, L\}$ into the sets $\mathcal{C}_1 = \{1 \le i \le L : C_i \subset \mathcal{T}_1\}$ and $\mathcal{C}_2 = \{1 \le i \le L : C_i \subset \mathcal{T}_2\}$ we obtain that

$$\alpha\big(\sigma\{X_{i,q_b} : i \in \Delta_1\}, \sigma\{X_{j,q_b} : j \in \Delta_2\}\big) \le 4\ell d_1 \phi(q_b). \tag{3.21}$$

Recall next that (see Corollary A.2 in [20]) for any two sub-σ-algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$,

$$\mathrm{Cov}(\eta_1, \eta_2) \le 8\|\eta_1\|_u \|\eta_2\|_v \big(\alpha(\mathcal{G}, \mathcal{H})\big)^{1 - \frac{1}{u} - \frac{1}{v}}, \tag{3.22}$$

whenever h_1 is $\mathcal{G}$-measurable, h_2 is $\mathcal{H}$-measurable and $1 < u, v \le \infty$ satisfy that $\frac{1}{u} + \frac{1}{v} < 1$ (where we set $\frac{1}{\infty} = 0$). The estimate (3.20) follows now exactly as in the paragraph preceeding equality (10) in [14], relying on (3.22) and on (3.21), in place of the mixing conditions from [14]. Indeed, writing $\mathcal{J} = \{1, 2, \dots, J\}$, setting $\Delta_1 = A_1$ and $\Delta_2 = \bigcup_{1<i\le J} A_i$ and applying (3.22) with $u = \frac{(1+\delta)k}{|\Delta_1|}$ and $v = \frac{(1+\delta)k}{|\Delta_2|}$ we obtain that

$$\left|\mathrm{Cov}\left(\prod_{i\in\Delta_1} X_{i,q_b}, \prod_{i\in\Delta_2} X_{i,q_b}\right)\right| \le 8\left\|\prod_{i\in\Delta_1} X_{i,q_b}\right\|_u \left\|\prod_{i\in\Delta_2} X_{i,q_b}\right\|_v \alpha^{1-\frac{1}{1+\delta}} \tag{3.23}$$

where $\alpha = \alpha(\sigma\{X_{i,q_b} : i \in \Delta_1\}, \sigma\{X_{j,q_b} : j \in \Delta_2\})$ and we also used that $\alpha \le 1$ and

$$\frac{1}{u} + \frac{1}{v} = \frac{|\Delta_1 \cup \Delta_2|}{k(1+\delta)} = \frac{r}{k(1+\delta)} \le \frac{1}{1+\delta}.$$

Using the Hölder inequality to estimate the norms on the right hand side of (3.23) and then repeating the above arguments with $\mathcal{J}_i = \{i, i+1, \dots, J\}$, $i = 2, 3, \dots, J$ in place of $\mathcal{J}$ we obtain (3.20), taking into account that $J = |\mathcal{J}| \le \sum_{i\in\mathcal{J}} |A_i| = r$. Using either (3.11) or (3.12) we conclude that all the conditions of Corollary 3.2 are satisfied under either Assumption 2.1 or Assumption 2.2, and the proof of Theorem 2.8 is complete.

3.3. *Proof of Theorems 2.3 and 2.6*

First, (2.9) from Theorem 2.3 follows from Theorem 2.8 and Lemma 2.3 in [30]. Next, for the purpose of proving Theorem 2.6, suppose that $D^2 > 0$. Then (2.14) follows by Lemma 6.2 in [10] (which is a consequence of Lemma 2.3 in [30]). Finally, let a_N, $N \ge 1$ be a sequence of real numbers so that

$$\lim_{N\to\infty} a_N = \infty \quad \text{and} \quad \lim_{N\to\infty} a_N N^{-\frac{1}{2+4\gamma}} = 0$$

where $\gamma = \gamma_1 = \frac{1}{\eta}$ under Assumption 2.1 and $\gamma = \gamma_2 = \gamma_1 + \lambda\zeta$ under Assumption 2.2. The variances v_N grow linearly fast in N and therefore by Theorem 2.8 and Theorem 1.1 in [10] the sequence $(a_N)^{-1} Z_N$, $N \ge 1$ satisfies the MDP with the speed $s_N = a_N^2$ and the rate function $I(x) = \frac{1}{2}x^2$. Since v_N/N converges to $D^2 > 0$ as $N \to \infty$, $|\mathbb{E}S_N|$ is bounded in N and I is continuous we derive that $(DN^{\frac{1}{2}} a_N)^{-1} S_N$, $N \ge 1$ satisfies the MDP stated in Theorem 2.6, and the proof of Theorem 2.6 is complete.

3.4. *Product functions case*

Consider the situation when F has the form

$$F(x_1, \dots, x_\ell) = \prod_{i=1}^{\ell} f_i(x_i).$$

We will describe here shortly how to prove Theorems 2.3 and 2.6 in the situations discussed at the end of Section 2.

3.4.1. *α-mixing case*

First, in the notations of Lemma 3.3, we obtain that (3.15) holds true for functions of the form $H(u) = \prod_{i=1}^{L} g_i(u_i)$ when all of the g_i's are bounded, where $\phi(m_i - n_{i-1})$ is replaced by $4\alpha(m_i - n_{i-1})$ for $i = 2, 3, \dots, L$. Indeed, setting

$$u^{(\mathcal{C}_j)} = \{u_i : i \in \mathcal{C}_j\} \quad \text{and} \quad G_j\big(u^{(\mathcal{C}_j)}\big) = \prod_{i\in\mathcal{C}_j} g_i(u_i), \quad j = 1, 2, \dots, s,$$

we derive from (2.21), exactly as in the proof of Corollary 1.3.11 in [19] (or Corollary 3.3 in [15]), that

$$\left|\mathbb{E}H(U_1, \dots, U_L) - \prod_{j=1}^{s} \mathbb{E}G_j\big(U(\mathcal{C}_j)\big)\right| \le 16\left(\prod_{j=1}^{L} \sup|g_j|\right) \sum_{i=2}^{L} \alpha(m_i - n_{i-1}). \tag{3.24}$$

Note that the derivation of (3.24) is indeed possible since (2.21) holds true for arbitrary bounded g_i's, appropriate U_i's and partitions $\mathcal{C}$'s. Relying on (3.24) we can approximate the left-hand side of (3.1) and therefore the results stated in Theorems 2.3 and 2.6 hold true with $\alpha(n)$ in place of $\phi(n)$.

We remark that (2.21) follows, in fact, by a repetitive application of (3.22) with $u = v = \infty$. Applying (3.22) with finite u's and v's we obtain similar estimates when the $g_i(U_i)$'s are not bounded but only satisfy certain moment conditions, where the product $\prod_{j=1}^{L} \sup |g_j|$ is replaced with an appropriate product of the form $\prod_{i=1}^{L} \|g_i(U_i)\|_q$ and $\alpha(m_t - n_{t-1})$ is replaced with $(\alpha(m_t - n_{t-1}))^{\zeta}$ for an appropriate $0 < \zeta < 1$. We refer the readers to the proof of (3.20) for the exact details. Relying on this "unbounded version" of (3.24), we can approximate the left-hand side of (3.1) and obtain results similar to the ones stated in Theorem 2.3 (ii) and Theorem 2.6 (ii), but with with $\alpha(n)$ in place of $\phi(n)$.

3.4.2. *Decay of correlations case*

Let T, $\mathcal{H}$ and $c(m), m \geq 1$ be as described at the end of Section 2. Let $n_1 < n_2 < \cdots < n_L$ and $g_1, \dots, g_L \in \mathcal{H}$. In the notations of Lemma 3.3, using (2.23) and the T-invariance of P, we obtain similarly to the above α-mixing case that

$$\left| \mathbb{E}_P \prod_{i=1}^{L} g_i \circ T^{n_i} - \prod_{j=1}^{s} \mathbb{E}_P \prod_{i \in \mathcal{C}_j} g_i \circ T^{n_i} \right| \leq 2dM^L \sum_{t=2}^{L} c(n_t - n_{t-1}) \tag{3.25}$$

where $M = \max\{\sup |g_i|, \|g_i\|_{\mathcal{H}} : i = 1, 2, \dots, L\}$. Note that when $\sum_{n=1}^{\infty} nc(n) < \infty$ then all the results stated in Theorem 2.7 are proved similarly to [17,26] and [25] relying on (3.25) instead of the mixing assumptions from there. The inequality (3.25) also yields appropriate estimates of the left-hand side of (3.1), and we conclude that that all the results stated in Theorem 2.3 (i) and Theorem 2.6 (i) hold true with $\beta_\kappa(n) \equiv 0$ and $c(n)$ in place of $\phi(n)$.

4. Exponential inequalities via martingale approximation-proof of Theorems 2.4 and 2.9

In this section we adapt the martingale approximation technique from [26] and approximate S_N in the L^∞ norm by martingales with bounded differences. As in [26] we first write

$$F(x_1, \dots, x_\ell) = \sum_{i=1}^{\ell} F_i(x_1, \dots, x_i) \tag{4.1}$$

where

$$F_\ell(x_1, \dots, x_\ell) = F(x_1, \dots, x_\ell) - \int F(x_1, \dots, x_{\ell-1}, z)\, d\mu(z)$$

and for $i = 1, 2, \dots, \ell - 1$,

$$\begin{aligned} F_i(x_1, \dots, x_i) = &\int F(x_1, \dots, x_i, z_{i+1}, \dots, z_\ell)\, d\mu(z_{i+1}) \cdots d\mu(z_\ell) \\ &- \int F(x_1, \dots, x_{i-1}, z_i, \dots, z_\ell)\, d\mu(z_i) \cdots d\mu(z_\ell). \end{aligned}$$

Then for each $1 \leq i \leq \ell$,

$$\int F_i(y_1, \dots, y_{i-1}, z)\, d\mu(z) = 0, \quad \forall y_1, \dots, y_{i-1}$$

where for $i = 1$ we used that $\bar{F} = 0$.

Next, recall that (see [3], Ch. 4) for any two sub-σ-algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$,

$$2\phi(\mathcal{G}, \mathcal{H}) = \sup\left\{ \left\| \mathbb{E}[g|\mathcal{G}] - \mathbb{E}g \right\|_\infty : g \in L^\infty(\Omega, \mathcal{H}, P), \|g\|_\infty \leq 1 \right\} \tag{4.2}$$

where $\phi(\mathcal{G}, \mathcal{H})$ is defined by (2.2). The following result is a version of Corollary 3.6 in [26] and Lemma 1.3.10 in [19] (see also Lemma 3.2 in [15]). It does not seem to be new but for readers' convenience and completeness we will prove it here.

Lemma 4.1. *Let $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ be two sub-σ-algebras of $\mathcal{F}$ and $d \in \mathbb{N}$. Let $f(\cdot, \omega) : \mathbb{R}^d \to \mathbb{R}$ be a random function so that $f(x, \omega)$ is $\mathcal{H}$-measurable for any fixed $x \in \mathbb{R}^d$ and P-a.s. for any $x, y \in \mathbb{R}^d$,*

$$\left| f(x, \omega) \right| \leq C \quad \text{and} \quad \left| f(x, \omega) - f(y, \omega) \right| \leq C|x - y|^\kappa \tag{4.3}$$

where $C > 0$ and $\kappa \in (0,1]$ are constants which do not depend on x, y and ω. Set $\tilde{f}(x,\omega) = \mathbb{E}[f(x,\cdot)|\mathcal{G}](\omega)$ and $\bar{f}(x) = \int f(x,\omega)\,dP(\omega) = \int \tilde{f}(x,\omega)\,dP(\omega)$. Then there exists a measurable set $\Omega' \subset \Omega$ so that $P(\Omega') = 1$, $\tilde{f}(x,\omega)$ is defined for all $\omega \in \Omega'$ and $x \in \mathbb{R}^d$ and

$$\sup_{x\in\mathbb{R}^d}\left|\tilde{f}(x,\omega) - \bar{f}(x)\right| \le 2C\phi(\mathcal{G},\mathcal{H}), \quad P\text{-a.s.} \tag{4.4}$$

In particular, for any $\mathbb{R}^d$-valued random variable X,

$$\left|\tilde{f}(X,\omega) - \bar{f}(X)\right| \le 2C\phi(\mathcal{G},\mathcal{H}), \quad P\text{-a.s.} \tag{4.5}$$

Proof. Let $\mathcal{A} = \{\mathcal{A}_i : i \in \mathcal{I}\}$ be a countable partition of $\mathbb{R}^d$ and denote its diameter by $\operatorname{diam}\mathcal{A}$. For each $i \in \mathcal{I}$ let $\mathbb{1}_{A_i}$ be the indicator function of A_i and choose some $a_i \in A_i$. Then by (4.3), P-a.s. for any $x \in \mathbb{R}^d$ we have

$$\left|f(x,\omega) - \sum_{i\in\mathcal{I}} \mathbb{1}_{A_i}(x) f(a_i,\omega)\right| \le C(\operatorname{diam}\mathcal{A})^\kappa.$$

Taking conditional expectations with respect to $\mathcal{G}$ and then the limit as $\operatorname{diam}\mathcal{A} \to 0$ we obtain the existence of Ω' as in the statement of the lemma. Fixing $\mathcal{A}$ and taking again conditional expectations with respect to $\mathcal{G}$ we derive that

$$\sup_{x\in\mathbb{R}^d}\left|\tilde{f}(x,\omega) - \sum_{i\in\mathcal{I}} \mathbb{1}_{A_i}(x)\tilde{f}(a_i,\omega)\right| \le C(\operatorname{diam}\mathcal{A})^\kappa, \quad P\text{-a.s.}$$

Similarly, we obtain by taking expectations that

$$\sup_{x\in\mathbb{R}^d}\left|\bar{f}(x) - \sum_{i\in\mathcal{I}} \mathbb{1}_{A_i}(x)\bar{f}(a_i)\right| \le C(\operatorname{diam}\mathcal{A})^\kappa.$$

Using (4.2) and (4.3) we deduce that for each i,

$$\left|\tilde{f}(a_i,\omega) - \bar{f}(a_i)\right| \le 2\left\|f(a_i,\cdot)\right\|_\infty \phi(\mathcal{G},\mathcal{H}) \le 2C\phi(\mathcal{G},\mathcal{H}), \quad P\text{-a.s.}$$

and therefore, P-a.s.,

$$\sup_{x\in\mathbb{R}^d}\left|\tilde{f}(x,\omega) - \bar{f}(x)\right| \le 2C\phi(\mathcal{G},\mathcal{H}) + 2C(\operatorname{diam}\mathcal{A})^\kappa.$$

Taking the limit as $\operatorname{diam}\mathcal{A} \to 0$ we obtain (4.4). □

Next, consider the random functions $F_{i,n,r}$ given by

$$F_{i,n,r}(x_1,\dots,x_{i-1},\omega) = \mathbb{E}\big[F_i(x_1,\dots,x_{i-1},\xi_n)|\mathcal{F}_{n-r,n+r}\big](\omega).$$

Note that in view of the uniform continuity of F these are indeed random functions, i.e. all the random variables $F_{i,n,r}(x_1,\dots,x_{i-1},\cdot)$, $x_1,\dots,x_{i-1} \in \mathbb{R}^\wp$ can be defined on a measurable set Ω' so that $P(\Omega') = 1$. Set

$$\begin{aligned} &Y_{i,in} = F(\xi_n,\xi_{2n},\dots,\xi_{in}) \quad \text{and} \quad Y_{i,m} = 0 \quad \text{if } m \notin \{in : n \in \mathbb{N}\} \quad \text{and} \\ &Y_{i,in,r} = F_{i,in,r}(\xi_{n,r},\xi_{2n,r},\dots,\xi_{(i-1)n,r},\omega) \quad \text{and} \quad Y_{i,m,r} = 0 \quad \text{if } m \notin \{in : n \in \mathbb{N}\} \end{aligned} \tag{4.6}$$

where we recall that $\xi_{m,r} = \mathbb{E}[\xi_m|\mathcal{F}_{m-r,m+r}]$ for any $m \ge 1$.

The following result is proved exactly as in the proof of Proposition 5.8 in [26] using Lemma 4.1 and the inequality $|F| \le K(1+\ell)$ instead of Corollary 3.6 (ii) and the moment assumptions from there.

Corollary 4.2. *Suppose that $\varphi := \sum_{n=0}^\infty \phi(n) < \infty$. Then there exists a constant $B > 0$ which depends only on ℓ so that for any $l \ge 0$ and $r \ge 0$,*

$$\sum_{n=l}^\infty \left\|\mathbb{E}[Y_{i,n,r}|\mathcal{F}_{-\infty,l+r}]\right\|_\infty \le BK(r+1+\varphi).$$

Now we introduced the martingales constructed in [17] relying on ideas originated in [26]. For any $1 \le i \le \ell$, $n \ge 0$ and $r \ge 0$ set $R_{i,n,r} = \sum_{s \ge n+1} \mathbb{E}[Y_{i,s,r}|\mathcal{F}_{-\infty,n+r}]$ and

$$W_{i,n,r} = Y_{i,n,r} + R_{i,n,r} - R_{i,n-1,r}.$$

Then when i and r are fixed $W_{i,n,r}$, $n \ge 1$ is a martingale difference with respect to the filtration $\{\mathcal{F}_{-\infty,n+r} : n \ge 1\}$ and by Corollary 4.2,

$$\|R_{i,n,r}\|_\infty \le 2BK(\varphi + r + 1) \tag{4.7}$$

and therefore there exists a constant $B_1 > 0$ which depends only on ℓ so that

$$\|W_{i,n,r}\|_\infty \le B_1 K(\varphi + r + 1).$$

Set $W^{(N)}_{i,n,r} = \mathbb{1}_{\{n \le iN\}} W_{i,n,r}$,

$$W^{(N)}_{n,r} = \sum_{i=1}^{\ell} W^{(N)}_{i,n,r},$$

$M^{(N)}_{i,n,r} = \sum_{m=1}^{n} W^{(N)}_{i,m,r}$ and

$$M^{(N,r)}_n = \sum_{m=1}^{n} W^{(N)}_{m,r} = \sum_{i=1}^{\ell} M^{(N)}_{i,n,r}.$$

Then when r and N are fixed $M^{(N,r)}_n$, $n \ge 1$ is a martingale (with respect to the above filtration) whose differences are bounded by $\ell B_1 K(\varphi + r + 1)$. We estimate now the L^∞-norm

$$\big\|S_N - M^{(N,r)}_{N\ell}\big\|_\infty.$$

We first write

$$S_N - M^{(N,r)}_{N\ell} = \sum_{i=1}^{\ell}\sum_{n=1}^{N}(Y_{i,in} - Y_{i,in,r}) + \sum_{i=1}^{\ell}(R_{i,N\ell,r} - R_{i,0,r})$$

where we used (4.1). By replacing ξ_{jn} with $\xi_{jn,r}$, $j = 1, 2, \dots, i$ in the definitions of $Y_{i,in}$ and $Y_{i,in,r}$, using the Hölder continuity of F and that $\xi_{\ell n,r}$ is $\mathcal{F}_{\ell n-r,\ell n+r}$-measurable we obtain that

$$|Y_{i,in} - Y_{i,in,r}| \le KB_2\beta^{\kappa}_{\infty}(r), \quad P\text{-a.s.}$$

for any $1 \le i \le \ell$, $n \in \mathbb{N}$ and $r \ge 0$, where $B_2 = B_2(\ell)$ is some constant which depends only on ℓ. Combining this with (4.7) we obtain that

$$\big\|S_N - M^{(N,r)}_{N\ell}\big\|_\infty \le B_3 K\big(N\beta^{\kappa}_{\infty}(r) + \varphi + r + 1\big) := \delta_2' \tag{4.8}$$

where $B_3 = B_3(\ell)$ is another constant, and the proof of Theorem 2.9 is complete. In order to prove Theorem 2.4, we first apply the Hoeffding–Azuma inequality (see, for instance, page 33 in [29]) and obtain that for any $\lambda > 0$,

$$\mathbb{E}e^{\lambda M^{(N,r)}_{N\ell}} \le e^{\lambda^2 \sum_{n=1}^{\ell N} \|W^{(N,r)}_n\|^2_\infty} \le e^{\ell N \delta_0^2 \lambda^2}$$

where $\delta_0 = B_1 K(\varphi + r + 1)$. Combining this with (4.8) we obtain (2.11). Next, by the Markov inequality for any random variable Z, $t_0 > 0$ and $\lambda > 0$ we have $P(Z \ge t_0) \le e^{-\lambda t_0}\mathbb{E}e^{\lambda Z}$. Taking $Z = S_N$, $t_0 = t + \delta_2$, using (2.11) and then optimizing by taking $\lambda = \frac{t}{2\ell N\delta_2^2}$ we obtain (2.12), and the proof of Theorem 2.4 is complete.

5. Nonlinear indexes

Let $q_1,\dots,q_\ell$ be functions which map $\mathbb{N}$ to $\mathbb{N}$, are strictly increasing on some ray $[R,\infty)$ and are ordered so that

$$q_1(n)<q_2(n)<\cdots<q_\ell(n)$$

for any sufficiently large n. For any $N\in\mathbb{N}$ consider the random variable

$$S_N=\sum_{n=1}^{N}\big(F(\xi_{q_1(n)},\xi_{q_2(n)},\dots,\xi_{q_\ell(n)})-\bar F\big) \tag{5.1}$$

where $\bar F$ is given by (2.7). We further assume that the difference $q_i(n)-q_{i-1}(n)$ tends to ∞ as $n\to\infty$ for any $i=1,2,\dots,\ell$, where $q_0\equiv 0$, though the situation when some of these differences are nonnegative constants can be considered, as well (see Section 3 in [18]). Next, for any $n,m\in\mathbb{N}$ set

$$\tilde\rho(n,m)=\tilde\rho_\ell(n,m)=\min_{1\le i,j\le\ell}\big|q_i(n)-q_j(m)\big|.$$

We will rely on the following

Assumption 5.1. There exists $Q\ge 1$ so that for any $1\le j\le\ell$ and $a,b\ge q_j(R)$,

$$\big|q_j^{-1}(a)-q_j^{-1}(b)\big|\le Q\big(1+|a-b|\big) \tag{5.2}$$

where q_j^{-1} is the inverse of the restriction of q_j to the ray $[R,\infty)$.

Set $\tilde A_s(n,N)=\{1\le m\le N:\tilde\rho(n,m)\le s\}$. When (5.2) holds true then for any $1\le n\le N$ and $s\ge 1$,

$$\big|\tilde A_s(n,N)\big|\le Q\ell^2(1+2s)\le 3\ell^2Qs$$

which means that (3.2) holds true with $c_0=3Q\ell^2$ and $u_0=1$. Condition (5.2) holds true, for instance, when all q_j's have the form $q_j(x)=[p_j(x)]$ where each p_j is a strictly increasing function whose inverse p_j^{-1} has bounded derivative on some ray $[K,\infty)$. For example we can take p_j's to be a polynomial with positive leading coefficient, exponential function etc.

We conclude that under Assumption 5.1, all the results stated in Theorem 2.8 hold true. Therefore, (2.9) holds true and all the results stated in Theorem 2.6 hold true when D^2 exists and it is positive. The limit D^2 exists when q_i's satisfy the conditions from [26] or, as in [18], when they are polynomials taking integer values on the integers. See [17] and [18] for conditions equivalent to $D^2>0$. Note also that for such q_i's Theorem 2.9 holds true, as well, since the martingale approximation method was applied in [26] and [17] successfully, and so the arguments from Section 4 can be repeated.

Remark 5.2. Let $q(n)$, $n\ge 1$ be a strictly increasing sequence of natural numbers, and consider the process $\tilde\xi_n$, $n\ge 1$ given by $\tilde\xi_n=\xi_{q(n)}$. Set $\tilde{\mathcal F}_{m,n}=\mathcal F_{q(m),q(n)}$ and let $\tilde\phi(n)$ and $\tilde\beta_q(n)$ be defined similarly to $\phi(n)$ and $\beta_q(n)$ but with the $\tilde{\mathcal F}_{m,n}$'s in place of the $\mathcal F_{m,n}$'s. Then $\tilde\beta_q(n)\le\beta_q(q(n))$ and $\tilde\phi(n)\le\phi(j(n))$, where

$$j(n)=\inf_{m\ge 1}\big(q(m+n)-q(m)\big).$$

When $q(n),j(n)\ge cn^l$ for some $l\ge 2$ and $c>0$ then the mixing and approximation coefficients $\tilde\phi(n)$ and $\tilde\beta_q(n)$ converge to 0 faster than $\phi(n)$ and $\beta_q(n)$, and by writing $s=q(s')\ge c(s')^l$ we can take $u_0=\frac{1}{l}$ in (3.2). Repeating the arguments from the proof of Theorem 2.8, we obtain similar estimates of $|\Gamma_k(\bar S_N)|$, but with $\gamma_1'=\frac{1}{\eta l^2}<\gamma_1$ in place of $\gamma_1=\frac{1}{\eta}$. The assumption that the distribution of (ξ_n,ξ_m) depends only on $n-m$ was only needed in order for D^2 to exist and for obtaining convergence rate towards it. Therefore, (2.9) and the corresponding estimate from Theorem 2.3 (ii) hold true with $\xi_{q(n)}$ and γ_1' in place of ξ_n and γ_1, respectively. If we know that the limit D^2 exists (after this replacement) then all the other results stated in Theorems 2.3 and 2.6 also hold true with $\frac{1}{\eta l^2}$ in place of $\frac{1}{\eta}$.

Consider, for instance, the case when q_i's are polynomials and $q(n) = n^l$ for some $l \geq 2$, namely, nonconventional sums of the form

$$\tilde{S}_N = \sum_{n=1}^{N} F(\xi_{p_1(n^l)}, \xi_{p_2(n^l)}, \dots, \xi_{p_\ell(n^l)}) \tag{5.3}$$

when all p_i's are polynomials. Then the limit D^2 exists (see [18]) and so all the results described above hold true.

6. Additional results

6.1. *The CLT and Berry–Esseen type estimates*

We recall first the following result (see Corollary 2.1 in [30]),

Lemma 6.1. *Let W be a random variable. Suppose that there exist $\gamma \geq 0$ and $\Delta > 0$ so that for any $k \geq 3$,*

$$\left|\Gamma_k(W)\right| \leq (k!)^{1+\gamma} \Delta^{-(k-2)}.$$

Let Φ be the standard normal distribution. Then,

$$\sup_{x\in\mathbb{R}}\left|P(W \leq x) - \Phi(x)\right| \leq c_\gamma \Delta^{-\frac{1}{1+2\gamma}}$$

where $c_\gamma = \frac{1}{6}(\frac{\sqrt{2}}{6})^{\frac{1}{1+2\gamma}}$.

Note that when $|\Gamma_k(W)| \leq C(k!)^{1+\gamma}\Delta^{-(k-2)}$, $k \geq 3$ for some constant $C \geq 1$ then the conditions of Lemma 6.1 are satisfied with ΔC^{-1} in place of Δ. This lemma together with the cumulants' estimates obtained in Theorem 2.8 yields convergence rates in the nonconventional CLT for $S_N/D\sqrt{N}$ which (when $\eta = 1$) are at best of order $N^{-\frac{1}{6}}$, since in our circumstances Δ is of order $N^{\frac{1}{2}}$ and $\gamma \geq 1$, where in the case when F is bounded we can take $\gamma = 1$. The rate $N^{-\frac{1}{6}}$ is better than the ones obtained in [17], which is important since the rates obtained in [15] and [19] do not apply to the cases considered in Section 5. Note that, in fact, we obtain here for the first time the CLT under condition (2.22) when F has the form (2.18).

Remark 6.2. Consider the case discussed in Remark 5.2 when all q_i's have the form $q_i(n) = p_i(n^l)$ for some polynomials $p_1, \dots, p_\ell$ and an integer $l \geq 2$, namely the sums $N^{-\frac{1}{2}}\tilde{S}_N$ where $\tilde{S}_N$ is defined in (5.3). Then under Assumption 2.1 we obtain (when $D^2 > 0$) closer to optimal rates. Indeed, in these circumstances Theorem 2.8 holds true with $\gamma_1' = \frac{1}{\eta l^2}$ in place of $\gamma_1 = \frac{1}{\eta}$ and so, using the equality $\Gamma_k(aW) = a^k\Gamma_k(W)$, $a \in \mathbb{R}$, we can apply now Lemma 6.1 with $\gamma = \gamma_1'$ and Δ of the form $\Delta = c\sqrt{N}$ and obtain rates of order $N^{-\frac{1}{2+4(\eta l^2)^{-1}}}$, which are better than $N^{-\frac{1}{6}}$ when $\eta = 1$.

6.2. *Moment estimates of Gaussian type*

Theorem 2.8 also implies the following

Theorem 6.3. *Suppose that the conditions of Theorem 2.8 hold true. Let Z be a random variable which is distributed according to the standard normal law. Then for any $p \geq 1$,*

$$\left|\mathbb{E}(\bar{S}_N)^p - \left(\mathrm{Var}(S_N)\right)^{\frac{p}{2}}\mathbb{E}Z^p\right| \leq (c_{0,1})^p (p!)^{1+\gamma} \sum_{1\leq u\leq \frac{p-1}{2}} N^u \frac{p^u}{(u!)^2}$$

where $c_{0,1} = \max(1, c_0)$, $\gamma = \gamma_1$ when Assumption 2.1 holds true, $\gamma = \gamma_2$ when Assumption 2.2 holds true and c_0, γ_1 and γ_2 are specified in Theorem 2.8.

Proof. The arguments below are based on the proof of Theorem 3 in [12]. By formula (1.53) in [30], for any $p \geq 1$ and $N \in \mathbb{N}$,

$$\mathbb{E}(\bar{S}_N)^p = \sum_{1 \leq u \leq \frac{p}{2}} \frac{1}{u!} \sum_{k_1+k_2+\cdots+k_u=p} \frac{p!}{k_1! \cdots k_u!} \Gamma_{k_1}(\bar{S}_N) \cdots \Gamma_{k_u}(\bar{S}_N).$$

Let $1 \leq u \leq \frac{p}{2}$. When $k_i = 1$ for some $1 \leq i \leq u$ then $\Gamma_{k_i}(\bar{S}_N) = \mathbb{E}\bar{S}_N = 0$ and so the corresponding summand $\frac{p! \prod_{i=1}^u \Gamma_{k_i}(\bar{S}_N)}{\prod_{i=1}^u (k_i!)}$ vanishes. When p is even and $u = \frac{p}{2}$ then the unique non-vanishing summand corresponds to the choice of $k_i = 2$, $i = 1, 2, \dots, u$ and it equals $(\mathrm{Var}(S_N))^{\frac{p}{2}} \mathbb{E}Z^p$. When p is odd then $\mathbb{E}Z^p = 0$, and therefore for any $p \geq 1$,

$$\begin{aligned} &\left|\mathbb{E}(\bar{S}_N)^p - \left(\mathrm{Var}(S_N)\right)^{\frac{p}{2}} \mathbb{E}Z^p\right| \\ &\quad \leq \sum_{1 \leq u \leq \frac{p-1}{2}} \frac{1}{u!} \sum_{k_1+k_2+\cdots+k_u=p} \frac{p!}{k_1! \cdots k_u!} \left|\Gamma_{k_1}(\bar{S}_N) \cdots \Gamma_{k_u}(\bar{S}_N)\right|. \end{aligned}$$

Applying the Hölder inequality to Euler's Γ function we obtain that $(k!)^p \leq (p!)^k$ for any integers k and p so that $1 \leq k \leq p$. Using Theorem 2.8 we derive that

$$\frac{|\Gamma_{k_1}(\bar{S}_N) \cdots \Gamma_{k_u}(\bar{S}_N)|}{k_1! \cdots k_u!} \leq N^u c_0^{\sum_{i=1}^u k_i - 2u} \left(\prod_{i=1}^u (k_i!)\right)^{\gamma} \leq (c_{0,1})^p (p!)^{\gamma}$$

for any $1 \leq k_1, \dots, k_u$ so that $\sum_{i=1}^u k_i = p$, where γ is described in the statement of Theorem 6.3. Thus,

$$\left|\mathbb{E}(\bar{S}_N)^p - \left(\mathrm{Var}(S_N)\right)^{\frac{p}{2}} \mathbb{E}Z^p\right| \leq (c_{0,1})^p (p!)^{1+\gamma} \sum_{1 \leq u \leq \frac{p-1}{2}} \frac{\mathcal{N}(u,p)}{u!}$$

where

$$\begin{aligned} \mathcal{N}(u,p) &:= \left|\left\{2 \leq k_1, \dots, k_u \leq p : \sum_{i=1}^u k_i = p\right\}\right| \\ &\leq \left|\left\{1 \leq k_1, \dots, k_u \leq p : \sum_{i=1}^u k_i = p\right\}\right| \leq \frac{p^u}{u!}. \end{aligned}$$

We conclude from the above estimates that for any integer $p \geq 1$,

$$\left|\mathbb{E}(\bar{S}_N)^p - \left(\mathrm{Var}(S_N)\right)^{\frac{p}{2}} \mathbb{E}Z^p\right| \leq (c_{0,1})^p (p!)^{1+\gamma} \sum_{1 \leq u \leq \frac{p-1}{2}} N^u \frac{p^u}{(u!)^2}$$

and the proof of Theorem 6.3 is complete. □

We note that this theorem yields an appropriate Rosenthal type inequality for the nonconventional sums $\bar{S}_N$ and that, in fact, makes the method of moments effective for them, which provides an additional proof of the nonconventional central limit theorem. See Remarks 4 and 5 in [12], where we also use (2.17) (which implies that $N^{-1}\mathrm{Var}(S_N)$ converges to D^2 as $N \to \infty$).

References

[1] J. Aaronson and M. Denker. Local limit theorems for Gibbs–Markov maps. *Stoch. Dyn.* **1** (2001) 193–237. MR1840194 https://doi.org/10.1142/S0219493701000114

[2] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*. *Lecture Notes in Mathematics* **470**. Springer, Berlin, 1975. MR0442989

[3] R. C. Bradley. *Introduction to Strong Mixing Conditions*, **1**. Kendrick Press, Heber City, 2007. MR2325295

[4] S. Chatterjee. An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc.* **53** (2016) 617–642. MR3544262 https://doi.org/10.1090/bull/1539

[5] S. Chatterjee and A. Dembo. Non linear large deviations. *Adv. Math.* **299** (2016) 396–450. MR3519474 https://doi.org/10.1016/j.aim.2016.05.017

[6] J. R. Chazottes. Fluctuations of observables in dynamical systems: From limit theorems to concentration inequalities. In *Nonlinear Dynamics New Directions, Vol. 11, Nonlinear Syst. Complex* 47–85. Springer, Berlin, 2015. MR3379397 https://doi.org/10.1007/978-3-319-09867-8_4

[7] J. R. Chazottes and S. Gouëzel. Optimal concentration inequalities for dynamical systems. *Comm. Math. Phys.* **316** (2012) 843–889. MR2993935 https://doi.org/10.1007/s00220-012-1596-7

[8] J. Dedecker, P. Doukhan, G. Lang, J. R. León, S. Louhichi and C. Prieur. *Weak Dependence: With Examples and Applications*. *Lecture Notes in Statistics* **190**. Springer-Verlag, Berlin, 2007. MR2338725

[9] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*, 2nd edition. *Applications of Mathematics* **38**. Springer, New York, 1998. MR1619036 https://doi.org/10.1007/978-1-4612-5320-4

[10] H. Döring and P. Eichelsbacher. Moderate deviations via cumulants. *J. Theoret. Probab.* **26** (2013) 360–385. MR3055808 https://doi.org/10.1007/s10959-012-0437-0

[11] P. Doukhan. *Mixing: Properties and Examples*. *Lecture Notes in Statistics* **85**. Springer, Berlin, 1994. MR1312160 https://doi.org/10.1007/978-1-4612-2642-0

[12] P. Doukhan and M. Neumann. Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stochastic Process. Appl.* **117** (2007) 878–903. MR2330724 https://doi.org/10.1016/j.spa.2006.10.011

[13] H. Furstenberg. Nonconventional ergodic averages. *Proc. Sympos. Pure Math.* **50** (1990) 43–56. MR1067751 https://doi.org/10.1090/pspum/050/1067751

[14] A. B. Grochakov. Upper estimates for semi invariants of the sum of multi-indexed random variables. *Discrete Math. Appl.* **5** (1995) 317–331. MR1361492 https://doi.org/10.1515/dma.1995.5.4.317

[15] Y. Hafouta. Stein's method for nonconventional sums. *Electron. Commun. Probab.* **23** (2018) Paper no. 38. MR3820128 https://doi.org/10.1214/18-ECP142

[16] Y. Hafouta and Yu. Kifer. A nonconventional local limit theorem. *J. Theoret. Probab.* **29** (2016) 1524–1553. MR3571253 https://doi.org/10.1007/s10959-015-0625-9

[17] Y. Hafouta and Yu. Kifer. Berry–Esseen type estimates for nonconventional sums. *Stoch. Model. Appl.* **126** (2016) 2430–2464. MR3505233 https://doi.org/10.1016/j.spa.2016.02.006

[18] Y. Hafouta and Yu. Kifer. Nonconventional polynomial CLT. *Stochastics* **89** (2017) 550–591. MR3590435 https://doi.org/10.1080/17442508.2016.1267181

[19] Y. Hafouta and Yu. Kifer. *Nonconventional Limit Theorems and Random Dynamics*. World Scientific, Singapore, 2018. MR3793181 https://doi.org/10.1142/10849

[20] P. G. Hall and C. C. Hyde. *Martingale Central Limit Theory and Its Application*. Academic Press, New York, 1980. MR0624435

[21] N. T. A. Haydn and F. Yang. Local escape rates for ϕ-mixing dynamical systems. Preprint. Available at arXiv:1806.07148.

[22] R. S. Kallabis and M. H. Neumann. An exponential inequality under weak dependence. *Bernoulli* **12** (2006) 333–350. MR2218558 https://doi.org/10.3150/bj/1145993977

[23] Yu. Kifer. Nonconventional limit theorems. *Probab. Theory Related Fields* **148** (2010) 71–106. MR2653222 https://doi.org/10.1007/s00440-009-0223-9

[24] Yu. Kifer. A nonconventional strong law of large numbers and fractal dimensions of some multiple recurrence sets. *Stoch. Dyn.* **12** (2012) 1150023. MR2926580 https://doi.org/10.1142/S0219493711500237

[25] Yu. Kifer. Strong approximations for nonconventional sums and almost sure limit theorems. *Stochastic Process. Appl.* **123** (2013) 2286–2302. MR3038505 https://doi.org/10.1016/j.spa.2013.02.009

[26] Yu. Kifer and S. R. S. Varadhan. Nonconventional limit theorems in discrete and continuous time via martingales. *Ann. Probab.* **42** (2014) 649–688. MR3178470 https://doi.org/10.1214/12-AOP796

[27] Yu. Kifer and S. R. S. Varadhan. Nonconventional large deviations theorems. *Probab. Theory Related Fields* **158** (2014) 197–224. MR3152784 https://doi.org/10.1007/s00440-013-0481-4

[28] V. Mayer, B. Skorulski and M. Urbański. *Distance Expanding Random Mappings, Thermodynamical Formalism, Gibbs Measures and Fractal Geometry*. *Lecture Notes in Mathematics* **2036**. Springer, Berlin, 2011. MR2866474 https://doi.org/10.1007/978-3-642-23650-1

[29] V. D. Milman and G. Schechtman. *Asymptotic Theory of Finite-Dimensional Normed Spaces*. *Lecture Notes in Mathematics* **1200**. Springer-Verlag, Berlin, 1986. (With an appendix by M. Gromov). MR0856576

[30] L. Saulis and V. A. Statulevicius. *Limit Theorems for Large Deviations*. Kluwer Academic, Dordrecht, Boston, 1991. MR1171883 https://doi.org/10.1007/978-94-011-3530-6

[31] L. S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* **7** (1998) 585–650. MR1637655 https://doi.org/10.2307/120960

[32] L. S. Young. Recurrence time and rate of mixing. *Israel J. Math.* **110** (1999) 153–188. MR1750438 https://doi.org/10.1007/BF02808180