

Article

Potential Well in Poincaré Recurrence

Miguel Abadi ^{1,*}, Vitor Amorim ²  and Sandro Gallo ^{3,*}

¹ Instituto de Matemática e Estatística, Universidade de São Paulo-USP, Rua do Matão, 1010 Butantã, São Paulo 05508-090, Brazil

² Instituto Federal de Educação, Ciência e Tecnologia de São Paulo-IFSP, Câmpus Araraquara, Rua Doutor Aldo Benedito Pierri, 250 Jardim Paulo Freire, Araraquara 14804-296, Brazil; vitoramorim@usp.br

³ Departamento de Estatística, Universidade Federal de São Carlos-UFSCar, Rua dos Bem-te-vis, São Carlos 13565-905, Brazil

* Correspondence: leugim@ime.usp.br (M.A.); sandro.gallo@ufscar.br (S.G.)

Abstract: From a physical/dynamical system perspective, the potential well represents the proportional mass of points that escape the neighbourhood of a given point. In the last 20 years, several works have shown the importance of this quantity to obtain precise approximations for several recurrence time distributions in mixing stochastic processes and dynamical systems. Besides providing a review of the different scaling factors used in the literature in recurrence times, the present work contributes two new results: (1) For ϕ -mixing and ψ -mixing processes, we give a new exponential approximation for hitting and return times using the potential well as the scaling parameter. The error terms are explicit and sharp. (2) We analyse the uniform positivity of the potential well. Our results apply to processes on countable alphabets and do not assume a complete grammar.

Keywords: potential well; poincaré recurrence times; exponential approximation; rare events; mixing processes



Citation: Abadi, M.; Amorim, V.; Gallo, S. Potential Well in Poincaré Recurrence. *Entropy* **2021**, *23*, 379. <https://doi.org/10.3390/e23030379>

Academic Editor: Sandro Vaienti

Received: 6 February 2021

Accepted: 18 March 2021

Published: 23 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The close relation between the Extreme Value Theory (EVT) and the statistical properties of Poincaré recurrence has been recently quite well explored. The starting point is that the exceedances of a stochastic process to a sequence of barrier values $a_n > 0, n \in \mathbb{N}$, can be considered as hittings to a sequence of nested sets. More precisely, if one defines the semi-infinite intervals:

$$A_n = (a_n, \infty),$$

and considers a sequence of random variables X_1, X_2, \dots , one has the equivalence:

$$\max\{X_1, \dots, X_t\} > a_n \quad \text{if and only if} \quad T_{A_n} \leq t,$$

where for any measurable set A , T_A denotes the smallest k such that $X_k \in A$. As the sequence of levels a_n diverges, the sets A_n are nested. This equivalence allows building a bridge between two historically independent theories: Extreme Value Theory (EVT) and Poincaré Recurrence Theory (PRT). While EVT focuses on the existence (and identification) of the limit of the distribution of the partial maxima and k -maxima, among others [1–5], the aim of recent works on PRT is to understand the statistical properties of the different notions of return times.

The present paper stands on the approach of PRT. Our interest is the statistics of visits of a random process $X_t, t \in \mathbb{N}$ to a given target measurable set. Asymptotic statistics are obtained by studying sequences of target sets $A_n, n \geq 1$, usually of a measure shrinking to zero. In this context and for certain classes of processes, hitting and return times with respect to a given sequence of target sets converge to the exponential distribution, modelling the unpredictability of rare events. However, this rough affirmation is full of

nuances that need to be established in very precise terms. It turns out that these details bring much information on the system.

For instance, for two observables having the same probability, the Ergodic theorem says that, macroscopically, their numbers of occurrences are about the same. However, these occurrences can appear scattered in a very different way along time. Under some strong mixing assumptions, it is a well-known fact of the literature that for nested sequences of observables with the same probability, the asymptotic observation of one of them can be distributed as a Poisson process while the other one will follow a compound Poisson process. Thus, the dichotomy Poisson/compound Poisson in the same system is determined also by the intrinsic properties of the target sets considered.

In the setting of the present paper, the target sets are finite strings of symbols (patterns). In this case, even if the process is a sequence of independent random variables, the successive occurrences of the string are not independent because the structure of the pattern itself enters the game, allowing or avoiding consecutive observations due to possible overlaps with itself. This leads to a dichotomy between aperiodic/periodic patterns that yields, in the limit of long patterns, the dichotomy Poisson/compound Poisson mentioned before. In passing, let us also mention that this dichotomy also exists in EVT where it is referred to as the phenomenon of clustering/non-clustering of maxima, and it has generated a great deal of research over the last two decades [3].

Let us now be more specific about what we are doing here. First, we operate in the context of discrete time stochastic processes with countable alphabet enjoying ϕ -mixing. Fix any point x , that is any right infinite sequence of symbols taken from the alphabet, and consider the nested sequence of neighbourhoods corresponding to the first n symbols of x , namely $A_n = (x_0, \dots, x_{n-1})$, $n \geq 1$. The main theorem of the paper, Theorem 1, gives explicit and computable error terms for the approximation of the hitting time distribution $\mu(T_{A_n} > t)$ and return time distribution $\mu_{A_n}(T_{A_n} > t)$, by exponential distributions whose parameter is explicit and depends on A_n .

The first main advantage of Theorem 1 is that it uses the potential well as the scaling parameter. In other words, the potential well is the probability, conditioned on starting from A_n , that the pattern A_n does not reappear at the first possible moment it could reappear. The use of this simple and well-defined quantity as the scaling parameter contrasts with previous works using parameters whose expressions are hardly explicit and even more hardly computable. Another advantage of Theorem 1 is that, unlike a whole body of literature obtaining almost sure results, our results hold for all x . This allows distinguishing different limiting distributions, as for example in the periodic/aperiodic dichotomy described above that almost sure results cannot detect. Finally, the error terms of our approximations are not in total variation distance, but in the stronger point-wise form with respect to the time scale.

The closest result of the literature, but restricted to return times, is the main theorem of [6]. Our Theorem 1, besides considering also the case of hitting times, also extends the class of processes to which the approximations apply. First, we do not assume the finiteness of the alphabet, allowing countably infinite alphabets. Second, our processes are not assumed to have a complete grammar, as was the case of [6]. In ergodic theory terminology, this means that we do not restrict to the full shift. Yet another contribution here is that we specify a tighter and simpler error term under the stronger assumption of ψ -mixing in both hitting and return time approximations.

Last but not least, Theorem 1 corrects the exponential approximation obtained by [6] for return times. Indeed, the error term of their Theorem 4.1 contains a mistake for small t 's.

The other main novelty of the present work is Theorem 2, stating that (1) the potential well almost surely converges to one as the size of the pattern diverges under ϕ -mixing and (2) the potential well is uniformly bounded away from zero when we have ψ -mixing or ϕ -mixing with summable function $\phi(n)$. Naturally, as a conditional probability, we know that the potential well belongs to the interval $[0, 1]$ for any $n \geq 1$ and any pattern A_n . However, it was proven that the potential well could be arbitrarily close to zero

for β -mixing processes, a slightly weaker mixing assumption than ϕ -mixing. Indeed, it was shown in [7] that for the binary renewal process, with specific choices of transition probabilities and target sets $A_n, n \geq 1$, the potential well of A_n vanishes as n diverges. Note that the border is thin between this β -mixing example and our Theorem 2 holding for ψ -mixing and ϕ -mixing with summable ϕ (see the review of [8] on the distinct mixing assumptions). We conjecture that the assumption of the summability of the ϕ rates can be dropped.

A fundamental quantity is the shortest possible return of the pattern A_n , denoted $\tau(A_n)$, which is in particular used to define the potential well. It is known [9,10] that under the assumptions of specification and positive entropy, $\tau(A_n)/n$ converges almost surely to one. Since we do not assume a complete grammar, $\tau(A_n)$ is not bounded above by n , and it could be strictly larger than n in general. Here, we prove an important technical result (Lemma 2) giving explicit uniform upper bounds, which hold for sufficiently large n and under either ϕ - or ψ -mixing.

To conclude about the importance of the present work as a whole, let us mention that our results are fundamental for the study of further recurrence quantities, such as the return time function [11,12] and the waiting time function [13,14], establishing a link with information theory. These random variables are known to satisfy a counterpart of the famous Shannon–McMillan–Breiman theorem (asymptotic equipartition property). In order to study the fluctuations of these limit theorems, for instance a large deviation principle, we need to control the return/hitting time exponential approximations for any point and any $t > 0$. This was particularly clear in [15,16], studying the fluctuations of the waiting time and return time, respectively. It is also interesting to notice that it was [17] who first pointed out the importance of seeking exponential approximations for any point x , and it was precisely to study the small and large fluctuations of the return time function.

The paper is organized as follows. We describe in Section 2 the setting of the paper in the context of PRT, defining carefully the types of exponential approximations we are interested in and explaining, including through an extensive bibliography, the role of the potential well as the scaling parameter. Section 3 contains the main results, and Section 4 is dedicated to their proofs.

2. Poincaré Recurrence Theory for Mixing Processes

2.1. The Framework of Mixing Processes

Consider a countable set \mathcal{A} that we call the alphabet. By \mathbb{N} , we denote the set of nonnegative integers and, by $\mathcal{X} := \mathcal{A}^{\mathbb{N}}$, the set of right infinite sequences $x = (x_0, x_1, \dots)$ of symbols taken from \mathcal{A} . Given a point $x \in \mathcal{X}$ and for any finite set $I \subset \mathbb{N}$, the cylinder sets with the base in I is defined as the set $A_I(x) := \{y \in \mathcal{X} : y_i = x_i, i \in I\}$. In the particular case where $I = \{0, \dots, n-1\}$, we write $A_n(x)$ and sometimes abuse notation, writing x_0^{n-1} . We endow \mathcal{X} with the σ -algebra \mathcal{F} generated by the class of cylinder sets $\{A_I : I \subset \mathbb{N}, |I| < \infty\}$. Furthermore, \mathcal{F}_I denotes the σ -algebra generated by $A_I(x), x \in \mathcal{X}$. For the special case in which $I = \{i, \dots, j\}, 0 \leq i \leq j < \infty$, we use the notation \mathcal{F}_i^j . We use the shorthand notation $a_i^j := (a_i, a_{i+1}, \dots, a_j), 0 \leq i \leq j < \infty$ for finite strings of consecutive symbols of \mathcal{A} . When necessary, $A_n(x)$ will be naturally identified with the sequence x_0^{n-1} .

The shift operator $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ shifts the point $x = (x_0, x_1, x_2, \dots)$ to the left by one coordinate, $(\sigma x)_i = x_{i+1}, i \geq 0$.

We consider a shift invariant (or stationary) probability measure μ on $(\mathcal{X}, \mathcal{F})$. For any $A \in \mathcal{F}$ of positive measure, $\mu_A(\cdot) := \frac{\mu(\{x \in \cdot \cap A\})}{\mu(A)}$ is the conditional measure μ restricted to A .

Our results are stated under two mixing conditions that we now define. For all $n \geq 1$, define:

$$\begin{aligned} \phi(n) &:= \sup_{i \in \mathbb{N}, A \in \mathcal{F}_0^i, B \in \mathcal{F}_{i+n}^\infty} \left| \frac{\mu(A \cap B)}{\mu(A)} - \mu(B) \right|, \\ \psi(n) &:= \sup_{i \in \mathbb{N}, A \in \mathcal{F}_0^i, B \in \mathcal{F}_{i+n}^\infty} \left| \frac{\mu(A \cap B)}{\mu(A)\mu(B)} - 1 \right|. \end{aligned}$$

Note that $\psi(n)$ and $\phi(n)$ are nonincreasing sequences, since $\mathcal{F}_0^i \subset \mathcal{F}_0^{i+1}$ for every $i \geq 0$.

Definition 1. We say that the measure μ on $(\mathcal{X}, \mathcal{F})$ is ϕ -mixing (resp. ψ -mixing) if $\phi(n)$ (resp. $\psi(n)$) goes to zero as n diverges. We say that μ is “summable ϕ -mixing” if it is ϕ mixing with $\sum_n \phi(n) < \infty$.

We refer to [8] for an exhaustive review of mixing properties and examples.

2.2. Recurrence Times and Exponential Approximations

The hitting time of a point y to a set $A \in \mathcal{F}$ is defined by:

$$T_A(y) = \inf\{k \geq 1 : \sigma^k(y) \in A\}.$$

For sets A of small measure (rare events) and under mixing conditions such as the ones introduced in the preceding subsection, it is expected that $\mu(T_A > t)$ is approximately exponentially distributed. This is what we call hitting time exponential approximation. Similarly, when we refer to return time, we mean that we study the approximation of $\mu_A(T_A > t)$, that is the measure of the same event, conditioned on the points starting in A .

In this paper, we are interested in the case where we fix any point x and consider $A_n(x)$ as the target set. When n diverges, the measure of $A_n(x)$ vanishes, leading to rare events. The scaling parameter of the exponential approximation depends on the point x .

The two main types of approximations that appeared in the literature when approximating the hitting/return time distributions around any point x of the phase space are a total variation distance type and a pointwise type.

- Type 1: Total variation distance. For any $x \in \mathcal{X}$,
 - Hitting times:

$$\sup_{t > 0} \left| \mu(T_A > t) - e^{-\mu(A)\theta(A)t} \right| \leq \epsilon(A),$$

- Return times:

$$\sup_{t > 0} \left| \mu_A(T_A > t) - \bar{\theta}(A)e^{-\mu(A)\theta(A)t} \right| \leq \epsilon(A).$$

- Type 2: Pointwise. For any $x \in \mathcal{X}$ and any $t > 0$,
 - Hitting times:

$$\left| \mu(T_A > t) - e^{-\mu(A)\theta(A)t} \right| \leq \epsilon(A, t),$$

- Return times:

$$\left| \mu_A(T_A > t) - \bar{\theta}(A)e^{-\mu(A)\theta(A)t} \right| \leq \epsilon(A, t).$$

Note that in the return time approximation, the parameters θ and $\bar{\theta}$ need not be equal. However, such approximation leads to:

$$\mathbb{E}_A(T_A) \approx \bar{\theta}(A) \frac{1}{\mu(A)\theta(A)}.$$

In view of Kac lemma, which, we recall, states that $\mathbb{E}_A(T_A) = \frac{1}{\mu(A)}$, the last display suggests that θ and $\bar{\theta}$ must be close.

2.3. Potential Well: Definition and Genealogy in PRT

As already explained, the potential well will be used as the scaling parameter in the exponential approximations of Types 1 and 2 defined above. In order to define it, we need first to define the shortest possible return of a set $A \in \mathcal{F}$ (to itself):

$$\tau(A) := \inf_{y \in A} \left\{ T_A(y) : \mu_A \left(\sigma^{-T_A(y)}(A) \right) > 0 \right\},$$

or, equivalently:

$$\tau(A) := \inf \left\{ k \geq 1 : \mu_A \left(\sigma^{-k}(A) \right) > 0 \right\}.$$

In the case where $A = A_n(x)$, we define $\tau_n(x) = \tau(A_n(x))$. The $\tau_n : \mathcal{X} \rightarrow \mathbb{N}$ constitutes a sequence of simple functions. Notice that the above alternative definitions of $\tau(A)$ involve the measure μ_A , while the traditional definition is completely topological. This is to account for the case where the measure does not have a complete grammar (We say that μ has a complete grammar if $\mu(A_n(x)) > 0$ for any $x \in \mathcal{X}$ and $n \geq 1$).

The first possible return time $\tau_n(x)$ is an object of independent interest, which was studied under several perspectives in the literature. Let us mention that its asymptotic concentration was proven by [9,10], large deviations in [18–20], and fluctuations in [21,22].

Obviously, by definition, $\mu_A(T_A \geq \tau(A)) = 1$. If for a point $x \in A$, we have $T_A(x) > \tau(A)$, we say that x escapes from A . The potential well of order n at x is precisely the proportional measure of points of A that escape from A :

$$\rho(A) := \mu_A(T_A > \tau(A)).$$

Since we are interested in the case where $A = A_n(x)$, we use the alternative notation $\rho(x_0^{n-1})$ instead of $\rho(A_n(x))$. Besides being explicitly computable in many situations, the potential well is physically meaningful and, as the scaling parameter, provides precise exponential approximations for recurrence times under suitable mixing assumptions.

We give below a small genealogy of scaling parameters that appeared in the literature that consider results holding for all points to get approximations for hitting/return times.

- As far as we know, the first paper to prove exponential approximations for hitting time statistics for all points is due to Aldous and Brown [23]. They obtained Type 1 approximations in the case of reversible Markov chains. The parameter used there was just the inverse of the expectation, which is mandatory to use when the approximating law is the exponential distribution. However, this does not bring information about the value of the expectation.
- Galves and Schmitt [24] obtained Type 1 approximations for hitting times in ψ -mixing processes. The major breakthrough there was that the authors provided an explicit formula for the parameter (denoted by $\lambda(A)$). This quantity could be viewed as the grandfather of ρ . Nonetheless, its explicit significance was not evident.
- References [25,26] gave exponential approximations (Type 1 and Type 2, respectively) of the distribution of hitting time around any point using a scaling parameter. In [26], however, only its existence and necessity were proven, the calculation of λ being intractable in general. The main problem is that $\lambda(A)$ depends on the recurrence property of the cylinder A up to large time scales (usually of the order of $\mu(A)^{-1}$).
- In order to circumvent this issue, Reference [25] also provided, in the context of approximations of Type 1, another scaling parameter, easier to compute, but with a slightly larger error term as a price to pay. It is defined as follows:

$$\zeta_s(x_0^{n-1}) := \mu_{x_0^{n-1}}(T_{x_0^{n-1}} > n/s).$$

This quantity depends on, at most, the $2n$ first coordinates of the process. $\zeta_s(x_0^{n-1})$ can be seen as the father of the potential well. Both works [25,26] led with processes enjoying ψ -mixing or summable ϕ -mixing.

- The use of the potential well ρ as the scaling parameter was firstly proposed by Abadi in [27], still in the context of an approximation of Type 1 for hitting and return times. More specifically, it is proven that, for exponentially α -mixing processes, λ and ζ (grandpa and father of ρ) can be well approximated by ρ .
- The first paper to really directly use ρ as the scaling parameter was [6], in which a Type 2 approximation for return times was obtained, with $\bar{\theta} = \theta = \rho$. The process is assumed to be ϕ -mixing.
- Focusing on proving exponential approximations for hitting and return times under the largest possible class of systems, and still for all points, Abadi and Saussol [28] returned to the approach of Galves and Schmitt. Their results hold under the α -mixing condition, which is the weakest hypothesis used up to date, but the scaling parameter is not explicit.
- Focussing on the specific class of binary renewal processes, Reference [7] proved a Type 1 approximation for hitting and return times using the potential well ρ . One interesting aspect concerning this work lies in the fact that the renewal process is β -mixing (weaker than the ϕ -mixing assumed by [6]). Moreover, the authors managed to use the renewal property to compute the limit of $\rho(A_n(x))$ for any point x . In other words, the approximating asymptotic law for hitting and return times was explicitly computed as a function of the parameters of the process. This result shows the usefulness of the potential well, an “easy to compute” scaling parameter.

3. Main Results

Theorem 1 below presents Type 2 approximations for hitting and return time under ϕ - and ψ -mixing conditions with the potential well as the scaling parameter and an explicit error term.

Before we can state this result, we first need to define the second order periodicity of string $A_n(x)$, which plays a crucial role for the size of the error term.

3.1. Second Order Periodicity

The short returns that we will define here are precisely those that are difficult to treat as (almost) independent. They not only depend on the correlation decay of the system, but also on the particular properties of the string itself. Technically, for an n -cylinder, short means returning in up to the order n steps.

Consider the cylinder A , and suppose $\tau(A) = k$. Write $n = qk + r$, where $q \in \mathbb{N}$ and $0 \leq r < k$, and note that the cylinder overlaps itself in all multiples of k smaller than n . The set $\mathcal{P}(A) := \{mk : 1 \leq m \leq q\}$ is the indexes of possible returns at multiples of $\tau(A)$, but returns can also occur at other time indexes after that. Let:

$$\mathcal{R}(A) = \{j \in \{qk + 1, \dots, qk + r - 1\} : \mu_A(\sigma^{-j}(A)) > 0\}.$$

A point $y \in A$ could only return to A before n at time indexes in $\mathcal{P}(A) \cup \mathcal{R}(A)$, but there is a crucial difference between them. A point that escapes from A cannot return in $\mathcal{P}(A)$, but it could return in $\mathcal{R}(A)$. Namely,

$$\mu_A(\sigma^{-\tau(A)}(A^c) \cap \sigma^{-j}(A)) = 0, \quad j \in \mathcal{P}(A),$$

while:

$$\mu_A(\sigma^{-\tau(A)}(A^c) \cap \sigma^{-j}(A)) > 0, \quad j \in \mathcal{R}(A).$$

We set n_A as the first possible return to A , among those points $x \in A$ that escape A at $\tau(A)$:

$$n_A = \min\{j : \mu_A(\sigma^{-\tau(A)}(A^c) \cap \sigma^{-j}(A)) > 0\}.$$

We refer to [6] for an example that illustrates these ideas under a complete grammar and a finite alphabet. In the general case, notice that n_A could be strictly larger than n , if one is not considering the full shift. For a complete example, consider the house of cards Markov chain. It has transition matrix Q on \mathbb{N} with entries $Q(i, 0) = q_i = 1 - Q(i, i + 1)$ where $q_i, i \geq 0$ is a sequence of real numbers taking values in the interval $(0, 1)$. Therefore, it is defined on an infinite alphabet and does not have a complete grammar, since several transitions are forbidden due to the sparse nature of Q . Consider the strings $A = 00010001000$ and $B = (n + 1) \cdots 2n$ (for some $n \geq 1$) generated by this Markov chain in the stationary measure. Then, we see that $\tau(A) = 4, \mathcal{R}(A) = \{9, 10\}$, and $n_A = 9$, while, on the other hand, $\tau(B) = 2n + 1 > n, \mathcal{R}(B) = \emptyset$, and $n_B = 3n + 2$.

3.2. Type 2 Approximations Scaled by the Potential Well

For any finite string A , let us denote by $A^{(k)}$ the suffix of A of size k . That is, if $A = x_0^{n-1}$, then $A^{(k)} = x_{n-k}^{n-1}$. When A is an n -cylinder we use the convention $\mu(A^{(j)}) = \mu(A^{(n)}) = \mu(A)$ for $j \geq n$.

By definition, $\phi(g)$ is finite for all $g \geq 1$. This is not the case for $\psi(g)$. Thus, for ψ -mixing measures, we define:

$$g_0 = g_0(\psi) := \inf\{g \geq 1 : \psi(g) < \infty\} - 1. \tag{1}$$

Now, for the error term, define:

$$(a) \ \epsilon_\psi(A) := n\mu(A^{(n_A - g_0)}) + \psi(n), \tag{2}$$

$$(b) \ \epsilon_\phi(A) := \inf_{1 \leq w \leq n^*} \left\{ (n + \tau(A))\mu(A^{(w)}) + \phi(\min\{n_A - w + 1, n\}) \right\}. \tag{3}$$

where $n^* := \min\{n, n_A\}$.

Note that cylinders A of size n verify that $n_A \geq n/2$, then ϵ_ψ is well defined for all $n > 2g_0$.

We use ϵ to denote either ϵ_ψ or ϵ_ϕ when the argument/statement is general.

Theorem 1. Consider a stationary measure μ on $(\mathcal{X}, \mathcal{F})$ enjoying either ϕ -mixing with $\sup_{A \in \mathcal{A}^n} \mu(A)\tau(A) \xrightarrow{n} 0$, or simply ψ -mixing. There exist five positive constants $C_i, i = 1, \dots, 5$, and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ and all $A \in \mathcal{A}^n$, the following inequalities hold.

- For all $t \geq 0$:

$$\left| \mu(T_A > t) - e^{-\rho(A)\mu(A)t} \right| \leq \begin{cases} C_1(\tau(A)\mu(A) + t\mu(A)\epsilon(A)) & t \leq [2\mu(A)]^{-1} \\ C_2\mu(A)t\epsilon(A)e^{-\mu(A)t(\rho(A) - C_3\epsilon(A))} & t > [2\mu(A)]^{-1}. \end{cases}$$

For all $t \geq \tau(A)$:

$$\begin{aligned} & \left| \mu_A(T_A > t) - \rho(A)e^{-\rho(A)\mu(A)(t - \tau(A))} \right| \\ & \leq \begin{cases} C_4\epsilon(A) & t \leq [2\mu(A)]^{-1} \\ C_5\mu(A)t\epsilon(A)e^{-\mu(A)t(\rho(A) - C_3\epsilon(A))} & t > [2\mu(A)]^{-1}. \end{cases} \end{aligned}$$

Theorem 1 (and its proof) were definitely inspired by [6] and their Theorem 4.1. However, let us first observe that our result provides the first statement of the literature for Type 2 hitting time approximations with the potential well as the scaling parameter. Moreover, contrary to [6], we do not assume a complete grammar nor a finite alphabet.

Let us make some further important observations concerning this theorem.

Remark 1. Under ϕ -mixing, the assumption $\sup \mu(A)\tau(A) \xrightarrow{n} 0$ can be dropped under certain circumstances. For instance, if the measure has a complete grammar, we have $\tau(A) \leq n$, and the

assumption is granted using Lemma 1. Another way is to assume that μ is summable ϕ -mixing, as mentioned after Lemma 2 in Section 4.

Remark 2. According to Lemma 1, if μ is ϕ -mixing (and a fortiori, ψ -mixing), there exist constants C and c such that $\mu(A) \leq Ce^{-cn}$ for all $n \geq 1$ and $A \in \mathcal{A}^n$. On the other hand, since $n_A \geq n/2$, we get $\mu(A^{(n_A - g_0)}) \leq Ce^{-c(n/2 - g_0)}$ for all $n > 2g_0$. Therefore, $\epsilon_\psi(A) \xrightarrow{n} 0$ uniformly. Further, if $\tau(A) \leq 2n$, it is enough to take $w = \lceil n/4 \rceil$ to obtain $\phi(\min\{n_A - w + 1, n\}) \leq \phi(\lfloor n/4 \rfloor)$ and $(n + \tau(A))\mu(A^{(w)}) \leq 3Cne^{-cn/4}$, which ensures $\epsilon_\phi(A) \xrightarrow{n} 0$ uniformly. This is the case, for instance, if one has a complete grammar. On the other hand, notice that $\tau(A) < n_A$. Hence, if $\tau(A) > 2n$, we take $w = n$ and get $\phi(\min\{n_A - w + 1, n\}) = \phi(n)$. Therefore, since $\tau(A)\mu(A) \xrightarrow{n} 0$, we also have in this case $\epsilon_\phi \xrightarrow{n} 0$.

Remark 3. Naturally, the statements under ψ -mixing are less general, but have smaller error terms. The error term is the same for $t > [2\mu(A)]^{-1}$ for both hitting and return times' approximations. The difference is for small t 's, due to the correlation arising from the conditional measure.

Remark 4. For application purposes involving data, it is essential to control all the constants involved in the statements. These constants can be accessed from the proof presented in Section 4.2, where we also make explicit the integer n_0 from which Theorem 1 holds (see (23)). If μ is ψ -mixing, we define $M := \psi(g_0 + 1) + 1$. In this case, $C_1 = 8M + 9$, $C_2 = 194M + 206$, $C_3 = 66M + 89$, $C_4 = 12M + 15$, and $C_5 = 197M + 220$. On the other hand, for the ϕ -mixing case, we have $C_1 = 9$, $C_2 = 143$, $C_3 = 61$, $C_4 = 14$, and $C_5 = 170$.

Remark 5. We show the sharpness of the error term in the return time approximation given by Theorem 1 with a simple example. Consider an i.i.d. process $(X_m)_{m \in \mathbb{N}}$ with alphabet \mathcal{A} . Take $b \in \mathcal{A}$ such that $\mu(b) = p$ and $x = (b, b, \dots) \in \mathcal{X} = \mathcal{A}^{\mathbb{N}}$. Thus, $A_n(x) = x_0^{n-1} = (b, b, \dots, b)$. Direct calculations give:

- $\mu(A_n) = p^n$
- $\tau(A_n) = 1$
- $\rho(A_n) = 1 - \mu_{A_n}(T_{A_n} = \tau(A_n)) = 1 - \mu_{A_n}(X_n = b) = 1 - p$
- $n_{A_n} = n$
- $\mu_{A_n}(T_{A_n} > n - 1) = \rho(A_n) = 1 - p$.

An i.i.d. process is trivially ψ -mixing with function ψ identically zero. Thus, Theorem 1 states that the error for small t 's is $\epsilon_\psi(A_n) = np^n$. On the other hand, by direct substitution using the above facts, we have for each $n \geq 2$:

$$\begin{aligned} \left| \mu_{A_n}(T_{A_n} > n - 1) - \rho(A_n)e^{-\rho(A_n)\mu(A_n)((n-1) - \tau(A_n))} \right| &= \left| (1 - p) - (1 - p)e^{-(1-p)p^n((n-1) - 1)} \right| \\ &= (1 - p) \left(1 - e^{-(1-p)p^n(n-2)} \right) \end{aligned}$$

which implies that the exact error in the approximation for the return time at $n - 1$ is of order $p^n n$, just as stated by Theorem 1.

Remark 6. The reader may notice a difference between Theorem 1 and Theorem 4.1 of [6] concerning the error term for small t 's for return time approximation. Indeed, their statement is incorrect as shown by the preceding example. We recall that the error term for small t 's plays a fundamental role when studying the return time spectrum, as was done by [15]. Theorem 1, besides correcting [6], is also fundamental to correct [15], which was based on the exponential approximations given by [6].

Remark 7. As an example where Theorem 1 can be applied while Theorem 4.1 of [6] cannot, let us consider the house of cards Markov chain defined in the previous subsection. We refer to [7] where it was explained that, if $q_i \geq \epsilon$, $i \geq 1$ for some $\epsilon > 0$, there exists a stationary process $(X_m)_{m \in \mathbb{N}}$ with matrix Q , and it is ϕ -mixing with exponentially decaying $\phi(n)$. Therefore, according to Remark 1,

it fits into the conditions of Theorem 1. However, since it is defined on an infinite alphabet and does not have a complete grammar, such processes are not covered by [6].

3.3. Uniform Positivity of the Potential Well

Theorem 1 says that the potential well can be used as the scaling parameter to obtain approximations for recurrence times around any point. We now ask about the possible values of this scaling parameter in its range $[0, 1]$.

Abadi and Saussol [29], in the more general case known up to now, proved that for α -mixing processes with at least polynomially decaying α rates, the distribution of the hitting and return time converges, almost surely, to an exponential with parameter of one. We refer to [8] for the precise definition of α -mixing, but the only important point for us it to know that summable ϕ -mixing implies α -mixing with at least polynomially decaying α rates. This fact, combined with Theorem 1, proves, indirectly, that for summable ϕ -mixing processes, the potential well converges almost surely to one, since both theorems must agree on the limiting distribution under these conditions. Theorem 2 Item (a) below states that the same holds for ϕ -mixing without any assumption on the rate.

On the other hand, for the renewal processes, with a certain tail distribution for the inter-arrival times, Abadi, Cardeño, and Gallo [7] proved that for the point $x = (0000\dots)$, the sequence of potential wells $\rho(x_0^{n-1})$ converges to zero. In this case, the scaling parameter has a predominant role, indicating the drastic change of scale of the occurrence of events. For instance, in this case, the mean hitting time is much larger than the mean return time:

$$\mathbb{E}(T_{x_0^{n-1}}) \approx \frac{1}{\rho(x_0^{n-1})\mu(x_0^{n-1})} \gg \frac{1}{\mu(x_0^{n-1})} = \mathbb{E}_{x_0^{n-1}}(T_{x_0^{n-1}}).$$

Such renewal processes are β -mixing (see [8] for the definition). Theorem 2 Item (b) below states that this cannot happen for ψ -mixing processes or summable ϕ -mixing processes.

Theorem 2. *Let μ be a stationary ϕ -mixing measure. Then:*

- (a) $\rho(x_0^{n-1}) \xrightarrow{n} 1$, almost surely.
- (b) If μ is ψ -mixing or summable ϕ -mixing, there exists $n_1 \geq 1$ such that:

$$\inf_{n \geq n_1, x_0^{n-1} \in \mathcal{A}^n} \rho(x_0^{n-1}) = \rho_- > 0.$$

If the alphabet, \mathcal{A} is finite; the set $\{\rho(A) : A \in \mathcal{A}^n, n < n_1\}$ is finite and has a strictly positive infimum, which implies that the infimum above can be taken over all $n \geq 1$.

4. Proofs of the Results

The statement of Theorem 1 is for ϕ and ψ and for hitting and return times. The case of return times under ϕ -mixing was already done by [6]. Our proof follows their method. In particular for the next subsection that lists a sequence of auxiliary results, some of them are not proven.

4.1. Preliminary Results

The following lemma plays a fundamental role in Theorems 1 and 2. It was originally proven in [25] assuming the summability of the function ϕ , an assumption that can be dropped.

Lemma 1. *Let μ be a ϕ -mixing measure. Then, there exists positive constants C and c such that for all $n \geq 1$ and all $A \in \mathcal{A}^n$, one has:*

$$\mu(A) \leq Ce^{-cn}.$$

Proof. We denote by $\lambda = \sup\{\mu(a) : a \in \mathcal{A}\} < 1$. Consider a positive integer k_0 , and for all $n \geq k_0$, write $n = k_0q + r$, with $1 \leq q \in \mathbb{N}$ and $0 \leq r < k_0$. Suppose $A = a_0^{n-1}$, and apply the ϕ -mixing property to obtain:

$$\begin{aligned} \mu(A) &\leq \mu\left(\bigcap_{j=0}^{q-1} \{\sigma^{-jk_0}(a_{jk_0})\}\right) \leq \mu\left(\bigcap_{j=0}^{q-2} \{\sigma^{-jk_0}(a_{jk_0})\}\right) (\phi(k_0) + \mu(a_{(q-1)k_0})) \\ &\leq \mu\left(\bigcap_{j=0}^{q-2} \{\sigma^{-jk_0}(a_{jk_0})\}\right) (\phi(k_0) + \lambda). \end{aligned}$$

Iterating this argument, one concludes:

$$\mu(A) \leq (\phi(k_0) + \lambda)^q.$$

Since $\phi(k) \xrightarrow{k} 0$, there exists $k_0 \in \mathbb{N}$ such that $\phi(k_0) + \lambda < 1$. Thus, for $n \geq k_0$ and observing that $q = \frac{n-r}{k_0} > \frac{n}{k_0} - \frac{k_0-1}{k_0}$:

$$\mu(A) \leq (\phi(k_0) + \lambda)^{-(k_0-1)/k_0} \left((\phi(k_0) + \lambda)^{1/k_0} \right)^n.$$

This covers the case $n \geq k_0$. By eventually enlarging the constant C , one covers the case $n < k_0$. This ends the proof. \square

For a complete grammar, one has $\tau(A_n(x)) \leq n$. Since we do not assume this, we need the following lemma, which provides upper bounds for $\tau(A)$ when μ is ψ -mixing or summable ϕ -mixing.

Lemma 2. Consider μ a ψ -mixing or summable ϕ -mixing measure. Then, there exists $n_2 \in \mathbb{N}$ such that for all $n \geq n_2$ and $A \in \mathcal{A}^n$,

- $\tau(A) \leq 2n$, for ψ ;
- $\tau(A) \leq -\frac{2}{\mu(A) \ln \mu(A)} + n$, for summable ϕ .

Proof. We start with the case ψ . For n large enough, we have $\psi(n) < 1$, which implies:

$$\mu(A \cap \sigma^{-2n}(A)) \geq \mu(A)^2(1 - \psi(n)) > 0$$

Since $\tau(A)$ is the smallest positive integer such that $\mu(A \cap \sigma^{-\tau(A)}(A)) > 0$, one has $\tau(A) \leq 2n$.

Now consider the ϕ -mixing case. The summability of ϕ ensures that for g large enough, we have $\phi(g) \leq 1/(g \ln g)$. Thus:

$$\mu(A \cap \sigma^{-g-n}(A)) \geq \mu(A)(\mu(A) - \phi(g)) \geq \mu(A) \left(\mu(A) - \frac{1}{g \ln g} \right).$$

Take $g = -\frac{2}{\mu(A) \ln \mu(A)}$. The rightmost parenthesis above becomes:

$$\mu(A) \left[1 - \frac{1}{2} \frac{-\ln \mu(A)}{(\ln(2) - \ln \mu(A) - \ln(-\ln \mu(A)))} \right]$$

which is positive for n large enough. \square

The multiplicative constant of two in both cases is technical and was chosen for the simplicity of the proof. Actually, it can be replaced by any constant strictly larger than one.

An irreducible aperiodic finite state Markov chain with some entry equal to zero shows that this constant cannot be taken equal to one in the ψ -mixing case. Whether this bound is optimal for the ϕ -mixing case is an open question. Note that Lemmas 1 and 2 imply that $\tau(A)\mu(A) \xrightarrow{n} 0$ uniformly.

The remaining results of this subsection hold for $n \geq n'$, where $n' = 1$ for the case of ϕ -mixing and:

$$n' := \inf\{n > 2g_0 : \psi(n) < 1\} \tag{4}$$

for the ψ -mixing case (see (1) for the definition of g_0).

Let us define:

$$M := \psi(g_0 + 1) + 1.$$

Proposition 1. *Let μ be a ψ -mixing measure. Then, for all $n \geq n'$, $A \in \mathcal{A}^n$ and $k \geq n_A$, the following inequality holds:*

$$\mu_A(\tau(A) < T_A \leq k) \leq M(k - n_A + 1)\mu(A^{(n_A - g_0)}).$$

Proof. By definition of n_A , we first note that $\mu_A(\tau(A) < T_A \leq k) = \mu_A(n_A \leq T_A \leq k)$. Consider the case in which $n_A \leq n + g_0$. In this case, for $j \geq n_A$, one trivially has:

$$\{T_A = j\} \subset \sigma^{-j}(A) \subset \sigma^{-j-(n-(n_A-g_0))}(A^{(n_A-g_0)}).$$

Thus:

$$\mu_A(n_A \leq T_A \leq k) \leq \mu_A\left(\bigcup_{n_A \leq j \leq k} \sigma^{-j-(n-(n_A-g_0))}(A^{(n_A-g_0)})\right).$$

Note that A and the union on the right hand-side in the above inequality are separated by a gap of length $g_0 + 1$. By ψ -mixing, one concludes that the left-hand side is bounded by:

$$(\psi(g_0 + 1) + 1)\mu\left(\bigcup_{n_A \leq j \leq k} \sigma^{-j-(n-(n_A-g_0))}(A^{(n_A-g_0)})\right) \leq M(k - n_A + 1)\mu(A^{(n_A-g_0)}).$$

For $n_A > n + g_0$, recall first the convention in Subsection 3.2, which states that $\mu(A^{(n_A-g_0)}) = \mu(A)$. In a similar way to the first case,

$$\begin{aligned} \mu_A(n_A \leq T_A \leq k) &\leq \mu_A\left(\bigcup_{n_A \leq j \leq k} \sigma^{-j}(A)\right) \\ &\leq M(k - n_A + 1)\mu(A). \end{aligned}$$

□

For the next proposition, recall that ϵ stands either for ϵ_ψ (2) or for ϵ_ϕ (3), according to the mixing property of the measure under consideration. Further, let us use the notation $T_A^{[t]} := T_A \circ \sigma^t$.

Proposition 2. *Let μ be a ϕ or ψ -mixing measure. Then, for all $n \geq n'$, $A \in \mathcal{A}^n$ and $t \geq \tau(A)$:*

$$|\mu_A(T_A > t) - \rho(A)\mu(T_A > t)| \leq C\epsilon(A),$$

where $C = 4$ for ϵ_ϕ and $C = 4(M + 1)$ for ϵ_ψ .

Proof. The proof for ϵ_ϕ can be found in Proposition 4.1 Item (b) of [6], and it remains valid even for a non-complete grammar and infinite alphabet. We observe that the error term defined therein is:

$$\epsilon'(A) = \inf_{1 \leq w \leq n_A} \left\{ (2n + \tau(A))\mu\left(A^{(w)}\right) + \phi(n_A - w + 1) \right\} \leq 2\epsilon_\phi(A),$$

which justifies $C = 4$ for this case.

Here, we prove the case ϵ_ψ in the same way. We start by assuming that $t \geq \tau(A) + 2n$. By the triangle inequality:

$$\begin{aligned} & |\mu_A(T_A > t) - \rho(A)\mu(T_A > t)| \\ & \leq \left| \mu_A\left(T_A > \tau(A); T_A^{[\tau(A)]} > t - \tau(A)\right) - \right. \\ & \quad \left. \mu_A\left(T_A > \tau(A); T_A^{[\tau(A)+2n]} > t - \tau(A) - 2n\right) \right| \end{aligned} \tag{5}$$

$$+ \left| \mu_A\left(T_A > \tau(A); T_A^{[\tau(A)+2n]} > t - \tau(A) - 2n\right) - \rho(A)\mu(T_A > t - \tau(A) - 2n) \right| \tag{6}$$

$$+ |\rho(A)\mu(T_A > t - \tau(A) - 2n) - \rho(A)\mu(T_A > t)|. \tag{7}$$

For the first modulus, by inclusion of sets:

$$(5) \leq \mu_A\left(T_A > \tau(A); T_A^{[\tau(A)]} \leq 2n\right) = \mu_A(\tau(A) < T_A \leq \tau(A) + 2n).$$

If $n_A > \tau(A) + 2n$, the last term is equal to zero. Otherwise, we apply Proposition 1 to obtain:

$$\begin{aligned} \mu_A(\tau(A) < T_A \leq \tau(A) + 2n) & \leq M(\tau(A) + 2n - n_A + 1)\mu\left(A^{(n_A - g_0)}\right) \\ & \leq 4Mn\mu\left(A^{(n_A - g_0)}\right) \end{aligned} \tag{8}$$

where the last inequality follows from Lemma 2.

By ψ -mixing, the modulus (6) is bounded by:

$$\rho(A)\mu(T_A > t - \tau(A) - 2n)\psi(n) \leq \psi(n).$$

Note that the modulus is not needed for (7), and by inclusion:

$$\begin{aligned} (7) & \leq \rho(A)\mu\left(T_A^{[t - \tau(A) - 2n]} \leq \tau(A) + 2n\right) \\ & = \rho(A)\mu(T_A \leq \tau(A) + 2n) \\ & \leq (2n + \tau(A))\mu(A) \\ & \leq 4n\mu\left(A^{(n_A - g_0)}\right) \end{aligned}$$

where the equality and second inequality follow from the stationarity of μ .

Therefore, for $t \geq \tau(A) + 2n$, the sum of (5), (6) and (7) is bounded by:

$$4Mn\mu\left(A^{(n_A - g_0)}\right) + \psi(n) + 4n\mu\left(A^{(n_A - g_0)}\right) \leq 4(M + 1)\epsilon_\psi(A).$$

We now consider the case where $\tau(A) \leq t < \tau(A) + 2n$. We have:

$$\begin{aligned} |\mu_A(T_A > t) - \rho(A)\mu(T_A > t)| & \leq |\mu_A(T_A > t) - \rho(A)| + |\rho(A) - \rho(A)\mu(T_A > t)| \\ & \leq \mu_A(\tau(A) < T_A \leq \tau(A) + 2n) + t\mu(A) \\ & \leq 4Mn\mu\left(A^{(n_A - g_0)}\right) + (\tau(A) + 2n)\mu\left(A^{(n_A - g_0)}\right) \\ & \leq 4(M + 1)\epsilon_\psi(A). \end{aligned}$$

The last inequality follows from (8). The other inequalities are straightforward. This ends the proof. \square

The next lemma establishes upper bounds for the tail distribution at the scale given by Kac’s lemma, namely $1/\mu(A)$. For technical reasons, we actually choose the scale:

$$f_A := 1/(2\mu(A)).$$

Lemma 3. *Let μ be a stationary measure. Then, for all $n \geq 1$, $A \in \mathcal{A}^n$, positive integer k , and $B \in \mathcal{F}_{kf_A}^\infty$, the following inequalities hold:*

- (a) $\mu(T_A > kf_A; B) \leq (\psi(n) + 1)^k \mu(T_A > f_A - 2n)^k \mu(B)$,
- (b) $\mu(T_A > kf_A; B) \leq (\mu(T_A > f_A - 2n) + \phi(n))^k (\mu(B) + \phi(n))$,
- (c) $\mu_A(T_A > kf_A; B) \leq (\psi(n) + 1)^k \mu(T_A > f_A - 2n)^{k-1} \mu(B)$.

Proof. We start by observing that $\{T_A > kf_A\} \subset \{T_A > kf_A - 2n\} \in \mathcal{F}_0^{kf_A - n}$. Thus, applying the ψ -mixing property, we get:

$$\mu(T_A > kf_A; B) \leq \mu(T_A > kf_A - 2n; B) \leq (\psi(n) + 1) \mu(T_A > kf_A - 2n) \mu(B) \tag{9}$$

Furthermore:

$$\{T_A > kf_A - 2n\} = \{T_A > (k-1)f_A; T_A^{[(k-1)f_A]} > f_A - 2n\}.$$

Now, one can take in particular $B = \{T_A^{[(k-1)f_A]} > f_A - 2n\} \in \mathcal{F}_{(k-1)f_A}^\infty$ and then apply (9) with $k - 1$ instead of k to get:

$$\begin{aligned} \mu(T_A > kf_A - 2n) &\leq (\psi(n) + 1) \mu(T_A > (k-1)f_A - 2n) \mu(T_A^{[(k-1)f_A]} > f_A - 2n) \\ &= (\psi(n) + 1) \mu(T_A > (k-1)f_A - 2n) \mu(T_A > f_A - 2n). \end{aligned}$$

The equality follows by stationarity. Iterating this argument, one concludes that:

$$\mu(T_A > kf_A - 2n) \leq (\psi(n) + 1)^{k-1} \mu(T_A > f_A - 2n)^k. \tag{10}$$

Applying the resulting inequality in (9), we get Statement (a). In a similar way, ϕ -mixing gives:

$$\mu(T_A > kf_A; B) \leq \mu(T_A > kf_A - 2n) (\mu(B) + \phi(n)).$$

Thus:

$$\begin{aligned} \mu(T_A > kf_A - 2n) &\leq \mu(T_A > (k-1)f_A - 2n) (\mu(T_A^{[(k-1)f_A]} > f_A - 2n) + \phi(n)) \\ &\leq (\mu(T_A > f_A - 2n) + \phi(n))^k. \end{aligned} \tag{11}$$

which ends the proof of (b).

The proof for (c) follows the same lines as Item (a), by observing that for $A, B \in \mathcal{F}_0^i$ and $C \in \mathcal{F}_{i+n}^\infty$, the ψ -mixing property implies $\mu_A(B; C) \leq \mu_A(B) \mu(C) (\psi(n) + 1)$. \square

The next proposition is the key to the proof of Theorem 1, and the idea is the following. We work under the time scale f_A . When $t = kf_A$, $k \in \mathbb{N}$, then we simply cut t into k pieces of equal size f_A . Then, the case of general $t = kf_A + r$, $r < f_A$ is approximated by its integer part kf_A . Technically, this is done in (b) and (a), respectively.

Proposition 3. *Let μ be a ϕ or ψ -mixing measure. Then, for all $n \geq n'$, $A \in \mathcal{A}^n$ and positive integer k , the following inequalities hold:*

- (a) For $0 \leq r \leq f_A$:

1. $|\mu(T_A > kf_A + r) - \mu(T_A > kf_A)\mu(T_A > r)| \leq C'(\psi(n) + 1)^{k-1}\mu(T_A > f_A - 2n)^k\epsilon_\psi(A)$
2. $|\mu(T_A > kf_A + r) - \mu(T_A > kf_A)\mu(T_A > r)| \leq C'(\mu(T_A > f_A - 2n) + \phi(n))^k\epsilon_\phi(A)$
3. $|\mu_A(T_A > kf_A + r) - \mu_A(T_A > kf_A)\mu(T_A > r)| \leq C'((\psi(n) + 1)\mu(T_A > f_A - 2n))^{k-1}\epsilon_\psi(A)$.

(b) For $k \geq 1$:

1. $|\mu(T_A > kf_A) - \mu(T_A > f_A)^k| \leq C'\epsilon_\psi(A)(k-1)(\psi(n) + 1)^{k-2}\mu(T_A > f_A - 2n)^{k-1}$
2. $|\mu(T_A > kf_A) - \mu(T_A > f_A)^k| \leq C'\epsilon_\phi(A)(k-1)(\mu(T_A > f_A - 2n) + \phi(n))^{k-1}$
3. $|\mu_A(T_A > kf_A) - \mu_A(T_A > f_A)\mu(T_A > f_A)^{k-1}| \leq C'\epsilon_\psi(A)(k-1)((\psi(n) + 1)\mu(T_A > f_A - 2n))^{k-2}$

where $C' = 2(M + 1)$ for the cases involving ψ and $C' = 4$ for ϕ .

Proof. We will prove Items (a)-1 and (a)-2 together. Initially, consider the case in which $r < 2n$. In this case, for all $n \geq n'$, we have:

$$\begin{aligned} & |\mu(T_A > kf_A + r) - \mu(T_A > kf_A)\mu(T_A > r)| \\ & \leq \left| \mu\left(T_A > kf_A, T_A^{[kf_A]} > r\right) - \mu(T_A > kf_A) \right| + \mu(T_A > kf_A)|1 - \mu(T_A > r)| \\ & \leq \mu\left(T_A > kf_A, T_A^{[kf_A]} \leq r\right) + \mu(T_A > kf_A)\mu(T_A \leq r). \end{aligned} \tag{12}$$

By Lemma 3-(a) and (10), the last sum is bounded by:

$$\begin{aligned} & ((\psi(n) + 1)\mu(T_A > f_A - 2n))^k\mu\left(T_A^{[kf_A]} \leq r\right) + \mu(T_A > kf_A - 2n)r\mu(A) \\ & \leq ((\psi(n) + 1)\mu(T_A > f_A - 2n))^k\mu(T_A \leq r) + (\psi(n) + 1)^{k-1}\mu(T_A > f_A - 2n)^k r\mu(A) \\ & \leq (\psi(n) + 1)^{k-1}\mu(T_A > f_A - 2n)^k r\mu(A)(M + 1) \\ & \leq 2(M + 1)(\psi(n) + 1)^{k-1}\mu(T_A > f_A - 2n)^k\epsilon_\psi(A) \end{aligned} \tag{13}$$

which gives us (a)-1. To get (a)-2 for $r < 2n$, we apply Lemma 3-(b) and (11) in a similar way. Thus, (12) is bounded by:

$$\begin{aligned} & (\mu(T_A > f_A - 2n) + \phi(n))^k(\mu(T_A \leq r) + \phi(n)) + (\mu(T_A > f_A - 2n) + \phi(n))^k\mu(T_A \leq r) \\ & \leq 4(\mu(T_A > f_A - 2n) + \phi(n))^k\epsilon_\phi(A). \end{aligned}$$

We now consider the case $r \geq 2n$. The triangle inequality gives us:

$$\begin{aligned} & |\mu(T_A > kf_A + r) - \mu(T_A > kf_A)\mu(T_A > r)| \\ & \leq \left| \mu\left(T_A > kf_A; T_A^{[kf_A]} > r\right) - \mu\left(T_A > kf_A; T_A^{[kf_A+2n]} > r - 2n\right) \right| \end{aligned} \tag{14}$$

$$+ \left| \mu\left(T_A > kf_A; T_A^{[kf_A+2n]} > r - 2n\right) - \mu(T_A > kf_A)\mu\left(T_A^{[kf_A+2n]} > r - 2n\right) \right| \tag{15}$$

$$+ \left| \mu(T_A > kf_A)\mu\left(T_A^{[kf_A+2n]} > r - 2n\right) - \mu(T_A > kf_A)\mu(T_A > r) \right|. \tag{16}$$

We proceed as in (5) and use Lemma 3-(a) to get:

$$\begin{aligned} (14) & \leq \mu\left(T_A > kf_A; T_A^{[kf_A]} \leq 2n\right) \\ & \leq ((\psi(n) + 1)(\mu(T_A > f_A - 2n))^k\mu(T_A \leq 2n) \\ & \leq 2n\mu(A)((\psi(n) + 1)(\mu(T_A > f_A - 2n))^k. \end{aligned} \tag{17}$$

For the case ϕ , we apply Lemma 3-(b) and get:

$$(14) \leq (\mu(T_A > f_A - 2n) + \phi(n))^k (2n\mu(A) + \phi(n)). \tag{18}$$

By ψ -mixing and (10):

$$\begin{aligned} (15) &\leq \mu(T_A > kf_A - 2n)\psi(n) \\ &\leq (\psi(n) + 1)^{k-1} \mu(T_A > f_A - 2n)^k \psi(n) \end{aligned} \tag{19}$$

Applying ϕ -mixing and (11):

$$(15) \leq (\mu(T_A > f_A - 2n) + \phi(n))^k \phi(n) \tag{20}$$

Finally, using the shift-invariance and the same arguments as above:

$$\begin{aligned} (16) &= \mu(T_A > kf_A)\mu(r - 2n < T_A \leq r) \\ &\leq 2n\mu(A)\mu(T_A > kf_A - 2n). \end{aligned} \tag{21}$$

Therefore, (10), (17), (19), and (21) give us:

$$(14) + (15) + (16) \leq 2(M + 1)(\psi(n) + 1)^{k-1} \mu(T_A > f_A - 2n)^k \epsilon_\psi(A)$$

and from (11), (18), (20), and (21), we get:

$$(14) + (15) + (16) \leq 4(\mu(T_A > f_A - 2n) + \phi(n))^k \epsilon_\phi(A)$$

which ends the proof of (a)-1 and (a)-2.

For the proof of (a)-3, we write a similar triangle inequality as above:

$$\begin{aligned} &|\mu_A(T_A > kf_A + r) - \mu_A(T_A > kf_A)\mu(T_A > r)| \\ &\leq \left| \mu_A(T_A > kf_A; T_A^{[kf_A]} > r) - \mu_A(T_A > kf_A; T_A^{[kf_A+2n]} > r - 2n) \right| \\ &+ \left| \mu_A(T_A > kf_A; T_A^{[kf_A+2n]} > r - 2n) - \mu_A(T_A > kf_A)\mu(T_A^{[kf_A+2n]} > r - 2n) \right| \\ &+ \mu_A(T_A > kf_A) \left| \mu(T_A^{[kf_A+2n]} > r - 2n) - \mu(T_A > r) \right|. \end{aligned}$$

Then, we follow the same as we did for (a)-1, but applying Item (c) of Lemma 3 and using the ψ -mixing property:

$$|\mu_A(B; C) - \mu_A(B)\mu(C)| \leq \mu_A(B)\mu(C)\psi(n)$$

where $A, B \in \mathcal{F}_0^i$ and $C \in \mathcal{F}_{i+n}^\infty$. For the case $r < 2n$, we use:

$$\begin{aligned} &|\mu_A(T_A > kf_A + r) - \mu_A(T_A > kf_A)\mu(T_A > r)| \\ &\leq \left| \mu_A(T_A > kf_A, T_A^{[kf_A]} > r) - \mu_A(T_A > kf_A) \right| + \mu_A(T_A > kf_A) |1 - \mu(T_A > r)| \end{aligned}$$

and proceed as we did in (13), applying again Lemma 3-(c). This ends Item (a).

We now come to the proof of Items (b)-1 and (b)-2. For $k = 1$, we have an equality. For $k \geq 2$, we get:

$$\begin{aligned} & \left| \mu(T_A > kf_A) - \mu(T_A > f_A)^k \right| \\ &= \left| \sum_{j=2}^k (\mu(T_A > jf_A) - \mu(T_A > (j-1)f_A)\mu(T_A > f_A))\mu(T_A > f_A)^{k-j} \right| \\ &\leq \sum_{j=2}^k |\mu(T_A > jf_A) - \mu(T_A > (j-1)f_A)\mu(T_A > f_A)|\mu(T_A > f_A)^{k-j}. \end{aligned} \tag{22}$$

We put $r = f_A$ in Item (a)-1 to obtain (b)-1:

$$\begin{aligned} (22) &\leq 2(M+1)\epsilon_\psi(A) \sum_{j=2}^k (\psi(n)+1)^{j-2} \mu(T_A > f_A - 2n)^{j-1} \mu(T_A > f_A)^{k-j} \\ &\leq 2(M+1)\epsilon_\psi(A)(k-1)(\psi(n)+1)^{k-2} \mu(T_A > f_A - 2n)^{k-1}. \end{aligned}$$

Furthermore, we get the inequality (b)-2, under ϕ -mixing, proceeding similarly as above:

$$\begin{aligned} (22) &\leq 4\epsilon_\phi(A) \sum_{j=2}^k (\mu(T_A > f_A - 2n) + \phi(n))^{j-1} (\mu(T_A > f_A - 2n) + \phi(n))^{k-j} \\ &= 4\epsilon_\phi(A)(k-1)(\mu(T_A > f_A - 2n) + \phi(n))^{k-1} \end{aligned}$$

Finally, we prove (b)-3 applying (a)-3 as follows:

$$\begin{aligned} & \left| \mu_A(T_A > kf_A) - \mu_A(T_A > f_A)\mu(T_A > f_A)^{k-1} \right| \\ &\leq \sum_{j=2}^k |\mu_A(T_A > jf_A) - \mu_A(T_A > (j-1)f_A)\mu(T_A > f_A)|\mu(T_A > f_A)^{k-j} \\ &\leq 2(M+1)\epsilon_\psi(A) \sum_{j=2}^k ((\psi(n)+1)\mu(T_A > f_A - 2n))^{j-2} \mu(T_A > f_A)^{k-j} \\ &\leq 2(M+1)\epsilon_\psi(A)(k-1)((\psi(n)+1)\mu(T_A > f_A - 2n))^{k-2}. \end{aligned}$$

□

The next two lemmas are classical results and are stated without proof. The first one establishes the reversibility of certain sets for stationary measures, and the second one is a discrete version of the mean value theorem, which follows with a straightforward computation.

Lemma 4. Let μ be a stationary measure. For all positive $i \in \mathbb{N}$, $n \geq 1$ and $A \in \mathcal{A}^n$:

$$\mu(T_A = i) = \mu(T_A > i - 1; A)$$

Lemma 5. Given $a_1, \dots, a_n, b_1, \dots, b_n$ real numbers such that $0 \leq a_i, b_i \leq 1$, the following inequality holds:

$$\left| \prod_{i=1}^n a_i - \prod_{i=1}^n b_i \right| \leq \sum_{i=1}^n |a_i - b_i| \left(\max_{1 \leq i \leq n} \{a_i, b_i\} \right)^{n-1} \leq \sum_{i=1}^n |a_i - b_i|.$$

4.2. Proof of Theorem 1

Theorem 1 contains eight statements, each statement corresponding to the choice of:

- recurrence time: hitting or return,

- mixing property: ψ or ϕ ,
- amplitude of t : smaller or larger than f_A .

Recall the definition of n' in (4). The proof of Theorem 1 holds for all $n \geq n_0$, where n_0 is explicitly given by:

$$n_0 := \inf \left\{ m \geq n'; \sup_{A \in \mathcal{A}^n} \mu(A)\tau(A) < 1/2, \forall n \geq m \right\} \tag{23}$$

which is finite since $\sup_{A \in \mathcal{A}^n} \mu(A)\tau(A) \xrightarrow{n} 0$. Then, in particular, we have $\tau(A) < f_A$ for all $n \geq n_0$ and $A \in \mathcal{A}^n$.

4.2.1. Proofs of the Statements for Small t 's

Here, we assume that $1 \leq t \leq f_A := [2\mu(A)]^{-1}$.

Proof of hitting time, ϕ and ψ together. Recall that $\epsilon(A)$ denotes $\epsilon_\phi(A)$ or $\epsilon_\psi(A)$, depending on whether the measure is ϕ or ψ -mixing.

Applying the inequality $|1 - e^{-x}| \leq x$ for $x \geq 0$, we obtain the statement for $1 \leq t \leq \tau(A)$ as follows:

$$\begin{aligned} \left| \mu(T_A > t) - e^{-\rho(A)\mu(A)t} \right| &\leq \left| \mu(T_A > t) - 1 \right| + \left| 1 - e^{-\rho(A)\mu(A)t} \right| \\ &\leq 2\tau(A)\mu(A). \end{aligned} \tag{24}$$

We consider now the case $\tau(A) < t \leq f_A$. For positive $i \in \mathbb{N}$, define:

$$p_i = \frac{\mu_A(T_A > i - 1)}{\mu(T_A > i - 1)}.$$

Then:

$$\begin{aligned} \frac{\mu(T_A > t)}{\mu(T_A > \tau(A))} &= \prod_{i=\tau(A)+1}^t \frac{\mu(T_A > i)}{\mu(T_A > i - 1)} = \prod_{i=\tau(A)+1}^t (1 - \mu(T_A = i | T_A > i - 1)) \\ &= \prod_{i=\tau(A)+1}^t (1 - \mu(A)p_i) \end{aligned} \tag{25}$$

where we used Lemma 4 in the last equality.

Thus, for $\tau(A) < t \leq f_A$, we apply (25) to obtain:

$$\begin{aligned} &\left| \mu(T_A > t) - e^{-\rho(A)\mu(A)t} \right| \\ &= \left| \mu(T_A > \tau(A)) \prod_{i=\tau(A)+1}^t (1 - \mu(A)p_i) - e^{-\rho(A)\mu(A)\tau(A)} \prod_{i=\tau(A)+1}^t e^{-\rho(A)\mu(A)} \right| \\ &\leq \left| \mu(T_A > \tau(A)) - e^{-\rho(A)\mu(A)\tau(A)} \right| + \left| \prod_{i=\tau(A)+1}^t (1 - \mu(A)p_i) - \prod_{i=\tau(A)+1}^t e^{-\rho(A)\mu(A)} \right| \\ &\leq 2\tau(A)\mu(A) + \sum_{i=\tau(A)+1}^t \left| 1 - \mu(A)p_i - e^{-\rho(A)\mu(A)} \right| \end{aligned} \tag{26}$$

where the two inequalities follow from Lemma 5 and (24).

On the other hand, by the triangle inequality:

$$\left| 1 - p_i\mu(A) - e^{-\rho(A)\mu(A)} \right| \leq |p_i - \rho(A)|\mu(A) + \left| 1 - \rho(A)\mu(A) - e^{-\rho(A)\mu(A)} \right|. \tag{27}$$

Since $|1 - x - e^{-x}| \leq \frac{x^2}{2}$ for all $0 \leq x \leq 1$, by doing $x = \rho(A)\mu(A)$, we get:

$$\left|1 - \rho(A)\mu(A) - e^{-\rho(A)\mu(A)}\right| \leq \frac{\rho(A)^2\mu(A)^2}{2} \leq \frac{\epsilon(A)\mu(A)}{2}.$$

Furthermore, still for $\tau(A) + 1 \leq i \leq f_A + 1$, Proposition 2 gives us:

$$|p_i - \rho(A)| = \left|\frac{\mu_A(T_A > i - 1)}{\mu(T_A > i - 1)} - \rho(A)\right| \leq \frac{C\epsilon(A)}{\mu(T_A > i - 1)} \leq 2C\epsilon(A),$$

where, for the last inequality, we used:

$$\mu(T_A > i - 1) = 1 - \mu(T_A \leq i - 1) \geq 1 - (i - 1)\mu(A) \geq 1 - f_A\mu(A) = \frac{1}{2}.$$

Thus, applying (27), we obtain for $\tau(A) + 1 \leq i \leq f_A + 1$:

$$\left|1 - p_i\mu(A) - e^{-\rho(A)\mu(A)}\right| \leq (2C + 1/2)\epsilon(A)\mu(A). \tag{28}$$

Therefore, (26) and (28) give us:

$$\begin{aligned} \left|\mu(T_A > t) - e^{-\rho(A)\mu(A)t}\right| &\leq 2\tau(A)\mu(A) + (2C + 1/2)(t - \tau(A))\epsilon(A)\mu(A) \\ &\leq (2C + 1/2)[\tau(A)\mu(A) + t\mu(A)\epsilon(A)] \end{aligned} \tag{29}$$

which concludes the statement of Theorem 1 for hitting time at small t 's (with either ϕ or ψ). \square

Proof for return time, ϕ and ψ together. We first note that the statement is trivial for $t = \tau(A)$, then we consider $t > \tau(A)$. By definition, we have $\mu_A(T_A > t) = p_{t+1}\mu(T_A > t)$. Then, we use again the triangle inequality to write:

$$\begin{aligned} &\left|\mu_A(T_A > t) - \rho(A)e^{-\rho(A)\mu(A)(t-\tau(A))}\right| \\ &\leq \mu(T_A > t)|p_{t+1} - \rho(A)| + \rho(A)\left|\mu(T_A > t) - e^{-\rho(A)\mu(A)(t-\tau(A))}\right|. \end{aligned} \tag{30}$$

As we saw before, the first modulus above is bounded by $2C\epsilon(A)$. On the other hand, we apply (25) to obtain for $\tau(A) < t \leq f_A$:

$$\left|\mu(T_A > t) - e^{-\rho(A)\mu(A)(t-\tau(A))}\right| = \left|\mu(T_A > \tau(A)) \prod_{i=\tau(A)+1}^t (1 - \mu(A)p_i) - \prod_{i=\tau(A)+1}^t e^{-\rho(A)\mu(A)}\right|.$$

This is bounded, applying Lemma 5, by:

$$\begin{aligned} &|\mu(T_A > \tau(A)) - 1| + \left|\prod_{i=\tau(A)+1}^t (1 - \mu(A)p_i) - \prod_{i=\tau(A)+1}^t e^{-\rho(A)\mu(A)}\right| \\ &\leq \tau(A)\mu(A) + (2C + 1/2)t\mu(A)\epsilon(A) \end{aligned}$$

where the last inequality follows from (26) and (28). Finally, notice that $t\mu(A) \leq f_A\mu(A) = 1/2$ and $\tau(A)\mu(A) \leq 2\epsilon(A)$ (use Lemma 2 for ψ). Therefore, we obtain from (30):

$$\left|\mu_A(T_A > t) - \rho(A)e^{-\rho(A)\mu(A)(t-\tau(A))}\right| \leq (3C + 9/4)\epsilon(A) \tag{31}$$

This concludes the statement of Theorem 1 for the return time at small t 's (with either ϕ or ψ). \square

4.2.2. Proof of the Statements for Large t 's

The proof for the return time for $t > f_A$ was given in [6] under ϕ -mixing, a finite alphabet, and a complete grammar. The proof still holds if one just assumes a countable alphabet and an incomplete grammar (recall Remark 2 for the uniform convergence to zero of the error term ϵ_ϕ). Thus, we focus on the hitting time under each mixing assumption and the return time only under ψ -mixing.

Proof of Theorem 1 for hitting times, for $t > f_A$. Write $t = kf_A + r$ with integer $k \geq 1$ and $0 \leq r < f_A$. Thus, we have:

$$\left| \mu(T_A > t) - e^{-\rho(A)\mu(A)t} \right| \leq \left| \mu(T_A > kf_A + r) - \mu(T_A > kf_A)\mu(T_A > r) \right| \tag{32}$$

$$+ \left| \mu(T_A > kf_A) - \mu(T_A > f_A)^k \right| \mu(T_A > r) \tag{33}$$

$$+ \left| \mu(T_A > f_A)^k - e^{-\rho(A)\frac{k}{2}} \right| \mu(T_A > r) \tag{34}$$

$$+ \left| e^{-\rho(A)\frac{k}{2}}\mu(T_A > r) - e^{-\rho(A)\mu(A)t} \right|. \tag{35}$$

In order to get an upper bound for the sum of (32) and (33), we analyse the ψ and ϕ cases separately and start with the ψ -mixing. Applying Items (a)-1 and (b)-1 of Proposition 3, that sum is bounded by:

$$\begin{aligned} &\leq C'\epsilon_\psi(A)(\psi(n) + 1)^{k-1}\mu(T_A > f_A - 2n)^k \left(1 + (k - 1)((\psi(n) + 1)\mu(T_A > f_A - 2n))^{-1}\right) \\ &\leq 2(M + 1)\epsilon_\psi(A)((\psi(n) + 1)\mu(T_A > f_A - 2n))^k 2k \\ &\leq 8(M + 1)\epsilon_\psi(A)\mu(A)t((\psi(n) + 1)\mu(T_A > f_A - 2n))^k. \end{aligned} \tag{36}$$

where the last two inequalities are justified by $\mu(T_A > f_A - 2n)^{-1} \leq \mu(T_A > f_A)^{-1} \leq 2$ and $k \leq 2\mu(A)t$.

On the other hand, applying (29) with $t = f_A - 2n$, we get:

$$\begin{aligned} \left| \mu(T_A > f_A - 2n) - e^{-\rho(A)\mu(A)(f_A - 2n)} \right| &= \left| \mu(T_A > f_A - 2n) - e^{-\frac{\rho(A)}{2} + 2n\rho(A)\mu(A)} \right| \\ &\leq (2C + 1/2)(\tau(A)\mu(A) + (f_A - 2n)\mu(A)\epsilon_\psi(A)) \\ &\leq (5C + 5/4)\epsilon_\psi(A) \end{aligned}$$

where we use $\tau(A)\mu(A) \leq 2\epsilon_\psi(A)$.

Furthermore, by the Mean Value Theorem (MVT):

$$\begin{aligned} \left| e^{-\frac{\rho(A)}{2} + 2n\rho(A)\mu(A)} - e^{-\frac{\rho(A)}{2}} \right| &\leq 2n\rho(A)\mu(A)e^{-\frac{\rho(A)}{2} + 2n\rho(A)\mu(A)} \\ &\leq 2n\mu(A)e^{2n\mu(A)} \leq \frac{11}{2}n\mu(A) \end{aligned}$$

since for $n \geq n_0$, we have $2n\mu(A) \leq 2 \sup \mu(A)\tau(A) \leq 1$.

Thus, it follows that:

$$\begin{aligned} &\left| (\psi(n) + 1)\mu(T_A > f_A - 2n) - e^{-\frac{\rho(A)}{2}} \right| \\ &\leq \psi(n) + \left| \mu(T_A > f_A - 2n) - e^{-\frac{\rho(A)}{2} + 2n\rho(A)\mu(A)} \right| + \left| e^{-\frac{\rho(A)}{2} + 2n\rho(A)\mu(A)} - e^{-\frac{\rho(A)}{2}} \right| \\ &\leq (5C + 27/4)\epsilon_\psi(A). \end{aligned}$$

Therefore:

$$((\psi(n) + 1)\mu(T_A > f_A - 2n))^k \leq \left(e^{-\frac{\rho(A)}{2}} + (5C + 27/4)\epsilon_\psi(A) \right)^k.$$

Since $e^x - 1 \geq x \forall x \in \mathbb{R}$, by doing $K = (5C + 27/4)e^{1/2}$, we get:

$$\begin{aligned} (e^{K\epsilon_\psi(A)} - 1) &\geq K\epsilon_\psi(A) \geq (5C + 27/4)\epsilon_\psi(A)e^{\frac{\rho(A)}{2}} \\ \implies e^{-\frac{\rho(A)}{2}}(e^{K\epsilon_\psi(A)} - 1) &\geq (5C + 27/4)\epsilon_\psi(A) \\ \implies e^{-\frac{\rho(A)}{2} + K\epsilon_\psi(A)} &\geq (5C + 27/4)\epsilon_\psi(A) + e^{-\frac{\rho(A)}{2}}. \end{aligned} \tag{37}$$

Now, using that $k = 2\mu(A)(t - r)$, we have:

$$\begin{aligned} ((\psi(n) + 1)\mu(T_A > f_A - 2n))^k &\leq \left(e^{-\frac{\rho(A)}{2} + K\epsilon_\psi(A)} \right)^k \\ &= e^{-\rho(A)\mu(A)t + \rho(A)\mu(A)r + 2K\epsilon_\psi(A)\mu(A)t - 2K\epsilon_\psi(A)\mu(A)r} \\ &\leq e^{-\mu(A)t(\rho(A) - 2K\epsilon_\psi(A))} e^{\mu(A)r} \\ &\leq e^{1/2} e^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))} \end{aligned} \tag{38}$$

where the last inequality follows from $e^{\mu(A)r} \leq e^{\mu(A)f_A}$.

Therefore, it follows from (36) that the sum of (32) and (33) is bounded by:

$$14(M + 1)\epsilon_\psi(A)\mu(A)te^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))}.$$

We now turn to the case of ϕ -mixing. We apply Items (a)-2 and (b)-2 of Proposition 3 to get an upper bound for the sum of (32) and (33):

$$\begin{aligned} &|\mu(T_A > kf_A + r) - \mu(T_A > kf_A)\mu(T_A > r)| + |\mu(T_A > kf_A) - \mu(T_A > f_A)|^k \mu(T_A > r) \\ &\leq 4\epsilon_\phi(A)(\mu(T_A > f_A - 2n) + \phi(n))^k (1 + (k - 1)(\mu(T_A > f_A - 2n) + \phi(n))^{-1}) \\ &\leq 4\epsilon_\phi(A)(\mu(T_A > f_A - 2n) + \phi(n))^k 2k \\ &\leq 16\epsilon_\phi(A)\mu(A)t(\mu(T_A > f_A - 2n) + \phi(n))^k. \end{aligned}$$

Similarly to the ψ -mixing case, one obtains:

$$(\mu(T_A > f_A - 2n) + \phi(n))^k \leq e^{1/2} e^{-\mu(A)t(\rho(A) - C_3\epsilon_\phi(A))}$$

which implies in the ϕ -mixing case that the sum of (32) and (33) is bounded by:

$$27\epsilon_\phi(A)\mu(A)te^{-\mu(A)t(\rho(A) - C_3\epsilon_\phi(A))}.$$

Now, we will treat the cases ψ and ϕ together to obtain upper bounds for (34) and (35). In order to get an upper bound for (34), we apply (29) with $t = f_A$:

$$\begin{aligned} \left| \mu(T_A > f_A) - e^{-\rho(A)\mu(A)f_A} \right| &= \left| \mu(T_A > f_A) - e^{-\frac{\rho(A)}{2}} \right| \\ &\leq (2C + 1/2)(\tau(A)\mu(A) + f_A\mu(A)\epsilon(A)) \\ &\leq (5C + 5/4)\epsilon(A). \end{aligned} \tag{39}$$

Thus, applying Lemma 5, we have:

$$\begin{aligned} & \left| \mu(T_A > f_A)^k - e^{-\rho(A)\frac{k}{2}} \right| \\ & \leq \sum_{i=1}^k \left| \mu(T_A > f_A) - e^{-\frac{\rho(A)}{2}} \right| \left(\max \left\{ \mu(T_A > f_A), e^{-\frac{\rho(A)}{2}} \right\} \right)^{k-1}. \end{aligned}$$

The max is bounded using (39) by:

$$e^{-\frac{\rho(A)}{2}} + (5C + 5/4)\epsilon(A).$$

Naturally, the absolute value is also bounded by using (39), and we get that the above sum is bounded above by:

$$k(5C + 5/4)\epsilon(A) \left(e^{-\frac{\rho(A)}{2}} + (5C + 5/4)\epsilon(A) \right)^{k-1}. \tag{40}$$

Recalling that $k = 2\mu(A)(t - r)$ and proceeding as we did for (37) and (38), we get the following upper bound for (34):

$$2(5C + 5/4)\epsilon(A)\mu(A)t e^{-\mu(A)t(\rho(A) - C_3\epsilon(A))} e^1 \leq 7(4C + 1)\epsilon(A)\mu(A)t e^{-\mu(A)t(\rho(A) - C_3\epsilon(A))}.$$

To conclude the proof for the hitting time, we apply (29) with $t = r$ to bound (35) as follows:

$$\begin{aligned} \left| e^{-\rho(A)\frac{k}{2}} \mu(T_A > r) - e^{-\rho(A)\mu(A)t} \right| &= e^{-\rho(A)\mu(A)t + \rho(A)\mu(A)r} \left| \mu(T_A > r) - e^{-\rho(A)\mu(A)r} \right| \\ &\leq (2C + 1/2) e^{-\rho(A)\mu(A)t + \mu(A)f_A} (\tau(A)\mu(A) + r\mu(A)\epsilon(A)) \\ &\leq (2C + 1/2) (\tau(A)\mu(A) + f_A\mu(A)\epsilon(A)) e^{-\rho(A)\mu(A)t} e^{1/2} \\ &\leq (17C + 5)\epsilon(A)\mu(A)t e^{-\mu(A)t(\rho(A) - C_3\epsilon(A))} \end{aligned}$$

where the term $\mu(A)t$ follows from $1 = 2\mu(A)f_A \leq 2\mu(A)t$. \square

Proof of Theorem 1 for the return time, for $t > f_A$ and under ψ -mixing. We use again the triangle inequality to write:

$$\begin{aligned} & \left| \mu_A(T_A > t) - \rho(A)e^{-\rho(A)\mu(A)(t - \tau(A))} \right| \\ & \leq \left| \mu_A(T_A > kf_A + r) - \mu_A(T_A > kf_A)\mu(T_A > r) \right| \tag{41} \end{aligned}$$

$$+ \left| \mu_A(T_A > kf_A) - \mu_A(T_A > f_A)\mu(T_A > f_A)^{k-1} \right| \mu(T_A > r) \tag{42}$$

$$+ \left| \mu_A(T_A > f_A)\mu(T_A > f_A)^{k-1} - \rho(A)e^{-\rho(A)\frac{k}{2}} \right| \mu(T_A > r) \tag{43}$$

$$+ \rho(A)e^{-\rho(A)\frac{k}{2}} \left| \mu(T_A > r) - e^{-\rho(A)\mu(A)(r - \tau(A))} \right|. \tag{44}$$

Applying Items (a)-3 and (b)-3 of Proposition 3, the sum of (41) and (42) is bounded by:

$$\begin{aligned} & 2(M + 1)\epsilon_\psi(A)((\psi(n) + 1)\mu(T_A > f_A - 2n))^{k-1} (1 + (k - 1)((\psi(n) + 1)\mu(T_A > f_A - 2n))^{-1}) \\ & \leq 2(M + 1)\epsilon_\psi(A)((\psi(n) + 1)\mu(T_A > f_A - 2n))^{k-1} 2k \\ & \leq 8(M + 1)\epsilon_\psi(A)\mu(A)t((\psi(n) + 1)\mu(T_A > f_A - 2n))^{k-1}. \end{aligned}$$

Replacing k by $k - 1$ in (38), the last term is bounded above by:

$$8(M + 1)\epsilon_\psi(A)\mu(A)t e^1 e^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))} \leq 22(M + 1)\epsilon_\psi(A)\mu(A)t e^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))}.$$

On the other hand, Lemma 5 gives us:

$$(43) \leq \left(\max \left\{ \mu_A(T_A > f_A), \mu(T_A > f_A), e^{-\rho(A)/2} \right\} \right)^{k-1} \left(\left| \mu_A(T_A > f_A) - \rho(A)e^{-\rho(A)/2} \right| + \sum_{i=1}^{k-1} \left| \mu(T_A > f_A) - e^{-\rho(A)/2} \right| \right)$$

The last sum is bounded by $(5C + 5/4)(k - 1)\epsilon_\psi(A)$ using (39). On the other hand, applying (31) with $t = f_A$ and the MVT, we obtain:

$$\begin{aligned} & \left| \mu_A(T_A > f_A) - \rho(A)e^{-\rho(A)/2} \right| \\ & \leq \left| \mu_A(T_A > f_A) - \rho(A)e^{-\rho(A)/2 + \rho(A)\mu(A)\tau(A)} \right| + \rho(A) \left| e^{-\rho(A)/2 + \rho(A)\mu(A)\tau(A)} - e^{-\rho(A)/2} \right| \\ & \leq (3C + 9/4)\epsilon_\psi(A) + \rho(A)\mu(A)\tau(A)e^{-\rho(A)(1/2 - \mu(A)\tau(A))} \\ & \leq (3C + 17/4)\epsilon_\psi(A) \end{aligned}$$

since $e^{-\rho(A)(1/2 - \mu(A)\tau(A))} \leq 1$ and $\mu(A)\tau(A) \leq 2\epsilon_\psi(A)$ for $n \geq n_0$.

Furthermore, the last inequality implies:

$$\mu_A(T_A > f_A) \leq \rho(A)e^{-\rho(A)/2} + (3C + 17/4)\epsilon_\psi(A) \leq e^{-\rho(A)/2} + (5C + 5/4)\epsilon_\psi(A)$$

and by (39), we get:

$$\max \left\{ \mu_A(T_A > f_A), \mu(T_A > f_A), e^{-\rho(A)/2} \right\} \leq e^{-\rho(A)/2} + (5C + 5/4)\epsilon_\psi(A).$$

Therefore, as we saw in (40), we have:

$$\begin{aligned} (43) & \leq (5C + 5/4)\epsilon_\psi(A)k \left(e^{-\rho(A)/2} + (5C + 5/4)\epsilon_\psi(A) \right)^{k-1} \\ & \leq 2(5C + 5/4)\epsilon_\psi(A)\mu(A)t e^1 e^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))} \\ & \leq (109M + 116)\epsilon_\psi(A)\mu(A)te^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))}. \end{aligned}$$

Finally, by doing $t = r$ in (29) and applying the MVT once again, we get:

$$\begin{aligned} & \left| \mu(T_A > r) - e^{-\rho(A)\mu(A)(r - \tau(A))} \right| \\ & \leq \left| \mu(\tau_A > r) - e^{-\rho(A)\mu(A)r} \right| + \left| e^{-\rho(A)\mu(A)r} - e^{-\rho(A)\mu(A)(r - \tau(A))} \right| \\ & \leq (2C + 1/2)(\tau(A)\mu(A) + r\mu(A)\epsilon_\psi(A)) + \rho(A)\mu(A)\tau(A)e^{-\rho(A)\mu(A)(r - \tau(A))} \\ & \leq (2C + 1/2)(2\epsilon_\psi(A) + f_A\mu(A)\epsilon_\psi(A)) + (7/2)\epsilon_\psi(A) \\ & \leq (5C + 19/4)\epsilon_\psi(A). \end{aligned}$$

The third inequality follows from $e^{-\rho(A)\mu(A)(r - \tau(A))} \leq e^{\rho(A)\mu(A)\tau(A)} \leq e^{1/2}$, since $n \geq n_0$. Now, just note that $\rho(A)\mu(A)\tau(A) \leq 2\epsilon_\psi(A)$.

Therefore, we finish the proof by obtaining the following upper bound:

$$\begin{aligned} (44) & \leq (5C + 19/4)\epsilon_\psi(A)e^{\rho(A)\mu(A)r}e^{-\rho(A)\mu(A)t} \\ & \leq (5C + 19/4)\epsilon_\psi(A)2\mu(A)t e^{f_A\mu(A)}e^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))} \\ & \leq (66M + 82)\epsilon_\psi(A)\mu(A)te^{-\mu(A)t(\rho(A) - C_3\epsilon_\psi(A))}. \end{aligned}$$

□

4.3. Proof of Theorem 2

Proof of Statement (a). For each $x \in \mathcal{X}$, we define:

$$\tau(x) := \sup\{\tau(x_0^{n-1}), n \geq 1\}.$$

Let $\mathcal{B} = \{x \in \mathcal{X}; \tau(x) = \infty\}$ be the set of aperiodic points of \mathcal{X} . For $x \in \mathcal{B}$, denote $A_n = A_n(x)$, and consider the case $\tau(A_n) < n$. Then, we have:

$$\begin{aligned} 1 - \rho(A_n) &= \mu_{A_n}(T_{A_n} = \tau(A_n)) \\ &= \mu_{A_n}\left(\sigma^{-n}\left(A_n^{(\tau(A_n))}\right)\right) \\ &\leq \mu_{A_n}\left(\sigma^{-n-\lfloor \tau(A_n)/2 \rfloor}\left(A_n^{\lceil \tau(A_n)/2 \rceil}\right)\right) \\ &\leq \mu\left(A_n^{\lceil \tau(A_n)/2 \rceil}\right) + \phi(\lfloor \tau(A_n)/2 \rfloor + 1). \end{aligned}$$

Since $x \in \mathcal{B}$, we have $\tau(A_n) \xrightarrow{n} \infty$, which implies that the last expression converges to zero. For the case $\tau(A_n) \geq n$, we use the same argument:

$$\begin{aligned} 1 - \rho(A_n) &= \mu_{A_n}\left(\sigma^{-\tau(A_n)}(A_n)\right) \\ &\leq \mu_{A_n}\left(\sigma^{-\tau(A_n)-\lfloor n/2 \rfloor}\left(A_n^{\lceil n/2 \rceil}\right)\right) \\ &\leq \mu\left(A_n^{\lceil n/2 \rceil}\right) + \phi(\lfloor n/2 \rfloor + 1) \end{aligned}$$

which also converges to zero. Therefore, $\rho(A_n) \xrightarrow{n} 1$. We conclude the proof by noting that $\mathcal{X} - \mathcal{B}$ is a countable set, and thus, $\mu(\mathcal{B}) = 1$. \square

Proof of Statement (b). By Lemma 2, for ψ -mixing or summable ϕ -mixing measures, there exists $n_0 \geq 1$ such that:

$$\forall n \geq n_0, \forall A \in \mathcal{A}^n, \mu(A)^{-1} > \tau(A).$$

Now, since $\mu_A(T_A > j), j \geq 1$ is a nonincreasing sequence, the potential well is larger than or equal to the arithmetic mean of the subsequent $\mu(A)^{-1} - \tau(A)$ elements:

$$\begin{aligned} \rho(A) = \mu_A(T_A > \tau(A)) &\geq \frac{1}{\mu(A)^{-1} - \tau(A)} \sum_{j=\tau(A)}^{\mu(A)^{-1}-1} \mu_A(T_A > j) \\ &\geq \frac{1}{\mu(A)^{-1}} \sum_{j=\tau(A)}^{\mu(A)^{-1}-1} \mu_A(T_A > j) \\ &= \sum_{j=\tau(A)}^{\mu(A)^{-1}-1} \mu(A; T_A > j) \\ &= \sum_{j=\tau(A)}^{\mu(A)^{-1}-1} \mu(T_A = j + 1). \end{aligned} \tag{45}$$

In the last equality, we used Lemma 4. By (45), one obtains:

$$\begin{aligned} \rho(A) &\geq \mu(T_A \leq \mu(A)^{-1}) - \mu(T_A \leq \tau(A)) \\ &= \mu(T_A \leq \mu(A)^{-1}) - \tau(A)\mu(A) \end{aligned} \tag{46}$$

where the equality follows by stationarity and the definition of $\tau(A)$.

By Lemmas 1 and 2, we know that $\tau(A)\mu(A) \xrightarrow{n} 0$ uniformly. Thus, it is enough to find a strictly positive lower bound for $\mu(T_A \leq \mu(A)^{-1})$. Let:

$$N = \sum_{j=1}^{\mu(A)^{-1}} \mathbb{1}_A \circ \sigma^j$$

which counts the number of occurrences of A up to $\mu(A)^{-1}$. By the so-called second moment method,

$$\mu(T_A \leq \mu(A)^{-1}) = \mu(N \geq 1) \geq \frac{\mathbb{E}(N)^2}{\mathbb{E}(N^2)}. \tag{47}$$

Stationarity gives $\mathbb{E}(N) = 1$. It remains to prove that $\mathbb{E}(N^2)$ is bounded above by a constant. Expanding N^2 , using stationarity and $\mathbb{E}(N) = 1$, we obtain:

$$\mathbb{E}(N^2) = 1 + 2 \sum_{j=1}^{\mu(A)^{-1}} (\mu(A)^{-1} - j) \mu(A \cap \sigma^{-j}(A)). \tag{48}$$

Let us first consider the ϕ -mixing case. For $j \geq n$, mixing gives $\mu(A \cap \sigma^{-j}(A)) \leq \mu(A)^2 + \mu(A)\phi(j - n + 1)$. Thus,

$$\sum_{j=n}^{\mu(A)^{-1}} (\mu(A)^{-1} - j) \mu(A \cap \sigma^{-j}(A)) \leq \frac{1}{2} + \sum_{\ell=0}^{\mu(A)^{-1}-n} \phi(\ell + 1) \tag{49}$$

where we used $\mu(A)^{-1} - j \leq \mu(A)^{-1}$ to get the last term.

For $1 \leq j \leq n - 1$, as before $A^{(j)} \subset A^{(\lceil j/2 \rceil)}$; thus:

$$\begin{aligned} \mu(A \cap \sigma^{-j}(A)) &= \mu(A \cap \sigma^{-n}(A^{(j)})) \\ &\leq \mu(A \cap \sigma^{-n-\lfloor j/2 \rfloor}(A^{(\lceil j/2 \rceil)})) \\ &\leq \mu(A) (\mu(A^{(\lceil j/2 \rceil)}) + \phi(\lfloor j/2 + 1 \rfloor)) \\ &\leq \mu(A) (Ce^{-c\lceil j/2 \rceil} + \phi(\lfloor j/2 + 1 \rfloor)). \end{aligned}$$

Therefore,

$$\sum_{j=1}^{n-1} (\mu(A)^{-1} - j) \mu(A \cap \sigma^{-j}(A)) \leq \sum_{j=1}^{n-1} (Ce^{-c\lceil j/2 \rceil} + \phi(\lfloor j/2 + 1 \rfloor)). \tag{50}$$

Therefore, by (49) and (50), the summability of ϕ concludes the proof for the ϕ -mixing case.

If μ is ψ -mixing, we separate the sum in (48) into three parts. First, recall the definition of g_0 in Section 3.2. For $1 \leq j \leq g_0$, we bound the sum as follows:

$$\sum_{j=1}^{g_0} (\mu(A)^{-1} - j) \mu(A \cap \sigma^{-j}(A)) \leq \sum_{j=1}^{g_0} \mu(A)^{-1} \mu(A) = g_0.$$

For $g_0 + 1 \leq j \leq g_0 + n - 1$, we have by ψ -mixing:

$$(\mu(A)^{-1} - j) \mu(A \cap \sigma^{-j}(A)) \leq \mu(A)^{-1} \mu(A \cap \sigma^{-n-g_0}(A^{(j-g_0)})) \leq M \mu(A)^{-1} \mu(A) \mu(A^{(\ell)})$$

where we denoted $\ell = j - g_0$. Thus:

$$\sum_{j=g_0+1}^{g_0+n-1} (\mu(A)^{-1} - j) \mu(A \cap \sigma^{-j}(A)) \leq M \sum_{\ell=1}^{n-1} Ce^{-c\ell}.$$

Finally, applying ψ -mixing again,

$$\sum_{j=g_0+n}^{\mu(A)^{-1}} (\mu(A)^{-1} - j) \mu(A \cap \sigma^{-j}(A)) \leq M \sum_{j=n+g_0}^{\mu(A)^{-1}} \mu(A)^{-1} \mu(A)^2 \leq M,$$

concluding the proof of the ψ -mixing case. \square

Author Contributions: Conceptualization, M.A., V.A., and S.G.; formal analysis, M.A., V.A., and S.G.; investigation, M.A., V.A., and S.G.; methodology, M.A., V.A., and S.G.; writing, original draft, M.A., V.A., and S.G.; writing, review and editing, M.A., V.A., and S.G. All authors read and agreed to the published version of the manuscript.

Funding: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Chamada Universal 439422/2018-3 and Bolsa de Produtividade 305096/2019-2), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Auxílio Regular 2019/23439-4).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work is part of the PhD Thesis of V.A. under the supervision of M.A. and S.G. at the PIPGEs (Programa Interinstitucional de Pós-Graduação em Estatística, UFSCar-USP). V.A. and S.G. thank IME-USP for hospitality during several stays, and M.A. thanks DEs-UFSCar for hospitality during several stays.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Freitas, A.C.M.; Freitas, J.M.; Todd, M. Hitting time statistics and extreme value theory. *Probab. Theory Relat. Fields* **2010**, *147*, 675–710. [\[CrossRef\]](#)
- Freitas, J.M. Extremal behaviour of chaotic dynamics. *Dyn. Syst.* **2013**, *28*, 302–332. [\[CrossRef\]](#)
- Lucarini, V.; Faranda, D.; de Freitas, J.M.M.; Holland, M.; Kuna, T.; Nicol, M.; Todd, M.; Vaienti, S. *Extremes and Recurrence in Dynamical Systems*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
- Leadbetter, M.R.; Lindgren, G.; Rootzén, H. *Extremes and Related Properties of Random Sequences and Processes*; Springer: Berlin/Heidelberg, Germany, 2012.
- Resnick, S.I. *Extreme Values, Regular Variation and Point Processes*; Springer: Berlin/Heidelberg, Germany, 2013.
- Abadi, M.; Vergne, N. Sharp error terms for return time statistics under mixing conditions. *J. Theoret. Probab.* **2009**, *22*, 18–37. [\[CrossRef\]](#)
- Abadi, M.; Cardeno, L.; Gallo, S. Potential well spectrum and hitting time in renewal processes. *J. Stat. Phys.* **2015**, *159*, 1087–1106. [\[CrossRef\]](#)
- Bradley, R.C. Basic properties of strong mixing conditions. A survey and some open questions. *Probab. Surv.* **2005**, *2*, 107–144. update of, and a supplement to, the 1986 original. [\[CrossRef\]](#)
- Afraimovich, V.; Chazottes, J.R.; Saussol, B. Pointwise dimensions for Poincaré recurrences associated with maps and special flows. *Discrete Contin. Dyn. Syst.* **2003**, *9*, 263–280.
- Saussol, B.; Troubetzkoy, S.; Vaienti, S. Recurrence, dimensions, and Lyapunov exponents. *J. Statist. Phys.* **2002**, *106*, 623–634. [\[CrossRef\]](#)
- Ornstein, D.S.; Weiss, B. Entropy and data compression schemes. *IEEE Trans. Inform. Theory* **1993**, *39*, 78–83. [\[CrossRef\]](#)
- Wyner, A.D.; Ziv, J. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inf. Theory* **1989**, *35*, 1250–1258. [\[CrossRef\]](#)
- Marton, K.; Shields, P. Almost sure waiting time results for weak and very weak bernoulli processes. In Proceedings of the 1994 IEEE International Symposium on Information Theory, Trondheim, Norway, 27 June–1 July 1994.
- Shields, P.C. Waiting times: Positive and negative results on the wyner-ziv problem. *J. Theor. Probab.* **1993**, *6*, 499–519. [\[CrossRef\]](#)
- Abadi, M.; Chazottes, J.; Gallo, S. The complete l^q -spectrum and large deviations for return times for equilibrium states with summable potentials. *arXiv* **2019**, arXiv:1902.03441.
- Chazottes, J.; Ugalde, E. Entropy estimation and fluctuations of hitting and recurrence times for Gibbsian sources. *Discrete Contin. Dyn. Syst. Ser. B* **2005**, *5*, 565–586. [\[CrossRef\]](#)
- Collet, P.; Galves, A.; Schmitt, B. Repetition times for Gibbsian sources. *Nonlinearity* **1999**, *12*, 1225–1237. [\[CrossRef\]](#)
- Abadi, M.; Cardeno, L. Rényi entropies and large deviations for the first match function. *IEEE Trans. Inform. Theory* **2015**, *61*, 1629–1639. [\[CrossRef\]](#)

19. Abadi, M.; Vaienti, S. Large deviations for short recurrence. *Discret. Cont. Dyn. Syst. Ser. A* **2008**, *21*, 729–747. [[CrossRef](#)]
20. Haydn, N.; Vaienti, S. The Rényi entropy function and the large deviation of short return times. *Ergod. Theory Dynam. Syst.* **2010**, *30*, 159–179. [[CrossRef](#)]
21. Abadi, M.; Gallo, S.; Rada-Mora, E.A. The shortest possible return time of β -mixing processes. *IEEE Trans. Inf. Theory* **2017**, *64*, 4895–4906. [[CrossRef](#)]
22. Abadi, M.; Lambert, R. The distribution of the short-return function. *Nonlinearity* **2013**, *26*, 1143–1162. [[CrossRef](#)]
23. Aldous, D.J.; Brown, M. Inequalities for rare events in time-reversible markov chains ii. *Stoch. Process. Their Appl.* **1993**, *44*, 15–25. [[CrossRef](#)]
24. Galves, A.; Schmitt, B. Inequalities for hitting times in mixing dynamical systems. *Random Comput. Dynam.* **1997**, *5*, 337–347.
25. Abadi, M. Exponential approximation for hitting times in mixing processes. *Math. Phys. Electron. J* **2001**, *7*, 1–19.
26. Abadi, M. Sharp error terms and necessary conditions for exponential hitting times in mixing processes. *Ann. Probab.* **2004**, *32*, 243–264. [[CrossRef](#)]
27. Abadi, M. Hitting, returning and the short correlation function. *Bull. Braz. Math. Soc.* **2006**, *37*, 593–609. [[CrossRef](#)]
28. Abadi, M.; Saussol, B. Hitting and returning to rare events for all alpha-mixing processes. *Stoch. Process. Their Appl.* **2011**, *121*, 314–323. [[CrossRef](#)]
29. Abadi, M.; Saussol, B. Almost sure convergence of the clustering factor in α -mixing processes. *Stochastics Dyn.* **2016**, *16*, 1660016. [[CrossRef](#)]