

LIMIT THEOREMS FOR CHAINS WITH UNBOUNDED VARIABLE LENGTH MEMORY WHICH SATISFY CRAMER CONDITION*

A. LOGACHOV^{1,2,**}, A. MOGULSKII¹ AND A. YAMBARTSEV³

Abstract. We consider a class of variable length Markov chains with a binary alphabet in which context tree is defined by adding finite trees with uniformly bounded height to the vertices of an infinite comb tree. Such type of Markov chain models the spike neuron patterns and also extends the class of persistent random walks. The main interest is the limiting properties of the empirical distribution of symbols from the alphabet. We obtain the strong law of large numbers, central limit theorem, and exact asymptotics for large and moderate deviations. The presence of an intrinsic renewal structure is the subject of discussion in the literature. Proofs are based on the construction of a renewals of the chain and the applying corresponding properties of the compound (or generalized) renewal processes.

Mathematics Subject Classification. 60F10, 60K05, 60K30.

Received February 3, 2019. Accepted February 5, 2022.

1. INTRODUCTION

Let $\mathcal{A} = \{0, 1, \dots, d\}$ be a set of $d + 1$ symbols (characters) – an alphabet. Here we consider a class of chains $\mathbf{r} = (r_i, i \in \mathbb{Z}) \in \mathcal{A}^{\mathbb{Z}}$ which is a special case of so-called *chains with unbounded variable length memory* or *variable length Markov chain* (VLMC). These chains began to be actively studied after they were first introduced by Jorma Rissanen [22] as an economic and universal way of data compression. The short and simple introduction to these processes can be found, for example, in [14], see also overview [7]. They are used for modeling data in computer science [22], in biology [1, 18], in neuro-biology [12, 13], in linguistics [15]. We do not pretend for at least some sort of full review on these chains, and we restrict ourselves to some papers known to us by the activity of the research group *NeuroMat* (IME-USP) under the guidance of Prof. A.Galves.

*This work was supported by the Basic Research Program of the Siberian Branch of the Russian Academy of Sciences, project No. FWNF-2022-0010. It is part of USP project Mathematics, computation, language and the brain, FAPESP project Research, Innovation and Dissemination Center for Neuromathematics grant 2013/07699-0. AL and AY thank FAPESP grant 2017/20482-0, AY also thanks CNPq and FAPESP grants 301050/2016-3 and 2017/10555-0, respectively.

Keywords and phrases: Variable length memory chain, regeneration scheme, compound renewal process, local limit theorem, large deviation principle, moderate deviation principle, rate function, Cramer condition.

¹ Laboratory of Probability Theory and Mathematical Statistics, Sobolev Institute of Mathematics, Siberian Branch of the Russian Academy of Sciences, Koptuga, 4, Novosibirsk 630090, Russian Federation.

² Department of High Mathematics, Siberian State University of Geosystems and Technologies, Plahotnogo str. 10, Novosibirsk 630108, Russian Federation.

³ Department of Statistics, Institute of Mathematics and Statistics (IME-USP), University of São Paulo, Rua do Matão 1010, CEP 05508-090, São Paulo SP, Brazil.

** Corresponding author: omboldovskaya@mail.ru

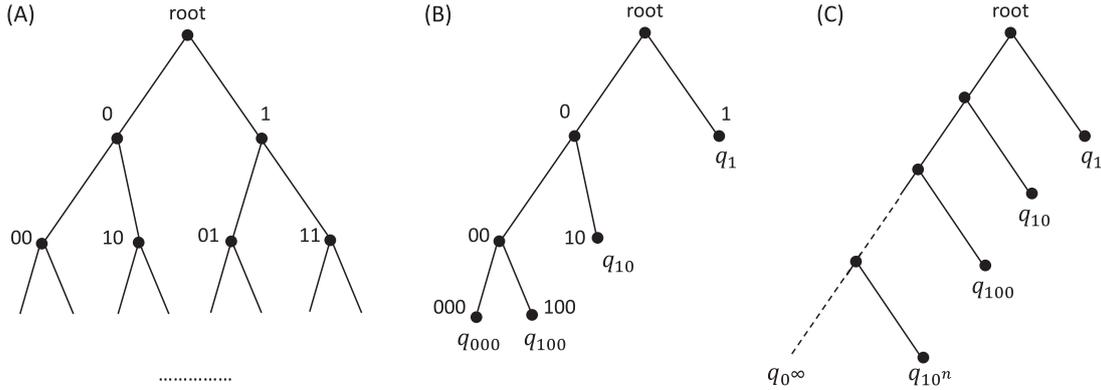


FIGURE 1. (A) the complete binary rooted tree, where non-rooted vertices codified by the finite binary sequences; (B) the example of a finite binary context tree; (C) the example of infinite binary comb context tree; the notation “ 0^n ” means the string of n repeating symbol “0”, and “ 0^∞ ” means the semi-infinite string of “0”s.

If we interpret \mathbb{Z} as a discrete time, then one can imagine such chains as a successive in time attribution of a character from the alphabet \mathcal{A} with probability which depends on the past (existing sequence of characters), or, more precisely, it depends on a part of the past which called a *context*. More formally, let us fix an initial configuration

$$r(0) = \{r_i(0)\}_{-\infty < i \leq 0}, \quad r_i(0) \in \mathcal{A},$$

where $r_i(0)$, $-\infty < i \leq 0$ take values from the finite alphabet \mathcal{A} , which is binary $\mathcal{A} := \{0, 1\}$ in our case. Let us denote $\mathcal{A}^{-\infty}$ the set of all such semi-infinite sequences. Thus, adding a character from the right, the state of the process on the n -th step is a configuration

$$r(n) = \{r_i(n)\}_{-\infty < i \leq n}, \quad r_i(n) \in \mathcal{A},$$

such that $\{r_i(n)\}_{-\infty < i \leq n-1} \equiv \{r_i(n-1)\}_{-\infty < i \leq n-1}$, for any n . Observe that the set

$$\mathcal{A}^* = \bigcup_{n=1}^{\infty} \mathcal{A}^n$$

of all finite sequences of symbols from \mathcal{A} can be represented as the non-root vertices of an infinite $(d + 1)$ -ary complete rooted tree, see Figure 1(A), for $d = 1, |\mathcal{A}| = 2$. Note that we read the sequence from the right to the left.

Given two integers $s \leq k$, we will denote by r_s^k the sequence (r_s, \dots, r_k) of symbols in \mathcal{A} . One call the process $r(n)$ the VLMC if for any semi-infinite configuration $\alpha = \{\alpha_i\}_{-\infty < i \leq 0} \in \mathcal{A}^{-\infty}$ such that $\{r_{i-n}(n)\}_{-\infty < i \leq n} = \alpha$ there exists $m(\alpha) \leq \infty$ (context memory function) such that

$$\mathbf{P}(r_{n+1}(n+1) = a \mid r(n) = \alpha) = \mathbf{P}(r_{n+1}(n+1) = a \mid r(n)_{n-m(\alpha)}^n = \alpha_{-m(\alpha)}^{-1}),$$

i.e. the probability to add the next symbol depends on the part (symbols from the site $n - m(\alpha)$ to the site n) of the hole history $r(n)$ (the theory allows to be the hole semi-infinite sequence $r(n)$). Note that in general $r(n)$ is not Markov chain. The role of matrix of transition probabilities for the Markov chain plays the context

tree in case of VLMC, where each vertex represent a context, and each leaf is associated with a probability distribution (on \mathcal{A}) of a new character.

To make the notion of VLMC more clear, let consider the finite context tree represented on Figure 1(B) with binary alphabet $\mathcal{A} = \{0, 1\}$. Here, suppose at time n we have a sequence $r(n) = \{r_i(n)\}_{-\infty < i \leq n}$. Then, the probability to add 1 on the site $n + 1$ is

$$\mathbf{P}(r_{n+1}(n+1) = 1 \mid r(n)) = \begin{cases} q_1, & \text{if } r_n(n) = 1; \\ q_{01}, & \text{if } r_n(n) = 0, r_{n-1}(n) = 1; \\ q_{001}, & \text{if } r_n(n) = 0, r_{n-1}(n) = 0, r_{n-2}(n) = 1; \\ q_{000}, & \text{if } r_n(n) = 0, r_{n-1}(n) = 0, r_{n-2}(n) = 0. \end{cases}$$

Note that a Markov chain with the state space \mathcal{A} is a particular case of these chains. In this case the context tree has height 1 – we should know only the last character in order to decide the distribution of a new character. The existence of a stationary measure on $\mathcal{A}^{\mathbb{Z}}$ compatible with a family of transition probabilities determined by a context tree is the question which naturally arises. This question was answered and the short review can be find in [14], where the methods of statistical inferences of context tree were also studied. See also [7], where the necessary and sufficient conditions for the existence and uniqueness of stationary distribution were established for some special the *infinite comb* (see Fig. 1(C)) and the *bamboo blossom* context trees.

In [16] the perfect simulation for such processes was constructed. The success of the algorithm (if the perfect simulation stops in a finite time) depends directly from the existence of a (finite) renewal time moment – the moment when the next and successive attributions of characters do not depend on the past. In the same paper the connection between renewal processes and the chains with variable length memory was established. This indicates to us that the large deviations results for this sort of chain can be obtained using the regeneration structure and the recently published results on large deviations for the (compound) renewal processes [21].

Let us now specify the process $r(n)$ we deal with in the paper. The process is VLMC with binary alphabet. Instead of providing the context tree, let us first define it in an alternative way.

In order to set the transition rules we fix a number $v \in \mathbb{N}$ (one of the parameters of the chain). Consider the set of all strings of v characters (from \mathcal{A}^v) ending with the character 1; the total number of such words is 2^{v-1} ; for any such word we assign their own order number j , $1 \leq j \leq 2^{v-1}$. Let us set the sequence of positive numbers on this set of words

$$\mathbf{p}_{kj} \in (0, 1), \quad k \in 0 \cup \mathbb{N}, \quad 1 \leq j \leq 2^{v-1}. \quad (1.1)$$

Now we are ready to describe the rules (transition probabilities) of adding a character from the right: let we have a configuration on the n -th step $r(n) = \{r_i(n)\}_{-\infty < i \leq n}$. Denote

$$m_n := \sup\{-\infty < i \leq n : r_i(n) = 1\}.$$

Then, at the next step the configuration $r(n) = \{r_i(n)\}_{-\infty < i \leq n}$ jump to the configuration

$$r(n+1) = \{r_i(n+1)\}_{-\infty < i \leq n+1}$$

by writing from the right a character 1, $r_{n+1}(n+1) = 1$, with probability $\mathbf{p}_{k_n j_n}$, where $k_n = n - m_n$, and j_n is the number of word, which forms the sequence

$$r_{m_n-v+1}(n), r_{m_n-v+2}(n), \dots, r_{m_n}(n).$$

Thus, the character 0, $r_{n+1}(n+1) = 0$, will be adding with probability $1 - \mathbf{p}_{k_n j_n}$. Note that the previous sequence do not change, that is $\{r_i(n+1)\}_{-\infty < i \leq n} \equiv r(n)$. Thus, the probability of the attributed character $r_{n+1}(n+1)$ depends on

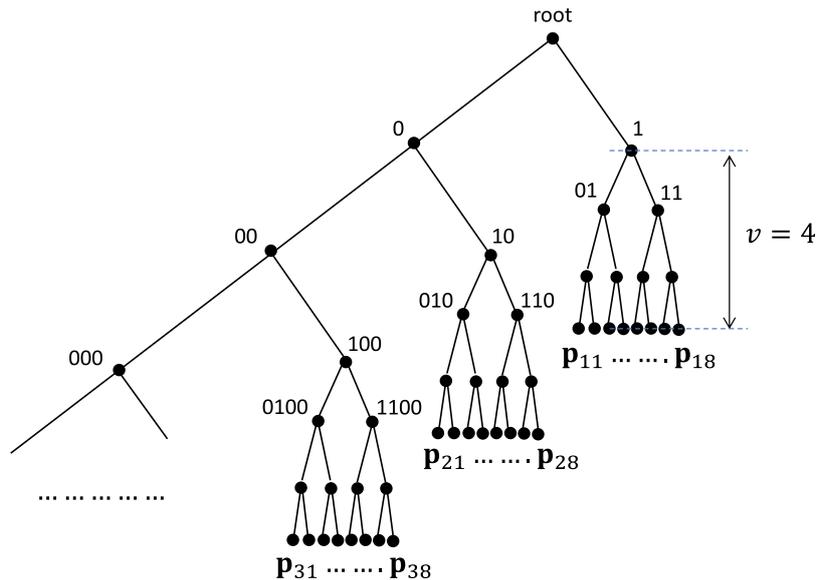


FIGURE 2. The context tree corresponding the studying process $r(n)$.

- (i) the distance to the nearest character 1;
- (ii) the word of $v - 1$ letters, which stand on the left from this 1.

The context tree corresponding to the defined process $r(n)$ for the case $v = 4$ is represented in Figure 2 which is extension of comb tree of Figure 1(C). The comb tree on Figure 1(C) correspond to the case $v = 1$ with probabilities $\mathbf{p}_{k1} = q_{10^{k-1}}$ for any $k \in \mathbb{N}$.

It is obvious that the random sequence $\{r_n(n)\}$ is not Markov chain, because the transition probability from $r_n(n)$ to $r_{n+1}(n + 1)$ can depend, generally speaking, not only on the character $r_n(n)$, but also on values $r_{n-j}(n - j)$ for arbitrarily large $j \geq 0$.

There are some motivations why we are restricted to that specific class of chains with unbounded variable length memory. First of all, in studies of neuronal spiking signals, the time of the last spiking of a neuron is, in some sense, a renewal time, and the distribution of time interval between two consecutive spikes indicate the neuronal activity of neighbors neurons. Thus, when we add information about the short (of size v) previous activity pattern of a neuron we try to extend the information about the neuronal dynamics and the limiting behavior is our interest in this context.

We would note also that the processes of such type correspond to the one-side comb-like context trees. Such a structure is very popular in various fields of research. But here we would like to mention the class of processes known in the literature as a *Persistent random walk* (PRW), a *Goldstein-Kac random walk*, or also a *correlated random walk*, see [5] for definitions and short historical introduction. The increments of these random walks are not independent but form a Markov chain and are closely related to the chains with unbounded variable length memory, see [5, 6]. These random walks can be considered as a chain with unbounded variable length memory with the special context tree on the binary alphabet called by the authors as a double (infinite) comb context tree [5].

Moreover, in [5] the authors also consider persistent random walk corresponding to the adding to the leaves of comb tree the context trees with success probabilities associated with the leaves of added trees. It means that increments of a random walk can be derived by an infinite binary context tree, which is a natural extension of studied persistent random walks. The authors noted that the renewal properties may still hold in this situation. In our case, the process corresponds to the infinite comb tree where each leaf receives a complete binary tree of height $v - 1$ with arbitrary success probabilities (1.1).

For such defined process $r(n)$ let $R(n)$ be the number of units adding from the right to the initial configuration $r(0)$ in n steps:

$$R(n) := \sum_{k=1}^n r_k(n) \equiv \sum_{k=1}^n r_k(k).$$

In terms of persistent (one dimensional) random walk on \mathbb{Z} with increments ± 1 , the position of random walk at time n is $2R(n) - n$. We are interested in the asymptotic behaviour of the process $R(n)$ when $n \rightarrow \infty$. In the next section we prove law of large numbers, central limit theorem, local limit theorem, we establish also the large and moderate deviations principles.

Further we suppose that the following condition **[A]** holds true. The condition **[A]** composed by two items.

[A.1] *Initial configuration $r(0)$ contains at least one unit.*

[A.2] *There exist constants $0 < \delta_2 < \delta_1 < 1$ such that for all $k \in 0 \cup \mathbb{N}$, $1 \leq j \leq 2^{v-1}$ the following inequalities holds*

$$\delta_2 \leq \mathbf{p}_{kj} \leq \delta_1.$$

The condition **[A.1]** is an obvious condition for existence of process and for implementation of transition probabilities. Note, however, that this condition can be omitted adding the probability $\mathbf{p}_\infty \in (0, 1)$ to attribute a character “1”, when the sequence consists of only zeros. Condition **[A.2]** gives us the possibility of construction of arithmetic compound (generalized) renewal process which satisfy Cramer moment condition **[C₀]** and the condition of arithmeticity of **[Z]** (see Sect. 2).

Note that the transition probabilities of VLMC can be considered as a special case of so-called g -function defining the chains of infinite order. A g -function is a probability kernel $g : \mathcal{A} \times \mathcal{A}^{\mathbb{N}} \rightarrow [0, 1]$ such that $\sum_{a \in \mathcal{A}} g(a | \alpha) = 1$ for any $\alpha \in \mathcal{A}^{\mathbb{N}}$: in our notation

$$\mathbf{P}(r_{n+1}(n+1) = a | r(n) = \alpha) = g(a | \alpha).$$

But we should note that all (to the best of our knowledge) results in the literature on chains of infinite order deals with the condition of continuity of a g -function: $g(a | \cdot)$ should be continuous for each $a \in \mathcal{A}$, *i.e.* for all $\varepsilon > 0$, there exists $n \geq 0$ so that

$$|g(a | \alpha) - g(a | \beta)| < \varepsilon$$

for all $\alpha = (a_i, i \in \mathbb{N}), \beta = (b_i, i \in \mathbb{N}) \in \mathcal{A}^{\mathbb{N}}$ with $a_i = b_i$ for each $i = 1, \dots, n$. This condition is not assumed here, although the continuity condition can be attained in some special cases of VLMC. Indeed, let us here recall the example considered in [16]. In our notations we consider the case $v = 1$, *i.e.* the corresponding context tree is the infinite binary comb context tree represented by Figure 1(C), and $\mathbf{p}_{11} = q_1, \mathbf{p}_{21} = q_{10}, \dots, \mathbf{p}_{n1} = q_{10^{n-1}}$ and $\mathbf{p}_\infty = q_{0^\infty}$. Let fix $\varepsilon_1 \neq \varepsilon_2 \neq \varepsilon_3 \in (0, 1)$, consider $q_{10^n} = \varepsilon_1$ if n is odd, $q_{10^n} = \varepsilon_2$ if n is even, and $q_{0^\infty} = \varepsilon_3$. It is straightforward to check that for any n

$$\sup_{\alpha, \beta: a_i = b_i, i=1, \dots, n} |g(a | \alpha) - g(a | \beta)| = \max\{|\varepsilon_1 - \varepsilon_2|, |\varepsilon_1 - \varepsilon_3|, |\varepsilon_2 - \varepsilon_3|\},$$

It follows that non of our processes have to be continuous. In particular, it means that the central limit theorem proved in [20] or the rough large deviation asymptotics from [8, 9] cannot be applied for considered here process $R(n)$. Moreover, under the continuity condition the limiting properties of $R(n)$ can be proved using the renewal structure constructed in [10]. In general, to study the limiting behavior of the sequence $R(n)$ requires a renewal structure different from that constructed in [10], see Section 2.

The paper is organized as follows. In Section 2 we introduce our definitions and notations and we formulate main results by Theorem 2.2: law of large numbers, central limit theorem, local moderate and large limit theorem and principle of moderate large deviations for $R(n)$. In Section 3 we prove Theorem 2.2. Section 4 contains the auxiliary lemmas and their proofs.

2. DEFINITIONS, NOTATIONS AND MAIN RESULTS

To formulate and prove the main result we need to recover the renewal structure of the studying process. The renewal structure is represented as *generalized renewal process* [2] known in the literature also as *renewal reward process*. Naturally, all limit theorems we prove here for $R(n)$ are related to limiting properties of the generalized renewal process. The results from [2] will be useful in order to prove the strong law of large numbers and central limit theorem, and for the large deviation theorems we apply the results from [21].

For any state $r(n)$ we correspond the pair

$$Y(n) := (Y_1(n), \mathbf{Y}_2(n)),$$

where $Y_1(n) := n - m_n$ (the distance to the nearest unit, or, what's the same, the number of zeros before the nearest unit),

$$\mathbf{Y}_2(n) := (r_{m_n-v+1}(n), r_{m_n-v+2}(n), \dots, r_{m_n}(n))$$

(sequence of the nearest unit and $v - 1$ letters from its left).

Note, that the pair $Y(n) := (Y_1(n), \mathbf{Y}_2(n))$ can transit with probability $\mathbf{p}_{k_n j_n}$ into the pair

$$Y(n+1) = (0, \mathbf{Y}_2(n+1)),$$

where

$$\mathbf{Y}_2(n+1) = \begin{cases} (0, \dots, 0, 1), & \text{if } Y_1(n) \geq v - 1, \\ (r_{m_n-v+Y_1(n)+2}(n), \dots, r_{m_n}(n), 0, \dots, 0, 1), & \text{if } Y_1(n) < v - 1, \end{cases}$$

and with probability $1 - \mathbf{p}_{k_n j_n}$ into the pair

$$Y(n+1) = (Y_1(n) + 1, \mathbf{Y}_2(n)).$$

In this way $Y(n+1)$ is a random function on $Y(n)$. Thus, the sequence $\{Y(n)\}$, $n \geq 0$ is homogeneous Markov chain with the phase state

$$\mathcal{Y} := \{y = (y_1, \mathbf{y}_2) : y_1 \in 0 \cup \mathbb{N}, \mathbf{y}_2 = (a_1, \dots, a_v), a_1 \in \mathcal{A}, \dots, a_{v-1} \in \mathcal{A}, a_v = 1\}.$$

Let us pick out the state

$$y_0 := (0, (0, \dots, 0, 1)).$$

Note that the chain can jump in one step from any state (y_1, \mathbf{y}_2) to the chosen state y_0 , if the coordinate y_1 not less than $v - 1$. Denote

$$\begin{aligned} \tau_1 &:= \min\{n > 0 : Y(n) = y_0\}, \\ \tau_k &:= \min\{n > \tau_1 + \dots + \tau_{k-1} : Y(n) = y_0\} - (\tau_1 + \dots + \tau_{k-1}), \quad 2 \leq k. \end{aligned}$$

Since $\{Y(n)\}$ is homogeneous Markov chain, the random variables $\tau_1, \dots, \tau_k, \dots$ are independent and, moreover, τ_k are identically distributed when $k \geq 2$.

Let ζ_k be the number of units added from the right during the time interval $\{T_{k-1} + 1, \dots, T_k\}$, where $T_0 := 0, T_k := T_{k-1} + \tau_k$ for $k \in \mathbb{N}$. In other words

$$\zeta_0 := 0, \quad \zeta_k := \sum_{n=T_{k-1}+1}^{T_k} r_n(n) \equiv R(T_k) - R(T_{k-1}) \quad \text{for } k \in \mathbb{N}.$$

By construction the random vectors $\xi_k := (\tau_k, \zeta_k)$, $k \in \mathbb{N}$ are independent, and ξ_k are identically distributed when $k \geq 2$. Let

$$\nu(0) := 0, \quad \nu(n) := \max\{k \geq 0 : T_k < n\} \quad \text{for } k \in \mathbb{N}.$$

The following process

$$Z(n) := \sum_{k=0}^{\nu(n)} \zeta_k.$$

is *generalized renewal process*, see, for example, [2] for definition and some properties. Naturally, we will highlight and use some specific properties which come from the definitions of $r(n)$ and $R(n)$.

Let the random vector $\xi := (\tau, \zeta)$ has the same distribution as vectors $\xi_k = (\tau_k, \zeta_k)$ for $k \geq 2$. Since $\zeta_k \leq \tau_k$ a.s. when $k \in \mathbb{N}$, then from Lemma 4.1 (see Sect. 4) it follows that for ξ_1 and ξ the Cramer's condition $[\mathbf{C}_0]$ holds true:

$$\mathbf{E}e^{\delta|\xi_1|} \leq \mathbf{E}e^{2\delta\tau_1} < \infty \quad \text{and} \quad \mathbf{E}e^{\delta|\xi|} \leq \mathbf{E}e^{2\delta\tau} < \infty, \quad \text{when } \delta < \rho/2,$$

where $\rho > 0$ is the constant from Lemma 4.1.

From Lemma 4.2 (see Sect. 4) we obtain that the vector ξ satisfies the arithmeticity condition $[\mathbf{Z}]$:

$[\mathbf{Z}]$ For any $u \in \mathbb{Z}^2$ the equality $f(2\pi u) = 1$ holds and for any $u \in \mathbb{R}^2 \setminus \mathbb{Z}^2$ the inequality $|f(2\pi u)| < 1$ holds, where for $u = (u_1, u_2) \in \mathbb{R}^2$ the function $f(u)$ is the characteristic function for ξ , i.e. $f(u) := \mathbf{E}e^{i(u_1\tau + u_2\zeta)}$.

Give the notation we need from the paper [21]

$$a := \frac{\mathbf{E}\zeta}{\mathbf{E}\tau}, \quad \sigma^2 := \frac{1}{a\tau} \mathbf{E}(\zeta - a\tau)^2, \quad \text{where } a_\tau := \mathbf{E}\tau.$$

The Laplace transform we need are defined as

$$\begin{aligned} \psi_1(\lambda, \mu) &:= \mathbf{E}e^{\lambda\tau_1 + \mu\zeta_1}, & \psi(\lambda, \mu) &:= \mathbf{E}e^{\lambda\tau + \mu\zeta}, \\ A(\lambda, \mu) &:= \ln \psi(\lambda, \mu), & (\lambda, \mu) &\in \mathbb{R}^2. \end{aligned}$$

The Legendre transform is

$$\begin{aligned} D(\theta, \alpha) &:= \sup_{(\lambda, \mu) \in \mathcal{A}^{\leq 0}} \{\lambda\theta + \mu\alpha\}, \quad \text{where } \mathcal{A}^{\leq 0} := \{(\lambda, \mu) : A(\lambda, \mu) \leq 0\}, \\ D(\alpha) &:= D(1, \alpha). \end{aligned}$$

Denote \mathfrak{B} the Borel σ -algebra of subsets of \mathbb{R} . For an arbitrary set $B \in \mathfrak{B}$ we denote $[B]$ and (B) its closure and interior correspondingly.

Now we give the definition of the Large Deviation Principle (LDP).

Definition 2.1. The sequence of random variables s_n satisfies LDP in \mathbb{R} with rate function $I = I(y) : \mathbb{R} \rightarrow [0, \infty]$ and normalized function $\varphi(n) : \lim_{n \rightarrow \infty} \varphi(n) = \infty$, if for any $c \geq 0$ the set $\{y \in \mathbb{R} : I(y) \leq c\}$ is compact and the following inequalities holds true: for any closed set $B \in \mathfrak{B}$

$$\limsup_{n \rightarrow \infty} \frac{1}{\varphi(n)} \ln \mathbf{P}(s_n \in B) \leq -I(B),$$

and for any open set $B \in \mathfrak{B}$

$$\liminf_{n \rightarrow \infty} \frac{1}{\varphi(n)} \ln \mathbf{P}(s_n \in B) \geq -I(B),$$

where $I(B) = \inf_{y \in B} I(y)$, $I(\emptyset) = \infty$.

Denote Φ_{0, σ^2} the normal distribution with parameters $(0, \sigma^2)$, and by \Rightarrow we denote the convergence in distribution.

Let us formulate now the main result of our work.

Theorem 2.2. *Let the condition [A] holds true. Then*

1. *(strong law of large numbers) When $n \rightarrow \infty$*

$$\frac{R(n)}{n} \rightarrow a \text{ a.s.}$$

2. *(central limit theorem) When $n \rightarrow \infty$*

$$\frac{R(n) - an}{\sqrt{n}} \Rightarrow \Phi_{0, \sigma^2}.$$

3. *(local theorem in regions of normal, moderate and large deviations) There exists $\Delta > 0$ such that if $x \in 0 \cup \mathbb{N}$, $\lim_{n \rightarrow \infty} \frac{x}{n} = \alpha_0$ and $|\alpha_0 - a| \leq \Delta$, then*

$$\mathbf{P}(R(n) = x) = \frac{1}{\sqrt{n}} \psi_1(\lambda(\alpha_0), \mu(\alpha_0)) C_H(1, \alpha_0) I(\alpha_0) e^{-nD(\frac{x}{n})} (1 + o(1)),$$

where

$$I(\alpha_0) = \sum_{l=1}^{\infty} e^{\lambda(\alpha_0)l} \mathbf{E}(e^{\mu(\alpha_0)R(l)}, \tau \geq l \mid Y(0) = y_0),$$

and $C_H(\theta, \alpha)$ is the positive function, which is continuous in a neighborhood of the point $(\theta, \alpha) = (1, \alpha_0)$ and it is known explicitly from Theorem 2.1 and 2.1A [21].

4. *(local theorem in regions of normal and moderate deviations) If $x \in 0 \cup \mathbb{N}$ and $\lim_{n \rightarrow \infty} \frac{x}{n} = a$, then the following equality holds true*

$$\mathbf{P}(R(n) = x) = \frac{1}{\sigma \sqrt{2\pi n}} e^{-nD(\frac{x}{n})} (1 + o(1)).$$

5. (**moderate deviations principle**) Let the sequence $\kappa := \kappa_n$, $\kappa \in \mathbb{R}$ satisfies the conditions

$$\lim_{n \rightarrow \infty} \frac{\kappa}{n} = 0, \quad \lim_{n \rightarrow \infty} \frac{\kappa}{\sqrt{n}} = \infty.$$

Then the sequence of random variables $\tilde{R}(n) := \frac{R(n) - an}{\kappa}$ satisfies LDP with normalized function $\varphi(n) := \frac{\kappa^2}{n}$ and rate function

$$I(y) := \frac{y^2}{2\sigma^2}.$$

The law of large numbers and central limit theorems for the compound renewal process were established in [2] (Thm. 11.5.2, pp. 332 and Thm. 10.6.2, pp. 311). We can recall but not use here the paper [4], where the limit theorems were established for a periodical renewal process and for the special case of compound renewal process, when, in our notations, the components τ and ζ are independent. The more general case, when τ and ζ can be dependent variables was considered in [3], where the central limit theorem was proved in the functional space. It means that the CLT for $R(n)$ can be also derived from [3].

Our proofs of large deviations are based on the paper [21], where the large deviations result was proved for the compound renewal process. The large deviations for the renewal processes is also a popular topic, and it was developed in several works. Without pretending to make a complete review we will recall some of these works and highlight the principal differences from the results of [21].

We should start, probably, with the classical renewal process which corresponds to our case when $\zeta_k \equiv 1$. In [23] the rough asymptotics of large deviations was obtained for this case. More general case was studied in [17] where again the rough asymptotics for the special type of compound renewal process was proved. The considered process corresponds to the case when the reward paid ζ_k (in terms of renewal reward processes) is a function of the time elapsed to accomplish it τ_k , *i.e.* $\zeta_k = F(\tau_k)$, where $F(\cdot)$ is a deterministic function. In our case, a more complex random dependence is allowed.

Finally, in all papers [8, 9, 17, 23] the rough asymptotics for large deviation was proved. These asymptotics can provide the results of the following type

$$\mathbf{P}(R(n) = ny) = e^{-nD(y)(1+o(1))}.$$

In [21] the *exact* asymptotics was established, which provide the result like

$$\mathbf{P}(R(n) = ny) = C(1 + o(1))e^{-nD(y)},$$

where the constant was found explicitly.

3. PROOF OF THEOREM 2.2

Proof of statements 1) and 2). Since $\zeta_k \leq \tau_k$ a.s. when $k \in \mathbb{N}$, then the following inequality holds

$$Z(n) \leq R(n) \leq Z(n) + \tau_{\nu(n)+1}. \quad (3.1)$$

Using Lemma 4.4 and Borel-Cantelli lemma, when $n \rightarrow \infty$, it is easy to see that

$$\frac{\tau_{\nu(n)+1}}{\sqrt{n}} \rightarrow 0 \text{ a.s.}$$

Thus the statements (1) and (2) follow from the inequality (3.1) and corresponding results for $Z(n)$ (see [2], Thm. 11.5.2, pp. 332 and Thm. 10.6.2, pp. 311). \square

Proof of statement 3). Consider

$$\begin{aligned} L_n(x) &:= \sum_{k=1}^{\infty} \sum_{l=1}^{[\ln^2 n]} \mathbf{P}(R(n) = x, T_k = n - l, \tau_{k+1} \geq l) \\ &= \sum_{k=1}^{\infty} \sum_{l=1}^{[\ln^2 n]} \sum_{s=0}^l \mathbf{P}(Z_k = x - s, R(n) - Z_k = s, T_k = n - l, \tau_{k+1} \geq l). \end{aligned} \quad (3.2)$$

Note that if $T_k = n - l$, then the random variable Z_k uniquely defined by the values of Markov chain $Y(m)$ when $m < n - l$, but the random variables $R(n) - Z_k$ and τ_{k+1} depend on the values of the chain $Y(m)$ when $m > n - l$. Therefore, by the inclusion

$$\{\omega : T_k = n - l\} \subseteq \{\omega : Y(n - l) = y_0\}$$

the following equality holds

$$\begin{aligned} &\mathbf{P}(Z_k = x - s, R(n) - Z_k = s, \tau_{k+1} \geq l \mid T_k = n - l) \\ &= \mathbf{P}(Z_k = x - s \mid T_k = n - l) \mathbf{P}(R(n) - Z_k = s, \tau_{k+1} \geq l \mid T_k = n - l). \end{aligned} \quad (3.3)$$

Applying (3.2) and (3.3) we obtain

$$\begin{aligned} L_n(x) &= \sum_{k=1}^{\infty} \sum_{l=1}^{[\ln^2 n]} \sum_{s=0}^l \mathbf{P}(Z_k = x - s, R(n) - Z_k = s, T_k = n - l, \tau_{k+1} \geq l) \\ &= \sum_{k=1}^{\infty} \sum_{l=1}^{[\ln^2 n]} \sum_{s=0}^l \mathbf{P}(Z_k = x - s \mid T_k = n - l) \mathbf{P}(R(n) - Z_k = s, \tau_{k+1} \geq l \mid T_k = n - l) \mathbf{P}(T_k = n - l) \\ &= \sum_{k=1}^{\infty} \sum_{l=1}^{[\ln^2 n]} \sum_{s=0}^l \mathbf{P}(Z_k = x - s, T_k = n - l) \mathbf{P}(R(n) - Z_k = s, \tau_{k+1} \geq l \mid T_k = n - l). \end{aligned} \quad (3.4)$$

Note that

$$\mathbf{P}(R(n) - Z_k = s, \tau_{k+1} \geq l \mid T_k = n - l) = \mathbf{P}(R(l) = s, \tau \geq l \mid Y(0) = y_0).$$

Thus, from equality (3.4) it follows that

$$L_n(x) = \sum_{l=1}^{[\ln^2 n]} \sum_{s=0}^l \mathbf{P}(R(l) = s, \tau \geq l \mid Y(0) = y_0) \sum_{k=1}^{\infty} \mathbf{P}(Z_k = x - s, T_k = n - l). \quad (3.5)$$

Since $\mathbf{P}(T_0 = n - l) = 0$ when $0 \leq l \leq [\ln^2 n]$, then from Theorem 2.2 [21] it follows, when $n \rightarrow \infty$, that

$$\begin{aligned} \sum_{k=1}^{\infty} \mathbf{P}(Z_k = x - s, T_k = n - l) &= \sum_{k=0}^{\infty} \mathbf{P}(Z_k = x - s, T_k = n - l) \\ &= \frac{1}{\sqrt{n}} \psi_1 \left(\lambda \left(\frac{\tilde{\alpha}}{\theta} \right), \mu \left(\frac{\tilde{\alpha}}{\theta} \right) \right) C_H(\theta, \tilde{\alpha}) e^{-nD(\theta, \tilde{\alpha})} (1 + o(1)), \end{aligned}$$

where

$$\tilde{\alpha} := \frac{x-s}{n}, \quad \theta := \frac{n-l}{n}.$$

Since the function $\psi_1(\lambda(\alpha), \mu(\alpha))$ is continuous in a neighborhood of the point $\alpha = a$, and the function $C_H(\theta, \alpha)$ is continuous in a neighborhood of the point $(\theta, \alpha) = (1, a)$, then, for sufficiently small Δ and $n \rightarrow \infty$ the following equality holds true

$$\sum_{k=1}^{\infty} \mathbf{P}(Z_k = x-s, T_k = n-l) = \frac{1}{\sqrt{n}} \psi_1(\lambda(\alpha_0), \mu(\alpha_0)) C_H(1, \alpha_0) e^{-nD(\theta, \tilde{\alpha})} (1 + o(1)). \quad (3.6)$$

Applying Lemma 4.6 (see Sect. 4) and considering that $0 \leq s \leq l \leq [\ln^2 n]$ and $|\alpha_0 - a| < \Delta$, from the equality (3.6) we obtain

$$\begin{aligned} & \sum_{k=1}^{\infty} \mathbf{P}(Z_k = x-s, T_k = n-l) \\ &= \frac{1}{\sqrt{n}} \psi_1(\lambda(\alpha_0), \mu(\alpha_0)) C_H(1, \alpha_0) e^{-nD(\frac{x}{n}) + (\lambda(\frac{x}{n}) + \varepsilon_n)l + (\mu(\frac{x}{n}) + \theta_n)s} (1 + o(1)). \end{aligned} \quad (3.7)$$

Show that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{l=1}^{[\ln^2 n]} \sum_{s=0}^l \mathbf{P}(R(l) = s, \tau \geq l | Y(0) = y_0) e^{(\lambda(\frac{x}{n}) + \varepsilon_n)l + (\mu(\frac{x}{n}) + \theta_n)s} \\ &= \sum_{l=1}^{\infty} e^{\lambda(\alpha_0)l} \mathbf{E}(e^{\mu(\alpha_0)R(l)}, \tau \geq l | Y(0) = y_0) =: I(\alpha_0). \end{aligned} \quad (3.8)$$

Due to the fact that

$$\lim_{n \rightarrow \infty} e^{(\lambda(\frac{x}{n}) + \varepsilon_n)l + (\mu(\frac{x}{n}) + \theta_n)s} = e^{\lambda(\alpha_0)l + \mu(\alpha_0)s}$$

the equality (3.8) will be proved if we can show that the series

$$\sum_{l=1}^{\infty} \sum_{s=0}^l \mathbf{P}(R(l) = s, \tau \geq l | Y(0) = y_0) e^{(\lambda(\frac{x}{n}) + \varepsilon_n)l + (\mu(\frac{x}{n}) + \theta_n)s} \quad (3.9)$$

converges.

Note that if $\tau \geq l$, then $\zeta \geq R(l)$, thus

$$\begin{aligned} & \sum_{l=1}^{\infty} \sum_{s=0}^l \mathbf{P}(R(l) = s, \tau \geq l | Y(0) = y_0) e^{(\lambda(\frac{x}{n}) + \varepsilon_n)l + (\mu(\frac{x}{n}) + \theta_n)s} \\ &= \sum_{l=1}^{\infty} e^{(\lambda(\frac{x}{n}) + \varepsilon_n)l} \mathbf{E}(e^{(\mu(\frac{x}{n}) + \theta_n)R(l)}, \tau \geq l | Y(0) = y_0) \\ &\leq \sum_{l=1}^{\infty} e^{(\lambda(\frac{x}{n}) + \varepsilon_n)l} \mathbf{E}(e^{(\mu(\frac{x}{n}) + \theta_n)\zeta}, \tau \geq l | Y(0) = y_0). \end{aligned} \quad (3.10)$$

Due to the Cramer's condition $[\mathbf{C}_0]$ for sufficiently small $\Delta > 0$ and sufficiently large n

$$\mathbf{E}(e^{2(\mu(\frac{x}{n})+\theta_n)\zeta} | Y(0) = y_0) < \infty$$

and there exists $\rho > 0$ such that

$$\mathbf{E}(e^{2(\lambda(\frac{x}{n})+\varepsilon_n+\rho)\tau} | Y(0) = y_0) < \infty.$$

Therefore, using Cauchy-Bunyakovsky-Schwarz and Chebyshev inequalities, we obtain

$$\begin{aligned} & \mathbf{E}(e^{(\mu(\frac{x}{n})+\theta_n)\zeta}, \tau \geq l | Y(0) = y_0) \\ & \leq (\mathbf{E}e^{2(\mu(\frac{x}{n})+\theta_n)\zeta} | Y(0) = y_0)^{\frac{1}{2}} (\mathbf{P}(\tau \geq l | Y(0) = y_0))^{\frac{1}{2}} \\ & \leq (\mathbf{E}e^{2(\mu(\frac{x}{n})+\theta_n)\zeta} | Y(0) = y_0)^{\frac{1}{2}} (\mathbf{E}e^{2(\lambda(\frac{x}{n})+\varepsilon_n+\rho)\tau} | Y(0) = y_0)^{\frac{1}{2}} e^{-(\lambda(\frac{x}{n})+\varepsilon_n+\rho)l} \\ & =: K e^{-(\lambda(\frac{x}{n})+\varepsilon_n+\rho)l}. \end{aligned} \tag{3.11}$$

Using (3.10), (3.11), we obtain

$$\sum_{l=1}^{\infty} \sum_{s=0}^l \mathbf{P}(R(l) = s, \tau \geq l | Y(0) = y_0) e^{(\lambda(\frac{x}{n})+\varepsilon_n)l + (\mu(\frac{x}{n})+\theta_n)s} \leq K \sum_{l=1}^{\infty} e^{-\rho l} < \infty.$$

Thus, the series (3.9) converges, hence the equality (3.8) holds true.

From (3.5), (3.7) and (3.8) it follows that

$$L_n(x) = \frac{1}{\sqrt{n}} \psi_1(\lambda(\alpha_0), \mu(\alpha_0)) C_H(1, \alpha_0) I(\alpha_0) e^{-nD(\frac{x}{n})} (1 + o(1)). \tag{3.12}$$

It is obvious that the following inequality holds

$$L_n(x) \leq \mathbf{P}(R(n) = x) \leq L_n(x) + \mathbf{P}(R(n) = x, \tau_{\nu(n)+1} \geq \ln^2 n). \tag{3.13}$$

From Lemma 4.5 (see Sect. 4) it follows that

$$\mathbf{P}(R(n) = x, \tau_{\nu(n)+1} \geq \ln^2 n) \leq \tilde{C} e^{-nD(\alpha) - \tilde{\gamma} \ln^2 n}. \tag{3.14}$$

From (3.12), (3.13) and (3.14) it follows that

$$\begin{aligned} \mathbf{P}(R(n) = x) & = (1 + o(1)) L_n(x) \\ & = \frac{1}{\sqrt{n}} \psi_1(\lambda(\alpha_0), \mu(\alpha_0)) C_H(1, \alpha_0) I(\alpha_0) e^{-nD(\frac{x}{n})} (1 + o(1)). \end{aligned} \tag{3.15}$$

□

Proof of statement 4). Due to the fact that $(\lambda(a), \mu(a)) = (0, 0)$ and the function $I(\alpha)$ is continuous in a neighborhood of the point $\alpha = a$ we will have

$$\begin{aligned} I(a) &= \sum_{l=1}^{\infty} e^{\lambda(a)l} \mathbf{E}(e^{\mu(a)R(l)}, \tau \geq l \mid Y(0) = y_0) \\ &= \sum_{l=1}^{\infty} \mathbf{P}(\tau \geq l \mid Y(0) = y_0) = \mathbf{E}(\tau \mid Y(0) = y_0). \end{aligned} \quad (3.16)$$

From Lemma 2.1 (see [21]) it follows that

$$C_H(1, a) = \frac{1}{\mathbf{E}(\tau \mid Y(0) = y_0) \sigma \sqrt{2\pi}}.$$

Hence, from (3.15) and (3.16) we obtain

$$\mathbf{P}(R(n) = x) = \frac{1}{\sigma \sqrt{2\pi n}} e^{-nD(\frac{x}{n})} (1 + o(1)).$$

□

Proof of statement 5). From Consequence 3.2 (see [19]) it follows that the sequence of random variables $\tilde{Z}(n) := \frac{Z(n) - an}{\kappa}$ satisfies LDP with normalized function $\varphi(n) = \frac{\kappa^2}{n}$ and rate function $I(y)$.

Using Lemma 4.4, for any $\varepsilon > 0$ we will have

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n}{\kappa^2} \ln \mathbf{P}(|\tilde{R}(n) - \tilde{Z}(n)| > \varepsilon) &\leq \lim_{n \rightarrow \infty} \frac{n}{\kappa^2} \ln \mathbf{P}(\tau_{\nu(n)+1} > \kappa \varepsilon) \\ &\leq \lim_{n \rightarrow \infty} \frac{n}{\kappa^2} \ln e^{-\frac{\rho}{4} \kappa \varepsilon} = - \lim_{n \rightarrow \infty} \frac{n \rho \varepsilon}{4 \kappa} = -\infty. \end{aligned}$$

Therefore from Theorem 4.2.13 (see [11]) we obtain that the sequences $\tilde{R}(n)$ and $\tilde{Z}(n)$ satisfy the same LDP. □

4. AUXILIARY RESULTS

Lemma 4.1. *For any $k, n \in \mathbb{N}$ the following inequality holds*

$$\mathbf{P}(\tau_k \geq n) \leq C e^{-\rho n}, \quad (4.1)$$

where

$$C := \frac{1}{1 - (1 - \delta_1)^{v-1} \delta_2}, \quad \rho := \frac{1}{v} \ln C.$$

Proof. Due to the fact that the process $Y(n)$ is markovian it is sufficient to prove Lemma 4.1 for τ_1 with an arbitrary initial condition. We fix some initial state $Y(0) = (y_1, \mathbf{y}_2)$. Since $C > 1$, $C e^{-\rho n} = C^{1 - \frac{n}{v}}$, then for $n \leq v$ the right-hand side of inequality (4.1) is not less than 1, therefore (4.1) obviously holds true.

We prove now the inequality (4.1) when $n \geq v + 1$. Denote $k := \lceil \frac{n}{v} \rceil$ and for $l = 0, 1, \dots, k - 1$ we consider the events

$$A_l := \{\omega : r_{vl+1}(n) = 0, \dots, r_{v(l+v-1)}(n) = 0, r_{v(l+v)}(n) = 1\}.$$

Denote B the complement to the set $\bigcup_{l=0}^{k-1} A_l$:

$$B := \bigcap_{l=0}^{k-1} \overline{A_l}.$$

Since it is obvious that

$$\{\tau_1 < n\} \supset \bigcup_{l=0}^{k-1} A_l,$$

then we obtain

$$\{\tau_1 \geq n\} \subseteq B.$$

Hence we have

$$\begin{aligned} P_n &:= \mathbf{P}(\tau_1 \geq n \mid Y(0) = (y_1, \mathbf{y}_2)) \\ &\leq \mathbf{P}(B \mid Y(0) = (y_1, \mathbf{y}_2)) = \mathbf{P}\left(\bigcap_{l=0}^{k-1} \overline{A_l} \mid Y(0) = (y_1, \mathbf{y}_2)\right) \\ &= \mathbf{P}(\overline{A_0} \mid Y(0) = (y_1, \mathbf{y}_2)) \prod_{l=1}^{k-1} \mathbf{P}\left(\overline{A_l} \mid \bigcap_{i=0}^{l-1} \overline{A_i}, Y(0) = (y_1, \mathbf{y}_2)\right). \end{aligned}$$

Since by condition **[A]** each cofactor in the right-hand side has the following upper bound

$$1 - (1 - \delta_1)^{v-1} \delta_2 = \frac{1}{C},$$

then we have

$$P_n \leq \left(\frac{1}{C}\right)^k \leq \left(\frac{1}{C}\right)^{\frac{n}{v}-1} = Ce^{-\rho n}.$$

□

Lemma 4.2. *For any $u \in \mathbb{Z}^2$ the equality $f(2\pi u) = 1$ holds true, and for any $u \in \mathbb{R}^2 \setminus \mathbb{Z}^2$ the inequality $|f(2\pi u)| < 1$ holds, where*

$$f(u) := \mathbf{E}e^{i(u_1\tau + u_2\zeta)}$$

is characteristic function for ξ .

Proof. Since τ and ζ are integers numbers, then, it is obvious, that for $u \in \mathbb{Z}^2$ the equality $f(2\pi u) = 1$ holds true. We show that for any $u \in \mathbb{R}^2 \setminus \mathbb{Z}^2$ the inequality $|f(2\pi u)| < 1$ holds true.

Suppose that it is not true, then there exists $(u_1, u_2) \in \mathbb{R}^2 \setminus \mathbb{Z}^2$ such that $|\mathbf{E}e^{2\pi i(u_1\tau + u_2\zeta)}| = 1$. Note that equality $|\mathbf{E}e^{2\pi i(u_1\tau + u_2\zeta)}| = 1$ is equivalent to the fact that there exists $k \in \mathbb{R}$ such that

$$2\pi(u_1\tau + u_2\zeta) = k \pmod{2\pi} \text{ a.s.}$$

From Condition [A] it follows that

$$\begin{aligned}\mathbf{P}(\zeta = 1, \tau = s + 1) &\geq (1 - \delta_1)^s \delta_2 > 0, \\ \mathbf{P}(\zeta = 2, \tau = s + 1) &\geq \delta_2 (1 - \delta_1)^{s-1} \delta_2 > 0, \\ \mathbf{P}(\zeta = 1, \tau = s + 2) &\geq (1 - \delta_1)^{s+1} \delta_2 > 0.\end{aligned}$$

Thus, if or hypothesis is true, then should exist $k_1 \in \mathbb{Z}$, $k_2 \in \mathbb{Z}$, $k_3 \in \mathbb{Z}$ such that the following inequalities holds true

$$\begin{cases} 2\pi(u_1(s+1) + u_2) = k + 2\pi k_1 \\ 2\pi(u_1(s+1) + 2u_2) = k + 2\pi k_2 \\ 2\pi(u_1(s+2) + u_2) = k + 2\pi k_3. \end{cases}$$

Divide each equality by 2π . Subtracting from the 2nd equality the 1st we obtain $u_2 = k_2 - k_1 \in \mathbb{Z}$; subtracting from the 3rd equality the 1st, we obtain $u_1 = k_3 - k_1 \in \mathbb{Z}$. The resulting contradiction completes the proof. \square

For the vector $(\tilde{\lambda}, \tilde{\mu})$ such that $\psi(\tilde{\lambda}, \tilde{\mu}) = 1$, we consider the sequence of random vectors $(\hat{\tau}_k, \hat{\zeta}_k)$, $k \in \mathbb{N}$, whose joint distribution is given as follows

$$\mathbf{P}((\hat{\tau}_1, \hat{\zeta}_1) \in A_1, \dots, (\hat{\tau}_k, \hat{\zeta}_k) \in A_k, \dots) := \frac{1}{\psi_1(\tilde{\lambda}, \tilde{\mu})} \prod_{k=1}^{\infty} \mathbf{E}(e^{\tilde{\lambda}\tau_k + \tilde{\mu}\zeta_k}; (\tau_k, \zeta_k) \in A_k). \quad (4.2)$$

Let $\hat{\tau}_0 := 0$, $\hat{\zeta}_0 := 0$, $\hat{\nu}(0) := 0$. Denote

$$\hat{T}_k := \sum_{l=0}^k \hat{\tau}_l, \quad \hat{\nu}(n) := \max\{k \geq 0 : \hat{T}_k < n\}.$$

Lemma 4.3. *Let $\gamma + \tilde{\lambda} + \tilde{\mu} < \rho$, then there exists the constant $\hat{C} > 0$, such that for any $n \in \mathbb{N}$ the following inequality holds true*

$$\mathbf{E}e^{\gamma\hat{\nu}(n)+1} < \hat{C}n.$$

Proof. Since random variables $\hat{\tau}_{k+1}$ and \hat{T}_k are independent, then

$$\begin{aligned}E_1 &:= \mathbf{E}e^{\gamma\hat{\nu}(n)+1} = \mathbf{E}(e^{\gamma\hat{\tau}_1}; \hat{\tau}_1 \geq n) + \sum_{k=1}^{\infty} \mathbf{E}(e^{\gamma\hat{\tau}_{k+1}}; \hat{T}_k < n \leq \hat{T}_{k+1}) \\ &\leq \frac{1}{\psi_1(\tilde{\lambda}, \tilde{\mu})} \mathbf{E}e^{\gamma\tau_1} + \mathbf{E}e^{\gamma\tau + \tilde{\lambda}\tau + \tilde{\mu}\zeta} \sum_{k=1}^{\infty} \mathbf{P}(\hat{T}_k < n).\end{aligned}$$

Due to the fact that by arithmeticity the inequality $\hat{T}_k \geq n$ a.s. when $k \geq n$. Therefore, using Lemma 4.1 and inequality $\tau_k \geq \zeta_k$ a.s., we obtain

$$\begin{aligned}E_1 &\leq \frac{1}{\psi_1(\tilde{\lambda}, \tilde{\mu})} \mathbf{E}e^{\gamma\tau_1} + \mathbf{E}e^{(\gamma + \tilde{\lambda} + \tilde{\mu})\tau} n \leq \frac{1}{\psi_1(\tilde{\lambda}, \tilde{\mu})} C \sum_{k=1}^{\infty} e^{(\gamma - \rho)k} + Cn \sum_{k=1}^{\infty} e^{(\gamma + \tilde{\lambda} + \tilde{\mu} - \rho)k} \\ &\leq \left(\frac{C}{\psi_1(\tilde{\lambda}, \tilde{\mu})} \frac{e^{(\gamma - \rho)}}{1 - e^{(\gamma - \rho)}} + \frac{C e^{(\gamma + \tilde{\lambda} + \tilde{\mu} - \rho)}}{1 - e^{(\gamma + \tilde{\lambda} + \tilde{\mu} - \rho)}} \right) n.\end{aligned}$$

\square

Lemma 4.4. *Let $\lim_{n \rightarrow \infty} \kappa_n = \infty$. Then for sufficiently large $n \in \mathbb{N}$ the following holds true*

$$\mathbf{P}(\tau_{\nu(n)+1} \geq \kappa_n) \leq e^{-\frac{\rho}{4}\kappa_n}.$$

Proof. Since random variables τ_{k+1} and T_k are independent, then

$$\begin{aligned} \mathbf{P}(\tau_{\nu(n)+1} \geq \kappa_n) &\leq \mathbf{P}(\tau_1 \geq \kappa_n) + \sum_{k=1}^{\infty} \mathbf{P}(\tau_{k+1} \geq \kappa_n, T_k < n \leq T_{k+1}) \\ &\leq \mathbf{P}(\tau_1 \geq \kappa_n) + \mathbf{P}(\tau \geq \kappa_n) \sum_{k=1}^{\infty} \mathbf{P}(T_k < n). \end{aligned}$$

Due to the arithmeticity the inequality $T_k \geq n$ holds true almost surely when $k \geq n$. By Lemma 4.1 and Chebyshev inequality for sufficiently large n we obtain

$$\mathbf{P}(\tau_{\nu(n)+1} \geq \kappa_n) \leq \frac{\mathbf{E}e^{\frac{\rho}{2}\tau_1}}{e^{\frac{\rho}{2}\kappa_n}} + n \frac{\mathbf{E}e^{\frac{\rho}{2}\tau}}{e^{\frac{\rho}{2}\kappa_n}} \leq e^{-\frac{\rho}{4}\kappa_n}.$$

□

Lemma 4.5. *There exist constants $\Delta > 0$, $\tilde{C} > 0$, $\tilde{\gamma} > 0$ such that for $x \in 0 \cup \mathbb{N}$, $\alpha := \frac{x}{n}$, $n \geq 1$, $|\alpha - a| \leq \Delta$ the following inequality holds true*

$$\mathbf{P}(R(n) = x, \tau_{\nu(n)+1} \geq \ln^2 n) \leq \tilde{C}e^{-nD(\alpha) - \tilde{\gamma}\ln^2 n}.$$

Proof. By Theorem 2.1, [21], it follows that for sufficiently small Δ there exists $\lambda(\alpha)$ and $\mu(\alpha)$ such that $(\lambda(\alpha), \mu(\alpha)) \in \mathcal{A}^{\leq 0}$, $A(\lambda(\alpha), \mu(\alpha)) = 0$ and

$$D(\alpha) = \lambda(\alpha) + \alpha\mu(\alpha).$$

Denote

$$B_n := \{\omega : \tau_{\nu(n)+1} \geq \ln^2 n\}.$$

We have

$$\begin{aligned} \mathbf{P}(R(n) = x, \tau_{\nu(n)+1} \geq \ln^2 n) \\ = \mathbf{P}(R(n) = x, B_n, \nu(n) = 0) + \mathbf{P}(R(n) = x, B_n, \nu(n) \geq 1) := \mathbf{P}_0 + \mathbf{P}_1. \end{aligned}$$

From Lemma 4.1 it follows that

$$\mathbf{P}_0 \leq \mathbf{P}(\tau_1 \geq n) \leq Ce^{-\rho n}.$$

Since the function $D(\alpha)$ is continuous in a neighborhood of the point $\alpha = a$ and $D(a) = 0$, then for sufficiently small $\Delta > 0$ and $\alpha : |\alpha - a| \leq \Delta$ the following inequality holds

$$\mathbf{P}_0 \leq Ce^{-\rho n} \leq Ce^{-D(\alpha)n - c\ln^2 n}. \quad (4.3)$$

Denote $Z_k := \sum_{l=1}^k \zeta_l$.

Estimate from above \mathbf{P}_1 . For $\lambda = \lambda(\alpha)$, $\mu = \mu(\alpha)$ we obtain

$$\begin{aligned} \mathbf{P}_1 &= \mathbf{P}(R(n) = x, B_n, \nu(n) \geq 1) = \sum_{k=1}^{\infty} \mathbf{P}(R(n) = x, B_n, \nu(n) = k) \\ &= \sum_{k=1}^{\infty} \mathbf{P}(R(n) = x, \tau_{k+1} \geq \ln^2 n, T_k < n \leq T_{k+1}) \\ &= e^{-D(\alpha)n} \sum_{k=1}^{\infty} \mathbf{E}(e^{-\lambda(T_{k+1}-n) - \mu(Z_{k+1}-x) + \lambda T_{k+1} + \mu Z_{k+1}}; R(n) = x, \tau_{k+1} \geq \ln^2 n, T_k < n \leq T_{k+1}). \end{aligned} \quad (4.4)$$

Note that if $T_{k+1} \geq n$ and $R(n) = x$, then $x + \zeta_{k+1} \geq Z_{k+1}$, therefore from (4.4) it is follows that

$$\begin{aligned} \mathbf{P}_1 &\leq e^{-D(\alpha)n} \sum_{k=1}^{\infty} \mathbf{E}(e^{|\lambda|\tau_{k+1} + |\mu|\zeta_{k+1} + \lambda T_{k+1} + \mu Z_{k+1}}; R(n) = x, \tau_{k+1} \geq \ln^2 n, T_k < n \leq T_{k+1}) \\ &\leq e^{-D(\alpha)n} \sum_{k=1}^{\infty} \mathbf{E}(e^{|\lambda|\hat{\tau}_{k+1} + |\mu|\hat{\zeta}_{k+1}}; \hat{\tau}_{k+1} \geq \ln^2 n, \hat{T}_k < n \leq \hat{T}_{k+1}), \end{aligned} \quad (4.5)$$

where the joint distribution of random variables $(\hat{\tau}_k, \hat{\zeta}_k)$, $k \in \mathbb{N}$ has the form (compare with (4.2))

$$\mathbf{P}((\hat{\tau}_1, \hat{\zeta}_1) \in A_1, \dots, (\hat{\tau}_k, \hat{\zeta}_k) \in A_k, \dots) := \frac{1}{\psi_1(\lambda(\alpha), \mu(\alpha))} \prod_{k=1}^{\infty} \mathbf{E}(e^{\lambda(\alpha)\tau_k + \mu(\alpha)\zeta_k}; (\tau_k, \zeta_k) \in A_k), \quad (4.6)$$

Making the summation in the inequality (4.5) we obtain

$$\mathbf{P}_1 \leq e^{-D(\alpha)n} \mathbf{E}(e^{|\lambda|\hat{\tau}_{\nu(n)+1} + |\mu|\hat{\zeta}_{\nu(n)+1}}; \hat{\tau}_{\nu(n)+1} \geq \ln^2 n).$$

Since $\tau_k \geq \zeta_k$ a.s., then from (4.6) it is follows that $\hat{\tau}_k \geq \hat{\zeta}_k$ a.s. Therefore, using Cauchy-Bunyakovsky-Schwarz inequality we obtain

$$\mathbf{P}_1 \leq e^{-D(\alpha)n} \left(\mathbf{E}e^{2(|\lambda|+|\mu|)\hat{\tau}_{\nu(n)+1}} \right)^{\frac{1}{2}} \left(\mathbf{P}(\hat{\tau}_{\nu(n)+1} \geq \ln^2 n) \right)^{\frac{1}{2}}. \quad (4.7)$$

From Consequence 2.1, [21], it is follows that $(\lambda(a), \mu(a)) = (0, 0)$, therefore due to the continuity in the point $\alpha = a$ of the function $D(\alpha)$ for sufficiently small $\Delta > 0$ the following inequality $3(|\lambda(\alpha)| + |\mu(\alpha)|) < \rho$ holds true.

Thus, from Lemma 4.3 it is follows that there exists a constant $C_1 > 0$ such that

$$\mathbf{E}e^{2(|\lambda|+|\mu|)\hat{\tau}_{\nu(n)+1}} \leq C_1 n.$$

Using Lemma 4.3 and Chebyshev inequality for some $C_2 > 0$, $\gamma > 0$ we obtain

$$\mathbf{P}(\hat{\tau}_{\nu(n)+1} \geq \ln^2 n) < C_2 e^{-\gamma \ln^2 n}.$$

From inequalities (4.3), (4.7) it follows that for sufficiently large n

$$\begin{aligned} \mathbf{P}(R(n) = x, \tau_{\nu(n)+1} \geq \ln^2 n) &\leq C e^{-D(\alpha)n - c \ln^2 n} + e^{-D(\alpha)n} \sqrt{C_1 C_2} \sqrt{n} e^{-\frac{\gamma}{2} \ln^2 n} \\ &\leq C e^{-D(\alpha)n - c \ln^2 n} + \sqrt{C_1 C_2} e^{-D(\alpha)n - \frac{\gamma}{4} \ln^2 n}. \end{aligned}$$

□

Lemma 4.6. *There exists $\Delta > 0$ such that if $\lim_{n \rightarrow \infty} \frac{x}{n} = \alpha_0$, $|a - \alpha_0| \leq \Delta$, then the following holds*

$$-nD\left(1 - \frac{m}{n}, \frac{x}{n} - \frac{y}{n}\right) = -nD\left(1, \frac{x}{n}\right) + \left(\lambda\left(\frac{x}{n}\right) + \varepsilon_n\right)m + \left(\mu\left(\frac{x}{n}\right) + \theta_n\right)y, \quad (4.8)$$

where functions $\varepsilon_n = \varepsilon_n(m, y)$, $\theta_n = \theta_n(m, y)$ satisfy

$$\beta_n := \max_{(m, y) \in \mathcal{B}_n} \{|\varepsilon_n(m, y)| + |\theta_n(m, y)|\} = o(1) \quad \text{when } n \rightarrow \infty,$$

where $\mathcal{B}_n := \{(m, y) \in \mathbb{Z}^2 : 1 \leq m \leq [n\kappa_n], 1 \leq |y| \leq [n\kappa_n]\}$, $\kappa_n = o(1)$ when $n \rightarrow \infty$.

Proof. For sufficiently small Δ if $|a - \alpha_0| \leq \Delta$, then the function $D(\alpha)$ is analytic in some neighborhood of the point α_0 . Therefore, due to the fact that $D(\theta, \alpha) = \theta D(\frac{\alpha}{\theta})$ the function $D(\theta, \alpha)$ is analytic in some neighborhood of the point $(1, \alpha_0)$. It means that for sufficiently large n the function can be represented as a Taylor series in a neighborhood of the point $(1, \frac{x}{n})$. Therefore we have

$$-nD\left(1 - \frac{m}{n}, \frac{x}{n} - \frac{y}{n}\right) = -nD\left(1, \frac{x}{n}\right) + \lambda\left(\frac{x}{n}\right)m + \mu\left(\frac{x}{n}\right)y - nM_1, \quad (4.9)$$

where M_1 is the remainder term in Lagrange's form

Denote

$$D''_{(1,1)}(x, y) := \frac{\partial^2}{\partial x^2} D(x, y), \quad D''_{(1,2)}(x, y) := \frac{\partial^2}{\partial x \partial y} D(x, y), \quad D''_{(2,2)}(x, y) := \frac{\partial^2}{\partial y^2} D(x, y).$$

Then there exist $u \in [1 - \frac{m}{n}, 1]$, $v \in [\frac{x}{n} - \frac{y}{n}, \frac{x}{n}]$ such that

$$|M_1| \leq 2 \max(|D''_{(1,1)}(u, v)|, |D''_{(1,2)}(u, v)|, |D''_{(2,2)}(u, v)|) \left(\frac{m^2}{n^2} + \frac{y^2}{n^2}\right) := K \left(\frac{m^2}{n^2} + \frac{y^2}{n^2}\right).$$

Therefore, from (4.9) it follows that there exist $\varepsilon_n(m, y)$, $\theta_n(m, y)$ such that

$$|\varepsilon_n(m, y)| \leq K \frac{m}{n}, \quad |\theta_n(m, y)| \leq K \frac{y}{n}$$

and the equality (4.8) holds. □

REFERENCES

- [1] G. Bejerano and G. Yona, Variations on probabilistic suffix trees: statistical modeling and prediction of protein families. *Bioinformatics* **17** (2001) 23–43.
- [2] A.A. Borovkov, Probability Theory. Springer (2013).
- [3] A.A. Borovkov, Functional limit theorems for compound renewal processes. *Siberian Math. J.* **60** (2019) 27–40.
- [4] E.V. Bulinskaya and A.I. Sokolova, Limit theorems for generalized renewal processes. *Theory Prob. Appl.* **62** (2018) 35–54.

- [5] P. C enac, A. Le Ny, B. de Loynes and Y. Offret, Persistent random walks. I. Recurrence versus transience. *J. Theor. Prob.* **31** (2018) 232–243.
- [6] P. C enac, B. Chauvin, F. Paccaut and N. Pouyanne, Variable length markov chains, persistent random walks: a close encounter. arXiv:1909.04475 (2019).
- [7] P. C enac, B. Chauvin, F. Paccaut and N. Pouyanne, Context trees, variable length Markov chains and dynamical sources. In *Seminaire de Probabilites XLIV* (pp. 1-39). Springer, Berlin, Heidelberg (2012).
- [8] J-R. Chazottes and E. Olivier, Relative entropy, dimensions and large deviations for g-measures. *J. Phys. A* **33** (2000) 675–689.
- [9] J.R. Chazottes and D. Gabrielli, Large deviations for empirical entropies of g-measures. *Nonlinearity* **18** (2005) 2545–2563.
- [10] F. Comets, R. Fern andez and P.A. Ferrari, Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Probab.* **12** (2002) 921–943.
- [11] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer, New York (1998).
- [12] A. Duarte and G. Ost, A model for neural activity in the absence of external stimuli. *Markov Processes Related Fields* **22** (2016) 37.
- [13] A. Duarte, A. Galves, E. L ocherbach and G. Ost, Estimating the interaction graph of stochastic neural dynamics. To be appear in *BERNOULLI* (2018).
- [14] A. Galves and E. L ocherbach, Stochastic chains with memory of variable length. “Festschrift in honour of the 75th birthday of Jorma Rissanen”, TICSP series **38** (2008) 117–133.
- [15] A. Galves, C. Galves, J.E. Garcia, N.L. Garcia and F. Leonardi, Context tree selection and linguistic rhythm retrieval from written texts. *Ann. Appl. Stat.* **6** (2012) 186–209.
- [16] S. Gallo, Chains with unbounded variable length memory: perfect simulation and a visible regeneration scheme. *Adv. Appl. Prob.* **43** (2011) 735–759.
- [17] R. Lefevre, M. Mariani and L. Zambotti, Large deviations for renewal processes. *Stochastic Process. Appl.* **121** (2011) 2243–2271.
- [18] F.G. Leonardi, A generalization of the PST algorithm: modeling the sparse nature of protein sequences. *Bioinformatics* **22** (2006) 1302–1307.
- [19] A.V. Logachov and A.A. Mogulskii, Anscombe-type theorem and moderate deviations for trajectories of a compound renewal process. *J. Math. Sci.* **229** (2018) 36–50.
- [20] G. Maillard and S. Sch opfer, A functional central limit theorem for regenerative chains. arXiv:0801.2263 (2008).
- [21] A.A. Mogulskii, Local theorems for arithmetic compound renewal processes when Cramer’s condition holds. *Siberian Electr. Math. Rep.* **16** (2019) 21–41.
- [22] J. Rissanen, A universal data compression system. *IEEE Trans. Inf. Theory* **29** (1983) 656–664.
- [23] J. Tiefeng, Large deviations for renewal processes. *Stochastic Process. Appl.* **50** (1994) 57–71.

Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/math-s2o-programme>