

# Number of visits in arbitrary sets for $\phi$ -mixing dynamics

Sandro Gallo<sup>1,a</sup>, Nicolai Haydn<sup>2,b</sup> and Sandro Vaienti<sup>3,c</sup>

<sup>1</sup>*Departamento de Estatística, Universidade Federal de São Carlos, CEP: 13565-905, São Paulo, Brasil, [sandro.gallo@ufscar.br](mailto:sandro.gallo@ufscar.br)*

<sup>2</sup>*Mathematics Department, USC, Los Angeles, 90089-2532, [nhaydn@math.usc.edu](mailto:nhaydn@math.usc.edu)*

<sup>3</sup>*Aix Marseille Université, Université de Toulon, CNRS, CPT, 13009 Marseille, France, [vaienti@cpt.univ-mrs.fr](mailto:vaienti@cpt.univ-mrs.fr)*

Received 4 September 2021; revised 11 July 2022; accepted 14 November 2022

**Abstract.** It is well-known that, for sufficiently mixing dynamical systems, the number of visits to balls and cylinders of vanishing measure is approximately Poisson compound distributed in the Kac scaling. Here we extend this kind of results when the target set is an arbitrary set with vanishing measure in the case of  $\phi$ -mixing systems. The error of approximation in total variation is derived using Stein–Chen method. An important part of the paper is dedicated to examples to illustrate the assumptions, as well as applications to temporal synchronisation of  $g$ -measures.

**Résumé.** Il est bien connu que, pour les systèmes dynamiques suffisamment mélangeant, la loi du nombre de visites dans les boules et les cylindres de mesure tendant vers zéro, est proche d'une loi de Poisson composée à l'échelle de Kac. Ici, nous étendons ce type de résultats lorsque l'ensemble cible est un ensemble arbitraire de mesure qui tend vers zéro, dans le cas des systèmes  $\phi$ -mélangeants. L'erreur d'approximation en variation totale est obtenue à l'aide de la méthode de Stein–Chen. Une partie importante de l'article est consacrée à des exemples pour illustrer les hypothèses, ainsi qu'à des applications à la synchronisation temporelle de  $g$ -mesures.

*MSC2020 subject classifications:* Primary 37B20; 37D35; secondary 37A25; 37A50; 60G70

*Keywords:* Poincaré recurrence; Synchronization of dynamical systems;  $g$ -measures; Mixing processes; Compound poisson distribution

## 1. Introduction

The recurrence in small sets, which could be seen alternatively as a *rare* or *extreme* event, turned out to have very rich probabilistic features and established itself as a major statistical property of dynamical systems. We consider in this paper the general situation of a measurable deterministic dynamical system and try to characterise the distribution of the number of visits to sets whose measure will tend to zero. Since the probability to visit the set coincides with its measure for ergodic systems, one should normalise the length of the trajectory with the measure of the set, in order to get meaningful asymptotic distributions. We called it, in the paper, the Kac's scaling. If the system loses memory fast enough in the future, which is achieved with relatively strong mixing properties, the number of visits of a trajectory of length  $n$  tends to follow a binomial distribution  $B(n, p_n)$ , where  $p_n$  is the measure of the small set. Kac's scaling requires that the product  $np_n$  equals asymptotically the constant  $t$  and therefore one gets a Poisson law of parameter  $t$  in the limit of large  $n$  for the number of visits up to time  $\frac{t}{p_n}$ . The implementation of this heuristic argument, for a given measure preserving dynamical system, requires not only mixing properties, as we said above, but also some control on the *nature* of the small sets. When the map acts on a metric space, the small set is usually taken as a ball around a given point  $z$  and with radius shrinking to zero. The *nature* of the point  $z$  could change the limit distribution. Suppose in fact that  $z$  is a periodic point; even if the system is mixing, the orbits starting or passing close to  $z$  tend to sojourn for a longer time in the small set. This produces an effect of *clusterization* which will alter the Poisson law into a more general *compound Poisson distribution*.

The aim of the present paper is to obtain such results for measurable dynamical systems and for a wide class of small sets. The latter are obtained by fixing an initial measurable generating partition and by taking its backward (and eventually forward for invertible systems) joins. An arbitrary countable disjoint union of elements of the join of order  $n$  will be a small set  $U_n$ . We will also assume that the sequence  $\{U_n\}_{n \geq 1}$  is nested and that it converges to a set of measure zero. The asymptotic distribution of successive visits to  $U_n$  will be assured by requiring that the invariant measure is  $\phi$ , or  $\psi$ -mixing with respect to the initial partition.

First of all we proceed to adapt the Stein–Chen method [8,15,47,49] to compare a given probability measure, in our case the distribution of the number  $W_n$  of visits to a set, to a compound Poisson distribution. This will give us an error for the total variation distance between the two distributions. Any compound Poisson distribution depends upon a set of parameters  $\lambda_l, l \geq 1$ . It has been shown in [29], that those parameters are related to another sequence  $\alpha_l, l \geq 1$ , (see Section 2.2) which quantify the distribution of higher order returns. Whenever the limits defining the  $\alpha_l$  exist and verify a summable condition, the error term given by the Stein–Chen method will go to zero, and therefore we recover the expected convergence to a compound Poisson law: this is the content of the main result, Theorem 5. Applications to concrete examples basically require to check two conditions on the system: (i) first of all the  $\phi$ , or  $\psi$ -mixing property, which enters the estimate of the error in the Stein–Chen approach; (ii) secondly the existence and summability of the  $\alpha_l$ , which instead depend on the system and on the choice of the nested sequence of small sets  $U_n$ . A similar program was carried over in [29], with a few substantial differences which in particular imply that the examples given in the present paper cannot be covered by the theory developed in [29]. The latter targets  $C^2$  local diffeomorphisms on smooth manifolds and satisfying a few geometrical and metric conditions, among which the most relevant are: (a) local hyperbolicity and distortion; (b) the annulus-type condition which allows to control the relative measure of the neighborhoods of the small sets; and finally (c) the decay of correlations which is stated in terms of Lipschitz against  $L^\infty$  norms. The technique of the proof of [29] was different from the Stein–Chen’ used here and it had a more geometric flavour, adapted to differentiable dynamical systems. In particular it was possible to handle partially hyperbolic maps and synchronisation of coupled map lattices. In the latter case and for the invariant absolutely continuous measure, it has been established that the returns to the diagonal is compound Poisson distributed where the coefficients are given by certain integrals along the diagonal. This example is reconsidered in this paper and compared with a different way to collect points close to each other in the attempt to synchronise two or more trajectories. In the spirit of the present paper, a neighborhood of the diagonal will be constructed with the elements of the join partition of increasing order, also called cylinders. As we said above, cylinders and union of cylinders will be our small sets. If the dynamical system is encoded in a symbolic space, we could transport our theory in the domain of symbolic dynamics and cover new panels of examples which are unattainable with the previous geometric approach. Among the applications investigated in the paper, we quote here the *House of cards process*, for which the distribution of the number of visits to runs of length above a given threshold is found to be Pólya–Aeppli, and a class of (not necessarily Markovian) regenerative processes for which we compute explicitly the parameters of the compound Poisson distribution. In particular we exhibit the existence of the quantity  $\alpha_1$  which takes on a particular role in *extreme value theory*, where it coincides with the *extremal index*.

An important part of the paper is dedicated to  $g$ -measures (see Section 4). These are equilibrium states with normalized potentials of the form  $\phi = \log g$ , where  $g$  is the  $g$ -function [35]. These objects form the counterpart, in the dynamical system setting, of the (possibly long memory) stochastic processes. For this class of models, we give mild sufficient conditions (strict positivity and summable variation), allowing us to apply our theorem for the number of visits in cylinders around periodic points. It has been recently shown [3] that for a particular class of  $g$ -measures called *renewal measures*, it is possible to show that the limit defining the extremal index does not exist even though the measure is  $\phi$ -mixing and this was due to an essential discontinuity of  $g$  in a given point. Here we will prove that uniform continuity is enough for the existence of the extremal index and, by discussing an example due to [25], that the lack of continuity of  $g$  does not prevent the existence of the parameter, leading to a Pólya–Aeppli distribution around any periodic point. We will then consider a decreasing cover by cylinders of the diagonal in the  $m$ -dimensional product space where a given  $g$ -measure is seen as the coupling of the  $m$  coordinates  $g$ -measures. This will allow us to study the synchronisation of the coordinates, what we called *temporal synchronisation for  $g$ -measures*. In the general case where the coordinate  $g$ -measures are not independent, we will show the convergence to a Pólya–Aeppli distribution whose parameter is related to the topological pressure of a given potential, see Theorem 12. It is interesting to observe that whenever the coordinate  $g$ -measures are independent (the uncoupling case), the previous parameter can be expressed in terms of the Renyi entropy of order  $m - 1$ . We also address the more general question of the interaction of possibly distinct  $g$ -measures and we propose two ways to construct such an interaction.

Finally, in a discussion section on synchronisation, we highlight a difference between the geometric approach of [29] and the symbolic approach of the present work. In a simple example of two uncoupled identical deterministic dynamical systems, we show that the asymptotic distribution of synchronisations is Pólya–Aeppli when we target the diagonal by cylinder sets (symbolic approach) but it is not Pólya–Aeppli when we target the diagonal by tubular neighbourhoods (geometric approach). We conclude the discussion by considering yet another situation, of two uncoupled copies of the same Markov chain on  $[0, 1]$ , with strong ergodicity conditions. We show that the asymptotic synchronisation always follows a pure Poisson distribution, meaning that there is no clustering phenomenon in that setting.

We previously compared our achievements with the results obtained in the paper [29]; several other contributions deserve to be quoted and we will give here a brief survey of them. We should first remind the seminal papers by [46] and [33] who showed that generic points have, in the limit, Poisson distributed return times if one uses cylinder neighbourhoods,

while at periodic points the return times distribution has a point mass at the origin which corresponds to the periodicity of the point. This dichotomy inspired and originated several successive works: it was proved in [1] for  $\phi$ -mixing systems in the symbolic setting, and in [28] for more general classes of dynamical systems with various kind of mixing properties. The latter paper derived also the error terms for the convergence to the limiting compound Poissonian distribution. The extension to  $\psi$ -mixing shifts was given in [38]; for  $\phi$ -mixing systems a recent contribution is provided in [39]. We also quote [13] who studied Poisson limit laws for subsystems of finite type in symbolic dynamics. The Chen–Stein method, which is at the base of the actual work, was firstly used in [27] for  $\phi$ -mixing measures and cylinder sets. A complementary approach to the statistics of the number of visits has been developed in the framework of extreme value theory, where it is more often called point process, or particular kinds of it as the marked point process associated to extremal observations corresponding to exceedances of high thresholds. See for instance [21–23] for applications to deterministic and random dynamical systems, and the book [41] for a panorama and an account on extreme value theory and point processes applied to dynamical systems. The distribution of the number of visits to vanishing balls has been studied for systems modeled by a Young tower: in [14] for the Hénon attractor, in [44] for some nonuniformly hyperbolic invertible dynamical systems, in [30] and [31] for polynomially decaying correlations. Recurrence in billiards provided recently several new contributions; for planar billiards in [43] and [24]; in [45] the spatio-temporal Poisson processes was obtained from recording not only the successive times of visits to a set, but also the positions.

The paper is structured as follows. We directly state the main results in Section 2, and follow with a discussion concerning assumptions and examples in Section 3. Next, Section 4 specializes to the case of  $g$ -measures. We provide one further discussion on synchronization in Section 5, and conclude with Sections 6 and 7 containing the proofs of all the results.

## 2. Main results

Let  $T$  be a measurable map on a measure space  $\Omega$  and  $\mu$  a  $T$ -invariant measure on  $\Omega$ . Moreover let  $\mathcal{A}$  be a countable measurable partition on  $\Omega$  and denote by  $\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}$  be the joins of  $\mathcal{A}$ . We assume that  $\mathcal{A}$  is *generating*, that is,  $\bigvee_{j=0}^{\infty} T^{-j} \mathcal{A}$  consists of singletons.<sup>1</sup> For every measurable set  $U$  we will denote by  $\mu_U$  with  $\mu(U) > 0$  the measure conditioned on (the points starting in)  $U$ , that is  $\mu_U(A) := \frac{\mu(U \cap A)}{\mu(U)}$ . As usual, for any collection of sets  $\mathcal{B}$  we denote by  $\sigma(\mathcal{B})$  the smallest  $\sigma$ -algebra generated by  $\mathcal{B}$ .

### 2.1. Distribution of the number of visits in a fixed set $U$

Initially, our interest will be to characterise the distribution of the number of visits to sets with small measure. To this end, we define for any fixed set  $U$  and for any  $t > 0$  the  $\mathbb{N}$ -valued random variable

$$(1) \quad W = \sum_{i=0}^{t/\mu(U)} \mathbb{1}_U \circ T^i,$$

which counts the number of visits to  $U$  in the Kac’s scaling  $t/\mu(U)$ . Although  $W$  depends on  $t$  and  $U$ , we do not explicit this dependence in the notation for the sake of simplicity. Our first theorem gives an upper bound on the total variation distance<sup>2</sup> between  $\mu(W \in \cdot)$  and a compound Poisson distribution  $\tilde{\nu}$  with parameters  $t\tilde{\lambda}_\ell$ ,  $\ell \geq 1$ , that is, a probability distribution with generating function

$$(2) \quad \varphi_{\tilde{\nu}}(z) = e^{\sum_{k \geq 1} t\tilde{\lambda}_k (e^{zk} - 1)}.$$

Naturally, the parameters  $\tilde{\lambda}_\ell$ ,  $\ell \geq 1$  will depend on the dynamic.

We stand in the world of  $\phi$ -mixing measures; we point out that the following mixing conditions are stated in terms of the one-sided join  $\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}$  which allows to open a gap between events separated in the future.

<sup>1</sup>When the map  $T$  is invertible, we ask  $\bigvee_{j=-\infty}^{\infty} T^{-j} \mathcal{A}$  to consist of singletons in order to guarantee that  $\mathcal{A}$  is generating. We need generating partitions since in the following we consider descending nested sequences of cylinders as “rare” sets, even collapsing around points.

<sup>2</sup>The total variation between two probability distributions  $P$  and  $Q$  on some measurable space  $(\Omega, \mathcal{F})$  is defined as  $\|P(\cdot) - Q(\cdot)\| = \sup_{B \in \mathcal{F}} |P(B) - Q(B)|$ .

**Definition 1.** We say a  $T$ -invariant probability measure  $\mu$  on  $\Omega$  is *left  $\phi$ -mixing* with respect to the partition  $\mathcal{A}$  if there exists a decreasing sequence  $\phi(k) \searrow 0$  so that for every  $n, m \geq 1$ ,  $U \in \sigma(\mathcal{A}^n)$  and  $V \in \sigma(\bigcup_{m=1}^{\infty} \mathcal{A}^m)$ :

$$(3) \quad \left| \frac{\mu(U \cap T^{-n-k}V)}{\mu(U)} - \mu(V) \right| \leq \phi(k).$$

Similarly we say  $\mu$  is *right  $\phi$ -mixing* if under the same conditions

$$(4) \quad \left| \frac{\mu(U \cap T^{-n-k}V)}{\mu(V)} - \mu(U) \right| \leq \phi(k).$$

For any  $i \geq 1$  and any  $1 \leq K \leq i$ , the variable counting the number of visits to  $U$  at a distance less or equal to  $K$  around  $i$  is

$$(5) \quad Z_i^{(K)} := \sum_{j=i-K}^{i+K} \mathbb{1}_U \circ T^j.$$

For  $x \in \Omega$  denote by  $A_j(x)$  the unique atom of  $\mathcal{A}^j$  which contains  $x$ . More generally for a set  $U \subset \Omega$  we put for its outer  $j$ -cylinder approximation of  $U$  ( $j \geq 1$ )

$$(6) \quad U^j := A_j(U) = \bigcup_{A \in \mathcal{A}^j, A \cap U \neq \emptyset} A.$$

Similarly for  $U \subset \Omega$  and integer then, for  $j < n$ , we also define the  $n$ -right  $j$ -cylinder approximation by:

$$(7) \quad \tilde{U}_n^j = T^{-(n-j)} A_j(T^{n-j}U) = T^{-(n-j)} \bigcup_{A \in \mathcal{A}^j, A \cap T^{n-j}U \neq \emptyset} A.$$

In the case when  $U \in \sigma(\mathcal{A}^n)$  (union of  $n$ -cylinders) then we shall write below  $\tilde{U}^j$  for  $\tilde{U}_n^j$ . (In Remark 5 we will give an example of a null set whose  $n$ -right  $j$ -cylinder approximation is the entire space for all  $j < n/2$ .) We are now ready to state our first main result where we denote by  $\phi^1(\ell) = \sum_{j=\ell}^{\infty} \phi(j)$  the tail sum of  $\phi$ .

**Theorem 2.** Let  $\mu$  be a  $T$ -invariant probability measure on  $\Omega$  which is right  $\phi$ -mixing with  $\phi$  summable. Then there exists a constant  $C_1$  so that, for any measurable set  $U \in \sigma(\mathcal{A}^n)$ , any  $t > 0$  and any  $K < t/\mu(U)$ , we have

$$(8) \quad \|\mu(W \in \cdot) - \tilde{\nu}_{K,U}\|_{TV} \leq C_1 t \inf_{K < \Delta < t/\mu(U)} \left( K \frac{\phi(\Delta - n)}{\mu(U)} + \Delta \mu(U) + \phi^1(K/2) + \sum_{j=K/2}^n \mu(U^j) \right),$$

( $U^j$  as defined in (6)) where  $\tilde{\nu}_{K,U}$  is the compound Poisson distribution with parameters  $t\tilde{\lambda}_\ell(K, U)$ ,  $\ell \geq 1$ , where

$$(9) \quad \tilde{\lambda}_\ell(K, U) := \frac{1}{\ell} \mathbb{E}(\mathbb{1}_{Z_i^{(K)} = \ell} | \mathbb{1}_U \circ T^i = 1), \quad \forall i \geq K.$$

If we assume left  $\phi$ -mixing instead of right  $\phi$ -mixing, the same statement holds after replacing the  $j$ -cylinder approximations  $U^j$  by the  $n$ -right  $j$ -cylinder approximations  $\tilde{U}^j$  (as defined in (7)) of  $U \in \sigma(\mathcal{A}^n)$ .

We now consider the case in which we have a stronger kind of mixing called  $\psi$ -mixing.

**Definition 3.** We say a  $T$ -invariant probability measure  $\mu$  on  $\Omega$  is  $\psi$ -mixing with respect to the partition  $\mathcal{A}$  if there exists a decreasing sequence  $\psi(k) \searrow 0$  so that for every  $n, m \geq 1$ ,  $U \in \sigma(\mathcal{A}^n)$  and  $V \in \sigma(\bigcup_{m=1}^{\infty} \mathcal{A}^m)$ :

$$\left| \frac{\mu(U \cap T^{-n-k}V)}{\mu(U)\mu(V)} - 1 \right| \leq \psi(k).$$

This stronger assumption naturally yields a stronger result.

**Theorem 4.** Let  $\mu$  be a  $T$ -invariant probability measure on  $\Omega$  which is  $\psi$ -mixing where  $\psi(j) \rightarrow 0$  as  $j \rightarrow \infty$ . Then there exists a constant  $C'_1$  so that, for any measurable set  $U \in \sigma(\mathcal{A}^n)$ , any  $t > 0$  and any  $K < t/\mu(U)$ , one has

$$(10) \quad \|\mu(W \in \cdot) - \tilde{v}_{K,U}\|_{TV} \leq C'_1 t \inf_{K < \Delta < t/\mu(U)} \left( \psi(\Delta - n) + \Delta\mu(U) + \sum_{j=K/2}^n \mu(U^j) \right),$$

where  $\tilde{v}_{K,U}$  is the compound Poisson distribution with parameters  $t\tilde{\lambda}_\ell(K, U)$ ,  $\ell \geq 1$  given by (9).

By symmetry of  $\psi$ -mixing, the same inequality holds with  $\tilde{U}^j$  instead of  $U^j$  on the RHS.

2.2. Asymptotic distribution of the number of visits in a nested sequence  $\{U_n\}_{n \geq 1}$

Now we will consider nested sequences of measurable sets  $U_1 \supset U_2 \supset \dots$  satisfying  $\mu(U_n) \rightarrow 0$ . We will denote by  $\Gamma$  the limiting null-set. Our interest is to study the convergence in distribution of

$$W_n := \sum_{i=0}^{t/\mu(U_n)} \mathbb{1}_{U_n} \circ T^i$$

as  $n$  diverges, for any  $t > 0$ .

Naturally, it is expected that, if in the Poisson compound approximations of the preceding theorems the involved parameters (9) converge and we can further control the error terms, then we would have a Poisson compound distribution in the limit, parametrised by the limiting parameters. The statement of such a result needs some more definitions on the entry/return time probabilities and the corresponding limiting quantities.

For a subset  $U \subset \Omega$  we define the first entry/return time  $\tau_U$  by  $\tau_U(x) = \min\{j \geq 1 : T^j x \in U\}$ . Similarly we get higher order returns by defining recursively  $\tau_U^\ell(x) = \tau_U^{\ell-1} + \tau_U(T^{\tau_U^{\ell-1}}(x))$  with  $\tau_U^1 = \tau_U$ . We also write  $\tau_U^0 = 0$  on  $U$ .

We now come back to our nested sequence of sets  $U_n$ ,  $n \geq 1$  and define (provided the limits exist) for  $k, L, n \geq 1$

$$(11) \quad \begin{aligned} \alpha_k(L, U_n) &:= \mu_{U_n}(\tau_{U_n}^{k-1} \leq L < \tau_{U_n}^k), \\ \alpha_k(L) &:= \lim_{n \rightarrow \infty} \alpha_k(L, U_n), \\ \alpha_k &:= \lim_{L \rightarrow \infty} \alpha_k(L). \end{aligned}$$

As promised, using Theorems 2 and 4, and under proper further assumptions, we establish that the limiting distribution of the number of visits to the  $U_n$ 's is asymptotically compound Poisson.

**Theorem 5.** Consider a nested sequence of sets  $U_n \in \sigma(\mathcal{A}^n)$ ,  $n \geq 1$ , converging to a null-set  $\Gamma$ . Suppose the  $T$ -invariant probability measure  $\mu$  satisfies:

- (1) either  $\psi$ -mixing, or right  $\phi$ -mixing with  $\phi$  summable,
- (2) there exists a vanishing sequence of positive real numbers  $a_k$ ,  $k \geq 1$  such that  $\sum_{i=k}^n \mu(U_n^i) \leq a_k$  for all sufficiently large  $n$ 's,
- (3)  $\sum_{k=1}^\infty k^2 a_k < \infty$  (and naturally that the  $\alpha_k$ ,  $k \geq 1$ , exist, see (11)).

Then, for every  $E \subset \mathbb{N}_0$  one has

$$\mu(W_n \in E) \longrightarrow \tilde{v}(E)$$

as  $n \rightarrow \infty$ , where  $\tilde{v}$  is the compound Poisson distribution with parameters  $t\tilde{\lambda}_\ell$ ,  $\ell \geq 1$  and

$$\tilde{\lambda}_k := \alpha_k - \alpha_{k+1}.$$

If in assumption (1) we rather assume left  $\phi$ -mixing with  $\phi$  summable, then we have to change  $U_n^i$  to  $\tilde{U}_n^i$  in (2), and the same statement applies.

**3. Discussion of the results**

In this section, we list a series of remarks concerning the results presented in the previous section, together with some example illustrating these remarks.

### 3.1. Concerning the assumptions

Here we discuss the assumptions of the above theorems.

- The results and proofs apply to invertible and non-invertible maps  $T$ .
- It is classical in recurrence theory for dynamical systems to require some mixing conditions on the dynamic. Here we have two alternative assumptions which are not included one in the other. For Theorem 5 for instance, we need either that the measure be  $\psi$ -mixing, or we require right (or left)  $\phi$ -mixing with polynomially decaying  $\phi$ . The difference between assuming right or left  $\phi$ -mixing is made in order to handle the case of invertible maps (see Remark 5 where, after the proof of Theorem 2, this is explained). Plenty of examples satisfying these assumptions can be found in the literature [9,10]. We will give some examples in Sections 3.7 and 4.
- The assumption (2) of Theorem 5 is necessary in our setting because in general, the  $U_n$ 's may be large unions of cylinders whose measures have to be controlled. It is clear that in the case where  $\Gamma$  is a point, this point is contained in a cylinder  $U_n$  and our mixing assumptions automatically imply that  $\mu(U_n)$  decays exponentially fast and thus satisfies the assumptions.

If  $U_n$  is the outer  $n$ -cylinder approximation of  $\Gamma$ , then  $(U_n)^j = U_j$  for any  $j \leq n$  and the condition simplifies to  $\sum_{j \geq n} \mu(U_j) \rightarrow 0$ .

- Finally, we need that the  $\alpha_k$ 's exist and decay sufficiently fast so that  $\sum_{k=1}^{\infty} k^2 \alpha_k < \infty$ . As we will explain, the existence/computation of the parameters  $\alpha_k, k \geq 1$  is not obvious in general, it is not granted by our mixing assumptions, and can only be, at most, guaranteed case by case.

### 3.2. Interpretation of the compound Poisson distribution

The definition (2) of the Poisson compound distribution is not the most common in the literature. Let us explain that it indeed coincides with the classical definition. Put  $r := \sum_{\ell} t \tilde{\lambda}_{\ell}$  and  $\lambda_{\ell} := t \tilde{\lambda}_{\ell} / r$ . (Proposition 6 below will give conditions under which we have that  $\lambda_{\ell} = \alpha_{\ell} - \alpha_{\ell+1}$ .) With these quantities, we have

$$\varphi_{\tilde{\nu}}(z) = e^{\sum_{k \geq 1} t \tilde{\lambda}_k (e^{z^k} - 1)} = e^{r \sum_{k \geq 1} \lambda_k (e^{z^k} - 1)} = e^{r(\sum_{k \geq 1} \lambda_k e^{z^k} - 1)}.$$

We recognise the moment generating function of the random variable  $Z = \sum_{i=1}^N X_i$  in which  $N \sim \text{Poisson}(r)$  and  $X_i, i \geq 1$  are i.i.d. integer valued r.v.'s with distribution

$$P_X(\ell) = \lambda_{\ell} = \frac{\tilde{\lambda}_{\ell}}{\sum_k \tilde{\lambda}_k}, \quad \ell \geq 1.$$

When  $\tilde{\lambda}_1 = \lambda$  and  $\tilde{\lambda}_k = 0, k \geq 2$ , we obtain the straight Poisson distribution with parameter  $t\lambda$ .

We can now make the relation with our results concerning the count of limiting returns to sets with small measure. The interpretation of the Poisson random variable  $N$  is that it gives the distribution of clusters which occur on a large timescale as suggested by Kac's formula. And the number of returns in each cluster is given by the i.i.d. random variables  $X_j$ 's. These returns are on a fixed timescale and they surely depend on the "shape" of the return sets, as observed in Section 5.

An important non-trivial compound Poisson distribution is the Pólya–Aeppli distribution which happens when the  $X_j$ 's are geometrically distributed with parameter  $1 - p$ , that is  $P_X(k) = (1 - p)p^{k-1}, k \geq 1$ .

For instance, when  $\tilde{\lambda}_{\ell} = (1 - p)^2 p^{\ell-1}$ , the compound Poisson distribution with parameters  $t \tilde{\lambda}_{\ell}, \ell \geq 1$ , is Pólya–Aeppli since  $\lambda_{\ell} = (1 - p)p^{\ell-1}$ . In this particular case we have moreover that  $N \sim \text{Poisson}(t(1 - p))$ . This specific case will be called "Pólya–Aeppli distribution with parameter  $t(1 - p)$ ". This means in explicit form that

$$\tilde{\nu}(\{k\}) = e^{-(1-p)t} \sum_{j=1}^k \binom{k-1}{j-1} \frac{((1-p)^2 t)^j}{j!} p^{k-j}.$$

Several asymptotic distributions will appear along the paper, Pólya–Aeppli or not, depending of the examples (and the setting).

### 3.3. Relation to the extremal index in extreme value theory

Assuming that the  $\alpha_k, k \geq 1$  exist and vanish as  $k$  diverges, we have that  $\sum_{\ell} \tilde{\lambda}_{\ell} = \sum_{\ell} (\alpha_{\ell} - \alpha_{\ell+1})$  telescopes to  $\alpha_1$ . This quantity,  $\alpha_1 := \lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty} \mu_{U_n}(K < \tau_{U_n})$ , is called the *extremal index* and has a particular importance in extreme

value theory [23]. Under some circumstances [5], it is equal to the inverse of the mean cluster size. Indeed, according to Section 3.2, the expected cluster size is given by

$$\sum_{\ell} \ell \frac{\tilde{\lambda}_{\ell}}{\sum_k \tilde{\lambda}_k} = \frac{\sum_{\ell} \ell \tilde{\lambda}_{\ell}}{\alpha_1} = \frac{\sum_{\ell} \alpha_{\ell}}{\alpha_1}.$$

It is explained in [29] (see for instance Theorem 2 and Remark 2 therein or see Proposition 6 below) that, if  $\sum_k k \sum_{\ell \geq k} \alpha_{\ell} < \infty$  (so in particular  $\alpha_k$  exists and vanishes as  $k$  diverges), then  $\sum_{\ell} \alpha_{\ell} = 1$  and we obtain the desired result  $\frac{1}{\alpha_1}$  for the mean size of a cluster.

An important issue however is to know, for given dynamical systems, whether or not the limits appearing in all these quantities actually exist. We will investigate this question in Sections 3.6 and 3.4 on some examples, and in Section 4 for the case of  $g$ -measures.

### 3.4. Example 1: The house of cards process

The house of cards process is a Markov chain on  $\mathcal{A} = \{0, 1, 2, \dots\}$  with transition matrix  $Q$  parametrized by a sequence of  $[0, 1]$  real numbers  $r_i, i \geq 0$ :

$$(12) \quad Q(i, j) = \begin{cases} r_i & \text{if } j = 0, \\ 1 - r_i & \text{if } j = i + 1. \end{cases}$$

It has a stationary version if and only if  $\sum_{i \geq 1} \prod_{j=0}^{i-1} (1 - r_j) < \infty$ , which is the condition ensuring that the expecting distance between two consecutive occurrences of a 0 is finite. In this case, the row vector  $\pi$  satisfying  $\pi Q = \pi$  is

$$\pi(k) = \pi(0) \prod_{i=0}^{k-1} (1 - r_i)$$

where

$$\pi(0) = \frac{1}{1 + \sum_{i \geq 1} \prod_{j=0}^{i-1} (1 - r_j)}.$$

For the stationary version of this Markov chain, we want to study the asymptotic distribution of the number of visits to runs of length  $n$  above a threshold  $l \geq 1$ .

We will use the stochastic process notation involving random variables, but in order to relate to the framework of Section 2, we could let  $\mu$  denote the measure on  $\mathcal{A}^{\mathbb{N}}$  associated to the stationary process. This measure is  $\sigma$ -invariant, where  $\sigma$  is the shift operator  $\sigma : \mathcal{A}^{\mathbb{N}} \circlearrowleft$  defined through  $(\sigma(x))_i = x_{i+1}$  for any  $x = (x_0 x_1 x_2 \dots) \in \mathcal{A}^{\mathbb{N}}$ . We are interested in studying the statistics of visits of this symbolic system in  $U_n = U_n(l) = \bigcap_{i=1}^n \sigma^{-i}[l, +\infty)$  as  $n$  diverges for some fixed  $l \geq 1$ .

Let  $\{X_i\}_{i \geq 0}$  be a stationary House of Cards Markov chain. By the Markov property, successive visits to 0 parse the process into independent blocks. Let us denote

$$T := \inf\{k \geq 1 : X_k = 0\},$$

and for any  $i \geq 0$

$$q_i(k) := \mathbb{P}(T = k | X_0 = i) = r_{i+k} \prod_{j=i}^{i+k-1} (1 - r_j)$$

the probability that the time elapsed until the next 0, starting with  $X_0 = i$ , be equal to  $k$ .

Recalling the definition (11) of  $\alpha_k(L, U_n)$ , we have

$$\alpha_{k+1}(L, U_n) \geq \mathbb{P}(T = k | X_0 \geq n + l) \mathbb{P}(X_i < n, i = 1, \dots, L - k | X_0 = 0),$$

and

$$\alpha_{k+1}(L, U_n) \leq \mathbb{P}(T = k | X_0 \geq n + l) \mathbb{P}(X_i < n, i = 1, \dots, L - k | X_0 = 0) + \mathbb{P}(T < k | X_0 \geq n + l) \mathbb{P}(T \geq n | X_0 = 0).$$

Naturally,  $T$  and  $X_i$  being a.s. finite, we have that, as  $n$  diverges,  $\mathbb{P}(X_i < n, i = 1, \dots, L - k | X_0 = 0)$  converges to 1 and  $\mathbb{P}(T \geq n | X_0 = 0)$  converges to 0. We will prove below that, if  $r_i \rightarrow r_\infty \in (0, 1)$ , then for any  $k \geq 1$  and any  $l \geq 1$

$$(13) \quad \lim_n \mathbb{P}(T = k | X_0 \geq n + l) = r_\infty (1 - r_\infty)^k$$

and therefore

$$\alpha_{k+1} := \lim_L \lim_{n \rightarrow \infty} \alpha_{k+1}(L, U_n) = \lim_{n \rightarrow \infty} \mathbb{P}_\pi(T = k | X_0 \in U_n) = r_\infty (1 - r_\infty)^k$$

exists and decays exponentially fast in  $k$ , which grants Condition (3) of theorem 5.

Moreover, under the assumption  $r_i \rightarrow r_\infty \in (0, 1)$  we have that the Markov chain is Doeblin, and thus automatically exponentially (right)  $\phi$ -mixing [9]. This grants condition (1) of Theorem 5. Moreover, in our case, we have  $U_n^j = U_j = \{X_i \geq l, i = 0, \dots, j - 1\}$ , thus

$$\mu(U_n^j) = \sum_{n \geq l} \pi(n) \mathbb{P}(T \geq j | X_0 = n) = \pi(0) \sum_{n \geq l} \prod_{i=0}^{j+n-1} (1 - r_i)$$

which is summable in  $j$  since for any  $\epsilon > 0$ ,  $r_i \geq r_\infty - \epsilon$  for large enough  $i$ 's, granting Condition (2) of Theorem 5.

Thus if  $r_i \rightarrow r_\infty \in (0, 1)$ , we can apply Theorem 5, which gives us that the number of visits to  $U_n$  is, asymptotically, Pólya–Aeppli distributed with parameter  $t(1 - r_\infty)$ . According to Section 3.3, the corresponding extremal index is  $1/r_\infty$ .

It only remains to prove the convergence (13). Let us compute

$$\begin{aligned} \mathbb{P}(T = k | X_0 \geq n + l) &= \frac{\sum_{i \geq n+l} \mathbb{P}(T = k | X_0 = i) \pi(i)}{\sum_{i \geq n+l} \pi(i)} \\ &= \frac{\sum_{i \geq n+l} r_{i+k} \prod_{j=i}^{i+k-1} (1 - r_j) \pi(i)}{\sum_{i \geq n+l} \pi(i)} \\ &= \frac{\sum_{i \geq n+l} r_{i+k} \prod_{j=i}^{i+k-1} (1 - r_j) \pi(0) \prod_{j=0}^{i-1} (1 - r_j)}{\sum_{i \geq n+l} \pi(0) \prod_{j=0}^{i-1} (1 - r_j)} \\ &= \frac{\sum_{i \geq n+k+l} r_i \prod_{j=0}^{i-1} (1 - r_j)}{\sum_{i \geq n+l} \prod_{j=0}^{i-1} (1 - r_j)}. \end{aligned}$$

By Stolz–Cesàro

$$\begin{aligned} \lim_n \mathbb{P}(T = k | X_0 \geq n + l) &= \lim_n \frac{r_{n+l+k} \prod_{j=0}^{n+l+k-1} (1 - r_j)}{\prod_{j=0}^{n+l-1} (1 - r_j)} \\ &= \lim_n r_{n+l+k} \prod_{j=n+l}^{n+l+k-1} (1 - r_j) = r_\infty (1 - r_\infty)^k \end{aligned}$$

as we said.

### 3.5. Return and entry times

An important task, in order to apply Theorem 5, is to prove that the involved limiting quantities exist and to compute the sequence  $\tilde{\lambda}_k, k \geq 1$  (or, equivalently,  $\lambda_k, k \geq 1$ , see Section 3.2), parameter of the asymptotic compound Poisson distribution. Here we give some alternative ways to prove these facts, by defining other quantities related to  $\tilde{\lambda}_k, k \geq 1$ , which are eventually easier to handle.

Let us define  $\hat{\alpha}_\ell(K, U_n) := \mu_{U_n}(\tau_{U_n}^{\ell-1} \leq K)$  and assume that  $\hat{\alpha}_\ell(K) = \lim_{n \rightarrow \infty} \mu_{U_n}(\tau_{U_n}^{\ell-1} \leq K)$  exist for  $K$  large enough. Since  $\{\tau_{U_n}^{\ell+1} \leq K\} \subset \{\tau_{U_n}^\ell \leq K\}$  we get that  $\hat{\alpha}_\ell(K) \geq \hat{\alpha}_{\ell+1}(K)$  for all  $\ell$  and in particular  $\hat{\alpha}_1(K) = 1$ . By monotonicity the limits  $\hat{\alpha}_\ell = \lim_{K \rightarrow \infty} \hat{\alpha}_\ell(K)$  exist and satisfy  $\hat{\alpha}_1 = 1$  and  $\hat{\alpha}_\ell \geq \hat{\alpha}_{\ell+1}$  for any  $\ell \geq 1$ . Now assume that moreover the limits  $p_i^{(\ell)} = \lim_{n \rightarrow \infty} \mu_{U_n}(\tau_{U_n}^{\ell-1} = i)$  of the conditional size of the level sets of the  $\ell^{\text{th}}$  return time  $\tau_{U_n}^\ell$  exist for  $i \geq 0$  (clearly  $p_i^{(\ell)} = 0$  for  $i \leq \ell - 2$ ). According to Lemma 1 in [29] one has, for  $\ell \geq 2$ ,

$$\hat{\alpha}_\ell = \sum_i p_i^{(\ell)}.$$

We also have, by definition, that  $\alpha_\ell = \hat{\alpha}_\ell - \hat{\alpha}_{\ell+1}$ ,  $\ell \geq 1$ . So the existence of the  $\alpha_\ell$ 's is granted once the  $\hat{\alpha}_\ell$ 's exist, moreover, according to what we just said

$$\alpha_\ell = \sum_i (p_i^{(\ell)} - p_i^{(\ell+1)}), \quad \ell \geq 2.$$

This relation also holds for  $\ell = 1$ . To see this, first recall that

$$\alpha_1 = \lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty} \mu_{U_n}(K < \tau_{U_n})$$

and observe that  $p_0^{(1)} = 1$  and  $p_i^{(1)} = 0, i \geq 1$ . It follows that we can write  $\sum_i (p_i^{(1)} - p_i^{(2)}) = 1 - \sum_i p_i^{(2)} = 1 - \hat{\alpha}_2 = \alpha_1$ .

Finally, we state without proof the following result which was proven in [29], and which gives an important characterization of  $\lambda_\ell$  under some conditions.

**Proposition 6 ([29]).** *Let  $U_n \subset \Omega$  be a nested sequence so that  $\mu(U_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Assume that the limits  $\hat{\alpha}_\ell(L) = \lim_{n \rightarrow \infty} \hat{\alpha}_\ell(L, U_n)$  exist for  $\ell = 1, 2, \dots$  and  $L$  large enough. Assume  $\sum_\ell \ell \hat{\alpha}_\ell < \infty$ , then*

$$\lambda_k = \frac{\alpha_k - \alpha_{k+1}}{\alpha_1}$$

where  $\alpha_k = \hat{\alpha}_k - \hat{\alpha}_{k+1}$ . In particular the limit defining  $\lambda_k$  exists.

### 3.6. Example 2: Regenerative processes

We recall that a stochastic process  $\{X_i\}_{i \geq 0}$  is a regenerative process if there exist random times  $T_1 < T_2 < \dots$  such that the sigma fields  $\sigma(X_{T_n}^{T_{n+1}-1}), n \geq 1$  are i.i.d. and independent of  $\sigma(X_0^{T_1-1})$  (we refer for instance to [6, Chapter VI] and [48]). So the model is completely defined if we specify the distribution of  $X_0^{T_1-1}$ , and  $X_{T_1}^{T_2-1}$ . Here we consider a particular case in which these vectors belong to  $\bigcup_{a \in \mathcal{A}} \bigcup_{k \geq 1} a^k$  where  $a^k$  denotes the vector  $(a, \dots, a)$  of  $k$  times the same symbol  $a$  concatenated, and  $\mathcal{A} \subset \mathbb{N}$ . In other words, the independent blocks are filled up with only one symbol as follows

$$X_0^\infty = \underbrace{X_0 \dots X_0}_{T_1 \text{ times}} \underbrace{X_{T_1} \dots X_{T_1}}_{T_2 - T_1 \text{ times}} \dots \underbrace{X_{T_n} \dots X_{T_n}}_{T_{n+1} - T_n \text{ times}} \dots$$

Specifically, we consider that for any  $a \in \mathcal{A}$  and  $k \geq 1$

$$\mathbb{P}(X_{T_1}^{T_2-1} = a^k) = p(a)q_a(k)$$

where  $\sum_a p(a) = 1$  and, for any  $a \in \mathcal{A}$ ,  $\sum_k q_a(k) = 1$ . A way to interpret the above formula is in a two-steps procedure. First we choose the symbol  $X_{T_1} = a$  independently of everything, with probability  $p(a)$ , and next, we choose the size  $k$  of the block, with probability  $q_a(k)$ . This is a particular instance of Semi-Markov process [16,34]. In particular, it is well-known that there exists a stationary version of the process if and only if the expectation of the blocks is finite, that is

$$v = \sum_{a \in \mathcal{A}} p(a)v_a := \sum_{a \in \mathcal{A}} p(a) \sum_{k \geq 1} kq_a(k) < \infty$$

where  $\nu_a$  is the expectation of the blocks of symbols  $a$ . Another known fact is that, for the process to be stationary, the distribution of  $X_0^{T_1-1}$  must be

$$\mathbb{P}(X_0^{T_1-1} = a^k) = \bar{p}(a)\bar{q}_a(k), \quad a \in \mathcal{A}, k \geq 1$$

where

$$\bar{p}(a) = \frac{p(a)\nu_a}{\nu} \quad \text{and} \quad \bar{q}_a(k) = \frac{\sum_{l \geq k} q_a(l)}{\nu_a}.$$

We want to study the distribution of the number of visits to states larger or equals to  $n$  when  $n$  gets large (we assume that  $\mathcal{A}$  is countably infinite).

As for the House of Cards Markov chain, we can relate to the framework of Section 2 by considering the symbolic measure space  $(\mathcal{A}^{\mathbb{N}}, \mathcal{B}, \mu, \sigma)$ , and this time, we consider the nested sets  $U_n = [n, +\infty)$ ,  $n \geq 1$ .

The regenerative structure was also present in the House of Cards Markov chain, since visits to 0 cut the realisation into independent blocks. However, regenerative processes need not be Markovian. The first step, if we want to use Theorem 5 is to investigate the mixing properties of this model.

**Proposition 7.** *For the regenerative process described above, inequality (4) (right  $\phi$ -mixing) holds for*

$$\phi(k) = 2 \sup_{a \in \mathcal{A}} \sum_{i > k} \bar{q}_a(i).$$

The proof of this proposition is not difficult but since we did not find it in the literature, we do it in Section 7.

An interesting case is the *Smith example*, in which  $q_a(a+1) = \frac{1}{a} = 1 - q_a(1)$ . This example was used by [29] as a case in which Proposition 6 cannot be used, because they proved that  $\hat{\alpha}_k = \frac{1}{2}$  for any  $k$ . In any case, we cannot use Proposition 7 neither, and thus we cannot conclude on the statistics of this model here.

To simplify the presentation, suppose now that  $q_a(l) = q(l)$  for any  $l \geq 1$  independently of  $a$ , in which case  $\bar{q}_a = \bar{q}$  for any  $a$  also. According to Proposition 7, Condition (1) of Theorem 5 is granted if the probability distribution  $\bar{q}$  has first moment.

On the other hand, it is not too complicated to see that  $\alpha_k(L, U_n)$  (see (11)) is close to the probability that the starting block, which is a block of a symbol in  $U_n$  (since we are conditioned on starting in  $U_n$ ), equals  $k$ . Indeed, using a similar reasoning as the one used for the house of cards Markov chain [29, Section 8.2], we get

$$\alpha_k(L, U_n) = \bar{q}(k) + \mathcal{O}(L\mu(U_n))$$

from which it follows that

$$\alpha_k = \bar{q}(k) = \frac{\sum_{l \geq k} q(l)}{\sum_k k q(k)}$$

since in the present case  $q_a(k) = q(k)$  for any  $k$ . So condition (3) of Theorem 5 is granted if we assume that  $\bar{q}$  has finite second moment, that is,  $\sum k^2 \bar{q}(k) < \infty$ . This in turns is granted if  $q$  has third moment, since (we put  $\nu_a = \nu$  for any  $a$ )

$$\sum_k k^2 \bar{q}(k) = \frac{1}{\nu} \sum_k k^2 \sum_{l \geq k} q(l) \leq \frac{1}{\nu} \sum_k \sum_{l \geq k} l^2 q(l) = \frac{1}{\nu} \sum_k k^3 q(k) < \infty.$$

In order to ensure condition (2), we will assume that  $n \sum_{i \geq n} p(i) \rightarrow 0$  as  $n$  diverges, which is the case for instance if the probability distribution  $p(a)$ ,  $a \in \mathcal{A}$  has first moment. Then, since in our case  $U_n^i = U_n$  for any  $i \leq n$ , we have for any  $1 \leq K$  and sufficiently large  $n$

$$\sum_{i=K}^n \mu(U_n^i) \leq n\mu(U_n) = n \sum_{i \geq n} p(i) \leq K \sum_{i \geq K} p(i).$$

Thus taking  $a_k = k \sum_{i \geq k} p(j)$ , the assumption (2) of Theorem 5 also holds.

We conclude that, if  $p(a)$ ,  $a \in \mathcal{A}$  has first moment and  $q(k)$ ,  $k \geq 1$  has third moment, then the number of visits to  $U_n$  is, asymptotically compound Poisson with parameter  $t\tilde{\lambda}_k$ ,  $k \geq 1$  where

$$\tilde{\lambda}_k = \frac{q(k)}{\sum_k kq(k)}.$$

As explained in Section 3.3,  $\alpha_1 = \sum_k \tilde{\lambda}_k$  which in this case gives  $1/\sum_k kq(k)$ . Thus the extremal index is  $1/\alpha_1 = \sum_k kq(k)$  which is the expected block size.

### 3.7. Number of visits around a point

Determining the limiting distribution of the number of visits when the limiting target set is a point and the nested sequence of sets is a sequence of cylinders containing the point is a classical question in the literature of recurrence theory [32]. It is however a nice way to illustrate Theorem 5.

Suppose that the mixing conditions of Theorem 5 are satisfied by the dynamic under consideration. Fix a point  $x \in \Omega$  and for any  $n \geq 1$  consider the  $n$ -cylinder  $A_n(x)$ , that is, the unique atom of  $\mathcal{A}^n$  containing  $x \in \Omega$ . In the notation of the previous section, we let  $U_n = A_n(x)$  (that is  $\Gamma = \{x\}$ ), and we ask what is the asymptotic distribution of visits to  $U_n$ ,  $n \geq 1$ . It is well-known that in this case there is a dichotomy according to whether  $x$  is aperiodic, in which case we have a Poisson distribution, or periodic, in which case we have a Polya–Aeppli distribution instead. Illustrating how to use our results in this simple example will be the opportunity to clarify technical details concerning notation and some involved limiting quantities.

Initially, for any measurable set  $U \subset \Omega$  we write  $\tau(U) = \inf_{y \in U} \tau_U(y)$  for the *period* of  $U$ . In other words,  $U \cap T^{-j}U = \emptyset$  for  $j = 1, \dots, \tau(U) - 1$  and  $U \cap T^{-\tau(U)}U \neq \emptyset$ . The proof of the following lemma is direct and can be found for instance in [29].

**Lemma 8.** *Let  $\mathcal{A}$  be a (finite) generating partition of  $\Omega$ . Then the sequence  $\tau(A_n(x))$ ,  $n = 1, 2, \dots$  is bounded if and only if  $x$  is a periodic point.*

We start with the case where  $x$  is a periodic point, with minimal period  $m$ , say. Let us compute the values  $\lambda_\ell$ . For  $n$  large enough one has  $\tau(A_n(x)) = \tau_\infty = m$  and therefore  $A_n(x) \cap \{\tau_{A_n(x)} = m\} = A_n(x) \cap T^{-m}A_n(x) = A_{n+m}(x)$ . Moreover  $p_i^{(\ell)} = \lim_{n \rightarrow \infty} \mu_{A_n(x)}(\tau_{A_n(x)}^{\ell-1} = i) = 0$  for  $i < m$ .

Assume the limit

$$(14) \quad p := p_m^{(2)} = \lim_{n \rightarrow \infty} \frac{\mu(A_{n+m}(x))}{\mu(A_n(x))}$$

exists, then one also has more generally

$$p_{(\ell-1)m}^{(\ell)} = \lim_{n \rightarrow \infty} \frac{\mu(A_{n+(\ell-1)m}(x))}{\mu(A_n(x))} = p^{\ell-1}.$$

All other values of  $p_i^{(\ell)}$  are zero, that is  $p_i^{(\ell)} = 0$  if  $i \neq (\ell - 1)m$ . Thus  $\hat{\alpha}_\ell = p_{(\ell-1)m}^{(\ell)} = p^{\ell-1}$  and consequently

$$\alpha_\ell = \hat{\alpha}_\ell - \hat{\alpha}_{\ell+1} = (1 - p)p^{\ell-1}$$

which is a geometric distribution and in particular implies that  $\sum_k k^2 \alpha_k < \infty$ , meaning that condition (3) of Theorem 5. Moreover, in the present case, we have  $U_n^i = U_n$  for any  $i \leq n$ , and since our mixing assumptions imply that the measure of cylinders decays exponentially fast in  $n$ , it follows that Assumption (2) of Theorem 5 is automatically granted. So by Theorem 5, we conclude that the random variable  $W$  has Pólya–Aeppli distribution with parameter  $t(1 - p)$  (see Section 3.2).

We now consider the case of a non-periodic point  $x$ . In this case, the increasing sequence  $\tau(A_n(x))$  goes to infinite as  $n \rightarrow \infty$ . Note that  $\mu_{A_n(x)}(\tau_{A_n(x)} \leq K) = 0$  for all  $n$  large enough so that  $K < \tau(A_n(x))$ . Hence  $\hat{\alpha}_2(K) = 0$  for all  $K$  which implies that  $\hat{\alpha}_2 = 0$  and consequently  $\hat{\alpha}_\ell = 0$  for all  $\ell \geq 2$ . Consequently in this case the extremal index is  $\alpha_1 = 1 - \hat{\alpha}_2 = 1$  and  $\alpha_k = 0$ ,  $k \geq 2$  so that  $\tilde{\lambda}_1 = 1$  and  $\tilde{\lambda}_k = 0$ ,  $k \geq 2$  and therefore  $W$  is Poisson( $t$ ) distributed.

### 4. The case of $g$ -measures

Let  $\mathcal{A} = \{1, 2, \dots, M\}$  be a finite alphabet and

$$\Sigma_B = \{x \in \mathcal{A}^{\mathbb{N}} : B_{x_i, x_{i+1}} = 1, \forall i \geq 1\} \subset \Sigma := \mathcal{A}^{\mathbb{N}}$$

where  $B$  is an aperiodic and irreducible  $M \times M$  matrix of 0's and 1's. Let  $\mathcal{F}$  denote the Borel  $\sigma$ -algebra of  $\Sigma_B$ . For any finite string  $a_1^n$  (shorthand notation for  $(a_1, \dots, a_n)$ ,  $a_i \in \mathcal{A}$ ) of symbols of  $\mathcal{A}$ , we let  $[a_1^n] := \{x \in \Sigma_B : x_i = a_i, i = 1, \dots, n\}$  denote the corresponding cylinder set. The Borel  $\sigma$ -algebra  $\mathcal{F}$  is generated by the cylinder sets. The shift operator  $\sigma : \Sigma_B \rightarrow \Sigma_B$  defined through  $(\sigma(x))_i = x_{i+1}$  for any  $x = (x_1 x_2 x_3 \dots) \in \Sigma_B$  is called sub-shift of finite type.

A measurable function  $g : \Sigma_B \rightarrow [0, 1]$  satisfying

$$(15) \quad \sum_{y:\sigma(y)=x} g(y) = \sum_{a \in \mathcal{A}} g(ax) = 1$$

for any  $x \in \Sigma_B$  is called a  $g$ -function. Let  $\mathcal{L}_g$  be the associated transfer operator given by

$$\mathcal{L}_g f(x) = \sum_{y:\sigma(y)=x} g(y) f(y) = \sum_{a \in \mathcal{A}} g(ax) f(ax),$$

for functions  $f : \Sigma_B \rightarrow \mathbb{R}$ , where  $ax$  is the concatenation of the symbol  $a$  with the sequence  $x$  (if admissible). A  $g$ -measure is a probability measure satisfying  $\mathcal{L}_g^* \mu = \mu$  [35] where  $\mathcal{L}_g^*$  is the dual of  $\mathcal{L}_g$ . This is equivalent [40] to  $\mu$  being  $\sigma$ -invariant and satisfying

$$\mathbb{E}_\mu(\mathbb{1}_{[a]} | \mathcal{F}_2^\infty)(x) = g(a\sigma(x)),$$

for any  $a \in \mathcal{A}$  and  $\mu$ -almost every  $x \in \Sigma_B$ . Yet another equivalent way is to define  $\mu$  is through the variational principle,

$$\mu \in \operatorname{argmax} \left\{ h_\nu + \int \log g \, d\nu : \nu \text{ is } \sigma\text{-invariant} \right\}$$

where  $h_\nu$  denotes the Kolmogorov–Sinai entropy. The maximum of the quantity above is called the topological pressure of  $\log g$  and denoted  $P(\log g)$ . It turns out that, since  $\sum_{y:\sigma(y)=x} g(y) = 1$ , we have  $P(\log g) = 0$ .

All the above can be stated in the framework of *equilibrium states* for a real function  $\varphi$  on  $\Sigma_B$ . This can be done simply by substituting  $g$  by  $e^\varphi$ , except for the restriction (15) which is put in the  $g$ -measure context to give a stochastic process flavour. We refer to [40,52] for the proofs of all the above equivalences and further details on the variational principle for generic potentials and  $g$ -functions.

An important characterisation of the regularity of  $g$  is its variation of order  $k \geq 1$

$$(16) \quad \operatorname{var}_k g := \sup \{ |g(x) - g(y)| : x_1^k = y_1^k \}.$$

The convergence  $\operatorname{var}_k g \rightarrow 0$  is equivalent to uniform continuity in the product topology. In this case,  $\mathcal{L}_g^* \mu = \mu$  has at least one solution [35]. Under the stronger assumption that  $\sum_k \operatorname{var}_k g < \infty$  and  $g > 0$ , there is a unique  $g$ -measure specified by  $g$  [40] and it enjoys  $\psi$ -mixing [52, see the proof of Theorem 3.2 therein].

#### 4.1. Visits close to a periodic point

In Section 3.7, we considered the case of visits around a point through cylinders. Concerning periodic points of minimal period  $m$ , the existence of the limit

$$(17) \quad p := p_m^{(2)} = \lim_{n \rightarrow \infty} \frac{\mu(A_{n+m}(x))}{\mu(A_n(x))}$$

was assumed in order to conclude the asymptotic distribution of the number of visits close to the point. Here we consider this question in the case of  $g$ -measures. We have the following proposition.

**Proposition 9.** *Consider a  $g$ -measure  $\mu$  and a point  $x \in \Sigma_B$  of prime period  $m \geq 1$ . If  $\prod_{i=0}^{m-1} g \circ \sigma^i$  is continuous at  $x$ , then the limiting parameter  $p$  defined through (17) exists and is given by  $\prod_{i=0}^{m-1} g(\sigma^i(x))$ . So in particular, if  $\lim_{k \rightarrow 0} \operatorname{var}_k g = 0$ , the limiting parameter exists for any periodic point. If moreover  $\sum_k \operatorname{var}_k g < \infty$  and  $g > 0$ , the limiting distribution of the number of visits around  $x$  has Pólya–Aepli distribution with parameter  $t(1 - p)$ .*

**Remark 1.** Let us mention that the existence of the limit was proven and computed for Axiom A by [46].

Let us now consider a specific class of  $g$ -measures, called *renewal measures*. Consider the space  $\Sigma = \{0, 1\}^{\mathbb{N}}$  and for any  $x \in \Sigma$  let  $\kappa(x) := \inf\{n \geq 0 : x_{n+1} = 1\}$  count the number of 0 until the first occurrence of a 1 in  $x$ . Now take a sequence of  $[0, 1]$ -valued real numbers  $q_i, i \geq 0$  and define the function  $\bar{g}$  by  $\bar{g}(1x) = q_{\kappa(x)} = 1 - \bar{g}(0x)$ . A  $g$ -measure corresponding to  $\bar{g}$  exists under some technical assumptions on the sequence  $q_i, i \geq 1$ , which are automatically granted if we assume that  $q_i \geq \epsilon$  for any  $i \geq 1$ .

**Proposition 10.** Consider a renewal measure with sequence of parameters  $\{q_i\}_{i \geq 1}$  which satisfies  $q_i \in [\epsilon, 1 - \epsilon]$  for some  $\epsilon > 0$ . Then, for any periodic point  $x \neq 0^\infty$ , the limit  $p$  defined by (14) exists and the limiting distribution of the number of visits around  $x$  has Pólya–Aeppli distribution with parameter  $t(1 - p)$ .

The same occurs for visits around  $0^\infty$  if, and only if,  $q_i, i \geq 1$  converge.

The main interest of this example lays in the fact that it is  $\phi$ -mixing.

#### 4.2. Existence of the extremal index

Due to its relation to the so-called extremal index, the question of (non-)existence of the limit (17) was investigated recently in [4]. According to Proposition 9, vanishing variation guarantees existence of the extremal index at any periodic point. So if we want to characterise systems for which the extremal index does not exist, we have to get out of the classical setting of  $g$ -measures in which  $g$  is assumed uniformly continuous. This is what we discuss now.

First, let us observe that Theorem 3.2 of [3] completely solved the question of the existence (and computation) of the limiting parameters in the case of renewal measures. Something interesting which is shown therein is that the renewal measure provides a simple situation in which  $p$  does not exist although the measure enjoys good mixing. This is the case if we take  $q_{\kappa(x)} = \epsilon_1$  if  $\kappa(x)$  is odd and  $\bar{g}(x) = \epsilon_2$  otherwise: the limit  $\lim \frac{\bar{\mu}(\{0^{n+1}\})}{\mu(\{0^n\})}$  does not exist (see also Proposition 10 above). With this choice of parameters, the measure is  $\phi$ -mixing with exponentially decaying rate  $\phi$ , but we easily see that  $\bar{g}$  has a discontinuity (with respect to the product topology) at the point  $0^\infty$ .

The first information we get from this example is that good mixing properties are not enough to ensure existence of the extremal index, and that this existence is perhaps related to the continuity properties of the  $g$ -function. Technically, the discontinuity of  $\bar{g}$  at  $0^\infty$  is an *essential discontinuity*, borrowing the terminology used in the context of statistical physics [19]. This is a discontinuity which cannot be removed by changing function  $\bar{g}$  on a null  $\bar{\mu}$ -measure subset of  $\Sigma$  [20].

More generally, for a  $g$ -measure  $\mu$  with  $g$ -function  $g$ , the non-existence of the limit  $\lim_n \mu([a_1^n]) / \mu([a_1^{n-1}])$  implies that  $g$  has an essential discontinuity at  $y_1^\infty$ , but the converse is not necessarily true. So even when we focus on the easier case of points of period 1, the non-existence of the extremal index implies on an essential discontinuity of  $g$  at  $a^\infty$ , but the converse is not true in general. This non-equivalence is spectacularly clear with the following example.

#### 4.3. The Furstenberg example

As far as we know, the following example is due to [25] (see Chapter 3.12 therein). On  $\mathcal{A}^{\mathbb{N}} = \{-1, +1\}^{\mathbb{N}}$ , take the product measure  $\mu$  with marginal  $\mu([+1]) = \epsilon = 1 - \mu([-1])$ . Next, consider the function  $\Pi : \{-1, +1\}^{\mathbb{N}} \rightarrow \{-1, +1\}^{\mathbb{N}}$  defined through  $(\Pi(x))_i = x_i x_{i+1}$ , the product of two consecutive coordinates. If  $\epsilon = 1/2$ , then  $\nu := \mu \circ \Pi^{-1}$  is also the product measure with marginal  $\nu([+1]) = \nu([-1]) = 1/2$ , but otherwise,  $\nu$  has a  $g$ -function which is essentially discontinuous everywhere [20,51]. Let us now write down its  $g$ -function  $g^*$ . Since our interest here is with respect to discontinuities, we will assume hereafter that  $\epsilon \neq 1/2$ .

For any fixed  $j \in \mathbb{N}$ , when the limit exists, let

$$d(x_j^\infty) = \lim_{n \rightarrow \infty} \frac{\#\{j \leq i \leq n, x_i = -1\}}{n}$$

denote the asymptotic density of  $-1$  in the sequence  $x_j^\infty$  with  $x_i \in \{-1, +1\}, i \geq j$ . For any fixed  $y \in \{-1, +1\}^{\mathbb{N}}$  the preimage set  $\Pi^{-1}(y)$  contains two elements, that we denote by  $x^+(y)$  for the one starting by  $+1$  and  $x^-(y)$  for the one starting by  $-1$ . Now, let

$$G := \{y \in \{-1, +1\}^{\mathbb{N}} : d(x^+(y)) = \epsilon \text{ or } d(x^-(y)) = \epsilon\}.$$

Then by the law of large numbers for the product measure  $\mu$ , we have that  $\nu(G) = 1$ . It is proved in [20] that for any  $y \in G$ ,

$$(18) \quad \frac{\nu([1y_2^n])}{\nu([y_2^n])} \rightarrow \begin{cases} \epsilon & \text{if } d(x^-(y)) = \epsilon, \\ 1 - \epsilon & \text{if } d(x^+(y)) = \epsilon. \end{cases}$$

This defines  $g^*(1\sigma(y))$  for  $y \in G$  by the Martingale Theorem. For  $y \in G^c$ ,  $g^*(1\sigma(y))$  may be defined arbitrarily, as this set has null  $\nu$ -measure and the choice will not affect the conclusion of everywhere essential discontinuity.

The following simple result proves that, despite of the terrible (dis)continuity properties of  $g^*$ , the limiting quantities needed to apply our theorems exist.

**Proposition 11.** *Consider the measure  $\nu$  with  $g$ -function  $g^*$  as defined above. Then, for any periodic point  $x$  of prime period  $m \geq 1$ , the limiting quantity  $p$  defined by (14) exists and the limiting distribution of the number of visits around  $x$  has Pólya–Aeppli distribution with parameter  $t(1 - p)$ .*

To conclude on this example, we observe that the value  $p$  is explicitly computed in the proof of the proposition (see (35)).

#### 4.4. Temporal synchronisation for $g$ -measures

Consider  $m$   $g$ -measures  $\mu_1, \dots, \mu_m$  on  $\Sigma_B$ , respectively with functions  $g_1, \dots, g_m$ , and define the product measure  $\mu_1 \otimes \dots \otimes \mu_m$  on  $\Omega = \Sigma_B^m$  or even  $\Sigma^m$ . Let  $\hat{\sigma} : \Omega \circlearrowleft$  be the shift map on the product space. For any  $n \geq 1$ ,  $S_n = \bigcup_{A \in \mathcal{A}^n} A^m$  is the  $n$ -cylinder neighbourhood of the diagonal  $\Delta = \{(x, \dots, x) : x \in \Sigma_B\} \subset \Sigma_B^m$  (see Figure 2 for a picture with  $m = 2$ ). Observe that

$$\hat{\sigma}^{-i} S_n = \{(x^{(1)}, \dots, x^{(m)}) : x_i^{(1)} \dots x_{n+i-1}^{(1)} = x_i^{(j)} \dots x_{n+i-1}^{(j)}, j = 2, \dots, m\},$$

that is, a visit in  $S_n$  can be interpreted as a synchronisation lasting  $n$  time units of the symbols of the dynamical systems, therefore justifying the name “temporal synchronisation”.

A natural first problem is, in the above “uncoupled”, or “non-interacting” setting of product measures, to study the distribution of the number of visits to  $S_n$  as  $n$  diverges, that is, visits to longer and longer synchronised pieces of orbits. However, it would be even more interesting to study the same question for interacting  $g$ -measures. In full generality, for any  $m \geq 2$ , any  $g$ -measure on the product space  $\Omega = (\mathcal{A}^m)^\mathbb{N}$  can be considered the coupling of the  $m$  coordinates  $g$ -measures. That is, we see  $\Omega$  as  $\Sigma^m$  instead of seeing it as  $(\mathcal{A}^m)^\mathbb{N}$ . In other words, in this general setting, we are studying the synchronisation of the coordinates.

Theorem 12 below is stated with this abstract approach because it is more general, and next, Corollary 13 will specialise to the non-interacting case. Finally, we will rapidly discuss two explicit ways to make  $g$ -measures interact.

Let  $W_n$  count the number of synchronisations on the Kac scaling

$$W_n := \sum_{i=0}^{t/\hat{\mu}(S_n)} \mathbb{1}_{S_n} \circ \hat{\sigma}^i.$$

**Theorem 12.** *As before let  $B$  be irreducible and aperiodic. Then*

- (1) *On the product space  $\Omega = \Sigma^m$ , for some  $m \geq 2$ , let  $\hat{\mu}$  be a  $\hat{\sigma}$ -invariant  $\hat{g}$ -measure. Assume that  $\hat{g} > 0$  has summable variation.*

*Then we have that*

$$p := \lim_n \frac{\hat{\mu}(S_n \cap \hat{\sigma}^{-1} S_n)}{\hat{\mu}(S_n)}$$

*exists and  $\hat{\mu}(W_n \in \cdot)$  converges to a Pólya–Aeppli distribution with parameter  $t(1 - p)$ .*

- (2) *On the product space  $\Omega = \Sigma_B^m$ , for some  $m \geq 2$ , let  $\hat{\mu}$  be a  $\hat{\sigma}$ -invariant  $\hat{g}$ -measure. Assume that the function  $g^\Delta : \Sigma_B \rightarrow \mathbb{R}$  has exponentially decaying variations, where  $g^\Delta(x) := \hat{g}(x, x, \dots, x)$ .*

*Then,  $\hat{\mu}(W_n \in \cdot)$  converges to a Pólya–Aeppli distribution with parameter  $t(1 - p)$  where  $p = e^{P(\log g^\Delta)} < 1$  with  $P(\log g^\Delta)$  being the topological pressure of  $\log g^\Delta$ .*

**Remark 2.** In the particular case in which  $\hat{g}$  only depends on the two first coordinates,  $\hat{\mu}$  is a Markov chain with matrix  $\hat{Q}(\hat{a}, \hat{b}) := \hat{g}(\hat{x})$  for any  $\hat{x} \in \Sigma_B^m$  such that  $\hat{x}_1 = \hat{b} \in \mathcal{A}^m$  and  $\hat{x}_2 = \hat{a} \in \mathcal{A}^m$ . In this case,  $P(\log g^\Delta)$  is  $\log \rho$  where  $\rho$  is the largest positive eigenvalue of the matrix  $Q^\Delta : \mathcal{A} \times \mathcal{A} \rightarrow (0, 1)$  defined by  $Q^\Delta(a, b) = \hat{Q}((a, \dots, a), (b, \dots, b))$ .

We have the following direct corollary of item (2) of the preceding theorem.

**Corollary 13.** *On the subshift space  $\Sigma_B$ , consider  $m \geq 2$  independent  $g$ -measures  $\mu_i, i = 1, \dots, m$ , with  $g$ -functions  $g^{(i)}$  satisfying  $g^{(i)} > 0$  and having exponentially vanishing variation.*

*Then,  $\hat{\mu}(W_n \in \cdot)$  converges to a Pólya–Aeppli distribution with parameter  $t(1 - p)$  where  $p = e^{P(\sum_i \log g^{(i)})}$ , with  $P(\sum_i \log g^{(i)})$  being the topological pressure of  $\sum_{i=1}^m \log g^{(i)}$ .*

The proof of this corollary is direct as, in the uncoupled case,  $\hat{g}(x_1, x_2, \dots, x_m) = \prod_{i=1}^m g^{(i)}(x_i)$  and therefore  $\hat{g}$  inherits the regularity and mixing properties of the  $g^{(i)}$ 's.

**Remark 3.** The Rényi entropy of order  $q \in \mathbb{R}$  is defined as the limit

$$\mathcal{R}_\mu(q) = - \lim_{n \rightarrow \infty} \frac{1}{qn} \log \sum_{[x_1^n]} \mu([x_1^n])^{q+1}$$

when it exists. Proposition 7 in [2] states that it exists and equals  $-\frac{P((1+q)\log g)}{q}$  as long as  $g$  is continuous. So in the case of the synchronization of  $m$  independent copies of the same  $g$ -measure, Corollary 13 states that the parameter of the Pólya–Aeppli asymptotic distribution is  $-(m - 1)\mathcal{R}_\mu(m - 1)$ .

**Remark 4.** In the particular case in which  $g^{(i)}, i = 1, \dots, m$  only depend on the two first coordinates,  $\mu^{(i)}, i = 1, \dots, m$  are Markov chains with matrices  $Q^{(i)}(a, b) := g^{(i)}(x)$  for any  $x$  such that  $x_1 = b$  and  $x_2 = a$ . In this case,  $P(\sum_i \log g^{(i)})$  is  $\log \rho$  where  $\rho$  is the largest positive eigenvalue of the matrix  $Q^\Delta$  defined through  $Q^\Delta(a, b) = \prod_{i=1}^m Q^{(i)}(a, b)$ .

As an example, consider  $m = 2$  independent Markov measures with  $\mathcal{A} = \{0, 1\}$  and take  $Q^{(1)}(0, 0) = 0.2, Q^{(1)}(1, 1) = 0.7, Q^{(2)}(0, 0) = 0.8, Q^{(2)}(1, 1) = 0.9$ . Then, we have  $Q^\Delta(0, 0) = 0.16 = Q^\Delta(0, 1), Q^\Delta(1, 0) = 0.03, Q^\Delta(1, 1) = 0.63$ , and in particular  $p = \frac{16}{25}$ .

Theorem 12 is abstract because it is not stated in terms of the interaction of (possibly distinct) given  $g$ -measures. The natural question now is *how to make  $g$ -measures interact?* A direct application of the *coupled map lattice approach* used for instance in [18,29] (see also Section 5.1 below) does not seem to make much sense in the setting of  $g$ -measures. An observation at this point is that we prefer to use the terminology “interacting  $g$ -measures” instead of “coupled  $g$ -measures”, because the second one has a precise definition in stochastic processes, which does not necessarily correspond to what we want here.

We consider two ways. The first way to make  $g$ -measures interact is through a coupling of their  $g$ -functions, coupling in the sense of stochastic processes as we now explain. Suppose we have  $m \geq 2$  possibly distinct  $g$ -functions  $g^{(1)}, \dots, g^{(m)}$  on  $\Sigma$ , and use the notation  $\vec{x}_k = (x_k(1), \dots, x_k(m)) \in \mathcal{A}^m, k \geq 1$ , and  $\vec{x} = (\vec{x}_1 \vec{x}_2 \dots)$ . Then, a  $g$ -function  $\hat{g}$  on  $\Sigma^m$  is said to be a coupling  $g$ -function of the  $g^{(k)}$ 's if, for any  $k = 1, \dots, m$ , any  $a \in \mathcal{A}$  and any  $(\vec{x}_2 \vec{x}_3 \dots)$

$$\sum_{\vec{x}_1 : x_1(k) = a} \hat{g}(\vec{x}) = g^{(k)}(ax_2(k)x_3(k) \dots).$$

The  $g$ -measure  $\hat{\mu}$  associated to  $\hat{g}$  is then automatically a coupling of the  $g$ -measures  $\mu^{(k)}$  associated to  $g^{(k)}, k = 1, \dots, m$ , in the sense that the  $k^{\text{th}}$  marginal of  $\hat{\mu}$  is equally distributed to  $\mu^{(k)}$  for any  $k$ . An example is given in Section 4.4.1.

A second way to make  $g$ -measures interact is, given  $m \geq 2$  possibly distinct  $g$ -functions  $g^{(1)}, \dots, g^{(m)}$ , to construct a  $g$ -function on the product space  $\Sigma^m$  parametrized by a tuning parameter  $\gamma \in [0, 1]$  indicating the strength of the interaction, that is, having the property that, when  $\gamma = 0, \hat{g} = \prod_{i=1}^m g^{(i)}$ , which corresponds to the non-interacting case. The resulting  $\hat{g}$  needs not to be a coupling of the  $g^{(k)}$ 's in the stochastic process meaning. We give a simple example in Section 4.4.2.

In order to simplify the presentation, we will restrict ourselves to the case where the  $g^{(i)}$ 's depend only on the two first coordinates and will construct  $\hat{g}$ 's having the same property. This means that instead of specifying  $g$ -functions  $g$ , we will specify matrices  $Q$ . Moreover, we assume that the  $Q^{(i)}$ 's and  $\hat{Q}$  have only strictly positive entry, ensuring that we are in force of all the conditions of Theorem 12, item (2).

Observe finally that, also according to Theorem 12, we only have to define the coupling  $\hat{g}$  on  $\Delta$ , which means that we only have to define  $Q^\Delta$ .

#### 4.4.1. Example 1: The maximal coupling

The *maximal coupling* is a classical coupling in the theory of stochastic processes, but is less known in dynamical systems. We refer to [11] for a complete definition, here we only define the coupling on  $\Delta$ , which is sufficient for our purposes. We start with matrices  $Q^{(i)}, i = 1, \dots, m$ , then the maximal coupling is defined on the diagonal as

$$\hat{Q}_{\max}((a, \dots, a), (b(1), \dots, b(m))) := \inf\{Q^{(i)}(a, b(i)), i = 1, \dots, m\}.$$

So, taking  $m = 2$  and coming back to the matrices used in Remark 4, we obtain  $Q_{\max}^{\Delta}(0, 0) = 0.2, Q_{\max}^{\Delta}(0, 1) = 0.2, Q_{\max}^{\Delta}(1, 0) = 0.1, Q_{\max}^{\Delta}(1, 1) = 0.7$  and therefore, according to Theorem 12 and Remark 2 we have asymptotically Polya–Aeppli distributed synchronizations with parameter  $p_{\max} = \rho_{\max}$ , the largest eigenvalue of  $Q_{\max}^{\Delta}$ . This gives  $p_{\max} = (\frac{9+\sqrt{33}}{20})$ , strictly larger than the value  $p = 16/25$  obtained in the uncoupled case considered in Remark 4.

The terminology *maximal* comes from the fact that  $\hat{Q}_{\max}$  puts as much probability as possible on the diagonal, that is, as much probability of agreement as possible in one step, still keeping the marginals equal to  $Q^{(i)}, i = 1, \dots, m$ . So it is natural that the synchronizations last longer than in the uncoupled case, and this is what  $p_{\max} > p$  means.

#### 4.4.2. Example 2: Parametrized coupling

Let  $Q^{(1)}$  and  $Q^{(2)}$  be two stochastic matrices on  $\mathcal{A} = \{0, 1\}$  and define transition probabilities  $q^{(1)}, q^{(2)} : \mathcal{A}^2 \times \mathcal{A} \rightarrow [0, 1]$  through

$$\begin{aligned} q^{(1)}((a(1), a(2)), 1) &= (1 - \gamma)Q^{(1)}(a(1), 1) + \gamma a(2), \\ q^{(2)}((a(1), a(2)), 1) &= (1 - \gamma)Q^{(2)}(a(2), 1) + \gamma a(1). \end{aligned}$$

Naturally, we put  $q^{(i)}((a(1), a(2)), 0) = 1 - q^{(i)}((a(1), a(2)), 1)$ .

Now, define

$$\hat{Q}_{\gamma}((a(1), a(2)), (b(1), b(2))) = q^{(1)}((a(1), a(2)), b(1))q^{(2)}((a(1), a(2)), b(2)).$$

Observe that  $\hat{Q}_{\gamma}$  is indeed a stochastic matrix on  $\mathcal{A}^2$ , and that when  $\gamma = 0$ , we get  $\hat{Q}_0((a(1), a(2)), (b(1), b(2))) = Q^{(1)}(a(1), b(1))Q^{(2)}(a(2), b(2))$  as we wanted.

As an example, consider the matrices used in Remark 4. According to Theorem 12 and Remark 4, we have asymptotically Polya–Aeppli with parameter  $p_{\gamma} = e^{P(\log Q_{\gamma}^{\Delta})}$  which equals the largest eigenvalue of  $Q_{\gamma}^{\Delta}$ . This yields

$$p_{\gamma} = \frac{1}{200} (79 + 2\gamma + 19\gamma^2 + \sqrt{2401 + 7996\gamma + 3006\gamma^2 - 7604\gamma^3 + 4201\gamma^4}).$$

So we notice that the interacting parameter  $\gamma$  modifies in a non-trivial way the parameter of the asymptotic distribution. Actually, in the present case, synchronisation increases as the parameter  $\gamma$  increases. (Observe that when  $\gamma = 0$ , we retrieve  $16/25$ , as in the non-interacting case, which is natural.)

The original inspiration here is as a toy model for two interacting neurons, in which the value 1 means that the neuron is spiking, and 0 means it is resting. We assume that each neuron, when they don't interact ( $\gamma = 0$ ), has a spiking dynamic given by the  $Q^{(i)}, i = 1, 2$ . When they interact ( $\gamma > 0$ ), the probability that a neuron spikes will depend not only on whether or not it just spiked, but also on whether or not the other neuron just spiked, this is what  $q^{(i)}$  models. More precisely, it models the effect of excitatory neurons, since the probability of spiking for one neuron increases when the other neuron has spiked:  $q^{(1)}((a(1), 1), 1) > Q^{(1)}(a(1), 1)$ . Next, given the past, we assume that the probability of spiking for each neuron is independent, this is why  $\hat{Q}_{\gamma}$  is the product  $q^{(1)}q^{(2)}$ .

Naturally, this model is very simple and does not represent correctly the complexity of a system of interacting neurons, however, it retains some features of a recent model introduced by [26]. Their model is not markovian, and include several physiological considerations, and it is a natural and interesting problem to analyse the synchronisation properties in their setting, using Theorem 12.

### 5. Discussion on synchronisation

The present paper is mainly concerned with the symbolic setting of deterministic dynamical systems. In the present section we make a small digression to discuss the difference, with regard to asymptotic synchronisation, between three situations: (1) the *geometric* approach of deterministic dynamical systems, (2) the *symbolic* approach of deterministic dynamical systems, and (3) Markov chain, a particular case of random dynamical systems.

We start this section by making a rapid overview of what is known in the geometric setting. The first application of recurrence type argument, like those used in this paper, to synchronisation was given in Section 4 of the article [37], where the authors explicitly computed a first order formula for the leading eigenvalue of the perturbed transfer operator, the perturbation being the small neighborhood around the diagonal. It was successively shown by [36] that such a perturbative formula was intimately related to the extremal index. This spectral approach to extreme value theory was developed in [18], which showed that the probability of the appearance of synchronization in chaotic coupled map lattices was related to the distribution of the maximum of a certain observable evaluated along almost all orbits. The statistics of the number of visits was proven in [29], with a technique different from the spectral approach: we recall it in the next subsection and then, by considering a very simple example in Section 5.2 we will exhibit clearly how different can be this approach from the symbolic approach developed in Section 4.4. We remind that an alternate probabilistic approach in a coupled maps setting is proposed in [12]. We conclude with Section 5.3, considering the case of continuous state Markov chains, a third situation, in which yet another behaviour is displayed.

5.1. Synchronisation of (un)coupled map lattices: Geometric approach

We shall consider coupled map lattices over uniformly expanding interval maps. Let  $T$  be a piecewise continuous map on the unit interval  $I = [0, 1]$  which is uniformly expanding, i.e. satisfies  $\inf |DT| > 1$ . We also assume that  $T^{-1}$  has only finitely many branches. Then we define the coupled map  $\hat{T}$  on  $\Omega = I^m$ , for some integer  $m \geq 2$  by

$$(19) \quad \hat{T}(\vec{x})_i = (1 - \gamma)T(x_i) + \gamma \sum_{j=1}^m M_{i,j}T(x_j) \quad \forall i = 1, 2, \dots, m,$$

for  $\vec{x} \in \Omega$ , where  $M$  is an  $m \times m$  stochastic matrix and  $\gamma \in [0, 1]$  is a coupling constant. The uncoupled case corresponds to  $\gamma = 0$  in which case  $\hat{T}$  is the product of  $m$  copies of  $T$ . For  $\nu > 0$  small, we put

$$(20) \quad S_\nu := \{ \vec{x} \in [0, 1]^m : |x_i - x_j| \leq \nu, \forall i, j \}$$

for a tubular neighbourhood of the diagonal  $\Gamma$  (see Figure 1 for a picture with  $m = 2$ ). Then we define as before  $\hat{\alpha}_{k+1} = \lim_{K \rightarrow \infty} \lim_{\nu \rightarrow 0} \hat{\alpha}_{k+1}(K, S_\nu)$  for the parameters of the limiting compound Poisson distribution which describes the synchronisation effect in the neighbourhood of the diagonal  $\Gamma$ .

It has been previously shown by [29] that if  $T$  is a piece-wise uniformly expanding map of the unit interval with finitely many branches satisfying a mild geometric condition along the diagonal and if  $\mu$  is an equilibrium state for a sufficiently regular potential function on  $\Omega$  then the compound Poisson parameters are given by

$$(21) \quad \hat{\alpha}_{k+1} = \frac{1}{(1 - \gamma)^{k(m-1)} \int_I h((x)^m) dx} \int_I \frac{h((x)^m)}{|DT^k(x)|^{m-1}} dx$$

where  $h : \Omega \rightarrow \mathbb{R}^+$  is the density function of  $\mu$  and  $(x)^m$  denotes the set of points on the diagonal.

Obtaining explicit results is still a complicated problem for coupled map lattices. So we now turn our attention to the case of *uncoupled* map lattices, that is, to the case where  $\gamma = 0$ . Such a situation was first investigated by [17]. They proved that, for an absolutely continuous measure of a piecewise expanding and smooth map of the circle, the asymptotic distribution of synchronisation is compound Poisson, and identified the limiting parameters  $\hat{\alpha}_k, k \geq 1$ .

So let us see how (21) looks like in the uncoupled case (we consider here the case  $m = 2$  to simplify) in the setting of interval transformations. Consider a partition  $\mathcal{A} = \{I_1, \dots, I_M\}$  of  $I := [0, 1)$  and a piecewise linear Markov transformation  $T$  which is continuous, monotone and uniformly expanding on each of the sub intervals  $I_i$ , that is  $\inf_{I_i} |T'_i| > 1$  for any  $i = 1, \dots, M$ , where  $T_i = T$  on  $I_i$ . Define the stochastic  $M \times M$  matrix  $Q$  by

$$Q_{i,j} = \begin{cases} 0 & \text{if } I_j \cap T(I_i) = \emptyset, \\ \frac{1}{|T'_i|} & \text{if } I_j \subset T(I_i). \end{cases}$$

We know in this case that the invariant density  $h$  which satisfies  $\mathcal{L}h = h$ , where  $\mathcal{L}$  is the transfer operator, is piecewise constant. Thus put  $h_i = h(x)$  for  $x \in I_i$  and consider the row vector  $\vec{h} = (h_1, \dots, h_M)$ . Then  $hQ = h$ .

According to [18] and [29], we then have to compute

$$\hat{\alpha}_{k+1} = \frac{\int_I \frac{h^2(x)}{|DT^k(x)|} dx}{\int_I h^2(x) dx}.$$

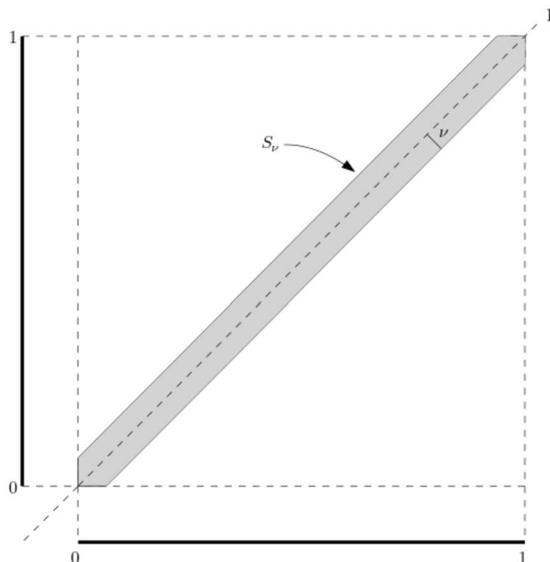


Fig. 1. A strip  $S_\nu$  around the diagonal  $\Gamma$  in  $[0, 1]^2$ .

For any  $(a_1, \dots, a_k) \in \{1, \dots, M\}^k$ , we use the notation  $x \in I_{(a_1, \dots, a_k)}$  for  $T^{i-1}x \in I_{a_i}$ ,  $i = 1, \dots, k$  and let

$$\mathcal{A}^k := \{(a_1, \dots, a_k) \in \{1, \dots, M\}^k : Q_{a_i, a_{i+1}} > 0, i = 1, \dots, k - 1\}.$$

If  $(a_1, \dots, a_k) \in \mathcal{A}^k$  then, using the chain rule,

$$|DT^k(x)| = \left| \prod_{i=1}^k DT(T^{i-1}x) \right| = \prod_{i=1}^k |DT_{a_i}|, \quad \forall x \in I_{(a_1, \dots, a_k)}$$

while on the other hand

$$\lambda(I_{(a_1, \dots, a_k)}) = \frac{1}{\prod_{i=1}^k |DT_{a_i}|}.$$

We can now compute

$$(22) \quad \int \frac{h^2(x)}{|DT^k(x)|} dx = \sum_{(a_1, \dots, a_k) \in \mathcal{A}^k} h_{a_1}^2 \frac{\lambda(I_{(a_1, \dots, a_k)})}{\prod_{i=1}^k |DT_{a_i}|}$$

and

$$\int_I h^2(x) dx = \sum_i h_i^2 \lambda(I_i).$$

For instance, consider the case of  $T(x) = 3x \bmod 1$ . Then we have  $h_i = 1$  and  $|DT_i| = 3$  for  $i = 1, 2, 3$ , and thus  $\int \frac{h^2(x)}{|DT^k(x)|} dx = 3^{-k}$  while  $\int_I h^2(x) dx = 1$ . We therefore obtain  $\hat{\alpha}_{k+1} = 3^{-k}$ , which means that the random variable  $W$  has Pólya–Aeppli distribution with parameter  $t(1 - 1/3)$  (see Section 3.2). A natural conjecture (having also in view Theorem 12) would be that this is the case for any uncoupled (sufficiently mixing) dynamical systems. However, as we show below, this is not necessarily the case.

### 5.2. Geometric vs. symbolic: An explicit example

As is probably clear by now, we call geometric approach when we measure the synchronisation through visits to  $S_\nu$ ,  $\nu > 0$  (see (20)), thin strips around the diagonal, as pictured in Figure 1. On the other hand, the symbolic approach, when we measure synchronisation through visits to cylinder sets covering the diagonal, is pictured in Figure 2.

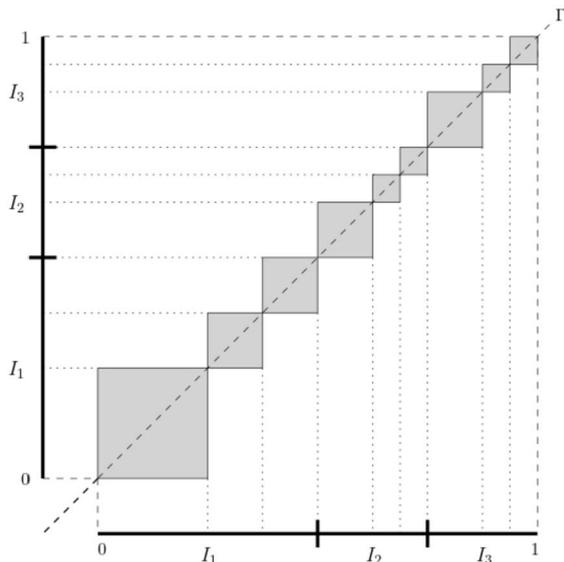


Fig. 2. The set  $S_2 = \bigcup_{A \in \mathcal{A}^2} (A \times A)$  where the partition  $\mathcal{A} = \{[0, \frac{1}{2}), [\frac{1}{2}, \frac{3}{4}), [\frac{3}{4}, 1)\}$ .

Here we show that both approaches yield different classes of distributions in general (although both compound Poisson), by mean of a simple example. Consider the following map  $T$  on the unit interval  $I$ :

$$(23) \quad T(x) = \begin{cases} 3x & \text{if } x \in I_1 := [0, 1/3), \\ 5/3 - 2x & \text{if } x \in I_2 := [1/3, 2/3), \\ -2 + 3x & \text{if } x \in I_3 := [2/3, 1). \end{cases}$$

It is a piecewise linear Markov transformation with

$$Q = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

5.2.1. Geometric approach

We know [18,29] that the asymptotic distribution of synchronisation is Poisson compound. We now calculate explicitly the parameters  $\hat{\alpha}_{k+1}$ ,  $k \geq 1$  using (22) which, in matricial form, gives

$$\int \frac{h^2(x)}{|DT^k(x)|} dx = \sum_{(a_1, \dots, a_k) \in \mathcal{A}^k} h_{a_1}^2 \frac{\lambda(I_{(a_1, \dots, a_k)})}{\prod_{i=1}^k |DT_{a_i}|} = \sum_{(a_1, \dots, a_k)} h_{a_1}^2 \prod_{i=1}^k Q_{a_i, a_{i+1}}^2 = \text{Trace}(v^h \hat{Q}^k)$$

where  $v^h = (h_1^2 \dots h_n^2)$  and  $\hat{Q}$  is the matrix with entries  $\hat{Q}_{i,j} = Q_{i,j}^2$ . The corresponding piecewise constant density with respect to Lebesgue is the row vector  $\vec{h} = (3/56/56/5)$ , and we then obtain  $v^h = (9/2536/2536/25)$ . With this example we get (using Mathematica online) that  $\text{Trace}(v^h \hat{Q}^k)$  equals

$$\frac{2^{-3k-1} 3^{2-2k} [(3\sqrt{145} - 23)(17 - \sqrt{145})^k + (23 + 3\sqrt{145})(17 + \sqrt{145})^k]}{25\sqrt{145}}$$

and

$$\int_I h^2(x) dx = \sum_i h_i^2 \lambda(I_i) = 27/25.$$

This yields

$$\hat{\alpha}_{k+1} = \frac{2^{-3k-1}3^{2-2k}[(3\sqrt{145} - 23)(17 - \sqrt{145})^k + (23 + 3\sqrt{145})(17 + \sqrt{145})^k]}{27\sqrt{145}}$$

which does not correspond to a geometric distribution and we do not have Pólya–Aeppli asymptotic distribution of synchronisation in a strip around the diagonal.

### 5.2.2. Symbolic approach

We can use Theorem 12 doing  $g(x) =: Q_{i,j}$  for any  $x$  such that  $x_1 = i, x_2 = j$  (i.e.  $g(x)$  depends only on the first two coordinates  $x_1, x_2$ ). According to Remark 4, we conclude that synchronisation distribution for cylinder neighbourhoods of the diagonal converges to Pólya–Aeppli with the parameter  $t(1 - \rho)$ , where  $\rho$ , the largest positive eigenvalue of  $\hat{Q}$ , equals  $\frac{1}{72}(17 + \sqrt{145})$  for this specific example.

### 5.3. Synchronisation of Markov chains

Let us in this section consider the quintessential random dynamical systems which are Markov chains. In fact a *random transformation* can in a simple way lead to a Markov chain in the following way (see for instance [7]). Take  $\{\omega_k\}_{k \in \mathbb{N}_0}$  a sequence of i.i.d. random variables with values in some  $\Omega$  that carries the probability measure  $\theta$ . We associate to each  $\omega \in \Omega$  a map  $T_\omega$  and the iteration of the unperturbed map  $T^n(x)$ , will be replaced by the composition of random maps  $T_{\omega_{n-1}} \circ \dots \circ T_{\omega_0}$ . These random transformations generate a stationary Markov chain  $\{Z_n\}_{n \geq 0}$  with transition probabilities, for any  $n \geq 1$ :

$$\mathbb{P}(Z_{n+1} \in A | Z_n = x) = \int_{\Omega} \mathbf{1}_A(T_\omega(x)) d\theta(\omega),$$

where  $A$  is a measurable set in some  $M$  and  $T_\omega$  is a map from  $M$  to  $M$ .

We therefore consider such a Markov chain  $\{X_n\}_{n \geq 0}$ , stationary, with continuous state space  $I \subset \mathbb{R}$  and transition probabilities  $p(x, A) = \mathbb{P}(X_{n+1} \in A | X_n = x)$  for measurable sets  $A$ . If the transition kernel has a density  $p(x, y)$ , that is if  $p(x, dy) = p(x, y) dy$ , then

$$\mathbb{P}(X_{n+1} \in A | X_n = x) = \int_A p(x, y) dy.$$

The invariant measure  $\mathbb{P}$  is then given by the transition probabilities  $p$  and an initial probability measure  $\rho$ . That is  $\mathbb{P}(X_0 \in A) = \int_A d\rho(y)$  and

$$d\mathbb{P}(x_0, x_1, x_2, \dots, x_n) = d\rho(x_0)p(x_0, dx_1)p(x_1, dx_2) \dots p(x_{n-1}, dx_n).$$

The probability measure  $\rho$  on  $I$  is invariant under the map  $T$  which is given by  $T\mu(A) = \int_I p(x, A) d\mu(x)$ , for all  $A$  measurable, and is the annealed invariant measure on  $I$ . The Markov chain satisfies the *Doeblin condition* if there exists a probability measure  $\nu$  and an  $\eta \in (0, 1)$  so that  $p(x, A) \geq (1 - \eta)\nu(A)$  for all measurable  $A$ . If  $p$  satisfies this condition then in the total variation norm  $\|p^n(x, \cdot) - \rho(\cdot)\|_{TV} \leq 2\eta^n$  uniformly in  $x$ .

We can now associate to  $\{X_n\}_{n \geq 0}$  another independent copy  $\{Y_n\}_{n \geq 0}$  and ask the distribution of the first synchronisation time of the two chains. The Markov chain  $(X_n, Y_n) \in I \times I, n \in \mathbb{N}_0$ , has transition probabilities  $p_2((x, y), (A, B)) = p(x, A)p(y, B), A, B$  measurable. Let us denote by  $\hat{T}$  the product map on  $I \times I$  and by  $\mathbb{P}_2 = \mathbb{P} \otimes \mathbb{P}$  its invariant probability measure. By using the procedure of Section 5.1, we are led to consider the direct product of these two chains and look at the couples of points which stay close to each other up to time  $n - 1$ . Let us therefore set  $\Delta_\delta := \{(x, y) \in I_2 = I \times I, |x - y| \leq \delta\}$  for a neighbourhood of the diagonal in  $I \times I$ .

Specifically, we will show that for Markov chains whose densities are bounded above and away from 0, the limiting distribution (as  $\delta \rightarrow 0$ ) of return times to  $\Delta_\delta$  is Poissonian, which means that orbits don't cluster over time and that there is no synchronisation effect.

In order to get a limiting compound Poisson distribution as  $\delta \rightarrow 0$  we want to use Theorem 3 of [29]. For that purpose let us put  $\mathfrak{X}_i = \mathbf{1}_{\Delta_\delta} \circ \hat{T}^i, i = 0, 1, 2, \dots$ , and  $\mathfrak{W}_a^b = \sum_{i=a}^b \mathfrak{X}_i$ . For simplicity we put  $\mathfrak{W} = \mathfrak{W}_0^N$ , where  $N = t/\mathbb{P}_2(\Delta_\delta)$  (take integer part) where  $t > 0$  is a parameter. We cut the time interval  $N$  into blocks of length  $2K + 1$  for some large  $K$  ( $K \ll N$ ) and put  $\mathfrak{Z} = \sum_{i=0}^{2K} \mathfrak{X}_i$ . If we put  $N' = N/(2K + 1)$  (assuming it being an integer) then  $\mathfrak{W} = \sum_{n=0}^{N'-1} \mathfrak{Z} \circ \hat{T}^{n(2K+1)}$ . We

now choose a gap  $\gamma \ll N'$  ( $\gamma \geq 2$ ) and want to estimate the quantities

$$\begin{aligned} \mathcal{R}_1 &= \sup_{\substack{0 < \gamma < M \leq N' \\ 0 < q < N' - \gamma - 1/2}} \left| \sum_{u=1}^{q-1} (\mathbb{P}_2(\mathfrak{Z} = u \wedge \mathfrak{W}_{\gamma(2K+1)}^{M(2K+1)} = q - u)) \right. \\ &\quad \left. - \mathbb{P}_2(\mathfrak{Z} = u) \mathbb{P}_2(\mathfrak{W}_{\gamma(2K+1)}^{M(2K+1)} = q - u) \right| \\ \mathcal{R}_2 &= \sum_{j=1}^{\gamma} \mathbb{P}_2(\mathfrak{Z} \geq 1 \wedge \mathfrak{Z} \circ \hat{T}^{(2K+1)j} \geq 1). \end{aligned}$$

If we denote by  $\tilde{\nu}$  the compound binomial distribution measure where the binomial part has values  $\mathfrak{p} = \mathbb{P}_2(\mathfrak{Z} \geq 1)$  and  $N' = N/(2K + 1)$  and the compound part has probabilities  $\lambda_\ell(K, \delta) = \mathbb{P}_2(\mathfrak{Z} = \ell)/\mathfrak{p}$ , then, by Theorem 3 of [29], there exists a constant  $c_1$ , independent of  $K$  and  $\gamma$ , such that

$$|\mathbb{P}_2(\mathfrak{W} = k) - \tilde{\nu}(\{k\})| \leq c_1 2(N'(\mathcal{R}_1 + \mathcal{R}_2) + \gamma \mathbb{P}_2(\mathfrak{X}_0 = 1)).$$

(Please note that Theorem 3 in the paper from 2020 which we use here contains a typo in the lower summation limit of  $j$  in the expression for  $\mathcal{R}_2$ : As we put it here the lower summation limit must be  $j = 1$  and not  $j = 2$  as printed there.) The proof of the following lemma is given at the end of the present section.

**Lemma 14.** *For any  $K' < K$  there exists a constant  $c_2$  so that*

$$N'(\mathcal{R}_1 + \mathcal{R}_2) \leq c_2 K \gamma \mathbb{P}_2(\Delta_\delta) + \eta^{K'} + \frac{K'}{K}.$$

If we put  $\gamma = \mathbb{P}_2(\Delta_\delta)^{-\beta}$  for some  $\beta \in (0, 1)$  then

$$\mathbb{P}_2(\mathfrak{W} = k) \longrightarrow \nu^K(\{k\}) + \mathcal{O}(\eta^{K'} + K'/K)$$

as  $\delta \rightarrow 0$ , where  $\nu^K$  is the compound Poisson distribution with parameters  $t\lambda_\ell(K)$  with

$$\lambda_\ell(K) = \lim_{\delta \rightarrow 0} \frac{\mathbb{P}_2(\mathfrak{Z} \geq 1)}{\mathbb{P}_2(\Delta_\delta)}$$

assuming the limits exist. Now put e.g.  $K' = \sqrt{K}$  and let  $K$  go to infinity. Then  $\nu^K$  converges to a compound Poisson distribution  $\nu$  with parameters  $t\lambda_\ell$ , where  $\lambda_\ell = \lim_{K \rightarrow \infty} \lambda_\ell(K)$  assuming the limits exist. Thus

$$\mathbb{P}_2(\mathfrak{W} = k) \longrightarrow \nu(\{k\}).$$

What we showed above is that, if the transition probabilities are given by a density  $p(x, y)$  satisfying the Doeblin condition (that is, if we assume that  $p$  is also bounded away from 0), then  $\mathfrak{W}$  converges in distribution to a compound Poisson distribution  $\nu$ . In order to show that  $\nu$  is in fact a straight Poisson distribution we want to show that  $\hat{\alpha}_2 = 0$ , which implies by monotonicity that  $\hat{\alpha}_\ell = 0$  for all  $\ell \geq 2$ . To that end, we shall further assume that the transition probabilities  $p(x, y)$  are bounded above by some constant  $\mathcal{K}$ . If  $p(x, y) \leq \mathcal{K}$  for all  $x, y$ , we also get  $p^n(x, y) \leq \mathcal{K}$  for all  $x, y$  and thus

$$\begin{aligned} \mathbb{P}_2(\Delta_\delta \cap \hat{T}^{-n} \Delta_\delta) &= \int \mathbb{1}_{\Delta_\delta}(x_0, y_0) \mathbb{1}_{\Delta_\delta}(x_n, y_n) d\rho(x_0) d\rho(y_0) p^n(x_0, x_n) p^n(y_0, y_n) dx_n dy_n \\ &\leq \mathcal{K} \mathbb{P}_2(\Delta_\delta) \mathfrak{m}(\Delta_\delta), \end{aligned}$$

where  $\mathfrak{m}$  is the Lebesgue measure on  $I \times I$ . Thus

$$\hat{\alpha}_2(K, \delta) = \frac{\mathbb{P}_2(\Delta_\delta \cap \bigcup_{n=1}^{2K} \hat{T}^{-n} \Delta_\delta)}{\mathbb{P}_2(\Delta_\delta)} \leq 2K \mathcal{K} \mathfrak{m}(\Delta_\delta) \longrightarrow 0$$

as  $\delta \rightarrow 0$  (since  $\mathfrak{m}(\Delta_\delta) \leq 2\delta$ ) for all  $K$ . Thus  $\hat{\alpha}_\ell = 0$  for all  $\ell \geq 2$  which implies that  $\nu$  is Poisson with parameter  $t$ .

Let us remark that the double limit first  $\delta \rightarrow 0$  and then  $K \rightarrow \infty$  can be synchronised by going along a sequence  $K_\delta \rightarrow \infty$  as  $\delta$  approaches 0 in such a way that  $K_\delta \mathbb{P}_2(\Delta_\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ . This was shown in [53, Proposition 6.2].

We conclude the section with the proof of the lemma.

**Proof of Lemma 14.** In order to estimate  $\mathcal{R}_1$  and the terms of  $\mathcal{R}_2$  for  $j \geq 2$  we use the Doeblin condition and the consequential exponential convergence to the initial distribution to obtain:

$$\begin{aligned} \mathbb{P}_2(\mathfrak{Z} \geq 1, \mathfrak{Z} \circ \hat{T}^{j(2K+1)} \geq 1) &= \int \mathbb{1}_{\mathfrak{Z} \geq 1}(z_0, \dots, z_{2K}) \mathbb{1}_{\mathfrak{Z} \geq 1}(z'_0, \dots, z'_{2K}) \\ &\quad d\rho_2(z_0) p_2(z_0, dz_1) \cdots p_2(z_{2K-1}, dz_{2K}) \\ &\quad p_2^{(j-1)(2K+1)}(z_{2K}, dz'_0) p_2(z'_0, dz'_1) \cdots p_2(z'_{2K-1}, dz'_{2K}) \\ &\leq \mathbb{P}_2(\mathfrak{Z} \geq 1) (\mathbb{P}_2(\mathfrak{Z} \geq 1) + 2\eta^{(j-1)(2K+1)}), \end{aligned}$$

where  $z_i = (x_i, y_i)$  and  $\rho_2 = \rho \times \rho$ . Consequently

$$\mathcal{R}_1 \leq \sum_{j=\gamma}^{\infty} \mathbb{P}_2(\mathfrak{Z} \geq 1) \eta^{(j-1)(2K+1)} \leq \mathbb{P}_2(\mathfrak{Z} \geq 1) \eta^{(\gamma-1)(2K+1)}.$$

For the estimate of  $\mathcal{R}_2$  we consider the case  $j = 1$  separately, choose  $K' < K$  and put  $\mathfrak{Z}' = \sum_{i=2K-K'}^{2K} \mathfrak{X}_i$ ,  $\mathfrak{Z}'' = \mathfrak{Z} - \mathfrak{Z}' = \sum_{i=0}^{2K-K'-1} \mathfrak{X}_i$ . Then, similarly as above, we obtain

$$\mathbb{P}_2(\mathfrak{Z}'' \geq 1, \mathfrak{Z} \circ \hat{T}^{2K+1} \geq 1) \leq \mathbb{P}_2(\mathfrak{Z}'' \geq 1) (\mathbb{P}_2(\mathfrak{Z} \geq 1) + 2\eta^{K'}).$$

Since

$$\mathbb{P}_2(\mathfrak{Z} \geq 1, \mathfrak{Z} \circ \hat{T}^{2K+1} \geq 1) \leq \mathbb{P}_2(\mathfrak{Z}'' \geq 1, \mathfrak{Z} \circ \hat{T}^{2K+1} \geq 1) + \mathbb{P}_2(\mathfrak{Z}' \geq 1)$$

we obtain

$$\mathcal{R}_2 \leq \mathbb{P}_2(\mathfrak{Z}'' \geq 1) (\mathbb{P}_2(\mathfrak{Z} \geq 1) + 2\eta^{K'}) + \mathbb{P}_2(\mathfrak{Z}' \geq 1) + 2 \sum_{j=2}^{\gamma} \mathbb{P}_2(\mathfrak{Z} \geq 1) \eta^{(j-1)(2K+1)}.$$

For the final estimate we use that  $N' = \frac{N}{2K+1} \leq \frac{N}{K} = \frac{t}{K\mathbb{P}_2(\Delta_\delta)}$  and  $\mathbb{P}_2(\mathfrak{Z} \geq 1) \leq (2K+1)\mathbb{P}_2(\Delta_\delta)$ ,  $\mathbb{P}_2(\mathfrak{Z}' \geq 1) \leq K'\mathbb{P}_2(\Delta_\delta)$ . Then as  $K' < K$

$$\begin{aligned} N'(\mathcal{R}_1 + \mathcal{R}_2) &\lesssim \frac{t}{K\mathbb{P}_2(\Delta_\delta)} (3K^2\mathbb{P}_2(\Delta_\delta)^2 + K\mathbb{P}_2(\Delta_\delta)\eta^{(\gamma-1)K} + 2K\mathbb{P}_2(\Delta_\delta)\eta^{K'} + K'\mathbb{P}_2(\Delta_\delta)) \\ &\lesssim K\mathbb{P}_2(\Delta_\delta) + \eta^{(\gamma-1)K} + \eta^{K'} + \frac{K'}{K} \end{aligned}$$

which implies the statement as  $K' < (\gamma - 1)K$ . □

## 6. Proofs of general theorems

Before we come to the proofs of Theorems 2, 4 and 5 we start this section with an important subsection which describes rapidly the classical Stein–Chen method, used to prove Theorem 2.

### 6.1. Stein–Chen method

We will use the Stein–Chen method as described in [47] to estimate how close a given probability measure  $\nu$  is to a compound Poisson distribution  $\tilde{\nu}$  for parameters  $t\tilde{\lambda}_\ell$ ,  $\ell = 1, 2, \dots$ , which satisfy  $\sum_\ell \tilde{\lambda}_\ell < \infty$ . On the space  $\mathcal{F} = \{f : \mathbb{N}_0 \rightarrow \mathbb{R}\}$  of functions on the non-negative integers, the Stein operator  $\mathcal{S} : \mathcal{F} \rightarrow \mathcal{F}$  is defined by

$$\mathcal{S}g(k) = kg(k) - \sum_{\ell=1}^{\infty} t\ell\tilde{\lambda}_\ell g(k + \ell).$$

For a given set  $E \subset \mathbb{N}_0$  one wants to find  $f$  so that

$$\mathfrak{S}f = \mathbb{1}_E - \tilde{\nu}(E)$$

where  $\tilde{\nu}$  is the compound Poisson distribution with parameters  $t\tilde{\lambda}_\ell$ . This last identity is the Stein equation and Theorem 3 of [8] guarantees that such a solution function  $f$  exists. Moreover Proposition 1 therein also states that, for given  $E \subset \mathbb{N}_0$  it satisfies  $f(k) \lesssim \frac{1}{k}$ .

Indeed if  $\int \mathfrak{S}f d\tilde{\nu} = 0$  for all bounded functions  $f$  on  $\mathbb{N}$ , then

$$\begin{aligned} 0 &= \sum_k \mathfrak{S}f(k) \tilde{\nu}(\{k\}) \\ &= \sum_k \left( kf(k) - \sum_{\ell=1}^{\infty} t\ell\tilde{\lambda}_\ell f(k+\ell) \right) \tilde{\nu}(\{k\}) \\ &= \sum_k f(k) \left( k\tilde{\nu}(\{k\}) - \sum_{\ell=1}^k t\ell\tilde{\lambda}_\ell \tilde{\nu}(\{k-\ell\}) \right) \end{aligned}$$

implies

$$k\tilde{\nu}(\{k\}) = \sum_{\ell=1}^k t\ell\tilde{\lambda}_\ell \tilde{\nu}(\{k-\ell\})$$

for every  $k$ . From this we conclude that  $\tilde{\nu}$  has the generating function

$$\varphi_{\tilde{\nu}}(z) = \exp \sum_{\ell} t\tilde{\lambda}_\ell (e^{z\ell} - 1)$$

which equals  $\exp \int_0^\infty (e^{zx} - 1) d\rho(x) = \exp(\varphi_\rho(z) - L)$  where  $\rho = \sum_{\ell} t\tilde{\lambda}_\ell \delta_\ell$ ,  $L = \sum_{\ell} t\tilde{\lambda}_\ell$  and  $\varphi_\rho(z) = \sum_{\ell} t\tilde{\lambda}_\ell e^{z\ell}$  is the generating function for the measure  $\rho$ . This implies that  $\tilde{\nu}$  is compound Poisson with parameters  $t\tilde{\lambda}_\ell$ .

### 6.2. Proof of Theorem 2

We are now ready to prove our first main result.

**Proof of Theorem 2.** Let  $\tilde{\nu}$  be the compound Poisson distribution for  $t\tilde{\lambda}_\ell$ ,  $\ell \in \mathbb{N}$  as defined in the statement of the theorem. For  $E \subset \mathbb{N}_0$  let again  $f$  be the solution of the Stein equation  $\mathfrak{S}f = \mathbb{1}_E - \tilde{\nu}(E)$ . Then for a probability measure  $\nu$  on  $\mathbb{N}_0$  we have

$$\begin{aligned} \nu(E) - \tilde{\nu}(E) &= \int \mathbb{1}_E d\nu - \int \mathbb{1}_E d\tilde{\nu} \\ &= \int (\mathbb{1}_E - \tilde{\nu}(E)) d\nu \\ &= \int \mathfrak{S}f d\nu \\ &= \int \left( kf(k) - \sum_{\ell=1}^{\infty} t\ell\tilde{\lambda}_\ell f(k+\ell) \right) d\nu(k). \end{aligned}$$

For some fixed set  $U \subset \Omega$  and  $t > 0$ , recall the definition (1) of the random variable  $W$  and put  $\nu(\cdot) = \mu(W \in \cdot)$ . We have

$$(24) \quad \mu(W \in E) - \tilde{\nu}(E) = \int \mathfrak{S}f d\nu = \mathbb{E}[Wf(W)] - \sum_{\ell=1}^{\infty} t\ell\tilde{\lambda}_\ell \mathbb{E}[f(W+\ell)].$$

Now let  $K$  and  $\Delta$  be (later taken to be large) numbers so that  $K \ll \Delta \ll t/\mu(U)$  and let us establish the following notation (with the obvious restrictions  $i - j \geq 0$  and  $i + j \leq t/\mu(U)$ ), where we used  $I_i := \mathbb{1}_U \circ T^i$ :

- (i) Close range interactions:  $Z_i = \sum_{j=-K}^K I_{i+j}$ . (The purpose of the double sided sum centred at  $i$  is to cover both cases, when  $T$  is non-invertible as well as the case when  $T$  is invertible.) Observe that was denoted  $Z_i^{(K)}$  in the beginning of the paper but we will omit the upper script to avoid overloaded notation.
- (ii) The gap terms

$$V_i^- = \sum_{j=K+1}^{K+\Delta} I_{i-j}, \quad V_i^+ = \sum_{j=K+1}^{K+\Delta} I_{i+j}$$

and  $V_i = V_i^- + V_i^+$  for the entire gap.

- (iii) The principal terms:

$$Y_i^- = \sum_{j>K+\Delta} I_{i-j}, \quad Y_i^+ = \sum_{j>K+\Delta} I_{i+j},$$

and  $Y_i = Y_i^- + Y_i^+$  for the entire principle term. We reforce that these summands are restricted by  $i - j \geq 0$  and  $i + j \leq t/\mu(U)$ , and are therefore not infinite.

In this way we decomposed  $W$  as  $W = Z_i + V_i + Y_i$  for every  $i = 1, \dots, N$ .

Now observe that, by translation invariance for any  $i \in (K, N - K)$

$$t\ell\tilde{\lambda}_\ell = t\mathbb{E}(\mathbb{1}_{Z_i=\ell} | I_i = 1) = \frac{t}{\mu(U)} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) = \sum_{i=1}^{t/\mu(U)} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}).$$

Thus

$$t\ell\tilde{\lambda}_\ell \mathbb{E}f(W + \ell) = \sum_{i=0}^{t/\mu(U)} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}f(W + \ell).$$

On the other hand we can naturally write

$$Wf(W) = \sum_i I_i f(W).$$

With this in hand, coming back to (24) we have

$$\begin{aligned} \mu(W \in E) - \tilde{v}(E) &= \sum_{i=0}^{t/\mu(U)} \left( \mathbb{E}[I_i f(W)] - \sum_{\ell} \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell}] \mathbb{E}[f(W + \ell)] \right) \\ &= \sum_{i=0}^{t/\mu(U)} \left( \sum_{\ell} \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} f(W)] - \sum_{\ell} \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell}] \mathbb{E}[f(W + \ell)] \right) \\ &= \sum_{i=0}^{t/\mu(U)} \left( \sum_{\ell} \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} f(Y_i + V_i + \ell)] - \sum_{\ell} \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell}] \mathbb{E}[f(W + \ell)] \right). \end{aligned}$$

We now split the error term on the right hand side into three parts as follows:

$$\begin{aligned} &\sum_{i=0}^{t/\mu(U)} \sum_{\ell} (\mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} f(Y_i + V_i + \ell)] - \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} f(Y_i + \ell)]) \\ &+ \sum_{i=0}^{t/\mu(U)} \sum_{\ell} (\mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} f(Y_i + \ell)] - \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell}] \mathbb{E}[f(Y_i + \ell)]) \\ &+ \sum_{i=0}^{t/\mu(U)} \sum_{\ell} (\mathbb{E}[I_i \mathbb{1}_{Z_i=\ell}] \mathbb{E}[f(Y_i + \ell)] - \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell}] \mathbb{E}[f(W + \ell)]) \\ &= A + B + C. \end{aligned}$$

We now proceed to show that each of the three terms can be upper bounded in order to give the bound stated by Theorem 2.

(i) For the first term we write  $A = \sum_{i,\ell} A_{i,\ell}$  where

$$(25) \quad \begin{aligned} |A_{i,\ell}| &= |\mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} (f(Y_i + V_i + \ell) - f(Y_i + \ell))]| \\ &\leq \|f'\| \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} V_i] \end{aligned}$$

where  $\|f'\| := \sup_k |f(k) - f(k - 1)|$ . Note that

$$\begin{aligned} \sum_{\ell=1}^{2K+1} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} V_i^+) &= \mathbb{E}(I_i \mathbb{1}_{Z_i \geq 1} V_i^+) \\ &\leq \mathbb{E}(I_i V_i^+) \\ &\leq \sum_{j=K+1}^{2n} \mu(U^{j/2} \cap T^{-j}U) + \sum_{j=2n+1}^{K+\Delta} \mu(U \cap T^{-j}U) \end{aligned}$$

where we recall that  $U^\ell = A_\ell(U) = \bigcup_{A \in \mathcal{A}^\ell, A \cap U \neq \emptyset} A$  is the outer  $\ell$ -approximation of  $U$  ( $\ell \leq n$ ). Therefore by the right  $\phi$ -mixing property (see (3))

$$(26) \quad \begin{aligned} \sum_{\ell} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} V_i^+) &\leq \mu(U) \left( \sum_{j=K+1}^{2n} [\mu(U^{j/2}) + \phi(j/2)] + \sum_{j=2n+1}^{K+\Delta} [\mu(U) + \phi(j-n)] \right) \\ &\leq \mu(U) \left( \sum_{j=K/2}^n \mu(U^j) + \Delta\mu(U) + \sum_{j=K/2}^{\infty} \phi(j) \right). \end{aligned}$$

One can also show that

$$\sum_{\ell} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} V_i^-) = \sum_{\ell} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} V_i^+).$$

Putting the above together we obtain ( $\phi^1$  is the tail sum of  $\phi$ )

$$\begin{aligned} |A| &\leq \sum_{i=0}^{t/\mu(U)} c_1 \|f'\| \mu(U) \left[ \Delta\mu(U) + \phi^1(K/2) + \sum_{j=K/2}^n \mu(U^j) \right] \\ &= c_2 \|f'\| t \left[ \Delta\mu(U) + \phi^1(K/2) + \sum_{j=K/2}^n \mu(U^j) \right]. \end{aligned}$$

On the other hand, we have  $\|f'\| = \mathcal{O}(1)$  since by Theorem 4 in [8]

$$|f(Y_i + V_i + \ell) - f(Y_i + \ell)| \leq c_3 \frac{1}{Y_i + \ell} = \mathcal{O}(1).$$

So

$$(27) \quad |A| \leq c_2 c_3 t \left[ \Delta\mu(U) + \phi^1(K/2) + \sum_{j=K/2}^n \mu(U^j) \right].$$

(ii) We now estimate  $B$ . We get

$$B = \sum_i \sum_{\ell} \sum_{a=0}^{\infty} f(a + \ell) (\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i=a}) - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i=a}))$$

$$= \sum_{i,\ell} \sum_{a=0}^N f(a+\ell) \sum_{a^-+a^+=a} (\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+}) - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+})),$$

as  $a \leq N$ . We want to sort the terms by their sign so that every level we have only two terms to which we can apply the mixing property. Put

$$\epsilon_{a^-,a^+} = \epsilon_{a^-,a^+}(i, \ell) = \text{sgn}(\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+}) - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+})),$$

and then  $|B_{i,\ell}| = B_{i,\ell}^- + B_{i,\ell}^+$ , where

$$B_{i,\ell}^+ = \sum_a |f(a+\ell)| \sum_{\substack{a^-+a^+=a \\ \epsilon_{a^-,a^+}=+1}} (\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+}) - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+}))$$

and

$$B_{i,\ell}^- = - \sum_a |f(a+\ell)| \sum_{\substack{a^-+a^+=a \\ \epsilon_{a^-,a^+}=-1}} (\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+}) - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a^+})).$$

Let us begin estimating  $B_{i,\ell}^+$ , the case of  $B_{i,\ell}^-$  will be done below. We partition the sum over  $a$  into segments of exponential progression. For that purpose let us put, using  $N = 1/\mu(U)$

$$g_m(\ell) = \max_{N2^{-m} \leq a < N2^{-m+1}} |f(a+\ell)|$$

which by Proposition 1 of [8] satisfies  $g_m \leq c_4 2^m/N$  for some constant  $c_4$ . We now get

$$\begin{aligned} B_{i,\ell}^+ &\leq \sum_{m=0}^{\lg N} g_m \sum_{a=2^{-m}N}^{2^{-m+1}N-1} \sum_{\substack{a^-=0 \\ \epsilon_{a^-,a^+}=+1}}^a (\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a-a^-}) \\ &\quad - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{Y_i^+=a-a^-})) \\ &\leq \sum_{m=0}^{\lg N} c_4 \frac{2^m}{N} \sum_{a^-=0}^{2^{-m+1}N-1} (\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)})) \end{aligned}$$

where the set

$$\begin{aligned} \mathcal{Y}_{i,m}^+(a^-) &= \bigcup_{\substack{2^{-m}N-a^- \leq a^+ < 2^{-m+1}N-a^- \\ \epsilon_{a^-,a^+}=+1}} \{Y_i^+ = a^+\} \\ &= \{2^{-m}N - a^- \leq Y_i^+ < 2^{-m+1}N - a^- : \epsilon_{a^-,a^+} = +1\} \end{aligned}$$

cuts out slices of exponential progression. By the mixing property

$$\begin{aligned} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) &= \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-}) \mathbb{P}(\mathcal{Y}_{i,m}^+(a^-)) \\ &\quad + \mathcal{O}^*(\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-}) \phi(\Delta - n)) \\ (28) \quad &= \mathbb{P}(Y_i^- = a^-) \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{P}(\mathcal{Y}_{i,m}^+(a^-)) \\ &\quad + \mathcal{O}^*(\mathbb{P}(Y_i^- = a^-) \mathbb{P}(\mathcal{Y}_{i,m}^+(a^-)) \phi(\Delta - n)) \\ &\quad + \mathcal{O}^*(\mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-}) \phi(\Delta - n)) \end{aligned}$$

where the symbol  $\mathcal{O}^*$  indicates an error term where the implied constant is 1 (i.e. if  $G = \mathcal{O}^*(\varepsilon)$  then  $|G| \leq \varepsilon$ ). Also

$$\mathbb{P}(Y_i^- = a^-) \mathbb{P}(\mathcal{Y}_{i,m}^+(a^-)) = \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) + \mathcal{O}^*(\mathbb{P}(Y_i^- = a^-) \phi(2\Delta - n)).$$

Thus, for  $m = 0, 1, 2, \dots, \lg N$ ,

$$\begin{aligned} & \sum_{a^-=0}^{2^{-m+1}N-1} \left| \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) - \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{E}(\mathbb{1}_{Y_i^-=a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) \right| \\ & \leq \phi(\Delta - n) \sum_{a^-=0}^{2^{-m+1}N-1} (\mathbb{P}(Y_i^- = a^-) \mathbb{P}(\mathcal{Y}_{i,m}^+(a^-)) + \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^-=a^-}) + \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{P}(Y_i^- = a^-)) \\ & \leq 3\phi(\Delta - n) \end{aligned}$$

and

$$B_{i,\ell}^+ \leq 2c_5\phi(\Delta - n) \sum_{m=0}^{\lg N} \frac{2^m}{N} \leq 2c_5\phi(\Delta - n).$$

Similarly one estimates the negative term  $B_{i,\ell}^-$  which yields the estimate  $B_{i,\ell}^- \leq c_5\phi(\Delta - n)$ . Along the way one uses the set

$$\mathcal{Y}_{i,m}^-(a^-) = \bigcup_{\substack{2^{-m}N - a^- \leq a^+ < 2^{-m+1}N - a^- \\ \epsilon_{a^-, a^+} = -1}} \{Y_i^+ = a^+\},$$

where we note that

$$\mathcal{Y}_{i,m}^-(a^-) \cup \mathcal{Y}_{i,m}^+(a^-) = \{2^{-m}N - a^- \leq Y_i^+ < 2^{-m+1}N - a^-\}.$$

Consequently

$$(29) \quad |B| \leq \sum_{i=0}^{\frac{t}{\mu(U)}} \sum_{\ell} (B_{i,\ell}^- + B_{i,\ell}^+) \leq \frac{4tc_5}{\mu(U)} \sum_{\ell} \phi(\Delta - n) \leq K \frac{c_6 t \phi(\Delta - n)}{\mu(U)}.$$

(iii) To estimate  $C$ , we first notice that, similarly to what we did for (25)

$$|\mathbb{E}(f(Y_i + \ell) - f(W + \ell))| \leq \|f'\| \mathbb{E}(I_i + Z_i + V_i) \leq c_7 \Delta \mu(U).$$

Thus

$$(30) \quad |C| \leq \sum_i \sum_{\ell} c_7 \Delta \mu(U) \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \leq \sum_i c_7 \Delta \mu(U) \mathbb{E}(I_i \mathbb{1}_{Z_i \geq 1}) \leq c_7 t \Delta \mu(U).$$

Combining (27), (29) and (30) we finally achieve that, for any  $K < \Delta < t/\mu(U)$

$$|\mu(W \in E) - \tilde{\nu}(E)| \leq C_1 t \left( K \frac{\phi(\Delta - n)}{\mu(U)} + \Delta \mu(U) + \phi^1(K/2) + \sum_{j=K/2}^n \mu(U^j) \right),$$

for some  $C_1 \geq c_2 c_3 + c_6 + c_7$  thus proving (8) in the right  $\phi$ -mixing case.

To get the corresponding conclusion in the left  $\phi$ -mixing case notice that in the estimates of  $A_{i,\ell}$  and  $A$  we have to replace  $U^j$  by  $\tilde{U}^j$  and obtain that the estimate (27) is modified to

$$|A| \leq c_2 c_3 t \left[ \Delta \mu(U) + \phi^1(K/2) + \sum_{j=K/2}^n \mu(\tilde{U}^j) \right].$$

The estimates that lead to the bound of the term  $B$  will be the same although the order of splitting and combining terms is the reverse. □

**Remark 5.** There are two places in the proof of Theorem 2 where the argument for the right  $\phi$ -mixing property has to be adjusted. The first one is in the estimate of  $A_{i,\ell}$  (specifically display (26)) and the second place occurs in (28) which has to be replaced by

$$\begin{aligned} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{\mathfrak{Y}_{i,m}^+(a^+)} \mathbb{1}_{Y_i^+=a^+}) &= \mathbb{P}(\mathfrak{Y}_{i,m}^+(a^+)) \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \mathbb{P}(Y_i^+ = a^+) \\ &+ \mathcal{O}^*(\phi(\Delta - n)) (\mathbb{P}(\mathfrak{Y}_{i,m}^+(a^+)) \mathbb{P}(Y_i^+ = a^+) + \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} \mathbb{1}_{Y_i^+=a^+})), \end{aligned}$$

where

$$\mathfrak{Y}_{i,m}^+(a^+) = \{2^{-m}N - a^+ \leq Y_i^- < 2^{-m+1}N - a^+ : \epsilon_{a^-,a^+} = +1\}.$$

Then similarly

$$\mathbb{P}(\mathfrak{Y}_{i,m}^+(a^+)) \mathbb{P}(Y_i^+ = a^+) = \mathbb{E}(\mathbb{1}_{\mathfrak{Y}_{i,m}^+(a^+)} \mathbb{1}_{Y_i^+=a^+}) + \mathcal{O}^*(\mathbb{P}(Y_i^+ = a^+) \phi(2\Delta - n)).$$

Summation over  $a^+$  and  $\ell$  then leads to the estimate (29).

To illustrate the situation suppose we are in the invertible case and  $\Omega$  is a shift space with map  $T = \sigma$  the left shift map. Let us assume that we have a left  $\phi$ -mixing invariant probability measure  $\mu$  on that two-sided shift space  $\Omega$ . In the two-sided case we have to use the two-sided joins  $\hat{\mathcal{A}}^n = \bigvee_{j=n/2}^{n/2-1} T^{-j} \mathcal{A}$  where for simplicity's sake we assume  $n$  to be even. Now for a zero-measure limiting set  $\Gamma$  let  $x = (\dots, x_{-1}, x_0, x_1, x_2, \dots) \in \Omega$  be an arbitrary point and let  $\Gamma = \{y \in \Omega : y_i = x_i \forall i \leq 0\}$  the local unstable leaf through  $x$ . We then let  $\hat{U}_n = \hat{A}_n(\Gamma) = \bigcup_{\hat{A} \in \hat{\mathcal{A}}^n, \Gamma \cap \hat{A} \neq \emptyset} \hat{A}$  and in order to apply Theorem 2 we use the invariance of the measure and consider  $U_n = \sigma^{-n/2} \hat{U}_n$  which satisfies  $U_n \in \sigma(\mathcal{A}^n)$  where  $\mathcal{A}^n$  is the standard one-sided join.

Now, since  $\mu$  has the left  $\phi$ -mixing property the sums of measures of sets  $U_n^i$  as on the RHS in the theorem would have to go to zero as  $n$  goes to infinity. However

$$U_n^i = \sigma^{-(n-i)} A_i(\sigma^{n-i} U_n) = \sigma^{-(n-i)} A_i(\sigma^{\frac{n}{2}-i} \hat{U}_n),$$

which implies that  $U_n^i = \Omega$  whenever  $i < \frac{n}{2}$ . This implies that in a case like this Theorem 2 does not give any meaningful result. The right  $\phi$ -mixing property avoids this problem.

This also explains how the proof has to be changed if we assume left  $\phi$ -mixing instead of right  $\phi$ -mixing: the only difference is in display (26) where we have to use  $\tilde{U}^{j/2}$  instead of  $U^{j/2}$ .

### 6.3. Proof of Theorem 4

This proof is similar to the one of Theorem 2 but allows for some simplifications which we outline below.

**Proof of Theorem 4.** As above let  $\tilde{\nu}$  be the compound Poisson distribution for  $t\tilde{\lambda}_\ell$ ,  $\ell \in \mathbb{N}$  as defined in the statement of the corollary and the preceding theorem. Also we let  $E \subset N_0$  and  $f$  the solution of the Stein equation  $\mathfrak{S}f = \mathbb{1}_E - \tilde{\nu}(E)$ .

For  $K \ll \Delta \ll t/\mu(U)$  we denote as above by  $Z_i$  the close range interactions, by  $V_i^\pm$  the gap terms and by  $Y_i^\pm$  the two halves of the principal terms.

As in the proof of the theorem we split the error into three parts:

$$\mu(W \in E) - \tilde{\nu}(E) = A + B + C,$$

where  $A$  and  $C$  cover short term gap interactions in the dependent and independent case and  $B$  is the error that comes from the principal term with long range interactions. We now proceed to show that each of the three terms can be upper bounded in order to give the stated bounded.

(i) For the first term we write  $A = \sum_{i,\ell} A_{i,\ell}$  where

$$|A_{i,\ell}| = |\mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} (f(Y_i + V_i + \ell) - f(Y_i + \ell))]| \leq \|f'\| \mathbb{E}[I_i \mathbb{1}_{Z_i=\ell} V_i].$$

The  $\psi$ -mixing property then yields ( $U^\ell = A_\ell(U)$ )

$$\sum_{\ell} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell} V_i^+) \leq \sum_{j=K+1}^{2n} \mu(U^{j/2} \cap T^{-j} U) + \sum_{j=2n+1}^{K+\Delta} \mu(U \cap T^{-j} U)$$

$$\begin{aligned} &\leq \mu(U) \left( \sum_{j=K+1}^{2n} \mu(U^{j/2})(1 + \psi(j/2)) + \sum_{j=2n+1}^{K+\Delta} \mu(U)(1 + \psi(j-n)) \right) \\ &\leq \mu(U) \left( 2 \sum_{j=K/2}^n \mu(U^j) + 2\Delta\mu(U) \right) \end{aligned}$$

for  $K$  large enough, and similarly for the left part of the gap  $V_i^-$ . Thus, since  $\|f'\| = \mathcal{O}(1)$  by Theorem 4 in [8],

$$|A| \leq c_1 t \left( \sum_{j=K/2}^n \mu(U^j) + 2\Delta\mu(U) \right).$$

(ii) We now estimate  $B$  which we split as before

$$B = \sum_i \sum_\ell B_{i,\ell},$$

$|B_{i,\ell}| = B_{i,\ell}^- + B_{i,\ell}^+$ , where for  $\epsilon = +1, -1$ :

$$B_{i,\ell}^\epsilon = \epsilon \sum_a |f(a + \ell)| \sum_{\substack{a^- + a^+ = a \\ \epsilon_{a^-, a^+} = \epsilon}} (\mathbb{E}(I_i \mathbb{1}_{Z_i = \ell} \mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{Y_i^+ = a^+}) - \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell}) \mathbb{E}(\mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{Y_i^+ = a^+}))$$

with

$$\epsilon_{a^-, a^+} = \epsilon_{a^-, a^+}(i, \ell) = \text{sgn}(\mathbb{E}(I_i \mathbb{1}_{Z_i = \ell} \mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{Y_i^+ = a^+}) - \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell}) \mathbb{E}(\mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{Y_i^+ = a^+})).$$

Let us begin estimating  $B_{i,\ell}^+$ , the case of  $B_{i,\ell}^-$  will be done below. We partition the sum over  $a$  into segments of exponential progression. For that purpose let us put, using  $N = 1/\mu(U)$

$$g_m(\ell) = \max_{N2^{-m} \leq a < N2^{-m+1}} |f(a + \ell)|$$

which by Proposition 1 of [8] satisfies  $g_m \leq c_1 2^m/N$  for some constant  $c_1$ . We now get

$$B_{i,\ell}^\epsilon \leq \epsilon \sum_{m=0}^{\lg N} c_1 \frac{2^m}{N} \sum_{a^-=0}^{2^{-m+1}N-1} (\mathbb{E}(I_i \mathbb{1}_{Z_i = \ell} \mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) - \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell}) \mathbb{E}(\mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}))$$

where as above

$$\begin{aligned} \mathcal{Y}_{i,m}^+(a^-) &= \bigcup_{\substack{2^{-m}N - a^- \leq a^+ < 2^{-m+1}N - a^- \\ \epsilon_{a^-, a^+} = +1}} \{Y_i^+ = a^+\} \\ &= \{2^{-m}N - a^- \leq Y_i^+ < 2^{-m+1}N - a^- : \epsilon_{a^-, a^+} = +1\} \end{aligned}$$

cuts out slices of exponential progression. By the  $\psi$ -mixing property

$$\begin{aligned} \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell} \mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) &= \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell} \mathbb{1}_{Y_i^- = a^-}) \mathbb{P}(\mathcal{Y}_{i,m}^+(a^-)) (1 + \mathcal{O}^*(\psi(\Delta - n))) \\ &= \mathbb{P}(Y_i^- = a^-) \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell}) \mathbb{P}(\mathcal{Y}_{i,m}^+(a^-)) (1 + \mathcal{O}^*(\psi(\Delta - n))) \\ &= \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell}) \mathbb{E}(\mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) (1 + \mathcal{O}^*(\psi(\Delta - n))) \end{aligned}$$

where the symbol  $\mathcal{O}^*$  indicates an error term where the implied constant is 1 (i.e. if  $G = \mathcal{O}^*(\epsilon)$  then  $|G| \leq \epsilon$ ). Thus

$$B_{i,\ell}^\epsilon \leq c_1 \sum_{m=0}^{\lg N} \frac{2^m}{N} \sum_{a^-=0}^{2^{-m+1}N-1} \mathbb{E}(\mathbb{1}_{Y_i^- = a^-} \mathbb{1}_{\mathcal{Y}_{i,m}^+(a^-)}) \mathbb{E}(I_i \mathbb{1}_{Z_i = \ell}) \psi(\Delta - n)$$

$$\begin{aligned} &\leq c_1 \psi(\Delta - n) \sum_{m=0}^{\lg N} \frac{2^m}{N} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \\ &\leq 2c_1 \psi(\Delta - n) \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \end{aligned}$$

and consequently

$$|B| \leq 2c_1 \psi(\Delta - n) \sum_{i=1}^N \sum_{\ell=1}^{2K+1} \mathbb{E}(I_i \mathbb{1}_{Z_i=\ell}) \leq 2c_1 \psi(\Delta - n) \sum_i \mathbb{E}(I_i \mathbb{1}_{Z_i \geq 1}) \leq 2c_3 t \psi(\Delta - n).$$

(iii) The estimate of the term  $C$  is exactly the one from Theorem 2,

$$|C| \leq c_2 t \Delta \mu(U).$$

Combining the estimates we end up with

$$|\mu(W \in E) - \tilde{\nu}(E)| \leq C'_1 t \inf_{K < \Delta < t/\mu(U)} \left( \psi(\Delta - n) + \Delta \mu(U) + \sum_{j=K/2}^n \mu(U^j) \right)$$

for some constant  $C'_1$ . □

#### 6.4. Proof of Theorem 5

Recall that we start with a nested sequence of sets  $U_n, n \geq 1$ . For  $K < t/\mu(U_n)$  we define  $Z_i^+ = \sum_{j=0}^K I_{i+j}$  and  $Z_i^- = \sum_{j=1}^K I_{i-j}$ , where we assume that  $i \geq K$ . Let us also define  $W^L = \sum_{\ell=0}^L I_\ell$ . In order to prove this result we first state the following lemma taken from [29].

**Lemma 15.** *Assume that the limits  $\alpha_k, k \geq 1$  (see (11)) exist and furthermore  $\sum_{k=1}^\infty k^2 \alpha_k < \infty$ . Then for every  $\eta > 0$  there exists an  $L_0$  so that for all  $L \geq L_0$ :*

$$|\mathbb{E}(\mathbb{1}_{Z_i^+=k} \mathbb{1}_{Z_i^-=\ell-k} I_i) - \mathbb{E}(\mathbb{1}_{Z_i^+=k'} \mathbb{1}_{Z_i^-=\ell-k'} I_i)| \leq \eta \mu(U)$$

for all  $n$  large enough (depending on  $L, \ell$ ).

We are now ready to prove the theorem.

**Proof of Theorem 5.** For  $E \subset \mathbb{N}_0$  and  $K < t/\mu(U_n)$

$$(31) \quad |\mu(W_n \in E) - \tilde{\nu}(E)| \leq |\mu(W_n \in E) - \tilde{\nu}_{K,U_n}(E)| + |\tilde{\nu}_{K,U_n}(E) - \tilde{\nu}(E)|$$

where  $\tilde{\nu}_{K,U_n}$  is as in the statement of Theorem 2. In order to prove Theorem 5 it is therefore enough to prove that both terms on the RHS converge to 0 as  $n \rightarrow \infty$  and  $K \rightarrow \infty$ . We proceed in two steps.

(1) We start proving that the second term on the RHS of (31) converges to 0. First recall the definitions of  $\alpha_k(L, U_n), \alpha_k(L)$  and  $\alpha_k$  given in (11) and that of  $\lambda_\ell(K, U_n)$  given in (9). We have that  $\tilde{\nu}$  and  $\tilde{\nu}_{K,U_n}$  are Poisson compounds with parameters  $\tilde{\lambda}_\ell := \alpha_k - \alpha_{k+1}$  and  $\lambda_\ell(K, U_n)$  respectively. So what has to be proved is that, provided  $\alpha_k$  exists, we have  $\alpha_k - \alpha_{k+1} = \lim_{K \rightarrow \infty} \lim_{n \rightarrow \infty} \lambda_\ell(K, U_n)$ , that is, the convergence of the parameters of the involved Poisson compounds distributions.

Observe that  $\alpha_k(K, U)$  can be written as  $\mathbb{E}(\mathbb{1}_{Z_0=k} | I_0)$ . On the other hand, by translation invariance,  $\mathbb{E}(\mathbb{1}_{Z_0=k} I_0) = \mathbb{E}(\mathbb{1}_{Z_i^+=k} I_i)$  for any  $i \geq 1$ . We therefore work on the later quantity. Consider the disjoint union

$$\{Z_i^+ = k\} \cap \{I_i = 1\} = \bigcup_{\ell=k}^\infty \{Z_i^+ = k\} \cap \{Z_i^- = \ell - k\} \cap \{I_i = 1\}.$$

By invariance, the expectations  $\mathbb{E}(\mathbb{1}_{Z_i^+=k} \mathbb{1}_{Z_i^-=\ell-k} | I_i = 1)$  are equal for all  $i$ . Let us note that in the conditions of Theorem 5, we have  $\sum_{k \geq 1} k^2 \alpha_k < \infty$ . Thus we can use Lemma 15 which states that if  $\eta > 0$ , then for all  $K$  large enough

$$|\mathbb{E}(\mathbb{1}_{Z_i^+=k} \mathbb{1}_{Z_i^-=\ell-k} I_i) - \mathbb{E}(\mathbb{1}_{Z_i^+=k'} \mathbb{1}_{Z_i^-=\ell-k'} I_i)| \leq \eta \mu(U)$$

for  $k, k' = 1, 2, \dots, \ell$ . Hence, since  $Z_i = Z_i^- + Z_i^+$

$$\mathbb{E}(\mathbb{1}_{Z_i^+=k} \mathbb{1}_{Z_i^-=\ell-k} I_i) = \frac{1}{\ell} \mathbb{E}(\mathbb{1}_{Z_i=\ell} I_i) (1 + \mathcal{O}(\eta)) = \tilde{\lambda}_\ell(K, U_n) \mu(U) (1 + \mathcal{O}(\eta))$$

and therefore

$$\mathbb{E}(\mathbb{1}_{Z_i^+=k} I_i) = \sum_{\ell=k}^{\infty} \mathbb{E}(\mathbb{1}_{Z_i^+=k} \mathbb{1}_{Z_i^-=\ell-k} I_i) = (1 + \mathcal{O}(\eta)) \mu(U) \sum_{\ell=k}^{\infty} \tilde{\lambda}_\ell(K, U_n).$$

According to what we said above, we therefore have

$$\begin{aligned} \alpha_k(K, U) &= \mathbb{E}(\mathbb{1}_{Z_0^+=k} | I_0) \\ &= \frac{\mathbb{E}(\mathbb{1}_{Z_0^+=k} I_0)}{\mu(U_n)} = \frac{\mathbb{E}(\mathbb{1}_{Z_i^+=k} I_i)}{\mu(U_n)} \\ &= (1 + \mathcal{O}(\eta)) \sum_{\ell=k}^{\infty} \tilde{\lambda}_\ell(K, U_n). \end{aligned}$$

So in particular  $\alpha_k(K, U_n) - \alpha_{k+1}(K, U_n) = (1 + \mathcal{O}(\eta)) \tilde{\lambda}_k(K, U_n)$ , valid for any positive  $\eta \rightarrow 0$ , thus provided the limit  $\alpha_k$  exists, we have  $\lim_K \lim_n \tilde{\lambda}_k(K, U_n) = \alpha_k - \alpha_{k+1}$ .

(2) In order to prove that the first term of the RHS of (31) converges to 0 we naturally use Theorem 2. Let  $\beta \in (0, 1)$  and choose  $\Delta = \mu(U_n)^{-\beta}$ , we get by (8) that  $|\mu(W_n \in E) - \tilde{v}_{K, U_n}(E)|$  is bounded above by

$$C_1 t \left( K \frac{\phi(\mu(U_n)^{-\beta})}{\mu(U)} + \mu(U_n)^{1-\beta} + \phi^1(K/2) + \sum_{j=K/2}^n \mu(U_n^j) \right)$$

where we recall that, by assumption, for any sufficiently large  $n$ 's, the fourth term is bounded above by  $a_{K/2}$ . The two first terms go to zero as  $n$  diverges with a suitable choice of  $\beta < 1$  so that  $\beta > \frac{1}{\gamma}$ . Then, taking  $K \rightarrow \infty$  we get by assumption that  $a_{K/2}$  and  $\phi^1(K/2)$  vanish as well (by summability of  $\phi$ ). This concludes the proof of the theorem.  $\square$

### 7. Proofs of the results of Sections 3 and 4

**Proof of Proposition 7.** For any two measurable sets  $B \in \sigma(X_0, \dots, X_{n-1})$  and  $A \in \sigma(X_{n+k-1}^\infty)$  with positive probability, let

$$\phi_{A,B}(k) := |\mathbb{P}(X \in B | X \in A) - \mathbb{P}(X \in B)|.$$

Defining  $l(i) := \sup\{j \geq 1 : T_j < i\}$  and using the regenerative property, we get

$$\begin{aligned} \phi_{A,B}(k) &= \left| \sum_{i=1}^{n+k} \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B | X \in A) - \mathbb{P}(X \in B) \right| \\ &= \left| \sum_{i=k+1}^{n+k} \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B | X \in A) + \sum_{i=1}^k \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B) \right. \\ &\quad \left. - \sum_{i=1}^{n+k} \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B) \right| \end{aligned}$$

$$\begin{aligned} &= \left| \sum_{i=k+1}^{n+k} \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B | X \in A) - \sum_{i=k+1}^{n+k} \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B) \right| \\ &\leq \sum_{i=k+1}^{n+k} \left| \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B | X \in A) - \mathbb{P}(n+k - T_{l(n+k)} = i, X \in B) \right| \\ &\leq \sum_{i=k+1}^{n+k} (\mathbb{P}(n+k - T_{l(n+k)} = i | X \in A) + \mathbb{P}(n+k - T_{l(n+k)} = i)). \end{aligned}$$

By definition,  $n+k-1 \in [T_{l(n+k)}, T_{l(n+k)+1} - 1]$  and therefore,

$$\mathbb{P}(n+k - T_{l(n+k)} = i | X \in A) \leq \sup_{b \in \mathcal{A}} \mathbb{P}(T_1 = i | X_0 = b).$$

Thus

$$\begin{aligned} \phi_{A,B}(k) &\leq \sum_{i>k} (\mathbb{P}(T_1 = i | X_0 = b) + \mathbb{P}(T_1 = i)) \\ &= \sum_{i>k} \left( \bar{q}_b(i) + \sum_a \bar{p}(a) \bar{q}_a(i) \right) \\ &\leq 2 \sup_{a \in \mathcal{A}} \sum_{i>k} \bar{q}_a(i). \end{aligned}$$

□

**Proof of Proposition 9.** In the symbolic setting, if  $x$  has prime period  $m$  we have by translation invariance

$$(32) \quad \frac{\mu(A_{n+m}(x))}{\mu(A_n(x))} = \frac{\mu([x_1^{n+m}])}{\mu([x_1^n])} = \frac{\mu([x_1^{n+m}])}{\mu([x_{m+1}^{n+m}])}.$$

If  $x$  is a continuity point for  $g^m := \prod_{i=0}^{m-1} g \circ \sigma^i$ , then

$$\text{var}_k g^m(x) := \sup \{ |g^m(y) - g^m(z)| : y, z \in [x_1^k] \} \xrightarrow{k \rightarrow \infty} 0.$$

On the other hand, we can write

$$(33) \quad \mu([x_1^{n+m}]) = \int_{[x_{m+1}^{n+m}]} g^m(x_1^m y) d\mu(y) = \mu([x_{m+1}^{n+m}]) [g^m(x) + O^*(\text{var}_{n-m} g(x))].$$

This means that, if  $g^m$  is continuous at the periodic point  $x$ , we have that  $\frac{\mu(A_{n+m}(x))}{\mu(A_n(x))}$  converges to  $g^m(x) := \prod_{i=0}^{m-1} g(\sigma^i x)$ , concluding the proof of existence and computation of  $p$ .

The existence of the limiting parameters is now proved, and according to the discussion of Section 3.7, the proof of the second statement follows automatically using our Theorem 5. Indeed, the assumptions of this theorem are granted since as we already said, under summable variation, the measure is  $\psi$ -mixing, and moreover, the assumption that  $g > 0$  implies that  $U_n^j = A_j(x)$  has exponentially decaying measure in  $j$  since a simple argument shows that  $\mu(A_j(x)) \leq (\sup g)^j$ . □

**Proof of Proposition 10.** All the properties we use here, concerning the renewal measure, are proved in [3]. First, the existence of the parameters follows from Theorem 3.2 therein. Under our conditions, the measure under study is left and right  $\phi$ -mixing with exponential rate  $\phi$  (the renewal process is reversible in the stochastic process sense). Finally, since  $p_i \in [\epsilon, 1 - \epsilon]$ , the same holds for  $g$ , which automatically implies that  $U_n^j = A_j(x)$  has exponentially decaying measure in  $j$  as in the preceding proof. □

**Proof of Proposition 11.** Let us first mention that, as observed before the statement of the proposition, if  $\epsilon = 1/2$ , we have that  $\nu$  is the product measure with marginal  $\nu([+1]) = \nu([-1]) = 1/2$ . Thus, it is in particular a  $g$ -measure with  $g$  depending on only one site and the statement, in this case, is a particular case of Proposition 9. In what follows, we therefore focus on the cases where  $\epsilon \neq 1/2$ .

First of all, let us observe that the example is  $\psi$ -mixing as it is a two-coordinates factor map of a product measure. So it satisfies condition (1) of Theorem 5 concerning the mixing properties. The measure of cylinders of size  $n$  decays exponentially fast, so the second condition of Theorem 5 is granted as well. It only remains to check the third condition, but it holds if we are able to prove that the limit (14) holds. This is what we prove below.

Consider a point  $y \in \{-1, +1\}^{\mathbb{N}}$  of prime period  $m \geq 1$ . By (32) we have to compute the limit of

$$\frac{\nu([y_1^{n+m}]) \nu([y_2^{n+m}])}{\nu([y_2^{n+m}]) \nu([y_3^{n+m}])} \cdots \frac{\nu([y_m^{n+m}])}{\nu([y_{m+1}^{n+m}])}.$$

Let us start computing the limit (and proving it exists) of

$$\frac{\nu([1y_2^n])}{\nu([y_2^n])}, \quad n \geq 1.$$

Observe that  $x^+(y_2^\infty)$  and  $x^-(y_2^\infty)$  are also periodic points. Let us denote by  $m'$  their common prime period. Denote by  $s_n$  the number of ones in

$$x^+(y_2^\infty)_2, x^+(y_2^\infty)_3, \dots, x^+(y_2^\infty)_{n+2}$$

(that is, of the  $n + 1$  first coordinates of  $x^+(y_2^\infty)$ ). For technical matters, we will write  $n$  as  $k_n(2m' - 1) + r_n$  where  $k_n := \lfloor n/(2m' - 1) \rfloor$  and  $r_n$  is the remaining part, strictly smaller than  $2m' - 1$ . Observe that, since  $2m'$  is a period of  $x^+(y_2^\infty)$ , we have  $s_n = k_n s_{2m'-1} + R_n$  where  $R_n := s_n - k_n s_{2m'-1} < s_{2m'-1}$ . Therefore

$$\frac{s_n}{n+1} = \frac{k_n s_{2m'-1} + R_n}{k_n(2m'-1) + r_n + 1} \rightarrow \frac{s_{2m'-1}}{2m'-1}.$$

On the other hand, a simple calculation (see [20]) gives that

$$\frac{\nu([1y_2^{n+1}])}{\nu([y_2^{n+1}])} = (1 - \epsilon) \frac{(\frac{\epsilon}{1-\epsilon}) + (\frac{\epsilon}{1-\epsilon})^{2(n+1)(\frac{1}{2} - \frac{s_n}{n+1})}}{1 + (\frac{\epsilon}{1-\epsilon})^{2(n+1)(\frac{1}{2} - \frac{s_n}{n+1})}}.$$

We therefore have the following limits according to the values of  $\epsilon$  and  $\frac{s_{2m'-1}}{2m'-1}$ :

$$(34) \quad \frac{\nu([1y_2^n])}{\nu([y_2^n])} \xrightarrow{n \rightarrow \infty} \begin{cases} \epsilon & \text{if } 0 < \left(\frac{1}{2} - \epsilon\right) \left(\frac{1}{2} - \frac{s_{2m'-1}}{2m'-1}\right), \\ 1 - \epsilon & \text{if } 0 > \left(\frac{1}{2} - \epsilon\right) \left(\frac{1}{2} - \frac{s_{2m'-1}}{2m'-1}\right). \end{cases}$$

(Obviously,  $\lim \frac{\nu([-1y_2^n])}{\nu([y_2^n])} = 1 - \lim \frac{\nu([1y_2^n])}{\nu([y_2^n])}$ .) The same limiting value (34) holds for

$$\frac{\nu([1y_i^{n+m}])}{\nu([y_{i+1}^{n+m}])}, \quad i = 3, \dots, m.$$

So let  $k$  be the number of  $+1$ 's in the period of  $y$  (which we recall is of size  $m$ ). According to (32) we can thus conclude,

$$(35) \quad p_m^{(2)} = \begin{cases} \epsilon^k (1 - \epsilon)^{m-k} & \text{if } 0 < \left(\frac{1}{2} - \epsilon\right) \left(\frac{1}{2} - \frac{s_{2m'-1}}{2m'-1}\right), \\ (1 - \epsilon)^k \epsilon^{m-k} & \text{if } 0 > \left(\frac{1}{2} - \epsilon\right) \left(\frac{1}{2} - \frac{s_{2m'-1}}{2m'-1}\right). \end{cases} \quad \square$$

**Proof of the first statement of Theorem 12.** For simplicity, we do the proof with  $m = 2$ , but the general case follows identically. By assumption  $\hat{\mu}$  is  $\psi$ -mixing and consequently the conditions (1) and (2) of Theorem 5 are satisfied. So if we prove that  $\hat{\alpha}_{k+1}$  exists for any  $k$  and satisfies  $\hat{\alpha}_{k+1} = p^k$  for some  $p \in (0, 1)$ , then we prove at once that Theorem 5 holds and that the asymptotic distribution is Polya–Aeppli as stated.

Recall that, for any  $n \geq 1$ ,  $S_n = \bigcup_{A \in \mathcal{A}^n} A^m$  is the  $n$ -cylinder neighbourhood of the diagonal  $\Delta = \{(x, \dots, x) : x \in \Sigma_B\} \subset \Sigma_B^m$ . Write

$$\hat{\alpha}_{k+1} := \frac{\hat{\mu}(\bigcap_{i=0}^k \hat{\sigma}^{-i} S_n)}{\hat{\mu}(S_n)} = \frac{\hat{\mu}(S_n \cap \hat{\sigma}^{-1} S_n)}{\hat{\mu}(S_n)} \frac{\hat{\mu}(S_n \cap \hat{\sigma}^{-1} S_n \cap \hat{\sigma}^{-2} S_n)}{\hat{\mu}(S_n \cap \hat{\sigma}^{-1} S_n)} \cdots \frac{\hat{\mu}(\bigcap_{i=0}^k \hat{\sigma}^{-i} S_n)}{\hat{\mu}(\bigcap_{i=0}^{k-1} \hat{\sigma}^{-i} S_n)}$$

which by translation invariance writes

$$\frac{\hat{\mu}(S_n \cap \hat{\sigma}^{-1} S_n)}{\hat{\mu}(\hat{\sigma}^{-1} S_n)} \frac{\hat{\mu}(S_n \cap \hat{\sigma}^{-1} S_n \cap \hat{\sigma}^{-2} S_n)}{\hat{\mu}(\hat{\sigma}^{-1} S_n \cap \hat{\sigma}^{-2} S_n)} \cdots \frac{\hat{\mu}(\bigcap_{i=0}^k \hat{\sigma}^{-i} S_n)}{\hat{\mu}(\bigcap_{i=1}^k \hat{\sigma}^{-i} S_n)}.$$

Now, for  $j = 1, \dots, k$  and  $n$  large enough

$$\frac{\hat{\mu}(\bigcap_{i=0}^j \hat{\sigma}^{-i} S_n)}{\hat{\mu}(\bigcap_{i=1}^j \hat{\sigma}^{-i} S_n)} = \frac{\hat{\mu}((x, y) : x_1^{n+j} = y_1^{n+j})}{\hat{\mu}((x, y) : x_2^{n+j} = y_2^{n+j})} =: u_{n+j}.$$

Let us assume for now that  $u_{n+1} = \frac{\hat{\mu}(S_n \cap \hat{\sigma}^{-1} S_n)}{\hat{\mu}(S_n)}$  converges, and let  $p$  denote the limit. Then for any  $j \geq 1$ ,  $u_{n+j} = \frac{\hat{\mu}(S_{n+j-1} \cap \hat{\sigma}^{-1} S_{n+j-1})}{\hat{\mu}(S_{n+j-1})} \rightarrow p$  and the limit defining  $\hat{\alpha}_{k+1}$  exists and it equals  $p^k$ . In other words, provided the limit  $p$  exists we always have, in the limit, a Pólya–Aeppli distribution with parameter  $t(1 - p)$ , as stated by the theorem.

So it only remains to prove the existence of the limit  $p$ . Consider the projection operator  $\Pi : \Sigma \rightarrow \{0, 1\}^{\mathbb{N}}$  defined through  $\Pi(x, y) = z$  where  $z_i = \mathbb{1}_{x_i=y_i}$ . With this we now have to check whether

$$\lim_n \frac{\hat{\mu} \circ \Pi^{-1}([1^{n+1}])}{\hat{\mu} \circ \Pi^{-1}([1^n])} = \lim_n \mathbb{E}_{\hat{\mu} \circ \Pi^{-1}}(\mathbb{1}_{[1]} | \mathcal{F}_2^n)(1^{(\infty)})$$

exists. Using [42, Proposition 5] we only have to prove that the measure  $\hat{\mu} \circ \Pi^{-1}$  has a continuous and strictly positive  $g$ -function. By assumption,  $\hat{g}$  is strictly positive and with summable variation. By Theorem 1.1 of [50], we automatically have that  $\hat{\mu} \circ \Pi^{-1}$  has an everywhere continuous and strictly positive  $g$ -function. This concludes the proof of the theorem.  $\square$

**Proof of the second statement of Theorem 12.** For simplicity, we do the proof with  $m = 2$ , but the general case follows identically. For that reason let  $\hat{\mu}$  be the  $\hat{g}$  measure on  $\Sigma_B \times \Sigma_B$ . As above, the first two conditions of Theorem 5 are granted under our assumptions. We will show that  $\hat{\alpha}_2$  exists by computing it, this will automatically grant  $\hat{\alpha}_{k+1} = p^k$  and the third condition of Theorem 5, and conclude our proof.

By conformality we have then for all finite words  $\alpha, \beta$  that

$$\hat{\mu}(\sigma[\alpha] \times \sigma[\beta]) = \int_{[\alpha] \times [\beta]} \hat{g}(x, y)^{-1} d\hat{\mu}(x, y).$$

In particular, if we put  $\hat{g}_k(x, y) = \prod_{j=0}^{k-1} \hat{g}(\sigma^j(x), \sigma^j(y))$ , then for  $k$ -words  $\alpha', \alpha''$  and  $n$ -words  $\beta', \beta''$  one has

$$\hat{\mu}([\alpha' \beta'] \times [\alpha'' \beta'']) = \hat{\mu}([\beta'] \times [\beta'']) \hat{g}_k(\alpha' \beta', \alpha'' \beta'') e^{\mathcal{O}(v_n^1)},$$

where  $v_n^1 = \sum_{j=n}^{\infty} v_j$  is the tailsum of  $v_n = \text{var}_n \hat{g}$  and  $\hat{g}(\gamma', \gamma'') = \sup_{(x,y) \in [\gamma'] \times [\gamma'']} \hat{g}(x, y)$ .

By assumption  $\hat{\mu}$  is  $\psi$ -mixing and consequently the conditions of Theorem 5 are satisfied if we prove that the following limit exists

$$\hat{\alpha}_{k+1} = \lim_{n \rightarrow \infty} \frac{\hat{\mu}(\bigcap_{i=0}^k \hat{\sigma}^{-i} S_n)}{\hat{\mu}(S_n)}.$$

Indeed

$$\hat{\mu}\left(\bigcap_{i=0}^k \hat{\sigma}^{-i} S_n\right) = \hat{\mu}(S_{n+k})$$

$$\begin{aligned}
 &= \sum_{\alpha \in \mathcal{A}^k} \sum_{\beta \in \mathcal{A}^n} \hat{\mu}([\alpha\beta] \times [\alpha\beta]) \\
 &= \sum_{\beta \in \mathcal{A}^n} \hat{\mu}([\beta] \times [\beta]) \sum_{\alpha \in \mathcal{A}^k} \hat{g}_k(\alpha\beta, \alpha\beta) e^{\mathcal{O}(v_n^1)}.
 \end{aligned}$$

Since on the other hand  $\hat{\mu}(\mathcal{S}_n) = \sum_{|\beta|=n} \hat{\mu}([\beta] \times [\beta])$  we get

$$\hat{\alpha}_{k+1} = \lim_{n \rightarrow \infty} \frac{\sum_{\beta \in \mathcal{A}^n} \hat{\mu}([\beta] \times [\beta]) \sum_{\alpha \in \mathcal{A}^k} \hat{g}_k(\alpha\beta, \alpha\beta) e^{\mathcal{O}(v_n^1)}}{\sum_{\beta \in \mathcal{A}^n} \hat{\mu}([\beta] \times [\beta])}.$$

For any  $n \geq 1$ , let us define on  $\Sigma_B$  the measure

$$\nu_n = \frac{1}{Z_n} \sum_{\gamma \in \mathcal{A}^n} \hat{\mu}([\gamma] \times [\gamma]) \delta_\gamma,$$

where  $\delta_\gamma$  is a point mass at an arbitrarily chosen point  $x_\gamma \in [\gamma] \subset \Sigma_B$  depending only on the last symbol of  $\gamma$  so that  $x_{\sigma\gamma} = \sigma(x_\gamma)$  and  $Z_n = \sum_{|\beta|=n} \hat{\mu}([\beta] \times [\beta])$  is the normalising factor. Acting on functions  $f : \Sigma_B \rightarrow \mathbb{R}$  we define the transfer operator  $\mathcal{L}$  by

$$\mathcal{L}f(x) = \sum_{a \in \mathcal{A}} g^\Delta(ax) f(ax),$$

where  $g^\Delta : \Sigma_B \rightarrow \mathbb{R}$  is given by  $g^\Delta(y) = \hat{g}(y, y)$ . Then for the action of  $\mathcal{L}^k$  on  $\nu_n$  we get

$$\begin{aligned}
 (\mathcal{L}^k \nu_n)(f) &= \int \sum_{|\beta|=k} g_k^\Delta(\beta x) f(\beta x) d\nu_n(x) \\
 &= \frac{1}{Z_n} \sum_{|\gamma|=n} \sum_{|\beta|=k} g_k^\Delta(\beta x_\gamma) \hat{\mu}([\gamma] \times [\gamma]) f(\beta x_\gamma) \\
 &= \frac{1}{Z_n} \sum_{|\alpha|=n+k} \hat{\mu}([\alpha] \times [\alpha]) f(\alpha x_\alpha) \\
 &= \frac{Z_{n+k}}{Z_n} e^{\mathcal{O}(v_n^1)} \nu_{n+k}(f),
 \end{aligned}$$

where we used that by conformality

$$\hat{\mu}([\gamma] \times [\gamma]) = \int_{[\beta\gamma] \times [\beta\gamma]} \hat{g}_k(x, y)^{-1} d\hat{\mu}(x, y)$$

which implies  $\hat{\mu}([\gamma] \times [\gamma]) g_k^\Delta(\beta x_\gamma) = \hat{\mu}([\beta\gamma] \times [\beta\gamma]) e^{\mathcal{O}(v_n^1)}$  as  $x_{\beta\gamma} = \beta x_\gamma$ . That is, we can write

$$\nu_n = c_{n,k} e^{\mathcal{O}(v_k^1)} \mathcal{L}^{n-k} \nu_k,$$

where  $c_{n,k}$  is a normalising constant.

Now let  $\nu$  be the unique conformal measure for  $e^{-P}\mathcal{L}$ , where  $P$  is the pressure of  $\log g^\Delta$  (on  $(\Sigma_B, \sigma)$ ). Evidently  $e^{-P}\mathcal{L}\nu = \nu$  and there is an associated positive eigenfunction  $h$  so that  $e^{-P}\mathcal{L}h = h$ . For simplicity's sake we assume the normalisation  $\nu(1) = \nu(h) = 1$ . Then

$$e^{-\ell P} \mathcal{L}^\ell(f) = h\nu(f) + \mathcal{O}(\lambda^\ell),$$

where  $\lambda < 1$  as  $\mathcal{L}$  is quasi compact which is a consequence of exponentially decaying variation of  $g^\Delta$ . Thus

$$e^{-\ell P} (\mathcal{L}^\ell \nu_k)(f) = e^{-\ell P} \nu_k(\mathcal{L}^\ell(f)) = \nu_k(h)\nu(f) + \mathcal{O}(\lambda^\ell)$$

and consequently for every  $k$  and function  $f$ :

$$\begin{aligned}\lim_{n \rightarrow \infty} v_n(f) &= \lim_{n \rightarrow \infty} c_{n,k} \mathcal{L}^{n-k} v_k(f) e^{\mathcal{O}(v_k^1)} \\ &= v_k(h) v(f) e^{\mathcal{O}(v_k^1)}.\end{aligned}$$

In particular for the constant function  $f = 1$  one has  $1 = \lim_{n \rightarrow \infty} v_n(1) = v_k(h) v(1) e^{\mathcal{O}(v_k^1)}$  which implies that  $v_k(h) = e^{-\mathcal{O}(v_k^1)}$ . If we let  $k \rightarrow \infty$  we obtain that  $v_k(h) \rightarrow 1$  which implies that in fact  $v_n \rightarrow v$  weakly.

Finally we obtain

$$\hat{\alpha}_{k+1} = \lim_{n \rightarrow \infty} \sum_{\alpha \in \mathcal{A}^k} \int_{\Sigma_B} g_k^\Delta(\alpha x) dv_n(x) = \sum_{\alpha \in \mathcal{A}^k} \int g_k^\Delta(\alpha x) dv(x) = \int \mathcal{L}^k 1(x) dv(x)$$

and consequently

$$\hat{\alpha}_{k+1} = v(\mathcal{L}^k 1) = e^{kP}$$

since  $\mathcal{L}^k(1) = e^{kP} h v(1) + R_k$  where  $R_k$  is orthogonal to  $h$ , that is  $v(R_k) = 0$ . This implies that the limiting distribution is Pólya–Aeppli since  $P = P(\log g^\Delta)$  is negative which follows from the fact that the pressure of  $\hat{g}$  is zero on the system  $(\Sigma_B^2, \hat{\sigma})$  and that the topological entropy of  $\hat{\sigma}$  is positive by the  $\psi$ -mixing property.  $\square$

## Acknowledgements

SG would like to thanks Frédéric Paccaut for discussions concerning the Furstenberg example.

## Funding

SG thanks the Centre de Physique Théorique of Marseille for hospitality during part of the elaboration of this work and FAPESP scholarship abroad (2017/07084-6) for financial support during this stay. SG also thanks the support of FAPESP (2019/23439-4) and CNPq Universal (439422/2018-3). This article was produced as part of the activities of FAPESP Research, Innovation and Dissemination Center for Neuromathematics (grant 2013/ 07699-0).

NH was supported by Université de Toulon and the Simons Foundation (ID 526571). The research of SV was supported by the project “Dynamics and Information Research Institute” within the agreement between UniCredit Bank and Scuola Normale Superiore di Pisa.

## References

- [1] M. Abadi. Exponential approximation for hitting times in mixing processes. *Math. Phys. Electron. J.* **7** (2001) Paper 19. [MR1871384](#)
- [2] M. Abadi, V. Amorim, J.-R. Chazottes and S. Gallo. Return-time-spectrum for equilibrium states with potentials of summable variation. *Ergodic Theory Dynam. Systems* **43** (8) (2023) 2489–2515. [MR4611739](#) <https://doi.org/10.1017/etds.2022.40>
- [3] M. Abadi, L. Cardéno and S. Gallo. Potential well spectrum and hitting time in renewal processes. *J. Stat. Phys.* **159** (5) (2015) 1087–1106. [MR3345411](#) <https://doi.org/10.1007/s10955-015-1216-y>
- [4] M. Abadi, A. C. M. Freitas and J. M. Freitas. Clustering indices and decay of correlations in non-Markovian models. *Nonlinearity* **32** (12) (2019) 4853. [MR4030595](#) <https://doi.org/10.1088/1361-6544/ab37b8>
- [5] M. Abadi, A. C. M. Freitas and J. M. Freitas. Dynamical counterexamples regarding the extremal index and the mean of the limiting cluster size distribution. *J. Lond. Math. Soc.* **102** (2) (2020) 670–694. [MR4171430](#) <https://doi.org/10.1112/jlms.12332>
- [6] S. Asmussen. *Applied Probability and Queues Stochastic Modelling and Applied Probability* 2nd edition. *Applications of Mathematics (New York)* **51**. Springer-Verlag, New York, 2003. [MR1978607](#)
- [7] W. Bahsoun, H. Hu and S. Vaienti. Pseudo-orbits, stationary measures and metastability. *Dyn. Syst.* **29** (3) (2014) 322–336. [MR3227776](#) <https://doi.org/10.1080/14689367.2014.890172>
- [8] A. D. Barbour, L. H. Chen and W.-L. Loh. Compound poisson approximation for nonnegative random variables via stein’s method. *The Annals of Probability* (1992) 1843–1866. [MR1188044](#)
- [9] R. C. Bradley. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2** (2) (2005) 107–144. [MR2178042](#) <https://doi.org/10.1214/154957805100000104>
- [10] R. C. Bradley. Introduction to strong mixing conditions, 2007. Kendrick press. [MR2325295](#)
- [11] X. Bressaud, R. Fernández and A. Galves. Decay of correlations for non-Hölderian dynamics. A coupling approach. *Electron. J. Probab.* **4** (3) (1999) 19. [MR1675304](#) <https://doi.org/10.1214/EJP.v4-40>
- [12] M. Carney, M. Holland and M. Nicol. Extremes and extremal indices for level set observables on hyperbolic systems. *Nonlinearity* **34** (2) (2021) 1136. [MR4228013](#) <https://doi.org/10.1088/1361-6544/abd85f>

- [13] J.-R. Chazottes, Z. Coelho and P. Collet. Poisson processes for subsystems of finite type in symbolic dynamics. *Stoch. Dyn.* **9** (03) (2009) 393–422. MR2566908 <https://doi.org/10.1142/S0219493709002713>
- [14] J.-R. Chazottes and P. Collet. Poisson approximation for the number of visits to balls in non-uniformly hyperbolic dynamical systems. *Ergodic Theory Dynam. Systems* **33** (1) (2013) 49–80. MR3009103 <https://doi.org/10.1017/S0143385711000897>
- [15] L. H. Y. Chen and A. Barbour. *Stein's Method and Applications*, 5. World Scientific, Singapore, 2005. MR2201882 <https://doi.org/10.1142/9789812567673>
- [16] E. Cinlar. Introduction to stochastic processes. *Courier Corporation*. (2013). MR0380912
- [17] Z. Coelho and P. Collet. Asymptotic limit law for the close approach of two trajectories in expanding maps of the circle. *Probab. Theory Related Fields* **99** (2) (1994) 237–250. MR1278884 <https://doi.org/10.1007/BF01199024>
- [18] D. Faranda, H. Ghoudi, P. Guiraud and S. Vaienti. Extreme value theory for synchronization of coupled map lattices. *Nonlinearity* **37** (7) (2018) 3326. MR3816758 <https://doi.org/10.1088/1361-6544/aabc8e>
- [19] R. Fernández. Gibbsianness and non-gibbsianness in lattice random fields. *Les Houches LXXXIII* (2005) 731–799. MR2581896 [https://doi.org/10.1016/S0924-8099\(06\)80052-1](https://doi.org/10.1016/S0924-8099(06)80052-1)
- [20] R. F. Ferreira, S. Gallo and F. Paccaut. Non-regular g-measures and variable length memory chains. *Nonlinearity* **33** (11) (2020) 6026. MR4164670 <https://doi.org/10.1088/1361-6544/aba0c5>
- [21] A. C. M. Freitas, J. M. Freitas and M. Magalhães. Convergence of marked point processes of excesses for dynamical systems. *J. Eur. Math. Soc.* **20** (9) (2018) 2131–2179. MR3836843 <https://doi.org/10.4171/JEMS/808>
- [22] A. C. M. Freitas, J. M. Freitas, M. Magalhães and S. Vaienti. Point processes of non stationary sequences generated by sequential and random dynamical systems. *J. Stat. Phys.* **181** (4) (2020) 1365–1409. MR4163505 <https://doi.org/10.1007/s10955-020-02630-z>
- [23] A. C. M. Freitas, J. M. Freitas and M. Todd. The compound Poisson limit ruling periodic extreme behaviour of non-uniformly hyperbolic dynamics. *Comm. Math. Phys.* **321** (2) (2013) 483–527. MR3063917 <https://doi.org/10.1007/s00220-013-1695-0>
- [24] J. M. Freitas, N. Haydn and M. Nicol. Convergence of rare event point processes to the Poisson process for planar billiards. *Nonlinearity* **27** (7) (2014) 1669. MR3232197 <https://doi.org/10.1088/0951-7715/27/7/1669>
- [25] H. Furstenberg. *Stationary Processes and Prediction Theory*, 44. Princeton University Press, Princeton, 1960. MR0140151
- [26] A. Galves and E. Löcherbach. Infinite systems of interacting chains with memory of variable length—a stochastic model for biological neural nets. *J. Stat. Phys.* **151** (5) (2013) 896–921. MR3055382 <https://doi.org/10.1007/s10955-013-0733-9>
- [27] N. Haydn and Y. Psiloyenis. Return times distribution for Markov towers with decay of correlations. *Nonlinearity* **27** (6) (2014) 1323. MR3215837 <https://doi.org/10.1088/0951-7715/27/6/1323>
- [28] N. Haydn and S. Vaienti. The distribution of return times near periodic orbits. *Probab. Theory Related Fields* **144** (2009) 517–542. MR2496441 <https://doi.org/10.1007/s00440-008-0153-y>
- [29] N. Haydn and S. Vaienti. Limiting entry and return times distribution for arbitrary null sets. *Comm. Math. Phys.* **378** (1) (2020) 149–184. MR4124984 <https://doi.org/10.1007/s00220-020-03795-0>
- [30] N. Haydn and K. Wasilewska. Limiting distribution and error terms for the number of visits to balls in non-uniformly hyperbolic dynamical systems. *Discrete Contin. Dyn. Syst.* **36** (5) (2016) 2585–2611. MR3485409 <https://doi.org/10.3934/dcds.2016.36.2585>
- [31] N. Haydn and F. Yang. A derivation of the poisson law for returns of smooth maps with certain geometrical properties. *Contemporary Mathematics Proceedings in memoriam Chernov*. (2017). MR3716090 <https://doi.org/10.1090/conm/698/14031>
- [32] N. T. Haydn. Entry and return times distribution. *Dyn. Syst.* **28** (3) (2013) 333–353. MR3170620 <https://doi.org/10.1080/14689367.2013.822459>
- [33] M. Hirata. Poisson law for axiom a diffeomorphisms. *Ergodic Theory Dynam. Systems* **13** (3) (1993) 533–556. MR1245828 <https://doi.org/10.1017/S0143385700007513>
- [34] J. Janssen and R. Manca. *Applied Semi-Markov Processes*. Springer Science & Business Media, Berlin, 2006. MR2175472
- [35] M. Keane. Strongly mixing g-measures. *Invent. Math.* **16** (1972) 309–324. MR0310193 <https://doi.org/10.1007/BF01425715>
- [36] G. Keller. Rare events, exponential hitting times and extremal indices via spectral perturbation. *Dyn. Syst.* **27** (1) (2012) 11–27. MR2903242 <https://doi.org/10.1080/14689367.2011.653329>
- [37] G. Keller and C. Liverani. Rare events, escape rates and quasistationarity: Some exact formulae. *J. Stat. Phys.* **135** (3) (2009) 519–534. MR2535206 <https://doi.org/10.1007/s10955-009-9747-8>
- [38] Y. Kifer and A. Rapaport. Poisson and compound Poisson approximations in conventional and nonconventional setups. *Probab. Theory Related Fields* **160** (3–4) (2014) 797–831. MR3278921 <https://doi.org/10.1007/s00440-013-0541-9>
- [39] Y. Kifer and F. Yang. Geometric law for numbers of returns until a hazard under  $\phi$ -mixing, 2018. ArXiv preprint. Available at arXiv:1812.09927. MR4344031 <https://doi.org/10.1007/s11856-021-2182-5>
- [40] F. Ledrappier. Principe variationnel et systèmes dynamiques symboliques. *Z. Wahrsch. Verw. Gebiete* **30** (1974) 185–202. MR0404584 <https://doi.org/10.1007/BF00533471>
- [41] V. Lucarini, D. Faranda, J. M. M. de Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, S. Vaienti et al. *Extremes and Recurrence in Dynamical Systems*. Wiley, New York, 2016. MR3558780 <https://doi.org/10.1002/9781118632321>
- [42] M. R. Palmer, W. Parry and P. Walters. Large sets of endomorphisms and of g-measures. In *The Structure of Attractors in Dynamical Systems* 191–210. Springer, Berlin, 1978. MR0518560
- [43] F. Pène and B. Saussol. Back to balls in billiards. *Comm. Math. Phys.* **293** (3) (2010) 837–866. MR2566164 <https://doi.org/10.1007/s00220-009-0911-4>
- [44] F. Pène and B. Saussol. Poisson law for some non-uniformly hyperbolic dynamical systems with polynomial rate of mixing. *Ergodic Theory Dynam. Systems* **36** (8) (2016) 2602–2626. MR3570026 <https://doi.org/10.1017/etds.2015.28>
- [45] F. Pène and B. Saussol. Spatio-temporal Poisson processes for visits to small sets. *Israel J. Math.* **240** (2) (2020) 625–665. MR4193145 <https://doi.org/10.1007/s11856-020-2074-0>
- [46] B. Pittskel. Poisson limit law for Markov chains. *Ergodic Theory Dynam. Systems* **11** (3) (1991) 501–513. MR1125886 <https://doi.org/10.1017/S0143385700006301>
- [47] M. Roos et al. Stein's method for compound Poisson approximation: The local approach. *Ann. Appl. Probab.* **4** (4) (1994) 1177–1187. MR1304780
- [48] P. C. Shields. *The Ergodic Theory of Discrete Sample Paths*. Graduate Studies in Mathematics **13**. American Mathematical Society, Providence, RI, 1996. MR1400225 <https://doi.org/10.1090/gsm/013>
- [49] C. Stein Approximate computation of expectations, 1986. IMS. MR0882007

- [50] E. Verbitskiy. On factors of  $g$ -measures. *Indag. Math. (N.S.)* **22** (3–4) (2011) 315–329. MR2853610 <https://doi.org/10.1016/j.indag.2011.09.001>
- [51] E. Verbitskiy. Hidden gibbs models: Theory and applications. Unpublished notes, 2015.
- [52] P. Walters. Ruelle's operator theorem and  $g$ -measures. *Trans. Amer. Math. Soc.* **214** (1975) 375–387. MR0412389 <https://doi.org/10.2307/1997113>
- [53] F. Yang. Rare event process and entry times distribution for arbitrary null sets on compact manifolds. *Ann. Inst. Henri Poincaré Probab. Stat.* **57** (2) (2021) 1103–1135. MR4260497 <https://doi.org/10.1214/20-aihp1109>