Fluctuations of the metric entropy for mixing measures

Nicolai Haydn * Sandro Vaienti †

November 21, 2010

Abstract

It is known from the theorem of Shannon-McMillan-Breiman that the measure of cylinder sets decays in the limit exponentially with a rate which is given by the metric entropy. Here we prove that the measure of cylinders satisfies a Central Limit Theorem for a wide class of mixing measures which do not have necessarily the Gibbs property. We moreover provide the rate of convergence (which is algebraic). As a consequence we can then also prove that the distribution of the first return time in cylinder sets is log-normally distributed and give the speed of convergence. We also show that the weak invariance principle and the law of the iterated logarithm hold for the convergence to the entropy in the Shannon-McMillan-Breiman and for the distribution of the first return time.

1 Introduction

In [13] we studied a large class of invariant measures for dynamical systems that do not have necessarily the Gibbs property and which we called (ϕ, f)-mixing maps (see Section 2 for definition). There we investigated in particular the statistics of multiple return times. Let's recall one of those results which motivated the present work. If (Ω, μ, T) is a measurable dynamical system, let $x \in \Omega$ be any point and $A_n(x) \in \mathcal{A}^n$ the cylinder around x belonging to the join partition $\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}$, where $\mathcal{A}$ is any measurable generating partition of (Ω, μ, T). We call $\tau_{A_n(x)}(y) = \min\left\{k > 0 : T^k y \in A_n(x), y \in A_n(x)\right\}$. We next introduce the distribution

$$\mu_{A_n(x)}\left(\left\{y : \tau_{A_n(x)}(y) > \frac{t}{\mu(A_n(x))}\right\}\right)$$

where $\mu_{A_n(x)}$ is the conditional measure on $A_n(x)$. Inthe limit of $n \to \infty$ and for μ-almost every x this distribution can be shown to converge to e^{-t} for the class of (ϕ, f)-mixing maps (for reasonable ϕ and f). We next consider the first return of the 'center' x of the cylinder $A_n(x)$ (which we call the *repeat function* in Sect. 7); for an ergodic measure of finite entropy $h(\mu)$ Ornstein and Weiss [20] have shown that

$$\lim_{n\to\infty} \frac{1}{n} \log \tau_{A_n(x)}(x) = h(\mu) \tag{1}$$

*Mathematics Department, USC, Los Angeles, 90089-1113. E-mail: <nhaydn@math.usc.edu>. This work was supported by a grant from the NSF and by a grant from the Université de Toulon et du Var, France.

†Centre de Physique Théorique, UMR 6207 du CNRS et des Universités d'Aix-Marseille I,II et Sud Toulon-Var, affilié à la FRUMAM, Luminy Case 907, F-13288 Marseille Cedex 9. E-mail: <vaienti@cpt.univ-mrs.fr>.

almost everywhere. It is a remarkable fact that in order to prove fluctuations in the convergence of the limit (1) one needs the e^{-t}-statistics for return times. This leads us to look at the distribution

$$\mu\left(\left\{x : \frac{\log \tau_{A_n(x)}(x) - nh(\mu)}{\sigma\sqrt{n}} > t\right\}\right) \tag{2}$$

as $n \to \infty$, where (the variance) $\sigma^2 > 0$ will be specified later on. It can be shown [27] and also Sections 4 and 5, that the e^{-t}-statistics of the first return times combined with a Central Limit Theorem for the fluctuation to the entropy in the Shannon-McMillan-Breiman theorem (SMB), will allow to get the convergence in law of (2) to the normal distribution. This kind of result has been proved for Gibbsian sources [8], for some maps of the interval with topological covering properties [21] and some non-uniformly hyperbolic maps on the interval [9, 4]. Kontoyiannis [18] has some results in the setting of finite-valued stationary strongly mixing processes with some sort of finite-order Markov chain approximation (the assumption on the coefficient 'γ' introduced by Ibragimov [17]). (If $\sigma = 0$ then there are some examples when the limiting distribution is not normal [8].) All these contributions (except that of Kontoyiannis), used explicitly a Gibbs type characterization of the measure μ which allows to rephrase the CLT for the convergence to the entropy in Shannon-McMillan-Breiman theorem in terms of a standard CLT for the involved potential. For recent results on the CLT in various settings for potentials (and in particular for Gibbs states) see [12, 19, 5]. Note that none of these results provide error terms.

Instead, the contribution of Kontoyiannis quoted above used a CLT directly for the processes $\log \mu(A_n(x))$ (without requiring the Gibbs property) which was proven by Ibragimov in the early sixties. In [17] Ibragimov worked on 'strongly mixing systems'[1] (see also Rosenblatt [25]). He works in a classical probabilitstic setting of stationary random variables X_j (with a finite number of states) where it is assumed that the function of conditional expectations $f_0 = \log p(x_0|x_{-1}x_{-2}\dots)$ is known. He assumes that this function is sufficiently well approximable, that is the variation $\psi(n) = E|f_0 - \log p(x_0|x_{-1}\dots x_{-n})|$ decays sufficiently fast. In the present paper we consider a typical dynamical systems setting: We assume mixing properties of the measure with are stronger that the 'strong mixing property' of Rosenblatt and Ibragimov, but in return do not make any assumption on the function f_0 and its regularity. (Although we suspect we have not been able to prove that the rate of decay of ψ as assumed by Ibragimov and others implies the mixing property of Definition 1.) We moreover obtain error estimates for the CLT. In fact our approach is more accessible to numerical exploitation that the purely probabilistic framework.

The main result of this paper is to prove a CLT for the convergence to the entropy in the SMB theorem for (ϕ, f)-mixing maps (Theorem 16). The proof uses characteristic functions and a classical short/long splitting to estimate the distance to the characteristic function of the normal distribution. A Berry-Esseen type argument then implies the main result. The central part of the paper however is to provide an expression for the variance in terms of the dynamical information function $I_n(x) = -\log \mu(A_n(x))$: It will be shown that the variance of the measure is given by the linear growth rate of the difference of the second moment and the square of the first moment (Proposition 14). The speed of convergence to the variance determines the error terms in the CLT. These results were announced in [14].

[1] Meaning that

$$|\mu(U \cap T^{-m-n}V) - \mu(U)\mu(V)| \le \alpha(m)$$

for all n-cylinders U and, where V is as in Definition 1 where the decay function $\alpha(m)$ goes to zero.

Our setting is rather general in that it does not require the invariant measure to be Gibbsian. For Gibbs measures, A Broise [3] has proved error terms for the CLT for a large class of expanding maps on the interval for which the Perron-Frobenius operator has a 'spectral gap' (previous results had been obtained by Rousseau-Egele [26]). Moreover, for Gibbsian sources of dispersing billiards Pène [22] has obtained similar estimates.

¿From section 4 to 7 we improve the CLT by showing some of its classical, although not trivial, consequences either for the processes $\log \mu(A_n(x))$ and for $\log \tau_{A_n(x)}(x)$.

In the fourth section we prove the law of iterated logarithm and in the fifth the weak invariance principle for the function $\log \mu(A_n(x))$. The latter follows easily from the CLT in the case of transitive Gibbs states and can also be proved for a large class of non-uniformly hyperbolic maps of the interval [9] and for the stationary strongly mixing processes studied in [18].

In the section 6 we prove the exponential law for the first return time and provide error estimates which in section 7 are finally then used to derive a CLT with error terms and a Weak Invariance Principle for the repeat time function.

2 Properties of (ϕ, f)-mixing measures

Let T be a map on a space Ω and μ an invariant probability measure on Ω. Moreover let $\mathcal{A}$ be a finite measurable partition of Ω and denote by $\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}$ its *n-th join* which also is a measurable partition of Ω for every $n \geq 1$. The atoms of $\mathcal{A}^n$ are called *n-cylinders*. Let us put $\mathcal{A}^* = \bigcup_n \mathcal{A}^n$ for the collection of all cylinders in Ω and put $|A|$ for the length of an n-cylinder $A \in \mathcal{A}^*$, i.e. $|A| = n$ if $A \in \mathcal{A}^n$.

We shall assume that $\mathcal{A}$ is generating, i.e. that the atoms of $\mathcal{A}^\infty$ are single points in Ω.

Definition 1 *Assume*
(i) $f : \mathcal{A}^* \to \mathbf{N}_0$ *so that* $f(A) \geq f(B)$ *if* $|A| \geq |B|$, $A, B \in \mathcal{A}^*$. *If* C *is a union of* n*-cylinders* C_j *(some* n*) then* $f(C) = \max_j f(C_j)$.
(ii) $\phi : \mathbf{N}_0 \to \mathbf{R}^+$ *is non-increasing.*

We say that the dynamical system (T, μ) *is* (ϕ, f)-mixing *if*

$$\left|\mu(U \cap T^{-m-n}V) - \mu(U)\mu(V)\right| \leq \phi(m)\mu(U)\mu(V)$$

for all $m \geq f(U)$*, measurable* V *(in the* σ*-algebra generated by* $\mathcal{A}^*$*) and* U *which are unions cylinders of the same length* n*.*[2]

Oftentimes the function f depends only on the length of the cylinders, that is $f(A) = f(|A|)$. The function ϕ determines the rate at which the mixing occurs and the *separation function* f specifies a lower bound for the size of the gap m that is necessary to get a good mixing property. In the special case when f is constant 0 (or some other constant) then (T, μ) is traditionally called ϕ-mixing. There is a tradeoff between the decay function ϕ and the separation function f. Typically one can achieve to have ϕ decay faster at the expense of f which as a consequence will be increasing faster.

Examples

[2]Probabilists like to refer to this mixing property as ψ-mixing. In this paper however we adher to the notation favoured by dynamicists and therefore use the letter ϕ.

1. Classical ϕ-mixing systems (see, e.g. [8]): $f = 0$.
2. Dispersing billiards [22]: f is linear.
3. Gibbs measures for rational maps with critical points: f is linear, ϕ is exponential.
4. Multidimensional piecewise continuous maps [21]: f depends on the individual cylinders.

2.1 General properties

For $r \geq 1$ and (large) N denote by $G_r(N)$ the r-vectors $\vec{v} = (v_1, \dots, v_r)$ for which $1 \leq v_1 < v_2 < \dots < v_r \leq N$. (The set $G_r(N)$ is the intersection of a cone in $\mathbf{Z}^r$ with a ball of radius N and centre at the origin.) Let t be a positive parameter, put $N = [t/\mu(W)]$ (the normalised time) and $W \subset \Omega$. Then the entries v_j of the vector $\vec{v} \in G_r(N)$ are the iterates at which all the points in $C_{\vec{v}} = \bigcap_{j=1}^r T^{-v_j} W$, hit the set W during the time interval $[1, N]$.

Lemma 2 *[13] Let (T, μ) be (ϕ, f)-mixing, let $r > 1$ be an integer and let $W_j \subset \Omega$, be unions of n_j-cylinders, $j = 1, \dots, r$.*

Then for all 'hitting vectors' $\vec{v} \in G_r(N)$ with return times $v_{j+1} - v_j \geq f(W_j) + n_j$ ($j = 1, \dots, r-1$) one has

$$\left| \frac{\mu\left(\bigcap_{j=1}^r T^{-v_j} W_j\right)}{\prod_{j=1}^r \mu(W_j)} - 1 \right| \leq (1 + \phi(d(\vec{v}, \vec{n})))^r - 1,$$

and $d(\vec{v}, \vec{n}) = \min_k (v_{k+1} - v_k - n_k)$.

The following exponential estimate has previously been shown for ϕ-mixing measures in [11] and for α-mixing measures in [1].

Lemma 3 *[13] There exists a $0 < \gamma_1 < 1$ so that for all $A \in \mathcal{A}^*$:*

$$\mu(A) \leq \gamma_1^{|A|}.$$

This lemma implies in particular that μ has positive entropy [13]. In remainder of this section let us assume that μ is a (ϕ, f)-mixing probability measure on Ω for the map $T : \Omega \to \Omega$ where $f(A)$ for all $A \in \mathcal{A}^*$ depends only on the length $|A|$ of the cylinder A. Hence we shall now write $f(n)$ where $n \in \mathbf{N}$ is the length of the cylinders. For Lemmas 4 and 5 we assume that f grows no more than linearly, that is, there exists a constant C_0 so that $f(n) \leq C_0 n$ for all n.

We are interested in the cylinders that return within very short time to themselves. In particular let us put

$$\mathcal{S}_n = \{A \in \mathcal{A}^n : A \cap T^j A \neq \emptyset \text{ for some } 1 \leq j < \kappa(n/2)\},$$

where $\kappa(m) = \left[\frac{m}{1+C_0}\right]$.

Lemma 4 *There exists a $0 < \gamma_2 < 1$ so that for all large enough n:*

$$\sum_{A \in \mathcal{S}_n} \mu(A) \leq \gamma_2^n.$$

Proof. If we put $\Delta = f(\kappa(\frac{n}{2}))$, then $f(j) + \Delta < \frac{n}{2}$ for $j = 1, \ldots, \kappa(\frac{n}{2}) - 1$. Let $A \in \mathcal{A}^n$ and assume that $A \cap T^j A \neq \emptyset$ for some $j < \kappa(\frac{n}{2})$. For $B \in \mathcal{A}^j$ so that $A \subset B$ one has by Lemma 3

$$\mu(A) \leq \mu(B \cap T^{-j-\Delta} T^{j+\Delta} A) \leq (1 + \phi(\Delta))\mu(B)\mu(T^{j+\Delta} A) \leq (1 + \phi(\Delta))\mu(B)\gamma_1^{n-j-\Delta},$$

where $T^{j+\Delta}A$ is a cylinder of length $n - j - \Delta$. Therefore

$$\sum_{A \in \mathcal{S}_n} \mu(A) \leq (1 + \phi(\Delta)) \sum_{j=1}^{\kappa(\frac{n}{2})-1} \gamma_1^{n-j-\Delta} \sum_{B \in \mathcal{A}^j} \mu(B) \leq c_1 \gamma_1^{n-\kappa(\frac{n}{2})-\Delta},$$

for a suitable constant $c_1 \leq \frac{1+\phi(\Delta)}{1-\gamma_1}$. Put $\gamma_2 = \sqrt{\gamma_1}$. □

Lemma 5 *There exists a $0 < \gamma_3 < 1$ so that for all large enough n:*

$$\mu\left(\left\{x \in \Omega : \tau_{A_n(x)}(x) < n\right\}\right) \leq \gamma_3^n.$$

Proof. If for $k < n$ put $\rho_k = \mu\left(\left\{x \in \Omega : \tau_{A_n(x)}(x) = k\right\}\right)$, then $\rho_n \leq \sum_{A \in \mathcal{A}^n} \mu(A \cap T^{-k}A)$, where in view of Lemma 4 it is enough to consider $\kappa(n) \leq k < n$. Put $\ell = \kappa(k)$ and we get by the mixing property

$$\mu(A \cap T^{-k}A) \leq \mu(A_\ell(A) \cap T^{-k}A) \leq (1 + \phi(k - \ell))\mu(A_\ell(A))\mu(A) \leq c_1\gamma_1^\ell \mu(A),$$

since $k - \ell \geq f(\ell)$, where $A_\ell(A)$ is the ℓ-cylinder that contains A. Hence

$$\rho_k \leq \sum_{A \in \mathcal{A}^n} \mu(A_\ell(A) \cap T^{-k}A) \leq c_1 \gamma_1^\ell$$

and therefore, by Lemma 4

$$\sum_{k=1}^{n-1} \rho_k \leq \gamma_2^n + \sum_{k=\kappa(n)}^{n-1} c_1 \gamma_1^{\frac{k}{1+C_0}} \leq \gamma_2^n + c_2 \gamma_1^{\frac{\kappa(n)}{1+C_0}}.$$

Now choose $\gamma_3 > \max\left(\gamma_2, \gamma_1^{\frac{1}{(1+C_0)^2}}\right)$ less than 1. □

The metric entropy h for an invariant measure μ is $h = \lim_{n\to\infty} \frac{1}{n} H(\mathcal{A}^n)$, where, as usual, $H(\mathcal{A}^n) = \sum_{A \in \mathcal{A}^n} -\mu(A) \log \mu(A)$.

Lemma 6 *Assume that $f(m) \leq C_1 m^\gamma$ for some constant C_1 and $0 \leq \gamma < 1$.*

Then for every $\gamma' \in (\gamma, 1)$ there exists a constant C_2 so that for all m:

$$\left| h - \frac{1}{m} H(\mathcal{A}^m) \right| \leq C_2 \frac{1}{m^{1-\gamma'}}.$$

Proof. We shall first show that

$$H(\mathcal{A}^{n+m+\Delta}) \geq H(\mathcal{A}^n) + H(\mathcal{A}^m) - \phi(\Delta), \tag{3}$$

if $\Delta \geq f(\min(n,m))$. Let us assume that $m \leq n$, $f(m) < m$ and put $\Delta = f(m)$, $\mathcal{B} = \mathcal{A}^m$, $\mathcal{C} = T^{-m-\Delta}\mathcal{A}^n$ and $\mathcal{D} = T^{-m}\mathcal{A}^\Delta$ we obtain (using the fact that T is (ϕ, f)-mixing in fourth line)

$$\begin{aligned}
H(\mathcal{A}^{n+m+\Delta}) &= \sum_{B\in\mathcal{B},C\in\mathcal{C},D\in\mathcal{D}} \mu(B\cap D\cap C)\log\frac{1}{\mu(B\cap D\cap C)} \\
&\geq \sum_{B\in\mathcal{B},C\in\mathcal{C},D\in\mathcal{D}} \mu(B\cap D\cap C)\log\frac{1}{\mu(B\cap C)} \\
&\geq \sum_{B\in\mathcal{B},C\in\mathcal{C}} \mu(B\cap C)\log\frac{1}{\mu(B)\mu(C)(1+\phi(\Delta))} \\
&= H(\mathcal{A}^m) + H(\mathcal{A}^n) - \log(1+\phi(\Delta)).
\end{aligned}$$

This proves the estimate (3).

Let $\gamma < \gamma' < \frac{1}{2}$ and assume that u is large enough (so that $u^{\gamma'} \geq f(u)$). Since

$$\left|H(\mathcal{A}^u) - H(\mathcal{A}^{u-\Delta})\right| \leq H(\mathcal{A}^u|\mathcal{A}^{u-\Delta}) \leq H(\mathcal{A}^\Delta) \leq c_1\Delta,$$

we get (with $\Delta = [u^{\gamma'}]$)

$$H(\mathcal{A}^{2u}) \geq 2H(\mathcal{A}^{u-[u^{\gamma'}]}) - \phi([u^{\gamma'}]) \geq 2H(\mathcal{A}^u) - \phi([u^{\gamma'}]) - 2c_1u^{\gamma'},$$

and by iteration the following lower bound:

$$\begin{aligned}
H(\mathcal{A}^{2^iu}) &\geq 2^iH(\mathcal{A}^u) - \sum_{j=0}^{i-1} 2^{i-1-j}\left(\phi([(2^ju)^{\gamma'}]) + 2c_1(2^ju)^{\gamma'}\right) \\
&\geq 2^iH(\mathcal{A}^u) - c_2u^{\gamma'}2^i, \qquad (4)
\end{aligned}$$

where the (small) terms $\phi([(2^ju)^{\gamma'}])$ have been absorbed by the constant c_2.

To get a lower bound for arbitrary integers n let m arbitrary (large enough) and let n be some large number. We have $n = km + r$ where $0 \leq r < m$ and consider the binary expansion of k: $k = \sum_{i=0}^{\ell}\epsilon_i 2^i$, where $\epsilon_i = 0, 1$ ($\epsilon_\ell = 1$, $\ell = [\log_2 k]$). We also put $k_j = \sum_{i=0}^{\ell-j}\epsilon_i 2^i$ ($k_0 = k$). Obviously $k_{j+1} < 2^{\ell-j}$. If we put $a(u) = H(\mathcal{A}^u)$ then $a(u) \sim u$. We separate the 'first' block of length $k_{j+1}m$ from the 'second' block of length $\epsilon_{\ell-j}2^{\ell-j}m$ by a gap of length $(k_{j+1}m)^{\gamma'}$ which we cut away from the first block)

$$\begin{aligned}
a(k_j) &= a(\epsilon_{\ell-j}2^{\ell-j}m + k_{j+1}m) \\
&\geq a(\epsilon_{\ell-j}2^{\ell-j}m) + a(k_{j+1}m - [(k_{j+1}m)^{\gamma'}]) - \phi([(k_{j+1}m)^{\gamma'}]) \\
&\geq a(\epsilon_{\ell-j}2^{\ell-j}m) + a(k_{j+1}m) - c_1(k_{j+1}m)^{\gamma'} - \phi([(k_{j+1}m)^{\gamma'}]) \\
&\geq a(\epsilon_{\ell-j}2^{\ell-j}m) + a(k_{j+1}m) - c_1 2^{(\ell-j)\gamma'}m^{\gamma'} - \phi([(k_{j+1}m)^{\gamma'}]),
\end{aligned}$$

for $j = 0, 1, \ldots, \ell-1$. Iterating this formula and summing over j yields ($c_3 \leq \max(c_1, c_2)$)

$$\begin{aligned}
a(km) &\geq \sum_{j=0}^{\ell}\left(\epsilon_{\ell-j}2^{\ell-j}a(m) - c_3\left(m^{\gamma'}2^{\ell-j} + (2^{\ell-j}m)^{\gamma'}\right) - \phi([(k_{j+1}m)^{\gamma'}])\right) \\
&\geq ka(m) - c_4m^{\gamma'}2^\ell,
\end{aligned}$$

where we used (4) in the first inequality and had the the ϕ-term absorbed by the constant c_4. As above we estimate the contribution made by the remainder r as follows

$$|a(n) - a(km)| \leq \sigma\left(\mathcal{A}^n | \mathcal{A}^{km}\right) \leq c_1 r \leq c_1 m.$$

Therefore, if we use the fact that $2^\ell \leq k$, divide

$$a(n) \geq ka(m) - c_4 m^{\gamma'} k - c_1 m$$

by n and let n go to infinity ($k \to \infty$), we obtain

$$h = \liminf_{n\to\infty} \frac{a(n)}{n} \geq \frac{a(m)}{m} - c_4 m^{\gamma'-1}$$

for all m large enough. This proves the lemma since $h \leq \frac{a(m)}{m}$. □

2.2 The variance of the information function

For a partition $\mathcal{P}$ let us define $K(\mathcal{P}) = \int_\Omega \log^2 \mu(P(x))\, d\mu(x)$ where $P(x)$ is the atom of $\mathcal{P}$ that contains the point x (the function K will be needed to express the variance in Theorem 16). Note that

$$K(\mathcal{P}) = \sum_{P\in\mathcal{P}} \psi_2(\mu(P)),$$

where $\psi_2(t) = t\log^2 t$. Similarly, for two partitions $\mathcal{B}$ and $\mathcal{C}$, one has the conditional quantity

$$K(\mathcal{C}|\mathcal{B}) = \sum_{B\in\mathcal{B}, C\in\mathcal{C}} \mu(B)\psi_2\left(\frac{\mu(B\cap C)}{\mu(B)}\right) = \sum_{B\in\mathcal{B}, C\in\mathcal{C}} \mu(B\cap C)\log^2 \frac{\mu(B\cap C)}{\mu(B)}.$$

Lemma 7 *For any two partitions $\mathcal{B}, \mathcal{C}$ for which $\frac{\mu(B\cap C)}{\mu(B)} \leq 1/e \;\; \forall \;\; B\in\mathcal{B}, C\in\mathcal{C}$:*
(i) $K(\mathcal{C}|\mathcal{B}) \leq K(\mathcal{C})$
(ii) $\sqrt{K(\mathcal{B}\vee\mathcal{C})} \leq \sqrt{K(\mathcal{C})} + \sqrt{K(\mathcal{B})}$.

Proof. (i) The function $\psi_2(t) = t\log^2 t$ is convex on the interval $[0, 1/e]$, that is $\sum_i \alpha_i\psi_2(x_i) \leq \psi_2(\sum_i \alpha_i x_i)$ for numbers $x_i \in [0, 1/e]$ and positive α_i for which $\sum_i \alpha_i = 1$. Hence with $\alpha_i = \mu(B)$ and $x_i = \frac{\mu(B\cap C)}{\mu(B)}$ one has

$$K(\mathcal{C}|\mathcal{B}) = \sum_{B,C} \mu(B)\psi_2\left(\frac{\mu(B\cap C)}{\mu(B)}\right) \leq \sum_C \psi_2(\mu(C)) = K(\mathcal{C}).$$

(ii) Minkowski's inequality on L^2 spaces yields

$$\begin{aligned}
\sqrt{K(\mathcal{B}\vee\mathcal{C})} &= \left(\sum_{B\in\mathcal{B}, C\in\mathcal{C}} \mu(B\cap C)\log^2 \mu(B\cap C)\right)^{\frac{1}{2}} \\
&\leq \left(\sum_{B,C} \mu(B\cap C)\log^2 \frac{\mu(B\cap C)}{\mu(B)}\right)^{\frac{1}{2}} + \left(\sum_{B,C} \mu(B\cap C)\log^2 \mu(B)\right)^{\frac{1}{2}} \\
&\leq \sqrt{K(\mathcal{C})} + \sqrt{K(\mathcal{B})}
\end{aligned}$$

(as $K(\mathcal{C}|\mathcal{B}) \le K(\mathcal{C})$), where in the last estimate we used the inequality from part (i) of this lemma. $\square$

Let us note that for $w \ge 1$ one can define the

$$K_w(\mathcal{P}) = \sum_{P \in \mathcal{P}} \mu(P) |\log \mu(P)|^w$$

($K_1 = H$ and $K_2 = K$) for which one proves the following result in the same way as Lemma 7.

Lemma 8 *For any two partitions $\mathcal{B}, \mathcal{C}$ for which $\frac{\mu(B \cap C)}{\mu(B)} \le e^{1-w} \ \ \forall \ \ B \in \mathcal{B}, C \in \mathcal{C}$:*
(i) $K_w(\mathcal{C}|\mathcal{B}) \le K_w(\mathcal{C})$
(ii) $K_w(\mathcal{B} \vee \mathcal{C})^{1/w} \le K_w(\mathcal{C})^{1/w} + K_w(\mathcal{B})^{1/w}$.

Let us note that Lemma 8(ii) implies that the sequence $a_m(w) = K_w(\mathcal{A}^m)^{1/w}$ is subadditive which implies that $\lim_{m \to \infty} \frac{1}{m} K_w(\mathcal{A}^m)^{1/w}$ exists and equals the $\liminf$. In particular $K_w(\mathcal{A}^m) \le c_1 m^w$ for all m and some constant c_1 which depends on w (we shall use this fact at the end of the proof of Theorem 16).

For $m = 1, 2, \ldots$ denote by $A_m(x)$ the atom in $\mathcal{A}^m$ which contains the point $x \in \Omega$. If we put X_m, $m = 1, 2, \ldots$, for the function (random variable) given by $X_m(x) = -\log \mu(A_m(x))$ then it's expected value is

$$\mu(X_m) = \sum_{B \in \mathcal{A}^m} -\mu(B) \log \mu(B) = H(\mathcal{A}^m).$$

The variance of X_m is

$$\sigma^2(X_m) = \mu\left((X_m - H(\mathcal{A}^m))^2\right) = K(\mathcal{A}^m) - H^2(\mathcal{A}^m).$$

For a partition $\mathcal{D}$ let us write $\sigma^2(\mathcal{D}) = K(\mathcal{D}) - H^2(\mathcal{D}))$ and similarly for conditional variance.

The function $I_n(x) = -\log A_n(x)$ is often called the nth "information function" on Ω. With this notation we have $H(\mathcal{A}^n) = \int_\Omega I_n \, d\mu$ and $\sigma^2(\mathcal{A}^n)$ is the variance of I_n.

Lemma 9 *Let $\mathcal{B}$ and $\mathcal{C}$ be two partitions. Then*

$$\sigma(\mathcal{B} \vee \mathcal{C}) \le \sigma(\mathcal{C}|\mathcal{B}) + \sigma(\mathcal{B}).$$

Proof. By Minkowski's inequality

$$\begin{aligned}
\sigma(\mathcal{B} \vee \mathcal{C}) &= \left(\sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{\mu(B)}{\mu(B \cap C)} - H(\mathcal{C}|\mathcal{B}) + \log \frac{1}{\mu(B)} - H(\mathcal{B}) \right)^2 \right)^{\frac{1}{2}} \\
&\le \left(\sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{\mu(B)}{\mu(B \cap C)} - H(\mathcal{C}|\mathcal{B}) \right)^2 \right)^{\frac{1}{2}} \\
&\qquad + \left(\sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{1}{\mu(B)} - H(\mathcal{B}) \right)^2 \right)^{\frac{1}{2}} \\
&= \sigma(\mathcal{C}|\mathcal{B}) + \sigma(\mathcal{B}).
\end{aligned}$$

This completes the proof. $\square$

The remainder of this section will be devoted to prove that $\frac{1}{n}\sigma^2(\mathcal{A}^n)$ is bounded uniformly in n (Lemma 12) and then to show that it converges (Proposition 14). For this purpose we will need the following pair of arithmetic lemmas.

Lemma 10 *Let $\gamma' \in (0, 1/2)$ and define for integer x ([.] denotes the integer part) :*

$$\begin{aligned} T(x) &= \left[\frac{x}{2} - \left(\frac{x}{2}\right)^{\gamma'}\right] \\ S(x) &= x - 2T(x). \end{aligned}$$

Then there exists an x_0 so that $T(x)^{\gamma'} \le S(x) \le 3T(x)^{\gamma'}$ for all $x > x_0$.

Proof. Put $x' = T(x)$ and $y = x - 2x' = S(x)$, then

$$\begin{aligned} y &= x - 2\left(\frac{x}{2} - \left(\frac{x}{2}\right)^{\gamma'}\right) + \epsilon \\ &= \epsilon + 2\left(\frac{x}{2}\right)^{\gamma'} \\ &= \epsilon + \alpha(x')\left(\frac{x}{2} - \left(\frac{x}{2}\right)^{\gamma'}\right)^{\gamma'} \\ &= \epsilon' + \alpha(x')x'^{\gamma'} \end{aligned}$$

where $0 \le \epsilon \le 2$, $1 \le \alpha(x') \le 3$, for all x large enough and $0 \le \epsilon' \le 5$. Hence $x'^{\gamma'} \le y \le 3x'^{\gamma'}$ for all x larger than some x_0. $\square$

Lemma 11 *Let $0 < \gamma' < 1$ and $c_2 \ge 1$ be given. For any integer $n_0 \ge 1$ and any sequence of integers $m_0, m_1, m_2, \ldots$, for which $0 \le m_j \le 3n_j^{\gamma'}$, where recursively $n_{j+1} = 2n_j + m_j$, one has*

$$n_0 2^j \le n_j \le p_j n_0 2^j,$$

where $p_j = \exp \sum_{\ell=0}^{j-1} \frac{3}{2^{\ell(1-\gamma')}}$ ($p_\infty < \infty$).

Proof. The lower bound is obvious. The upper bound is shown by induction. Since $p_j \ge 1$, we get

$$\begin{aligned} n_{j+1} &\le 2n_j + 3n_j^{\gamma'} \\ &\le 2p_j n_0 2^j + 3(p_j n_0 2^j)^{\gamma'} \\ &= 2^{j+1} p_j n_0 \left(1 + \frac{3}{(p_j n_0 2^j)^{1-\gamma'}}\right) \\ &\le 2^{j+1} p_j n_0 \left(1 + \frac{3}{2^{j(1-\gamma')}}\right) \\ &\le p_{j+1} 2^{j+1} n_0. \end{aligned}$$

Lemma 12 *Assume that ϕ is summable and that $f(m) \le C_3 m^\gamma$, $m \in \mathbf{N}$, for some C_3 and $\gamma \in [0, \frac{1}{2})$.*

Then there exists a constant C_4 so that

$$0 \le K(\mathcal{A}^n) - H^2(\mathcal{A}^n) \le C_4 n.$$

Proof. The lower inequality follows from Schwarz's inequality:

$$H(\mathcal{A}^n) = \sum_{A \in \mathcal{A}^n} \mu(A)(-\log \mu(A)) \le \left(\sum_{A \in \mathcal{A}^n} \mu(A) \log^2 \mu(A) \right)^{\frac{1}{2}} = \sqrt{K(\mathcal{A}^n)},$$

which implies $K(\mathcal{A}^n) - H^2(\mathcal{A}^n) \ge 0$ for all n.

Let $\gamma' \in (\gamma, \frac{1}{2})$. Let n be some (large) integer and $n' = 2n + m$, where $n^{\gamma'} < m < 3n^{\gamma'}$ (for n large enough so that $n^{\gamma'} \ge 3f(n)$ where the factor 3 is needed in the footnote). Put $\mathcal{B} = \mathcal{A}^n$, $\mathcal{C} = T^{-m-n}\mathcal{A}^n$. Then

$$\begin{aligned} H(\mathcal{B} \vee \mathcal{C}) &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \log \frac{1}{\mu(B \cap C)} \\ &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{1}{\mu(B)} + \log \frac{1}{\mu(C)} + \rho_1(B, C) \right) \\ &= H(\mathcal{B}) + H(\mathcal{C}) + \rho_1. \end{aligned}$$

Here we used the mixing property $\mu(B \cap C) = \mu(B)\mu(C)(1 + \rho_2(B, C))$, where $|\rho_i| \le \phi(m), i = 1, 2$. We thus obtain

$$\begin{aligned} \sigma^2(\mathcal{B} \vee \mathcal{C}) &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{1}{\mu(B \cap C)} - H(\mathcal{B} \vee \mathcal{C}) \right)^2 \\ &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{1}{\mu(B)} + \log \frac{1}{\mu(C)} - H(\mathcal{B}) - H(\mathcal{C}) + \rho_3 \right)^2 \end{aligned}$$

where the absolute value of $\rho_3 = \rho_2 - \rho_1$ is bounded by $2\phi(m)$. By Minkowski's inequality :

$$\sqrt{E(\mathcal{B}, \mathcal{C})} - 2\phi(m) \le \sigma(\mathcal{B} \vee \mathcal{C}) \le \sqrt{E(\mathcal{B}, \mathcal{C})} + 2\phi(m),$$

where

$$\begin{aligned} E(\mathcal{B}, \mathcal{C}) &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{1}{\mu(B)} + \log \frac{1}{\mu(C)} - H(\mathcal{B}) - H(\mathcal{C}) \right)^2 \\ &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{1}{\mu(B)} - H(\mathcal{B}) \right)^2 \\ &\quad + \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left(\log \frac{1}{\mu(C)} - H(\mathcal{C}) \right)^2 + 2F(\mathcal{B}, \mathcal{C}) \\ &= \sigma^2(\mathcal{B}) + \sigma^2(\mathcal{C}) + 2F(\mathcal{B}, \mathcal{C}). \end{aligned}$$

We estimate the last term, $2F(\mathcal{B},\mathcal{C})$, as follows

$$\begin{aligned} F(\mathcal{B},\mathcal{C}) &= \sum_{B\in\mathcal{B},C\in\mathcal{C}} \mu(B\cap C)\left(\log\frac{1}{\mu(B)}-H(\mathcal{B})\right)\left(\log\frac{1}{\mu(C)}-H(\mathcal{C})\right) \\ &= \sum_{B\in\mathcal{B},C\in\mathcal{C}} \mu(B)\mu(C)(1+\rho_2(B,C))\left(\log\frac{1}{\mu(B)}-H(\mathcal{B})\right)\left(\log\frac{1}{\mu(C)}-H(\mathcal{C})\right). \end{aligned}$$

By Schwarz's inequality

$$\begin{aligned} |F(\mathcal{B},\mathcal{C})| &\le \|\rho_2\|\sum_{B\in\mathcal{B},C\in\mathcal{C}} \mu(B)\mu(C)\left|\log\frac{1}{\mu(B)}-H(\mathcal{B})\right|\cdot\left|\log\frac{1}{\mu(C)}-H(\mathcal{C})\right| \\ &\le \phi(m)\sigma(\mathcal{B})\sigma(\mathcal{C}). \end{aligned}$$

Hence

$$\sigma(\mathcal{B}\vee\mathcal{C}) \le \sqrt{\sigma^2(\mathcal{C})+\sigma^2(\mathcal{B})+2\phi(m)\sigma(\mathcal{C})\sigma(\mathcal{B})}+2\phi(m), \tag{5}$$

and since $\sigma(\mathcal{B})=\sigma(\mathcal{C})=\sigma(\mathcal{A}^n)$, we get

$$\sigma(\mathcal{B}\vee\mathcal{C}) \le \sigma(\mathcal{A}^n)\sqrt{2}\sqrt{1+\phi(m)}+2\phi(m).$$

Next we fill the gap of length m for which we use Lemmas 9 and 7[3]:

$$|\sigma(\mathcal{A}^{2n+m})-\sigma(\mathcal{B}\vee\mathcal{C})| \le \sigma(\mathcal{A}^m|\mathcal{B}\vee\mathcal{C}) \le \sqrt{K(\mathcal{A}^m)} \le c_1 m \le c_2 n^{\gamma'}$$

($c_2\le 3c_1$). Then (if $\phi(m)\le 1$)

$$\begin{aligned} \sigma(\mathcal{A}^{n'}) &\le \sqrt{2}\,\sigma(\mathcal{A}^n)(1+\phi(m))+c_3 n^{\gamma'} \\ &\le \sqrt{2}\,\sigma(\mathcal{A}^n)\left(1+\frac{c_4}{n^{\gamma'}}+\frac{c_3 n^{\gamma'}}{\sigma(\mathcal{A}^n)}\right), \end{aligned} \tag{6}$$

($c_3\le 2+2c_2$), where we used that $\phi(m)\le\frac{c_4}{m}$ for all m and some c_4 (as ϕ is summable).

Moreover let us note that if $\sigma(\mathcal{A}^{n'})>\sqrt{n'}$, then

$$\sigma(\mathcal{A}^n) \ge \frac{\sigma(\mathcal{A}^{n'})-c_3 n^{\gamma'}}{\sqrt{2}\,(1+\phi(m))} \ge \frac{1}{3}\left(\sqrt{n'}-c_3 n^{\gamma'}\right) \ge c_5\sqrt{n'}$$

---

[3] We are allowed to use Lemma 7 because $\mu(A)/\mu(B\cap C)\le 1/e$. To see this consider the gap which is of length m and let $D\in T^{-n-[m/3]}\mathcal{A}^{[m/3]}$ so that $A\subset B\cap D\cap C$. Then since $m/3\ge f(n)$ we get by Lemma 2

$$|\mu(B\cap D\cap C)-\mu(B)\mu(D)\mu(C)| \le \text{const.}\mu(B)\mu(C)\mu(D)\phi(m/3)$$

and by the (ϕ,f)-mixing property

$$|\mu(B\cap D\cap C)-\mu(B\cap C)\mu(D)| \le \text{const.}\mu(B\cap C)\mu(D)\phi(m/3).$$

This implies then that

$$\frac{\mu(A)}{\mu(B\cap C)} \le \frac{\mu(B\cap D\cap C)}{\mu(B\cap C)} \le \text{const.}\mu(D)\phi(m/3) \le \text{const.}\gamma_1^{m/3}$$

which is $<1/e$ if m is large enough.

for some positive c_5 and for all n larger than some N. By Lemma 11 we thus obtain

$$\frac{n^{\gamma'}}{\sigma(\mathcal{A}^n)} \leq \frac{n^{\gamma'}}{c_5\sqrt{n'}} \leq \frac{n^{\gamma'}}{c_5\sqrt{2n}} \leq c_6 n^{-(\frac{1}{2}-\gamma')}.$$

We apply this estimate in (6) as follows:

$$\begin{aligned}\sigma(\mathcal{A}^{n'}) &\leq \sqrt{2}\,\sigma(\mathcal{A}^n)\left(1 + c_4 n^{-\gamma'} + c_3 c_6 n^{-(\frac{1}{2}-\gamma')}\right)\\ &\leq \sqrt{2}\,\sigma(\mathcal{A}^n)\left(1 + c_7 n^{-\alpha}\right),\end{aligned} \tag{7}$$

where $\alpha = \min(\gamma', \frac{1}{2} - \gamma')$. Let us assume that $N \leq x_0$, where x_0 is from Lemma 10. Then for $j = 1, 2, 3, \ldots$ we define $q_i = c_8 \exp \sum_{\ell=-1}^{j-1} c_7 2^{-\alpha\ell}$, where $c_8 \geq 1$ is so that $\sigma(\mathcal{A}^j) \leq c_8\sqrt{j}$ for $j \leq 3x_0$.

Let n be an arbitrary (large enough) integer and construct a finite sequence of integers $\hat{n}_0, \hat{n}_1, \hat{n}_2, \ldots, \hat{n}_r$, so that $\hat{n}_0 = n$ and $\hat{n}_{j+1} = T(\hat{n}_j)$, $j = 0, 1, 2, \ldots, r-1$, where T was defined in Lemma 10 and r is such that $x_0 \leq \hat{n}_r \leq 3x_0 < \hat{n}_{r-1}$. If we put $n_j = \hat{n}_{r-j}$, $m_j = n_{j+1} - 2n_j$ $(j = 0, 1, \ldots, r)$, then by Lemma 10 $n_j^{\gamma'} \leq m_j \leq 3n_j^{\gamma'}$.

We will now show that $\sigma(\mathcal{A}^{n_r}) \leq q_r\sqrt{3p_\infty x_0 n_r}$. Suppose $\sigma(\mathcal{A}^{n_j}) \leq q_j\sqrt{3p_\infty x_0}\,2^{j/2}$ for $j \leq k$ $(k < r)$. If $\sigma(\mathcal{A}^{n_{k+1}}) \leq 2^{(k+1)/2}$ then obviously $\sigma(\mathcal{A}^{n_{k+1}}) \leq q_{k+1}\sqrt{3p_\infty x_0}\,2^{(k+1)/2}$ (induction step is trivial) and if $\sigma(\mathcal{A}^{n_k+1}) > 2^{(k+1)/2}$ then we complete the induction step with (7) as follows:

$$\sigma(\mathcal{A}^{n_{k+1}}) \leq \sqrt{2}\, q_k\sqrt{3p_\infty x_0}\,2^{k/2}\left(1 + c_7 2^{-k\alpha}\right) \leq q_{k+1}\sqrt{3p_\infty x_0}\,2^{(k+1)/2}.$$

Since $\sigma(\mathcal{A}^{n_0}) \leq q_0\sqrt{3p_\infty x_0}\,2^{n_0/2}$ by choice Lemma 11, we conclude that $\sigma(\mathcal{A}^{n_j}) \leq q_j\sqrt{3p_\infty x_0}\,2^{j/2}$ holds for all $j \leq r$ and in particular $\sigma(\mathcal{A}^{n_r}) \leq q_r\sqrt{3p_\infty x_0}\,2^{r/2} \leq q_r\sqrt{3p_\infty x_0 n_r}$. Since n was arbitrary we thus have shown that $\sigma(\mathcal{A}^n) \leq c_9\sqrt{n}$ for all n and some $c_9 \leq q_\infty\sqrt{3p_\infty x_0}$. Put $C_4 = c_9^2$. □

Corollary 13 *The limit*

$$k = \lim_{n\to\infty} \frac{1}{n^2} K(\mathcal{A}^n)$$

exists and is equal to h^2.

Proposition 14 *The limit*

$$\sigma_\infty^2 = \lim_{m\to\infty} \frac{1}{m}\sigma^2(\mathcal{A}^m)$$

exists and is finite. Moreover

$$\left|\sigma_\infty^2 - \frac{\sigma^2(\mathcal{A}^m)}{m}\right| \leq C_5\left(\phi(m^{\gamma'}) + m^{-(\frac{1}{2}-\gamma')}\right),$$

where $\gamma < \gamma' < \frac{1}{2}$ and C_5 depends on γ'.

Proof. By Lemma 12 we only know that $\frac{1}{m}\sigma^2(\mathcal{A}^m)$ is bounded uniformly in m. As in the proof of Lemma 12 let $\gamma < \gamma' < \frac{1}{2}$ and assume that u is large enough (so that $u^{\gamma'} \geq f(u)$. We have that

$$\left|\sigma(\mathcal{A}^u) - \sigma(\mathcal{A}^{u-[u^{\gamma'}]})\right| \leq \sigma\left(\mathcal{A}^u | \mathcal{A}^{u-[u^{\gamma'}]}\right) \leq \sqrt{K\left(\mathcal{A}^{[u^{\gamma'}]}\right)} \leq c_1 u^{\gamma'}.$$

By (6) we get

$$\begin{aligned}\sigma(\mathcal{A}^{2u}) &\le \sqrt{2}\,\sigma(\mathcal{A}^{u-[u^{\gamma'}]})\left(1+\phi([u^{\gamma'}])+c_2u^{-(\frac{1}{2}-\gamma')}\right)\\ &\le \sqrt{2}\,\sigma(\mathcal{A}^{u})\left(1+\phi([u^{\gamma'}])+c_2u^{-(\frac{1}{2}-\gamma')}\right)+3u^{\gamma'}\\ &\le \sqrt{2}\,\sigma(\mathcal{A}^{u})\left(1+\phi([u^{\gamma'}])+c_3u^{-(\frac{1}{2}-\gamma')}\right),\end{aligned}$$

and by iteration we obtain the following upper bound:

$$\sigma(\mathcal{A}^{2^iu}) \le 2^{i/2}\sigma(\mathcal{A}^u)\prod_{j=0}^{i-1}\left(1+\phi([(2^ju)^{\gamma'}])+c_3(2^ju)^{-(\frac{1}{2}-\gamma')}\right)\le$$
$$2^{i/2}\sigma(\mathcal{A}^u)\left(1+c_4\left(\phi(u^{\gamma'})+u^{-(\frac{1}{2}-\gamma')}\right)\right),$$

where we used that $\phi([(2^ju)^{\gamma'}])\le \text{const}.2^{-j\gamma'}\phi(u^{\gamma'})$.

To get an upper bound for arbitrary integers n we proceed as in the proof of Lemma 6. We let $n=km+r$ $(0\le r<m)$ and consider the binary expansion of k: $k=\sum_{i=0}^{\ell}\epsilon_i2^i$ $(\epsilon_i=0,1,$ $\epsilon_\ell=1,\ \ell=[\log_2 k])$. Put $k_j=\sum_{i=0}^{\ell-j}\epsilon_i2^i$ $(k_0=k)$ and $a(u)=\sigma(\mathcal{A}^u)$, then $a(u)\le C_4\sqrt{u}$ (by Lemma 12) and using the inequality (5) yields

$$\begin{aligned}a^2(k_j) &= a^2(\epsilon_{\ell-j}2^{\ell-j}m+k_{j+1}m)\\ &\le a^2(\epsilon_{\ell-j}2^{\ell-j}m)+a^2(k_{j+1}m-[(k_{j+1}m)^{\gamma'}])\\ &\quad +3\phi((k_{j+1}m)^{\gamma'})C_4^2\sqrt{\epsilon_{\ell-j}2^{\ell-j}m}\sqrt{k_{j+1}m-[(k_{j+1}m)^{\gamma'}]}\\ &\le a^2(\epsilon_{\ell-j}2^{\ell-j}m)+a^2(k_{j+1}m)+c_1(k_{j+1}m)^{\gamma'}\\ &\quad +c_5(k_{j+1}m)^{-\gamma'}\sqrt{\epsilon_{\ell-j}2^{\ell-j}m}\sqrt{k_{j+1}m}\\ &\le a^2(\epsilon_{\ell-j}2^{\ell-j}m)+a^2(k_{j+1}m)+c_12^{(\ell-j)\gamma'}m^{\gamma'}+c_5m^{1-\gamma'}2^{(\ell-j)(1-\gamma')}\\ &\le a^2(\epsilon_{\ell-j}2^{\ell-j}m)+a^2(k_{j+1}m)+c_62^{(\ell-j)\beta}m^{\beta}\end{aligned}$$

$(0\le j<\ell)$, where $\beta=\max(\gamma',1-\gamma')<1$. Iterating this formula and summing over j yields

$$\begin{aligned}a^2(km) &\le \sum_{j=0}^{\ell}\left(a^2(\epsilon_{\ell-j}2^{\ell-j}m)+c_62^{(\ell-j)\beta}m^{\beta}\right)\\ &\le \sum_{j=0}^{\ell}\epsilon_{\ell-j}2^{\ell-j}a^2(m)\left(1+c_4\left(\phi(m^{\gamma'})+m^{-(\frac{1}{2}-\gamma')}\right)\right)+c_72^{\ell\beta}m^{\beta}\\ &\le ka^2(m)\left(1+c_4\left(\phi(m^{\gamma'})+m^{-(\frac{1}{2}-\gamma')}\right)ht\right)+c_8(km)^{\beta},\end{aligned}$$

where we used (8) in the second inequality. Since (by Lemma 9) $|a(n)-a(km)|\le\sigma\left(\mathcal{A}^n|\mathcal{A}^{km}\right)\le c_1r\le c_1m$ we get

$$\begin{aligned}\limsup_{n\to\infty}\frac{a^2(n)}{n} &\le \limsup_{k\to\infty}\frac{ka^2(m)\left(1+c_4\left(\phi(m^{\gamma'})+m^{-(\frac{1}{2}-\gamma')}\right)\right)+c_8n^{\beta}+c_1m}{km}\\ &\le \left(1+c_4\left(\phi(m^{\gamma'})+m^{-(\frac{1}{2}-\gamma')}\right)\right)\frac{a^2(m)}{m}\end{aligned}$$

for all m large enough. Hence as $m \to \infty$ the right hand side can be replaced by a lim inf. This proves that the limit $\sigma_\infty = \lim_{n\to\infty} \frac{a^2(m)}{m}$ exists. We moreover have proven the upper bound

$$\sigma_\infty^2 \leq \frac{a^2(m)}{m} + C_5 \left(\phi(m^{\gamma'}) + m^{-(\frac{1}{2}-\gamma')} \right).$$

The lower bound is obtained in the same way. □

2.3 Higher order moments of the information function

In the proof of Theorem 16 we will need and estimate on some higher order moment of the information function. In the following lemma we shall estimate the third order. Below it will become clear that indeed we can estimate all orders although we will not need any of the estimates here.

In the following lemma we shall in fact estimate the quantity

$$N_w(\mathcal{P}) = \sum_{P\in\mathcal{P}} \mu(P) \left| \log \frac{1}{\mu(P)} - H(\mathcal{P}) \right|^w.$$

Lemma 15 *Assume that ϕ is summable and that $f(m) \leq C_3 m^\gamma$, $m \in \mathbf{N}$, for some C_3 and $\gamma \in [0, \frac{1}{2})$.*

Then there exists a constant C_6 so that

$$N_3(\mathcal{A}^n) \leq C_6 n^{3/2} \log n.$$

Proof. Let $\gamma' \in (\gamma, \frac{1}{2})$, n be some (large) integer and $n' = 2n + m$, where $n^{\gamma'} < m < 3n^{\gamma'}$ ($n^{\gamma'} \geq f(n)$). Put $\mathcal{B} = \mathcal{A}^n$, $\mathcal{C} = T^{-m-n}\mathcal{A}^n$. With ρ_i, $i = 1, 2, 3, 4$ as in the proof of Lemma 12, we obtain (as $H(\mathcal{B} \vee \mathcal{C}) = H(\mathcal{B}) + H(\mathcal{C}) + \rho_2$) with Minkowsky's inequality (on L^3 spaces)

$$\begin{aligned} N_3^{\frac{1}{3}}(\mathcal{B} \vee \mathcal{C}) &= \left(\sum_{B\in\mathcal{B}, C\in\mathcal{C}} \mu(B \cap C) \left| \log \frac{1}{\mu(B \cap C)} - H(\mathcal{B} \vee \mathcal{C}) \right|^3 \right)^{\frac{1}{3}} \\ &\leq E_3^{\frac{1}{3}}(\mathcal{B}, \mathcal{C}) + 2\phi(m) \end{aligned}$$

where

$$\begin{aligned} E_3(\mathcal{B}, \mathcal{C}) &= \sum_{B\in\mathcal{B}, C\in\mathcal{C}} \mu(B \cap C) \left| \log \frac{1}{\mu(B)} + \log \frac{1}{\mu(C)} - H(\mathcal{B}) - H(\mathcal{C}) \right|^3 \\ &\leq \sum_{B\in\mathcal{B}, C\in\mathcal{C}} \mu(B \cap C) \left| \log \frac{1}{\mu(B)} - H(\mathcal{B}) \right|^3 \\ &\quad + \sum_{B\in\mathcal{B}, C\in\mathcal{C}} \mu(B \cap C) \left| \log \frac{1}{\mu(C)} - H(\mathcal{C}) \right|^3 + 3F_{12}(\mathcal{B}, \mathcal{C}) + 3F_{21}(\mathcal{B}, \mathcal{C}) \\ &= N_3(\mathcal{B}) + N_3(\mathcal{C}) + 3F_{12}(\mathcal{B}, \mathcal{C}) + 3F_{21}(\mathcal{B}, \mathcal{C}). \end{aligned}$$

We estimate the term $F_{12}(\mathcal{B},\mathcal{C})$ as follows

$$\begin{aligned}
F_{12}(\mathcal{B},\mathcal{C}) &= \sum_{B\in\mathcal{B},C\in\mathcal{C}} \mu(B\cap C)\left|\log\frac{1}{\mu(B)}-H(\mathcal{B})\right|\left(\log\frac{1}{\mu(C)}-H(\mathcal{C})\right)^2\\
&\le \sum_{B\in\mathcal{B},C\in\mathcal{C}} \mu(B)\mu(C)(1+\|\rho_3\|)\left|\log\frac{1}{\mu(B)}-H(\mathcal{B})\right|\left(\log\frac{1}{\mu(C)}-H(\mathcal{C})\right)^2\\
&\le \sum_{B\in\mathcal{B},C\in\mathcal{C}} \mu(B)\sigma^2(\mathcal{C})(1+\phi(m))\left|\log\frac{1}{\mu(B)}-H(\mathcal{B})\right|\\
&\le (1+\phi(m))\sigma^2(\mathcal{C})\sigma(\mathcal{B}).
\end{aligned}$$

(in the last line we used Schwarz's inequality). Hence, by Lemma 12

$$F_{12}(\mathcal{B},\mathcal{C}) \le c_1 n^{\frac{3}{2}}$$

and similarly

$$F_{21}(\mathcal{B},\mathcal{C}) \le c_1 n^{\frac{3}{2}},$$

for some c_1. Hence

$$N_3^{\frac{1}{3}}(\mathcal{B}\vee\mathcal{C}) \le \sqrt[3]{N_3(\mathcal{C})+N_3(\mathcal{B})+6c_1 n^{\frac{3}{2}}}+2\phi(m).$$

To fill in the gap of length m we use Lemma 9 and the estimate on K_3:

$$\left|N_3(\mathcal{A}^{2n+m})-N_3(\mathcal{B}\vee\mathcal{C})\right| \le N_3(\mathcal{A}^{2n+m}|\mathcal{B}\vee\mathcal{C}) \le K_3(\mathcal{A}^m) \le c_2 n^{3\gamma'}.$$

Hence

$$N_3(\mathcal{A}^{n'}) \le N_3(\mathcal{B}\vee\mathcal{C})+c_2 n^{3\gamma'} \le N_3(\mathcal{C})+N_3(\mathcal{B})+c_3 n^{\frac{3}{2}} = 2N_3(\mathcal{A}^n)+c_3 n^{\frac{3}{2}}.$$

Let n be an arbitrary (large enough) integer and construct as in the proof of Lemma 12 numbers n_j, m_j $(j=0,1,\dots,r)$ satisfying $n_{j+1}=2n_j+m_j$ and $n_j^{\gamma'} \le m_j \le 3n_j^{\gamma'}$. By induction one then shows that

$$N_3(\mathcal{A}^{n_r}) \le c_4 r n_r^{\frac{3}{2}},$$

for some c_4 chosen so that $N_3(\mathcal{A}^{n_0}) \le c_4 r n_0^{\frac{3}{2}}$. As $r \sim \log n_r$ we get that $N_3(\mathcal{A}^n) \le C_6 n^{\frac{3}{2}} \log n$ for a suitable constant C_6 and all n. □

In fact, in general we get

$$N_w(\mathcal{A}^n) \le C_6(w) n^{w/2} \log^{w-2} n.$$

where the constant $C_6(w)$ depends on the order w.

3 The Central Limit Theorem for Shannon-McMillan-Breiman

We are interested in the limiting behaviour of the function

$$\Xi_n(t) = \mu\left(\left\{x\in\Omega : \frac{-\log\mu(A_n(x))-nh}{\sigma\sqrt{n}} \ge t\right\}\right)$$

for real valued t and a suitable σ, where h is the metric entropy of μ. The Central Limit Theorem states that this quantity converges to the normal distribution $N(t)=\frac{1}{\sqrt{2\pi}}\int_t^\infty e^{-s^2/2}\,ds$ as n goes to infinity if there exists a suitable σ which is positive. Indeed:

Theorem 16 *Assume that ϕ is summable, $|\mathcal{A}| < \infty$ and that there exists a $\gamma \in [0, \frac{1}{2})$ so that $f(n) \le C_1 n^\gamma$ (for some C_1) for all n.*

Then the limit which defines the variance σ converges:

$$\sigma^2 = \lim_{m\to\infty} \frac{K(\mathcal{A}^m) - H^2(\mathcal{A}^m)}{m},$$

and moreover

$$|\Xi_n(t) - N(t)| \le C_7 \frac{1}{n^\delta}$$

for all t and all $\delta < \min(\frac{1}{4}, \frac{1}{2} - \gamma)$, provided $\sigma \neq 0$.

Proof. The proof proceeds in two stages. We first represent $\log\mu(A_n(x))$ by a sum of 'random variable' that have some independence property. In its course we obtain a representation by a shorter sum of random variables and estimate the error we make in adjusting the cutoff value t. In the second part we reduce to the case where we have independent random variables to which we can apply Berry-Esseen's estimate.

By Lemma 14 $|\sigma_\infty - \sigma_m| \le C_5(\phi(\Delta) + m^{-(\frac{1}{2}-\gamma')})$ where $\sigma_m^2 = \frac{\sigma^2(\mathcal{A}^m)}{m}$.

Assume that $n = rm' - \Delta$ where $m' = m + \Delta$ ($\Delta = f(m)$) and $j = 0, 1, \dots, r$ put $Z_j(x) = -\log\mu(A_{jm+(j-1)\Delta}(x)) - H(\mathcal{A}^{jm+(j-1)\Delta})$. In particular $Z_r(x) = -\log\mu(A_n(x)) - H(\mathcal{A}^n)$ and $Z_1(x) = -\log\mu(A_m(x)) - H(\mathcal{A}^m)$. Let $\varphi(\lambda) = \int e^{i\lambda Z_1}\, d\mu$ and define

$$D_n(\lambda) = \left| \int e^{i\lambda Z_r}\, d\mu - \varphi(\lambda)^r \right|.$$

Then

$$D_n(\lambda) \le \sum_{j=0}^{r-1} |\varphi(\lambda)|^j \left| \int \left(e^{i\lambda Z_{j+1}} - \varphi(\lambda) e^{i\lambda Z_j} \right) d\mu \right|.$$

For every j (note that μ is T-invariant):

$$\int \left(e^{i\lambda \tilde{Z}_{j+1}} - \varphi(\lambda) e^{i\lambda Z_j} \right) d\mu \;=\; \int \left(e^{i\lambda Z_1} e^{i\lambda(\tilde{Z}_{j+1} - Z_1 - Z_j\circ T^{m'})} - \varphi(\lambda) \right) e^{i\lambda Z_j \circ T^{m'}}\, d\mu$$

where

$$\begin{aligned} \tilde{Z}_{j+1}(x) &= -\log\mu(C_{j+1}(T^{jm'}x)) - H(\mathcal{A}^{(j+1)m+j\Delta}) \\ C_{j+1}(x) &= A_m(x) \cap T^{-m'} A_{jm+(j-1)\Delta}(T^{m'}x). \end{aligned}$$

Then

$$\left|\mu(C_{j+1}(x)) - \mu(A_m(x))\mu(A_{jm+(j-1)\Delta}(T^{m'}x))\right| \le \phi(\Delta)\mu(A_m(x))\mu(A_{jm+(j-1)\Delta}(T^{m'}x)).$$

Next, for $\alpha > 0$ let us define

$$\mathcal{G}_j = \left\{ A \in \mathcal{A}^{(j+1)m+j\Delta} : \mu(C_{j+1}(A)) \le e^{n^\alpha} \mu(A) \right\}$$

and $\mathcal{F}_j = \mathcal{A}^{(j+1)m+j\Delta} \setminus \mathcal{G}_j$. Since the (disjoint) union over all possible sets C_j covers the whole space Ω (and thus has measure 1) and there are no more than M^Δ ($M = |\mathcal{A}|$) different ways to fill the 'gaps' we obtain (for large enough n)

$$\mu(F_j) \le e^{-n^\alpha} M^\Delta \sum_{C_{j+1} \in \mathcal{A}^m \vee T^{-m'} \mathcal{A}^{jm+(j-1)\Delta}} \mu(C_{j+1}) \le e^{-n^\alpha} M^\Delta,$$

where $F_j = \bigcup_{A\in\mathcal{F}_j} A$. Similarly, if for $\vartheta \in (0,1)$ we define the disjoint 'slices' of $\mathcal{F}_j$:

$$\mathcal{F}_{j,k} = \left\{ A \in \mathcal{A}^{(j+1)m+j\Delta} : \mu(A) < \vartheta^k e^{-n^\alpha} \mu(C_{j+1}(A)) \le \vartheta^{-1}\mu(A) \right\}$$

(obviously $\mathcal{F}_j = \bigcup_{k=0}^\infty \mathcal{F}_{j,k}$), then

$$\mu(F_{j,k}) \le \vartheta^k e^{-n^\alpha} M^\Delta,$$

$k = 0, 1, \ldots$, where $F_{j,k} = \bigcup_{A\in\mathcal{F}_{j,k}} A$. Otherwise, if $A \in \mathcal{G}_j$ then

$$1 \le \frac{\mu(C_{j+1}(A))}{\mu(A)} \le e^{n^\alpha}$$

and thus

$$0 \le Z_{j+1} - \tilde{Z}_{j+1} \le n^\alpha.$$

as $\log\mu(C_{j+1}(A)) - \log\mu(A) = Z_{j+1} - \tilde{Z}_{j+1}$. (Similarly $|Z_{j+1} - \tilde{Z}_{j+1}| \le \log(\vartheta^k e^{n^\alpha})$ on $F_{j,k}$.) Moreover, since

$$\left| \frac{\mu(C_{j+1}(A))}{\mu(A_m(A))\mu(A_{jm+(j-1)\Delta}(T^{m'}A))} - 1 \right| \le \phi(\Delta)$$

we get

$$\left| \tilde{Z}_{j+1} - Z_1 - Z_j \circ T^{m'} \right| \le \phi(\Delta) + h_j, \tag{8}$$

where $h_j = \left| H(\mathcal{A}^{(j+1)m+j\Delta}) - H(\mathcal{A}^m) - H(\mathcal{A}^{jm+(j-1)\Delta}) \right|$. Next notice that

$$\begin{aligned}
\left| \int_{F_j} \left(e^{i\lambda Z_{j+1}} - e^{i\lambda \tilde{Z}_{j+1}} \right) d\mu \right| &= \left| \int_{F_j} e^{i\lambda Z_{j+1}} \left(1 - e^{i\lambda(\tilde{Z}_{j+1} - Z_{j+1})} \right) d\mu \right| \\
&\le \lambda \sum_{k=0}^\infty \int_{F_{j,k}} |\tilde{Z}_{j+1} - Z_{j+1}|\, d\mu \\
&\le \lambda \sum_{k=0}^\infty |\log(\vartheta^k e^{-n^\alpha})| \mu(F_{j,k}) \\
&\le c_1 \lambda e^{-n^\alpha} M^\Delta
\end{aligned}$$

for some constant c_1, and therefore (where $G_j = \Omega \setminus F_j$)

$$\begin{aligned}
\left| \int \left(e^{i\lambda Z_{j+1}} - e^{i\lambda \tilde{Z}_{j+1}} \right) d\mu \right| &\le \left| \int_{F_j} \left(e^{i\lambda Z_{j+1}} - e^{i\lambda \tilde{Z}_{j+1}} \right) d\mu \right| + \left| \int_{G_j} \left(e^{i\lambda Z_{j+1}} - e^{i\lambda \tilde{Z}_{j+1}} \right) d\mu \right| \\
&\le c_1 \lambda e^{-n^\alpha} M^\Delta + \left| \int_{G_j} e^{i\lambda Z_{j+1}} \left(1 - e^{i\lambda(\tilde{Z}_{j+1} - Z_{j+1})} \right) d\mu \right| \\
&= c_1 \lambda e^{-n^\alpha} M^\Delta + \left| \int_{G_j} e^{i\lambda Z_{j+1}} \mathcal{O}(\lambda n^\alpha)\, d\mu \right|
\end{aligned}$$

(as $\tilde{Z}_{j+1} - Z_{j+1} = \mathcal{O}(n^\alpha)$ on the 'good set'), provided λn^α is small.

Hence

$$\left|\int \left(e^{i\lambda Z_{j+1}} - e^{i\lambda \tilde{Z}_{j+1}}\right) d\mu\right| \le c_2\lambda(M^\Delta e^{-n^\alpha} + n^\alpha).$$

Since by (8) $e^{i\lambda(\tilde{Z}_{j+1} - Z_1 - Z_j\circ T^{m'})} = 1 + \mathcal{O}(\lambda(\phi(\Delta) + h_j))$ we thus obtain (for all j)

$$\begin{aligned}
&\left|\int \left(e^{i\lambda Z_{j+1}} - \varphi(\lambda)e^{i\lambda Z_j\circ T^{m'}}\right) d\mu\right| \\
&\qquad\le \left|\int \left(e^{i\lambda Z_{j+1}} - e^{i\lambda \tilde{Z}_{j+1}}\right) d\mu\right| + \left|\int \left(e^{i\lambda \tilde{Z}_{j+1}} - \varphi(\lambda)e^{i\lambda Z_j}\right) d\mu\right| \\
&\qquad\le c_3\lambda(M^\Delta e^{-n^\alpha} + n^\alpha + \phi(\Delta) + h_j) + |\psi_{1j}(\lambda)|
\end{aligned}$$

where

$$\begin{aligned}
\psi_{1j}(\lambda) &= \int e^{i\lambda(Z_1 + Z_j\circ T^{m'})}\, d\mu - \int e^{i\lambda Z_1}\, d\mu \int e^{i\lambda Z_j}\, d\mu \\
&= \sum_{n=0}^{\infty} \frac{(i\lambda)^n}{n!}\left(\mu((Z_1 + Z_j\circ T^{m'})^n) - \sum_{p_1+p_j=n} \frac{n!}{p_1!p_j!}\mu(Z_1^{p_1})\mu(Z_j^{p_j})\right) \\
&= \sum_{n=0}^{\infty} \frac{(i\lambda)^n}{n!} \sum_{p_1+p_j=n} \frac{n!}{p_1!p_j!}\left(\mu\left(Z_1^{p_1}(Z_j^{p_j}\circ T^{m'})\right) - \mu(Z_1^{p_1})\mu(Z_j^{p_j})\right) \\
&= \sum_{n=2}^{\infty} (i\lambda)^n \sum_{p_1+p_j=n} \frac{1}{p_1!p_j!} \\
&\quad\times \sum_{B_1\in\mathcal{B}_1, B_j\in\mathcal{B}_j} \prod_{k=1,j} \left(\log\frac{1}{\mu(B_k)} - H(\mathcal{B}_k)\right)^{p_k} (\mu(B_1\cap B_j) - \mu(B_1)\mu(B_j)) \\
&= \sum_{n=2}^{\infty} (i\lambda)^n \sum_{p_1+p_j=n} \frac{1}{p_1!p_j!} \\
&\quad\times \sum_{B_1\in\mathcal{B}_1, B_j\in\mathcal{B}_j} \prod_{k=1,j} \left(\log\frac{1}{\mu(B_k)} - H(\mathcal{B}_k)\right)^{p_k} \rho(B_1, B_j)\mu(B_1)\mu(B_j)
\end{aligned}$$

where (by definition) $Z_k(B) = \log\frac{1}{\mu(B)} - H(\mathcal{B}_k)$, $B \in \mathcal{B}_k$ and where we have put $\mathcal{B}_1 = \mathcal{A}^m$, $\mathcal{B}_j = T^{-m'}\mathcal{A}^{jm+(j-1)\Delta}$. By Lemma 2 we have $\mu(B_1\cap B_j) - \mu(B_1)\mu(B_j) = \rho(B_1, B_j)\mu(B_1)\mu(B_j)$ where $|\rho| \le c_4\phi(\Delta)$. A rearrangement of the sums yields

$$\begin{aligned}
\psi_{1j}(\lambda) &= \sum_{B_1\in\mathcal{B}_1, B_j\in\mathcal{B}_j} \rho(B_1, B_j)\mu(B_1)\mu(B_j) \\
&\quad\times \left(\sum_{p_1,p_j=0}^{\infty} \frac{1}{p_1!p_j!}(i\lambda Z_1(B_1))^{p_1}(i\lambda Z_j(T^{m'}B_j))^{p_j} - i\lambda Z_1(B_1) - i\lambda Z_j(T^{m'}B_j) - 1\right) \\
&= \sum_{B_1\in\mathcal{B}_1, B_j\in\mathcal{B}_j} \rho(B_1, B_j)\mu(B_1)\mu(B_j)\left(e^{i\lambda Z_1(B_1)}e^{i\lambda Z_j(T^{m'}B_j)} - 1\right)
\end{aligned}$$

and thus

$$|\psi_{1j}(\lambda)| \le 2\|\rho\|.$$

To get an estimate for small λ differentiation yields

$$\begin{aligned}\frac{d}{d\lambda}\psi_{1j}(\lambda) &= \sum_{B_1\in\mathcal{B}_1, B_j\in\mathcal{B}_j} \rho(B_1,B_j)\mu(B_1)\mu(B_j) \\ &\quad\times\left(iZ_1(B_1)e^{i\lambda Z_1(B_1)}\left(e^{i\lambda Z_j(T^{m'}B_j)}-1\right)+iZ_j(T^{m'}B_j)e^{i\lambda Z_j(T^{m'}B_j)}\left(e^{i\lambda Z_1(B_1)}-1\right)\right)\end{aligned}$$

and therefore by Lemma 12[4]

$$\begin{aligned}|\psi'_{1j}(\lambda)| &\le 2\|\rho\|(\mu(|Z_1|)+\mu(|Z_j|)) \\ &\le 2\|\rho\|(\sigma(|Z_1|)+\sigma(|Z_j|)) \\ &\le c_5\|\rho\|\sqrt{jm+(j-1)\Delta} \\ &\le c_6\|\rho\|\sqrt{n},\end{aligned}$$

and consequently

$$|\psi_{1j}(\lambda)| \le c_7\phi(\Delta)\min(|\lambda|,1) \le c_7\phi(\Delta)|\lambda|$$

since we shall assume that $|\lambda| = |t|/s_n \le 1$, where s_n is defined below ($s_n \sim \sigma_\infty\sqrt{n}$).

This gives us ($h = \max_j h_j$)

$$\begin{aligned}D_n(\lambda) &\le c_8|\lambda|\sum_{j=0}^{r-1}|\varphi(\lambda)|^j\left(M^\Delta e^{-n^\alpha}+n^\alpha+\phi(\Delta)+h_j\right) \\ &\le c_8\frac{|\lambda|}{1-|\varphi(\lambda)|}\left(M^\Delta e^{-n^\alpha}+n^\alpha+\phi(\Delta)+h\right) \\ &\le c_9|\lambda|\left(M^\Delta e^{-n^\alpha}+n^\alpha+\phi(\Delta)+h\right)\end{aligned}$$

Let us approximate the variable Z_r by the sum $\tilde{U}$ of r independent variables $\tilde{Z}_1\circ T^{jm'}$, $j=0,\dots,r-1$, each of which has the same distribution as Z_1. Denote by $\phi_{Z_r}(\lambda) = \int e^{i\lambda Z_r}\,d\mu$ the characteristic function of Z_r and by $\varphi_{\tilde{U}}$ the characteristic function of $\tilde{U}$ (i.e. $\phi_{\tilde{U}}(\lambda) = \phi(\lambda)^r$). Lemma 9.4.1 of [6] applied to the sum $\tilde{U}$ yields (for $\delta = 1$)

$$\left|\varphi_{\tilde{U}}\left(\frac{t}{s_n}\right)-e^{-t^2/2}\right| \le 16\Gamma_3\left(\frac{|t|}{s_n}\right)^3 e^{-t^2/3},$$

provided $|t| \le \frac{s_n^3}{36\Gamma_3}$, where $\Gamma_3 = \sum_{j=0}^{r-1}\mu(|\tilde{Z}_1\circ T^{jm'}|^3)$ and $s_n^2 = \sum_j\sigma^2(\tilde{Z}_1\circ T^{jm'}) = r\sigma^2(Z_1) = rm\sigma_m^2 \sim rm\sigma_\infty^2$. By Lemma 15 $\Gamma_3 \le c_{10}rm^{3/2}\log m$.

By the triangle inequality ($\lambda = t/s_n$ and assuming $t\ge 0$)

$$\begin{aligned}\left|\varphi_{Z_r}(\lambda)-e^{-t^2/2}\right| &\le D_n(\lambda)+16\Gamma_3\left(\frac{t}{s_n}\right)^3 3e^{-t^2/3} \\ &\le c_9\frac{t}{\sqrt{n}}\left(2M^\Delta e^{-n^\alpha}+n^\alpha+\phi(\Delta)+h\right)+c_{11}\frac{t^3}{r^{1/2}}e^{-t^2/3}\log m\end{aligned}$$

[4] also note that for any partition $\mathcal{P}$ one has

$$\sum_{P\in\mathcal{P}}\mu(P)\left|\log\frac{1}{\mu(P)}-H(\mathcal{P})\right| \le \left(\sum_{P\in\mathcal{P}}\mu(P)\left(\log\frac{1}{\mu(P)}-H(\mathcal{P})\right)^2\right)^{\frac{1}{2}} = \sqrt{K(\mathcal{P})-H(\mathcal{P})} = \sigma(\mathcal{P}).$$

(for $t \leq c_{12} r^{3/2}/\log m$). Berry-Esseen's estimate yields with some τ

$$\begin{aligned}
&\left|\mu\left(x : \frac{Z_r(x)}{s_n} > t\right) - N(t)\right| \\
&\quad\leq \frac{2}{\pi}\int_0^\tau \left|\varphi_{Z_r}(\lambda) - e^{-t^2/2}\right| \frac{dt}{t} + \frac{c_{13}}{\tau} \\
&\quad\leq c_{14}\left(\frac{\tau}{\sqrt{n}}(2M^\Delta e^{-n^\alpha} + n^\alpha + \phi(\Delta) + h) + \log m \int_0^\tau \frac{t^2}{r^{1/2}} e^{-t^2/3} dt\right) + \frac{c_{13}}{\tau} \\
&\quad\leq c_{14}\frac{\tau}{\sqrt{n}}(2M^\Delta e^{-n^\alpha} + n^\alpha + \phi(\Delta) + h) + c_{15}\frac{\log m}{r^{1/2}} + \frac{c_{12}}{\tau}
\end{aligned}$$

where $c_{13} \leq \frac{24}{\sqrt{2}\pi^{3/2}}$. Let $\beta \in (0,1)$ and $r \sim n^\beta$. This implies that $m \sim n^{1-\beta}$ and for $\gamma' > \gamma$ we then have $\Delta \sim n^{\gamma'(1-\beta)}$. Moreover for $\alpha > \gamma'(1-\beta)$ we put $\alpha' = \frac{1}{2}(\alpha - \gamma'(1-\beta))$ which can be made arbitrarily small with a suitable choice of α. Hence $M^\Delta e^{-n^\alpha} \leq \text{const.} e^{-n^{\alpha'}}$. Hence if we choose $\tau \sim n^{\frac{1}{4}-\frac{\alpha}{2}}$, then

$$\left|\mu\left(x : \frac{Z_r(x)}{s_n} > t\right) - N(t)\right| \leq c_{16} \max\left(n^{\frac{\alpha}{2}-\frac{1}{4}}, n^{-\frac{\beta}{2}}\log n\right) \leq c_{17} n^{-\delta} \tag{9}$$

where $\delta < \frac{1}{4}$ (independent of the value of γ where we choose $\beta < 1$ arbitrarily close to 1) and c_{17} depends on δ. Here we also used that according to Lemma 6 $|h| \leq C_2 m^{\gamma'} \leq \text{const.} n^{\gamma'(1-\beta)}$ (which implies $|h|\tau/\sqrt{n} \leq \text{const.} n^{\frac{\alpha}{2}-\frac{1}{4}}$).

One has

$$\mu\left(x : \frac{Z_r(x)}{s_n} > t\right) = \mu\left(x : \frac{Z_r(x)}{\sqrt{n}\,\sigma_\infty} > t'\right)$$

where by Proposition 14 $\sigma_m - \sigma_\infty = \mathcal{O}\left(m^{-(\frac{1}{2}-\gamma')}\right)$ (recall that $\phi(k) \leq \text{const.}\frac{1}{k}$). Since in equation (9) the principal term becomes smaller than the error term for $|t| \geq c_{18}\log n$ for some c_{18}, we get

$$|t - t'| \leq c_{19} m^{-(\frac{1}{2}-\gamma')}\log n$$

for all $|t| \geq c_{18}\log n$. Moreover, since by Lemma 6 $|Z_r - (|\log\mu(A_n)| - nh)| \leq C_2 n^{\gamma'}$ and γ' can be chosen arbitrarily close to γ, we thus obtain ($c_{20} \leq c_{17} + c_{19}$)

$$\left|\mu\left(x : \frac{|\log\mu(A_n(x))| - nh}{\sqrt{n}\,\sigma_\infty} > t\right) - N(t)\right| \leq c_{20} n^{-\delta}$$

for all $\delta < \min\left(\frac{1}{4}, \frac{1}{2} - \gamma\right)$ where $\gamma \in [0, \frac{1}{2})$. □

ϕ-mixing maps: Here $\gamma = 0$ and thus we get the bound $\delta < \frac{1}{4}$ independent of the decay of ϕ (ϕ summable).

4 The Law of the Iterated Logarithm

The Central Theorem only establishes the convergence of the distribution of the quantity $\log\mu(A_n))/\sigma\sqrt{n}$. We can now use the rate of convergence to conclude the Law of the Iterated Logarithm applies to the Shannon-McMillan-Breiman theorem.

Theorem 17 *Assume that ϕ is summable, $\mathcal{A}$ a finite partition and $f(n) \leq C_1 n^\gamma$ for some C_1 and $\gamma \in [0, \frac{1}{2})$.*

If $\sigma > 0$ then

$$\limsup_{n\to\infty} \frac{|\log \mu(A_n(x))| - nh(\mu)}{\sigma\sqrt{2n \log\ \log n}} = 1$$

almost everywhere.

A similar statement is true for the lim inf where the limit is then equal to -1 almost everywhere. This theorem follows from Theorem 16 and [23] where it is proven that the LIL is implied if the CLT converges at least at the rate $\frac{1}{(\log n)^{1-\varepsilon}}$ for some $\varepsilon > 0$.

5 The Weak Invariance Principle for Shannon-McMillan-Breiman

The central limit theorem could be improved to get what is called the weak invariance principle (WIP). Such a principle has been obtained for a large class of observables and for a large class of dynamical systems by Chernov in [5]. We prove here the WIP for the function $-\log \mu(A_n(x))$. Let us first recall what the WIP says.

For each $x \in \Omega$ we construct the random variable $W_{n,x}(t)$ for $t \in [0, 1]$ by putting

$$W_{n,x}(k/n) = \frac{-\log \mu(A_k(x)) - kh}{\sigma\sqrt{n}}$$

($h = h(\mu)$ is the metric entropy of μ) extending linearly on each of the subintervals $\left[\frac{k}{n}, \frac{k+1}{n}\right]$. For each x, $W_{n,x}$ is therefore an element of the space $\mathcal{C} = C_\infty([0, 1])$ of the continuous function on $[0, 1]$ topologised with the supremum norm. If we denote with D_n the distribution of $W_{n,x}$ on $\mathcal{C}$, namely

$$D_n(H) = \mu\left(\{x \in \Omega : W_{n,x} \in H\}\right)$$

where H is a Borel subset of $\mathcal{C}$, then the WIP asserts that the distribution D_n converges weakly to the Wiener measure. This means that $\log \mu(A_n(x)) - nh$ is for large n, and after a suitable normalization distributed approximately as the position at time $t = 1$ of a particle in Brownian motion [2]. Recently there has been a great interest in the WIP in relation to the mixing properties of dynamical systems (see also [10, 9, 24]). In the following we assume that $|\mathcal{A}| < \infty$.

Theorem 18 *The information function $-\log \mu(A_n(x))$ satisfies the Weak Invariance Principle provided the variance σ^2 is positive.*

Proof. Let $\tilde{S}_i = -\log \mu(A_i(x)) - ih(\mu)$. We have verify two conditions ([2] Theorem 8.1), namely

(i) The tightness condition (10): We have to show that there exists a $\lambda > 0$ so that for every $\varepsilon > 0$ there exists an N_0 so that

$$\mu\left(\max_{0\leq i\leq n} |\tilde{S}_i| > 2\lambda\sqrt{n}\right) \leq \frac{\varepsilon}{\lambda^2} \tag{10}$$

for all $n \geq N_0$.

(ii) That the finite-dimensional distributions of $\tilde{S}_i$ converge to those of the Wiener measure.

(i) *Proof of tightness*: Let us put $S_i = -\log\mu(A_i(x)) - H(\mathcal{A}^i)$. By Lemma 6 $S_i - \tilde{S}_i = \mathcal{O}(i^\gamma)$. In the usual way (cf. e.g. [2]) we get

$$\mu\left(\max_{0\le i\le n}|S_i| > 2\lambda\sqrt{n}\right) \le \mu\left(|S_n| > \lambda\sqrt{n}\right) + \sum_{i=0}^{n-1}\mu\left(E_i\cap\{|S_i - S_n| \ge \lambda\sqrt{n}\}\right),$$

where E_i is the set of points x so that $|S_i(x)| > 2\lambda\sqrt{n}$ and $|S_k(x)| \le 2\lambda\sqrt{n}$ for $k = 0,\dots,i-1$. Note that E_i lies in the σ-algebra generated by $\mathcal{A}^i$. Also the sets E_i are pairwise disjoint. To estimate $\mu\left(E_i\cap\{|S_i - S_n| \ge \lambda\sqrt{n}\}\right)$ let us put $C_{n,i}(x) = A_i(x)\cap T^{-i-\Delta}A_{n-i-\Delta}(T^{i+\Delta}x)$ $(i < n-\Delta)$ where $\Delta = f(i)$ is the length of the gap, and use the mixing property

$$\mu(C_{n,i}(x)) \le \mu(A_i(x))\mu(A_{n-i-\Delta}(T^{i+\Delta}x))\left(1+\phi(ta)\right).$$

Similar to the proof of Theorem 16 we say an n-cylinder A_n is 'good' if $\mu(A_n) \le \mu(C_{n,i}(A_n)) \le e^{n^\alpha}\mu(A_n)$, where $\alpha < \frac{1}{2}$ will be determined later. The set $B_{n,i}$ of cylinders that are not good has total measure $\mu(B_{n,i}) \le e^{-n^\alpha}M^\Delta$, where $M = |\mathcal{A}|$ (cf. Thm 16). To estimate $|S_i - S_n|$ let us first do the upper estimate:

$$\begin{aligned}
S_n &= \log\frac{1}{\mu(A_n)} - H(\mathcal{A}^n)\\
&\ge \log\frac{1}{\mu(C_{n,i})} - H(\mathcal{A}^n)\\
&\ge \log\frac{1}{\mu(A_i)\mu(A_{n-i-\Delta})(1+\phi(\Delta))} - H(\mathcal{A}^i) - H(\mathcal{A}^{n-i-\Delta}) - H(\mathcal{A}^\Delta)\\
&= S_i - S_{n-i-\Delta} - H(\mathcal{A}^\Delta) - \phi(\Delta).
\end{aligned}$$

Hence $S_i - S_n \le S_{n-i-\Delta} + c_1\Delta$. To get the lower bound we estimate as follows:

$$\begin{aligned}
S_n &\le \log\frac{1}{e^{-n^\alpha}\mu(C_{n,i})} - H(\mathcal{A}^n)\\
&\le n^\alpha + S_i - S_{n-i-\Delta} - H(\mathcal{A}^\Delta) - \phi(\Delta).
\end{aligned}$$

Since for the gap $\Delta = f(i) \le c_2 i^\gamma \le c_2 n^\gamma$ and $\gamma < \frac{1}{2}$ we get

$$|(S_i - S_n) - S_{n-i-\Delta}| \le c_3 n^\alpha \tag{11}$$

where $\gamma \le \alpha < \frac{1}{2}$. In particular we thus get $\mu(B_{n,i}) \le e^{-c_4 n^\alpha}$ for some $0 < c_4 < 1$ and all n large enough.

Now, using the mixing property, we get by the Central Limit Theorem 16 $\mu\left(|S_{n-i-\Delta}| \ge \frac{\lambda}{2}\sqrt{n}\right) \le 2N(\frac{\lambda}{2}) + c_5 n^{-\delta}$ $(c_5 < \infty)$ and therefore

$$\begin{aligned}
\mu\left(E_i\cap\{|S_i - S_n| \ge \lambda\sqrt{n}\}\right) &\le \mu\left(E_i\cap T^{-i-\Delta}\left\{|S_{n-i-\Delta}| \ge \lambda\sqrt{n} - c_3 n^\alpha\right\}\right) + \mu(B_{n,i})\\
&\le \mu(E_i)\mu\left(|S_{n-i-\Delta}| \ge \frac{\lambda}{2}\sqrt{n}\right)(1+\phi(\Delta)) + \mu(B_{n,i})\\
&\le \mu(E_i)\left(2N\left(\frac{\lambda}{2}\right) + c_5 n^{-\delta}\right)(1+\phi(\Delta)) + e^{-c_4 n^\alpha}\\
&\le c_6\mu(E_i)\left(e^{-\lambda} + n^{-\delta}\right) + e^{-c_4 n^\alpha}
\end{aligned}$$

($c_6 > 0$), for all $i < n - c_2 n^\gamma$ and all n large enough (so that $\lambda\sqrt{n} - c_3 n^\alpha \geq \frac{\lambda}{2}\sqrt{n}$).

For $n - c_2 n^\gamma \leq i < n$ let us consider those cylinders A_i for which $\mu(A_n) \leq \mu(A_i) \leq e^{n^\alpha}\mu(A_n)$. The total measure of those cylinders that don't satisfy the inequality is bounded by $e^{-n^\alpha} M^\Delta \leq e^{-c_4 n^\alpha}$. This implies that

$$S_n = \log\frac{1}{\mu(A_n)} - H(\mathcal{A}^n) \geq \log\frac{1}{\mu(A_i)} - H(\mathcal{A}^i) - H(\mathcal{A}^{n-i}) \geq S_i - c_7\Delta,$$

and on the other hand also (by Lemma 6)

$$S_n \leq \log\frac{1}{e^{n^\alpha}\mu(A_i)} - H(\mathcal{A}^i) - H(\mathcal{A}^{n-i}) + C_2 n^\gamma.$$

For large i we thus obtain (for some positive c_8)

$$|S_i - S_n| \leq n^\alpha + c_\Delta + C_2 n^\gamma \leq c_8 n^\alpha$$

except on a set of measure $\leq e^{-n^\alpha} M^\Delta$. Since $c_8 n^\alpha \leq \lambda\sqrt{n}$ for all large enough n we get

$$\mu(|S_i - S_n| \geq \lambda\sqrt{n}) \leq e^{-n^\alpha} M^\Delta \leq e^{-c_4 n^\alpha}$$

for $n - c_2 n^\gamma \leq i < n$.

Finally we obtain (as $\mu\left(|S_n| > \lambda\sqrt{n}\right) \leq 2N(\lambda) + c_5 n^{-\delta}$)

$$\begin{aligned} \mu\left(\max_{0\leq i\leq n}|S_i| > 2\lambda\sqrt{n}\right) &\leq 2N(\lambda) + c_5 n^{-\delta} + (n + c_2 n^\gamma)e^{-c_4 n^\alpha} + c_6 e^{-\lambda}\sum_{i=0}^{n-c_2 n^\gamma}\mu(E_i) \\ &\leq c_9\left(n^{-\delta} + e^{-\lambda}\right). \end{aligned}$$

This proves the tightness condition (10), since for every $\varepsilon > 0$ one can find a $\lambda > 1$ so that the quadratic estimate holds.

(ii) *Proof of the finite-dimensional distribution convergence:* Let us put

$$X_n(t,x) = \frac{1}{\sigma\sqrt{n}}\left(\tilde{S}_{[nt]}(x) + (nt - [nt])\left(\tilde{S}_{[nt]+1}(x) - \tilde{S}_{[nt]}(x)\right)\right)$$

for $t \in [0,1]$. X_n is a random variable defined on Ω and with values in $\mathcal{C}$.

We have to show that the distribution of $(X_n(t,x), X_n(t,x) - X_n(s,x))$ converges to $(\mathcal{N}(0,t), \mathcal{N}(0,t-s))$ $(0 \leq s < t)$ as $n \to \infty$, where $\mathcal{N}(0,t)$ is the normal distribution with zero mean and variance t^2. To prove this as well as the convergence of higher finite dimensional distributions it is sufficient ([2], Theorem 3.2) to show that $X_n(t,x) - X_n(s,x)$ converges to $\mathcal{N}(0, t-s)$. We obtain by Lemma 6 and (11)

$$\begin{aligned} \tilde{S}_{[nt]} - \tilde{S}_{[ns]} &= S_{[nt]} - S_{[ns]} + \mathcal{O}\left((nt)^\gamma\right) \\ &= S_{[n(t-s)]-\Delta} + \mathcal{O}\left((c_3+1)(nt)^\gamma\right) \end{aligned}$$

and by the Central Limit Theorem 16

$$\mu\left(x \in \Omega : \frac{\tilde{S}_{[nt]} - \tilde{S}_{[ns]}}{\sigma\sqrt{n}} \geq \lambda\right)$$

$$
\begin{aligned}
&= \mu\left(x : \frac{S_{[n(t-s)]-\Delta}}{\sigma\sqrt{n}} \geq \lambda + \mathcal{O}\left(\frac{(nt)^\gamma}{\sqrt{n}}\right)\right) + \mathcal{O}(\mu(B_{[nt],[ns]}))\\
&= \mu\left(x : \frac{S_{[n(t-s)]-\Delta}}{\sigma\sqrt{[n(t-s)]-\Delta}} \geq \lambda\sqrt{\frac{n}{[n(t-s)]-\Delta}} + \mathcal{O}\left(n^{\gamma-\frac{1}{2}}\right)\right) + e^{-c_4(nt)^\alpha}\\
&= N\left(\lambda\sqrt{\frac{n}{[n(t-s)]-\Delta}}\right) + \mathcal{O}\left(n^{\gamma-\frac{1}{2}}\right)\\
&= N\left(\frac{\lambda}{\sqrt{t-s}}\right) + \mathcal{O}\left(n^{\gamma-\frac{1}{2}}\right)
\end{aligned}
$$

as $\Delta \leq c_2 n^\gamma$ and where the implied constants are uniformly in n (for all n large enough). Hence $\tilde{S}_{[nt]} - \tilde{S}_{[ns]}$ and therefore $X_n(t,x) - X_n(s,x)$ converges in distribution to $\mathcal{N}(0, \sqrt{t-s})$ as $n \to \infty$.

□

6 Some Results on the First Return Time

Let $W \subset \Omega$ and define the *return time function*

$$\tau_W(x) = \min\{k \geq 1 : T^k x \in W\}.$$

τ_W measures the first entry time for points outside W and (for the first return time for points in W. This function is finite almost everywhere with respect to ergodic measures and satisfies by a theorem of Kac the identity $\int_W \tau_W(x)\, d\mu(x) = 1$ for any ergodic probability measure μ and measurable W. Let us also define the *shortest return time function*

$$\tau(A) = \min_{x \in A} \tau_A(x)$$

which measures the shortest return time within the set A (see [15, 16]). By definition $A \cap T^{-k} A = \emptyset$ for $k = 1, 2, \ldots, \tau(A) - 1$.

In this section we deduce better error estimates for the distribution of the first return time, in particular, compared to the results of [13] we want to remove the factor e^t in the error estimate for the distribution of the first return. We follow a very successful scheme developed by Galves and Schmitt [11].

For an n-cylinder A let us define $\mathcal{O}_S = \{y \in \Omega : \tau_A(y) \geq S\}$.

Lemma 19 *Let A be an n-cylinder so that $\tau(A) \geq \kappa(n/2)$. Then there exists a constant C_8 and $\gamma_4 \in (0,1)$ so that*

$$S\mu(A) \geq 1 - \mu(\mathcal{O}_S) \geq S\mu(A)\,(1 - C_8 \gamma_4^n).$$

Proof. The upper bound is obvious. To prove the lower bound put $B_0 = A$ and define for $j = 1, \ldots, S$

$$B_j = T^{-j}A \setminus \bigcup_{k=0}^{j-1}(T^{-j}A \cap T^{-k}A) \subseteq T^{-j}\left(A \setminus \bigcup_{k=0}^{j-\kappa(n/2)}(A \cap T^{-k+j}A)\right),$$

since by assumption $A\cap T^{-k}A=\emptyset$ for $k<\kappa(n/2)$. As the complement of $\mathcal{O}_S$ is the disjoint union of B_j, we get by invariance of the measure $\mu(B_j)\ge\mu(A)-\sum_{k=\kappa(n/2)}^j\mu(A\cap T^{-k}A)$. For $k\ge n$, we get by the mixing property

$$\mu(A\cap T^{-k}A)\le\mu(A_\ell(A)\cap T^{-k}A)\le c_1\mu(A)\mu(A_\ell(A))\le c_1\mu(A)\gamma_1^\ell,$$

by Lemma 3, where $\ell=\kappa(k)=\left[\frac{k}{1+C_0}\right]$. Thus, for $j\ge 1$:

$$\mu(B_j)\ge\mu(A)-c_1\sum_{k=\kappa(n/2)}^{j}\mu(A)\gamma_1^\ell\ge\mu(A)\left(1-C_8\gamma_1^{\frac{n}{(1+C_0)^2}}\right),$$

(for some C_8), and since $\mu(B_0)=\mu(A)$:

$$1-\mu(\mathcal{O}_r)=\sum_{j=0}^{S}\mu(B_j)\ge\sum_{j=0}^{S}\mu(A)\left(1-C_8\gamma_1^{\frac{n}{(1+C_0)^2}}\right)\ge S\mu(A)\left(1-C_8\gamma_1^{\frac{n}{(1+C_0)^2}}\right).$$

Put $\gamma_4=\gamma_1^{\frac{1}{(1+C_0)^2}}$. □

Lemma 20 *Assume that ϕ is summable and $f(m)\le C_1m^\gamma$ $(0\le\gamma<1)$.*

Then there exists a constant $\gamma_5<1$ so that for all $r>0$, all n large enough and all n-cylinders A for which $\tau(A)\ge\kappa(n/2)$:

$$\left|\mu\left(\mathcal{O}_{[r/\mu(A)]}\right)-e^{-r}\right|\le(r+1)\gamma_5^n.$$

Proof. We use the decomposition

$$\mathcal{O}_R=\mathcal{O}_{S-\Delta}\cap T^{-(S-\Delta)}\mathcal{O}_\Delta\cap T^{-S}\mathcal{O}_{R-S},$$

where $\Delta\ge f(S)$ the length of the 'gap' (assuming that $S<R$ and $S-\Delta$ are large enough). Thus, by T-invariance of μ and the mixing property:

$$\begin{aligned}\left|\mu(\mathcal{O}_R)-\mu(\mathcal{O}_{S-\Delta}\cap T^{-S}\mathcal{O}_{R-S})\right| &\le \mu\left(T^{-(S-\Delta)}\mathcal{O}_\Delta^c\cap T^{-2S}\mathcal{O}_{R-2S}\right)\\ &\le (1+\phi(S))\mu(\mathcal{O}_\Delta^c)\mu(\mathcal{O}_{R-2S})\\ &\le c_1\Delta\mu(A)\mu(\mathcal{O}_{R-2S}) \end{aligned}\tag{12}$$

since $\mu(\mathcal{O}_\Delta^c)\le\Delta\mu(A)$ ($\mathcal{O}^c$ denotes the complement of $\mathcal{O}$) , and similarly

$$\left|\mu(\mathcal{O}_{S-\Delta}\cap T^{-S}\mathcal{O}_{R-S})-\mu(\mathcal{O}_{S-\Delta})\mu(\mathcal{O}_{R-S})\right| \le \phi(\Delta)\mu(\mathcal{O}_{S-\Delta})\mu(\mathcal{O}_{R-S}).\tag{13}$$

Moreover

$$|\mu(\mathcal{O}_S)-\mu(\mathcal{O}_{S-\Delta})|\le c_1\Delta\mu(A).\tag{14}$$

Equations (12), (13) and (14) combined yield (by twice applying the triangle inequality)

$$|\mu(\mathcal{O}_R)-\mu(\mathcal{O}_{R-S})\mu(\mathcal{O}_S)|\le c_1\Delta\mu(A)\left(\mu(\mathcal{O}_{R-2S})+\mu(\mathcal{O}_{R-S})\right)+\phi(\Delta)\mu(\mathcal{O}_{S-\Delta})\mu(\mathcal{O}_{R-S}).$$

Let $\alpha'\in(0,1)$, $r=\mu(A)^{1-\alpha'}$, $R=[r/\mu(A)]\sim\mu(A)^{-\alpha'}$, $m=R/S\sim\mu(A)^{\alpha'(\alpha-1)}$ where $S=[R^\alpha]\sim\mu(A)^{-\alpha\alpha'}$ for $\alpha\in(0,1)$ to be chosen below. Then, since $\phi(\Delta)\le c_4\Delta^{-1}$ (summability of ϕ), $\Delta \le C_1 S^\gamma \sim \mu(A)^{-\alpha\gamma\alpha'}$, $\mu(\mathcal{O}_S) \ge 1 - S\mu(A) \sim 1 - \mu(a)^{1-\alpha\alpha'} \ge \frac{1}{2}$ say. If we assume the induction hypothesis $\mu(\mathcal{O}_{R-2S})/\mu(\mathcal{O}_{R-S}) \le c_2$ for some $c_2 > 1$ then if $\gamma\alpha\alpha' = \frac{1}{2}$ one has

$$\left|\frac{\mu(\mathcal{O}_R)}{\mu(\mathcal{O}_{R-S})\mu(\mathcal{O}_S)} - 1\right| \le c_1 \Delta \frac{\mu(A)}{\mu(\mathcal{O}_S)} \left(\frac{\mu(\mathcal{O}_{R-2S})}{\mu(\mathcal{O}_{R-S})} + 1\right) + \phi(\Delta)\frac{\mu(\mathcal{O}_{S-\Delta})}{\mu(\mathcal{O}_S)} \le c_3 \sqrt{\mu(A)},$$

where we used that $\frac{\mu(\mathcal{O}_{S-\Delta})}{\mu(\mathcal{O}_S)} \le$ const.. Since $\frac{\mu(\mathcal{O}_R)}{\mu(\mathcal{O}_S)^m} = \prod_{k=1}^{m-1} \frac{\mu(\mathcal{O}_{(k+1)S})}{\mu(\mathcal{O}_{kS})\mu(\mathcal{O}_S)}$ we get

$$\left|\frac{\mu(\mathcal{O}_R)}{\mu(\mathcal{O}_S)^m} - 1\right| \le c_4 \sum_{k=1}^{m-1} \left|\frac{\mu(\mathcal{O}_{(k+1)S})}{\mu(\mathcal{O}_{kS})\mu(\mathcal{O}_S)} - 1\right| \le c_4 c_3 m \sqrt{\mu(A)} \le c_5 \mu(A)^\beta$$

where $\beta = \frac{1}{2}(1 - \frac{1}{\gamma}(\frac{1}{\alpha} - 1))$ is positive for suitable α' and $\alpha = \frac{1}{2\gamma\alpha'}$). Since $\mu(A)$ decays exponentially fast in n, we obtain in particular that $\frac{\mu(\mathcal{O}_{R-S})}{\mu(\mathcal{O}_R)} \le \frac{1}{\mu(\mathcal{O}_S)}(1 + c_6 c_5 \mu(A)^\beta)$ (for some universal constant c_6) satisfying the induction hypothesis for the next iterate.

By Lemma 19

$$\begin{aligned}
\left|e^{Sm\mu(A)} - \mu(\mathcal{O}_S)^m\right| &\le m\left|e^{S\mu(A)} - \mu(\mathcal{O}_S)\right| \\
&\le c_7 m \max\left(C_8 \gamma_4^n S\mu(A), (S\mu(A))^2\right) \\
&\le c_8 \mu(A)^{\beta'},
\end{aligned}$$

for some $\beta' > 0$, and therefore by Lemma 3 ($Sm\mu(A) = r$)

$$\left|\mu(\mathcal{O}_R) - e^{-r}\right| \le \left|\mu(\mathcal{O}_R) - \mu(\mathcal{O}_S)^m\right| + \left|\mu(\mathcal{O}_S)^m - e^{-r}\right| \le c_9 \gamma_5^n,$$

for some $\gamma_5 < 1$. □

Lemma 21 *Let $a > 0$. Then there exist $\gamma_6 < 1$ and a constant C_9 so that for all $A \in \mathcal{A}^n$ ($\tau(A) \ge \kappa(n/2)$) and $s \ge e^{an}$:*

$$\left|\mu\left(\{x \in A : R_n(x) \ge s\}\right) - \mu(A)e^{-\mu(A)s}\right| \le C_9 \mu(A)\gamma_6^n \left(\mu(A)s + 1\right).$$

Proof. If ϕ is summable then we have by Lemma 20 for every n-cylinder A for which $\tau(A) \ge \kappa(n/2)$:

$$\left|\mu(\mathcal{O}_s - e^{-s\mu(A)}\right| \le c_1 \gamma_5^n (s\mu(A) + 1) \tag{15}$$

for some positive β. We can assume that $t > \kappa(n/2) + \Delta$, $\Delta \ge f(\kappa(n/2))$, which then gives us

$$\begin{aligned}
&\left|\mu(A \cap \{x : R_n(x) \ge s\}) - \mu(A)\mu(\mathcal{O}_s)\right| \\
&\quad \le \mathcal{E} + \mu(A)\left|\mu(\mathcal{O}_{s-n-\Delta}) - \mu(\mathcal{O}_s)\right| + \left|\mu(A \cap T^{-n-\Delta}\mathcal{O}_{s-n-\Delta}) - \mu(A)\mu(\mathcal{O}_{s-n-\Delta})\right| \\
&\quad \le \mathcal{E} + (n+\Delta)\mu(A)^2 + \phi(\Delta)\mu(A)\mu(\mathcal{O}_{s-n-\Delta}),
\end{aligned}$$

where ($\kappa(j) = \left[\frac{j}{1+C_0}\right]$) by Lemmas 4 and 3 (as $\tau(A) \le \kappa(n/2)$)

$$\begin{aligned}
\mathcal{E} &= \left|\mu(A \cap \mathcal{O}_s) - \mu(A \cap T^{-n-\Delta}\mathcal{O}_{t-n-\Delta})\right| \\
&\le \sum_{j=\kappa(n/2)}^{n+\Delta} \mu(A_{\kappa(j)}(A) \cap A) \\
&\le \mu(A) \sum_{j=\kappa(n/2)}^{n+\Delta} (1 + \phi(\kappa(j)))\mu(A_{\kappa(j)}(A)) \\
&\le c_2 \gamma'^n \mu(A).
\end{aligned}$$

for some c_2. We put $\gamma' = \gamma_1^{\frac{1}{(1+C_0)^2}}$ (i.e. $\gamma'^n = \gamma_1^{\kappa(\kappa(n/2))}$) and $A_\ell(A)$ for the ℓ-cylinder that contains A.

If we choose $\Delta = [\min(\gamma_1^{-n/2}, e^{an})]$ then

$$|\mu(A \cap \{x : R_n(x) \geq s\}) - \mu(A)\mu(\mathcal{O}_s)| \leq c_3 \gamma_6^n \mu(A),$$

for any $\gamma_6 > \max(\gamma', \sqrt{\gamma_1}, e^{-a})$ ($\gamma_6 < 1$ and c_3 chosen suitably) as $\phi(\Delta) < 1/\Delta$ for all large enough Δ. With equation (15) we get the statement. □

7 The CLT and WIP for Repeat times

For $x \in \Omega$ and $n = 1, 2, \ldots$, we denote by $A_n(x)$ the (unique) atom in $\mathcal{A}^n$ which contains x. The 'repeat function' R_n which is then given by $R_n(x) = \tau_{A_n(x)}(x)$.

Theorem 22 *Assume that ϕ is summable and that there exists a $\gamma \in [0, \frac{1}{2})$ so that $f(n) \leq C_1 n^\gamma$ for all n.*

Then, if $\sigma > 0$:

$$\left|\mu\left(\left\{x \in \Omega : \frac{\log R_n(x) - nh}{\sigma\sqrt{n}} \geq t\right\}\right) - N(t)\right| \leq C_{10}\frac{1}{n^\delta},$$

where $\sigma = \lim_{m\to\infty} \frac{\sigma(\mathcal{A}^m)}{\sqrt{m}}$ and $\delta < \min(\frac{1}{4}, \frac{1}{2} - \gamma)$.

By a general result of B Saussol [27] the CLT for $\log \mu(A_n(x))$ and the exponential law for the first return time implies the Central Limit Theorem for the repeat time. Here however we are interested in the rate of the convergence, which in particular required us to obtain in the last section error estimates for the well-know exponential limiting statistics of the first return time.

Proof. For $s_n = e^{nh+t\sqrt{n}}$ then we get by Lemma 21 (with $a = h/2$)

$$\left|\mu\left(\{x \in A : R_n(x) \geq s_n\}\right) - \mu(A)e^{-\mu(A)s_n}\right| \leq C_9\mu(A)\gamma_6^n\left(\mu(A)e^{nh+t\sqrt{n}} + 1\right).$$

Let $\varepsilon = n^{-\delta}$, where according to Theorem 16 $\delta < \min(\frac{1}{4}, \frac{1}{2} - \gamma)$. If η is so that $N(\eta) < \varepsilon$ ($\eta <$ const.$|\log\varepsilon| \sim \log n$) then $\Omega_\eta = \left\{x \in \Omega : \frac{\log\mu(A_n(x)) + nh}{\sigma\sqrt{n}} > \eta\right\}$ (Ω_η is a union of n-cylinders) has measure less than $c_1 n^{-\delta}$ for all n large enough. Hence, if Y_n is the random variable defined as

$$Y_n = \sum_{A\in\mathcal{A}^n} e^{-\mu(A)e^{nh+t\sqrt{n}}}\chi_A$$

(χ_A is the characteristic function of A) we obtain by Lemma 5

$$\begin{aligned}
&\left|\mu\left(\{x \in \Omega : R_n(x) \geq s_n\}\right) - \int Y_n(x)\,d\mu(x)\right| \\
&\quad\leq c_1 n^{-\delta} + C_9\gamma_6^n \sum_{A\in\mathcal{A}^n, A\subset\Omega_\eta^c} \mu(A)\left(\mu(A)e^{nh+t\sqrt{n}} + 1\right) + \gamma_3^n \\
&\quad\leq c_1 n^{-\delta} + C_9\gamma_6^n(e^{\eta\sigma\sqrt{n}+t\sqrt{n}} + 1) \\
&\quad\leq c_1 n^{-\delta} + \gamma_6^{n/2},
\end{aligned}$$

if $t \le \eta$ (assuming $\gamma_6 \ge \gamma_3$) for all large enough n, since $\mu(A_n(x))e^{nh} \le e^{\eta\sigma\sqrt{n}}$ for all $x \notin \Omega_\eta$ and large n. By Markov's inequality

$$\begin{aligned}
\int Y_n(x)\,d\mu(x) &\ge e^{-e^{-\varepsilon\sqrt{n}}}\mu\left(\left\{x : \log Y_n(x) \ge -e^{-\varepsilon\sqrt{n}}\right\}\right) \\
&\ge (1-e^{-\varepsilon\sqrt{n}})\mu\left(\left\{x \in \Omega : \frac{|\log\mu(A_n(x))| - nh}{\sigma\sqrt{n}} \ge t - \frac{\varepsilon}{\sigma}\right\}\right) \\
&\ge N\left(t - \frac{\varepsilon}{\sigma}\right) - C_7\frac{1}{n^\delta} - e^{-n^{\frac{1}{4}}} \\
&\ge N(t) - c_2\frac{1}{n^\delta}
\end{aligned}$$

by Theorem 16. For $t > \eta$ note that $N(t) \le N(\eta) \le \text{const.}n^{-\delta}$ and

$$\mu\left(\left\{x \in \Omega : R_n(x) \ge e^{nh+t\sqrt{n}}\right\}\right) \le \mu\left(\left\{x \in \Omega : R_n(x) \ge e^{nh+\eta\sqrt{n}}\right\}\right).$$

For the upper bound we estimate as follows:

$$\begin{aligned}
\mu(Y_n) &\le e^{-e^{-\varepsilon\sqrt{n}}}\mu\left(\left\{x : \log Y_n(x) < -e^{-\varepsilon\sqrt{n}}\right\}\right) + \mu\left(\left\{x : \log Y_n(x) \ge -e^{-\varepsilon\sqrt{n}}\right\}\right) \\
&\le N(t) + c_3\frac{1}{n^\delta},
\end{aligned}$$

for $t \le \eta$ where we used that $\log Y_n(x) = -\mu(A_n(x))e^{nh+t\sqrt{n}}$ and hence $Y_n \le 1$ on Ω_η. For $t > \eta$ we do as above. This proves the theorem. □

Theorem 23 *The WIP holds for the repeat time $R_n(x)$.*

Proof. We shall show that

$$\lim_{n\to\infty}\frac{\log(R_n(x)A_n(x))}{n^\beta} = 0 \tag{16}$$

almost everywhere for all positive β. Let $r(n) = \nu n^\beta$ and

$$Z_n = \{x : \log(R_n(x)\mu(A_n(x))) \le -r(n)\}.$$

Lemmas 4 and 21 (with $t = e^{-r(n)}/\mu(A)$ and using the fact that $\mu(A) \ge e^{-an}$ for some $a > 0$) then yield

$$\begin{aligned}
\mu(Z_n) &= \sum_{A\in\mathcal{A}^n}\mu(A)\mu_A(\{x \in A : R_n(x)\mu(A(x)) \le e^{-r(n)}\}) \\
&\le \sum_{A\in\mathcal{A}^n}\mu(A)\left(\left|\mu_A(\{x \in A : R_n(x)\mu(A(x)) \ge e^{-r(n)}\}) - e^{-e^{-r(n)}}\right| + \left(1 - e^{-e^{-r(n)}}\right)\right) \\
&\le \sum_{A\in\mathcal{A}^n,\tau(A)<\kappa(n/2)}\mu(A) + \sum_{A\in\mathcal{A}^n,\tau(A)\ge\kappa(n/2)} C_{10}\gamma_6^n\mu(A)\left(\mu(A)e^{-r(n)} + 1\right) + e^{-r(n)} \\
&\le \gamma_2^n + c_1\gamma_6^n + e^{-r(n)} \\
&\le c_2e^{-r(n)}
\end{aligned}$$

for some c_2. Hence $\sum_n \mu(Z_n)$ is finite and by Borel-Cantelli there exists a function $N(x)$ which is almost everywhere finite so that $\log(R_n(x)A_n(x)) > -r(n)$ for all $n \ge N(x)$. Therefore

$$\liminf_n \frac{\log(R_n(x)A_n(x))}{n^\beta} > -\nu$$

for any positive ν. In a similar way one shows that $\limsup_n \frac{\log(R_n(x)A_n(x))}{n^\beta} \leq 0$. Hence (16) has been proven. By standard measure theoretical arguments it then follows that

$$\mu\left(\left\{x : \max_{\ell\leq n} \frac{|\log\mu(A_m(x)) + \log R_n(x)|}{\sigma\sqrt{n}} \geq \epsilon\right\}\right) \longrightarrow 0$$

and by Theorem 18 and Theorem 4.1 of [2] the WIP for the repeat time R_n follows. □

References

[1] M Abadi: Exponential Approximation for Hitting Times in Mixing Processes; Math. Phys. Elec. J., 7, N0 2, (2001)

[2] P Billingsley: *Convergence in Probability Measures*; Wiley 1968

[3] A Broise: Transformations dilatantes de l'intervalle et théorèmes limites; Asterisque 238, 1996

[4] H Bruin and S Vaienti: Return times for unimodal maps; Fundamenta Mathematicae 176 (2003), 77–94

[5] N. Chernov *Limit theorems and Markov approximations for chaotic dynamical systems*; Prob. Th. Rel. Fields **101** (1995), 321–362

[6] Y S Chow and H Teicher: *Probability Theory, Independence, Interchangeability, Martingales*, Springer, 2nd edition 1988

[7] P Collet and A Galves: Repetition times for Gibbsian sources; Nonlinearity 12 (1999) 1225–1237

[8] P Collet, A Galves and B Schmitt: Fluctuations of repetition times for Gibbsian sources; Nonlinearity 12 (1999) 1225–1237

[9] P Ferrero, N Haydn and S Vaienti: Entropy fluctuations for parabolic maps; Nonlinearity, 16, (2003), 1203–1218

[10] M Field, I Melbourne and A Török: Decay of Correlations, Central Limit Theorems and Approximations by Brownian Motion for Compact Lie Group Extensions; Ergod. Theor. & Dynam. Syst. 23 (2003), 87–110

[11] A Galves and B Schmitt: Inequalities for hitting times in mixing dynamical systems; Random and Computational Dynamics 5 (1997), 337–347

[12] M Gordin: The central limit theorem for stationary processes; Soviet Math. Doklady 10 (1969), 1174–1176

[13] N T A Haydn and S Vaienti: The limiting distibution and error terms for return times of dynamical systems; Discrete and Continuous Dynamical Systems, 10, (2004), 584–616

[14] N T A Haydn and S Vaienti: The distribution of the measure of cylinder sets for non-Gibbsian measures; in *Complex Dynamics and Related Topics*, International Press 2003, 147–162

[15] M Hirata: Poisson law for the dynamical systems with the "self-mixing" conditions; *Dynamical Systems and Chaos*, Vol. 1 (Worlds Sci. Publishing, River Edge, New York (1995), 87–96

[16] M Hirata, B Saussol and S Vaienti: Statistics of return times: a general framework and new applications. Commun. Math. Phys. 206 (1999), 33–55

[17] I A Ibragimov: Some limit theorems for stationary processes; Theory Prob. Appl. 7 (1962), 349–382

[18] I Kontoyiannis: Asymptotic Recurrence and Waiting Times for Stationary Processes; J. Theor. Prob. 11 (1998), 795–811

[19] C Liverani: Central Limit theorem for deterministic systems; Intern. Congress on Dyn. Syst., Montevideo 1995 (Proc. Research Notes in Math. Series), Pitman (1996), 56–75

[20] Ornstein and Weiss; Entropy and Data Compression Schemes; IEEE Transactions on Information Theory 39 (1993), 78–83

[21] F Paccaut: Propriétés Statistiques de Systèmes Dynamiques Non Markovian; PhD Thesis Dijon 2000

[22] F Pène: Rates of Convergence in the CLT for Two-Dimensional Dispersive Billiards; Commun. Math. Phys. 225 (2002), 91–119

[23] V V Petrov: On a relation between an estimate of the remainder in the central limit theorem and the law of the iterated logarithm; Teor. Veroyatn. Primen. 11 (3), 514–518. English translation: Theor. Probab. Appl. 11(3), 454–458.

[24] M Pollicott and R Sharp: Invariance Principle for interval maps with and indifferent fixed point; Commun. Math. Phys. 229 (2002), 337–346

[25] M Rosenblatt: A central limit theorem and a mixing condition; Proc. Nat. Acad. USA 42 (1956), 412–413

[26] J Rousseau-Egele: Un théorème de la limite locale pour une classe de transformations dilatants et monotone par morcaux; Annals of Probability 3 (1983), 772–788

[27] B Saussol: Exponential statistics of return times; Nonlinearity 14 (2001), 179–191