Probab. Theory Relat. Fields (2013) 155:463–486
DOI 10.1007/s00440-011-0404-1

A strong invariance principle for nonconventional sums

Yuri Kifer

Received: 23 December 2010 / Revised: 25 October 2011 / Published online: 8 December 2011

Abstract In Kifer and Varadhan (Nonconventional limit theorems in discrete and continuous time via martingales, 2010) we obtained a functional central limit theorem (known also as a weak invariance principle) for sums of the form $\sum_{n=1}^{[Nt]} F\big(X(n), X(2n), \ldots, X(kn), X(q_{k+1}(n)), X(q_{k+2}(n)), \ldots, X(q_\ell(n))\big)$ (normalized by $1/\sqrt{N}$) where $X(n), n \geq 0$ is a sufficiently fast mixing vector process with some moment conditions and stationarity properties, F is a continuous function with polynomial growth and certain regularity properties and $q_i, i > k$ are positive functions taking on integer values on integers with some growth conditions which are satisfied, for instance, when q_i's are polynomials of growing degrees. This paper deals with strong invariance principles (known also as strong approximation theorems) for such sums which provide their uniform in time almost sure approximation by processes built out of Brownian motions with error terms growing slower than $\sqrt{N}$. This yields, in particular, an invariance principle in the law of iterated algorithm for the above sums. Among motivations for such results are their applications to multiple recurrence for stochastic processes and dynamical systems as well, as to some questions in metric number theory and they can be considered as a natural follow up of a series of papers dealing with nonconventional ergodic averages.

Keywords Strong approximations · Limit theorems · Martingale approximation · Mixing · Dynamical systems

Mathematics Subject Classification (2000) Primary 60F15; Secondary 60G42 · 37D20 · 60F17

Y. Kifer (✉)
Institute of Mathematics, Hebrew University, 91904 Jerusalem, Israel
e-mail: kifer@math.huji.ac.il

1 Introduction

Nonconventional ergodic theorems attracted substantial attention in ergodic theory (see, for instance, [1,6]). From a probabilistic point of view ergodic theorems are laws of large numbers for stationary processes and once they are established it is natural to study deviations from the average. The most celebrated result of this kind is the central limit theorem. In [12] we obtained a functional limit theorem for expressions of the form

$$\xi_N(t) = 1/\sqrt{N} \sum_{1\le n\le Nt} \big(F\big(X(q_1(n)),\ldots,X(q_\ell(n))\big) - \bar{F}\big) \tag{1.1}$$

and for the corresponding continuous time expressions of the form

$$\xi_N(t) = 1/\sqrt{N} \int_0^{Nt} \big(F\big(X(q_1(s)),\ldots,X(q_\ell(s))\big) - \bar{F}\big)ds \tag{1.2}$$

where $X(n), n \ge 0$'s is a sufficiently fast mixing vector valued process with some moment conditions and stationarity properties, F is a continuous function with polinomial growth and certain regularity properties, $\bar{F} = \int F d(\mu \times \cdots \times \mu)$, μ is the distribution of $X(0)$, $q_j(t) = jt,\ j \le k$ and $q_j,\ j > k$ are positive functions taking on integer values on integers in the discrete time case with some growth conditions which are satisfied, for instance, when q_i's are polynomials of growing degrees though we actually need much less. A substantially more restricted central limit theorem for expressions of this sort was obtained in [11].

Functional central limit theorems are called nowadays also weak invariance principles while for more than 40 years now (since probably Strassen's work [16]) probabilists were interested also in strong invariance principles called also strong approximation theorems. The latter provides almost sure or in average approximation of a sum of N random variables by a Brownian motion or, more generally, by a Gaussian process with an error term growing slower than $\sqrt{N}$ which yields as a result both the central limit theorem and the law of iterated logarithm, as well as other limiting results which are clear or easy to prove for Gaussian processes.

We will show in this paper that the sums $\Xi(Nt) = \sqrt{N}\xi_N(t)$ appearing in (1.1) can be represented as $\sum_{1\le i\le \ell} \Xi_i(Nt)$ where each $\Xi_i(Nt)$ can be approximated with an error term of order $N^{\frac{1}{2}-\alpha}$, $\alpha > 0$ by a process $\sigma_i B_i(t)$ where $\sigma_i \ge 0$ is a constant and B_i is a Brownian motion. This result yields also a law of iterated logarithm type result saying that with probability one all limit points as $N \to \infty$ of the sequence $\xi_N(t)(\log\log N)^{-1/2},\ t \in [0,1]$ of paths belong to a compact set.

Our methods employ the martingale approximation machinery from [12], enhanced so that to obtain appropriate error estimates, together with the technique from [14] which involves partition into blocks and Skorokhod embedding of martingales into a Brownian motion (the latter was first used for similar purposes in [16]). Observe that the summands in (1.1) depend strongly on the future and martingale methods start working only after we force "the future to become present". By this reason the role of martingales in our nonconventional framework was not selfevident at the beginning but their effective use initiated in [12] opened a wide vista for proving various limit theorems in this setup. It was shown in [12] that ξ_N converges weakly to a Gaussian process and it would be interesting to obtain a strong approximation of $\sqrt{N}\xi_N(t)$ by such Gaussian process but this would require to deal with multi dimensional approximations where the Skorokhod embedding we rely on does not work. Observe that since the 1960ies several other methods were developed to provide approximation of sums of random variables by a Brownian motion. Among them is the quantile method (see, for instance, [13]) which provides essentially optimal approximation but works only for independent random variables and by this reason does not seem applicable to our setup. Another method developed by Stein (see its recent account in [5]) also yields nearly optimal error estimates but it is not yet clear whether it can be adapted to our situation. The advantages of yet another method based on estimates of conditional characteristic functions (see, for instance, [4]) lie in its applicability to the multidimensional situation where, for instance, the Skorokhod embedding does not work well, but complications in the use of characteristic functions exhibited in [11] make applicability of this method in our setup doubtful.

As in [12] our results hold true when, for instance, $X(n) = T^n f$ where $f = (f_1, \dots, f_\wp)$, T is a mixing subshift of finite type, a hyperbolic diffeomorphism or an expanding transformation taken with a Gibbs invariant measure (see for instance, [2]) and some other dynamical systems, as well, as in the case when $X(n) = f(\xi_n),\ f = (f_1, \dots, f_\wp)$ where ξ_n is a Markov chain satisfying the Doeblin condition (see [10]) considered as a stationary process with respect to its invariant measure. The main known application of the above type results is to multiple recurrence when we employ our limit theorems for the random variable which counts returns of the stochastic process under consideration to given sets. In this case the function F above is a product of some coordinate functions in which we plug in corresponding $X(q_i(n)) = \mathbb{I}_{A_j}(\eta(q_i(n)))$ where $\eta(m)$ is either $T^m x$ in the dynamical systems case or ξ_m in the Markov chain case and $\mathbb{I}_A$ is the indicator of a set A. This yields also applications to metric number theory providing limit theorems, for instance, for the number $M_N(x)$ of n not exceeding N such that the digits at places $q_i(n)$, $i \le \ell$ in the m-base or continued fraction expansion of x belong to a chosen subset of digits. As it is well known the former expansions can be obtained via the multiplication by m (expanding) transformation while the latter via the Gauss map of the interval and both dynamical systems are exponentially fast ψ mixing with respect to their invariant Lebesgue or Gauss measure, respectively (see, for instance, [8]).

2 Preliminaries and main results

Our setup consists of a $\wp$-dimensional stochastic process $\{X(n), n = 0, 1, \dots\}$ on a probability space $(\Omega, \mathcal{F}, P)$ and of a family of σ-algebras $\mathcal{F}_{kl} \subset \mathcal{F},\ -\infty \le k \le l \le \infty$ such that $\mathcal{F}_{kl} \subset \mathcal{F}_{k'l'}$ if $k' \le k$ and $l' \ge l$. The dependence between two sub σ-algebras $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$ is measured often via the quantities

$$\varpi_{q,p}(\mathcal{G}, \mathcal{H}) = \sup\{\|E\big[g|\mathcal{G}\big] - E[g]\|_p : g \text{ is } \mathcal{H} - \text{measurable and } \|g\|_q \le 1\}, \tag{2.1}$$

where the supremum is taken over real functions and $\|\cdot\|_r$ is the $L^r(\Omega, \mathcal{F}, P)$-norm. Then more familiar α, ρ, ϕ and ψ-mixing (dependence) coefficients can be expressed in the form (see [3], Ch. 4),

$$\begin{aligned}\alpha(\mathcal{G}, \mathcal{H}) &= \frac{1}{4}\varpi_{\infty,1}(\mathcal{G}, \mathcal{H}),\ \ \rho(\mathcal{G}, \mathcal{H}) = \varpi_{2,2}(\mathcal{G}, \mathcal{H})\\ \phi(\mathcal{G}, \mathcal{H}) &= \frac{1}{2}\varpi_{\infty,\infty}(\mathcal{G}, \mathcal{H}) \text{ and } \psi(\mathcal{G}, \mathcal{H}) = \varpi_{1,\infty}(\mathcal{G}, \mathcal{H}).\end{aligned}$$

The relevant quantities in our setup are

$$\varpi_{q,p}(n) = \sup_{k\geq 0} \varpi_{q,p}(\mathcal{F}_{-\infty,k}, \mathcal{F}_{k+n,\infty}) \tag{2.2}$$

and accordingly

$$\alpha(n) = \frac{1}{4}\varpi_{\infty,1}(n),\ \rho(n) = \varpi_{2,2}(n),\ \phi(n) = \frac{1}{2}\varpi_{\infty,\infty}(n) \text{ and } \ \psi(n) = \varpi_{1,\infty}(n).$$

Our assumptions will require certain speed of decay as $n \to \infty$ of both the mixing rates $\varpi_{q,p}(n)$ and the approximation rates defined by

$$\beta_p(n) = \sup_{m\geq 0} \|X(m) - E\big(X(m)|\mathcal{F}_{m-n,m+n}\big)\|_p. \tag{2.3}$$

In what follows we can always extend the definitions of $\mathcal{F}_{kl}$ given only for $k, l \geq 0$ to negative k by defining $\mathcal{F}_{kl} = \mathcal{F}_{0l}$ for $k < 0$ and $l \geq 0$. Furthermore, we do not require stationarity of the process $X(n), n \geq 0$ assuming only that the distribution of $X(n)$ does not depend on n and the joint distribution of $\{X(n), X(n')\}$ depends only on $n - n'$ which we write for further references by

$$X(n) \stackrel{d}{\sim} \mu \text{ and } (X(n), X(n')) \stackrel{d}{\sim} \mu_{n-n'} \text{ for all } n, n' \tag{2.4}$$

where $Y \stackrel{d}{\sim} Z$ means that Y and Z have the same distribution.

Next, let $F = F(x_1, \ldots, x_\ell)$, $x_j \in \mathbb{R}^\wp$ be a function on $\mathbb{R}^{\wp\ell}$ such that for some $\iota, K > 0, \kappa \in (0, 1]$ and all $x_i, y_i \in \mathbb{R}^\wp$, $i = 1, \ldots, \ell$,

$$|F(x_1, \ldots, x_\ell) - F(y_1, \ldots, y_\ell)| \leq K\left(1 + \sum_{j=1}^{\ell}|x_j|^\iota + \sum_{j=1}^{\ell}|y_j|^\iota\right)\sum_{j=1}^{\ell}|x_j - y_j|^\kappa \tag{2.5}$$

and

$$|F(x_1, \ldots, x_\ell)| \leq K\left(1 + \sum_{j=1}^{\ell}|x_j|^\iota\right). \tag{2.6}$$

The above assumptions are motivated by the desire to include, for instance, functions F polynomially dependent on their arguments. To simplify formulas we assume a 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 always can replace F by $F - \bar{F}$.

Our setup includes also a sequence of increasing functions $q_1(n) < q_2(n) < \cdots < q_\ell(n)$ taking on integer values on integers and such that the first k of them are $q_j(n) = jn,\ j \le k$ whereas the remaining ones grow faster in n. We assume that for $k+1 \le i \le \ell$,

$$q_i(n+1) - q_i(n) \ge n^\delta \tag{2.8}$$

for some $\delta > 0$ and all $n \ge 2$ while for $i \ge k$ and any $\epsilon > 0$,

$$\liminf_{n\to\infty} (q_{i+1}(\epsilon n) - q_i(n)) > 0 \tag{2.9}$$

which is equivalent in view of (2.8) to

$$\liminf_{n\to\infty} (q_{i+1}(\epsilon n) - q_i(n)) = \infty. \tag{2.10}$$

In order to give a detailed statement of our main result as well as for its proof it will be essential to represent the function $F = F(x_1, x_2, \dots, x_\ell)$ in the form

$$F = F_1(x_1) + \cdots + F_\ell(x_1, x_2, \dots, x_\ell) \tag{2.11}$$

where for $i < \ell$,

$$\begin{aligned} F_i(x_1, \dots, x_i) = &\int F(x_1, x_2, \dots, x_\ell)\, d\mu(x_{i+1}) \cdots d\mu(x_\ell) \\ &- \int F(x_1, x_2, \dots, x_\ell)\, d\mu(x_i) \cdots d\mu(x_\ell) \end{aligned} \tag{2.12}$$

and

$$F_\ell(x_1, x_2, \dots, x_\ell) = F(x_1, x_2, \dots, x_\ell) - \int F(x_1, x_2, \dots, x_\ell)\, d\mu(x_\ell)$$

which ensures, in particular, that

$$\int F_i(x_1, x_2, \dots, x_{i-1}, x_i)\, d\mu(x_i) \equiv 0 \quad \forall \quad x_1, x_2, \dots, x_{i-1}. \tag{2.13}$$

These enable us to write

$$\Xi(t) = \sum_{i=1}^{\ell} \Xi_i(t) \tag{2.14}$$

where for $1 \le i \le \ell$,

$$\Xi_i(t) = \sum_{1 \le n \le t} F_i(X(q_1(n)), \dots, X(q_i(n))). \tag{2.15}$$

The decomposition of $\Xi(t)$ above is different from [12] since we work here with each $\Xi_i(t)$ separately remaining all the time within a one-dimensional framework and do not care about multi-dimensional covariances.

For each $\theta > 0$ set

$$\gamma_\theta^\theta = \|X\|_\theta^\theta = E|X(n)|^\theta = \int \|x\|^\theta d\mu. \tag{2.16}$$

Our main result relies on

Assumption 2.1 With $d = (\ell - 1)\wp$ there exist $p, q \ge 1$ and $\delta, m > 0$ with $\delta < \kappa - \frac{d}{p}$ satisfying

$$\sum_{n=0}^{\infty} n\varpi_{q,p}(n) < \infty, \tag{2.17}$$

$$\sum_{r=0}^{\infty} r^{4+\delta}\beta_q^\delta(r) < \infty, \tag{2.18}$$

$$\gamma_m < \infty \ , \gamma_{2q(\iota+2)} < \infty \text{ with } \frac{1}{2+\delta} \ge \frac{1}{p} + \frac{\iota+2}{m} + \frac{\delta}{q}. \tag{2.19}$$

Following [14] we will write $Z(t) \ll a(t)$ a.s. for a family of random variables $Z(t), t \ge 0$ and a positive function $a(t), t \ge 0$ if $\limsup_{t\to\infty} |Z(t)/a(t)| < \infty$ almost surely (a.s.)

Theorem 2.2 *Suppose that Assumption 2.1 holds true. Then without changing their (own but may be not joint) distributions the processes $\Xi_i(t)$, $t \ge 0$, $i = 1, \dots, \ell$ can be redefined on a richer probability space where there exist also standard Brownian motions $B_i(t)$, $t \ge 0$, $i = 1, \dots, \ell$ such that for some constants $\alpha > 0$ and $\sigma_i \ge 0$, $i = 1, \dots, \ell$,*

$$\Xi_i(t) - \sigma_i B_i(t) \ll t^{\frac{1}{2}-\alpha} \ a.s.. \tag{2.20}$$

As usual (see [14,9]), relying on the well known invariance principle in the law of iterated logarithm for the Brownian motion (see [15]) we obtain immediately from the above theorem the following result.

Corollary 2.3 *Let K_i be the compact set of absolutely continuous functions x in $C[0,1]$ with $x(0) = 0$ and $\int_0^1 \dot{x}^2(u)du \le \sigma_i^2$ and set $\zeta_{i,t}(u) = (2t \ln\ln t)^{-1/2}\Xi_i(tu)$, $u \in [0,1]$. Then the family $\zeta_{i,t}$, $t \ge 3$ is relatively compact in the topology of uniform convergence and as $t \to \infty$ the set of all a.s. limit points of $\zeta_{i,t}$ coincides with K_i. Let K be the compact set of functions $x \in C[0,1]$ which can be written in the form*

$x(u)=\sum_{1\le i\le \ell} x_i(u)$ with $x_i\in K_i,\ i=1,\dots,\ell$. Set $\zeta_t(u)=(2t\ln\ln t)^{-1/2}\Xi(tu)$, $u\in[0,1]$. Then the family $\zeta_t,\ t\ge 3$ is relatively compact in the topology of uniform convergence and as $t\to\infty$ the set of all a.s. limit points of ζ_t is contained in K.

In order to understand our assumptions observe that $\varpi_{q,p}$ is non-increasing in q and non-decreasing in p. Hence, for any pair $p,q\ge 1$,

$$\varpi_{q,p}(n)\le\psi(n).$$

Furthermore, by the real version of the Riesz–Thorin interpolation theorem (see, for instance, [7], Sect. 9.3) if $\theta\in[0,1],\ 1\le p_0,p_1,q_0,q_1\le\infty$ and

$$\frac{1}{p}=\frac{1-\theta}{p_0}+\frac{\theta}{p_1},\ \frac{1}{q}=\frac{1-\theta}{q_0}+\frac{\theta}{q_1}$$

then

$$\varpi_{q,p}(n)\le 2(\varpi_{q_0,p_0}(n))^{1-\theta}(\varpi_{q_1,p_1}(n))^{\theta}.$$

Since, clearly, $\varpi_{q_1,p_1}\le 2$ for any $q_1\ge p_1$ it follows for pairs $(\infty,1)$, $(2,2)$ and (∞,∞) that for all $q\ge p\ge 1$,

$$\varpi_{q,p}(n)\le(2\alpha(n))^{\frac{1}{p}-\frac{1}{q}},\ \varpi_{q,p}(n)\le 2^{1+\frac{1}{p}-\frac{1}{q}}(\rho(n))^{1-\frac{1}{p}+\frac{1}{q}}$$
$$\text{and } \varpi_{q,p}(n)\le 2^{1+\frac{1}{p}}(\phi(n))^{1-\frac{1}{p}}.$$

We observe also that by the Hölder inequality for $q\ge p\ge 1$ and $\alpha\in(0,p/q)$,

$$\beta_q(r)\le 2^{1-\alpha}[\beta(p,r)]^{\alpha}\gamma^{1-\alpha}_{\frac{pq(1-\alpha)}{p-q\alpha}}$$

with γ_θ defined in (2.16). Thus, we can formulate Assumption 2.1 in terms of more familiar $\alpha,\ \rho,\ \phi$, and ψ–mixing coefficients and with various moment conditions.

The strategy of the proof of Theorem 2.2 consists of several steps. First, we split the sum $\Xi_i(t)$ into a sum of "big" and "small" growing blocks so that the total contribution of small block can be disregarded and their sole purpose is to provide sufficient separation between big blocks. Growing blocks will enable us to approximate their members by conditional expectations as in (2.3) with increasing precision which differs from [12] and is an important point in obtaining our estimates. In spite of the fact that big blocks still remain strongly dependent in our setup the technique of [12] enables us to treat them as if they were weakly dependent. Namely, employing appropriate estimates from [12] we construct a martingale approximation of sums of big blocks with an error sufficient for our purposes. Finally, we rely on the Skorokhod embedding of martingales into a Brownian motion and estimate the distance between the embedded process and the Brownian motion.

We observe that though the Skorokhod embedding preserves distribution of each one dimensional martingale it does not preserve, in general, joint distributions of several martingales when we employ it simultaneously to ℓ of them as in our case. By this reason we obtain strong approximations (2.20) for each $\Xi_i(t)$ but we do not obtain a strong approximation of the sum $\Xi(t)$ by the Gaussian process $\sum_{i=1}^{\ell}\sigma_i B_i(t)$ which according to [12] is the weak limit of processes $N^{-1/2}\Xi(Nt)$ as $N\to\infty$. In fact, this is connected with multidimensional strong approximation theorems where the Skorokhod embedding is not applicable while other methods employed usually in these circumstances do not seem to work in our nonconventional setup.

The conditions of Theorem 2.2 hold true for many important models. Let, for instance, ξ_n be a Markov chain on a space M satisfying the Doeblin condition (see, for instance, [10], p.p. 367–368) and $f_j,\ j=1,\dots,\ell$ be bounded measurable functions on the space of sequences $x=(x_i,\ i=0,1,2,\dots,\ x_i\in M)$ such that $|f_j(x)-f_j(y)|\le Ce^{-cn}$ provided $x=(x_i)$, $y=(y_i)$ and $x_i=y_i$ for all $i=0,1,\dots,n$ where $c,C>0$ do not depend on n and j. In fact, some polynomial decay in n will suffice here, as well. Let $X(n)=(X_1(n),\dots,X_\ell(n))$ with $X_j(n)=f_j(\xi_n,\xi_{n+1},\xi_{n+2},\dots)$ and take σ-algebras $\mathcal{F}_{kl},\ k<l$ generated by $\xi_k,\xi_{k+1},\dots,\xi_l$ then the conditions will be satisfied considering $\{\xi_n,\ n\ge 0\}$ with its invariant measure as a stationary process. In fact, our conditions hold true for a more general class of processes, in particular, for Markov chains whose transition operator has a spectral gap which leads to an exponentially fast decay of the ρ-mixing coefficient. In fact, these conditions already hold true under sufficiently fast polynomial decay of dependence coefficients and we refer the reader to [3] for more examples of fast mixing stochastic processes.

Remark 2.4 Formally, (2.4) requires some stationarity and, for instance, if we consider a Markov chain ξ_n satisfying the Doeblin condition but whose initial distribution differs from its invariant measure then (2.4) does not hold true for $X(n)=f(\xi_n)$. Still, a slight modification makes our method to work so that Theorem 2.2 remains valid. In order to do this we consider another probability measure Π on the space $(\Omega,\mathcal{F})$ and require the weak stationarity (2.4) with respect to Π, i.e. $X(n)\Pi=\mu$ and $(X(n),X(n'))\Pi=\mu_{n-n'}$. In addition, we modify the definition of the dependence coefficient $\varpi_{q,p}$ in (2.1) taking the conditional expectation of g there with respect to the probability P while the unconditional expectation of g taking with respect to Π. It is easy to see that under the same assumptions as above but with modified (2.1) and (2.4) our proof will still go through.

Important classes of processes satisfying our conditions come from dynamical systems. Let T be a C^2 Axiom A diffeomorphism (in particular, Anosov) in a neighborhood of an attractor or let T be an expanding C^2 endomorphism of a Riemannian manifold M (see [2]), f_j's be either Hölder continuous functions or functions which are constant on elements of a Markov partition and let $X(n)=(X_1(n),\dots,X_\ell(n))$ with $X_j(n)=f_j(T^n x)$. Here the probability space is $(M,\mathcal{B},\mu)$ where μ is a Gibbs invariant measure corresponding to some Hölder continuous function and $\mathcal{B}$ is the Borel σ-field. Let ζ be a finite Markov partition for T then we can take $\mathcal{F}_{kl}$ to be the finite σ-algebra generated by the partition $\cap_{i=k}^{l}T^i\zeta$. In fact, we can take here not only Hölder continuous f_j's but also indicators of sets from $\mathcal{F}_{kl}$. A related example corresponds to T being a topologically mixing subshift of finite type which means that T is the left shift on a subspace Ξ of the space of one-sided sequences $\xi=(\xi_i,i\ge 0)$, $\xi_i=1,\dots,l_0$ such that $\xi\in\Xi$ if $\pi_{\xi_i\xi_{i+1}}=1$ for all $i\ge 0$ where $\Pi=(\pi_{ij})$ is an $l_0\times l_0$ matrix with 0 and 1 entries and such that Π^n for some n is a matrix with positive entries. Again, we have to take in this case f_j to be Hölder continuous bounded functions on the sequence space above, μ to be a Gibbs invariant measure corresponding to some Hölder continuous function and to define $\mathcal{F}_{kl}$ as the finite σ-algebra generated by cylinder sets with fixed coordinates having numbers from k to l. The exponentially fast ψ-mixing is well known in the above cases (see [2]). Among other dynamical systems with exponentially fast ψ-mixing we can mention also the Gauss map $Tx = \{1/x\}$ (where $\{\cdot\}$ denotes the fractional part) of the unit interval with respect to the Gauss measure G (see [8]). The latter enables us to consider the number $N_a(x, n)$, $a = (a_1, \dots, a_\ell)$ of m's between 0 and n such that the $q_j(m)$-th digit of the continued fraction of x equals certain integer a_j, $j = 1, \dots, \ell$. Then Theorem 2.2 implies certain strong approximation theorem and a law of iterated logarithm for $N_a(x, n)$ considered as a random variable on the probability space $((0, 1], \mathcal{B}, G)$. In fact, our results rely only on sufficiently fast α- or ρ-mixing which holds true for wider classes of dynamical system, in particular, those with a spectral gap (such as many one dimensional not necessarily uniformly expanding maps) which ensures an exponentially fast ρ-mixing.

3 Blocks and martingale approximation

The following result which appears as Corollary 3.6 from [12] (improving in several respects Lemma 3.1 from [11]) will be a base for our estimates.

Proposition 3.1 *Let $\mathcal{G}$ and $\mathcal{H}$ be σ-subalgebras on a probability space $(\Omega, \mathcal{F}, P)$, X and Y be d-dimensional random vectors and $f = f(x, \omega)$, $x \in \mathbb{R}^d$ be a collection of random variables measurable with respect to $\mathcal{H}$ and satisfying*

$$\|f(x,\omega)-f(y,\omega)\|_q \le C_1(1+|x|^\iota+|y|^\iota)|x-y|^\kappa \text{ and } \|f(x,\omega)\|_q \le C_2(1+|x|^\iota) \tag{3.1}$$

where $g \ge 1$. Set $g(x) = Ef(x, \omega)$ and $\tilde{f}(x, \omega) = E(f(x, \cdot)|\mathcal{G})(\omega)$. Then

$$\|E(f(X,\cdot)|\mathcal{G})-g(X)\|_\upsilon \le c(1+\|X\|^{\iota+2}_{b(\iota+2)})(\varpi_{q,p}(\mathcal{G},\mathcal{H})+\|X-E(X|\mathcal{G})\|^\delta_q), \tag{3.2}$$

provided $\frac{1}{\upsilon} \ge \frac{1}{p}+\frac{1}{b}+\frac{\delta}{q}$ and $\kappa-\frac{d}{p} > \delta > 0$ with $c = c(C_1, C_2, \iota, \iota', \kappa, \delta, p, q, \upsilon, d) > 0$ depending only on parameters in brackets. Moreover, let $x = (v, z)$ and $X = (V, Z)$, where V and Z are d_1 and $d - d_1$-dimensional random vectors, respectively, and let $f(x, \omega) = f(v, z, \omega)$ satisfy (3.1) in $x = (v, z)$. Set $\tilde{g}(v) = Ef(v, Z(\omega), \omega)$. Then

$$\begin{aligned}\|E(f(V,Z,\cdot)|\mathcal{G}) - \tilde{g}(V)\|_\upsilon &\le c(1+\|X\|^{\iota+2}_{b(\iota+2)})\\ &\times\big(\varpi_{q,p}(\mathcal{G},\mathcal{H}) + \|V-E(V|\mathcal{G})\|^\delta_q + \|Z-E(Z|\mathcal{H})\|^\delta_q\big).\end{aligned} \tag{3.3}$$

Furthermore,

$$\begin{aligned}&\|\tilde{f}(X(\omega),\omega) - \tilde{f}(Y(\omega),\omega) - g(X) + g(Y)\|_\upsilon\\ &\quad\le c\,\varpi_{q,p}(\mathcal{G},\mathcal{H})\big(1+\|X\|^{\iota+2}_m + \|Y\|^{\iota+2}_m\big)\|X-Y\|^\delta_q.\end{aligned} \tag{3.4}$$

We will use the following notations

$$
\begin{aligned}
&F_{i,r,n}(x_1,x_2,\ldots,x_{i-1},\omega)=E\big(F(x_1,x_2,\ldots,x_{i-1},X(n))|\mathcal{F}_{n-r,n+r}\big),\\
&\quad X_r(n)=E\big(X(n)|\mathcal{F}_{n-r,n+r}\big),\ \ Y_i(q_i(n))=F_i(X(q_1(n)),\ldots,X(q_i(n)))\quad\text{and}\\
&\quad Y_i(j)=0\quad\text{if}\quad j\neq q_i(n)\quad\text{for any } n,\ \ Y_{i,r}(q_i(n))=F_{i,r,q_i(n)}(X_r(q_1(n)),\ldots,\\
&\quad X_r(q_{i-1}(n)),\omega)\quad\text{and}\quad Y_{i,r}(j)=0\quad if\quad j\neq q_i(n)\quad\text{for any } n.
\end{aligned}
\tag{3.5}
$$

Next, we fix some positive numbers $4\eta<2\theta<\tau<1/2$ which will be specified later on and following [14] introduce pairs of "big" and "small" increasing blocks defining for each i random variables $V_i(j)$ and $W_i(j)$ inductively so that

$$
\begin{aligned}
&V_i(1)=Y_{i,1}(q_i(1)),\ W_i(1)=Y_{i,1}(q_i(2)),\ a(1)=0,\ b(1)=1\text{ and for } j>1,\\
&\ a(j)=b(j-1)+[(j-1)^\theta],\ \ b(j)=a(j)+[j^\tau],\ \ r(j)=[j^\eta],\\
&V_i(j)=\sum_{a(j)<l\le b(j)}Y_{i,r(j)}(q_i(l))\text{ and } W_i(j)=\sum_{b(j)<l\le a(j+1)}Y_{i,r(j)}(q_i(l)).
\end{aligned}
\tag{3.6}
$$

Observe that unlike [12] but following [14] the parameter $r(j)$ grows with j increasing precision of conditional expectations approximations. Let $\nu_i(t)=\max\{j:\ b(j)+[j^\theta]\le t\}$ which is the number of full small blocks in the sum $\Xi_i(t)$. We will see that the small blocks $W_i(j),\ j=1,2,\ldots$ make negligible contributions to the sum Ξ_i and can be disregarded while the big blocks $V_i(j),\ j=1,2,\ldots$ are widely separated which enables us to exploit fully our mixing assumptions. Observe that unlike the sums appearing in standard limit theorems these big blocks are strongly (and not weakly) dependent but as in [12] we will see by means of Proposition 3.1 that only sufficient separation between $q_i(l)$ for different l's plays the role.

Next, set

$$
R_i(m)=\sum_{j=m+1}^{\infty}E\big(V_i(j)|\mathcal{G}_m\big)
\tag{3.7}
$$

and $M_i(m)=V_i(m)+R_i(m)-R_i(m-1)$ where $\mathcal{G}_m=\mathcal{F}_{-\infty,q_i(b(m))+r(m)}$. Observe that if $a(j)<l\le b(j)$ and $j\ge m+1$ then $X=\big(X_{r(j)}(q_1(l)),\ldots,X_{r(j)}(q_{i-1}(l))\big)$ is $\mathcal{F}_{-\infty,q_{i-1}(l)+r(j)}$ measurable while $f(x,\omega)=F_{i,r(j),q_i(l)}(x_1,\ldots,x_{i-1},\omega)$ is $\mathcal{F}_{q_i(l)-r(j),\infty}$ measurable. Hence, by (3.2) considered with $\mathcal{G}=\mathcal{F}_{-\infty,\max(q_{i-1}(l)+r(j),q_i(b(m))+r(m))}$ and $\mathcal{H}=\mathcal{F}_{q_i(l)-r(j),\infty}$ we obtain that

$$
\|E\big(Y_{i,r(j)}(q_i(l))|\mathcal{G}_m\big)\|_{2+\delta}\le C\big(\varpi_{q,p}(d_{i,j}(l))+\beta_q^\delta(d_{i,j}(l))\big)
\tag{3.8}
$$

where p and q satisfy conditions of Proposition 3.1 with $\upsilon=2+\delta$ and Assumption 2.1, $C>0$ does not depend on $i,\,j,\,l,\,m$ and

$$
\begin{aligned}
d_{i,j}(l)&=\min(q_i(l)-q_{i-1}(l)-2r(j),\ q_i(l)-q_i(b(m))-r(j)-r(m))\\
&\ge l-b(m)-2r(j)\ge a(j)-b(m)-2r(j)
\end{aligned}
\tag{3.9}
$$

taking into account that under our assumptions

$$q_i(l) - q_{i-1}(l) \geq l \quad \text{and} \quad q_i(l) - q_i(m) \geq l - m \tag{3.10}$$

provided l is large enough. Thus, for $j \geq m+1$,

$$\begin{aligned} &\left\| E(V_i(j)|\mathcal{G}_m)\right\|_{2+\delta} \\ &\leq C \sum_{a(j)<l\leq b(j)} \left(\varpi_{q,p}(l-b(m)-2r(j)) + \beta_q^\delta(l-b(m)-2r(j))\right) \end{aligned} \tag{3.11}$$

and

$$\begin{aligned} \|R_i(m)\|_{2+\delta} \leq C \sum_{j=m+1}^{\infty} \sum_{a(j)<l\leq b(j)} &\big(\varpi_{q,p}(l-b(m)-2r(j)) \\ &+\beta_q^\delta(l-b(m)-2r(j))\big) \leq \tilde{C} < \infty \end{aligned} \tag{3.12}$$

for some constant $\tilde{C} > 0$. In particular, the series (3.7) converges in $L^{2+\delta}(\Omega, \mathcal{F}, P)$, and so the definition of $R_i(m)$ makes sense. Observe also that $M_i(m)$ is $\mathcal{G}_m$ measurable and

$$E(M_i(m)|\mathcal{G}_{m-1}) = E(V_i(m) + R_i(m)|\mathcal{G}_{m-1}) - R_i(m-1) = 0$$

which means that $(M_i(m), \mathcal{G}_m),\ m = 1, 2, \dots$ is a martingale difference sequence.

We are going to replace the sum $\Xi_i(t)$ by the martingale $\sum_{1\leq m\leq \nu_i(t)} M_i(m)$ and it will be crucial for our purposes to estimate the corresponding error. In order to make the first step in this direction we set

$$I_1(m) = \sum_{1\leq j\leq m} (V_i(j) - M_i(j))$$

and relying on (3.12) it follows that

$$\left\|I_1(m)\right\|_2 = \|R_i(\nu_i(m))\|_2 + \|R_i(0)\|_2 \leq 2\tilde{C} \tag{3.13}$$

for some constant $\tilde{C} > 0$. By Chebyshev's inequality

$$P\left\{|I_1(m)| \geq \frac{1}{2}m^{\frac{1}{2}+\varepsilon}\right\} \leq 16\tilde{C}^2 m^{-(1+2\varepsilon)}. \tag{3.14}$$

Observe that

$$[t] \geq \sum_{1\leq j\leq \nu_i(t)} [j^\tau] \geq \int_0^{\nu_i(t)} u^\tau du = (1+\tau)^{-1}(\nu_i(t))^{1+\tau},$$

and so

$$\nu_i(t) \leq \big((1+\tau)[t]\big)^{1/1+\tau} \leq 2[t]^{1/1+\tau}. \tag{3.15}$$

Hence, taking $m = \nu_i([t]) = \nu_i(t)$ and $\varepsilon = \frac{1}{2}\left(\frac{1}{2}\tau - \theta\right) + \frac{1}{2}\tau\left(\frac{1}{2} - \theta\right) > 0$ we obtain by (3.14) and (3.15) that

$$\begin{aligned} P\{|I_1(\nu_i([t]))| \geq [t]^{\frac{1}{2}(1-\theta)}\} \leq P\Bigg\{&|I_1(\nu_i([t]))| \\ &\geq \left(\frac{1}{2}\nu_i([t])\right)^{\frac{1}{2}+\varepsilon}\Bigg\} \leq 16\tilde{C}^2(\nu_i([t]))^{-(1+2\varepsilon)}. \end{aligned} \tag{3.16}$$

Therefore, by the Borel–Cantelli lemma we conclude that for some $n_0 = n_0(\omega)$ and all $\nu_i(t) \geq n_0$,

$$|I_1(\nu_i(t))| \leq t^{\frac{1}{2}(1-\theta)} \quad \text{a.s.} \tag{3.17}$$

Observe that

$$t < 2\sum_{1\leq j\leq \nu_i(t)+1} [j^\tau] \leq \int_0^{\nu_i(t)+1} (u^\tau + 1)du \leq 4\tau^{-1}(\nu_i(t)+1)^{1+\tau},$$

and so

$$\nu_i(t) \geq (\tau t/2)^{1/1+\tau} - 1. \tag{3.18}$$

Hence, if $t \geq 2\tau^{-1}(n_0+1)^{1+\tau}$ then (3.17) holds true.

Next, set

$$I_2(m) = \sum_{a(m+1)\leq l<a(m+2)} |Y_i(q_i(l))|.$$

Since $a(\nu_i(t)+1) \leq t < a(\nu_i(t)+2)$ then

$$\left|\sum_{a(\nu_i(t))\leq l\leq t} Y_i(q_i(l))\right| \leq I_2(\nu_i(t)). \tag{3.19}$$

By (2.6) and (2.19),

$$\|Y_i(q_i(l))\|_{2+\delta} \leq C < \infty \tag{3.20}$$

for some $C > 0$ independent of l and since by the construction $a(m+2) - a(m+1) = (m+1)^\tau + (m+1)^\theta$ we see that

$$\|I_2(m)\|_{2+\delta} \le \sum_{a(m+1)\le l<a(m+2)} \|Y_i(q_i(l))\|_{2+\delta} \le 2C(m+1)^\tau. \tag{3.21}$$

By (3.15) and Chebyshev's inequality

$$P\{|I_2(\nu_i(t))| \ge [t]^{\frac{1}{2}(1-\varepsilon)}\} \le P\left\{|I_2(\nu_i(t))| \ge \left(\frac{1}{2}\nu_i(t)\right)^{\frac{1}{2}(1+\tau)(1-\varepsilon)}\right\} \le \tilde{C}(\nu_i(t))^{-1-\beta} \tag{3.22}$$

for some $\tilde{C} > 0$ independent of t where we assume that $\tau \le \frac{1}{4}\min(\delta, 1)$ and take $\varepsilon = \frac{1}{16}\min(\delta, 1)$, $\beta = \frac{1}{128}\min(\delta, 1)$. As in (3.17) we conclude using (3.18) and the Borel–Cantelli lemma that

$$I_2(\nu_i(t)) \le t^{\frac{1}{2}(1-\varepsilon)} \quad \text{a.s.} \tag{3.23}$$

for all $t \ge t_0$ and some random variable $t_0 = t_0(\omega) < \infty$.

Next, we estimate contribution of the small blocks. Let $l > j$ then $W_i(j)$ is measurable with respect to $\mathcal{G} = \mathcal{F}_{-\infty, q_i(a(j+1))+r(j)}$ provided $j^\eta > 2$, and so applying (3.2) with such $\mathcal{G}$, $f(x_1, \dots, x_{i-1}, \omega) = F_{i,r(l),q_i(n)}(x_1, \dots, x_{i-1}, \omega)$ where $b(l) < n \le a(l+1)$, $\mathcal{H} = \mathcal{F}_{q_i(b(l))-r(l),\infty}$ we obtain by (2.6), (2.19), (3.8) and (3.10) that for n large enough,

$$\begin{aligned} \left|EW_i(j)Y_{i,r(l)}(q_i(n))\right| &= \left|E\left(W_i(j)E(Y_{i,r(l)}(q_i(n))|\mathcal{G})\right)\right| \le C_1\|W_i(j)\|_2 \\ &\times\left(\varpi_{q,p}(n - r(l) - a(j+1) - r(j)) + \beta_q^\delta(n - r(l) - a(j+1) - r(j))\right) \end{aligned} \tag{3.24}$$

for some $C_1 > 0$ independent of j, n, l satisfying the conditions above. Since by (2.6), (2.19) and the definition of blocks,

$$\|W_i(j)\|_2 \le \sum_{b(j)<l\le a(j+1)} \|Y_{i,r(j)}(q_i(l))\|_2 \le C_2[j^\theta] \tag{3.25}$$

for some $C_2 > 0$ independent of j, then

$$\begin{aligned} |E\,(W_i(j)W_i(l))\,| &\le C_1C_2[j^\theta][l^\theta] \\ &\times\left(\varpi_{q,p}\left(\sum_{j<m\le l}([m^\tau] - 2[m^\eta])\right) + \beta_q^\delta\left(\sum_{j<m\le l}([m^\tau] - 2[m^\eta])\right)\right). \end{aligned} \tag{3.26}$$

Hence, by (2.17), (2.18), (3.25) and (3.26) for any positive integers $m < n$,

$$E\left(\sum_{m<l\le n} W_i(l)\right)^2 \le \sum_{m<l\le n}\left(EW_i^2(l) + 2\sum_{m<j<l}|E(W_i(j)W_i(l))|\right)$$
$$\le C_3 \sum_{m<l\le n} l^{2\theta} \le C_4(n^{1+2\theta} - m^{1+2\theta}) \tag{3.27}$$

for some $C_3, C_4 > 0$ independent of m and n. It follows by Theorem A1 from [14] together with (3.15) that

$$\left|\sum_{1\le m\le \nu_i(t)} W_i(j)\right| \ll (\nu_i(t))^{\frac{1}{2}+\theta}\log^3\nu_i(t) \le 2t^{\frac{1}{2}-\varepsilon} \quad \text{a.s.} \tag{3.28}$$

where $\varepsilon < \left(\frac{1}{2}\tau - \theta\right)(1+\tau)^{-1}$ and t is large enough.

Next, set

$$I_3(m) = \left|\sum_{1\le j\le m}\sum_{a(j)<l\le a(j+1)}\left(Y_i(q_i(l)) - Y_{i,r(j)}(q_i(l))\right)\right|$$

We observe that (2.3), (2.5), (2.19) and Hölder's inequality lead to a simplified version of (3.4) which yields (cf. Lemma 3.12 in [12]) that,

$$\|Y_i(q_i(l)) - Y_{i,r(j)}(q_i(l))\|_2 \le C\beta_q^\delta(r(j)) \tag{3.29}$$

for some $q, \delta > 0$ satisfying (2.19) and for a constant $C > 0$ independent of j. Hence, if $\tau \le 4\eta + \delta$ then by (2.18),

$$\|I_3(\nu_i(t)\|_2 \le \tilde{C} < \infty \tag{3.30}$$

for some constant $\tilde{C} > 0$ independent of t. Proceeding in the same way as in (3.17) we obtain that for some random variable $t_0 = t_0(\omega)$,

$$|I_3(\nu_i(t))| \le t^{\frac{1}{2}(1-\theta)} \quad \text{a.s.} \tag{3.31}$$

whenever $t \ge t_0$. Finally, collecting (3.17), (3.23), (3.28) and (3.31) we conclude that

$$|\Xi_i(t) - \sum_{1\le j\le \nu_i(t)} M_i(j)| \ll t^{\frac{1}{2}-\varepsilon} \tag{3.32}$$

for some $\varepsilon > 0$.

4 Completing the proof via Skorokhod embedding

A martingale version of the Skorokhod embedding (representation) theorem (see [16], Theorem 4.3 and [9], Theorem A1) applied to our martingale $\mathcal{M}_i(m) = \sum_{1\le j\le m} M_i(j)$ yields that if $\{B_i(t),\ t \ge 0\}$ is a standard Brownian motion then there exist non-negative random variables $T_j = T_{i,j}$ such that the processes

$$\left\{ B_i\left(\sum_{1\le j\le m} T_j\right),\ m \ge 1\right\} \qquad \{\mathcal{M}_i(m),\ m \ge 1\} \tag{4.1}$$

have the same distributions. Hence, without loss of generality we can redefine $\{M_i(j),\ j \ge 1\}$ by

$$M_i(m) = B_i\left(\sum_{1\le j\le m} T_j\right) - B_i\left(\sum_{1\le j\le m-1} T_j\right) \tag{4.2}$$

and can keep the same notations for both $M_i(m)$ and $\mathcal{M}_i(m)$. In fact, we will redefine also the processes $X(n)$, $V_i(m)$, $W_i(m)$ we had before on a richer and common with $M_i(m)$ probability space so that all marginal and joint distributions remain intact. Furthermore, the embedding theorem cited above yields that if $\mathcal{A}_m$ is the σ-algebra generated by $\{B_i(t),\ 0 \le t \le \sum_{1\le j\le m} T_j\}$ then T_m is $\mathcal{A}_m$ measurable, $B_i\left(\sum_{1\le j\le m} T_j + s\right) - B_i\left(\sum_{1\le j\le m} T_j\right)$ is independent of $\mathcal{A}_m$ for any $s > 0$,

$$E(T_m|\mathcal{A}_{m-1}) = E(M_i^2(m)|\mathcal{A}_{m-1}) = E(M_i^2(m)|\mathcal{G}_{m-1}) = E(M_i^2(m)|\tilde{\mathcal{G}}_{m-1}) \tag{4.3}$$

and

$$E(T_m^u|\mathcal{A}_{m-1}) \le c_u E(|M_i(m)|^{2u}|\mathcal{A}_{m-1}) \tag{4.4}$$

where $c_u > 0$ depends only on $u \ge 1$, $\mathcal{A}_m \supset \mathcal{G}_m \supset \tilde{\mathcal{G}}_m = \sigma\{M_i(j),\ 1 \le j \le m\}$ and $\mathcal{G}_m$ is the same as in (3.7).

In order to exploit the representation

$$\mathcal{M}_i(m) = B_i\left(\sum_{1\le j\le m} T_j\right) \tag{4.5}$$

we have to establish a strong law of large numbers with appropriate error estimates for sums of T_j's in the form

$$\left|\sum_{1\le t\le \nu_i(t)} T_j - \sigma_i^2 t\right| = O(t^{1-\lambda}) \quad \text{a.s.} \tag{4.6}$$

for some $\lambda > 0$ and $\sigma_i \geq 0$. This would imply that

$$\left| B_i \left(\sum_{1 \leq j \leq \nu_i(t)} T_j \right) - B_i(\sigma_i^2 t) \right| \ll t^{\frac{1}{2} - \tilde{\lambda}} \quad \text{a.s.} \tag{4.7}$$

for some $\tilde{\lambda} < \frac{1}{2}\lambda$. Indeed, set $\tau_i(t) = \sum_{1 \leq j \leq \nu_i(t)} T_j$. Then (4.6) means that $|\tau_i(t) - \sigma_i^2 t| \leq Q t^{1-\lambda}$ for some random variable $Q = Q(\omega) < \infty$ a.s. Introducing the events $\Omega_N = \{Q \leq N\}$ we obtain

$$A(t) = |B_i(\tau_i(t)) - B_i(\sigma_i^2 t)| \mathbb{I}_{\Omega_N} \leq A_1(t) + A_2(t) + A_3(t)$$

where

$$A_1(t) = \sup_{0 \leq s \leq N t^{1-\lambda}} |B_i(\sigma_i^2 t + s) - B_i(\sigma_i^2 t)|,$$
$$A_2(t) = |B_i(\sigma_i^2 t) - B_i(\sigma_i^2 t - N t^{1-\lambda})| \ \text{ and}$$
$$A_3(t) = \sup_{0 \leq s \leq N t^{1-\lambda}} |B_i(\sigma_i^2 t - N t^{1-\lambda} + s) - B_i(\sigma_i^2 t - N t^{1-\lambda})|.$$

By the martingale moment inequalities for the Brownian motion

$$E A_j^{2m}(t) \leq C_m N^m t^{m(1-\lambda)}, \ \ j = 1, 2, 3$$

where $C_m > 0$ depends only on $m \geq 1$. Thus

$$P\{A(n) > n^{\frac{1}{2} - \tilde{\lambda}}\} \leq 3^{2m-1} C_m N^m n^{-m(\lambda - 2\tilde{\lambda})}.$$

Choose $\tilde{\lambda} < \frac{1}{2}\lambda$ and $m \geq 2(\lambda - 2\tilde{\lambda})^{-1}$ then $n^{-m(\lambda - 2\tilde{\lambda})} \leq n^{-2}$, and so the probabilities above form a converging series. Hence, by the Borel–Cantelli lemma there exists $n_0 = n_0(\omega) < \infty$ such that

$$A(n) \leq n^{\frac{1}{2} - \tilde{\lambda}} \quad \text{a.s. for all} \ \ n \geq n_0.$$

Since $\nu_i(t)$, and so also $\tau_i(t)$, can change only at integer t and since $\Omega_N \uparrow \tilde{\Omega}$ as $N \uparrow \infty$ with $P(\tilde{\Omega}) = 1$ we conclude that, indeed, (4.6) implies (4.7). Finally, redefining without changing distributions all processes once again we can replace $B_i(\sigma^2 t)$ by $\sigma_i B_i(t)$ arriving at the assertion of Theorem 2.2.

We start deriving (4.6) by writing

$$\sum_{1 \leq j \leq m} (T_j - M_i^2(j)) = D^{(1)}(m) - D^{(2)}(m), \tag{4.8}$$

where

$$D^{(1)}(m) = \sum_{1 \leq j \leq m} (T_j - E(T_j | \mathcal{A}_{j-1})), \ \ D^{(2)}(m) = \sum_{1 \leq j \leq m} (M_i^2(j) - E(M_i^2(j) | \mathcal{G}_{j-1})),$$

and using (4.3) in order to have (4.8). Set $R^{(1)}(j) = T_j - E(T_j|\mathcal{A}_{j-1})$ then $(R^{(1)}(j), \mathcal{A}_j)_{j\geq 1}$ is a martingale differences sequence. By (3.12), (3.20) and (4.4) for any $j \geq 1$,

$$\begin{aligned} E|R^{(1)}(j)|^{1+\frac{1}{2}\delta} &\leq 2E|T_j|^{1+\frac{1}{2}\delta} \leq 2c_{1+\frac{1}{2}\delta}E|M_i(j)|^{2+\delta} \\ &\leq C(1+E|V_i(j)|^{2+\delta}) \leq \tilde{C}j^{(2+\delta)\tau} \end{aligned}$$

for some $C, \tilde{C} > 0$ independent of j. Observe that $(j^{-(1+\tau+\varepsilon)}R^{(1)}(j), \mathcal{A}_j)_{j\geq 1}$ is also a martingale differences sequence and assume that $\tau \leq \delta/4$ and $\tau + \varepsilon \leq 1/4$. Then

$$\sum_{j=1}^{\infty} j^{-(1+\frac{1}{2}\delta)(1+\tau-\varepsilon)}E|R^{(1)}(j)|^{1+\frac{1}{2}\delta} \leq \tilde{C}\sum_{j=1}^{\infty} j^{-(1+\frac{\delta}{8})} < \infty,$$

and so by the standard result on martingale series (see Theorem 2.17 in [9]),

$$\sum_{1\leq j\leq\infty} j^{-(1+\tau-\varepsilon)}R^{(1)}(j) \quad \text{converges a.s.}$$

Hence, by Kronecker's lemma

$$m^{-(1+\tau-\varepsilon)}\sum_{j=1}^{m} R^{(1)}(j) = m^{-(1+\tau-\varepsilon)}D^{(1)}(m) \to 0 \quad \text{a.s.} \quad \text{as } m \to \infty,$$

and so by (3.15),

$$t^{-(1-\frac{\varepsilon}{1+\tau})}|D^{(1)}(\nu_i(t))| \leq 4(\nu(t))^{-(1+\tau-\varepsilon)}|D^{(1)}(\nu_i(t))| \to 0 \text{ a.s. as } t \to \infty. \tag{4.9}$$

Setting $R^{(2)}(j) = M_i^2(j) - E(M_i^2(j)|\mathcal{G}_{j-1})$ we obtain that $(R^{(2)}(j), \mathcal{G}_j)_{j\geq 1}$ is a martingale differences sequence, as well, and by (3.12) and (3.20),

$$E|R^{(2)}(j)|^{1+\frac{1}{2}\delta} \leq 2E|M_i(j)|^{2+\delta} \leq C(1+E|V_i(j)|^{2+\delta}) \leq \tilde{C}j^{(2+\delta)\tau}$$

for some $C, \tilde{C} > 0$ independent of j. Thus, in the same way as above, we see that

$$t^{-(1-\frac{\varepsilon}{1+\tau})}|D^{(2)}(\nu_i(t))| \to 0 \text{ a.s. as } t \to \infty. \tag{4.10}$$

It follows from (4.8)–(4.10) that in order to obtain (4.6) it suffices to show that there exists $\sigma_i \geq 0$ such that

$$\left| \sum_{1\leq j\leq \nu_i(t)} M_i^2(j) - \sigma_i^2 t \right| = O(t^{1-\lambda}) \quad \text{a.s.} \tag{4.11}$$

for some $\lambda > 0$. By the definition of $M_i(j)$ and the Cauchy inequality,

$$\left|\sum_{1\le j\le m}(M_i^2(j)-V_i^2(j))\right|\le (A_i(m))^{1/2}\left(2\left(\sum_{1\le j\le m}V_i^2(j)\right)^{1/2}+(A_i(m))^{1/2}\right) \tag{4.12}$$

where $A_i(m)=\sum_{1\le j\le m}\rho_i^2(j)$, $\rho_i(j)=R_i(j)-R_i(j-1)$ and $E|A_i(m)|\le m\tilde{C}^2$ by (3.12). Fix $\beta>0$ and for each $l\ge 1$ set $m_l=[l^{2/\beta}]$ then by Chebyshev's inequality

$$P\{|A_i(m)|\ge m_l^{1+\beta}\}\le \tilde{C}^2 m_l^{-\beta}\le \tilde{\tilde{C}}l^{-2} \tag{4.13}$$

for some $\tilde{\tilde{C}}>0$ independent of l. Therefore, by the Borel–Cantelli lemma for all $l\ge l_0=l_0(\omega)<\infty$,

$$|A_i(m_l)|<m_l^{1+\beta}\quad \text{a.s.}$$

If $m_l\le \nu_i(t)<m_{l+1}$ and $l\ge l_0$ then by (3.15),

$$|A_i(\nu_i(t))|\le |A_i(m_{l+1})|<m_{l+1}^{1+\beta}<(\nu_i(t))^{1+\beta}\left(\frac{m_{l+1}}{m_l}\right)^{1+\beta}\le Ct^{\frac{1+\beta}{1+\tau}}\ \text{a.s.}$$

where $C>0$ does not depend on l. Choosing $\beta=\tau/2$ we obtain that

$$A_i(\nu_i(t))\le Ct^{1-\frac{\tau}{2(1+\tau)}}\quad \text{a.s.} \tag{4.14}$$

for all $t\ge t_0=t_0(\omega)<\infty$ where in view of (3.18) we can take $t_0=\frac{2}{\tau}((l_0+1)^{2/\beta}+2)^{1+\tau}$.

It follows from (4.12) and (4.14) that in order to obtain (4.11) it remains to show that

$$\left|\sum_{1\le j\le \nu_i(t)}V_i^2(j)-\sigma_i^2 t\right|=O(t^{1-\lambda})\quad \text{a.s.} \tag{4.15}$$

for some $\lambda>0$. Next, we will make yet another reduction showing that (4.15) will follow if

$$\left|\left(\sum_{1\le j\le t}Y_i(q_i(l))\right)^2-\sigma_i^2 t\right|=O(t^{1-\lambda})\quad \text{a.s.} \tag{4.16}$$

for some $\lambda>0$. A transition from (4.16) to (4.15) proceeds in the same way as in Lemma 7.3.5 of [14] but for readers' convenience we sketch also here the corresponding argument.

First, we write

$$\left| E\left(\sum_{1\le l\le t} Y_i(q_i(l))\right)^2 - \sum_{1\le j\le \nu_i(t)} V_i^2(j)\right| \le J_1(t) + J_2(\nu_i(t)) \tag{4.17}$$

where

$$J_1(t) = \left| E\left(\sum_{1\le l\le t} Y_i(q_i(l))\right)^2 - \sum_{1\le j\le \nu_i(t)} EV_i^2(j)\right|$$

and

$$J_2(m) = \left|\sum_{1\le j\le m}\big(V_i^2(j) - EV_i^2(j)\big)\right|.$$

Next,

$$J_1(t) \le J_{11}(\nu_i(t)) + J_{12}(\nu_i(t)) + J_{13}(\nu_i(t)) + J_{14}(\nu_i(t)) + J_{15}(\nu_i(t)) \tag{4.18}$$

where

$$J_{11}(m) = 2\sum_{1\le j<\tilde{j}\le m} |EV_i(j)V_i(\tilde{j})|,\quad J_{12}(m) = E\left(\sum_{1\le j\le m} W_i(j)\right)^2,$$
$$J_{13}(m) = E(I_2(m))^2,\quad J_{14}(m) = E(I_3(m))^2 \text{ and}$$
$$J_{15}(m) = 2\left\|\sum_{1\le j\le m} V_i(j)\right\|_2 \big(J_{12}^{1/2}(m) + J_{13}^{1/2}(m) + J_{14}^{1/2}(m)\big)$$

with I_2 and I_3 the same as in (3.21) and (3.30), respectively. Using (3.8)–(3.11) we obtain similarly to (3.24)–(3.27) that

$$\begin{aligned} J_{11}(m) \le 2C &\sum_{1\le j<\tilde{j}\le m} \|V_i(j)\|_2 \sum_{a(\tilde{j})<l\le b(\tilde{j})} \big(\varpi_{q,p}(l-b(j)-2r\tilde{j}) \\ &+\beta_q^\delta(l-b(j)-2r\tilde{j})\big) \le 2C \sum_{1\le j<\tilde{j}\le m} \tilde{j}^\tau \sum_{a(\tilde{j})<l\le b(\tilde{j})} \\ &\times\big(\varpi_{q,p}(l-b(j)-2r\tilde{j})) + \beta_q^\delta(l-b(j)-2r\tilde{j}))\big) \le \tilde{C}m \end{aligned} \tag{4.19}$$

for some $C, \tilde{C} > 0$ independent of m. For $J_{12}(m)$, $J_{13}(m)$ and $J_{12}(m)$ we already have appropriate estimates in (3.27), (3.21) and (3.30), respectively. Employing (3.2) from Proposition 3.1 together with Assumption 2.1 in order to estimate $a_{ln} = |EY_i(q_i(l))Y_i(q_i(n))|$ we see (see (4.31) and (4.34) below as well as Lemma 5.1 from [12]) that $\sum_{1\le l<n\le t} a_{ln}$ is of order $O(t)$, and so for $m \le \nu_i(t)$,

$$E\left(\sum_{1\le j\le m} V_i(j)\right)^2 \le \sum_{1\le l\le t} Y_i^2(q_i(l)) + 2\sum_{1\le l<n\le t} a_{ln} \le Ct \tag{4.20}$$

for some $C > 0$ independent of t. Combining (3.15), (3.21), (3.27), (3.30) and (4.18)–(4.20) we obtain that

$$J_{11}(\nu_i(t)) \le \tilde{C}t^{1-\varepsilon} \tag{4.21}$$

for some $\tilde{C} > 0$ independent of t where $\varepsilon = (\tau - 2\theta)/(1+\tau)$.

In order to estimate $J_2(t)$ we set

$$U_i(j) = \begin{cases} V_i^2(j) - EV_i^2(j) & \text{if } \ |V_i^2(j) - EV_i^2(j)| \le j^{1+\sigma} \\ 0 & \text{otherwise} \end{cases}$$

where $\sigma \in [\tau, \frac{\delta}{4}]$ will be further specified later on. Observe that

$$\begin{aligned} P\{U_i(j) \ne V_i^2(j) - EV_i^2(j)\} &= P\{|V_i^2(j) - EV_i^2(j)| > j^{1+\sigma}\} \\ &\le 2^{1+\frac{\delta}{2}} j^{-(1+\sigma)(1+\frac{\delta}{2})} \le 2^{1+\frac{\delta}{2}} j^{-(1+\frac{\delta}{2})(1+\sigma-2\tau)}. \end{aligned} \tag{4.22}$$

Since $\sigma \ge 2\tau$ then the power of j in the right hand side of (4.22) is less than -1, and so by the Borel–Cantelli lemma with probability one the event $\{U_i(j) \ne V_i^2(j) - EV_i^2(j)\}$ can occur only finite number of times. Hence, the asymptotical behavior as $m \to \infty$ of $J_2(m)$ and of $J_3(m) = |\sum_{1\le j\le m} U_i(j)|$ is the same (up to a random variable independent of m) and it suffices to estimate the latter. Set $U_i^*(j) = U_i(j) - EU_i(j)$. Using (3.11) we obtain that for $j < j'$,

$$|EU_i^*(j)U_i^*(j')| \le (jj')^{1+\sigma}\left(\varpi_{q,p}\left(j^\theta + \sum_{j<m\le j'} m^\tau\right) + \beta_q^\delta\left(j^\theta + \sum_{j<m\le j'} m^\tau\right)\right). \tag{4.23}$$

Next,

$$E(U_i^*(j))^2 \le 2j^{(1+\sigma)(1-\frac{\delta}{2})} E|U_i^*(j)|^{1+\frac{\delta}{2}} \le 32j^{(1+\sigma)(1-\frac{\delta}{2})} E|V_i(j)|^{2+\delta} \le Cj^{1+2\tau-\varepsilon} \tag{4.24}$$

for some $C > 0$ independent of j where $\varepsilon = \frac{\delta}{2} - \sigma + \frac{\delta\sigma}{2} - \tau\delta$ and we choose σ and τ so small that $\varepsilon \ge \delta/8$. It follows from (2.17), (2.18), (4.23) and (4.24) that for some $\tilde{C} > 0$ independent of n and m,

$$E\left(\sum_{j=m+1}^{n}U_i^*(j)\right)^2\le\tilde{C}(n^{2+2\tau-\varepsilon}-m^{2+2\tau-\varepsilon})\tag{4.25}$$

and applying again Theorem A1 from [14] we obtain by (3.15) and (4.25) that

$$\left|\sum_{1\le j\le\nu_i(t)}U_i^*(j)\right|\ll(\nu_i(t))^{1+\tau-\frac{1}{2}\varepsilon}\le 2t^{1-\frac{\varepsilon}{2(1+\tau)}}\text{ a.s.}\tag{4.26}$$

Hence, $J_2(\nu_i(t))\ll t^{1-\frac{\varepsilon}{2(1+\tau)}}$ a.s., as well.

Finally, it remains to establish (4.16). In fact, existence of the limit

$$\lim_{t\to\infty}t^{-1}E\left(\sum_{1\le n\le t}Y_i(q_i(n))\right)^2=\sigma_i^2$$

and its computation is given in Propositions 4.1 and 4.5 from [12] and we only have to explain the estimate (4.16) which is actually hidden inside the proof there. If $i\le k$ then the above limit has the form (see Proposition 4.1 in [12]),

$$\sigma_i^2=\sum_{l=-\infty}^{\infty}a_i(l)\text{ with }a_i(l)=\int F_i(x_1,\dots,x_i)F_i(y_1,\dots,y_i)\prod_{1\le u\le i}d\mu_{ul}(x_u,y_u)\tag{4.27}$$

where μ_n is the same as in (2.4) and $d\mu_0(x,y)=\delta_{xy}d\mu(x)$ is the measure supported by the diagonal. If $i>k$ then (see Proposition 4.5 in [12]),

$$\sigma_i^2=\int F_i^2(x_1,\dots,x_i)d\mu(x_1)\cdots d\mu(x_i).\tag{4.28}$$

We have

$$E\left(\sum_{1\le n\le t}Y_i(q_i(n))\right)^2=\sum_{1\le n,n'\le t}b_i(n,n')\tag{4.29}$$

where

$$b_i(n,n')=EF_i\big(X(q_1(n)),\dots,X(q_i(n))\big)F_i(X(q_1(n')),\dots,X(q_i(n'))\big).$$

If $i\le k$ then for each integer m we consider $b_i(n,n')$ with $in-in'=im$. Assume that $|im|\le\frac{1}{4}\max(n,n')$ then we can apply (3.3) of Proposition 3.1 with $\mathcal{G}=\mathcal{F}_{-\infty,(i-3/4)\max(n,n')}$, $\mathcal{H}=\mathcal{F}_{(i-1/2)\max(n,n'),\infty}$, $V=\big(X(n),\dots,X(i-1)n;X(n'),\dots,X((i-1)n')\big)$ and $Z=(X(in),X(in'))$ which gives that

$$\Big|b_i(n,n') - \int EF_i(X(n),\dots,X((i-1)n),x)$$
$$\times F_i(X(n'),\dots,X((i-1)n'),y)d\mu_{im}(x,y)\Big|$$
$$\le C_1\left(\varpi_{q,p}\left(\frac{1}{4}\max(n,n')\right)+\beta_q^\delta\left(\frac{1}{4}\max(n,n')\right)\right)$$

for some $C_1 > 0$ independent of n, n'. Repeating these estimates i times we obtain that

$$|b_i(n,n') - a_i(m)| \le C_1\left(\varpi_{q,p}\left(\frac{1}{4}\max(n,n')\right)+\beta_q^\delta\left(\frac{1}{4}\max(n,n')\right)\right).$$

If $|im| > \max(n,n')$, say $im > \max(n,n')$, then applying (3.3) of Proposition 3.1 with $\mathcal{G} = \mathcal{F}_{-\infty,\max(in',(i-1)n)+\frac{1}{16}n}$, $\mathcal{H} = \mathcal{F}_{(i-1/16)n,\infty}$, $V = (X(n),\dots,X((i-1)n); X(n'),\dots,X(in'))$ and $Z = X(in)$ which yields that

$$|b_i(n,n')| \le C_2(\varpi_{q,p}(n/16)+\beta_q^\delta(n/16))$$

for some $C_2 > 0$ independent of n. The same estimate holds true if $im < -n/4$ with n' in place of n, and so we can replace above n by $\max(n,n')$. Next, we want to show that a similar estimate holds true for $a_i(m)$ when $|im| > \max(n,n')$, assuming first that $im > \max(n,n')$. Since for $in - in' = im$,

$$a_i(m) = \int EF_i(x_1,\dots,x_{i-1},X(in))F_i(y_1,\dots,y_{i-1},X(in'))\prod_{1\le u\le i-1} d\mu_{um}(x_u,y_u)$$

we can apply (3.3) of Proposition 3.1 with $\mathcal{G} = \mathcal{F}_{-\infty,in'+\frac{1}{16}n}$, $\mathcal{H} = \mathcal{F}_{(i-\frac{1}{16})n,\infty}$, $V = (x_1,\dots,x_{i-1}; y_1,\dots,y_{i-1},X(in'))$ and $Z = X(in)$ which yields that

$$|a_i(m)| \le C_3(\varpi_{q,p}(n/16)+\beta_q^\delta(n/16))$$

for some $C_3 > 0$ independent of n. If $im < -\frac{n}{4}$ then we obtain a similar estimate with n replaced by n', and so we can replace n in the above estimate by $\max(n,n')$. Collecting the above estimates we obtain that if $i \le k$ and $in - in' = im$ for an integer m then

$$|b_i(n,n') - a_i(m)| \le C_4\left(\varpi_{q,p}\left(\frac{1}{16}\max(n,n')\right)+\beta_q^\delta\left(\frac{1}{16}\max(n,n')\right)\right) \tag{4.30}$$

for some $C_4 > 0$ independent of n. By (2.17), (2.18) and (4.30) we obtain that for $i \le k$,

$$\left|\sum_{1\le n,n'\le t} b_i(n,n') - \sigma_i^2\right| \le C_5 t^{1-\delta} \tag{4.31}$$

for some $C_5 > 0$ independent of t.

Next, we consider the case $i \geq k+1$. It follows from (2.8) that if $n \neq n'$ and $\max(n, n')$ is large enough then $|q_i(n) - q_i(n')| \geq (\max(n, n'))^{\delta}$. Hence, relying on (3.3) in Proposition 3.1 it is easy to see similarly to above that in this case

$$|b_i(n, n')| \leq C_6 \left(\varpi_{q,p} \left(\frac{1}{4} |q_i(n) - q_i(n')| \right) + \beta_q^{\delta} \left(\frac{1}{4} |q_i(n) - q_i(n')| \right) \right) \tag{4.32}$$

for some $C_6 > 0$ independent of n and n'. In order to estimate the difference between $b_i(n, n)$ and σ_i^2 from (4.28) we use that $q_i(n) - q_{i-1}(n) \geq n$ for large n which yields that

$$\begin{aligned} &|b_i(n, n) - \int E F_i^2(X(n), \ldots, X((i-1)n), x) d\mu(x)| \\ &\leq C_7 \left(\varpi_{q,p} \left(\frac{1}{4} n \right) + \beta_q^{\delta} \left(\frac{1}{4} n \right) \right) \end{aligned}$$

for some $C_7 > 0$ independent of n where we rely on (3.3) from Proposition 3.1 with $\mathcal{G} = \mathcal{F}_{-\infty, (i-\frac{3}{4})n}$, $\mathcal{H} = \mathcal{F}_{(i-\frac{1}{4})n, \infty}$, $V = (X(n), \ldots, X((i-1)n))$ and $Z = X(in)$. Repeating this estimate i times we obtain that

$$\left| \sum_{0 \leq n \leq t} b_i(n, n) - t\sigma_i^2 \right| \leq C_8 \sum_{0 \leq n \leq t} \left(\varpi_{q,p} \left(\frac{1}{4} n \right) + \beta_q^{\delta} \left(\frac{1}{4} n \right) \right) \tag{4.33}$$

for some $C_8 > 0$ independent of t. This together with (2.8), (2.17), (2.18) and (4.32) yields that

$$\left| \sum_{0 \leq n, n' \leq t} b_i(n, n') - t\sigma_i^2 \right| \leq C_9 t^{1-\delta} \tag{4.34}$$

for some $C_9 > 0$ independent of t. Finally, (4.29), (4.31) and (4.34) yields (4.16) completing the proof of Theorem 2.2. □

References

1. Bergelson, V.: Weakly mixing PET. Ergod. Th. Dynam. Sys. **7**, 337–349 (1987)
2. Bowen, R.: Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lecture Notes in Math., vol. 470. Springer, Berlin (1975)
3. Bradley, R.C.: Introduction to Strong Mixing Conditions. Kendrick Press, Heber City (2007)
4. Berkes, I., Philipp, W.: Approximation theorems for independent and weakly dependent random vectors. Ann. Probab. **7**, 29–54 (1979)
5. Chatterjee, S.: A new approach to strong embeddings. Probab. Theory Rel. Fields (to appear)
6. Furstenberg, H.: Nonconventional ergodic averages. Proc. Symp. Pure Math. **50**, 43–56 (1990)
7. Garling, D.J.H.: Inequalities: a Journey into Linear Analysis. Cambridge Univ. Press, Cambridge (2007)
8. Heinrich, L.: Mixing properties and central limit theorem for a class of non-identical piecewise monotonic C^2-transformations. Mathematische Nachricht. **181**, 185–214 (1996)

9. Hall, P., Heyde, C.C.: Martingale Limit Theory and its Application. Acad. Press, New York (1980)
10. Ibragimov, I.A., Linnik, Yu.V.: Independent and Stationary Sequences of Random Variables. Wolters–Noordhoff, Groningen (1971)
11. Kifer, Yu.: Nonconventional limit theorems. Probab. Th. Rel. Fields **148**, 71–106 (2010)
12. Kifer, Yu., Varadhan, S.R.S.: Nonconventional limit theorems in discrete and continuous time via martingales (Preprint, 2010). arXiv:1012.2223
13. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent RV's and the sample DF. II. Z. Warsch. Verw. Geb. **34**, 33–56 (1976)
14. Philipp, W., Stout, W.: Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables. Memoirs Amer. Math. Soc., vol. 161. Amer. Math. Soc., Providence (1975)
15. Strassen, V.: An invariance principle for the law of the iterated logarithm. Z. Wahrsch. Verw. Geb. **3**, 211–226 (1964)
16. Strassen, V.: Almost sure behavior of sums of independent random variables and martingales. In: Proc. Fifth Berkeley Symp. Math. Stat. Probab., vol.II, Part 1, pp 315–343 (1967)