Sharp error terms for Poisson statistics under mixing conditions: A new approach *

Miguel Abadi† Nicolas Vergne‡

Abstract

We describe the statistics of the number of occurrences of a string of symbols in a stochastic process: Chosen a string A of length n, we prove that the number of visits to A up to time t, denoted by N_t, has approximately a Poisson distribution. We provide a sharp error for this approximation. Contrarily to previous works who presente uniform error terms based on the total variation distance, our error is point-wise. As a byproduct we obtain approximations for all the moments of N_t. Our result holds for processes that verify the ϕ-mixing condition. The error term is explcitely expressed as function of ϕ and then easily computable. We breafly extend our result to the weaker α-mixing case.

Keyword: Mixing, recurrence, rare event, number of visits, Poisson distribution.

1 Introduction

This paper describes the statistics of occurrence times of a string of symbols in a mixing stochastic process with a finite alphabet. For $n \in I\!N$, we consider a fixed string of n symbols. We prove an upper bound for the difference between the law of the number of occurrences of the string in a long sequence and a Poisson law. Our result stands for ϕ-mixing and strong or α-mixing processes (see definitions below), each one with its corresponding error.

The first result about number of visits is obviusly the convergence of the binomial distribution to the Poisson distribution. Recently, motivated by the statistical analysis of data sources coming from different areas such as physics, biology, computer science, linguistics among other there was a major interes to generalise this convergence in various sense:

*Work done as part of the activities of the project Projeto Temático-PRONEX *"Stochastic behavior, critical phenomena and rhythmic patterns identification in natural languages"*. Pros. FAPESP 03/09930-9.

†IMECC, Universidade de Campinas, Pça Sérgio Buarque de Holanda 651 Cid. Univ. CP 6065, 13083-859, Campinas SP, Brazil. miguel@ime.unicamp.br

‡Université d'Evry Val d'Essonne, Département Mathématiques, Laboratoire Statistique et Génome, 91 000 Evry, France. vergne@genopole.cnrs.fr

(a) dependent process

(b) explicit rate of convergence

(c) different kind of observables

The pioneer paper considering (a) is that of Doeblin ([13]), who studied the Poisson approximation for the Gauss transformation. There is abundant litterature on this subjet in the dynamical systems contex. See for instance Galves and Schmitt [15] and the references there in.

Probably the most used tool to attack (b) is the Chen-Stein method introduced by Chen ([10]). There is also abundant litterature on this subject (see e.g. [6], [7], [8].) The principal feature of this method is that it provides only uniform bounds for the rate of convergence based on the total variation distance. As far us we know, this method was only implemented in processes that verify the Markov property. Wether it is usefull in other context is an open question for us. We are only aware of one work which provides point-wise rate of convergence. Haydn and Vaienti ([16]) prove a rate of convergence using the method of factorial moments. The result holds for $(\psi - f)$-mixing processes. The bound decreases factorially fast on k but it holds only for those values of k that does not exceed the inverse of some (positive) power of the measure of the n-string.

Our result tends to give bring some light over (a); (b); and (c).

With respect to (b) we prove an upper bound for the rate of convergence of the number of occurrences of a fixed string to the Poisson law, namely,

$$\lim_{I\!P(A)\to 0} I\!P\left(N_{t/I\!P(A)} = k\right) = \frac{e^{-t}t^k}{k!} ,$$

where N_t is the number of visits of the process to the string A up to time t.

The error bound we obtain decreases factorially fast as a function of k for *all* values of k. This control on the tail distribution of N_t allows us to obtain an approximation for *all the moments* of N_t by those of a Poisson random variable.

Our approach relies on a sharp result proved by Abadi ([1]) that states that for any string that does not overlaps itself,

$$I\!P(N_{t/I\!P(A)} = 0) \approx e^{-t} .$$

A crucial point is that, if A is any string, $N_{t/I\!P(A)}$ could not be well approximated by a Poisson law. An example of this fact is shown by Hirata ([17]), where it is proven that for periodic points, the asymptotic limit law of $\left\{N_{t/I\!P(A)} = 1\right\}$ differs of the one-level Poisson law. When this happens, Abadi and Vergne ([5], Theorem 2) show that the law of τ_A is different from the exponential. Moreover, Theorem 24 in the same paper shows that A occurs in clumps with geometric size, which says that N_t is not Poisson distributed.

Our result is established with its own error term. This error is explicitly expressed as function of the mixing rate. As we said, tt turns out that the error term depends on the overlapping properties of A. We state some basic facts about overlapping useful to prove our theorem. More on that topic can be find in [5].

With respect to (a) we establish our result under the mixing conditions. Mixing is a large family of processes. For instance, irreducible and aperiodic finite state Markov chain are known to be ψ-mixing with exponential decay. Moreover, Gibbs states which have summable variations are ψ-mixing (see [18]). They have exponential decay if they have Hölder continuous potential (see [9]). We refer the reader to Doukhan ([14]) for a source of examples of mixing processes. However, sometimes the ψ-mixing condition is very restricted hypothesis difficult to test. We establish our result under the more general ϕ-mixing condition. Even when the result is weaker, we find interesting to present also the α-mixing case. In this case we need to impose an extra condition in order for the theorem to hold: the string needs to be non-overlapping (up to some fraction of its own lenght).

With respect to (c), since any observable can be constructed as a union of strings, we focus our work on them.

Our result is applied in a forthcoming paper: In [5] the authors applied the Poisson approximation to test a method for detecting strings of high frequency in DNA and protein sequences.

This paper is organized as follows. In section 2 we establish our framework. In section 3 we collect some definitions and properties of overlapping. In section 4 we state and prove the convergence of the number of occurrences to a Poisson law. The statement are for ϕ-mixing processes. Since the proof of the α-mixing case is similar and easily obtained from that one, we briefly treat this case section 5.

2 Framework and notations

Let $\mathcal{E}$ be a finite set. Put $\Omega = \mathcal{E}^{Z}$. For each $x = (x_m)_{m \in Z} \in \Omega$ and $m \in \mathbb{Z}$, let $X_m : \Omega \to \mathcal{E}$ be the m-th coordinate projection, that is $X_m(x) = x_m$. We denote by $T : \Omega \to \Omega$ the one-step-left shift operator, namely $(T(x))_m = x_{m+1}$.

We denote by $\mathcal{F}$ the σ-algebra over Ω generated by strings. Moreover we denote by $\mathcal{F}_I$ the σ-algebra generated by strings with coordinates in I, $I \subseteq \mathbb{Z}$.

For a subset $A \subseteq \Omega$ we say that $A \in \mathcal{C}_n$ if and only if

$$A = \{X_0 = a_0; \ldots ; X_{n-1} = a_{n-1}\} \ ,$$

with $a_i \in \mathcal{E}, \ i = 0, \ldots, n-1$.

We consider an invariant probability measure $I\!P$ over $\mathcal{F}$. We shall assume without loss of generality that there is no singleton of probability 0.

We say that the process $(X_m)_{m \in Z}$ is ϕ-mixing if the sequence

$$\phi(l) = \sup |I\!P_B(C) - I\!P(C)| \ ,$$

converges to zero. The supremum is taken over B and C such that $B \in \mathcal{F}_{\{0,.,n\}}, n \in I\!N, I\!P(B) > 0, C \in \mathcal{F}_{\{m \geq n+l+1\}}$.

Similarly, we say that the process $(X_m)_{m\in Z}$ is α-mixing if the sequence

$$\alpha(l) = \sup |I\!P(B \cap C) - I\!P(B)I\!P(C)| \ ,$$

converges to zero. The supremum is taken over B and C such that $B \in \mathcal{F}_{\{0,.,n\}}, n \in I\!N, C \in \mathcal{F}_{\{m \geq n+l+1\}}$.

For two measurables V and W, we denote as usual $I\!P(V|W) = I\!P_W(V) = I\!P(V;W) / I\!P(W)$ the conditional measure of V given W. We write $I\!P(V;W) = I\!P(V \cap W)$. We also write $V^c = \Omega \backslash V$, for the complement of V.

We use the probabilistic notation: $\{X_n^m = x_n^m\} = \{X_n = x_n, \ldots, X_m = x_m\}$. For a n-string $A = \{X_0^{n-1} = x_0^{n-1}\}$ and $1 \leq w \leq n$, we write $A^{(w)} = \{X_{n-w}^{n-1} = x_{n-w}^{n-1}\}$ for the w-string belonging to the σ-algebra $\mathcal{F}_{\{n-w,\ldots,n-1\}}$ and consisting of the *last* w symbols of A.

The mean of a r.v. X is denoted by $I\!E(X)$. Wherever it is not ambiguous we will write C and c for different positive constants even in the same sequence of equalities/inequalities. For brevity we put $(a \vee b) = \max\{a, b\}$ and $(a \wedge b) = \min\{a, b\}$.

3 Overlapping

Definition 1 *Let* $A \in \mathcal{C}_n$. *We define the* periodicity *of* A *(with respect to* T*) as the number* $\tau(A)$ *defined as follows:*

$$\tau(A) = \min \left\{ k \in \{1, \ldots, n\} \mid A \cap \ T^{-k}(A) \neq \emptyset \right\} \ .$$

Let us write $n = q\,p + r$, with $\tau(A) = p$, $q = [n/p]$ and $0 \leq r < p$. Thus

$$A = \left\{ X_0^{p-1} = X_p^{2p-1} = \ldots = X_{(q-1)p}^{qp-1} = a_0^{p-1} \ ; \ X_{qp}^{n-1} = a_0^{r-1} \right\} \ .$$

For instance

$$A = (\overbrace{\texttt{aaaabb}}^{\text{period}}\overbrace{\texttt{aaaabb}}^{\text{period}}\overbrace{\texttt{aaa}}^{\text{rest}}) \ . \tag{1}$$

Thus, consider the set of overlapping positions of A:

$$\mathcal{O}(A) = \left\{ k \in \{1, \ldots, n-1\} \mid A \cap \ T^{-k}(A) \neq \emptyset \right\} \ .$$

Split $\mathcal{O}(A)$ in a disjoint union of $\{\tau(A), \ldots, [n/\tau(A)]\tau(A)\}$ and $\mathcal{R}(A)$ where

$$\mathcal{R}(A) = \left\{ k \in \{[n/\tau(A)]\tau(A) + 1, \ldots, n-1\} \mid A \cap \ T^{-k}(A) \neq \emptyset \right\} \ .$$

Put $r_A = \#\mathcal{R}(A)$. The cardinal of $\mathcal{O}(A)$ is then $\sigma(A) = [n/\tau(A)] + r_A$.

4 Poisson approximation

For $1 \le t' < t$ integers, let $N_{t'}^t = \sum_{i=t'}^t 1\!\!1_{T^{-i}A}$, so that, $N_{t'}^t$ counts the number of occurrences of A between t' and t. For the sake of simplicity we write $N_t = N_1^t$. With some abuse of notation we also put $(-1)! = 1$.

Theorem 2 *Let $(X_m)_{m\in Z}$ be a ϕ-mixing process. There exists a constant $C > 0$, such that for all $A \in \mathcal{C}_n$, and all non negative integer k, the following inequality holds:*

$$\left| I\!\!P\left(N_{t/I\!\!P(A)} = k\right) - \frac{e^{-t}t^k}{k!}\right| \le Ce(A)g(A,k) \ ,$$

with $e(A) = e_1(A) + e_2(A)$,

$$e_1(A) \stackrel{def}{=} \inf_{1\le w\le \tau(A)} \left[(\sigma(A)+n)I\!\!P(A^{(w)}) + \phi\left((\tau(A)) - w\right)\right] \ , \tag{2}$$

$$e_2(A) \stackrel{def}{=} \inf_{n\le \ell\le 1/I\!\!P(A)} \left[\ell I\!\!P(A) + \frac{\phi(\ell)}{I\!\!P(A)}\right] + \phi(n) \ , \tag{3}$$

and

$$g(A,k) \stackrel{def}{=} \begin{cases} \frac{(2\lambda)^{k-1}}{(k-1)!} & k \notin \left\{\frac{\lambda}{e(A)}, \dots, \frac{t}{I\!\!P(A)}\right\} \\ \frac{(2\lambda)^{k-1}}{\left(\frac{\lambda}{e(A)}\right)! \left(\frac{1}{e(A)}\right)^{k-\frac{1}{e(A)}-1}} & k \in \left\{\frac{\lambda}{e(A)}, \dots, \frac{t}{I\!\!P(A)}\right\} \end{cases} ,$$

where $\lambda \stackrel{def}{=} t\left[1 + \frac{\phi(\ell_A)}{I\!\!P(A)}\right]$ and ℓ_A is the ℓ that defines $e_2(A)$.

Below we state several remarks to better understand the error term of the theorem.

Remark 3 *Clearly $e(A)$ is the uniform error term and $g(A,k)$ is the error factor that provides the control on the tail distribtion.*

Remark 4 *$e_1(A)$ is the error that arises from the short correlations of the process while $e_2(A)$ is the error that arises from long ones.*

Remark 5 *$I\!\!P(A_n) \le Ce^{-cn}$ (see [1].) $\phi(n)$ goes to zero by hypothesis. Therefore $e_1(A)$ is small if $\tau(A)$ is large enough to chose a w between 1 and $\tau(A)$ such that Ce^{-cw} and $\phi(\tau(A) - w)$ are small.*

Remark 6 *Take a sequence of n-strings A_n with n diverging. $e_1(A) \to 0$ if $\tau(A_n)$ also diverges with n faster than $\ln n$ (since $I\!\!P(A_n)$ decays exponentially fast.)*

Remark 7 *$e_2(A) \to 0$ as $n \to \infty$ if the sequence $\phi(\ell)$ is summable.*

Remark 8 *Collet et al. ([11]) proved that for exponentially ψ-mixing processes there exist positive constants C and c such that $I\!\!P(A \in \mathcal{C}_n \ ; \ \tau_A \leq n/3) \leq Ce^{-cn}$. Abadi ([1]) proved that for ϕ-mixing processes $I\!\!P(A \in \mathcal{C}_n \ ; \ \tau_A \leq sn) \leq Ce^{-cn}$ with $s \in (0,1)$ that just depends on the cardinality of the alphabet of the process. Abadi and Vaienti ([3]) proved the above inequality for Gibbs measures and for* any *value of s (with $c = c(s)$.) This shows that Theorem 2 holds for typical (in the sense of $\tau(A)$) strings. Taking limit on the length of the strings along infinite sequences, we get that the Poisson limit law holds almost everywhere.*

Remark 9 *When $\tau(A)$ is not large enough, the return time is better approximated by a mixture of a Dirac measure at the origin and an exponential law as shown by Abadi and vergne ([5], Theorem 2). Therefore, the numbers of occurrences of the string can not be Poisson distributed.*

Remark 10 *When $e_2(A)$ is small, so is $\phi(\ell)/I\!\!P(A)$. Therefore λ is just the parameter of the Poisson law with a small correction factor $1 + \phi(\ell)/I\!\!P(A)$. Thus $\lambda/e(A)$ is a large number (smaller or equal to $t/I\!\!P(A)$.) For $k \leq \lambda/e(A)$ or $k \geq t/I\!\!P(A)$ we get tha $g(A,k)$ decays factorially fast. For k in the strip $\lambda/e(A)$ to $t/I\!\!P(A)$ we don't get $k!$ but something that we could call "truncated facorial": just get $(1/e(A))!$ times $k - (1/e(A))$ factors $1/e(A)$.*

The point-wise error term given in Theorem 2 allows us to estimate the moments of $N_{t/I\!\!P(A)}$ by those of a r.v. with Poisson distribution.

Corollary 11 *Let $\beta > 0$. Let Z be a r.v. with Poisson distribution of parameter $t > 0$. Under the conditions of Theorem 2*

$$\left| I\!\!E\left(N^{\beta}_{t/I\!\!P(A)}\right) - I\!\!E(Z^{\beta})\right| \leq C_{\beta}\ e(A) \ ,$$

where C_{β} is a constant that just depends on β.

4.1 Preparatory results

The next lemma says that the occurrence of two copies of A very close have small probability.

Lemma 12 *Let $(X_m)_{m\in Z}$ be a ϕ-mixing process. Then, for all $A \in \mathcal{C}_n$ the following inequalities hold:*

$$I\!\!P_A\left(\bigcap_{j=1}^{2n-1} T^{-j}A\right) \leq e_1(A) \ ,$$

and all $\ell \geq 2n$

$$I\!\!P_A\left(\bigcap_{j=2n}^{\ell} T^{-j}A\right) \leq \ell I\!\!P(A) + \phi(n) \ ,$$

Proof By the overlapping properties of A one has

$$A \bigcap_{j=1}^{2n-1} T^{-j}A = A \cap \left(\bigcap_{j \in \mathcal{O}(A)} \bigcap_{j=n}^{2n-1} T^{-j}A \right) .$$

The cardinal of the intersection is $\sigma(A) + n$. Now since $T^{-j}A \subseteq T^{-j}A^{(w)}$ for any $1 \le w \le \tau(A)$, the first part of the lemma follows using the ϕ-mixing property with $V = A$ and $W = \bigcap_{j\in\mathcal{O}(A)} \bigcap_{j=n}^{2n-1} T^{-j}A^{(w)}$.

The cardinal of the intersection in the second statement of the lemma is $\ell - n + 1$. The second part of the lemma follows using the ϕ-mixing property as in the previous case. □

Definition 13 *Given $A \in \mathcal{C}_n$, and $j \in I\!N$, we define the j-*occurrence time *of A as the r.v. $\tau_A^{(j)} : \Omega \to I\!N \cup \{\infty\}$, defined as follows: For any $x \in \Omega$, define $\tau_A^{(1)}(x) = \inf\{k \ge 1 : T^k(x) \in A\}$ and for $j \ge 2$*

$$\tau_A^{(j)}(x) = \inf\{k > \tau_A^{(j-1)}(x) : T^k(x) \in A\} .$$

The next proposition says that the measure of all the configurations where there are no two occurrences of A very close, is close to the product measure.

Proposition 14 *Let $(X_m)_{m\in Z}$ be a ϕ-mixing process. Then, for all $A \in \mathcal{C}_n$, all $0 \le t_1 < t_2 < \ldots < t_k \le t$, and all $k \in I\!N$, for which*

$$\min_{2\le j\le k} \{t_j - t_{j-1}\} > 2(\ell_A + n) ,$$

(ℓ_A defined in Theorem 2) the following inequality holds:

$$\left| I\!P\left(\bigcap_{j=1}^{k} \tau_A^{(j)} = t_j ; \tau_A^{(k+1)} > t \right) - I\!P(A)^k \prod_{j=1}^{k+1} I\!P\left(t_j - t_{j-1} - 2(\ell_A + n)\right) \right|$$
$$\le \quad 5k \left(I\!P(A) + \phi(\ell_A)\right)^k e(A) .$$

Proof We prove the proposition by induction on k. For shorthand notation put $\overline{\ell}_A = 2(\ell_A + n)$, $\Delta_1 = t_1$, $\Delta_{k+1} = t - t_k$, $\Delta_i = t_i - t_{i-1}$ and $\mathcal{P}_i = I\!P\left(\tau_A > \Delta_i - \overline{\ell}_A\right)$; $i = 1, \ldots, k+1$.

For $k = 1$, the triangle inequality gives

$$\left| I\!P\left(\tau_A = t_1 \ ; \ \tau_A^{(2)} > t\right) - I\!P(A) \prod_{j=1}^{2} \mathcal{P}_j \right| \tag{4}$$
$$\le \quad \left| I\!P\left(\tau_A = t_1 \ ; \ \tau_A^{(2)} > t\right) - I\!P\left(\tau_A = t_1 \ ; \ N_{t_1+\ell_A+n}^{t} = 0\right) \right| \tag{5}$$
$$+ \quad \left| I\!P\left(\tau_A = t_1 \ ; \ N_{t_1+\ell_A+n}^{t} = 0\right) - I\!P\left(\tau_A = t_1\right) \mathcal{P}_2 \right| \tag{6}$$
$$+ \quad \left| I\!P\left(A \ ; \ \tau_A > t_1 - 1\right) - I\!P\left(A \ ; \ N_{n+\ell_A}^{t_1-1} = 0\right) \right| \ \mathcal{P}_2 \tag{7}$$
$$+ \quad \left| I\!P\left(A \ ; \ N_{n+\ell_A}^{t_1-1} = 0\right) \mathcal{P}_2 - I\!P(A) \prod_{j=1}^{2} \mathcal{P}_j \right| . \tag{8}$$

In (7) we used that by stationarity $\mathbb{P}(\tau_A = t) = \mathbb{P}(A; \tau_A > t-1)$. Term (5) is equal to

$$\begin{aligned}
& \mathbb{P}\left(\tau_A = t_1; \bigcup_{i=t_1+1}^{t_1+\ell_A+n-1} T^{-i}A; N^t_{t_1+\ell_A+n} = 0\right) \\
\leq \; & \mathbb{P}\left(T^{-t_1}A; \bigcup_{i=t_1+1}^{t_1+\ell_A+n-1} T^{-i}A\right) \\
= \; & \mathbb{P}\left(A; \bigcup_{i=1}^{\ell_A+n-1} T^{-i}A\right) \; . \qquad (9)
\end{aligned}$$

We divide the above sum in those terms with $1 \leq i < 2n$, and $2n \leq i \leq \ell_A + n$. Lemma 12 implies

$$\mathbb{P}\left(A; \bigcup_{i=1}^{2n-1} T^{-i}A\right) \leq \mathbb{P}(A)e_1(A) \; .$$

and,

$$\mathbb{P}\left(A; \bigcup_{i=2n+1}^{\ell_A+n} T^{-i}A\right) \leq \mathbb{P}(A)\left(\ell_A \mathbb{P}(A) + \phi(n)\right) \; .$$

Term (6) is bounded using the mixing property by $\phi(\ell_A)\mathbb{P}(A)$. Analogous computations are used to bound terms (7) and (8). This shows that (4) is bounded by $2e(A)\mathbb{P}(A)$.

Now let us suppose that the proposition holds for $k-1$ and let us prove it for k. We use a triangle inequality where the terms involved are defined below. We briefly comment the idea behind each term. For brevity denote for each non negative i, $\mathcal{S}_i = \left\{\tau_A^{(i)} = t_i\right\}$. Thus we have

$$\left|\mathbb{P}\left(\bigcap_{j=1}^{k} \mathcal{S}_j \; ; \; \tau_A^{(k+1)} > t\right) - \mathbb{P}(A)^k \prod_{j=1}^{k+1} \mathcal{P}_j\right| \leq I + II + III + IV + V \; .$$

In I we open a gap of length $\ell_A + n$ at the left of the k-th occurrence of A, namely, between coordinates $t_k - (\ell_A + n)$ and $t_k - 1$.

$$\begin{aligned}
I \; & \stackrel{def}{=} \; \left|\mathbb{P}\left(\bigcap_{j=1}^{k} \mathcal{S}_j; \tau_A^{(k+1)} > t\right) - \mathbb{P}\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j; N^{t_k-(\ell_A+n)}_{t_{k-1}+1} = 0; T^{-t_k}A; N^t_{t_k+1} = 0\right)\right| \\
& = \; \mathbb{P}\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j; N^{t_k-(\ell_A+n)}_{t_{k-1}+1} = 0; \bigcup_{i=t_k-(\ell_A+n)+1}^{t_k-1} T^{-i}A; T^{-t_k}A; N^t_{t_k+1} = 0\right) \qquad (10) \\
& \leq \; \mathbb{P}\left(\bigcap_{j=1}^{k-1} T^{-t_j}A \; ; \bigcup_{i=t_k-(\ell_A+n)+1}^{t_k-1} T^{-i}A \; ; \; T^{-t_k}A\right) \; .
\end{aligned}$$

As with (9) we split the above sum in terms with $t_k-(\ell_A+n)+1 \le i \le t_k-2n$, $t_k-2n+1 \le i \le t_k-1$. We recall that by hypothesis $\Delta_i > \overline{\ell}_A$ for all $i=1,\ldots,k$. As in Lemma 12 we have

$$
\begin{aligned}
& I\!\!P\left(\bigcap_{j=1}^{k-1} T^{-t_j}A; \bigcup_{i=t_k-(\ell_A+n)+1}^{t_k-2n} T^{-i}A; T^{-t_k}A \right) \\
\le \ & I\!\!P\left(\bigcap_{j=1}^{k-1} T^{-t_j}A; \bigcup_{i=t_k-(\ell_A+n)+1}^{t_k-2n} T^{-i}A \right) \left(I\!\!P(A) + \phi(n) \right) .
\end{aligned}
$$

The ϕ-mixing property over the left most factor in the right hand side of the above inequality, we get that it is bounded by

$$
I\!\!P\left(\bigcap_{j=1}^{k-1} T^{-t_j}A \right) \left(\ell_A I\!\!P(A) + \phi(\ell_A) \right) .
$$

Iterating this procedure we get

$$
I\!\!P\left(\bigcap_{j=1}^{k-1} T^{-t_j}A \right) \le \left(I\!\!P(A) + \phi(\ell_A) \right)^{k-1} .
$$

Similarly

$$
I\!\!P\left(\bigcap_{j=1}^{k-1} T^{-t_j}A; \bigcup_{i=t_k-2n+1}^{t_k-1} T^{-i}A; T^{-t_k}A \right) \le \left(I\!\!P(A) + \phi(\ell_A) \right)^{k} e_1(A) .
$$

In II we apply the ϕ-mixing property to factorize the probability in the right hand side of the modulus in I. Then we iterated the ϕ-mixing property to obtain the last inequality.

$$
\begin{aligned}
II \ & \stackrel{def}{=} \ \left| I\!\!P\left(\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j; N_{t_{k-1}+1}^{t_k-(\ell_A+n)} = 0 \right); \left(T^{-t_k}A; N_{t_k+1}^{t} = 0 \right) \right) - \right. \\
& \left. - I\!\!P\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j \ ; \ N_{t_{k-1}+1}^{t_k-(\ell_A+n)} = 0 \right) I\!\!P\left(A \ ; \ N_1^{t-t_k} = 0 \right) \right| \\
& \le \ I\!\!P\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j \ ; \ N_{t_{k-1}+1}^{t_k-(\ell_A+n)} = 0 \right) \phi(\ell_A) \\
& \le \ I\!\!P\left(\bigcap_{j=1}^{k-1} T^{-t_j}A \right) \phi(\ell_A) \\
& \le \ \left(I\!\!P(A) + \phi(\ell_A) \right)^{k} \frac{\phi(\ell_A)}{I\!\!P(A)} .
\end{aligned}
$$

In III we "fill-up" the gap we opened in I

$$\begin{aligned} III &\stackrel{def}{=} \left| I\!P\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j ; N_{t_{k-1}+1}^{t_k-(\ell_A+n)} = 0\right) - I\!P\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j ; N_{t_{k-1}+1}^{t_k-1} = 0\right)\right| I\!P\left(A; N_1^{t-t_k}=0\right) \\ &\leq I\!P\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j \ ; \ N_{t_{k-1}+1}^{t_k-(\ell_A+n)} = 0 \ ; \ \bigcup_{t_k-(\ell_A+n)+1}^{t_k-1} T^{-i}A\right) I\!P(A) \\ &\leq I\!P\left(\bigcap_{j=1}^{k-1} T^{-t_j}A; \bigcup_{i=t_k-(\ell_A+n)+1}^{t_k-1} T^{-i}A\right) I\!P(A) \\ &\leq \left(I\!P(A)+\phi(\ell_A)\right)^k 2\ell_A I\!P(A) \ . \end{aligned}$$

In IV we use the inductive hypothesis

$$\begin{aligned} IV &\stackrel{def}{=} \left| I\!P\left(\bigcap_{j=1}^{k-1} \mathcal{S}_j ; N_{t_{k-1}+1}^{t_k-1} = 0\right) - I\!P(A)^{k-1}\prod_{j=1}^{k}\mathcal{P}_j\right| I\!P\left(A; N_1^{t-t_k}=0\right) \\ &\leq C(k-1)\left(I\!P(A)+\phi(\ell_A)\right)^{k-1} e(A) I\!P(A) \ . \end{aligned}$$

In V we use $k=1$ to get

$$\begin{aligned} V &\stackrel{def}{=} I\!P(A)^{k-1}\prod_{j=1}^{k}\mathcal{P}_j \left| I\!P\left(A \ ; \ N_1^{t-t_k}=0\right) - I\!P(A)\mathcal{P}_{k+1}\right| \\ &\leq I\!P(A)^k \ 2e(A) \ . \end{aligned}$$

Summing the bounds above we end the proof of the proposition. □

4.2 Proof of Theorem 2 and Corollary 11.

Proof of Theorem 2. Take $t \in I\!N$. Let us write for the sake of simplicity $N = N_t$. For $k = 0$ note that $I\!P(N=0) = I\!P(\tau_A > t)$. By Theorem 1 in Abadi ([2]) one has

$$|I\!P(\tau_A > t)\, e^{-\xi_A I\!P(A)t}| \leq e(A)(I\!P(A)t \vee 1)e^{-\xi_A I\!P(A)t} \ , \tag{11}$$

with a certain $\xi_A > 0$. Moreover, it follows in the proof of Theorem 2 in Abadi and Vergne ([4]) that $|\xi_A - \zeta_A| \leq e_1(A)$ where $\zeta_A = I\!P_A(\zeta_A > \tau(A))$. Finally $|\zeta_A - 1| = I\!P_A(\zeta_A = \tau(A)) \leq e_1(A)$ by Lemma 12. This concludes the proof for $k = 0$.

For $k > t$ we have that $I\!P(N=k) = 0$. Then

$$\begin{aligned} \left| I\!P(N=k) - \frac{e^{-tI\!P(A)}(tI\!P(A))^k}{k!}\right| &= \frac{e^{-tI\!P(A)}(tI\!P(A))^k}{k!} \\ &\leq \frac{(tI\!P(A))^{k-1}}{(k-1)!}\ I\!P(A) \ . \end{aligned}$$

Let us consider k with $1 \le k \le t$. The idea of the proof is the following: Consider a realization $x = (x_m)_{m \in Z}$ of the process $(X_m)_{m \in Z}$ such that the sequence $(x_1, \ldots, x_t)$ contains exactly k occurrences of A. These occurrences can appear in clusters or isolated one from each other. We prove that realizations with aisolated A's give the approximation to the Poisson law and realizations with clustered A's have small measure. We now formalize this idea. Given $1 \le t_1 < \ldots < t_k \le t$, let us define the following measurable set:

$$\mathcal{T}(t_1, \ldots, t_k) = \bigcap_{j=1}^{k} \left\{ \tau_A^{(j)} = t_j \right\} \bigcap \left\{ \tau_A^{(k+1)} > t \right\} .$$

As in Proposition 14 we put $\Delta_j = t_j - t_{j-1}$, for $j = 2, \ldots, k$. Put also $\Delta_1 = t_1$ and $\Delta_{k+1} = t - t_k$. Define $I(\mathcal{T}(t_1, \ldots, t_k)) = \min \{\Delta_j \mid 2 \le j \le k\}$. As before put $\bar{\ell}_A = 2(\ell_A + n)$. Let us divide $\{N = k\}$ in two sets

$$B_k = \bigcup_{I(\mathcal{T}(t_1,\ldots,t_k)) < \bar{\ell}_A} \mathcal{T}(t_1, \ldots, t_k) \quad \text{and} \quad G_k = \bigcup_{I(\mathcal{T}(t_1,\ldots,t_k)) \ge \bar{\ell}_A} \mathcal{T}(t_1, \ldots, t_k) .$$

Since $\{N = k\} = B_k \cup G_k$, disjoint union, we have

$$\begin{aligned} & \left| I\!P\,(N = k) - \frac{e^{-tI\!P(A)} t^k I\!P(A)^k}{k!} \right| \\ \le \quad & I\!P\,(B_k) + \left| I\!P\,(G_k) - \frac{e^{-tI\!P(A)} t^k I\!P(A)^k}{k!} \right| . \end{aligned} \tag{12}$$

We will prove that both quantities in the right hand side of (12) are small.

Proof: configurations with clusters have small measure.

We will prove an upper bound for $I\!P\,(B_k)$. First observe that if we fix t_1 then

$$\bigcup_{t_2,\ldots,t_k} \mathcal{T}(t_1, t_2, \ldots, t_k) \subseteq \bigcup_{\substack{i=2,\ldots,k \\ t_i = t_{i-1},\ldots,t_{i-1}+\bar{\ell}_A}} \bigcap_{j=1}^{k} T^{-t_j} A .$$

Now define

$$C(\mathcal{T}(t_1, \ldots, t_k)) = \sum_{j=2}^{k} 1\!\!1_{\{\Delta_j > \bar{\ell}_A\}} + 1 .$$

The above quantity compute how many clusters there are in a given $\mathcal{T}(t_1, \ldots, t_k)$. Suppose that $C(\mathcal{T}(t_1, \ldots, t_k)) = 1$ and fix the position t_1. Each occurrence inside the unique cluster (with the exception of the most left one which is fixed at t_1) can appear at distance d of the previous one, with $d \in \mathcal{O}(A)$ or $n \le d \le \bar{\ell}_A$. Therefore, the iterative argument of the ϕ-mixing property used to bound (10) leads to the bound

$$\begin{aligned} & I\!\!P\left(\bigcup_{\substack{i=2,\ldots,k \\ t_i=t_{i-1},\ldots,t_{i-1}+\overline{\ell}_A}} \bigcap_{j=1}^{k} T^{-t_j} A \right) \qquad (13) \\ \leq\ & I\!\!P(A)\ \left(e_1(A) + \overline{\ell}_A I\!\!P(A) + \phi(n)\right)^{k-1} \\ \leq\ & I\!\!P(A) e(A)^{k-1} \ . \end{aligned}$$

Suppose now that $C(\mathcal{T}(t_1,\ldots,t_k)) = i$. Assume also that the most left occurrence of these i clusters occurs at $1 \leq t(1) < \ldots < t(i) \leq t$ fixed. By the same argument used in (13), we have the inequalities

$$\begin{aligned} & I\!\!P\left(\bigcup_{\{t_1,\ldots,t_k\}\setminus\{t(1),\ldots,t(i)\}} \mathcal{T}(t_1,\ldots,t_k) \right) \\ \leq\ & I\!\!P(A) \left(I\!\!P(A) + \phi(\ell_A)\right)^{i-1} e(A)^{k-i} \\ \leq\ & \left(I\!\!P(A) + \phi(\ell_A)\right)^{i} e(A)^{k-i} \ . \qquad (14) \end{aligned}$$

To obtain an upper bound for $I\!\!P\left(B_k\right)$ we must sum the above bound over all $\mathcal{T}(t_1,\ldots,t_k)$ such that $C(\mathcal{T}(t_1,\ldots,t_k)) = i$ with i that runs from 1 to $k-1$.

Fixed $C(\mathcal{T}(t_1,\ldots,t_k)) = i$, the locations of the most left occurrences of A of each one of the i clusters can be chosen at most in $\binom{t}{i}$ many ways.

The cardinality of each one of the i clusters can be arranged in $\binom{k-1}{i-1}$ many ways. (This corresponds to break the interval $(1/2, k+1/2)$ in i intervals at points chosen from $\{1+1/2,\ldots,k-1/2\}$.)

Collecting these information and (14) we have that $I\!\!P\left(B_k\right)$ is bounded by

$$\sum_{i=1}^{k-1} \binom{t}{i} \binom{k-1}{i-1} \left(I\!\!P(A) + \phi(\ell_A)\right)^{i} e(A)^{k-i} \leq e(A)^k \max_{1\leq i\leq k-1} \frac{\gamma^i}{i!} \sum_{i=1}^{k-1} \binom{k-1}{i-1},$$

where $\gamma = tI\!\!P(A)\left[1 + \phi(\ell_A)/I\!\!P(A)\right]/e(A)$. The maximum in the above expression is reached at $(k-1 \wedge \gamma)$. The most right sum is bounded by 2^{k-1}. Therefore we have

$$I\!\!P\left(B_k\right) \leq e(A). \begin{cases} \frac{(2\gamma e(A))^{k-1}}{(k-1)!} & k-1 < \gamma \\ \frac{2^{k-1}(\gamma e(A))^{\gamma}}{\gamma!\left(\frac{1}{e(A)}\right)^{k-\gamma-1}} & k \geq \gamma \end{cases} . \qquad (15)$$

This ends the proof of the bound for $I\!\!P\left(B_k\right)$.

Proof: A's isolated provide the Poisson limit law.

We can bound the most right term on the right-hand side of (12) by the following triangular inequality:

$$\sum_{\mathcal{T}(t_1,\ldots,t_k)\in G_k} \left| I\!P\left(\bigcap_{j=1}^{k} \tau_A^{(j)} = t_j \ ; \ \tau_A^{(k+1)} > t\right) - I\!P(A)^k \prod_{j=1}^{k+1} \mathcal{P}_j \right| \tag{16}$$

$$+ \quad I\!P(A)^k \sum_{\mathcal{T}(t_1,\ldots,t_k)\in G_k} \left| \prod_{j=1}^{k+1} \mathcal{P}_j - \prod_{j=1}^{k+1} e^{-(\Delta_j - \overline{\ell}_A) I\!P(A)} \right| \tag{17}$$

$$+ \quad I\!P(A)^k \# G_k \left| e^{-(t-(k+1)\overline{\ell}_A) I\!P(A)} - e^{-t I\!P(A)} \right| \tag{18}$$

$$+ \quad \left| \frac{\# G_k \ k!}{t^k} - 1 \right| \ \frac{e^{-t I\!P(A)} t^k I\!P(A)^k}{k!} \ . \tag{19}$$

By a simple combinatorial argument we get the bounds

$$\frac{(t - k(n + \overline{\ell}_A))^k}{k!} \le \binom{t - k(n + \overline{\ell}_A - 1) - 1}{k} \le \# G_k \le \binom{t}{k} \le t^k / k! \ . \tag{20}$$

Moreover, the leading term in (16) is bounded using Proposition 14. Thus (16) is bounded by

$$5 \ \frac{t^k}{(k-1)!} \ (I\!P(A) + \phi(\ell_A))^k \ e(A) \ .$$

The difference between the leading factors in (17) is bounded as follows: again by (11)

$$|\mathcal{P}_j - e^{-\xi_A I\!P(A)(\Delta_j - \overline{\ell}_A)}| \le C e_1(A) \ .$$

As stated at the beginning of the proof one has $|\xi_A - 1| \le e_1(A)$. Therefore (17) is bounded by

$$\frac{t^k}{k!} I\!P(A)^k (k+1) \max_{1 \le j \le k+1} \left| \mathcal{P}_j - e^{-(\Delta_j - \overline{\ell}_A) I\!P(A)} \right| \le \frac{k+1}{k} \frac{(t I\!P(A))^k}{(k-1)!} C e_1(A) \ .$$

(18) is bounded using the Mean Value Theorem by

$$\frac{t^k I\!P(A)^k}{k!} (k+1) \overline{\ell}_A I\!P(A) \le \frac{k+1}{k} \frac{(t I\!P(A))^k}{(k-1)!} 4 \ell_A I\!P(A) \ .$$

The left hand side of (20) and the Mean Value Theorem provide a bound for the difference below

$$\left| \frac{\# G_k \ k!}{t^k} - 1 \right| \le \left| \frac{(t - k(n + \overline{\ell}_A))^k}{t^k} - 1 \right| \le \frac{k \ k(n + \overline{\ell}_A)}{t} \le k \ .$$

So, (19) is bounded by

$$\frac{(t I\!P(A))^k}{(k-1)!} \ 4 \ell_A I\!P(A) \ .$$

Summing the bounds obtained for (16), (17), (18) and (19) we get the desired bound for the difference in the right hand term of inequality (12). The exchange of variables $\widetilde{t} = t\mathbb{P}(A)$ ends the proof of the theorem. □

Proof of Corollary 11. By definition

$$\begin{aligned}\left|\mathbb{E}\left(N^{\beta}_{t/\mathbb{P}(A)}\right) - \mathbb{E}(Z^{\beta})\right| &= \left|\sum_{k\geq 0} k^{\beta}\mathbb{P}\left(N_{t/\mathbb{P}(A)} = k\right) - \sum_{k\geq 0} k^{\beta}\frac{e^{-t}t^{k}}{k!}\right| \\ &\leq \sum_{k\geq 0} k^{\beta}\left|\mathbb{P}\left(N_{t/\mathbb{P}(A)} = k\right) - \frac{e^{-t}\,t^{k}}{k!}\right| .\end{aligned}$$

The summability in k of k^{β} times the error term in Theorem 2 ends the proof of the corollary. □

5 α-mixing processes

Theorem 15 *Let $(X_m)_{m\in Z}$ be α-mixing process. For all $\tau(A) \geq C_1 n$ with a constant $C_1 \in (0,1)$, the following inequality holds:*

$$\left|\mathbb{P}\left(N_{t/\mathbb{P}(A)} = k\right) - \frac{e^{-t}t^{k}}{k!}\right| \leq C_2 e^{\alpha}_{k}(A) g^{\alpha}(A,k),$$

with

$$g^{\alpha}(A,k) \stackrel{def}{=} \begin{cases} \frac{2^{k-1}}{(k-1)!} & k \notin \left\{\frac{t}{e^{\alpha}(A)},\ldots,\frac{2t}{n\mathbb{P}(A)}\right\} \\ \dfrac{2^{k-1}}{\left(\frac{1}{e^{\alpha}_1(A)}\right)!\,\left(\frac{1}{e^{\alpha}_1(A)}\right)^{k-\frac{1}{e^{\alpha}_1(A)}-1}} & k \in \left\{\frac{t}{e^{\alpha}_1(A)},\ldots,\frac{2t}{n\mathbb{P}(A)}\right\} \end{cases},$$

and

$$e^{\alpha}_{k}(A) \stackrel{def}{=} \inf_{1\leq w\leq n_A}\left\{(r_A + n_A)\mathbb{P}(A^{(w)}) + \frac{\alpha\left(n_A - w\right)}{\mathbb{P}(A)^{k}}\right\} .$$

Furthermore, assume that for a fixed $k \in \mathbb{N}$, $e^{\alpha}_{k}(A_n) \to 0$ as $n \to \infty$. Then, $N_{t/\mathbb{P}(A_n)}$ converges in distribution to a Poisson law for $N_{t/\mathbb{P}(A_n)} = j$, for all $0 \leq j \leq k$.

Remark 16 *The condition $\alpha(n_A - w)/\mathbb{P}(A_n)^k \to 0$ as $n \to \infty$ for all k basically means that α must decay super-exponentially fast in order to have convergence for* all *the values of k. However, we usually are interested in the convergence for not too large values of k, say $k \leq C$, for a certain positive constant C. In that case, in order to have convergence we need an α decaying exponentially fast with constant c such that $c/m > C(h + \epsilon)$, with $\epsilon > 0, m > 2$, where h is the entropy of the process, where we used the Shannon-Mac-Millan-Breiman Theorem ([12]) to make $\mathbb{P}(A) \approx e^{-hn}$ for almost every A.*

Proof Take a realization with k occurrences of A at $t_1, \ldots, t_k$. Assume first that $k = 2$. Take any w with $1 \le w \le (t_2 - t_1 \wedge n)$. The α-mixing property gives the upper bound

$$I\!P(T^{-t_1}A \cap T^{-t_2}A) \le \begin{cases} I\!P(A)I\!P(A^w) + \alpha(t_2 - t_1 - w) & \text{for } t_2 - t_1 < 2n \\ I\!P(A)^2 + \alpha(t_2 - t_1 - n) & \text{for } t_2 - t_1 \ge 2n \end{cases} .$$

Now choose w_i as w above for each $i = 2, \ldots, k$ (or $w_i = n$ if $t_2 - t_1 \ge 2n$). Put $w_{\min} = \min\{w_i \mid 2 \le i \le k\}$. Iterating the above procedure of the α-mixing property one has

$$\begin{aligned} I\!P\left(\bigcap_{i=1}^{k} T^{-t_i}A\right) &\le I\!P(A)\prod_{i=2}^{k} I\!P(A^{(w_i)}) + \sum_{j=0}^{k-1}\alpha(t_j - t_{j-1} - w_j)\prod_{i=0}^{j} I\!P(A^{(w_i)}) \\ &\le I\!P(A)I\!P(A^{(w_{\min})})^{k-1} + C\alpha(n_A - w_{\min}). \end{aligned}$$

This bound can be used to prove a similar result to Proposition 14 in the α-mixing context. Theorem 1 in [2] can be replaced by Theorem 17 in the same paper which establishes that

$$\sup_{t\ge 0} |I\!P(\tau_A > t) - e^{-\lambda(A)I\!P(A)t}| \le n\sqrt{I\!P(A)} ,$$

with a certain parameter $\lambda(A)$. It is easy to follows the proof of Lemma 19 in the same paper to show that $|\lambda(A) - \zeta_A| \le e_1^{\alpha}(A)$. The condition $\tau(A) \ge C_1 n$ implies that $|\zeta_A - 1| \le e_1^{\alpha}(A)$. The rest of the proof follows as in the ϕ-mixing case. □

Acknowledgements MA thanks Ministere de la Recherche et Ministere de l'Education for support during part of this work. The authors thank P. Ferrari, A. Galves, B. Prum and P. Shields for useful discussions.

References

[1] M. Abadi. Exponential approximation for hitting times in mixing processes. *Math. Phys. Elec. J.* **7**, 2, (2001).

[2] M. Abadi. Sharp error terms and necessary conditions for exponential hitting times in mixing processes. *Ann. Probab.* **32**, 1A, (2004) 243-264.

[3] M. Abadi and S. Vaienti. Large deviation for short recurrence. *Preprint.* (2006).

[4] M. Abadi and N. Vergne. Sharp error terms for return time statistics under mixing conditions. *Preprint.* (2006).

[5] M. Abadi and N. Vergne. Poisson approximation in biological context. *Preprint.* (2006).

[6] R. Arratia, L. Goldstein and L. Gordon. Two moments suffice for Poisson approximation: the Chen-Stein method. *Ann. Probab,* **17** 9 - 25 (1989).

[7] R. Arratia, L. Goldstein and L. Gordon. Poisson approximation and the Chen-Stein method. *Stat. Science,* **5** 4. 403 - 434 (1990).

[8] Barbour A. D. Barbour, L. Holst and S. Janson. Poisson approximation. *Oxford, University Press, New York* (1992).

[9] R. Bowen. Equilibrium states and the ergodic theory of Anosov diffeomorphisms. *Lecture Notes in Math,* **470.** *Springer-Verlag, New York.* (1975).

[10] L. H. Y. Chen (1975). Poisson approximation for dependent trials. *Ann. Prob.* **3**, 534-545.

[11] P. Collet, A. Galves and B. Schmitt. Repetition times for gibbsian sources. *Nonlinearity* **12**, (1999) 1225-1237.

[12] I. Cornfeld, S. Fomin and Y. Sinai. Ergodic theory. *Grundlähren der Mathematischen Wissenschaften,* **245.** *Springer-Verlag, New York.* (1982).

[13] W. Doeblin. Remarques sur la théorie métrique des fractions continues. *Compositio Math.* **7**, (1940) 353-371.

[14] P. Doukhan. Mixing. Properties and examples. *Lecture Notes in Statistics* **85,** *Springer-Verlag.* (1995).

[15] A. Galves and B. Schmitt. Inequalities for hitting times in mixing dynamical systems. *Random Comput. Dyn.* **5**, (1997) 337-348.

[16] N. Haydn and S. Vaienti. The limiting distribution and error terms for return times of dynamical systems. *Discrete Contin. Dyn. Syst.***10**, 3, (2004) 589–616.

[17] M. Hirata. Poisson law for Axiom A diffeomorphism. *Ergod. Th. Dyn. Sys.* **13**, (1993) 533-556.

[18] P. Walters. Ruelle's operator theorem and g-measures. *Trans. Amer. Math. Soc.* **214**, (1975) 375-387.