A NONCONVENTIONAL STRONG LAW OF LARGE NUMBERS AND FRACTAL DIMENSIONS OF SOME MULTIPLE RECURRENCE SETS

YURI KIFER

INSTITUTE OF MATHEMATICS
THE HEBREW UNIVERSITY OF JERUSALEM

Abstract. We provide conditions which yield a strong law of large numbers for expressions of the form $1/N\sum_{n=1}^{N}F\big(X(q_1(n)),\cdots,X(q_\ell(n))\big)$ where $X(n),n\geq 0$'s is a sufficiently fast mixing vector process with some moment conditions and stationarity properties, F is a continuous function with polinomial growth and certain regularity properties and $q_i, i>m$ are positive functions taking on integer values on integers with some growth conditions. Applying these results we study certain multifractal formalism type questions concerning Hausdorff dimensions of some sets of numbers with prescribed asymptotic frequencies of combinations of digits at places $q_1(n),...,q_\ell(n)$.

1. Introduction

Nonconventional ergodic theorems which attracted substantial attention in ergodic theory (see, for instance, [3], [12] and [2]) studied the limits of expressions having the form $1/N\sum_{n=1}^{N}T^{q_1(n)}f_1\cdots T^{q_\ell(n)}f_\ell$ where T is a weakly mixing measure preserving transformation, f_i's are bounded measurable functions and q_i's are polynomials taking on integer values on the integers. While, for instance, [3] and [12] were interested in L^2 convergence, other papers such as [2] provided conditions for almost sure convergence in such ergodic theorems. Originally, these results were motivated by applications to multiple recurrence for dynamical systems taking functions f_i being indicators of some measurable sets.

Introducing stronger mixing or weak dependence conditions enabled us in [20] and [21] to obtain central limit theorems and invariance principles for even more general expressions of the form

$$\frac{1}{\sqrt{N}}\sum_{n=1}^{[Nt]}\big(F(X(q_1(n)),...,X(q_\ell(n))-\bar{F}\big) \tag{1.1}$$

where $X(n)$, $n\geq 0$ is a sufficiently fast mixing vector valued process with some moment conditions and stationarity properties, F is a locally Hölder continuous function with polinomial growth, $\bar{F}=\int Fd(\mu\times\cdots\times\mu)$ and μ is the distribution

Date: December 14, 2010.
2000 *Mathematics Subject Classification.* Primary: 60F15 Secondary: 37C45, 37A30, 60G48, 37D35.
Key words and phrases. strong law of large numbers, nonconventional ergodic averages, mixingales, dynamical systems.

of $X(0)$. In order to ensure existence of limiting variances and covariances we had to impose another assumption concerning the functions $q_j(n),\ j\geq 1$ saying that $q_j(n)=jn$ for $j=1,...,k$ while $q_j(n),\ j\geq k$ are positive functions taking on integer values on integers with some (faster than linear) growth conditions.

In this paper we are concerned with strong laws of large numbers (SLLN) for expressions of the form

$$\frac{1}{N}\sum_{n=1}^{N}F(X(q_1(n)),...,X(q_\ell(n)) \tag{1.2}$$

which can be proved under milder conditions that those required for central limit theorem type results. We still impose some mixing or weak dependence conditions but now the functions $q_j(n),\ n\geq 1$ are allowed to be of much more general form than in [21], in particular, because we do not have to take care about limiting variances. Recall, that the machinery of nonconventional ergodic theorems employed in [3], [12], [2] and other papers can only work when the functions $q_j,\ j=1,...,\ell$ are polinomials while our methods do not require any algebraic structure of them. We pay a price for this, namely, imposing stronger mixing assumptions which are satisfied though for important classes of stochastic processes and dynamical systems.

In order to obtain our strong laws of large numbers we represent the sum in (1.2) as a sum of certain mixingales and then rely on the SLLN for mixingales obtained in [22]. Another approach which works in this situation under more or less the same assumptions is a martingale approximation similar to [21] together with a SLLN for martingales (see, for instance, Section 2.6 in [16]).

Among more specific applications of our setup we can consider $F(x_1,...,x_\ell)=x_1^{(1)}\cdots x_\ell^{(\ell)},\ x_j=(x_j^{(1)},...,x_j^{(\ell)}),\ X(n)=(X_1(n),...,X_|ell(n)),\ X_j(n)=\mathbb{I}_{A_j}(T^nx)$ for a dynamical system $\{T^n\}$ or $X_j(n)=\mathbb{I}_{A_j}(\xi_n)$ for a Markov chain $\{\xi_n\}$ where $\mathbb{I}_A$ is the indicator of a set A. Then the expression (1.2) measures the frequency of arrivals of T^nx or of ξ_n to the sets A_j at the respective times $q_j(n)$. Recall, that the m-base and continued fraction expansions can be obtained via the multiplication by m and the Gauss transformations, i.e. $Tx=\{mx\}$ and $Tx=\{1/x\}$, respectively, which are both exponentially fast ψ-mixing with respect to many invariant measures (see [15] and [1]) and satisfy our assumptions. Denote by $\zeta_j(x)$ the j-th digit of $x\in[0,1)$ in one of these expansions. Then we can study the frequency of k-th such that the ℓ-tuple $(\zeta_{q_1(k)}(x),...,\zeta_{q_\ell(k)}(x))$ coincides with a prescribed ℓ-tuple of digits $(a_1,...,a_\ell)$. For a full Lebesgue measure of points $x\in[0,1)$ such frequencies are determined by our SLLN and other frequencies may occur only for x belonging to sets of zero measure. This leads to an interesting question about Hausdorff dimensions of such exceptional sets which we study in the last section of this paper.

2. Preliminaries and main results

Our setup consists of a $\wp$-dimensional stochastic process $\{X(n),n=0,1,...\}$ on a probability space $(\Omega,\mathcal{F},P)$ and of a family of σ-algebras $\mathcal{F}_{kl}\subset\mathcal{F},\ 0\leq k\leq l\leq\infty$ where we assume that $\mathcal{F}_{00}$ is a trivial σ-field and $\mathcal{F}_{kl}\subset\mathcal{F}_{k'l'}$ if $k'\leq k$ and $l'\geq l$. We extend $\mathcal{F}_{kl}$ also to negative $k\geq-\infty$ by defining $\mathcal{F}_{kl}=\mathcal{F}_{0l}$ for $k<0$ and $l\geq 0$. The dependence between two sub σ-algebras $\mathcal{G},\mathcal{H}\subset\mathcal{F}$ is measured often via the quantities

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

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 [9], Ch. 4),

$$\alpha(\mathcal{G},\mathcal{H})=\tfrac{1}{4}\varpi_{\infty,1}(\mathcal{G},\mathcal{H}),\ \rho(\mathcal{G},\mathcal{H})=\varpi_{2,2}(\mathcal{G},\mathcal{H})$$
$$\phi(\mathcal{G},\mathcal{H})=\tfrac{1}{2}\varpi_{\infty,\infty}(\mathcal{G},\mathcal{H}) \text{ and } \psi(\mathcal{G},\mathcal{H})=\varpi_{1,\infty}(\mathcal{G},\mathcal{H}).$$

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

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

Furthermore, we do not require stationarity of the process $X(n),n\geq 0$ assuming only that the distribution μ of $X(n)$ does not depend on n which we write for further references by

$$X(n)\overset{d}{\sim}\mu\tag{2.4}$$

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

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

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

and

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

Our assumptions on F are motivated by the desire to include, for instance, products $F(x_1,...,x_\ell)=x_{11}x_{22}\cdots x_{\ell\ell}$, where $x_i=(x_{i1},...,x_{i\ell})\in\mathbb{R}^\ell$, which are important in the study of multiple recurrence as described in Introduction.

Our setup includes also a sequence of positive functions $q_1(n)<q_2(n)<\cdots<q_\ell(n)$ taking on integer values on integers and such that for some positive $\varepsilon\leq 1$,

$$q_i(n)\geq q_{i-1}(n)+\varepsilon n,\ i=2,...,\ell \text{ and } q_i(n+1)\geq q_i(n)+\varepsilon \text{ for all } n\geq 1.\tag{2.7}$$

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

$$F=F_0+F_1(x_1)+\cdots+F_\ell(x_1,x_2,\dots,x_\ell)\tag{2.8}$$

where

$$F_0=\bar{F}=\int F(x_1,...,x_\ell)\,d\mu(x_1)\cdots d\mu(x_\ell),\tag{2.9}$$

$$F_i(x_1,\dots,x_i)=\int F(x_1,x_2,\dots,x_\ell)\,d\mu(x_{i+1})\cdots d\mu(x_\ell) \\ -\int F(x_1,x_2,\dots,x_\ell)\,d\mu(x_i)\cdots d\mu(x_\ell) \tag{2.10}$$

for $0<i<\ell$ 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

$$\int F_i(x_1,x_2,\dots,x_{i-1},x_i)\,d\mu(x_i)\equiv 0 \quad \forall \quad x_1,x_2,\dots,x_{i-1}. \tag{2.11}$$

These enable us to write

$$S(N)=\sum_{n=1}^{N}F(X(q_1(n)),...,X(q_\ell(n)))=\sum_{i=0}^{\ell}S_i(N) \tag{2.12}$$

where $S_0(N)=N\bar F$ and for $1\le i\le\ell$,

$$S_i(N)=\sum_{1\le n\le N}F_i(X(q_1(n)),\dots,X(q_i(n))). \tag{2.13}$$

Following [22] we say that a sequence $\{a_n,\,n\ge 0\}$ is of size $-1/2$ if there exists a positive eventually nondecreasing sequence $\{L_n,\,n\ge 0\}$ such that

$$\sum_{n\ge 0}(nL_n)^{-1}<\infty,\; L_n-L_{n-1}=O(L_n/n) \text{ and } a_n=O\big((n^{1/2}L_n)^{-1}\big).$$

For instance, any sequence with asymptotics $O\big(n^{1/2}\log n(\log\log n)^{1+\delta}\big)^{-1}$ for some $\delta>0$ is of size $-1/2$. For each $r>0$ set

$$\gamma_r^r=\|X\|_r^r=E|X(n)|^r=\int\|x\|^r d\mu. \tag{2.14}$$

Our main result relies on

2.1. **Assumption.** With $d=(\ell-1)\wp$ there exist $p,q\ge 1$ and $\theta,m>0$ such that $\theta<\kappa-\frac{d}{p}$,

$$\frac{1}{2}\ge\frac{1}{p}+\frac{\iota+2}{m}+\frac{\theta}{q} \text{ and } \gamma_m+\gamma_{2q(\iota+2)}<\infty \tag{2.15}$$

and the sequence $\varpi_{p,q}(n)+\beta_q^\theta(n)$, $n\ge 1$ is of size $-1/2$.

2.2. **Theorem.** *Suppose that Assumption 2.1 holds true. Then with probability one*

$$\lim_{N\to\infty}\frac{1}{N}S(N)=\bar F. \tag{2.16}$$

Our method relies on estimates from [21] which enable us to view for each $i\ge 1$ the sequence of pairs $\{F_i(X(q_1(n)),...,X(q_i(n))),\,\mathcal{F}_{-\infty,q_i(n)}\}_{n=1}^\infty$ as a mixingale sequence, and so a strong law of large numbers for mixingales from [22] can be employed. This gives an almost sure convergence of $\frac{1}{N}S_i(N)$ to 0 and by (2.12) Theorem 2.2 follows. Another approach which works in our situation is to rely on a martingale approximation of $S_i(N)$ similarly to [21] and then to employ a strong law of large numbers for martingales (see, for instance, Section 2.6 in [16]). This method has to deal with approximations of $F_i(X(q_1(n)),...,X(q_i(n)))$ by their conditional expectations and in order to avoid double limits as in [21] we can make this approximations with increasing in n precision.

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 $p, q \geq 1$,

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

Furthermore, by the real version of the Riesz–Thorin interpolation theorem (see, for instance, [14], Section 9.3) if $\delta \in [0,1]$, $1 \leq p_0, p_1, q_0, q_1 \leq \infty$ and

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

then

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

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

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

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

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

with γ_r defined in (2.14). Thus, we can formulate Assumption 2.1 in terms of more familiar α, ρ, ϕ, and ψ–mixing coefficients and with various moment conditions.

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, [17], p.p. 367–368) and f_j, $j = 1, ..., \ell$ be bounded measurable functions on the space of sequences $x = (x_i,\ i = 0, 1, 2, ...,\ x_i \in M)$ such that $|f_j(x) - f_j(y)| \leq Ce^{-cn}$ provided $x = (x_i)$, $y = (y_i)$ and $x_i = y_i$ for all $i = 0, 1, ..., n$ where $c, C > 0$ do not depend on n and j. In fact, some polinomial decay in n will suffice here, as well. Let $X(n) = (X_1(n), ..., X_\ell(n))$ with $X_j(n) = f_j(\xi_n, \xi_{n+1}, \xi_{n+2}, ...)$ and take σ-algebras $\mathcal{F}_{kl}$, $k < l$ generated by $\xi_k, \xi_{k+1}, ..., \xi_l$ then our condition will be satisfied considering $\{\xi_n,\ n \geq 0\}$ 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 probability has a spectral gap which leads to an exponentially fast decay of the ρ-mixing coefficient.

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 compact Riemannian manifold M (see [8]), f_j's be Hölder continuous functions and let $X(n) = (X_1(n), ..., X_\ell(n))$ 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 $\xi = (\xi_i, i \geq 0), \xi_i = 1, ..., l_0$ such that $\xi \in \Xi$ if $\pi_{\xi_i \xi_{i+1}} = 1$ for all $i \geq 0$ 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 [8]). 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 $G(\Gamma) = \frac{1}{\ln 2}\int_\Gamma \frac{1}{1+x}dx$ (see [15]), as well as with respect to many other Gibbs invariant measures (see [1]). The latter enables us to consider the number $N_a(x,n)$, $a = (a_1, ..., a_\ell)$ of k's between 0 and n such that the $q_j(k)$-th digit of the continued fraction of x equals certain integer $a_j, j = 1, ..., \ell$. Then Theorem 2.2 implies a strong law of large numbers 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. We will show how to derive from Theorem 2.2 the following result.

2.3. **Corollary.** *Let T be either a C^2 Axiom A diffeomorphism on a compact Riemannian manifold M considered in a neighborhood of an attractor or a C^2 expanding endomorphisms of a compact Riemannian manifold M or the Gauss map of the unit interval and let μ be an equilibrium state (Gibbs measure) corresponding to a Hölder continuous function in the first two cases or an exponentially fast ψ-mixing T-invariant (in particular, Gauss') measure (see Corollary 4.7.8 in* [1]*) in the latter case. Let $X_j(n) = f_j(T^n x)$, $j = 1, ..., \ell$ where f_j is either a continuous function or $f_j(x) = \mathbb{I}_{\Gamma_j}(x)$ where Γ_j is a measurable set whose boundary $\partial\Gamma_j$ has zero μ-measure. Finally, let $F = F(x_1, ..., x_\ell)$ satisfies conditions of Theorem 2.2 which means just that F is Hölder continuous since its arguments are bounded here. Then the conclusion of Theorem 2.2 holds true.*

Next, we discuss a continuous time version of our theorem. Our continuous time setup consists of a $\wp$-dimensional process $X(t)$, $t \geq 0$ on a probability space $(\Omega, \mathcal{F}, P)$ whose one dimensional distributions do not depend on time and of a family of σ-algebras $\mathcal{F}_{st} \subset \mathcal{F}$, $-\infty \leq s \leq t \leq \infty$ such that $\mathcal{F}_{st} \subset \mathcal{F}_{s't'}$ if $s' \leq s$ and $t' \geq t$. For all $t \geq 0$ we set

$$\varpi_{q,p}(t) = \sup_{s\geq 0} \varpi_{q,p}(\mathcal{F}_{-\infty,s}, \mathcal{F}_{s+t,\infty}) \tag{2.17}$$

and

$$\beta(p,t) = \sup_{s\geq 0} \|X(s) - E\big[X(s)|\mathcal{F}_{s-t,s+t}\big]\|_p. \tag{2.18}$$

where $\varpi_{q,p}(\mathcal{G}, \mathcal{H})$ is defined by (2.1). It will suffice for our purposes to rely on Assumtion 2.1 concerning $\varpi_{q,p}(t)$ and $\beta(p,t)$ considered only for integer t. Let $q_1(t) < q_2(t) < \cdots < q_\ell(t)$ be increasing positive functions satisfying the conditions (2.7) with t in place of n. Set

$$S(t) = \int_0^t F(X(q_1(s)), ..., X(q_\ell(s)))ds = \sum_{i=0}^{\ell} S_i(t) \tag{2.19}$$

where $S_0(t) = t\bar{F}$,

$$S_i(t) = \int_0^t F_i(X(q_1(s)), \dots, X(q_i(s)))ds \tag{2.20}$$

and F, $\bar{F}$, F_i are the same as in (2.5), (2.6) and (2.8)–(2.11). Then we obtain

2.4. **Corollary.** *Under the conditions above with probability one*

$$\lim_{t\to\infty} \frac{1}{t} S(t) = \bar{F}.$$

Next, we discuss the fractal dimensions part of this paper. Recall that the multifractal formalism deals with computations of Hausdorff dimensions of sets having the form

$$\{x : \lim_{n\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n x) = \rho\}.$$

In our setup it is natural to study Hausdorff dimensions of more general sets

$$G_\rho = \{x : \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} F(f_1(T^{q_1(n)}x), ..., f_\ell(T^{q_\ell(n)}x)) = \rho\},$$

say, under the conditions of Corollary 2.3. When

$$\rho = \int ... \int F(f_1(x_1), ..., f_\ell(x_\ell)) d\mu(x_1) \cdots d\mu(x_\ell)$$

then $\mu(G_\rho) = 1$ by Corollary 2.3 while otherwise $\mu(G_\rho) = 0$ and it is natural to inquire about the Hausdorff dimension of G_ρ.

We will not study here this general problem but consider a more specific question about Hausdorff dimensions of sets of numbers with prescribed frequencies of specific combinations of digits in m-expansions. Namely, for any $x \in [0, 1]$ and an integer $m > 1$ we can write

$$x = \sum_{i=1}^{\infty} \frac{a_{i-1}(x)}{m^i} \text{ where } a_j(x) \in \{0, 1, ..., m-1\},\ j = 0, 1, ...$$

and we allow zero tails of expansions but not tails consisting of all $(m-1)$'s. This convention affects only a countable number of points, and so it does not influence Hausdorff dimensions computations. For each $x \in [0, 1]$ and an ℓ-word $\alpha = (\alpha_1, \alpha_2, ..., \alpha_\ell) \in \{0, 1, ..., m-1\}^\ell$ define

$$N_\alpha(x, n) = \#\{k > 0, k \le n : (a_{q_1(k)}(x), ..., a_{q_\ell(k)}(x)) = \alpha\} \tag{2.21}$$

where $\#\Gamma$ denotes the number of elements in the set Γ. Denote by $\mathcal{A}_\ell = \{0, 1, ..., m-1\}^\ell$ the set of all ℓ-words and let $p_\alpha \ge 0$, $\alpha \in \mathcal{A}_\ell$ satisfy $\sum_{\alpha\in\mathcal{A}_\ell} p_\alpha = 1$. For such a probability vector $p = (p_\alpha,\ \alpha \in \mathcal{A}_\ell) \in \mathbb{R}^{m^\ell}$ define

$$U_p = \{x \in (0, 1) : \lim_{n\to\infty} \frac{1}{n} N_\alpha(x, n) = p_\alpha \text{ for all } \alpha \in \mathcal{A}_\ell\}. \tag{2.22}$$

We want to deal with the question of computation of the Hausdorff dimension $HD(U_p)$ of U_p. When $\ell = 1$ and $q_1(k) = k$ we arrive at the classical question studied in [3] and [10] by combinatorial means and in [6] via the ergodic theory.

In order to relate the limit of $n^{-1}N_\alpha(x, n)$ to the nonconventional strong law of large numbers (ergodic theorem) discussed before define the transformation $Tx = \{mx\}$ where $\{\cdot\}$ denotes the fractional part. Identifying 0 and 1 we can view T as an expanding map of the circle. Now $a_i(x)=a_0(T^ix)$ and if $\alpha=(\alpha_1,\alpha_2,...,\alpha_\ell)\in\mathcal{A}_\ell$ and $\Gamma_j=\{x:\,a_0(x)=j\}$ then

$$N_\alpha(x,n)=\sum_{k=1}^n\mathbb{I}_{\Gamma_{\alpha_1}}(T^{q_1(k)}x)\mathbb{I}_{\Gamma_{\alpha_2}}(T^{q_2(k)}x)\cdots\mathbb{I}_{\Gamma_{\alpha_\ell}}(T^{q_\ell(k)}x). \tag{2.23}$$

Taking into account that $\{\Gamma_j,\,j=0,1,...,m-1\}$ is the Markov partition for T in this simple situation we arrive at the setup of Corollary 2.3 with $F(x_1,...,x_\ell)=x_1x_2\cdots x_\ell$ and $f_j(x)=\mathbb{I}_{\Gamma_{\alpha_j}}(x),\,j=1,...,\ell$. Observe that in place of the dynamical systems setup described above we could rely in this situation on the fact that that the digits $a_n,\,n\geq 0$ are independent identically distributed (i.i.d.) random variables with respect to the Lebesgue measure on $[0,1]$, and so $\mathbb{I}_{\Gamma_{\alpha_j}}\circ T^n=\mathbb{I}_{a_n=\alpha_j},\,i=1,...,\ell,n=0,1,...$ are also i.i.d. random variables so that mixing conditions of Assumption 2.1 trivially hold true. The following result answers our question in a specific situation.

2.5. **Proposition.** *Suppose that $q_1(k)=k$ for all k and there exists a probability vector $r=(r_0,r_1,...,r_{m-1})$ such that $p_\alpha=\prod_{i=1}^\ell r_{\alpha_i}$ for any $\alpha=(\alpha_1,...,\alpha_\ell)\in\mathcal{A}_\ell$. Then*

$$HD(U_p)=\frac{-\sum_{j=0}^{m-1}r_j\ln r_j}{\ln m} \tag{2.24}$$

with the convention $0\ln 0=0$.

2.6. **Remark.** In view of (2.23) for any T-invariant probability measure μ on $[0,1]$ with mixing properties fulfilling conditions of Theorem 2.2 it follows that μ-almost everywhere

$$\lim_{n\to\infty}\frac{1}{n}N_\alpha(x,n)=\prod_{i=1}^\ell\mu(\Gamma_{\alpha_i}).$$

Hence, if $p=(p_\alpha,\,\alpha\in\mathcal{A}_\ell)$ and there exists no probability vector $r=(r_0,r_1,...,r_{m-1})$ such that $p_\alpha=\prod_{i=1}^\ell r_{\alpha_i}$ then $\mu(U_p)=0$ for any μ as above, and so such μ cannot be used for computation of the Hausdorff dimension of U_p (by one of methods where measures are involved) which complicates the study in this case.

Now, consider a bit more complex situation. For each $x\in[0,1]$ and $\alpha=(\alpha_1,\alpha_2,...,\alpha_\ell)$, $\beta=(\beta_1,\alpha_2,...,\beta_\ell)\in\{0,1,...,m-1\}^\ell$ set

$$N_{\alpha,\beta}(x,n)=\#\{k>0,k\leq n:\,(a_{q_1(k)}(x),...,a_{q_\ell(k)}(x))=\alpha \text{ and } (a_{q_1(k)+1}(x),...,a_{q_\ell(k)+1}(x))=\beta\}$$

and for each nonnegative matrix $P=(p_{\alpha\beta},\,\alpha,\beta\in\mathcal{A}_\ell)$ with $\sum_{\alpha,\beta}p_{\alpha\beta}=1$ define

$$U_P=\{x\in(0,1):\,\lim_{n\to\infty}\frac{1}{n}N_{\alpha,\beta}(x,n)=p_{\alpha,\beta}\text{ for all }\alpha,\beta\in\mathcal{A}_\ell\}. \tag{2.25}$$

Again, we can write $N_{\alpha\beta}$ in the form suitable for application of Theorem 2.2, namely,

$$N_{\alpha,\beta}(x,n)=\sum_{k=1}^n\mathbb{I}_{\Gamma_{\alpha_1\beta_1}}(T^{q_1(k)}x)\mathbb{I}_{\Gamma_{\alpha_2\beta_2}}(T^{q_2(k)}x)\cdots\mathbb{I}_{\Gamma_{\alpha_\ell\beta_\ell}}(T^{q_\ell(k)}x) \tag{2.26}$$

where $\Gamma_{ij}=\{x:\,a_0(x)=i,\,a_1(x)=j\}$. Then we obtain the following result.

2.7. **Proposition.** *Suppose that* $q_1(k) = k$ *and there exists a nonnegative matrix* $R = (r_{ij};\, i,j = 0,1,...,m-1)$ *satisfying the following conditions:*

(i) some power of R *is a positive matrix; (ii)* $\sum_{i,j} r_{ij} = 1$, $q_i = \sum_{j=0}^{m-1} r_{ij} = \sum_{j=0}^{m-1} r_{ij}$*; (iii)* $p_{\alpha\beta} = \prod_{i=1}^{\ell} r_{\alpha_i\beta_i}$.

Then $q = (q_0, q_1, ..., q_{m-1})$ *is a positive stationary vector of the* $m \times m$ *irredicible aperiodic probability matrix* $Q = (q_{ij})$, $q_{ij} = q_i^{-1} r_{ij}$ *and under the convention* $0 \ln 0 = 0$,

$$HD(U_P) = \frac{-\sum_{i,j=0}^{m-1} q_i q_{ij} \ln q_{ij}}{\ln m}. \tag{2.27}$$

2.8. **Remark.** Somewhat surprisingly Proposition 2.5 and 2.7 claim that in our circumstances the sets U_p and U_P have the same Hausdorff dimensions as if we were prescribing frequencies not of the whole ℓ-words or pairs of such words but just of their first digits or pairs of their first digits.

2.9. **Remark.** It is easy to see that unless $\sum_\beta p_{\alpha\beta} = \sum_\beta p_{\beta\alpha}$ the set U_P is empty, and so the condition (ii) in Proposition 2.7 is a necessary one.

Next, we consider a similar to Proposition 2.5 problem concerning integer digits $a_0(x), a_1(x), ... > 0$ of infinite continued fraction expansions

$$\cfrac{1}{a_0(x) + \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + ...}}}$$

for irrational numbers $x \in (0,1)$. We define again $N_\alpha(x,n)$ and U_p by (2.21) and (2.22) taking into account that now there are infinitely many words $\alpha = (\alpha_1, ..., \alpha_\ell) \in \{1,2,3,...\}^\ell = \mathcal{A}_\ell$ and, correspondingly, we have to prescribe infinitely many frequencies $p_\alpha \ge 0$, $\alpha \in \mathcal{A}_\ell$ with $\sum_{\alpha \in \mathcal{A}_\ell} p_\alpha = 1$. We recall that the Gauss map $Tx = \{\frac{1}{x}\}$ acts so that $a_i(Tx) = a_{i+1}(x)$, $i = 0,1,2,...$, and so $N_\alpha(x,n)$ can be represented again in the form (2.23). For each infinite probability vector $\bar{r} = (r_1, r_2, ...)$ denote by $\mathcal{N}(\bar{r})$ the set of T-invariant ergodic probability measures μ such that

$$\int |\log x| s\mu(x) < \infty \text{ and } \mu[(j+1)^{-1}, j^{-1}) = r_j \text{ for all } j \ge 1. \tag{2.28}$$

Here, $[(j+1)^{-1}, j^{-1}) = \{x \in (0,1) : a_0(x) = j\}$ and for any n we set $I(i_0, i_1, ..., i_{n-1}) = \{x \in (0,1) : a_0(x) = i_0, ..., a_{n-1}(x) = i_{n-1}\}$ which is called a rank-n basic interval. Denote by $\hat{\mathcal{N}}(\bar{r})$ the subset of $\mathcal{N}(\bar{r})$ consisting of measures ν such that for ν-almost all x and all $\alpha = (\alpha_1, ..., \alpha_\ell) \in \mathcal{A}_\ell$,

$$\lim_{n\to\infty} \frac{1}{n} N_\alpha(x,n) = \prod_{i=1}^{\ell} r_{\alpha_i}. \tag{2.29}$$

By (2.23) and (2.28) we see that $\hat{\mathcal{N}}(\bar{r})$ contains all measures $\nu \in \mathcal{N}(\bar{r})$ with sufficient mixing which make the process $X_\alpha(n) = X_\alpha(x,n) = (\mathbb{I}_{\Gamma_{\alpha_1}}(T^n x), ..., \mathbb{I}_{\Gamma_{\alpha_\ell}}(T^n x))$ on the probability space $((0,1), \nu)$ to satisfy conditions of Theorem 2.2. We observe that not only the Gauss measure $G(\Gamma) = \frac{1}{\ln 2}\int_\Gamma \frac{dx}{1+x}$, which is exponentially fast ψ-mixing according to [15], but also many other T-invariant Gibbs measures constructed in [23] have sufficiently good mixing properties to satisfy conditions of

Theorem 2.2. Actually, the rank-1 basic intervals form a Markov partition for T whose action is essentially equivalent to the full shift on a sequence space with infinite alphabet. For such Markov transformations Corollary 4.7.8 from [1] gives conditions for their Gibbs invariant measures to be exponentially fast ψ-mixing.

2.10. **Proposition.** *Suppose that $q_1(k)=k$ and there exists an infinite probability vector $\bar r=(r_0,r_1,...)$ such that $p_\alpha=\prod_{i=1}^\ell r_{\alpha_i}$ for any $\alpha=(\alpha_1,...,\alpha_\ell)\in\mathcal{A}_\ell$. Then*

(2.30)
$$\max\big(\frac{1}{2},\sup_{\nu\in\hat{\mathcal{N}}(\bar r)}\frac{h_\nu}{2\int|\ln x|d\nu(x)}\big)\leq HD(U_p)\leq\max\big(\frac{1}{2},\sup_{\nu\in\mathcal{N}(\bar r)}\frac{h_\nu}{2\int|\ln x|d\nu(x)}\big)$$

where h_μ denotes the entropy of T with respect to μ and "sup" is set to be zero if $\mathcal{N}(\bar r)=\emptyset$.

2.11. **Remark.** All results of this paper can be extended under appropriate conditions to random transformations and processes in random (dynamical) environment. Namely, suitable (random) mixing conditions can be introduced similarly to [19] and the corresponding relative strong law of large numbers can be proved relying on martingale approximations constructed combining methods of [19] and [21]. A relative version of Proposition 2.5 can be proved in the spirit of random base expansions from [18].

3. Mixingale representation and proof of SLLN

We rely on the following result which is part of Corollary 3.6 from [21].

3.1. **Lemma.** *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*

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

where $g\geq 1$. Set $g(x)=Ef(x,\omega)$. Then

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

provided $\kappa-\frac{d}{p}>\theta>0$, $\frac{1}{\upsilon}\geq\frac{1}{p}+\frac{1}{b}+\frac{\theta}{q}$ with $c=c(C,\iota,\kappa,\theta,p,q,\upsilon,d)>0$ 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

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

Set $\bar Y_i(n)=F_i(X(q_1(n)),...,X(q_i(n)))-EF_i(X(q_1(n)),...,X(q_i(n)))$ and denote $\mathcal{G}_n^{(i)}=\mathcal{F}_{-\infty,q_i(n)}$ for $n\geq 0$ while taking $\mathcal{G}_n^{(i)}$ to be the trivial σ-algebra $\{\emptyset,\Omega\}$ for $n<0$. Then by (2.5), (2.6), (2.15) and (3.3) of Lemma 3.1 we obtain that for some $C_1>0$ and all n,m and $i=1,...,\ell$,

$$\|E(\bar Y_i(n)|\mathcal{G}^{(i)}_{n-m})\|_2\leq C_1\big(\varpi_{p,q}(\rho_i(m,n))+\beta_q^\theta(\rho_i(m,n))\big) \tag{3.4}$$

where p,q,θ satisfy conditions of Assumption 2.1 and

$$\rho_i(m,n)=\min\big([\frac{q_i(n)-q_{i-1}(n)}{3},]\,[\frac{q_i(n)-q_i(n-m)}{3}]\big).$$

Observe that $E(\bar{Y}_i(n)|\mathcal{G}^{(i)}_{n-m})=0$ if $m>n$ and if $0\leq m\leq n$ then $\rho_i(m,n)\geq[\varepsilon m/3]$ by (2.7). Hence,

$$\|E(\bar{Y}_i(n)|\mathcal{G}^{(i)}_{n-m})\|_2\leq C_1\big(\varpi_{p,q}([\varepsilon m/3])+\beta_q^\theta([\varepsilon m/3])\big). \tag{3.5}$$

It follows also from (2.3) and (2.5)–(2.7) and the Hölder inequality (see Lemmas 4.1 and 4.2 together with Theorem 4.4 from [21]) that

$$\begin{aligned}\big\|\bar{Y}_i(n)-E(\bar{Y}_i(n)|\mathcal{G}_{n+m})\|_2\leq K\big\|\big(1+\textstyle\sum_{i=1}^\ell(|X(q_i(n))|^\iota \\ +|E(X(q_i(n))|\mathcal{G}^{(i)}_{n+m})|^\iota)\big)\textstyle\sum_{i=1}^\ell\big|X(q_i(n))-E(X(q_i(n))|\mathcal{G}^{(i)}_{n+m})\big|^\kappa\big\|_2\leq C_2\beta_q^\theta(m)\end{aligned} \tag{3.6}$$

for some $C_2>0$. The estimates (3.5) and (3.6) yield that $\bar{Y}_i(n)$, $n\geq 1$ is a mixingale sequence as defined in [22] and under Assumption 2.1 the conditions of Corollary 1.9 from there are satisfied yielding that with probability one for $i=1,...,\ell$,

$$\lim_{N\to\infty}\frac{1}{N}(\Xi_i(N)-E\Xi_i(N))=0. \tag{3.7}$$

Set $a_i(n)=(q_{i-1}(n)+q_i(n))/2$. By (2.11) and (3.3) we obtain that

$$\begin{aligned}\big|EF_i(X(q_1(n)),...,X(q_i(n)))\big|=\big|EE\big(F_i(X(q_1(n)),...,X(q_i(n)))|\mathcal{F}_{-\infty,a_i(n)}\big)\big| \\ \leq C\big(\varpi_{q,p}([\tfrac{1}{4}(q_i(n)-q_{i-1}(n))])+\beta_q^\delta([\tfrac14(q_i(n)-q_{i-1}(n))])\big)\to 0 \text{ as } n\to\infty\end{aligned} \tag{3.8}$$

by (2.7) and Assumption 2.1. It follows that for $i=1,...,\ell$,

$$\lim_{N\to\infty}\frac{1}{N}E\Xi_i(N)=0,\ \text{ and so }\ \lim_{N\to\infty}\frac{1}{N}\Xi_i(N)=0 \text{ a.s.} \tag{3.9}$$

by (3.7) yielding Theorem 2.2 in view of (2.12). □

In order to derive Corollary 2.3 we recall that Hölder continuous functions can be uniformly approximated by functions which are constant on elements $\cap_{i=-n}^n T^iG_{k_i}$ of the partition $\bigvee_{i=-n}^n T^i\zeta$ (where G_k are elements of a Markov partition ζ) with an error decaying exponentially fast in n. Thus Theorem 2.2 holds when $X_j(n)=f_j(T^nx)$ and f_j, $j=1,...,\ell$ are Hölder continuous. Then Theorem 2.2 holds true also for continuous functions f_j, $j=1,...,\ell$ since they can be uniformly approximated by Hölder continuous ones and F is Hölder continuous. Next, let $f_j=\mathbb{I}_{\Gamma_j}$ with $\mu(\partial\Gamma_j)=0$. Given a Markov partition ζ denote by $\tilde\Gamma_j^{(l)}$ the set consisting of elements of the partition $\bigvee_{i=-l}^l T^i\zeta$ which intersect Γ_j. Here we assume that Γ_j lie on a hyperbolic invariant set itself though the argument can be easily extended to a neighborhood of a hyperbolic attractor. For any $\delta>0$ there exists l_δ such that $\mu(\tilde\Gamma_j^{(l)}\setminus\Gamma_j)<\delta$ for each $l\geq l_\delta$. For such an l set $g_j=\mathbb{I}_{\tilde\Gamma_j^{(l)}}$, $j=1,...,\ell$. Since F is Hölder continuous we obtain that

$$\begin{aligned}\big|\textstyle\sum_{n=1}^N F(f_1(T^nx),...,f_\ell(T^nx))-\sum_{n=1}^N F(g_1(T^nx),...,g_\ell(T^nx))\big| \\ \leq C\textstyle\sum_{j=1}^\ell\sum_{n=1}^N(\mathbb{I}_{\tilde\Gamma_j^{(l)}}(T^nx)-\mathbb{I}_{\Gamma_j}(T^nx))\end{aligned} \tag{3.10}$$

and the remaining part of Corollary 2.3 follows by the ergodic theorem applied to the right hand side of (3.10). □

In the continuous time case we set

$$\bar{Y}_i(n)=\int_n^{n+1}\big(F_i(X(q_1(s)),...,X(q_\ell(s)))-EF_i(X(q_1(s)),...,X(q_\ell(s)))\big)ds \tag{3.11}$$

and check similarly to (3.4)–(3.6) that $\big(\bar Y_i(n),\mathcal{G}_n^{(i)}\big)_{n=1}^\infty$ is a mixingale sequence where $\mathcal{G}_n^{(i)}$ is the same as in (3.4)–(3.6). Now Corollary 2.4 follows from [22] as before. $\square$

4. Application to fractal dimensions

Set

$$N_i^{(1)}(x,n)=\#\{k>0,k\le n:\, a_k(x)=i\},$$
$$N_{ij}^{(1)}(x,n)=\#\{k>0,k\le n:\, a_k(x)=i,\, a_{k+1}(x)=j\},$$
$$U_r^{(1)}(x,n)=\{x\in(0,1):\, \lim_{n\to\infty}\frac 1nN_i^{(1)}(x,n)=r_i \text{ for all } i\} \text{ and}$$
$$U_R^{(1)}(x,n)=\{x\in(0,1):\, \lim_{n\to\infty}\frac 1nN_{ij}^{(1)}(x,n)=r_{ij} \text{ for all } i,j\}.$$

Since

$$N_i^{(1)}(x,n)=\sum_{\alpha_2,...,\alpha_n}N_{i\alpha_2...\alpha_n}(x,n) \text{ and}$$
$$N_{ij}^{(1)}(x,n)=\sum_{\alpha_2,...,\alpha_n,\beta_2,...,\beta_n}N_{i\alpha_2...\alpha_n;j\beta_2...\beta_n}(x,n)$$

then for

$$r_i=\sum_{\alpha_2,...,\alpha_n}p_{i\alpha_2...\alpha_n} \text{ and } r_{ij}=\sum_{\alpha_2,...,\alpha_n,\beta_2,...,\beta_n}p_{i\alpha_2...\alpha_n;j\beta_2...\beta_n}$$

we obtain that

$$U_p(x,n)\subset U_r^{(1)}(x,n) \text{ and } U_P(x,n)\subset U_R^{(1)}(x,n)$$

provided $p=(p_\alpha,\,\alpha\in\mathcal{A}_\ell)$ and $P=(p_{\alpha\beta},\,\alpha,\beta\in\mathcal{A}_\ell)$. Hence, the upper bounds of Propositions 2.5, 2.7 and 2.10 follow from the corresponding upper bounds from [5], [10] and [13]. Still, we provide below an argument yielding the upper bounds in Propositions 2.5 and 2.7 by the reason explained in Remark 4.1.

Denote by Ξ the space of sequences $\xi=(\xi_0,\xi_1,...)$ with $\xi_i\in\{0,1,...,m-1\}$ for all $i\ge 0$. For each probability vector $r=(r_0,r_1,...,r_{m-1})$ denote by $\mu_r=(r_0,r_1,...,r_{m-1})^{\mathbb N}$ the corresponding product measure on Ξ, i.e. the probability mesure which gives the weight $r_{\alpha_0}r_{\alpha_1}\cdots r_{\alpha_n}$ to each cylinder set $\Xi_{\alpha_0\alpha_1...\alpha_n}=\{\xi=(\xi_0,\xi_1,...)\in\Xi:\,\xi_i=\alpha_i \text{ for } i=0,1,...,n\}$. Observe that the map $\varphi:\Xi\to[0,1]$ acting by the formula $\varphi(\xi)=\sum_{i=1}^\infty m^{-i}\xi_{i-1}$ is one-to-one except for a countable set of points and since μ_r has no atoms φ maps μ_r to an atomless measure $\varphi\mu_r$ on $[0,1]$. Since μ_r is invariant with respect to the left shift $\theta:\Xi\to\Xi$ acting by $\theta(\xi)=\tilde\xi$ with $\tilde\xi_i=\xi_{i+1}$ then $\varphi\mu_r$ is invariant with respect to $Tx=\{mx\}$ and φ provides an isomorphism between (Ξ,μ_r,θ) and $([0,1],\varphi\mu_r,T)$. Clearly, the conditions of Theorem 2.2 are satisfied here and applying it (see also Remark 2.6) we conclude from (2.23) that for $\varphi\mu_r$ almost all $x\in[0,1]$,

$$\lim_{n\to\infty}\tfrac 1nN_\alpha(x,n)=\lim_{n\to\infty}\tfrac 1n\sum_{k=1}^n\mathbb{I}_{\Gamma_{\alpha_1}}(T^{q_1(k)}x)\mathbb{I}_{\Gamma_{\alpha_2}}(T^{q_2(k)}x)\times\cdots\times\mathbb{I}_{\Gamma_{\alpha_\ell}}(T^{q_\ell(k)}x)=\prod_{i=1}^\ell\varphi\mu_r(\Gamma_{\alpha_i})=\prod_{i=1}^\ell r_{\alpha_i}. \tag{4.1}$$

It follows that

$$\varphi\mu_r(U_p)=1 \quad\text{if}\quad p=(p_\alpha,\,\alpha\in\mathcal{A}_\ell)\quad\text{and}\quad p_\alpha=\prod_{i=1}^\ell r_{\alpha_i} \text{ whenever } \alpha=(\alpha_1,...,\alpha_\ell). \tag{4.2}$$

Suppose that $r_{i_j} > 0$, $j = 1, ..., k$ while $r_i = 0$ if $i \neq i_j$ for any j. Set

$$U^+ = \{x \in U_p : \, a_j(x) \in \{i_1, i_2, ..., i_k\} \text{ for any } j = 0, 1, 2...\}.$$

Then by (4.2) and the definition of μ_r,

$$\varphi\mu_r(U^+) = 1. \tag{4.3}$$

Observe that $I_{\alpha_0\alpha_1\ldots\alpha_{n-1}} = \varphi\Xi_{\alpha_0\alpha_1\ldots\alpha_{n-1}}$ is a subinterval of $[0,1]$ and let $I_n(x) = I_{\alpha_0\alpha_1\ldots\alpha_{n-1}}$ if $x \in I_{\alpha_0\alpha_1\ldots\alpha_{n-1}}$. Set $m_j(x,n) = \#\{i \geq 0, i < n : \, a_i(x) = j\}$. Then for any $x \in U^+$ we can write

$$\ln \varphi\mu_r(I_n(x)) = \sum_{j=0}^{m-1} m_j(x,n) \ln r_j. \tag{4.4}$$

Clearly, $|I_n(x)| = m^{-n}$ where $|I|$ denotes the length of I. Observe that if $x \in U_p$ with $p = (r_0, r_1, ..., r_{m-1})$ then

$$\lim_{n\to\infty} \frac{1}{n} m_j(x,n) = \lim_{n\to\infty} \frac{1}{n} \sum_{0 \leq \alpha_2,\ldots,\alpha_\ell \leq m-1} N_{j\alpha_2,\ldots,\alpha_\ell}(x,n) = r_j. \tag{4.5}$$

Hence, for any $x \in U^+$,

$$\lim_{n\to\infty} \frac{\ln \varphi\mu_r(I_n(x))}{\ln |I_n(x)|} = -\frac{\sum_{j=0}^{m-1} r_j \ln r_j}{\ln m} \tag{4.6}$$

which together with (4.2) implies (see Theorem 14.1 in [6] or Section 10.1 in [11]) that

$$HD(U_p) \geq HD(U^+) = -\frac{\sum_{j=1}^{k} r_{i_j} \ln r_{i_j}}{\ln m} = -\frac{\sum_{i=1}^{m-1} r_i \ln r_i}{\ln m} \tag{4.7}$$

with the convention $0 \ln 0 = 0$.

Set $l = m - k$ which is the number of $j \in \{0, 1, ..., m-1\}$ such that $r_j = 0$. Choose $\delta > 0$ so small that

$$r_j > \delta k^{-1} \text{ if } r_j > 0 \;\; \ln(\delta l^{-1}) \leq k^{-1} \sum_{j: r_j > 0} \ln(r_j - \delta k^{-1}). \tag{4.8}$$

Set $r^{(\delta)} = (r_0^{(\delta)}, r_1^{(\delta)}, ..., r_{m-1}^{(\delta)})$ where $r_j^{(\delta)} = r_j - \delta k^{-1}$ if $r_j > 0$ and $r_j^{(\delta)} = \delta l^{-1}$ if $r_j = 0$. Observe that by (4.8),

$$\sum_{j=0}^{m-1} r_j \ln r_j^{(\delta)} \geq \sum_{j=0}^{m-1} r_j^{(\delta)} \ln r_j^{(\delta)}. \tag{4.9}$$

Set

$$W^{(\delta)} = \big\{x \in [0,1] : \limsup_{n\to\infty} \big(-\frac{1}{n} \ln \varphi\mu_{r^{(\delta)}}(I_n(x))\big) \leq -\sum_{j=0}^{m-1} r_j^{(\delta)} \ln r_j^{(\delta)}\big\}.$$

where $\mu_{r^{(\delta)}}$ is the Bernoulli measure constructed by $r^{(\delta)}$ in the same way as μ_r is constructed by r. As in (4.4),

$$\ln \varphi\mu_{r^{(\delta)}}(I_n(x)) = \sum_{j=0}^{m-1} m_j(x,n) \ln r_j^{(\delta)}, \tag{4.10}$$

and so by (4.4), (4.5), (4.9) and (4.10),

$$U_p \subset \{x \in [0,1] : \lim_{n\to\infty}\big(-\frac{1}{n}\ln \varphi\mu_{r^{(\delta)}}(I_n(x))\big) = -\sum_{j=0}^{m-1} r_j \ln r_j^{(\delta)}\} \subset W^{(\delta)}. \tag{4.11}$$

If $U_{p^{(\delta)}}$ is constructed by $p^{(\delta)} = (p_\alpha^{(\delta)},\, \alpha \in \mathcal{A}_\ell)$ with $p_\alpha^{(\delta)} = \prod_{i=1}^{\ell} r_{\alpha_i}^{(\delta)}$ and $\alpha = (\alpha_1, ..., \alpha_\ell)$ in the same way as U_p is constructed by p then similarly to (4.2) it follows that $\varphi\mu_{r^{(\delta)}}(U_{p^{(\delta)}}) = 1$ and since $U_{p^{(\delta)}} \subset W^{(\delta)}$ we conclude from here and (4.11) that

$$\varphi\mu_{r^{(\delta)}}(W^{(\delta)}) = 1 \quad \text{and} \quad HD(U_p) \le HD(W^{(\delta)}). \tag{4.12}$$

Again, since $|I_n(x)| = m^{-n}$ then it follows from the definition of $W^{(\delta)}$ by the well known argument (see Theorem 2.3 in [5] or the proof of Theorem 14.1 in [6] or Proposition 4.9 in [11] which also can be adapted to our situation) that

$$HD(W^{(\delta)}) \le -\frac{\sum_{j=0}^{m-1} r_j^{(\delta)} \ln r_j^{(\delta)}}{\ln m}. \tag{4.13}$$

Letting $\delta \to 0$ we obtain

$$HD(U_p) \le -\frac{\sum_{j=0}^{m-1} r_j \ln r_j}{\ln m}$$

which together with (4.7) completes the proof of Proposition 2.5. □

4.1. **Remark.** Many papers and several books disregard the fact that the argument in the first part of the proof above due to Billingsley works only when all r_j's are positive while without this assumption it leads only to the lower bound of dimension. This gap was noticed and repaired first only in [18] (though it appeared in later papers, as well). The problem here is that when, say, $r_{j_0} = 0$ then $\varphi\mu_r(I_n(x)) = 0$ provided $a_i(x) = j_0$ for some $i < n$ and for such x the right hand side of (4.4) becomes $-\infty$ which leads nowhere. In other words, the measure $\varphi\mu_r$ "disregards" such points while, on the other hand, the set of points x which have zero frequency of appearences of j_0 in their m-expansions is not countable and it cannot be disregarded in the Hausdorff dimension computation. In order to prove the result for general probability vectors $(r_0, ..., r_{m-1})$ it is necessary to obtain here an appropriate upper bound for the Hausdorff dimension either by a combinatorial argument not related to Billingsley's ergodic theory one as in [10] or by a simpler perturbation argument above due to my student Z.Hellman which appeared in a more general form in [18].

Next, we prove Proposition 2.7. Since for some n the matrix R^n is a positive matrix then, clearly, each $q_i = \sum_j r_{ij}$ must be positive and for each i, j there exists a sequence $i_1, i_2, ..., i_{n-1}$ such that $r_{ii_1} r_{i_1 i_2} \cdots r_{i_{n-1} j} > 0$. Then $q_{ii_1} q_{i_1 i_2} \cdots q_{i_{n-1} j} > 0$, and so Q^n is a positive matrix, as well. Clearly, $\sum_i q_i q_{ij} = q_j$, and so q is the unique stationary vector of Q. Set $\Xi_Q = \{\xi = (\xi_0, \xi_1, ...) : q_{i,i+1} > 0 \text{ for all } i \ge 0\}$ where $Q = (q_{ij},\, i, j = 0, 1, ..., m-1)$. Let μ_Q be the Markov measure on $\Xi_{q,Q}$ which assigns the weight $q_{\alpha_0} q_{\alpha_0\alpha_1} q_{\alpha_1\alpha_2} \cdots q_{\alpha_{n-1}\alpha_n}$ to each cylinder set $R_{\alpha_0\alpha_1...\alpha_n}$ with all $\alpha_i \in \mathcal{A}_+$. Then μ_Q is invariant with respect to the left shift on $\Xi_{q,Q}$ and its image $\varphi\mu_Q$ on $[0,1]$ is invariant with respect to T. Under assumptions of Proposition 2.7 the probability matrix Q is a transition matrix of an exponentially fast ψ-mixing (finite) Markov chain (satisfying Doeblin's condition), and so the conditions of

Theorem 2.2 hold true here. We can also rely on Corollary 2.3 since μ_Q is a Gibbs measure for the left shift on Ξ constructed by the function $\psi(\xi) = -\ln q_{\xi_0\xi_1}$, $\xi = (\xi_0, \xi_1, ...)$ (see [8]). Since $\varphi\mu_Q(\Gamma_{ij}) = q_i q_{ij} = r_{ij}$ we conclude from here together with (2.16), (2.26) and the definition of U_P that $\varphi\mu_Q(U_P) = 1$. If $V_Q = \varphi\Xi_Q$ then taking into account that $\mu_Q(\Xi_Q) = 1$ we obtain also that $\varphi\mu_Q(U_P \cap V_Q) = 1$.

Now, for any $x \in V_Q$ and $I_n(x)$ as above

$$\ln \varphi\mu_Q(I_n(x)) = \ln q_{a_0(x)} + \sum_{i,j=0}^{m-1} m_{ij}(x,n) \ln q_{ij} \tag{4.14}$$

where $m_{ij}(x,n) = \#\{k \geq 0, k < n : a_{k-1}(x) = i \text{ and } a_k(x) = j\}$. If $x \in U_P$ then similarly to (4.5),

$$\lim_{n\to\infty} \frac{1}{n} m_{ij}(x,n) = r_{ij} = q_i q_{ij}. \tag{4.15}$$

It follows that for any $x \in U_P \cap V_Q$,

$$\lim_{n\to\infty} \frac{\ln \varphi\mu_Q(I_n(x))}{\ln |I_n(x)|} = -\frac{\sum_{i,j=0}^{m-1} q_i q_{ij} \ln q_{ij}}{\ln m}. \tag{4.16}$$

and so similarly to (4.7),

$$HD(U_P) \geq HD(U_P \cap V_Q) = -\frac{\sum_{i,j=0}^{m-1} q_i q_{ij} \ln q_{ij}}{\ln m}. \tag{4.17}$$

For the lower bound above we dealt only with points $x \in V_Q$ where $q_{a_i(x)a_{i+1}(x)} > 0$ for all $i \geq 0$. In order to obtain the upper bound we employ again a perturbation argument which in this case seems to be new. Let l_i, $i = 0, 1, ..., m-1$ be the number of $j = 0, 1, ..., m-1$ such that $r_{ij} = 0$ and set $k_i = m - l_i$. Choose $\delta > 0$ so small that for all $i, j = 0, 1, ..., m-1$,

$$r_{ij} > k_i^{-1}\delta \text{ if } r_{ij} > 0 \text{ and } \ln(l_i^{-1}\delta) \leq k_i^{-1} \sum_{j: r_{ij} > 0} \ln((r_{ij} - k_i^{-1}\delta) q_i^{-1}). \tag{4.18}$$

Set $r_{ij}^{(\delta)} = r_{ij} - k_i^{-1}\delta$ if $r_{ij} > 0$ and $r_{ij}^{(\delta)} = l_i^{-1}\delta$ if $r_{ij} = 0$. Observe that $\sum_{j=1}^{m-1} r_{ij}^{(\delta)} = q_i$ and define $q_{ij}^{(\delta)} = r_{ij}^{(\delta)} q_i^{-1}$ yielding a positive $m \times m$ probability matrix $Q^{(\delta)} = (q_{ij}^{(\delta)})$. By (4.18) we have

$$\sum_{i,j=0}^{m-1} r_{ij} \ln q_{ij}^{(\delta)} \geq \sum_{i,j=0}^{m-1} r_{ij}^{(\delta)} \ln q_{ij}^{(\delta)}. \tag{4.19}$$

Set

$$W^{(\delta)} = \{x \in [0,1] : \limsup_{n\to\infty} \big(-\tfrac{1}{n} \ln \varphi\mu_{Q^{(\delta)}}(I_n(x))\big) \\ \leq -\textstyle\sum_{i,j=0}^{m-1} r_{ij}^{(\delta)} \max(1, q_i^{(\delta)} q_i^{-1}) \ln q_{ij}^{(\delta)}$$

where $\mu_{Q^{(\delta)}}$ is the Markov measure constructed by $Q^{(\delta)}$ and its unique stationary vector $q^{(\delta)}$ (i.e. $q^{(\delta)}Q^{(\delta)} = q^{(\delta)}$) in the same way as μ_Q was constructed by Q and q. As in (4.14),

$$\ln \varphi\mu_{Q^{(\delta)}}(I_n(x)) = \ln q_{a_0(x)}^{(\delta)} + \sum_{i,j=0}^{m-1} m_{ij}(x,n) \ln q_{ij}^{(\delta)}, \tag{4.20}$$

and so by (4.14), (4.15), (4.19) and (4.20),

$$U_P \subset \{x \in [0,1]: \lim_{n\to\infty}\big(-\frac{1}{n}\ln \varphi\mu_{Q^\delta}(I_n(x))\big) = -\sum_{i,j=0}^{m-1} r_{ij}\ln q_{ij}^{(\delta)}\} \subset W^{(\delta)}. \tag{4.21}$$

Let

$$\hat U = \{x \in [0,1]: \lim_{n\to\infty}\frac{1}{n}N_{\alpha\beta}(x,n) = \prod_{i=1}^{\ell} q_{\alpha_i}^{(\delta)} q_{\alpha_i\alpha_j}^{(\delta)} \text{ for all } \alpha,\beta \in \mathcal{A}_\ell\}.$$

Then for any $x \in \hat U$,

$$\lim_{n\to\infty}\frac{1}{n}m_{ij}(x,n) = \lim_{n\to\infty}\frac{1}{n}\sum_{0\le\alpha_2,\beta_2,...,\alpha_\ell,\beta_\ell\le m-1} N_{i,\alpha_2,...,\alpha_\ell,j,\beta_2,...,\beta_\ell}(x,n) = q_i^{(\delta)}q_{ij}^{(\delta)} \tag{4.22}$$

and by (4.20) for any $x \in \hat U$,

$$\lim_{n\to\infty}\frac{1}{n}\ln\varphi\mu_{Q^{(\delta)}}(I_n(x)) = \sum_{i,j=0}^{m-1} q_i^{(\delta)}q_{ij}^{(\delta)}\ln q_{ij}^{(\delta)} = \sum_{i,j=0}^{m-1} q_i^{(\delta)}q_i^{-1}r_{ij}^{(\delta)}\ln q_{ij}^{(\delta)}. \tag{4.23}$$

By Theorem 2.2 (or by Corollary 2.3) we obtain that $\mu_{Q^{(\delta)}}(\hat U) = 1$ and since $\hat U \subset W^{(\delta)}$ by (4.23) and the definition of $W^{(\delta)}$ it follows that $\mu_{Q^{(\delta)}}(W^{(\delta)}) = 1$. Relying again on Theorem 2.3 in [5] (or see the proof of Theorem 14.1 in [6]) we conclude that

$$HD(U_P) \le HD(W^{(\delta)}) \le -\frac{\sum_{i,j=0}^{m-1} r_{ij}^{(\delta)}\max(1, q_i^{(\delta)}q_i^{-1})\ln q_{ij}^{(\delta)}}{\ln m}. \tag{4.24}$$

Since q is the unique probability vector satisfying $qQ = q$ then $q^{(\delta)} \to q$ as $\delta \to 0$ and letting $\delta \to 0$ in (4.24) we arrive at

$$HD(U_P) \le -\frac{\sum_{i,j=0}^{m-1} r_{ij}\ln q_{ij}}{\ln m}. \tag{4.25}$$

which together with (4.17) completes the proof of Proposition 2.7. □

Concerning Proposition 2.10 we explained already at the beginning of this section that the upper bound there follows from the upper bound derived in [13]. Next, we obtain the lower bound

$$HD(U_p) \ge \frac{h_\nu}{2\int|\ln x|}d\nu(x)$$

for any $\nu \in \hat{\mathcal{N}}(\bar r)$ in the same way as in Theorem 1 from [7] since in addition to arguments there concerning continued fractions themselves we need only that $\nu(U_p) = 1$ (actually, already $\nu(U_p) > 0$ is enough) which follows from (2.29).

The remaining bound $HD(U_p) \ge \frac{1}{2}$ can be proved similarly to Section 4 in [13]. Namely, we construct first points $z \in U_p$ with $a_n(z) \le n$ for all $n \ge 1$. In order to do this choose probability vectors $(r_1^{(n)}, r_2^{(n)}, ...)$ such that $r_k^{(n)} > 0$ when $1 \le k \le n$, $\sum_{k=1}^n p_k^{(n)} = 1$ and $\lim_{n\to\infty} r_k^{(n)} = r_k$ for any $k \ge 1$. Consider independent integer valued random variables $Y_1, Y_2, ...$ such that $P\{Y_n = k\} = r_k^{(n)}$. Applying Theorem 2.2 we conclude similarly to (4.1) that for any ℓ-word $\alpha = (\alpha_1, ..., \alpha_\ell) \in \mathcal{A}_\ell$ and

P-almost all ω,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\mathbb{I}_{\alpha_1}(Y_{q_1(k)}(\omega))\mathbb{I}_{\alpha_2}(Y_{q_2(k)}(\omega))\cdots\mathbb{I}_{\alpha_\ell}(Y_{q_\ell(k)}(\omega))=\prod_{i=1}^{\ell}r_{\alpha_i}. \tag{4.26}$$

Now, in order to satisfy our conditions we can take any z whose continued fraction expansion have digits $a_n(z)=Y_n(\omega)$, $n=1,2,...$ with ω such that (2.26) holds true.

Next, let $0\le m(k)\le\ell$ be integers such that $k^2+m(k)\ne q_i(k)$ for all $k\ge1$ and $i=1,...,\ell$. For $z\in U_p$ constructed above and $b>1$ define the set

$$G_z(b)=\{x\in(0,1):\ a_{k^2+m(k)}(x)\in(b^{k^2},2b^{k^2})\text{ and }$$
$$a_n(x)=a_n(z)\text{ if }n\ne k^2+m(k)\text{ for some }k\}.$$

Then, clearly, $G_z(b)\subset U_p$. Following [13] we construct a measure μ on $G_z(b)$ setting for each rank-m basic interval $I_m(x)$ containing x,

$$\mu(I_m(x))=\prod_{k=1}^{n}b^{-k^2}$$

provided $n^2\le m<(n+1)^2$. Now, in the same way as in [13] we can show that for any $\theta>0$ there exists $b>1$ such that for all $x\in G_z(b)$,

$$\liminf_{r\to0}\frac{\ln\mu(x-r,x+r)}{\ln r}\ge\frac{1}{2}-\theta.$$

It follows (see, for instance, Theorem 2.3 in [5] or Proposition 4.9 in [11]) that $HD(U_p)\ge HD(G_z(b))\ge\frac{1}{2}-\theta$ and since $\theta>0$ is arbitrary we obtain the required bound. Finally, we observe that if $\sum_{j=1}^{\infty}r_j\ln j=\infty$ then $HD(U_p)=\frac{1}{2}$ which follows from the latter lower bound and the upper bound of [13]. $\square$

References

[1] J. Aaronson, *An Introduction to Infinite Ergodic Theory*, Amer. Math. Soc., Providence, 1997.

[2] I. Assani, *Multiple recurrence and almost sure convergence for weakly mixing dynamical systems*, Israel J. Math. 103, 111–124 (1998).

[3] V. Bergelson, *Weakly mixing PET*, Ergod. Th.& Dynam. Sys. 7, 337–349 (1987).

[4] A. S. Besicovitch, *On the sum of digits of real numbers represented in the dyadic system*, Math. Annalen 110, 321-330 (1934).

[5] P. Billingsley, *Hausdorff dimension in probability theory II*, Ill. J. Math. 5, 291–298 (1961).

[6] P. Billingsley, *Ergodic Theory and Information*, Wiley, New York, 1965.

[7] P. Billingsley and I. Henningsen, *Hausdorff dimension of some continued-fraction sets*, Z. Wahrsch. verw. Geb. 31, 163–173 (1975).

[8] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math. 470, Springer–Verlag, Berlin, 1975.

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

[10] H. G. Eggleston, *The fractional dimension of a set defined by decimal properties*, Quart.J.Math. 20, 31-36 (1949).

[11] K. Falconer, *Fractal Geometry. Mathematical Foundations and Applications,* 2nd ed., Wiley, Chichester, 2003.

[12] H. Furstenberg, *Nonconventional ergodic averages*, Proc. Symp. Pure Math. 50, 43–56 (1990).

[13] A.-H. Fan, L. Liao and J.-H. Ma, *On the frequency of partial quotients of regular continued fractions*, Math. Proc. Camb. Phil. Soc. 148 (2010), 179–192.

[14] D.J.H. Garling, *Inequalities: a Journey into Linear Analysis*, Cambridge Univ. Press, Cambridge (2007).

[15] L. Heinrich, *Mixing properties and central limit theorem for a class of non-identical piecewise monotonic C^2-transformations*, Mathematische Nachricht. 181, 185–214 (1996).

[16] P. Hall and C.C. Heyde, *Martingale Limit Theory and its Application*, Acad. Press, New York (1980).

[17] I.A. Ibragimov and Yu.V. Linnik, *Independent and Stationary Sequences of Random Variables*, Wolters–Noordhoff, Groningen (1971).

[18] Yu. Kifer, *Fractal dimensions and random transformations*, Trans. Amer. Math. Soc., 348, 2003–2038 (1996).

[19] Yu. Kifer, *Limit theorems for random transformations and processes in random environments*, Trans. Amer. Math. Soc., 350, 1448–1518 (1998).

[20] Yu. Kifer, *Nonconventional limit theorems*, Probab. Th. Rel. Fields, 148, 71–106 (2010).

[21] Yu. Kifer and S.R.S. Varadhan, *Nonconventional limit theorems in discrete and continuous time via martingales*, Preprint, arXiv: 1012.2223.

[22] D.L. McLeish, *A maximal inequality and dependent strong laws*, Ann. Probab. 3 (1975), 829-839.

[23] P. Walters, *Invariant measures and equilibrium states for some mappings which expand distances*, Trans. Amer. Math. Soc. 236 (1978), 121–153.

E-mail address: `kifer@math.huji.ac.il`

Institute of Mathematics, Hebrew University, Jerusalem 91904, Israel