On Large Deviation Property of Recurrence Times

Siddharth Jain

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, India 208016
Email: sidjain@iitk.ac.in

Rakesh K. Bansal

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, India 208016
Email: rkb@iitk.ac.in

Abstract—We extend the study by Ornstein and Weiss on the asymptotic behavior of the normalized version of recurrence times and establish the large deviation property for a certain class of mixing processes. Further, an estimator for entropy based on recurrence times is proposed for which large deviation behavior is proved for stationary and ergodic sources satisfying similar mixing conditions.

I. INTRODUCTION

For a stationary and ergodic source with finite alphabet, the asymptotic relationship between probability of an n length sequence and entropy has been well established by Shannon-McMillan-Breiman Theorem [16]. Later, Ornstein and Weiss [15] established a similar expression relating recurrence times to entropy. Kontoyiannis [12] related recurrence times and probability of an n length sequence for Markov sources by showing that $\lim_{n\to\infty} \log[R_n(X)P(X_1^n)] = o(n^\beta)$ $a.s.$, for any $\beta > 0$. Here, $R_n(X)$ and $P(X_1^n)$ represent random variables for recurrence time and probability of an n-length block generated by a source X, respectively. Further, in [12, Corollary 2], he also identified a class of processes for which central limit theorem (CLT) and law of iterated logarithm (LIL) hold true for recurrence times.

The question of large deviations for Shannon-Mcmillan-Breiman Theorem has been successfully answered in literature under certain mixing conditions [16]. Motivated by Kontoyiannis' results and the satisfaction of large deviation property for Shannon-McMillan-Breiman Theorem, it is natural to ask under what conditions the asymptotic recurrence times relation satisfies large deviation property. Chazottes and Ugalde [6] in 2005 established partial large deviations results on recurrence times for Gibbsian sources. In this paper, we have found a class of processes for which large deviation property holds for recurrence times.

For an i.i.d source, Shannon-McMillan-Breiman Theorem satisfies large deviation property by direct application of Cramer's Theorem [8]. However for Ornstein and Weiss recurrence times result, even for an i.i.d. source Cramer's Theorem is not applicable. This makes the analysis of large deviation property for recurrence times non-trivial even for the i.i.d case. Hence, in order to answer the question of large deviations for recurrence times one needs to look more closely into the recurrence time statistics.

Maurer [14] studied the behavior of recurrence time statistics under the assumption of non-overlapping recurrence blocks for i.i.d sources. Later, Abadi and Galves [3] studied a similar non-overlapping scenario for ψ-mixing processes and established an exponential bound on the recurrence time distribution. Moreover, they also brought out the contrast between overlapping and non-overlapping case. In the context of overlapping $R_n(x)$, there are several references that show convergence in distribution of $R_n(x)P(x_1^n)$ to an exponentially distributed random variable for a certain class of stationary and ergodic processes [2][4][7][9][10]. Kim [11] also studied the behavior of conditional distribution of $R_n(X)P(X_1^n)$ given the n-length block $X_1^n = x_1^n$ and established an exponential bound on its distribution for these two classes of sources i) ψ-mixing, ii) ϕ-mixing with summable coefficients. In this paper, in order to do our analysis we have used this exponential bound on conditional distribution established by Kim [11].

The rest of the paper is organized as follows. In section II we state preliminary results on recurrence time statistics and mixing processes. In section III, we state our main theorems for the large deviation property of recurrence times. In section IV, we give proofs of these theorems and their corollaries. In section V, we define an estimator for entropy based on recurrence times and prove large deviation property for it. In section VI, we present our conclusion.

II. PRELIMINARIES

Let $\{X_n\}_{n=-\infty}^{n=\infty}$ be a stationary and ergodic process defined on the space of infinite sequences $(\mathcal{A}_{-\infty}^{\infty}, \sigma, P)$. Here $\mathcal{A}$ is a finite set of alphabets, σ is the sigma field generated by finite dimensional cylinders and P is the probability measure. For simplicity of notation, we will use X for $\{X_n\}_{n=-\infty}^{n=\infty}$.
X is called ψ-mixing if

$$\sup_{A\epsilon\sigma_{-\infty}^n, \ B\epsilon\sigma_{n+l}^\infty} \frac{|P(A \cap B) - P(A)P(B)|}{P(A)P(B)} \leq \psi(l). \quad (1)$$

Here, $\psi(l)$ is a decreasing sequence converging to 0 and σ_i^j denotes the sigma algebra generated by $X_i^j = X_i X_{i+1}....X_j$ and it is called ϕ-mixing if

$$\sup_{A\epsilon\sigma_{-\infty}^n, \ B\epsilon\sigma_{n+l}^\infty} \frac{|P(A \cap B) - P(A)P(B)|}{P(A)} \leq \phi(l). \quad (2)$$

Here, $\phi(l)$ is a decreasing sequence converging to 0.
Let $\{x_n\}_{n=-\infty}^{n=\infty}$ denote a particular realization of X. Now define, the first return time (recurrence time) of x_1^n to be:

$$R_n(x) = \min\{j \geq 1 : x_1^n = x_{-j+1}^{-j+n}\}.$$

As a dual of recurrence time $R_n(x)$, match length $L_m(x)$ is defined as follows:

$$L_m(x) = \max\{j \geq 1 : x_1^j = x_{-k+1}^{-k+j}, \ k = 1, 2, ..., m\}.$$

Observation 1 [12]: $R_n(x) > m \Leftrightarrow L_m(x) < n$.
On asymptotic behavior of $R_n(x)$ and $L_m(x)$ following holds,
Ornstein and Weiss [15]
For X with entropy rate $H(X)$, with probability 1,

$$\lim_{n \to \infty} \frac{\log R_n(X)}{n} = H(X); \qquad \lim_{m \to \infty} \frac{\log m}{L_m(X)} = H(X).$$

Further, in this paper unless stated otherwise, i) we use H to represent the entropy rate of the source X. ii) A ϕ-mixing process is assumed to be ϕ-mixing in both forward and backward directions.
Kim's Theorem [11]
For X satisfying ψ-mixing condition or ϕ-mixing condition with summable coefficients,

$$P(R_n(X) > t | X_1^n = x_1^n) > \xi_x e^{-t\xi_x P(x_1^n)}(1 - 2\sqrt{C_x(\xi_x P(x_1^n)t \vee 1)}) \quad \forall \ t > 0. \tag{3}$$

$$P(R_n(X) > t | X_1^n = x_1^n) < \xi_x e^{-t\xi_x P(x_1^n)}[1 + K(x,t) + 2C_x(\xi_x t P(x_1^n) \vee 1)] \ \forall \ t \geq \rho_x. \tag{4}$$

where $C_x = C\{\inf_{n \leq \triangle \leq 1/P(x_1^n)}[\triangle P(x_1^n) + *(\triangle)]\}$ ($C > 0$ is a constant, $*$ represents ψ or ϕ), $\rho_x = \frac{2\sqrt{C_x}}{(\sqrt{1+C_x}+\sqrt{C_x})\xi_x P(x_1^n)}$.

$$K(x,t) = 2\sqrt{C_x(t\xi_x P(x_1^n) \vee 1)(1 + C_x(t\xi_x P(x_1^n) \vee 1))},$$

and $C_x \to 0$ ($as \ n \to \infty$) and $\xi_x \in [E_1, E_2], (0 < E_1 < 1 < E_2 < \infty)$. $a_1 \vee a_2$ means $\max\{a_1, a_2\}$. Following additional properties as listed in [11] and originally proved in [7][1] hold for ϕ-mixing processes:

1) For an exponentially ϕ-mixing process, $\forall x_1^n \in \mathcal{A}^n$, there exists a positive constant D_o and $\Gamma > 0$, s.t. $\forall n \geq n_o$

$$C_x \leq D_o e^{-\Gamma n}. \tag{5}$$

2) Let $B_n(s)$ be the set of $x_1^n \in \mathcal{A}^n$; such that $R_n(x) < \frac{n}{s}$. Then, for any ϕ-mixing process, there exists $s \in \mathcal{N}$ ($\mathcal{N}$ being the set of natural numbers), and two positive constants D_1 and d_1 such that

$$P(\{x : x_1^n \in B_n(s)\}) \leq D_1 e^{-d_1 n}. \tag{6}$$

3) For exponentially ϕ-mixing processes for every $x_1^n \in \mathcal{A}^n \backslash B_n(s)$

$$|\xi_x - 1| < D_2 e^{-d_2 n}. \quad (for \ n \ large \ enough) \tag{7}$$

Here, D_2 and d_2 are constants.
Now, we state a Lemma which is required in the proof of Theorem 4 stated in section III. Let A_1, A_2 and A_3 be three sets such that $A_3 = A_1 \cap A_2$. Suppose $P(A_1) > 1 - p_1 e^{-p_2 n}$ and $P(A_2) > 1 - q_1 e^{-q_2 n}$, where p_1, p_2, q_1 and q_2 are positive constants. Then, we have
Lemma 1: $P(A_3) > 1 - (p_1 + q_1)e^{-\min\{p_2, q_2\}n}$.

Lemma 1 is proved in the appendix.
Definition [16]
X is said to have exponential rates for entropy if for every $\epsilon > 0$, we have

$$P(\{x_1^n : 2^{-n(H+\epsilon)} \leq P(x_1^n) \leq 2^{-n(H-\epsilon)}\}) \geq 1 - r(\epsilon, n). \tag{8}$$

where $-\frac{1}{n} \ln r(\epsilon, n)$ is bounded away from 0 or in other words $r(\epsilon, n) = e^{-k(\epsilon)n}$, where $k(\epsilon)$ is a real valued positive function of ϵ.
Theorem 1 [16]
1) I.I.D., ergodic Markov and ψ-mixing processes all have exponential rates for entropy.
2) An aperiodic and irreducible Markov chain is ψ-mixing.

Remark 1: In [16], the ψ-mixing condition used is weaker as to what we have defined in Eq. (1). So, Theorem 1 also holds for processes satisfying the stronger ψ-mixing condition given in Eq. (1).
Theorem 2 [5]
1) If a process is ψ-mixing, then it is also ϕ-mixing.
2) If a Markov Chain is ϕ-mixing then it is exponentially ϕ-mixing.

Corollary 1: From Theorem 1 and 2, it follows that an aperiodic and irreducible Markov Chain is exponentially ϕ-mixing and has exponential rates for entropy.

III. MAIN THEOREMS

Theorem 3: For a process satisfying ψ-mixing condition or ϕ-mixing condition with summable coefficients and with exponential rates for entropy,

$$P(\frac{\log R_n(X)}{n} > H + \epsilon) \leq e^{-f(\epsilon)n} \quad \forall \ n \geq N(\epsilon).$$

where, $f(\epsilon)$ is a real positive valued function for all $\epsilon > 0$ and $f(0) = 0$.
Corollary 2: Under the conditions of Theorem 3, we have

$$P(\frac{\log m}{L_m(X)} > H + \epsilon) \leq e^{-f(\epsilon)\frac{\log m}{H+\epsilon}} \quad \forall \ m \geq M(\epsilon).$$

Theorem 4: For an exponentially ϕ-mixing process,

$$P(\frac{\log R_n(X)}{n} < H - \epsilon) \leq e^{-g(\epsilon)n} \quad \forall \ n \geq N'(\epsilon).$$

where $g(\epsilon)$ is a real positive valued function for all $\epsilon > 0$ and $g(0) = 0$.
Corollary 3: Under the conditions of Theorem 4, we have

$$P(\frac{\log m}{L_m(X)} < H - \epsilon) \leq e^{-g(\epsilon)\frac{\log m}{H-\epsilon}} \quad \forall \ m \geq M'(\epsilon).$$

Theorem 3 and 4 are combined in the form of
Theorem 5 (Large Deviation Property for Recurrence Times)
For an exponentially ϕ-mixing process with exponential rates for entropy,

$$P(|\frac{\log R_n(X)}{n} - H| > \epsilon) \leq 2e^{-I(\epsilon)n} \quad \forall n \geq N''(\epsilon).$$

where, $I(\epsilon) = \min\{f(\epsilon), g(\epsilon)\}$ and $N''(\epsilon) = \max\{N(\epsilon), N'(\epsilon)\}$.

Remark 2: From, Corollary 1, it can be inferred that the quantity $\frac{\log R_n(X)}{n}$ for an aperiodic and irreducible Markov chain satisfies Large Deviation Property.

IV. PROOFS

Proof of Theorem 3:
Let $A_n^{(\delta)}$ be a set of n long sequences defined as,

$$A_n^{(\delta)} = \{x_1^n : 2^{-n(H+\delta)} \le P(x_1^n) \le 2^{-n(H-\delta)}\}.$$

Now,

$$P(\frac{\log R_n(X)}{n} > H + \epsilon) = P(R_n(X) > 2^{n(H+\epsilon)})$$

$$= \sum_{y \epsilon \mathcal{A}^n} P(y) P(R_n(X) > 2^{n(H+\epsilon)} | X_1^n = y)$$

$$= \sum_{y \epsilon A_n^{(\delta)}} P(y) P(R_n(X) > 2^{n(H+\epsilon)} | X_1^n = y)$$

$$+ \sum_{y \epsilon A_n^{(\delta)c}} P(y) P(R_n(X) > 2^{n(H+\epsilon)} | X_1^n = y)$$

$$< \sum_{y \epsilon A_n^{(\delta)}} P(y)[\xi_y e^{-2^{n(H+\epsilon)}\xi_y P(y)}[1 + K(y, 2^{n(H+\epsilon)})$$

$$+ 2C_y(\xi_y 2^{n(H+\epsilon)} P(y) \vee 1)]] + \sum_{y \epsilon A_n^{(\delta)c}} P(y) \quad (a)$$

$$< \sum_{y \epsilon A_n^{(\delta)}} P(y)[E_2 e^{-2^{n(H+\epsilon)}E_1 P(y)}[1 + V$$

$$+ 2d(E_2 2^{n(H+\epsilon)} P(y) \vee 1)] + \sum_{y \epsilon A_n^{(\delta)c}} P(y) \quad (b) \qquad (9)$$

where

$$V = 2\sqrt{d(2^{n(H+\epsilon)}E_2 P(y) \vee 1)(1 + d(2^{n(H+\epsilon)}E_2 P(y) \vee 1))}.$$

(a) follows from the use of inequality (4) and Remark 6 as stated in Appendix. (b) follows from using the fact that $\xi_y \epsilon [E_1, E_2]$ and $C_y \to 0$ as $n \to \infty \Rightarrow C_y < d \; \forall \; y$ and n large enough, where $d > 0$ is an arbitrary constant. For $y \epsilon A_n^{(\delta)}$, we have

$$2^{n(\epsilon-\delta)} \le 2^{n(H+\epsilon)} P(y) \le 2^{n(\epsilon+\delta)}.$$

For every $\epsilon > 0$, choose $\delta = \frac{\epsilon}{2}$. Consequently, we have

$$2^{\frac{n\epsilon}{2}} \le 2^{n(H+\epsilon)} P(y) \le 2^{\frac{3n\epsilon}{2}} \; \forall y \epsilon A_n^{(\frac{\epsilon}{2})}. \qquad (10)$$

Also, $2^{\frac{3n\epsilon}{2}} E_2 > 1$ since $E_2 > 1$. Hence, using (9) and (10) we have,

$$P(\frac{\log R_n(X)}{n} > H + \epsilon) \le \sum_{y \epsilon A_n^{(\frac{\epsilon}{2})}} P(y)[E_2 e^{-E_1 2^{\frac{n\epsilon}{2}}}(1+$$

$$2\sqrt{dE_2 2^{\frac{3n\epsilon}{2}}(1 + dE_2 2^{\frac{3n\epsilon}{2}})} + 2dE_2 2^{\frac{3n\epsilon}{2}})] + \sum_{y \epsilon A_n^{(\frac{\epsilon}{2})c}} P(y). \qquad (11)$$

Using (8) and (11), for processes having exponential rates for entropy and satisfying ψ-mixing condition or ϕ-mixing condition with summable coefficients, we have

$$P(\frac{\log R_n(X)}{n} > H + \epsilon) \le \left\{ E_2 e^{-E_1 2^{\frac{n\epsilon}{2}}}[1 + 2dE_2 2^{\frac{3n\epsilon}{2}} \right.$$

$$\left. + 2\sqrt{dE_2 2^{\frac{3n\epsilon}{2}}(1 + dE_2 2^{\frac{3n\epsilon}{2}})}] \right\} + r(\frac{\epsilon}{2}, n) < e^{-f(\epsilon)n}. \qquad (12)$$

This completes the proof of Theorem 3.

Remark 3: Since the first term on the right hand side of inequality (12) stated above rapidly (super exponentially) converges to 0, $f(\epsilon)$ behaves in a similar manner as $-\frac{\ln r(\frac{\epsilon}{2}, n)}{n} = k(\frac{\epsilon}{2})$. (Also see Remark 7 as stated in Appendix)

Proof of Corollary 2: From Observation 1, we have

$$R_n(x) > 2^{n(H+\epsilon)} \Leftrightarrow L_{2^{n(H+\epsilon)}}(x) < n$$

$$\Rightarrow P(L_{2^{n(H+\epsilon)}}(X) < n) = P(R_n(X) > 2^{n(H+\epsilon)}) < e^{-f(\epsilon)n}.$$

$\forall \; n \ge N(\epsilon)$. Now, letting $m = 2^{n(H+\epsilon)}$, we have

$$P(L_m(X) < \frac{\log m}{H + \epsilon}) < e^{-f(\epsilon)\frac{\log m}{H+\epsilon}} \quad \forall \; m \ge M(\epsilon)$$

$$\Rightarrow P(\frac{\log m}{L_m(X)} > H + \epsilon) < e^{-f(\epsilon)\frac{\log m}{H+\epsilon}}.$$

Proof of Theorem 4: Let $A_n^{(\frac{\epsilon}{2})}$ be the same set as considered in the proof of Theorem 3. For each $y \epsilon A_n^{(\frac{\epsilon}{2})}$, we have

$$2^{-\frac{3n\epsilon}{2}} \le P(y) 2^{n(H-\epsilon)} \le 2^{-\frac{n\epsilon}{2}}. \qquad (13)$$

Now,

$$P(\frac{\log R_n(X)}{n} < H - \epsilon) = 1 - P(\frac{\log R_n(X)}{n} \ge H - \epsilon)$$

$$\le 1 - P(\frac{\log R_n(X)}{n} > H - \epsilon)$$

$$= 1 - P(R_n(X) > 2^{n(H-\epsilon)})$$

$$= 1 - \sum_{y \epsilon \mathcal{A}^n} P(y) P(R_n(X) > 2^{n(H-\epsilon)} | X_1^n = y)$$

$$< 1 - \sum_{y \epsilon \mathcal{A}^n} P(y)[\xi_y e^{-P(y)\xi_y 2^{n(H-\epsilon)}}(1-$$

$$2\sqrt{C_y(\xi_y P(y) 2^{n(H-\epsilon)} \vee 1)})] \quad (a)$$

$$< 1 - \sum_{y \epsilon A_n^{(\frac{\epsilon}{2})}} P(y)[\xi_y e^{-E_2 2^{-\frac{n\epsilon}{2}}}(1 - 2\sqrt{C_y(E_2 2^{-\frac{n\epsilon}{2}} \vee 1)})] \quad (b)$$

$$= 1 - \sum_{y \epsilon A_n^{(\frac{\epsilon}{2})}} P(y)[\xi_y e^{-E_2 2^{-\frac{n\epsilon}{2}}}(1 - 2\sqrt{C_y})] \quad (c) \qquad (14)$$

Here, (a) follows from (3), (b) follows from the fact that $\xi_y \epsilon [E_1, E_2]$ and inequality (13). Also in (b) the negative term contributed by sequences belonging to the set $A_n^{(\frac{\epsilon}{2})c}$ is ignored because we are looking at an upper bound. (c) follows because eventually $E_2 2^{-\frac{n\epsilon}{2}} < 1$, since $E_2 2^{-\frac{n\epsilon}{2}} \to 0$ $(as \; n \to \infty)$.

To proceed further, we introduce the following notations, let $A_1 = \mathcal{A}^n \backslash B_n(s); A_2 = A_n^{(\frac{\epsilon}{2})}$. From (6) and (8), we have

$P(A_1) > 1 - D_1 e^{-d_1 n}$ and $P(A_2) > 1 - e^{-k(\frac{\epsilon}{2})n}$ respectively for processes with exponential rates for entropy. Let $A_3 = A_1 \cap A_2$.

Therefore from (14), we have

$$P(\frac{\log R_n(X)}{n} < H - \epsilon) < 1-$$
$$\sum_{y \epsilon A_3} P(y)\xi_y e^{-E_2 2^{-\frac{n\epsilon}{2}}}(1 - 2\sqrt{C_y})-$$
$$\sum_{y \epsilon A_2 \backslash A_3} P(y)\xi_y e^{-E_2 2^{-\frac{n\epsilon}{2}}}(1 - 2\sqrt{C_y})$$
$$\leq 1 - \sum_{y \epsilon A_3} P(y)(1 - D_2 e^{d_2 n})e^{-E_2 2^{-\frac{n\epsilon}{2}}}(1 - 2\sqrt{D_o}e^{-\frac{\Gamma n}{2}}) \ (d)$$
$$= 1 - P(A_3)(1 - D_2 e^{-d_2 n})e^{-E_2 2^{-\frac{n\epsilon}{2}}}(1 - 2\sqrt{D_o}e^{-\frac{\Gamma n}{2}})$$
$$\leq 1 - \left[e^{-E_2 2^{-\frac{n\epsilon}{2}}}(1 - (D_1 + 1)e^{-\min\{d_1,k(\epsilon)\}n})\right.$$
$$\left.(1 - D_2 e^{-d_2 n})(1 - 2\sqrt{D_o}e^{-\frac{\Gamma n}{2}})\right] \ (e)$$
$$\leq 1 - (1 - C'e^{-u(\epsilon)n})e^{-E_2 2^{-\frac{n\epsilon}{2}}}$$
$$(15)$$

Here, (d) follows from (7) and (5) and ignoring the negative contribution made by the sequences in the set $A_2 \backslash A_3$. (e) follows from Lemma 1. $C' > 0$ (constant) and $u(\epsilon)$ (positive valued function $\forall \ \epsilon > 0$ and 0 if $\epsilon = 0$) are obtained after simplification of (e). Now, using extended mean value theorem for the function e^{-z},

$$e^{-E_2 2^{-\frac{n\epsilon}{2}}} = 1 - E_2 2^{-\frac{n\epsilon}{2}} + \frac{e^{-c}}{2}E_2^2 2^{-n\epsilon}.$$

Here, $c \ \epsilon \ (0, E_2 2^{-\frac{n\epsilon}{2}})$. Therefore, we have

$$e^{-E_2 2^{-\frac{n\epsilon}{2}}} \geq 1 - E_2 2^{-\frac{n\epsilon}{2}}. \tag{16}$$

Hence, using (16) in (15), we get

$$P(\frac{\log R_n(X)}{n} < H - \epsilon) < e^{-g(\epsilon)n}. \tag{17}$$

where $g(\epsilon)$ is a positive valued function $\forall \ \epsilon > 0$ and $g(0) = 0$. This completes the proof of Theorem 4.

Proof of Corollary 3: Using Observation 1, we have

$$L_{2^{n(H-\epsilon)}}(x) > n \Rightarrow R_n(x) \leq 2^{n(H-\epsilon)}$$
$$\Rightarrow P(L_{2^{n(H-\epsilon)}}(X) > n) \leq P(R_n(X) \leq 2^{n(H-\epsilon)}) < e^{-g(\epsilon)n}.$$

$\forall \ n \geq N'(\epsilon)$. Now, letting $m = 2^{n(H-\epsilon)}$, we have

$$P(L_m(X) > \frac{\log m}{H - \epsilon}) < e^{-g(\epsilon)\frac{\log m}{H - \epsilon}} \quad \forall \ m \geq M'(\epsilon)$$
$$\Rightarrow P(\frac{\log m}{L_m(X)} < H - \epsilon) < e^{-g(\epsilon)\frac{\log m}{H - \epsilon}}.$$

Remark 4: Note that in the first step in Eq. (14) we have a term $1 - P(\frac{\log R_n(X)}{n} > H - \epsilon) = P(R_n(X) \leq 2^{n(H-\epsilon)})$. Further, in the proof of Theorem 4, the bound $e^{-g(\epsilon)n}$ is obtained on this term. Hence, there is no ambiguity in using the exponential bound obtained in Theorem 4 on $P(R_n(X) \leq 2^{n(H-\epsilon)})$.

V. ESTIMATOR FOR ENTROPY

Motivated by experimental results on estimators based on match lengths given in [13], we propose an estimator based on *recurrence times* as given below:

Estimator: Consider $R_{n,i}(X) = R_n(T^i X)$.

Define: $J_n(X) = \frac{1}{Q(n)}\sum_{i=1}^{Q(n)} \frac{\log R_{n,i}(X)}{n}$.

Proposition 1: If $Q(n)$ is of the polynomial order, then for processes which are exponentially ϕ-mixing and have exponential rates for entropy, $\lim_{n \to \infty} J_n(X) = H$ a.s. with $J_n(X)$ satisfying large deviation property. The proof of the proposition is given below:

$$P(J_n(X) > H + \epsilon) = P(\frac{1}{Q(n)}\sum_{i=1}^{Q(n)} \frac{\log R_{n,i}(X)}{n} > H + \epsilon)$$
$$\leq \sum_{i=1}^{Q(n)} P(\frac{\log R_{n,i}(X)}{n} > H + \epsilon) \ (a)$$
$$< \sum_{i=1}^{Q(n)} e^{-f(\epsilon)n} \quad \forall \ n \geq N(\epsilon) \ (b)$$
$$= Q(n)e^{-f(\epsilon)n}.$$
$$(18)$$

Here, step (a) follows from Remark 8 given in appendix and step (b) follows from the stationarity of the source X and Theorem 3. Similarly,

$$P(J_n(X) < H - \epsilon) = P(\frac{1}{Q(n)}\sum_{i=1}^{Q(n)} \frac{\log R_{n,i}(X)}{n} < H - \epsilon)$$
$$\leq \sum_{i=1}^{Q(n)} P(\frac{\log R_{n,i}(X)}{n} < H - \epsilon) \ (a)$$
$$< \sum_{i=1}^{Q(n)} e^{-g(\epsilon)n} \quad \forall \ n \geq N'(\epsilon) \ (b)$$
$$= Q(n)e^{-g(\epsilon)n}.$$
$$(19)$$

Here, step (a) follows from Remark 8 and step (b) follows from the stationarity of the source X and Theorem 4. Therefore, combining (18) and (19), we have

$$P(|J_n(X) - H| > \epsilon) < 2Q(n)e^{-I(\epsilon)n} \ \forall \ n \geq N''(\epsilon) \tag{20}$$

where $N''(\epsilon) = \max\{N(\epsilon), N'(\epsilon)\}$.

For $Q(n)$ of polynomial order, we have

$$\sum_{n=1}^{\infty} P(|J_n(X) - H| > \epsilon) < \sum_{n=1}^{N''(\epsilon)-1} P(|J_n(X) - H| > \epsilon)$$
$$+ \sum_{n=N''(\epsilon)}^{\infty} 2Q(n)e^{-I(\epsilon)n}$$
$$< N''(\epsilon) + \sum_{n=N''(\epsilon)}^{\infty} 2Q(n)e^{-I(\epsilon)n}$$
$$< \infty.$$
$$(21)$$

Hence, by Borel-Cantelli Lemma

$$\lim_{n \to \infty} J_n(X) = H \qquad a.s. \qquad (22)$$

Remark 5: The bounds we establish on convergence rates are loose, we *conjecture* that our proposed estimator will converge to entropy rate at a faster rate than $2e^{-I(\epsilon)n}$.

VI. CONCLUSION

In this paper, we have proved the Large deviation property for the normalized version of recurrence times for exponentially ϕ-mixing processes. Further, we have also shown this property to hold for our proposed estimator of entropy based on recurrence times. As a future work, it will be interesting to answer if there are faster rate functions than $f(\epsilon)$ and $g(\epsilon)$ in this context, and further on what more classes of processes large deviation property holds for normalized version of recurrence times. Also, one can conduct experimental or theoretical studies comparing the convergence rates of the estimator based on match length given in [13] and that based on recurrence times proposed in this paper.

APPENDIX

Remark 6: Note that in step (a), inequality (4) has been used, however it is important to check if it can be applied. This is verified below:

$$\rho_y = \frac{2\sqrt{C_y}}{(\sqrt{1+C_y} + \sqrt{C_y})\xi_y P(y)} \quad \forall \ y \ \epsilon \ A_n^{(\delta)}$$

Using lower bounds on $P(y)$ and ξ_y, we have

$$\rho_y \leq \frac{2\sqrt{C_y}\ 2^{n(H+\delta)}}{(\sqrt{1+C_y} + \sqrt{C_y})E_1} \quad \forall \ y \ \epsilon \ A_n^{(\delta)}.$$

Since $C_y \to 0 \ (as \ n \to \infty)$, for a given $d' > 0$, $\frac{2\sqrt{C_y}}{\sqrt{1+C_y}+\sqrt{C_y}} < d'$ for n large enough. Now, we choose d' such that $0 < d' < E_1$. Since eventually δ is chosen to be less than ϵ, we have

$$\rho_y < 2^{n(H+\delta)} < 2^{n(H+\epsilon)} \ \forall \ y \ \epsilon \ A_n^{(\delta)}.$$

Remark 7: Note that, though we prove Theorem 3 under the restriction of certain mixing conditions and using inequality (4), it can also be proved using Markov Inequality and Kac's Lemma under no restriction of mixing. However, the super exponential behavior shown by first term in the proof of Theorem 3 (see Inequality (12)) is not evident from this alternative proof for mixing sources considered. Due to space limitations, we have omitted this proof.

Remark 8: Let $Z_1, Z_2, ..., Z_m$ be m real valued random variables. Consider the following probability, $P(\frac{1}{m}\sum_{i=1}^{m} Z_i > r)$

and set $E_i = \{\omega : Z_i(\omega) > r\}$. Now,

$$P(\frac{1}{m}\sum_{i=1}^{m} Z_i > r) \leq P(\cup_{i=1}^{m} E_i)$$

$$\leq \sum_{i=1}^{m} P(E_i) \quad (Union \ Bound)$$

$$= \sum_{i=1}^{m} P(Z_i > r).$$

Similarly, by changing '>' sign with '<' accordingly, it can be proved that

$$P(\frac{1}{m}\sum_{i=1}^{m} Z_i < r) \leq \sum_{i=1}^{m} P(Z_i < r).$$

Proof of Lemma 1:

$$P(A_1 \cup A_2) \leq 1 \Rightarrow P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq 1$$
$$\Rightarrow P(A_1 \cap A_2) \geq P(A_1) + P(A_2) - 1$$
$$\Rightarrow P(A_1 \cap A_2) > 1 - p_1 e^{-p_2 n} + 1 - q_1 e^{-q_2 n} - 1$$
$$\Rightarrow P(A_1 \cap A_2) > 1 - (p_1 e^{-p_2 n} + q_1 e^{-q_2 n})$$
$$\Rightarrow P(A_1 \cap A_2) > 1 - (p_1 + q_1)e^{-\min\{p_2, q_2\}n}.$$

REFERENCES

[1] M. Abadi, "Exponential Approximation for hitting times in mixing processes," Math. Phys. Electron. J. 7, Paper No. 2, 19 p., 2001.
[2] M. Abadi, " Sharp error terms and necessary conditions for exponential hitting times in mixing processes," Ann. Probabality, vol. 32 (1A), pp. 243-264, 2004.
[3] M. Abadi and A. Galves, "A version of Maurer's conjecture for stationary ψ-mixing processes," Nonlinearity, vol. 17, pp. 1357-1366, 2004.
[4] M. Abadi and N. Vergne, "Sharp Error Terms for Return Time statistics under mixing conditions," Journal of Theoretical Probability, vol. 22, no. 1, pp. 18-37, 2009.
[5] R. Bradley, "Basic Properties of Strong Mixing Conditions " in Dependence in Probability and Statistics edited by E. Eberlein and M. S. Taqqu, 1986.
[6] J.-R. Chazottes and E. Ugalde, "Approximation and Fluctuations of Hitting and Recurrence times for Gibbsian measures, " Discrete and Continuous Dynamical Systems B, vol. 5, no. 3, pp. 565-586, 2005.
[7] P. Collet, A. Galves and B. Schimitt, "Repetition times for Gibbsian sources," Nonlinearity, vol. 12, pp. 1225-1237, 1999.
[8] A. Dembo and O. Zeitouni, "Large Deviations Techniques and Applications," Springer-Verlag New Nork Inc. 1998.
[9] A. Galves and B. Schimitt, "Inequalites for hitting times in mixing dynamical systems," Random Comput. Dynam., vol. 5, pp. 337-347, 1997.
[10] M. Hirata, B. Saussol and S. Vaienti, "Statistics of return times: a general framework and new applications," Comm. Math. Phys., vol. 206, pp. 33-35, 1999.
[11] D. H. Kim, "Convergence of Recurrence of Blocks for Mixing Processes," Osaka Journal of Mathematics, Vol. 49 (1), pp. 1-20, 2012.
[12] I. Kontoyiannis, "Asymptotic Recurrence and Waiting Times for Stationary Processes," Journal of Theoretical Probability, vol. 11 (3), pp. 795-811, July 1998.
[13] I. Kontoyiannis, P. H. Algoet, Yu. M. Suhov and A. J. Wyner, "Nonparametric Entropy Estimation for Stationary Processes and Random Fields, with Applications to English Text," IEEE Trans. Inform. Theory, vol. IT-44 (3), pp. 1319-1327, May 1998.
[14] U. M. Maurer, "A universal statistical test for random bit generators," J. Cryptology, Vol. 5, pp. 89-105, 1992.
[15] D. Ornstein and B.Weiss, "Entropy and data compression schemes," IEEE Trans. Inform. Theory, vol. IT-39 (1), pp. 78-83, Jan. 1993.
[16] P. C. Shields, "The ergodic theory of discrete sample paths," AMS Graduate Studies in Mathematics, Amer. Math. Soc., 1996.