The Almost Sure Invariance Principle for Beta-Mixing Measures

Nicolai Haydn

Abstract The theorem of Shannon–McMillan–Breiman states that for every generating partition on an ergodic system of finite entropy the exponential decay rate of the measure of cylinder sets equals the metric entropy almost everywhere. In addition the measure of n-cylinders is in various settings known to be lognormally distributed in the limit. In this paper the logarithm of the measure of n-cylinder, the information function, satisfies the almost sure invariance principle in the case in which the measure is β-mixing. We get a similar result for the recurrence time. Previous results are due to Philipp and Stout who deduced the ASIP when the measure is strong mixing and satisfies an $\mathscr{L}^1$-type Gibbs condition. We also prove the ASIP for the recurrence time.

Keywords Shannon–McMillan–Breiman · Beta-mixing measures · Almost sure invariance principle · Central limit theorem · Recurrence time

1 Introduction

Let μ be a T-invariant probability measure on a space Ω on which the map T acts measurably. For a measurable partition $\mathcal{A}$ one forms the nth join $\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}$ which forms a finer partition of Ω. (The atoms of $\mathcal{A}^n$ are traditionally called *n-cylinders*.) For $x \in \Omega$ we denote by $A_n(x) \in \mathcal{A}^n$ the n-cylinder which contains x. The Theorem of Shannon–McMillan–Breiman (see e.g. [18,24]) then states that for μ-almost every x in Ω the limit

$$\lim_{n \to \infty} \frac{-\log \mu(A_n(x))}{n}$$

exists and equals the metric entropy $h(\mu)$ provided the entropy is finite in the case of a countable infinite partition. The convergence is uniform only in degenerate cases (see [6] for an example). This theorem was proved for finite partitions in increasing degrees of generality

N. Haydn (✉)
Mathematics Department, USC, Los Angeles 90089-1113, USA
e-mail: nhaydn@usc.edu

in the years 1948–1960 first for finite partitions and then for countably infinite partitions. For a setting on metric spaces and with Bowen balls instead of cylinders, Brin and Katok [2] proved a similar almost sure limiting result.

Related to the SMB theorem are recurrence and waiting times for which limiting result were proven by Ornstein and Weiss [20,21] and Nobel and Wyner [19] respectively. Here we are interested in a more detailed description of the limiting distribution of the information function $I_n(x) = -\log \mu(A_n(x))$ around its mean value. These properties are of interest when evaluating the efficiency of compression algorithms in information theory.

In 1962 Ibragimov [15] proved the Central Limit Theorem for SMB for measures that are strongly mixing (in Rosenblatt's sense [27]) and satisfy an $\mathscr{L}^1$-type Gibbs condition, that is, he proved that I_n is in the limit lognormally distributed. Various improvements followed although most of them following Ibragimov's arguments or assume that the measure is Gibbs. For instance, Collet et al. [6] proved that I_n is lognormally distributed in the limit for ψ-mixing Gibbs measures, Paccot [22] for interval maps with suitable topological covering properties. For other results see for instance [3,5,10,17,23]. For Gibbs measures on non-uniformly expanding systems such results have been obtained in [4,8]. For (ψ, f)-mixing measures a CLT was proven in [14], for rational maps with critical points in the Julia set in [13] and for β-mixing maps in [12]. This latter result does not require Ibragimov's $\mathscr{L}^1$-Gibbs condition, but in return asks for the somewhat stronger mixing property, that is β-mixing instead of the strong mixing property.

In the setting of Ibragimov, Philipp and Stout [25] then proved the almost sure invariance principle under similar conditions although with faster decay and better rates of approximability of the conditional entropy function. Kontoyiannis [16] then used this result to prove the almost sure invariance principle, CLT and the law of the iterated logarithm LIL for recurrence and waiting times, strengthening the result of Nobel and Wyner [19]. Also Han [11] proved the ASIP for SMB in the case of exponentially ψ-mixing systems following Philipp and Stout. In the present paper we prove the ASIP for SMB for measures that are β-mixing. Here we don't require the $\mathscr{L}^1$-Gibbs property of Ibragimov and Philipp and Stout. Also we allow for countably infinite partitions. These two aspects are the novelties of the present paper. For finite partitions we also deduct the ASIP for the recurrence function (Theorem 6).

In Sect. 2 we define mixing conditions and state the main theorem. In Sect. 3 we show existence of the variance and give the rate of convergence. We also obtain estimates on the growthrate of the higher order moments of the centred information function. These estimates are important in Sect. 4 where we prove the ASIP following the road laid out in [25].

2 Main Results

Let T be a map on a space Ω and μ a probability measure on Ω. Moreover let $\mathcal{A}$ be a (possibly infinite) 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 $\mathcal{A}^* = \bigcup_{n=1}^{\infty} \mathcal{A}^n$ for the collection of all cylinders in Ω and put $|A|$ for the length of a 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.

2.1 Mixing

The main assumption in the results described here is on the mixing property of the invariant measure. Here we use the following:

Definition 1 We say the invariant probability measure μ is β-mixing[1] if there exists a decreasing function $\beta : \mathbb{N} \to \mathbb{R}^+$ which satisfies $\beta(\Delta) \to 0$ as $\Delta \to \infty$ so that

$$\sum_{(B,C) \in \mathcal{A}^n \times T^{-\Delta-n}\mathcal{A}^m} |\mu(B \cap C) - \mu(B)\mu(C)| \le \beta(\Delta)$$

for every every $n, m, \Delta > 0$.

Other Kinds of Mixing: For comparison purposes we list here some other kinds of mixing which are commonly used in dynamics. Below U is always in the σ-algebra generated by $\mathcal{A}^n$ and V lies in the σ-algebra generated by $\mathcal{A}^*$ (see also [7]). The limiting behaviour described is as the length of the 'gap' Δ tends to infinity:

1. ψ-*mixing:* $\sup_n \sup_{U,V} \left| \dfrac{\mu(U \cap T^{-\Delta-n}V)}{\mu(U)\mu(V)} - 1 \right| \to 0.$

2. *Left ϕ-mixing:* $\sup_n \sup_{U,V} \left| \dfrac{\mu(U \cap T^{-\Delta-n}V)}{\mu(U)} - \mu(V) \right| \to 0.$

3. *Strong mixing* [15,27] *(also called α-mixing):* $\sup_n \sup_{U,V} \left| \mu(U \cap T^{-\Delta-n}V) - \mu(U)\mu(V) \right| \to 0.$

4. *Uniform mixing* [27,28]: $\sup_n \sup_{U,V} \left| \dfrac{1}{k} \sum_{j=1}^k \mu(U \cap T^{-n-j}V) - \mu(U)\mu(V) \right| \to 0$ as $k \to \infty$.

One can also have *right ϕ-mixing* when $\sup_n \sup_{U,V} \left| \frac{\mu(U \cap T^{-\Delta-n}V)}{\mu(V)} - \mu(U) \right| \to 0$ as $\Delta \to \infty$. Clearly ψ-mixing is the strongest mixing property and implies the other kinds of mixing. The next strongest is ϕ-mixing, then comes strong mixing and uniform mixing is the weakest. The β mixing property is stronger that the strong mixing property but is implied by the ϕ-mixing property.

One says μ has the weak Bernoulli property (with respect to the partition $\mathcal{A}$) if for every $\varepsilon > 0$ there exists an $N(\varepsilon)$ so that

$$\sum_{B \in \mathcal{A}^n} |\mu(B \cap C) - \mu(B)\mu(C)| \le \varepsilon\mu(C)$$

for every $C \in T^{-\Delta-n}\mathcal{A}^m$, $\Delta > N$ and $n, m \in \mathbb{N}$ (see e.g. [24]). We see that the β-mixing property implies the weak Bernoulli property. The rate β determines how fast the function $N(\varepsilon)$ grows as ε goes to zero, where to be precise $N(\varepsilon) = \beta^{-1}(\varepsilon)$.

For a partition $\mathcal{A}$ we have the (n-th) *information function* $I_n(x) = -\log \mu(A_n(x))$, where $A_n(x)$ denotes the unique n-cylinder that contains the point $x \in \Omega$, whose moments are

$$K_w(\mathcal{A}) = \sum_{A \in \mathcal{A}} \mu(A)|\log \mu(A)|^w = \mathbb{E}(I_n^w),$$

$w \ge 0$ not necessarily integer. (For $w = 1$ one traditionally writes $H(\mathcal{A}) = K_1(\mathcal{A}) = \sum_{A \in \mathcal{A}} -\mu(A)\log \mu(A)$.) If $\mathcal{A}$ is finite then $K_w(\mathcal{A}) < \infty$ for all w. For infinite partitions the theorem of Shannon–McMillan–Breiman requires that $H(\mathcal{A})$ be finite to ensure finiteness of the entropy. We will require finiteness of a larger than fifth moment $K_w(\mathcal{A})$ for some $w > 5$ (not necessarily integer).

---

[1] In [12] we used the term 'uniform strong mixing' for what is commonly called β-mixing. Here we adhere to the standard terminology.

2.2 Results

Our main result is the following theorem.

Theorem 2 *Let μ be a β-mixing invariant probability measure on Ω with respect to a countably infinite, measurable and generating partition $\mathcal{A}$ which satisfies $K_w(\mathcal{A}) < \infty$ for some $w > 5$. Assume that β decays at least polynomially with power $p > 7 + \frac{30}{w-5}$.*

If the variance is positive, then the information function $I_n(x) = -\log \mu(A_n(x))$ satisfies the almost sure invariance principle for any error exponent $\delta \leq \frac{1}{3} - \frac{10}{3p+30}$. That is, there exists a Brownian motion $B(n)$ such that

$$I_n = \sigma B(n) + \mathcal{O}\left(n^{\frac{1}{2}-\delta}\right)$$

almost surely for any $\delta \leq \frac{1}{3} - \frac{10}{3p+30}$. Moreover the variance σ^2 is given by

$$\sigma^2 = \lim_{n\to\infty} \frac{K_2(\mathcal{A}^n) - H^2(\mathcal{A}^n)}{n}$$

where the limit exists (and is strictly positive if the partition is infinite).

The variance σ^2 is determined in Proposition 11 and essentially only requires finiteness of the second moment $K_2(\mathcal{A})$ although the rate of convergence uses that we have higher moments available. We obtain the following special cases using (8).

Corollary 3 *Suppose $\sigma > 0$. Then:*

(i) *If $w = \infty$, e.g. if $\mathcal{A}$ is finite, and β decays at least polynomially with power $p > 7$, then $I_n = \sigma B(n) + \mathcal{O}(n^{\frac{1}{2}-\delta})$ almost surely for any $\delta < \frac{1}{3} - \frac{10}{3p+30}$.*

(ii) *If β decays superpolynomially then I_n satisfies the ASIP with exponent $\delta < \frac{1}{3}$ for any $w > 5$.*

Philipp and Stout [25] Theorem 9.1 proved the ASIP under the condition that the measure is strong mixing where $\phi(\Delta) = \mathcal{O}(\Delta^{-336})$ and requires that the $\mathcal{L}^1$-norms of the differences $f - f_n$ decay polynomially with power ≥ 48, where $f = \lim_{n\to\infty} f_n$ and

$$f_n(x) = \log \mathbb{P}(x_0|x_{-1}x_{-2}\ldots x_{-n}).$$

The ASIP holds then for any $\delta < \frac{1}{294}$.

If $N(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-s^2/2} ds$ denotes the normal distribution and $h(\mu)$ the metric entropy of μ then we also have the CLT ([12]):

Theorem 4 *Let μ be a β-mixing probability measure on Ω with respect to a countably finite, measurable and generating partition $\mathcal{A}$ which satisfies $K_w(\mathcal{A}) < \infty$ for some $w > 4$. Assume that β decays at least polynomially with power $> 6 + \frac{20}{w-4}$.*

If $\sigma > 0$ then

$$\mathbb{P}\left(\frac{I_n - nh}{\sigma\sqrt{n}} \leq t\right) = N(t) + \mathcal{O}(n^{-\kappa})$$

for all t and all

(i) *$\kappa < \frac{1}{10} - \frac{3}{5}\frac{w}{(p+2)(w-2)+6}$ if β decays polynomially with power p,*

(ii) *$\kappa < \frac{1}{10}$ if β decays super polynomially.*

The limiting result follows immediately from Theorem 2 and the rate of convergence was obtained in [12] using Stein's method. There is an incomplete argument in the variance and higher order estimates in [12] which are presented here in complete form (and also because we need higher than fourth moment) and in fact here we obtain better lower bounds on the power p than claimed in [12]. As a consequence, the error term for the variance has power $\frac{1}{4}$ or better. All other estimates remain unchanged. Another consequence of Theorem 2 is the Law of the Iterated Logarithm:

Corollary 5 *Under the assumptions of Theorem 4:*

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

almost everywhere.

And similarly for the lim inf where the limit then equals -1 almost everywhere. Also in [12] we had proven the weak invariance principle WIP which then required to prove tightness and independence. It now follows directly from Theorem 2 (although under slightly stronger assumptions).

2.3 Examples

(I) Bernoulli shift: For the Bernoulli measure μ over the full shift space $\Sigma = \mathbb{N}^{\mathbb{Z}}$ over the infinite alphabet $\mathbb{N}$ generated by the weights $p_1, p_2, \ldots$ ($\sum_j p_j = 1$), the entropy is then $h(\mu) = \sum_j p_j |\log p_j|$ and the variance is

$$\sigma^2 = \frac{1}{2} \sum_{i,j} p_i p_j \log^2 \frac{p_i}{p_j}$$

assuming that $\sum_i p_i \log^2 p_i < \infty$. If moreover $\sum_i p_i |\log p_i|^{5+\epsilon} < \infty$ for some $\epsilon > 0$ then we conclude the ASIP for $I_n(x) = -\log\mu(A_n(x))$ as β decays exponentially fast.

(II) Markov shift: If μ is the Markov measure on $\Sigma = \mathbb{N}^{\mathbb{Z}}$ generated by an infinite probability vector $\mathbf{p} = (p_1, p_2, \ldots)$ ($p_j > 0$, $\sum_j p_j = 1$) and an infinite stochastic matrix P ($\mathbf{p}P = \mathbf{p}$, $P\mathbf{1} = \mathbf{1}$) then the entropy is $h(\mu) = \sum_{i,j} -p_i P_{ij} \log P_{ij}$ [29] and the variance [12,26,31] is

$$\sigma^2 = \frac{1}{2}\sum_{ijk\ell} p_i P_{ij} p_k P_{k\ell} \log^2 \frac{P_{ij}}{P_{k\ell}} + 4\sum_{k=2}^{\infty}\sum_{\mathbf{x}\in\mathcal{A}^k} \mu(\mathbf{x})\left(\log P_{x_1 x_2} \log P_{x_{k-1} x_k} - h^2\right)$$

where the terms in brackets on the RHS decay exponentially fast. Then if $\sum_{i,j} p_i P_{i,j} |\log p_i P_{i,j}|^{5+\epsilon} < \infty$ for some $\epsilon > 0$ then I_n satisfies the ASIP as β decays exponentially fast ($p = \infty$). Naturally we get the ASIP for any Markov measure over a finite alphabet.

2.4 Recurrence time

Let $\Omega \subset \mathcal{A}^{\mathbb{Z}}$ be a symbolic system with the shift map T. We shall assume that $\mathcal{A}$ is a finite alphabet. We denote by

$$R_n(x) = \min\{k \geq 1 : T^{-k}x \in A_n(x)\}$$

the nth *recurrence time* of x. For a symbolic system where T is the shift map on a symbolic space $\Sigma \subset \mathcal{A}^{\mathbb{Z}}$ the recurrence time $R_n(\mathbf{x}) = \min\{k \geq 1 : x_{-k}\cdots x_{-k+n} = x_0\cdots x_n\}$

of the point $\mathbf{x} = (\dots, x_0, x_1, \dots)$ is the time is takes to see the first word of length n again as one goes to the 'left' (in this way we stay consistent with the notation used in [16] and some other places). Ornstein and Weiss [20,21] showed that for ergodic measures $\lim_{n \to \infty} \frac{1}{n} |\log R_n(x)| = h(\mu)$ almost everywhere improving on [30] where the convergence was shown to be in measure. Collet, Galves and Schmitt [6] proved the central limit theorem for Gibbs measures which are exponentially ψ-mixing. For finite alphabet processes Kontoyiannis [16] then proved the ASIP (for $\delta < \frac{1}{294}$) under the assumption that μ be α-mixing with α decaying at least with power 336 and that $\| f - f_n \|_{\mathscr{L}^1}$ decays with power 48. Here we obtain the following result which frees us from any condition on the sequence $\{ f_n : n \}$.

Theorem 6 *Assume $\mathcal{A}$ is a finite alphabet, μ is β-mixing where β decays at least least polynomially with power $p > 7$. If $\sigma > 0$ then*

$$R_n = \sigma B(n) + \mathcal{O}(n^{\frac{1}{2} - \delta})$$

almost surely for all $\delta < \frac{1}{8}$.

As a result of the ASIP we have the CLT for the recurrence time:

Corollary 7 *If $\mathcal{A}$ is finite and μ is β-mixing with β decaying at least least polynomially with power $p > 7$, then*

$$\mathbb{P}\left(\frac{\log R_n - n h(\mu)}{\sigma \sqrt{n}} \geq t \right) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{s^2}{2}} \, ds$$

if $\sigma > 0$.

3 Variance and higher moments

Before we prove the existence of the variance and bound the higher moments of the centred information function we shall summarise some known results which will be needed along the way.

3.1 The information function

Denote by $A_n(x)$ the atom in $\mathcal{A}^n$ ($n = 1, 2, \dots$) which contains the point $x \in \Omega$. It was then shown in [9] for ψ-mixing measures that $\sup_{A \in \mathcal{A}^n} \mu(A)$ decays exponentially fast as $n \to \infty$. For ϕ-mixing measures this was shown in [1] if $\phi(k)$ decays exponentially but is not necessarily true otherwise. In [12] we then showed that for a β-mixing measure μ one has:

(i) $\sup_{A \in \mathcal{A}^n} \mu(A) = \mathcal{O}(n^{-p})$ if β is polynomially decreasing with exponent $p > 0$;
(ii) $\sup_{A \in \mathcal{A}^n} \mu(A) = \mathcal{O}(\theta^{\sqrt{n}})$ for some $\theta \in (0, 1)$ if β is exponentially decreasing.

The metric entropy h for the invariant measure μ is $h = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{A}^n)$, where $\mathcal{A}$ is a generating partition of Ω (cf. [18]), provided $H(\mathcal{A}) < \infty$. For $w \geq 1$ put $\eta_w(t) = t \log^w \frac{1}{t}$ ($\eta_w(0) = 0$) and define

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

for partitions $\mathcal{B}$. Similarly one has the conditional quantity ($\mathcal{C}$ is a partition):

$$K_w(\mathcal{C}|\mathcal{B}) = \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B) \eta_w \left(\frac{\mu(B \cap C)}{\mu(B)} \right) = \sum_{B,C} \mu(B \cap C) \left| \log \frac{\mu(B \cap C)}{\mu(B)} \right|^w.$$

In the following we always assume that if $K_w(\mathcal{A}) < \infty$ then we also have $\mu(A) \leq e^{-w} \; \forall \, A \in \mathcal{A}$. This can be achieved by passing to a higher join. The assumption is convenient as it allows to use convexity arguments which are implicit in some of the properties and estimates we use. We will need the following result.

Lemma 8 *[13] For any two partitions $\mathcal{B}, \mathcal{C}$ for which $K_w(\mathcal{B}), K_w(\mathcal{C}) < \infty$*

 (i) $K_w(\mathcal{C}|\mathcal{B}) \leq K_w(\mathcal{C})$,
 (ii) $K_w(\mathcal{B} \vee \mathcal{C})^{1/w} \leq K_w(\mathcal{C}|\mathcal{B})^{1/w} + K_w(\mathcal{B})^{1/w}$,
 (iii) $K_w(\mathcal{B} \vee \mathcal{C})^{1/w} \leq K_w(\mathcal{C})^{1/w} + K_w(\mathcal{B})^{1/w}$.

In [12] it was shown that as a consequence

$$K_w(\mathcal{A}^n) \leq C_2 n^w. \tag{1}$$

The variance of the information function I_n is given by $\sigma_n^2 = \sigma^2(I_n) = K_2(\mathcal{A}^n) - H_n^2$ where $H_n = \mathbb{E}(I_n) = H(\mathcal{A}^n)$. For a partition $\mathcal{B}$ we write $J_\mathcal{B}$ for the centered information function given by $J_\mathcal{B}(B) = -\log \mu(B) - H(\mathcal{B})$, $B \in \mathcal{B}$ (i.e. $\int J_\mathcal{B} \, d\mu = 0$). Its variance is $\sigma^2(\mathcal{B}) = \sum_{B \in \mathcal{B}} \mu(B) J_\mathcal{B}(B)^2$. For two partitions $\mathcal{B}$ and $\mathcal{C}$ we put

$$J_{\mathcal{C}|\mathcal{B}}(B \cap C) = \log \frac{\mu(B)}{\mu(B \cap C)} - H(\mathcal{C}|\mathcal{B})$$

for $(B, C) \in \mathcal{B} \times \mathcal{C}$. That is $J_{\mathcal{C}|\mathcal{B}} = J_{\mathcal{B} \vee \mathcal{C}} - J_\mathcal{B}$ and $\sigma(\mathcal{C}|\mathcal{B}) = \sigma(J_{\mathcal{C}|\mathcal{B}})$. By [12]

$$\sigma(\mathcal{B} \vee \mathcal{C}) \leq \sigma(\mathcal{C}|\mathcal{B}) + \sigma(\mathcal{B}). \tag{2}$$

As a consequence of Lemma 8(i) one also has $K_w(\mathcal{B} \vee \mathcal{C}|\mathcal{B}) = K_w(\mathcal{C}|\mathcal{B}) \leq K_w(\mathcal{C})$ which in particular implies $\sigma(\mathcal{B} \vee \mathcal{C}|\mathcal{B}) = \sigma(\mathcal{C}|\mathcal{B}) \leq \sqrt{K_2(\mathcal{C})}$. Let us put $\rho(B, C) = \mu(B \cap C) - \mu(B)\mu(C)$. The following technical lemma is central to get the variance of μ and bounds on the higher moments of $J_n = I_n - H_n$.

Lemma 9 *Let μ be β-mixing and assume that $K_w(\mathcal{A}) < \infty$. Then for every $\gamma > 1$ and $a \in [1, w)$ there exists a constant C_1 so that*

$$\sum_{B \in \mathcal{A}^m, C \in T^{-\Delta-m} \mathcal{A}^n} \mu(B \cap C) \left| \log \left(1 + \frac{\rho(B, C)}{\mu(B)\mu(C)} \right) \right|^a$$

$$\leq C_1 \left(\beta(\Delta)(m + n)^{(1+a)\gamma} + (m + n)^{a\gamma - w(\gamma-1)} \right)$$

for $\Delta < \min(n, m)$ and for all $n = 1, 2, \ldots$.

We also have the following estimates for the approximations of $H(\mathcal{A}^n)$.

Lemma 10 *Under the assumptions of Lemma 9 the following applies:*

(I) For every $\gamma > 1$ there exists a constant C_2 so that for all n:

$$\left| H(\mathcal{A}^n \vee T^{-\Delta-n} \mathcal{A}^n) - 2H(\mathcal{A}^n) \right| \leq C_2 \left(\beta(\Delta) n^{2\gamma} + n^{\gamma - (\gamma-1)w} \right). \tag{3}$$

(II) There exists a constant C_3 so that

$$\left| \frac{1}{m} H(\mathcal{A}^m) - h \right| \leq C_3 \frac{1}{m^\zeta} \tag{4}$$

for all m, where

(i) $\zeta \in (0, 1 - \frac{2w}{p(w-1)})$ if β decays polynomially with power $p > \frac{2w}{w-1}$,
(ii) $\zeta \in (0, 1)$ if β decays faster than polynomially.

3.2 The variance

In this section we prove the existence of the variance as given in Theorem 2 and moreover obtain convergence rates.

Proposition 11 *Let μ be β-mixing and assume that $K_w(\mathcal{A}) < \infty$ for some $w > 2$. Let $\alpha \in (0, \frac{1}{2})$ and assume that β is at least polynomially decaying with power $p \geq \frac{3\frac{w}{\alpha} - w - 1}{w-2}$. Then the limit*

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

exists and is finite. Moreover there exists a constant C_4 so that

$$\left| \sigma^2 - \frac{\sigma^2(\mathcal{A}^n)}{n} \right| \leq \frac{C_4}{n^{\frac{1}{2} - \alpha}}.$$

If the partition $\mathcal{A}$ is infinite, then σ is strictly positive.

Proof Let us put $\mathcal{B} = \mathcal{A}^n$, $\mathcal{C} = T^{-n-\Delta}\mathcal{A}^n$. The gap Δ will be chosen to be $[n^\alpha]$ for some $\alpha \in (0, \frac{1}{2})$. We also assume here that β decays polynomially with power p, that is $\beta(\Delta) = \mathcal{O}(\Delta^{-p})$. Then by (3)

$$H(\mathcal{B} \vee \mathcal{C}) = 2H(\mathcal{A}^n) + \mathcal{O}\left(\beta(\Delta)n^{2\gamma} + n^{\gamma - (\gamma - 1)w} \right)$$

$$= 2H(\mathcal{A}^n) + \mathcal{O}\left(n^{-\alpha p + 2\gamma} + n^{\gamma - (\gamma - 1)w} \right)$$

where $\gamma > 1$ is arbitrary. The optimal value for γ is $\gamma = \frac{\alpha p + w}{w + 1}$ ($\alpha p > 1$) which yields the exponent $\theta_1 = \frac{-\alpha p(w-1) + 2w}{w + 1}$. That is $H(\mathcal{B} \vee \mathcal{C}) = 2H(\mathcal{A}^n) + \mathcal{O}(n^{\theta_1})$. We get for the variance

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

$$= \sum_{B,C} \mu(B \cap C) \left(J_\mathcal{B}(B) + J_\mathcal{C}(C) + \mathcal{O}(n^{\theta_1}) - \log\left(1 + \frac{\rho(B,C)}{\mu(B)\mu(C)} \right) \right)^2.$$

By Minkowski's inequality:

$$\left| \sigma(\mathcal{B} \vee \mathcal{C}) - \sqrt{N(\mathcal{B},\mathcal{C})} \right| \leq c_1 n^{\theta_1} + \sqrt{F(\mathcal{B},\mathcal{C})}$$

($c_1 > 0$) where (by Lemma 9 with $a = 2$)

$$F(\mathcal{B},\mathcal{C}) = \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \log^2\left(1 + \frac{\rho(B,C)}{\mu(B)\mu(C)} \right) \leq c_2 \left(\beta(\Delta)n^{3\gamma} + n^{2\gamma - (\gamma - 1)w} \right),$$

for any $\gamma > 1$ which when optimised yields the value $\gamma = \frac{\alpha p + w}{w+1}$. Then with $\theta_2 = \frac{-\alpha p(w-2)+3w}{w+1}$ we get $F(\mathcal{B}, \mathcal{C}) = \mathcal{O}(n^{\theta_2})$. The principal term is

$$
\begin{aligned}
N(\mathcal{B}, \mathcal{C}) &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C)\,(J_\mathcal{B}(B) + J_\mathcal{C}(C))^2 \\
&= \sum_{B,C} \mu(B \cap C)\,\big(J_\mathcal{B}(B)^2 + J_\mathcal{C}(C)^2\big) + 2R(\mathcal{B}, \mathcal{C}) \\
&= \sigma^2(\mathcal{B}) + \sigma^2(\mathcal{C}) + 2R(\mathcal{B}, \mathcal{C}).
\end{aligned}
$$

Since $J_\mathcal{B}$ and $J_\mathcal{C}$ have average zero the remainder term is

$$
\begin{aligned}
R(\mathcal{B}, \mathcal{C}) &= \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) J_\mathcal{B}(B) J_\mathcal{C}(C) \\
&= \sum_{B,C} (\mu(B)\mu(C) + \rho(B, C)) J_\mathcal{B}(B) J_\mathcal{C}(C) \\
&= \sum_{B,C} \rho(B, C) J_\mathcal{B}(B) J_\mathcal{C}(C).
\end{aligned}
$$

In order to estimate R, put $\mathcal{L} = \{(B, C) \in \mathcal{B} \times \mathcal{C} : \mu(B \cap C) \geq 2\mu(B)\mu(C)\}$ and write $R = R^+ + R^-$ where

$$
R^+(\mathcal{B}, \mathcal{C}) = \sum_{(B,C) \in \mathcal{L}} \rho(B, C) J_\mathcal{B}(B) J_\mathcal{C}(C),
$$

$$
R^-(\mathcal{B}, \mathcal{C}) = \sum_{(B,C) \in \mathcal{L}^c} \rho(B, C) J_\mathcal{B}(B) J_\mathcal{C}(C).
$$

For $(B, C) \in \mathcal{L}$ one has $\rho(B, C) = \mu(B \cap C) - \mu(B)\mu(C) \geq \frac{1}{2}\mu(B \cap C)$ and therefore, using Hölder's inequality twice $\left(\frac{1}{r} + \frac{1}{s} = 1 \text{ and } \frac{1}{u} + \frac{1}{v} = 1 \text{ such that } su, sv \leq w\right)$,

$$
\begin{aligned}
|R^+(\mathcal{B}, \mathcal{C})| &\leq \sum_{(B,C) \in \mathcal{L}} \frac{\rho(B, C)}{\mu(B \cap C)} \big|J_\mathcal{B}(B) J_\mathcal{C}(C)\big| \mu(B \cap C) \\
&\leq \left(\sum_{(B,C) \in \mathcal{L}} \left(\frac{\rho(B, C)}{\mu(B \cap C)}\right)^r \mu(B \cap C)\right)^{\frac{1}{r}} \left(\sum_{(B,C) \in \mathcal{L}} \big|J_\mathcal{B}(B) J_\mathcal{C}(C)\big|^s \mu(B \cap C)\right)^{\frac{1}{s}} \\
&\leq 2^{\frac{1}{r}} \left(\sum_{B,C} |\rho(B, C)|\right)^{\frac{1}{r}} \left(\left(\sum_B |J_\mathcal{B}(B)|^{su} \mu(B)\right)^{\frac{1}{u}} \left(\sum_C |J_\mathcal{C}(C)|^{sv} \mu(C)\right)^{\frac{1}{v}}\right)^{\frac{1}{s}} \\
&\leq c_3 \beta(\Delta)^{\frac{1}{r}} n^2
\end{aligned}
$$

since $\frac{\rho(B,C)}{\mu(B \cap C)} \leq 1 \, \forall\, (B, C) \in \mathcal{L}$, the β-mixing property $\sum_{B,C} |\rho(B, C)| \leq \beta(\Delta)$ and where we used the a priori estimates $\sum_B |J_\mathcal{B}(B)|^{su} \mu(B) = \mathcal{O}(n^{us})$ (similarly for the sum over C). We proceed similarly for the second part of the error term using the a priori estimate

$\sigma^2(\mathcal{A}^n) \le c_4 n^2$:

$$|R^-(\mathcal{B}, \mathcal{C})| \le \sum_{(B,C)\in\mathcal{L}^c} \frac{|\rho(B,C)|}{\mu(B)\mu(C)} \left| J_{\mathcal{B}}(B) J_{\mathcal{C}}(C) \right| \mu(B)\mu(C)$$

$$\le \left(\sum_{(B,C)\in\mathcal{L}^c} \left(\frac{|\rho(B,C)|}{\mu(B)\mu(C)} \right)^2 \mu(B)\mu(C) \right)^{\frac{1}{2}}$$

$$\times \left(\sum_{(B,C)\in\mathcal{L}^c} \left| J_{\mathcal{B}}(B) J_{\mathcal{C}}(C) \right|^2 \mu(B)\mu(C) \right)^{\frac{1}{2}}$$

$$\le \left(\sum_{B,C} |\rho(B,C)| \right)^{\frac{1}{2}} \sigma(\mathcal{B})\sigma(\mathcal{C})$$

$$\le c_4 \beta(\Delta)^{\frac{1}{r}} n^2,$$

where we used that $\mu(B \cap C) < 2\mu(B)\mu(C)$ implies $|\rho(B,C)| \le \mu(B)\mu(C)$. Hence with $s = \frac{w}{2}$ which is the largest possible value so that $\frac{1}{r} = \frac{w-2}{w}$ is the smallest possible and also $u = v = 2$ say, we obtain that $|R(\mathcal{B}, \mathcal{C})| = \mathcal{O}(n^{2-\alpha p \frac{w-2}{w}})$ and therefore

$$\sigma(\mathcal{B} \vee \mathcal{C}) \le \sqrt{\sigma^2(\mathcal{C}) + \sigma^2(\mathcal{B}) + \mathcal{O}(n^{2-\alpha p \frac{w-2}{w}})} + \mathcal{O}(n^{\theta_2/2}). \tag{5}$$

To fill the gap of length Δ estimates (2) and (1) yield

$$|\sigma(\mathcal{A}^{2n+\Delta}) - \sigma(\mathcal{B} \vee \mathcal{C})| \le \sigma(T^{-n}\mathcal{A}^{\Delta}|\mathcal{B}\vee\mathcal{C}) \le \sqrt{K_2(T^{-n}\mathcal{A}^{\Delta})} = \sqrt{K_2(\mathcal{A}^{\Delta})} \le c_6\Delta = \mathcal{O}(n^{\alpha}).$$

as $\Delta = [n^{\alpha}]$. We want to demand that $n^{1-\alpha p \frac{w-2}{2w}}, n^{\theta_2/2} = \mathcal{O}(n^{\alpha})$ which is achieved if $p \ge p_1 = \frac{2w}{w-2}\frac{1-\alpha}{\alpha}$ and $p \ge p_2 = \frac{3\frac{w}{\alpha}-w-1}{w-2}$ respectively. Since $p_2 > p_1$ we get the assumption $p \ge p_2$. Then, as $\sigma(\mathcal{B}) = \sigma(\mathcal{C}) = \sigma_n = \sigma(\mathcal{A}^n)$, one has

$$\sigma_{2n+[n^{\alpha}]} = \sqrt{2\sigma_n^2 + \mathcal{O}(n^{2\alpha})} + \mathcal{O}(n^{\alpha}),$$

Since $\alpha < \frac{1}{2}$ one has $\sigma_k^2 \le c_{10}k$ for all k and some constant c_{10}.

In order to get the rate of convergence let n_0 be given put recursively $n_{j+1} = 2n_j + [n_j^{\alpha}]$ ($j = 0, 1, 2, \dots$). Then $2^j n_0 \le n_j \le 2^j n_0 \prod_{i=0}^{j-1}\left(1 + \frac{1}{2}n_i^{\alpha-1}\right)$ where the product is bounded by

$$1 \le \prod_{i=0}^{j-1}\left(1 + \frac{1}{2}n_i^{\alpha-1}\right) \le \prod_{i=0}^{j-1}\left(1 + \frac{1}{n_0^{1-\alpha}2^{(1-\alpha)i+1}}\right) \le \exp\frac{c_{10}}{n_0^{1-\alpha}}.$$

On the other hand as $\sigma_{n_{j+1}} = \sqrt{2\sigma_{n_j}^2 + \mathcal{O}(n_j^{2\alpha})} + \mathcal{O}(n_j^{\alpha})$, which yields

$$\sigma_{n_{j+1}}^2 = 2\sigma_{n_j}^2 + \mathcal{O}(n_j^{2\alpha}) + \mathcal{O}(n_j^{\alpha})\sqrt{2\sigma_{n_j}^2 + \mathcal{O}(n_j^{2\alpha})} = \left(\sqrt{2}\sigma_{n_j} + \mathcal{O}(n_j^{\alpha})\right)^2$$

and consequently

$$\sigma_{n_{j+1}} = \sqrt{2}\,\sigma_{n_j} + \mathcal{O}(n_j^{\alpha}).$$

Thus

$$\sigma_{n_j} = 2^{\frac{j}{2}}\sqrt{n_0} + \sum_{k=0}^{j-1} 2^{\frac{j-1-k}{2}} \mathcal{O}(n_k^{\alpha})$$

and since $\sqrt{n_j} = 2^{\frac{j}{2}}\sqrt{n_0}e^{\mathcal{O}(n_0^{\alpha-1})}$ we get (as $\alpha < \frac{1}{2}$)

$$\frac{\sigma_{n_j}}{\sqrt{n_j}} = \frac{\sigma_{n_0}}{\sqrt{n_0}}e^{\mathcal{O}(n_0^{\alpha-1})} + \sum_{k=0}^{j-1}\frac{2^{\frac{j-1-k}{2}}2^{\alpha k}n_0^{\alpha}\mathcal{O}(1)}{j^{\frac{1}{2}}\sqrt{n_0}} = \frac{\sigma_{n_0}}{\sqrt{n_0}}e^{\mathcal{O}\left(n_0^{\alpha-1}\right)} + \mathcal{O}\left(n_0^{\frac{1}{2}-\alpha}\right).$$

Hence

$$\frac{\sigma_{n_j}^2}{n_j} = \frac{\sigma_{n_0}^2}{n_0} + \mathcal{O}\left(\frac{1}{n_0^{\frac{1}{2}-\alpha}}\right).$$

Taking $\limsup$ as $j \to \infty$ and $n_0 \to \infty$ shows that the limit $\sigma^2 = \lim_n \frac{\sigma_n^2}{n}$ exists and satisfies moreover $\left|\sigma^2 - \frac{\sigma_n^2}{n}\right| \leq C_4 n^{-(\frac{1}{2}-\alpha)}$ for some C_4.

Positivity of σ in the case of an infinite partition was shown in [12]. □

For finite partitions the measure has variance zero if it is a Gibbs state for a potential which is a coboundary.

3.3 Higher order moments

We will need estimates on the higher moments of J_n which we denote by

$$M_\ell(\mathcal{B}) = \sum_{B \in \mathcal{B}} \mu(B)|J_{\mathcal{B}}(B)|^{\ell}.$$

the wth (absolute) moment of the function $J_{\mathcal{B}}$. By Minkowski's inequality (on $\mathscr{L}^{\ell}$ spaces)

$$M_\ell^{\frac{1}{\ell}}(\mathcal{B} \vee \mathcal{C}) = \sqrt[\ell]{\mu\left(\left|J_{\mathcal{C}|\mathcal{B}} + J_{\mathcal{B}}\right|^{\ell}\right)} \leq \sqrt[\ell]{\mu(|J_{\mathcal{C}|\mathcal{B}}|^{\ell})} + \sqrt[\ell]{\mu(|J_{\mathcal{B}}|^{\ell})} = M_\ell^{\frac{1}{\ell}}(\mathcal{C}|\mathcal{B}) + M_\ell^{\frac{1}{\ell}}(\mathcal{B}),$$

where $M_\ell(\mathcal{C}|\mathcal{B}) = \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C)|J_{\mathcal{C}|\mathcal{B}}(B \cap C)|^{\ell}$ are the conditional moments. It follows from Corollary 1 that the absolute moments for the joins $\mathcal{A}^n$ have the rough a priori estimate $M_\ell(\mathcal{A}^n) \leq K_\ell(\mathcal{A}^n) \leq C_2 n^{\ell}$. The purpose of the next proposition is to reduce the power from ℓ to $\frac{\ell}{2}$.

Proposition 12 *Let μ be β-mixing and assume that $K_w(\mathcal{A}) < \infty$ for some $w > 4$. Also assume that β decays at least polynomially with power p.*

Let ℓ be an integer strictly smaller than w, then there exists a constant C_5 so that for all $q \leq \ell$

$$M_q(\mathcal{A}^n) \leq C_5 n^{\frac{q}{2}}$$

provided $p \geq \frac{w(\ell+2)-\ell}{w-\ell}$.

Proof The statement is true for $\ell = 2$ by Proposition 11. We prove the result by induction for integers $\ell < [w]$. Assume the estimate is true for $k \leq \ell - 1$ and we will now prove it for ℓ.

With $\mathcal{B} = \mathcal{A}^n$, $\mathcal{C} = T^{-\Delta-n}\mathcal{A}^n$ and the gap $\Delta = [\sqrt{n}]$ ($\alpha = \frac{1}{2}$ in the previous notation) we get by (3) $H(\mathcal{B} \vee \mathcal{C}) = H(\mathcal{B}) + H(\mathcal{C}) + \mathcal{O}(\beta(\Delta)n^{2\gamma} + n^{\gamma-(\gamma-1)w})$ for $\gamma > 1$ which as in the proof of Proposition 11 optimised gives $H(\mathcal{B} \vee \mathcal{C}) = H(\mathcal{B}) + H(\mathcal{C}) + \mathcal{O}(n^{\theta_1})$ ($\theta_1 = \frac{-p(w-1)+4w}{2(w+1)}$). With Minkowsky's inequality

$$
M_\ell^{\frac{1}{\ell}}(\mathcal{B} \vee \mathcal{C}) = \left(\sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left| \log \frac{1}{\mu(B \cap C)} - H(\mathcal{B} \vee \mathcal{C}) \right|^\ell \right)^{\frac{1}{\ell}}
$$
$$
= \hat{M}_\ell^{\frac{1}{\ell}}(\mathcal{B}, \mathcal{C}) + \mathcal{O}(n^{\theta_1}) + F_\ell^{\frac{1}{\ell}}(\mathcal{B}, \mathcal{C}),
$$

where

$$
\hat{M}_\ell(\mathcal{B}, \mathcal{C}) = \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \, |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^\ell
$$

and by Lemma 9 (with $a = \ell$) we get the error estimate

$$
F_\ell(\mathcal{B}, \mathcal{C}) = \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) \left| \log \left(1 + \frac{\rho(B, C)}{\mu(B)\mu(C)} \right) \right|^\ell = \mathcal{O}\left(\beta(\Delta)n^{(\ell+1)\gamma} + n^{\ell\gamma-(\gamma-1)w} \right)
$$

where the value of $\gamma > 1$ is optimised when $\gamma = \frac{p+2w}{2(w+1)}$ which then yields $F_\ell(\mathcal{B}, \mathcal{C}) = \mathcal{O}(n^{\theta_\ell})$ where $\theta_\ell = \frac{-p(w-\ell)+2(\ell+1)w}{2(w+1)}$. We want to achieve that $n^{\theta_1} = \mathcal{O}(n^{\frac{1}{2}})$ and, more generally, $n^{\theta_\ell} = \mathcal{O}(n^{\frac{\ell}{2}})$. That is $\theta_\ell \leq \frac{\ell}{2}$ and this is satisfied if (as by assumption) $p \geq p_1 = \frac{w(\ell+2)-\ell}{w-\ell}$. Hence $F(\mathcal{B}, \mathcal{C}) = \mathcal{O}(n^{\frac{\ell}{2}})$ and moreover

$$
M_\ell(\tfrac{1}{\ell}\mathcal{B} \vee \mathcal{C}) = \hat{M}_\ell^{\frac{1}{\ell}}(\mathcal{B}, \mathcal{C}) + \mathcal{O}(n^{\frac{1}{2}}). \tag{6}
$$

We further approximate

$$
\hat{M}_\ell(\mathcal{B}, \mathcal{C}) = M_\ell^\times(\mathcal{B}, \mathcal{C}) + R_\ell(\mathcal{B}, \mathcal{C})
$$

where

$$
M_\ell^\times(\mathcal{B}, \mathcal{C}) = \sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B)\mu(C) \, |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^\ell
$$

is the principal term and the error term $R_\ell = R_\ell^+ + R_\ell^-$ is split into a sum of two terms as in the proof of Proposition 11. Put $\mathcal{L} = \{(B, C) \in \mathcal{B} \times \mathcal{C} : \mu(B \cap C) \geq 2\mu(B)\mu(C)\}$ and then

$$
R_\ell^+(\mathcal{B}, \mathcal{C}) = \sum_{(B,C) \in \mathcal{L}} \rho(B, C) \, |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^\ell,
$$
$$
R_\ell^-(\mathcal{B}, \mathcal{C}) = \sum_{(B,C) \in \mathcal{L}^c} |\rho(B, C)| \, |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^\ell.
$$

We now estimate the two terms separately as follows:

(I) For $(B, C) \in \mathcal{L}$ one has $\rho(B, C) = \mu(B \cap C) - \mu(B)\mu(C) \geq \frac{1}{2}\mu(B \cap C)$ and therefore, by Hölder's inequality ($\frac{1}{s} + \frac{1}{r} = 1$),

$$
\begin{aligned}
R_\ell^+(\mathcal{B}, \mathcal{C}) &= \sum_{(B,C) \in \mathcal{L}} \frac{\rho(B, C)}{\mu(B \cap C)} |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^\ell \, \mu(B \cap C) \\
&\leq \left(\sum_{(B,C) \in \mathcal{L}} \left(\frac{\rho(B, C)}{\mu(B \cap C)} \right)^r \mu(B \cap C) \right)^{\frac{1}{r}} \\
&\quad \times \left(\sum_{(B,C) \in \mathcal{L}} |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^{s\ell} \, \mu(B \cap C) \right)^{\frac{1}{s}} \\
&\leq 2^{\frac{1}{r}} \left(\sum_{B,C} |\rho(B, C)| \right)^{\frac{1}{r}} \hat{M}_{s\ell}(\mathcal{B}, \mathcal{C})^{\frac{1}{s}} \\
&\leq 2^{\frac{1}{r}} \beta(\Delta)^{\frac{1}{r}} \hat{M}_{s\ell}(\mathcal{B}, \mathcal{C})^{\frac{1}{s}}
\end{aligned}
$$

$(s\ell \leq w)$ since $\frac{\rho(B,C)}{\mu(B \cap C)} \leq 1 \, \forall (B, C) \in \mathcal{L}$ and where we used the β-mixing property $\sum_{B,C} |\rho(B, C)| \leq \beta(\Delta)$. We now use the a priori estimate $M_q(\mathcal{A}^n) \leq c_2 n^q$ to bound $\hat{M}_{s\ell}(\mathcal{B}, \mathcal{C})$. Using Minkowsky's inequality we get the rough a priori bound for $q \leq w$:

$$
\begin{aligned}
\hat{M}_q^{\frac{1}{q}}(\mathcal{B}, \mathcal{C}) &= \left(\sum_{B \in \mathcal{B}, C \in \mathcal{C}} \mu(B \cap C) |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^q \right)^{\frac{1}{q}} \\
&\leq \left(\sum_{B \in \mathcal{B}} \mu(B) |J_\mathcal{B}(B)|^q \right)^{\frac{1}{q}} + \left(\sum_{C \in \mathcal{C}} \mu(C) |J_\mathcal{C}(C)|^q \right)^{\frac{1}{q}} \\
&= 2 M_q^{\frac{1}{q}}(\mathcal{A}^n) \\
&\leq 2c_2 n
\end{aligned}
$$

and consequently $R_\ell^+(\mathcal{B}, \mathcal{C}) = \mathcal{O}(\beta(\Delta)^{\frac{1}{r}} n^\ell)$.

(II) We proceed similarly for the second part of the error term ($\frac{1}{s} + \frac{1}{r} = 1$):

$$
\begin{aligned}
|R_\ell^-(\mathcal{B}, \mathcal{C})| &\leq \sum_{(B,C) \in \mathcal{L}^c} \frac{|\rho(B, C)|}{\mu(B)\mu(C)} |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^\ell \, \mu(B)\mu(C) \\
&\leq \left(\sum_{(B,C) \in \mathcal{L}^c} \left(\frac{|\rho(B, C)|}{\mu(B)\mu(C)} \right)^r \mu(B)\mu(C) \right)^{\frac{1}{r}} \\
&\quad \times \left(\sum_{(B,C) \in \mathcal{L}^c} |J_\mathcal{B}(B) + J_\mathcal{C}(C)|^{s\ell} \, \mu(B)\mu(C) \right)^{\frac{1}{s}}
\end{aligned}
$$

$$\leq \left(\sum_{B,C} |\rho(B,C)| \right)^{\frac{1}{r}} M_{s\ell}^{\times}(B,C)^{\frac{1}{s}}$$

$$\leq \beta(\Delta)^{\frac{1}{r}} M_{s\ell}^{\times}(B,C)^{\frac{1}{s}},$$

where we used that $\mu(B \cap C) < 2\mu(B)\mu(C)$ implies $|\rho(B,C)| \leq \mu(B)\mu(C)$. Using the a priori estimate $M_q(\mathcal{A}^n) \leq c_2 n^q$, we obtain by Minkowsky for $q \leq w$:

$$(M_q^{\times}(B,C))^{\frac{1}{q}} \leq \left(\sum_{B \in \mathcal{B}} \mu(B) |J_{\mathcal{B}}(B)|^q \right)^{\frac{1}{q}} + \left(\sum_{C \in \mathcal{C}} \mu(C) |J_{\mathcal{C}}(C)|^q \right)^{\frac{1}{q}}$$

$$= 2 M_q^{\frac{1}{q}}(\mathcal{A}^n)$$

$$\leq 2 c_2 n$$

and consequently $R_\ell^{-}(B,C) = \mathcal{O}(\beta(\Delta)^{\frac{1}{r}} n^\ell)$.

The two parts (I) and (II) combined yield

$$|R_\ell(\mathcal{B} \vee \mathcal{C})| \leq c_3 n^{\ell - \frac{p}{2r}}$$

where we choose s and r such that $s\ell \leq w$ and $\ell - \frac{p}{2r} \leq \frac{\ell}{2}$. For $s = \frac{w}{\ell}$ (largest possible) and $r = \frac{w}{w-\ell}$ this requires $p \geq p_2 = \frac{w\ell}{w-\ell}$ which is satisfied by the assumption since $p_1 > p_2$. Then

$$M_\ell(\mathcal{B} \vee \mathcal{C}) = M_\ell^{\times}(\mathcal{B},\mathcal{C}) + \mathcal{O}(n^{\frac{\ell}{2}})$$

and now look more closely at the principal term $M_\ell^{\times}$. Using the induction hypothesis $M_k(\mathcal{A}^n) = \mathcal{O}(n^{\frac{k}{2}})$ for $k \leq \ell - 1$ we obtain for ℓ integer

$$M_\ell^{\times}(\mathcal{B},\mathcal{C}) \leq M_\ell(\mathcal{B}) + M_\ell(\mathcal{C}) + \sum_{k=1}^{\ell-1} \binom{\ell}{k} M_k(\mathcal{B}) M_{\ell-k}(\mathcal{C}) \leq 2 M_\ell(\mathcal{A}^n) + c_4 n^{\frac{\ell}{2}}$$

and consequently

$$M_\ell^{\frac{1}{\ell}}(\mathcal{B} \vee \mathcal{C}) = \left(2 M_\ell(\mathcal{A}^n) + \mathcal{O}(n^{\frac{\ell}{2}}) \right)^{\frac{1}{\ell}} + \mathcal{O}(n^{\frac{1}{2}}).$$

To fill in the gap of length Δ we use Lemma 8(iii), the estimate (1) on K_ℓ and the fact that $\Delta \sim \sqrt{n}$:

$$\left| M_\ell^{\frac{1}{\ell}}(\mathcal{A}^{2n+\Delta}) - M_\ell^{\frac{1}{\ell}}(\mathcal{B} \vee \mathcal{C}) \right| \leq M_\ell^{\frac{1}{\ell}}(\mathcal{A}^{2n+\Delta} | \mathcal{B} \vee \mathcal{C}) \leq K_\ell^{\frac{1}{\ell}}(\mathcal{A}^{\Delta}) = \mathcal{O}(\Delta) = \mathcal{O}(n^{\frac{1}{2}}).$$

which implies

$$M_\ell^{\frac{1}{\ell}}(\mathcal{A}^{2n+\Delta}) \leq \left(2 M_\ell(\mathcal{A}^n) + c_5 n^{\frac{\ell}{2}} \right)^{\frac{1}{\ell}} + c_6 n^{\frac{1}{2}}$$

for constants c_5, c_6. Given n_0, put recursively $n_{j+1} = 2n_j + [\sqrt{n_j}]$ ($\Delta_j = [\sqrt{n_j}]$ are the gaps), then for a constant c_7 large enough so that $(2 + c_5/c_7)^{\frac{1}{\ell}} + c_6/c_7^{\frac{1}{\ell}} \leq \sqrt{2}$ (which is possible as $\ell > 2$) we obtain

$$M_\ell(\mathcal{A}^{n_j}) \leq c_7 n_j^{\frac{\ell}{2}}$$

for all j. Increasing the constant c_7 allows us to extend the estimate to all n with a constant C_5. This completes the inductions step. If ℓ is the largest integer strictly smaller than w, then we can use Hölder's inequality to extend the estimate $M_q(\mathcal{A}^n) \leq c_7 n^{\frac{q}{2}}$ to arbitrary values of $q \leq \ell$. □

4 Proof of the ASIP (Theorem 2)

Let α denote a number between 0 and 1 and $\ell < w$ an integer. We decompose $J_n = I_n - H(\mathcal{A}^n) = \sum_{j=1}^{Q_n}(y_j + z_j)$ where $|z_j| = \Delta_j$ and $\Delta_j = [n_j^\alpha]$ is the length of the gaps where the length $|y_j| = n_j$ will be chosen to be $[\sqrt{j}]$. Then $n = \sum_{j=1}^{Q_n}(n_j + \Delta_j) + r_n$ with remainder $r_n < n_{Q_n+1} + \Delta_{Q_n+1}$. In particular $n \asymp \sum_{j=1}^{Q_n} \sqrt{j} \asymp Q_n^{\frac{3}{2}}$ implies that $Q_n \asymp n^{\frac{2}{3}}$. We put $\hat{\mathcal{A}}^n = \bigvee_{j=1}^{Q_n} T^{-N_j}\mathcal{A}^{n_j}$ where we put $N_j = \sum_{i=1}^{j-1}(n_i + \Delta_i)$ and $N_1 = 0$. We have $N_j \asymp \sum_{i=1}^{j-1} \sqrt{i} \asymp j^{\frac{3}{2}}$.

Lemma 13 *Let $\ell \geq 5$ and $p \geq \frac{2(w+2)-\ell}{w-\ell}$. Then*

$$J_n = \sum_{j=1}^{Q_n} J_{n_j} \circ T^{N_j} + \mathcal{O}(n^{1-\delta_1})$$

almost surely for any $\delta_1 \leq \min(\frac{p(w-2)-5w+1}{2p(w-2)+2w+2}, \frac{1-\alpha}{3})$.

Proof We proceed in three steps. First we cut 'gaps' of lengths Δ_j, then we use the β-mixing property to separate the long blocks of lengths n_j and in the last part we adjust the averaging term (entropy).

(I) Since $J_{\mathcal{A}^n|\hat{\mathcal{A}}^n} = J_{\mathcal{A}^n \vee \hat{\mathcal{A}}^n} - J_{\hat{\mathcal{A}}^n} = J_{\mathcal{A}^n} - J_{\hat{\mathcal{A}}^n}$ (as $\mathcal{A}^n$ refines $\hat{\mathcal{A}}^n$) we obtain

$$\|J_{\mathcal{A}^n} - J_{\hat{\mathcal{A}}^n}\|_a^a = \|J_{\mathcal{A}^n|\hat{\mathcal{A}}^n}\|_a^a = M_a\left(\mathcal{A}^n|\hat{\mathcal{A}}^n\right) \leq \sum_{j=1}^{Q_n} M_a(\mathcal{A}^{\Delta_j})$$

and using Proposition 12 for $1 < a \leq \ell$ (as $p \geq \frac{2(\ell+2)-\ell}{w-\ell}$ by assumption)

$$\|J_{\mathcal{A}^n} - J_{\hat{\mathcal{A}}^n}\|_a^a \leq c_1 \sum_{j=1}^{Q_n} n_j^{\frac{a\alpha}{2}} \leq c_1 \sum_{j=1}^{Q_n} j^{\frac{a\alpha}{4}} \leq c_2 Q_n^{\frac{a\alpha}{4}+1} = \mathcal{O}\left(n^{\frac{a\alpha}{6}+\frac{2}{3}}\right).$$

By Tchebycheff

$$\mathbb{P}(|J_{\mathcal{A}^n} - J_{\hat{\mathcal{A}}^n}| > \epsilon_n) \leq \frac{\|J_{\mathcal{A}^n|\hat{\mathcal{A}}^n}\|_a^a}{\epsilon_n^a} \leq c_3 \frac{n^{\frac{a\alpha}{6}+\frac{2}{3}}}{n^{a(1-\delta)}}$$

where we put $\epsilon_n = n^{1-\delta}$. This is summable if $a(1-\delta) - \frac{a\alpha}{6} - \frac{2}{3} > 1$ which is satisfied we we choose $a = \ell$ and as $\ell \geq 5$ this is satisfied for any $\alpha < 1$ and $\delta < \frac{1}{3}$. Thus by Borel–Cantelli

$$J_{\mathcal{A}^n} = J_{\hat{\mathcal{A}}^n} + \mathcal{O}(n^{1-\delta})$$

almost surely for any $\delta \leq \frac{1}{3}$.

(II) Now let us put $D_k = \bigvee_{j=1}^{k-1} T^{-N_j} \mathcal{A}^{n_j}$, i.e. recursively we have $D_{k+1} = D_k \vee T^{-N_k} \mathcal{A}^{n_k}$ and also $D_{Q_n+1} = \hat{\mathcal{A}}^n$. Then by Lemma 9 (with identification $\Delta = \Delta_{k-1}, n = N_\ell - \Delta_{k-1}, m = n_k, n + m \le N_{k+1}$ and not necessarily the same number a as in part (I))

$$\|I_{D_{k+1}} - I_{D_k} - I_{n_k} \circ T^{N_k}\|_a^a$$

$$= \sum_{D \in D_k, A \in T^{-N_k} \mathcal{A}^{n_k}} \mu(D \cap A) \left| \log \left(\frac{1}{\mu(D \cap A)} - \log \frac{1}{\mu(A)} - \log \frac{1}{\mu(D)} \right) \right|^a$$

$$= \mathcal{O}\left(\beta(\Delta_{k-1}) N_{k+1}^{(1+\alpha)\gamma} + N_{k+1}^{\alpha\gamma - w(\gamma-1)} \right)$$

for any $\gamma > 1$ and $1 < a < w$. As $\beta(\Delta) = \mathcal{O}(\Delta^{-p})$

$$\|I_{D_{k+1}} - I_{D_k} - I_{n_k} \circ T^{N_k}\|_a^a = \mathcal{O}\left(n_{k-1}^{-\alpha p} N_k^{(1+a)\gamma} + N_k^{a\gamma - w(\gamma-1)} \right)$$

$$= \mathcal{O}\left(k^{-\frac{\alpha p}{2}} k^{\frac{3}{2}\gamma(1+a)} + k^{\frac{3}{2}(a\gamma - w(\gamma-1))} \right)$$

which implies by Minkowski that

$$\left\| I_{\hat{\mathcal{A}}^n} - \sum_{j=1}^{Q_n} I_{n_j} \circ T^{N_j} \right\|_a = \mathcal{O}(1) \sum_{k=1}^{Q_n} \left(k^{(-\frac{\alpha p}{2} + \frac{3}{2}\gamma(1+a))\frac{1}{a}} + k^{(\frac{3}{2}(a\gamma - w(\gamma-1))\frac{1}{a}} \right)$$

$$= \mathcal{O}\left(n^{(\gamma(1+a) - \frac{\alpha p}{3} + \frac{2}{3})\frac{1}{a}} + n^{((a\gamma - w(\gamma-1) + \frac{2}{3})\frac{1}{a}} \right)$$

Thus

$$\mathbb{P}\left(\left| I_{\hat{\mathcal{A}}^n} - \sum_{j=1}^{Q_n} I_{n_j} \circ T^{N_j} \right| > \epsilon_n \right) \le \frac{1}{\epsilon_n^a} \left\| I_{\hat{\mathcal{A}}^n} - \sum_{j=1}^{Q_n} I_{n_j} \circ T^{N_j} \right\|_a^a$$

is summable over $n \in \mathbb{N}$ if $(\gamma > 1, 1 < a < w)$

$$\begin{cases} \gamma(1+a) - \frac{1}{3}\alpha p + \frac{2}{3} + \delta < 0 \\ a\gamma - w(\gamma-1) + \frac{2}{3} + \delta < 0 \end{cases}$$

where we chose $\epsilon_n = n^{1-\delta}$. The conditions on γ and a are satisfied for any $\delta \le \frac{1}{3}$ since $w > 5$ and $p > 4$. Hence by Borel–Cantelli

$$\left| I_{\hat{\mathcal{A}}^n} - \sum_{j=1}^{Q_n} I_{n_j} \circ T^{N_j} \right| \le n^{1-\delta}$$

for all n large enough almost everywhere.

(III) The entropies are estimated using Lemma 4 as follows:

$$\left| H(\mathcal{A}^n) - \sum_{j=1}^{Q_n} H(\mathcal{A}^{n_j}) \right| \le \sum_{j=1}^{Q_n} H(\mathcal{A}^{\Delta_j}) \le c_4 Q_n^{\frac{\alpha}{2}+1} \le c_5 n^{\frac{\alpha}{3} + \frac{2}{3}}$$

as $H(\mathcal{A}^{\Delta_j}) = \mathcal{O}(\Delta_j) = \mathcal{O}(n_j^\alpha) = \mathcal{O}(j^{\frac{\alpha}{2}})$. The RHS is bounded by $n^{1-\delta_1}$ as $\delta_1 \le \frac{1-\alpha}{3}$. $\square$

With $y_j = J_{n_j} \circ T^{N_j}$ we get $\mathbb{E}|y_j|^\ell \le n_j^{\frac{\ell}{2}} \le j^{\frac{\ell}{4}}$ by Proposition 12. By Proposition 11 then

$$\sigma^2(y_j) = \mathbb{E}(y_j^2) = n_j \sigma^2 + \mathcal{O}(n_j^{1-\eta})$$

where $\eta = \frac{p(w-2)-5w+1}{2p(w-2)+2w+2}$. This then implies (as $n_j \sim \sqrt{j}$)

$$\sum_{j=1}^{Q_n} \mathbb{E}(y_j^2) = \sigma^2 \sum_j n_j + \sum_{j=1}^{Q_n} \mathcal{O}(n_j^{1-\eta})$$

$$= \sigma^2 n + \mathcal{O}\left(\sum_{j=1}^{Q_n} n_j^\alpha\right) + \sum_{j=1}^{Q_n} \mathcal{O}\left(j^{\frac{1-\eta}{2}}\right)$$

$$= n\sigma^2 + \mathcal{O}(n^{\frac{\alpha+2}{3}} + n^{1-\frac{\eta}{3}}) \tag{7}$$

for $\delta \leq \min\left(\frac{p(w-2)-5w+1}{2p(w-2)+2w+2}, \frac{1-\alpha}{3}\right)$ as $n - \sum_j n_j = \sum_{j=1}^{Q_n} \mathcal{O}(j^{\frac{\alpha}{2}}) = \mathcal{O}(Q_n^{\frac{\alpha}{2}+1}) = \mathcal{O}(n^{\frac{\alpha+2}{3}})$.

Lemma 14 *If $\ell \geq 5$ and $p \geq \frac{w(\ell+2)-\ell}{w-\ell}$. Then*

$$\sum_{j=1}^{Q_n} y_j^2 = n\sigma^2 + \mathcal{O}(n^{1-\delta_2})$$

almost surely, for any $\delta_2 < \min(\frac{1}{3}, (1-\frac{4}{\ell})\frac{\alpha p}{6})$.

Proof To use Gal-Kuksma's estimate as given in Lemma A1 of [25] directly we put

$$x_j = y_j^2 - \mathbb{E}(y_j^2)$$

and then obtain for $1 \leq m < m' \leq Q_n$:

$$\mathbb{E}\left(\sum_{j=m}^{m'} x_j\right)^2 = \mathbb{E}\left(\sum_{j=m}^{m'} y_j^2 - \mathbb{E}(y_j^2)\right)^2$$

$$= \sum_{j=m}^{m'} \left(\mathbb{E}(y_j^4) - (\mathbb{E}(y_j^2))^2\right) + \sum_{j \neq i} \left(\mathbb{E}(y_j^2 y_i^2) - \mathbb{E}(y_j^2)\mathbb{E}(y_i^2)\right)$$

$$= I + II.$$

For the second term, $i \neq j$, we use Lemma 7.2.1 from [25]: if $i < j$ then

$$\left|\mathbb{E}(y_j^2 y_i^2) - \mathbb{E}(y_j^2)\mathbb{E}(y_i^2)\right| \leq \beta(\Delta_j)^{\frac{1}{u}} \|y_i^2\|_r \|y_j^2\|_s$$

where $\frac{1}{u} + \frac{1}{r} + \frac{1}{s} = 1$. For the terms on the RHS we get by Proposition 12

$$\|y_i^2\|_s = \left(\mathbb{E}(y_i^{2s})\right)^{\frac{1}{s}} \leq c_1(n_i^s)^{\frac{1}{s}} \leq c_1\sqrt{i}$$

under the assumption that $2s, 2r \leq \ell$ (which requires $p \geq \frac{w(\ell+2)-\ell}{w-\ell}$) and obtain

$$\left|\mathbb{E}(y_i^2 y_j^2) - \mathbb{E}(y_i^2)\mathbb{E}(y_j^2)\right| \leq \beta(j^{\frac{\alpha}{2}})^{\frac{1}{u}} n_i n_j = \mathcal{O}(j^{-\frac{\alpha p}{2}}\sqrt{ij})$$

Thus for any $1 \le m < m' \le Q_n$ we get

$$II \le c_2 \sum_{m \le i < j \le m'} j^{-\frac{\alpha p}{2u}} \sqrt{ij}$$

$$\le c_2 \sum_{i=m}^{m'-1} \sqrt{i} \sum_{j=i+1}^{m'} j^{\frac{1}{2} - \frac{1}{2}\frac{\alpha p}{u}}$$

$$\le c_3 \sum_{i=m}^{m'-1} \sqrt{i} (m'^{\frac{3}{2} - \frac{1}{2}\frac{\alpha p}{u}} - i^{\frac{3}{2} - \frac{1}{2}\frac{\alpha p}{u}})$$

$$\le c_4 (m'^\zeta - m^\zeta),$$

where $\zeta = 3 - \frac{1}{2}\frac{\alpha p}{u}$. As $\mathbb{E}(y_j^4) = \mathcal{O}(n_j^2)$ we bound the first term (I) using Proposition 12

$$I \le \sum_{j=m}^{m'} \left| \mathbb{E}(y_j^4) - (\mathbb{E}(y_j^2))^2 \right| = \sum_{j=m}^{m'} \mathcal{O}\left(n_j^2 + n_j^2 \right) = \sum_{j=m}^{m'} \mathcal{O}(j) = \mathcal{O}(m'^2 - m^2).$$

With $\zeta' = \max(\zeta, 2)$ by [25] Lemma A.1 for any $\bar{\delta} > 0$ there exist a constant c_5 such that

$$\sum_{j=1}^{Q_n} x_j = \sum_{j=1}^{Q_n} \left(y_j^2 - \mathbb{E}(y_j^2) \right) \le c_5 Q_n^{\frac{\zeta'}{2}} \log^{2+\bar{\delta}} Q_n = \mathcal{O}(n^{1-\delta_2})$$

for any $\delta_2 < \min(\frac{1}{3}, (1 - \frac{4}{\ell})\frac{\alpha p}{6})$ almost surely where we have chosen $s = r = \frac{\ell}{2}$ which implies $\frac{1}{u} = 1 - \frac{1}{r} - \frac{1}{s} = 1 - \frac{4}{\ell}$ which is positive as $\ell \ge 5$ by assumption. Thus

$$\sum_{j=1}^{Q_n} y_j^2 = \sum_{j=1}^{Q_n} \mathbb{E}(y_j^2) + \mathcal{O}(n^{1-\delta_2}) = \sigma^2 n + \mathcal{O}(n^{1-\delta_2})$$

almost surely. $\qquad\qquad\square$

We now do the Martingale decomposition. Let $\mathscr{F}_j = \sigma(\mathcal{A}^{h_j})$, where $h_j = N_j + n_j$. Then $\mathscr{F}_j \subset \mathscr{F}_{j+1} \subset \mathscr{F}_{j+2} \subset \cdots$ and introduce Y_j by $y_j = Y_j + u_j - u_{j+1}$ where

$$u_j = \sum_{k=0}^{\infty} \mathbb{E}(y_{j+k}|\mathscr{F}_{j-1}).$$

Lemma 15 *There exists a constant C_6 such that*

$$\|u_j\|_q^q \le C_6 j^{\frac{1}{4} - \frac{\alpha p}{2}(1 - \frac{1}{q} - \frac{1}{\ell})}$$

for $q < \frac{1}{4}(3\alpha p(1 - \frac{1}{q} - \frac{1}{\ell}) - 1)$, provided $p \ge \frac{w(\ell+2)-\ell}{w-\ell}$.

Proof Since $\mathscr{F}_{j-1}$ 'lives' on the first h_{j-1} coordinates we obtain by [25] Lemma 7.2.1

$$\mathbb{E}\left(|\mathbb{E}(y_{j+k}|\mathscr{F}_{j-1})|^q \right) = \mathbb{E}\left(|y_{j+k}| \cdot |\mathbb{E}(y_{j+k}|\mathscr{F}_{j-1})|^{q-1} \right)$$

$$\le \|y_{j+k}\|_r \left\| |\mathbb{E}(y_{j+k}|\mathscr{F}_{j-1})|^{q-1} \right\|_s \beta^{\frac{1}{u}}(h_{j+k} - n_{j+k} - h_{j-1})$$

for $\frac{1}{r} + \frac{1}{s} + \frac{1}{u} = 1$, where

$$h_{j+k} - h_{j-1} = \sum_{i=j}^{j+k-1} (n_i + \Delta_i) + \Delta_{j-1}$$

$$= \mathcal{O}(1) \left(\sum_{i=j}^{j+k-1} \left(\sqrt{i} + i^{\frac{\alpha}{2}} \right) + (j-1)^{\frac{\alpha}{2}} \right)$$

$$= \mathcal{O}(1) \left((j+k-1)^{\frac{3}{2}} - j^{\frac{3}{2}} + j^{\frac{\alpha}{2}} \right)$$

Now

$$\left\| \left| \mathbb{E}(y_{j+k}|\mathscr{F}_{j-1}) \right|^{q-1} \right\|_s = \left(\mathbb{E} \left| \mathbb{E}(y_{j+k}|\mathscr{F}_{j-1}) \right|^{s(q-1)} \right)^{\frac{1}{s}}$$

and let s be so that $s(q-1) = q$ i.e. $s = \frac{q}{q-1}$. Then

$$\left(\mathbb{E} \left| \mathbb{E}(y_{j+k}|\mathscr{F}_{j-1}) \right|^{s(q-1)} \right)^{\frac{1}{q}} \leq \|y_{j+k}\|_r \beta^{\frac{1}{u}} (h_{j+k} - n_{j+k} - h_{j-1})$$

where we used that $1 - \frac{1}{s} = 1 - \frac{q-1}{q} = \frac{1}{q}$. By Proposition 12 $\mathbb{E}|y_{j+k}|^r \leq c_1 n_{j+k}^{\frac{r}{2}} \leq c_1 (j+k)^{\frac{r}{4}}$ which implies $\|y_{j+k}\|_r \leq c_2 (j+k)^{\frac{1}{4}}$ provided $r \leq \ell$. Since $(j+k-1)^{\frac{3}{2}} - j^{\frac{3}{2}} \geq k^{\frac{3}{2}}$ one has

$$\beta(h_{j+k} - n_{j+k} - h_{j-1}) \leq \beta \left(k^{\frac{3}{2}} + j^{\frac{\alpha}{2}} \right) \leq c_3 \left(k^{\frac{3}{2}} + j^{\frac{\alpha}{2}} \right)^{-p} \leq c_3 \min \left(j^{-\frac{\alpha p}{2}}, k^{-\frac{3p}{2}} \right)$$

and

$$\left\| \mathbb{E}(y_{j+k}|\mathscr{F}_{j-1}) \right\|_q^q \leq c_2 c_3 (j+k)^{\frac{1}{4}} \left(k^{\frac{3}{2}} + j^{\frac{\alpha}{2}} \right)^{-p}$$

and consequently with $r = \ell$ (largest possible) and $\frac{1}{u} = 1 - \frac{1}{q} - \frac{1}{\ell}$

$$\|u_j\|_q \leq c_2 c_3 \sum_{k=0}^{\infty} \left((j+k)^{\frac{1}{4}} \left(k^{\frac{3}{2}} + j^{\frac{\alpha}{2}} \right)^{-\frac{p}{u}} \right)^{\frac{1}{q}} = \mathcal{O} \left(j^{\frac{1}{4} - \frac{\alpha p}{2}(1 - \frac{1}{q} - \frac{1}{\ell})} \right)^{\frac{1}{q}}$$

provided p is large enough, so that $\frac{1}{4q}(1 - \frac{2\alpha p}{u}) < -1$ which is satisfied by assumption. $\square$

Lemma 16 *Let $\ell \geq 5$ and assume $p \geq \max(\frac{3}{2\alpha}, \frac{9}{\alpha(\ell-2)}, \frac{w(\ell+2)-\ell}{w-\ell})$. Then*

$$\sum_{j=1}^{Q_n} Y_j^2 = \sigma^2 n + \mathcal{O}(n^{1-\delta_3})$$

almost surely for every $\delta_3 < \min \left(\frac{1}{3}, \frac{2}{3}\alpha p - 1, (1 - \frac{4}{\ell})\frac{\alpha p}{6} \right)$.

Proof Put $v_j = u_j - u_{j+1}$, then $Y_j = y_j - v_j$ and $Y_j^2 = y_j^2 - 2v_j y_j + v_j^2$. We estimate the two terms separately. First by Lemma 15 with $(q = 2)$

$$\mathbb{E}(v_j^2) = \mathbb{E}(u_j - u_{j+1})^2 \leq 4(\max(\|u_j\|_2, \|u_{j+1}\|_2))^2 \leq c_1 j^{\frac{1}{2} - \alpha p}$$

$(c_1 \leq 4C_9)$ provided $p(\frac{1}{2} - \frac{1}{\ell}) > \frac{9}{2\alpha}$ (which is satisfied by assumption) and, using Cauchy-Schwarz,

$$\mathbb{E}|y_j v_j| \leq \|y_j\|_2 \|v_j\|_2 \leq c_1 \sqrt{\mathbb{E}(y_j^2)} j^{\frac{1}{4} - \frac{1}{2}\alpha p} \leq c_2 n_j^{\frac{1}{2}} j^{\frac{1}{4} - \frac{1}{2}\alpha p} \leq c_2 j^{\frac{1}{2} - \frac{1}{2}\alpha p}.$$

Then

$$\mathbb{E}\left(\sum_{j=1}^{Q_n} v_j^2\right) = \sum_{j=1}^{Q_n} \mathbb{E}(v_j^2) \leq c_1 \sum_{j+1}^{Q_n} j^{\frac{1}{2} - \alpha p} \leq c_4 Q_n^{\frac{3}{2} - \alpha p} \leq c_5 n^{1 - \frac{2}{3}\alpha p}$$

and with $\epsilon_n = n^{1 - \delta_3'}$

$$\mathbb{P}\left(\sum_{j=1}^{Q_n} v_j^2 \geq \epsilon_n\right) \leq \frac{1}{\epsilon_n} \mathbb{E}\left(\sum_{j=1}^{Q_n} v_j^2\right) \leq c_5 \frac{n^{1 - \frac{2}{3}\alpha p}}{\epsilon_n} \leq c_5 n^{\delta_3' - \frac{2}{3}\alpha p}.$$

If $\delta_3' - \frac{2}{3}\alpha p < -1$ then by Borel–Cantelli

$$\sum_{j=1}^{Q_n} v_j^2 < n^{1 - \delta_3'}$$

for all large enough n almost surely. It now follows from Lemma 14 that

$$\sum_{j=1}^{Q_n} Y_j^2 = \sum_{j=1}^{Q_n} y_j^2 + \mathcal{O}(n^{1 - \delta_3'}) = \sigma^2 n + \mathcal{O}(n^{1 - \delta_3})$$

almost surely for all $\delta_3 < \min\left(\frac{1}{3}, \frac{2}{3}\alpha p - 1, (1 - \frac{4}{\ell})\frac{\alpha p}{6}\right)$. $\square$

Lemma 17 *Let $\ell \geq 3$, then*

$$\sum_{j=1}^{Q_n} \left(\mathbb{E}(Y_j^2 | \mathscr{F}_{j-1}) - Y_j^2\right) \leq n^{1 - \delta_4}$$

almost surely, for any $\delta_4 < \frac{2}{3}(1 - \frac{2}{\ell})$.

Proof If we put $R_j = \mathbb{E}(Y_j^2 | \mathscr{F}_{j-1}) - Y_j^2$, a Martingale difference, then by Minkowsky and Proposition 12

$$\mathbb{E}|R_j^q| \leq \mathbb{E}|Y_j|^{2q} \leq \mathbb{E}|y_j|^{2q} + \mathbb{E}|v_j|^{2q} \leq C_5 n_j^q + C_6 j^{(\frac{1}{4} - \frac{1}{2}\alpha p)q} \leq c_1 j^{\frac{q}{2}}$$

provided $2q \leq \ell$, and therefore

$$\sum_{j=1}^{\infty} j^{-q\gamma} \mathbb{E}|R_j|^q \leq c_1 \sum_{j=1}^{\infty} j^{-q\gamma} j^{\frac{q}{2}} < \infty$$

if $q(\frac{1}{2} - \gamma) < -1$ i.e. $\gamma > \frac{1}{2} + \frac{1}{q}$. Then $\sum_j j^{-\gamma} R_j$ converges almost surely and therefore by Kronecker's lemma

$$\sum_{j=1}^{Q_n} R_j = \mathcal{O}(Q_n^\gamma) = \mathcal{O}(n^{\frac{2}{3}\gamma})$$

almost surely if $\frac{2}{3}\gamma = 1 - \delta_4$, where $\delta_4 < \frac{2}{3}(1 - \frac{1}{q})$ and where we chose $q = \frac{\ell}{2}$ (the largest possible value). (Note $\frac{2}{3}(1 - \frac{1}{q}) > \frac{1}{3}$.) □

Proof of Theorem. 2 By the Skorokhod representation theorem there exist T_j such

$$\sum_{j=1}^{Q_n} Y_j = X\left(\sum_{j=1}^{Q_n} T_j\right)$$

almost surely, where X is the Brownian motion and where $\mathbb{E}(T_j|\mathscr{F}_{j-1}) = \mathbb{E}(Y_j^2|\mathscr{F}_{j-1})$ a.s. and $\mathbb{E}(T_j^2) \le \mathbb{E}|Y_j|^{2q}$ for $q > 1$. Then, we conclude as in Philipp and Stout [25]

$$\sum_{j=1}^{Q_n} T_j - \sigma^2 n = \sum_{j=1}^{Q_n}\left(T_j - \mathbb{E}(T_j|\mathscr{F}_{j-1})\right) + \sum_{j=1}^{Q_n}\left(\mathbb{E}(T_j|\mathscr{F}_{j-1}) - Y_j^2\right) + \sum_{j=1}^{Q_n} Y_j^2 - \sigma^2 n.$$

For the first term on the RHS we use that

$$\mathbb{E}\left|T_j - \mathbb{E}(T_j|\mathscr{F}_{j-1})\right|^2 \le \mathbb{E}|T_j|^q \le \mathbb{E}|Y_j|^{2q},$$

for the second term we use Lemma 17 and the third term was estimated in Lemma 16. Notice that since $\ell \ge 5$ we get that $\frac{2}{3}(1 - \frac{2}{\ell}) > \frac{1}{3}$ (Lemma 17). Hence

$$\delta < \min\left(\frac{p(w-2) - 5w + 1}{2p(w-2) + 2w + 2}, \sup_{\alpha\in(0,1)} \min\left(\frac{1-\alpha}{3}, \frac{\alpha p}{30}, \frac{2\alpha p}{3} - 1\right)\right).$$

To get the statement of the theorem let us look at the second term and notice that the last two entries in it agree when $\alpha p = \frac{30}{19}$ and there they produce the value $\frac{1}{19}$. Hence the supremum is realised at $\alpha = \frac{10}{p+10}$ and its value equals $\frac{1}{3} - \frac{10}{3p+30}$. Consequently

$$\delta < \min\left(\frac{p(w-2) - 5w + 1}{2p(w-2) + 2w + 2}, \frac{1}{3} - \frac{10}{3p + 30}\right) = \frac{1}{3} - \frac{10}{3p + 30} \tag{8}$$

for $w > 5$ and $p > 7 + \frac{30}{w-5}$. □

Proof of Corollary 3. Part (i) is obvious and in part (ii) we let p go to infinity which leads to the condition $\delta < \frac{1}{3}$. □

Lemma 18 *Assume $\mathcal{A}$ is a finite alphabet, μ is β-mixing where β decays at least least polynomially with power $p > 0$. Then*

$$\mathbb{E}\left|I_n - \mathbb{E}(I_n|\mathcal{C}_\infty)\right| = \mathcal{O}(n^\eta)$$

for any $\eta > \frac{1}{p+1}$, where $\mathcal{C}_\infty = \bigvee_{j=-\infty}^{-1} T^j \mathcal{A}$.

Proof Let $\mathcal{B} = \mathcal{A}^n$ and $\mathcal{C} = T^{-m}\mathcal{A}^m = \bigvee_{j=-m}^{-1} T^j \mathcal{A}$. We want to estimate $\int_\Omega |I_n - \mathbb{E}(I_n|\mathcal{C})|\, d\mu$ and then let m go to infinity. Let $\alpha > 0$ and put $\Delta = n^\alpha$. Put $\mathcal{L}^+ = \{(B, C) \in \mathcal{B} \times \mathcal{C} : \frac{\mu(B\cap C)}{\mu(B)\mu(C)} \ge 2\}$ and similarly $\mathcal{L}^- = \{(B, C) : \frac{\mu(B\cap C)}{\mu(B)\mu(C)} < 2\}$. Then with $\phi(x) = |\log x|x$ we get

$$F^- = \sum_{(B,C)\in\mathcal{L}^-} \mu(B \cap C)\log\frac{\mu(B \cap C)}{\mu(B)\mu(C)} \le \sum_{(B,C)\in\mathcal{L}^-} \phi\left(\frac{\mu(B \cap C)}{\mu(B)\mu(C)}\right)\mu(B)\mu(C) \le 2$$

as $\phi(x) \leq 2$ for $x \in (0, 2]$. On $\mathcal{L}^+$ we proceed as follows:

$$
\begin{aligned}
F^+ &= \sum_{(B,C)\in\mathcal{L}^+} \mu(B \cap C) \log \frac{\mu(B \cap C)}{\mu(B)\mu(C)} \\
&= \sum_{\tilde{C}\in\tilde{\mathcal{C}}} \sum_{(B,C)\in\mathcal{L}^+, C\subset\tilde{C}} \mu(B \cap C) \left(\log \frac{\mu(B \cap C)}{\mu(B \cap \tilde{C})} + \log \frac{\mu(B \cap \tilde{C})}{\mu(B)\mu(\tilde{C})} + \log \frac{\mu(\tilde{C})}{\mu(C)} \right) \\
&\leq \sum_{\tilde{C}\in\tilde{\mathcal{C}}} \sum_{(B,C)\in\mathcal{L}, C\subset\tilde{C}} \mu(B \cap C) \left| \log \frac{\mu(B \cap \tilde{C})}{\mu(B)\mu(\tilde{C})} \right| + H(\mathcal{B} \vee \mathcal{C}|\mathcal{B} \vee \tilde{C}) + H(\mathcal{C}|\tilde{C})
\end{aligned}
$$

Where $\mathcal{L} = \mathcal{L}^- \cup \mathcal{L}^+$ and $\tilde{C} = T^{-m}\mathcal{A}^{m-\Delta}$ and $\tilde{C} \in \tilde{\mathcal{C}}$ is understood to be chosen so that $C \subset \tilde{C}$. For the two error terms we readily estimate $H(\mathcal{C}|\tilde{C}) \leq H(\mathcal{A}^\Delta) = \mathcal{O}(\Delta)$ and similarly $H(\mathcal{B} \vee \mathcal{C}|\mathcal{B} \vee \tilde{C}) \leq H(\mathcal{A}^\Delta) = \mathcal{O}(\Delta)$. For the principal term let us put as above $\tilde{\mathcal{L}}^+ = \{(B, \tilde{C}) \in \mathcal{B} \times \tilde{C} : \frac{\mu(B\cap\tilde{C})}{\mu(B)\mu(\tilde{C})} \geq 2\}$ and similarly $\tilde{\mathcal{L}}^- = \{(B, \tilde{C}) : \frac{\mu(B\cap\tilde{C})}{\mu(B)\mu(\tilde{C})} < 2\}$. Then as above we estimate

$$
\tilde{F}^- = \sum_{(B,\tilde{C})\in\tilde{\mathcal{L}}^-} \mu(B \cap \tilde{C}) \log \frac{\mu(B \cap \tilde{C})}{\mu(B)\mu(\tilde{C})} \leq 2,
$$

Now let $\gamma > 1$ and put $\mathcal{T} = \{(B, \tilde{C}) \in \mathcal{B} \times \tilde{C} : \frac{\mu(B\cap\tilde{C})}{\mu(B)\mu(\tilde{C})} \geq e^{n^\gamma}\}$. Then, assuming $\alpha \leq 1$,

$$
\mu \left(\bigcup_{(B,\tilde{C})\in\mathcal{T}} B \cap \tilde{C} \right) \leq |\mathcal{A}|^\Delta e^{-n^\gamma} = \mathcal{O}(e^{-n^\gamma})
$$

for all n large enough as $\frac{\mu(B\cap\tilde{C})}{\mu(B)\mu(\tilde{C})} \leq \frac{1}{\mu(B)}$ implies $\mu(B) \leq e^{-n^\gamma}$ for all $(B, \tilde{C}) \in \mathcal{T}$. Hence

$$
\begin{aligned}
\tilde{F}^+ &= \sum_{(B,\tilde{C})\in\tilde{\mathcal{L}}^+} \mu(B \cap \tilde{C}) \log \frac{\mu(B \cap \tilde{C})}{\mu(B)\mu(\tilde{C})} \\
&\leq n^\gamma \sum_{(B,\tilde{C})\in\tilde{\mathcal{L}}^+\setminus\mathcal{T}} \mu(B \cap \tilde{C}) + \mathcal{O}(e^{-n^\gamma}) \\
&\leq 2n^\gamma \sum_{(B,\tilde{C})\in\tilde{\mathcal{L}}^+\setminus\mathcal{T}} |\mu(B \cap \tilde{C}) - \mu(B)\mu(\tilde{C})| + \mathcal{O}(e^{-n^\gamma}) \\
&= \mathcal{O}(n^\gamma \beta(\Delta)),
\end{aligned}
$$

where in the last step we used that $\mu(B \cap \tilde{C}) \geq 2\mu(B)\mu(\tilde{C})$ for all $(B, \tilde{C}) \in \tilde{\mathcal{L}}^+$. Moreover note that $F^+ + F^- = H(\mathcal{B}) + H(\mathcal{C}) - H(\mathcal{B} \vee \mathcal{C})$ is non-negative by subadditivity of the entropy function. Similarly $\tilde{F}^+ + \tilde{F}^- \geq 0$. In particular we also receive the lower bound $\tilde{F}^- = \mathcal{O}(n^\gamma \beta(\Delta))$. We have thus shown that

$$
\begin{aligned}
\mathbb{E}|I_\mathcal{B} - \mathbb{E}(I_\mathcal{B}|\mathcal{C})| &= \sum_{B,C} \mu(B \cap C) \left| \log \frac{\mu(B \cap \tilde{C})}{\mu(B)\mu(\tilde{C})} \right| \\
&= \mathcal{O}(n^\gamma \beta(\Delta)) + \mathcal{O}(\Delta) \\
&= \mathcal{O}(n^{\gamma-\alpha p} + n^\alpha)
\end{aligned}
$$

which is optimised for the value $\alpha = \frac{\gamma}{p+1}$ where $\gamma > 1$ is arbitrary. This estimate is true for any value of m. Letting m go to infinity we obtain that $\mathbb{E}\,|I_\mathcal{B} - \mathbb{E}(I_\mathcal{B}|\mathcal{C}_\infty)| = \mathcal{O}(n^\eta)$ for any $\eta > \frac{1}{p+1}$, where $\mathcal{C}_\infty = \bigvee_{j=-\infty}^{-1} T^j \mathcal{A}$. $\qquad\square$

Let us note that the conclusion of this lemma is much weaker than the requirement of summability of $\|f - f_n\|$ which is required by Ibragimov and also used by Kontoyiannis.

Proof of Theorem. 6 In [16, Theorem 1] it was shown that for all ergodic measures one has $\limsup_{n\to\infty} n^{-\zeta} (\log R_n(x) - I_n(x)) \leq 0$ and also $\liminf_{n\to\infty} n^{-\zeta} (\log R_n(x) - \mathbb{E}(I_n|\mathcal{C}_\infty)) \geq 0$ for a.e. x and for all $\zeta > 0$. This combined with Lemma 18 proves the theorem. $\qquad\square$

References

1. Abadi, M.: Exponential approximation for hitting times in mixing stochastic processes. Math. Phys. Electron. J. **7**(2) (2001)
2. Brin, M., Katok, A.: On local entropy; Geometric Dynamics, pp. 30–38, Springer LNM #1007 (1983)
3. Broise, A: Transformations dilatantes de l'intervalle et théorèmes limites; Asterisque #238 (1996)
4. Bruin, H., Vaienti, S: Return times for unimodal maps. Forum Math. **176**, 77–94 (2003)
5. Chernov, N.: Limit theorems and Markov approximations for chaotic dynamical systems. Prob. Th. Rel. Fields **101**, 321–362 (1995)
6. Collet, P., Galves, A., Schmitt, B.: Fluctuations of repetition times for Gibbsian sources. Nonlinearity **12**, 1225–1237 (1999)
7. Doukhan, P: Mixing: properties and examples. Lecture Notes in Statistics, vol. 85. Springer, Berlin (1995)
8. Ferrero, P., Haydn, N., Vaienti, S.: Entropy fluctuations for parabolic maps. Nonlinearity **16**, 1203–1218 (2003)
9. Galves, A., Schmitt, B.: Inequalities for hitting times in mixing dynamical systems. Random Comput. Dyn. **5**, 337–348 (1997)
10. Gordin, M.: The central limit theorem for stationary processes. Soviet Math. Doklady **10**, 1174–1176 (1969)
11. Han, Guanyue: Limit theorems in hidden Markov models. IEEE Trans. Inf. Theory **59**(3), 1311–1328 (2013)
12. Haydn, N: The central limit theorem for uniformly strong mixing measures. Stoch. Dyn. (2012). doi:10.1142/S0219493712500062
13. Haydn, N., Vaienti, S.: Fluctuations of the metric entropy for mixing measures. Stoch. Dyn. **4**, 595–627 (2004)
14. Haydn, N., Vaienti, S: The distribution of the measure of cylinders for non-Gibbsian measures; Complex Dynamics and Related Topics 147–162, New Studies in Advanced Mathematics #5 (2003)
15. Ibragimov, I.A.: Some limit theorems for stationary processes. Theory Prob. Appl. **7**, 349–382 (1962)
16. Kontoyiannis, I.: Asymptotic recurrence and waiting times for stationary processes. J. Theor. Prob. **11**, 795–811 (1998)
17. Liverani, C.: Central Limit theorem for deterministic systems; Intern. Congress on Dyn. Syst., Montevideo. Proc. Research Notes in Math. Series. Pitman **1996**, 56–75 (1996)
18. Mañé, R.: Ergodic Theory and Differentiable Dynamics. Springer, Berlin (1987)
19. Nobel, A., Wyner, A.: A recurrence theorem for dependent processes with applications to data compression. IEEE **38**(5), 1561–1564 (1992)
20. Ornstein and Weiss; Entropy and Data Compression Schemes. IEEE Trans. Inf. Theory **39**, 78–83 (1993)
21. Ornstein and Weiss; Entropy and Recurrence Rates for: Stationary Random Fields. IEEE Trans. Inf. Theory **48**(6), 1694–1697 (2002)
22. Paccaut, F.: Propriétés Statistiques de Systèmes Dynamiques Non Markovian; Thèse (Doctorat) Dijon 2000
23. Pène, F.: Rates of convergence in the CLT for two-dimensional dispersive Billiards. Commun. Math. Phys. **225**, 91–119 (2002)
24. Petersen, K: Ergodic Theory: Cambridge studies in advanced mathematics #2 (1983)
25. Philipp, W., Stout, W: Almost sure invariance principles for partial sums of weakly dependent random variables. AMS Mem. **2**(2), 161 (1975)
26. Romanovskii, V.I.: Discrete Markov Chains. Wolters-Noordhoff Publishing Groningen (1970)

27. Rosenblatt, M.: A central limit theorem and a strong mixing condition. Proc. Nat. Acad. Sci. USA **42**, 43–47 (1956)
28. Rosenblatt, M: Markov Processes. Structure and Asymptotic Behavior. Springer, Grundlagen #184, (1971)
29. Walters, P.: An Introduction to Ergodic Theory. Springer, Berlin (1981)
30. Wyner, A., Ziv, J.: Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. IEEE Trans. Inf. Theory **35**, 1250–8 (1989)
31. Yushkevich, A.: On limit theorems connected with the concept of the entropy of Markov chains (in Russian). Uspehi Math. Nauk **8**, 177–180 (1953)