



**HAL**  
open science

# Weak law of large numbers for some Markov chains along non homogeneous genealogies

Vincent Bansaye, Chunmao Huang

► **To cite this version:**

Vincent Bansaye, Chunmao Huang. Weak law of large numbers for some Markov chains along non homogeneous genealogies. 2013. hal-00824085

**HAL Id: hal-00824085**

**<https://hal.science/hal-00824085>**

Preprint submitted on 21 May 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Weak law of large numbers for some Markov chains along non homogeneous genealogies

Vincent Bansaye <sup>\*</sup>, Chunmao Huang <sup>†</sup>

May 21, 2013

## Abstract

We consider a population with non-overlapping generations, whose size goes to infinity. It is described by a discrete genealogy which may be time non-homogeneous and we pay special attention to branching trees in varying environments. A Markov chain models the dynamic of the trait of each individual along this genealogy and may also be time non-homogeneous. Such models are motivated by transmission processes in the cell division, reproduction-dispersion dynamics or sampling problems in evolution. We want to determine the evolution of the distribution of the traits among the population, namely the asymptotic behavior of the proportion of individuals with a given trait. We prove some quenched laws of large numbers which rely on the ergodicity of an auxiliary process, in the same vein as [21, 15]. Applications to time inhomogeneous Markov chains lead us to derive a backward (with respect to the environment) law of large numbers and a law of large numbers on the whole population until generation  $n$ . A central limit is also established in the transient case.

*Key words.* Non-homogeneous Markov chain, random environment, branching processes, law of large numbers.

*MSC classes.* 60J05, 60J20, 60J80, 60J85, 60F05.

## 1 Introduction

In this paper, we consider Markov chains which are indexed by discrete trees. This approach is motivated in particular by the study of structured populations. The tree is thus describing the genealogy of the population in discrete time, with non overlapping generations and the nodes of the tree are the individuals. We consider a trait in the population, which could be the location of the individual, its phenotype, its genotype or any biological characteristic. Letting this trait evolve as a Markov chain and be transmitted to the offspring with a random transition leads us to consider a Markov chain indexed by the genealogical tree. Such a process can also be regarded as a branching particle system where the offspring of each particle is given by the genealogy and the associated traits by the Markov chain.

Let  $(\mathcal{X}, B_{\mathcal{X}})$  be a measurable space. The process starts with an initial single individual  $\emptyset$  with trait  $X(\emptyset) \in \mathcal{X}$  whose distribution is  $\nu$ . The initial individual  $\emptyset$  produces a random number  $N = N(\emptyset)$  of particles of generation 1, denoted by  $1, 2, \dots, N$ , with traits determined by

---

<sup>\*</sup>CMAP, Ecole Polytechnique, Palaiseau, France

<sup>†</sup>CMAP, Ecole Polytechnique, Palaiseau, France. Corresponding author at sasamao02@gmail.com.

$$\mathbb{P}(X(1) \in dx_1, \dots, X(k) \in dx_k | N = k, X(\emptyset) = x) = p^{(k)}(0)(x, dx_1, \dots, dx_k),$$

where for each  $k, n \in \mathbb{N}$  and  $x \in \mathcal{X}$ ,  $p^{(k)}(n)(x, \cdot)$  is a probability measure on  $(\mathcal{X}^k, B_{\mathcal{X}^k})$ . More generally, each individual  $u = u_1 \cdots u_n$  of generation  $n$  whose trait is  $X(u)$  yields  $N(u)$  offspring in generation  $n + 1$ , denoted by  $u1, u2, \dots, uN(u)$ , whose traits are determined by

$$\begin{aligned} \mathbb{P}(X(u1) \in dx_1, \dots, X(uk) \in dx_k | N(u) = k, X(u) = x) \\ = p^{(k)}(n)(x, dx_1, \dots, dx_k). \end{aligned}$$

The individuals of each generation evolve independently, so the process enjoys the branching property.

The evolution of a time homogeneous Markov chain indexed by a binary tree is well known thanks to the works of [4, 21]. As soon as the Markov chain along a random lineage of the binary tree is ergodic, a law of large numbers holds. It yields the convergence for the proportions of individuals in generation  $n$  whose trait has some given value. More specifically, this asymptotic proportion is given by the stationary measure of the ergodic Markov chain. This convergence holds in probability in general, and under additional assumptions on the speed of convergence of the Markov chain or uniform ergodicity, it also holds almost surely. Such results have been extended and modified to understand the (random) transmission of some biological characteristic of dividing cells such as cellular aging, cell damages, parasite infection... In particular, [6] considered non ergodic Markov chains and rare events associated to the Markov chain for cell division with parasite infection. In [15], the authors considered a Markov chain indexed by a Galton-Watson tree, which is motivated by cellular aging when the cells may die. In the same vein, the almost sure convergence in the case of bifurcating autoregressive Markov chain is achieved in [9] via martingale arguments. Such results have been extended recently and one can see the works of De Saporta, Gégout- Petit, Marsalle, Blandin and al. For biological motivations in this vein, we also refer to [37, 35].

In this paper, we consider similar questions for the case where both the genealogical tree and the Markov chain along the branches are time non-homogeneous. In particular, we are motivated by the fact that the cell division is affected by the media. This latter is often time non-homogeneous, which may be due to the variations of the available resources or the environment, a medical treatment... Such phenomena are well known in biology from the classical studies of Gause about Paramecium or Tilman about diatoms. The cell genealogy may be modeled by a Galton-Watson process in a varying (or random) environment. It is quite straightforward to extend the weak law of large numbers to the case of non-homogeneous genealogies if the branching events are symmetric and independant (each child obtains an i.i.d. copy) and the Markov chain along the branches is time homogeneous. However, as the convergence of non-homogeneous Markov chains is a delicate problem, we need to consider new limit theorems to understand the evolution of the traits in the cell population. As stated in the next section, the asymptotic proportion can still be characterized as the stationary probability of an auxiliary Markov chain, in the same vein as [15]. It yields a natural interpretation of the repartition of the traits as a stationary probability and the description of the lineage of a typical individual, which then can be easily simulated. A large literature also exists concerning asymptotic behavior of even-odd Markov chains along time homogeneous trees (see e.g. [33]), with different motivations. We stress that in our model the trait of the cell does not influence its division, which means that the genealogical tree may be random but does not depend on the evolution of the Markov chain along its branches. When such a dependence holds (in continuous time, with fixed environments), some many to one formulas can be found in [22] and asymptotic proportions were briefly considered in [7].

Letting the trait be (replaced by) the location of the individual, the process considered here is more usually called a *Branching Markov Chain*. The particular case that the motion of each

individual has i.i.d. increments, i.e. *branching random walks*, has been largely studied from the pioneering works of Biggins [10, 11, 12]. The density occupation, the law of large numbers, central limit theorems, large deviations results and the positions of the extremal particles have been considered. These results, such as recurrence, transience or survival criteria, have been partially extended to random environment both in time and space, see e.g. [20, 30, 24, 13, 14, 39, 28]. Here, we consider both the case of non i.i.d. displacements and non homogeneous environments. We mainly focus on (positive) recurrent branching random walks. This provides some tractable models for reproduction-dispersion of species which evolve in a spatially and temporally non-homogeneous environment and compact state space. With our assumptions, the time environment may influence both the reproduction and the dispersion. One can figure out the effects of the humidity and the enlightenment for the reproduction of plants and the wind for pollination. The spatial environment (such as the intensity of wind, the relief...) may here influence only the dispersion. The fact that the space location does not influence the reproduction events requires space homogeneity of some environmental parameters, such as the light exposure, quality of the ground. An extension of our results to space dependent reproduction is a challenging problem. As a motivating article in this vein in ecology, we refer for instance to [18].

Finally, such models might be a first step to consider evolution processes on larger time scale with time inhomogeneity. The trait would then correspond to a phenotype or a genotype. The fact that the branching event does not depend on the trait (neutral theory) may hold in some cases or be used as a zero hypothesis, see for example [26]. More generally, the non-homogeneity of the branching rates in the genealogies raises many difficulties but has various motivations. As an example, we refer to [36] for discussions on time non-homogeneity for extinction and speciation.

In next two sections, we state the results of this paper and consider some applications. Firstly (Section 2), we give a very general statement which ensures the convergence in probability of the proportions of individuals with a given trait as time goes to infinity. The fact that the common ancestor of two individuals is not recent yields a natural setting for law of large numbers. Here the genealogical tree may be very general but the assumptions that we need are often not satisfied and the asymptotic proportions are not explicit. That's why we focus next on time non-homogeneous tree with branching properties (Section 3). That allows us to get a many to one formula, in the same vein as [21, 7, 22, 16] (see Lemma 5.1). We can then state a forward law of large numbers, which requires ergodic convergence of an auxiliary time non-homogeneous Markov chain. This latter is still not easily satisfied. Thus, we give a backward analog of this result, which yields quite general sufficient conditions linked to ergodicity of Markov chains with stationary Markov transitions [32]. To get assumptions which are easy to check and applications with stationary ergodic environments, we also provide a weak law of large numbers for the whole population. Finally, we derive a central limit theorem and apply it to some branching random walks in random environments. The rest of the paper (Sections 4 and 5) is dedicated to the proofs.

**Notations.** In the whole paper, we need the following notations. If  $u = u_1 \cdots u_n$  and  $v = v_1 \cdots v_m$ , then  $|u| = n$  is the length of  $u$  and  $uv = u_1 \cdots u_n v_1 \cdots v_m$ . We denote by

$$\mathbb{T} \in \cup_{m=0}^{\infty} \{1, 2, \dots\}^m$$

the generation tree rooted at  $\emptyset$  and we define by

$$\mathbb{T}_n := \{u \in \mathbb{T} : |u| = n\}$$

the set of all individuals in generation  $n$ . Let

$$Z_n = \sum_{u \in \mathbb{T}_n} \delta_{X(u)}$$

be the counting measure of particles of generation  $n$ . In fact, for any measurable set  $A$  of  $\mathcal{X}$ ,

$$Z_n(A) := \#\{u \in \mathbb{T}_n : X(u) \in A\}$$

denotes the number of individuals whose trait belongs to  $A$ . Our aim is to obtain the asymptotic behavior of this quantity. In particular, we write

$$N_n := Z_n(\mathcal{X})$$

and we shall consider the asymptotic proportion of individual whose trait belongs to  $A$ , which is given by  $Z_n(A)/N_n$ .

For two different individuals  $u, v$  of a tree, write  $u < v$  if  $u$  is an ancestor of  $v$ , and denote by  $u \wedge v$  the nearest common ancestor of  $u$  and  $v$  in the means that  $|w| \leq |u \wedge v|$  if  $w < u$  and  $w < v$ .

## 2 Weak law of large numbers for non-homogeneous trees

In this section, the genealogical tree  $\mathbb{T}$  is fixed (non random). We require that the size of the population in generation  $n$  goes to infinity as  $n \rightarrow \infty$ .

We consider a transition kernel  $(p^{(k)}(n)(x, dx_1, \dots, dx_k) : k, n \geq 0)$ . Then the Markov chain  $X$  along the tree  $\mathbb{T}$  is specified recursively by

$$\begin{aligned} \mathbb{E} \left[ \prod_{u \in \mathbb{T}_n} F_u(X(u1), \dots, X(uN(u))) \middle| (X(u) : |u| \leq n) \right] \\ = \prod_{u \in \mathbb{T}_n} \int F_u(x_1, \dots, x_{N(u)}) p^{N(u)}(n)(x, dx_1, \dots, dx_k). \end{aligned} \quad (2.1)$$

where  $(F_u : u \in \mathbb{T}) \in \mathcal{B}_b^{\mathbb{T}}$  and  $\mathcal{B}_b$  is the set of bounded measurable functions from  $\cup_{k \geq 0} \mathcal{X}^k$  to  $\mathbb{R}$ .

The trees rooted at  $u$  are defined similarly:

$$\mathbb{T}(u) := \{v : uv \in \mathbb{T}\}, \quad \mathbb{T}_n(u) := \{v : uv \in \mathbb{T}, |v| = n\}$$

$$Z_n^{(u)} := \sum_{uv \in \mathbb{T}_n(u)} \delta_{X(uv)}, \quad N_n(u) := Z_n^{(u)}(\mathcal{X}).$$

**Proposition 2.1.** *Let  $A \in \mathcal{B}_{\mathcal{X}}$ . We assume that*

- (i)  $N_n \rightarrow \infty$  as  $n \rightarrow \infty$ ;
- (ii)  $\limsup_{n \rightarrow \infty} \mathbb{P}(|U_n \wedge V_n| \geq K) \rightarrow 0$  as  $K \rightarrow \infty$ , where  $U_n, V_n$  are two individuals uniformly and independently chosen in  $\mathbb{T}_n$ ;
- (iii) there exists  $\mu(A) \in \mathbb{R}$  such that for all  $u \in \mathbb{T}$  and  $x \in \mathcal{X}$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( X(U_n^{(u)}) \in A \middle| X(u) = x \right) = \mu(A),$$

where  $U_n^{(u)}$  denotes an individual uniformly chosen in  $\mathbb{T}_n(u)$ .

Then

$$\frac{Z_n(A)}{N_n} \rightarrow \mu(A) \quad \text{in } L^2.$$

The Assumption (ii) means that the common ancestor of two individuals chosen randomly is at the beginning of the tree. Then assumption (iii) ensures that any sampling is giving the same distribution. The assumptions (i – ii) hold for many classical genealogies, such as branching genealogies (see below), Wright Fischer (or Moran) genealogies when we let the size of the population  $N$  go to infinity. The assumption (iii) is difficult to obtain in general. We first give a simple example where it holds. The next section is giving better sufficient conditions, in the branching framework. We also provide below a result which weakens the Assumption (ii), since the most recent common ancestor can be in the middle of the tree. It requires a stronger ergodicity along the branches than Assumption (iii).

**Example 2.1. Symmetric independent kernels.** The Assumption (iii) becomes clear in the symmetric and homogeneous case. More precisely it holds if

$$p^{(k)}(n)(x, dx_1, \dots, dx_k) = \prod_{i=1}^k p(x, dx_i)$$

and  $\mathbb{P}_x(Y_n \in A) \rightarrow \mu(A)$  as  $n \rightarrow \infty$ , for every  $x$ , where  $Y_n$  is a Markov chain with transition kernel  $p$ . This problem is related to the ergodicity of  $Y$ . Sufficient conditions for the ergodicity of a Markov chain are known in literature, and we refer e.g. to [29].

**Proposition 2.2.** *Let  $A \in B_{\mathcal{X}}$ . We assume that*

- (i)  $N_n \rightarrow \infty$  as  $n \rightarrow \infty$ ;
- (ii)  $\limsup_{n \rightarrow \infty} \mathbb{P}(|U_n \wedge V_n| \geq n - K) \rightarrow 0$  as  $K \rightarrow \infty$ , where  $U_n, V_n$  are two individuals uniformly and independently chosen in  $\mathbb{T}_n$ ;
- (iii) there exists  $\mu(A) \in \mathbb{R}$  such that

$$\lim_{n \rightarrow \infty} \sup_{u \in \mathbb{T}, x \in \mathcal{X}} \left| \mathbb{P} \left( X(U_n^{(u)}) \in A \mid X(u) = x \right) - \mu(A) \right| = 0,$$

where  $U_n^{(u)}$  denotes an individual uniformly chosen in  $\mathbb{T}_n(u)$ .

Then

$$\frac{Z_n(A)}{N_n} \rightarrow \mu(A) \quad \text{in } L^2.$$

We note that the assumption (ii) is satisfied for any tree  $\mathbb{T}$  where each individual has at most  $q$  (constant) offspring. Considering the symmetric and homogeneous case described in Example 2.1, (iii) is satisfied when a strong ergodicity holds. It is the case for example in finite state space or under Doeblin type conditions. In this situation, Proposition 2.2 can be applied.

### 3 Quenched Law of large numbers for branching Markov chains in random environment

In this section, the genealogical tree  $\mathbb{T}$  may be random. The population evolves following a branching process in random environment (BPRE), described as follows. Let  $\xi = (\xi_0, \xi_1, \dots)$  be a sequence of random variables taking values in some measurable space  $\Omega$ , which will come in

applications below from a stationary and ergodic process. Each  $\xi_n$  corresponds to a probability distribution on  $\mathbb{N} = \{0, 1, 2, \dots\}$ , denoted by  $p(\xi_n) = \{p_k(\xi_n) : k \geq 0\}$ . This infinite vector  $\xi$  is called a *random environment*.

We consider now random measurable transition kernels  $(p_{\xi_n}^{(k)}(x, dx_1, \dots, dx_k) : k, n \geq 0)$ , which are indexed by the  $n$ th environment component  $\xi_n$ . The process  $X$  is a Markov chain along the random tree  $\mathbb{T}$  with transition kernels  $p$ . Conditionally on  $(\xi, \mathbb{T})$ , the process is constructed following (2.1). More specifically, the successive offspring distributions are  $\{p(\xi_n)\}$ , so that the number of offspring  $N(u)$  of individual  $u$  of generation  $n$  is distributed as  $p(\xi_n)$  and the traits of its offspring  $\{X(ui)\}$  are determined by

$$\begin{aligned} \mathbb{P}_\xi(X(u1) \in dx_1, \dots, X(uk) \in dx_k | N(u) = k, X(u) = x) \\ = p_{\xi_n}^{(k)}(x, dx_1, \dots, dx_k). \end{aligned}$$

We note that the offspring number  $N(u)$  does not depend on the parent's trait  $X(u)$ , and the offspring traits  $\{X(ui) : i = 1, \dots, N(u)\}$  may depend on  $N(u)$ ,  $X(u)$  and  $\xi_n$ .

Given  $\xi$ , the conditional probability will be denoted by  $\mathbb{P}_\xi$  and the corresponding expectation by  $\mathbb{E}_\xi$ . The total probability will be denoted by  $\mathbb{P}$  and the corresponding expectation by  $\mathbb{E}$ . As usual,  $\mathbb{P}_\xi$  is called *quenched law*, and  $\mathbb{P}$  *annealed law*.

Let  $\mathcal{F}_0 = \mathcal{F}(\xi) = \sigma(\xi_0, \xi_1, \dots)$  and  $\mathcal{F}_n = \mathcal{F}_n(\xi) = \sigma(\xi_0, \xi_1, \dots, (N(u) : |u| < n))$  be the  $\sigma$ -field generated by the random variables  $N(u)$  with  $|u| < n$ , so that  $N_n$ , the size of the population in generation  $n$ , are  $\mathcal{F}_n$ -measurable. Denote

$$m_n = \sum_k p_k(\xi_n) \quad \text{for } n \geq 0,$$

$$P_0 = 1 \quad \text{and} \quad P_n = m_0 \cdots m_{n-1} \quad \text{for } n \geq 1.$$

Thus, for every  $n \in \mathbb{N}$ ,  $P_n = \mathbb{E}_\xi N_n$ . It is well known that the normalized population size

$$W_n = \frac{N_n}{P_n}$$

is a nonnegative martingale with respect to  $\mathcal{F}_n$ , so the limit

$$W = \lim_{n \rightarrow \infty} W_n$$

exists a.s. and  $\mathbb{E}_\xi W \leq 1$ .

In the rest of this section, we make the following assumptions

**Assumption 3.1.** (i) *The environment  $\xi = (\xi_0, \xi_1, \dots)$  is a stationary ergodic sequence.*

(ii) *We assume that  $\mathbb{P}(m_0 = 0) = 0$ ,  $\mathbb{P}(p_0(\xi_0) = 1) < 1$  and  $\mathbb{E}(\log m_0) < \infty$ .*

(iii) *We focus on the supercritical non degenerated case*

$$\mathbb{E}(\log m_0) > 0, \quad \mathbb{E} \left( \log \frac{\mathbb{E}_\xi N^2}{m_0^2} \right) < \infty. \quad (3.1)$$

The first assumption allows to get asymptotic results on the size of the population. The second assumption avoids some degenerated cases. Denoting by

$$q(\xi) := \mathbb{P}_\xi(N_n = 0 \text{ for some } n)$$

the extinction probability, it is well known that the non-extinction event  $\{N_n \rightarrow \infty\}$  has quenched probability  $1 - q(\xi)$ . Moreover, the condition  $\mathbb{E}(\log m_0) \leq 0$  implies that  $q(\xi) = 1$  a.s., whereas  $\mathbb{E}(\log m_0) > 0$  (supercritical case) yields

$$q(\xi) < 1 \quad \text{a.s.}$$

The last assumption ensures that the random variable  $W$  is positive on the non-extinction event. We refer to [2, 3] for the statements and proofs of these results.

### 3.1 Forward weak law of large numbers in generation $n$

We first give a forward law of large numbers in generation  $n$  for the model introduced above, with the help of an auxiliary Markov process constructed as follows. Let

$$P_{\xi_n}^{(k,i)}(x, \cdot) = p_{\xi_n}^{(k)}(x, \mathcal{X}^{i-1} \times \cdot \times \mathcal{X}^{k-i})$$

and the random transition probability

$$Q_n(x, \cdot) = Q(T^n \xi; x, \cdot) := \frac{1}{m_n} \sum_{k=0}^{\infty} p_k(\xi_n) \sum_{i=1}^k P_{\xi_n}^{(k,i)}(x, \cdot).$$

We note that for each  $\xi \in \Omega$ , the Markov transition kernel  $Q(\xi; \cdot, \cdot)$  is a function from  $\mathcal{X} \times B_{\mathcal{X}}$  into  $[0, 1]$  satisfying:

- for each  $x \in \mathcal{X}$ ,  $Q(\xi; x, \cdot)$  is a probability measure on  $(\mathcal{X}, B_{\mathcal{X}})$ ;
- for each  $A \in B_{\mathcal{X}}$ ,  $Q(\xi; \cdot, A)$  is a  $B_{\mathcal{X}}$ -measurable function on  $\mathcal{X}$ .

Given the environment  $\xi$ , we define an auxiliary Markov chain in varying environment  $Y$ , whose transition probability in generation  $j$  is  $Q_j$  :

$$\mathbb{P}_{\xi}(Y_{j+1} = y | Y_j = x) = Q_j(x, y).$$

As usual, we denote by  $\mathbb{P}_{\xi, x}$  the quenched probability when the process  $Y$  starts from the initial value  $x$ , and by  $\mathbb{E}_{\xi, x}$  the corresponding expectation.

As stated below, the convergence of the measure  $Z_n(\cdot)$  normalized comes from the ergodic behavior of  $Y_n$ . In the same vein as [15], we have

**Theorem 3.2.** *Let  $A \in B_{\mathcal{X}}$ . We assume that there exists a sequence  $(\mu_{\xi, n}(A))_n$  such that for almost every  $\xi$  and for each  $r \in \mathbb{N}$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}^{Tr \xi, x}(Y_{n-r} \in A) - \mu_{\xi, n}(A) = 0 \quad \text{for every } x \in \mathcal{X}, \quad (3.2)$$

where  $T\xi = (\xi_1, \xi_2, \dots)$  if  $\xi = (\xi_0, \xi_1, \dots)$ . Then we have for almost all  $\xi$ ,

$$\frac{Z_n(A)}{P_n} - \mu_{\xi, n}(A)W \rightarrow 0 \quad \text{in } \mathbb{P}_{\xi}\text{-}L^2, \quad (3.3)$$

and conditionally on the non-extinction event,

$$\frac{Z_n(A)}{N_n} - \mu_{\xi, n}(A) \rightarrow 0 \quad \text{in } \mathbb{P}_{\xi}\text{-probability}. \quad (3.4)$$

This forward result theorem is adapted to the underlying branching genealogy. The proof is deferred to the next section, where a more general is obtained. The condition (3.2) holds if the auxiliary Markov chain is weakly ergodic, for suitable sets  $A$ . For sufficient (and necessary) conditions of weak ergodicity in the non-homogeneous case, we refer in particular to [31].

Let us now give more trackable results. We derive a first result of (quenched forward) weak law of large numbers, under a stronger assumption, and two examples in simple cases (homogeneous case).

**Corollary 3.3.** *Let  $A \in B_{\mathcal{X}}$ . We assume that there exists  $\mu(A) \in \mathbb{R}$  such that for almost all  $\xi$ ,*

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\xi, x}(Y_n \in A) = \mu(A) \quad \text{for every } x \in \mathcal{X}. \quad (3.5)$$



Then we have for almost all  $\xi$ ,

$$\frac{Z_n(A)}{P_n} \rightarrow \mu(A)W \quad \text{in } \mathbb{P}_\xi\text{-}L^2, \quad (3.6)$$

and conditionally on the non-extinction event,

$$\frac{Z_n(A)}{N_n} \rightarrow \mu(A) \quad \text{in } \mathbb{P}_\xi\text{-probability.} \quad (3.7)$$

We now give two examples, where we can check (3.5) by considering models for which the associated auxiliary chain  $Y$  is time homogeneous. We also note that assuming a uniform convergence in with respect to  $x$ , we can get an almost sure convergence following the proof of Theorem 2 in [3].

**Example 3.1. Homogeneous Markov chains along Galton Watson trees.** We focus here on the case when the time environment is non random, i.e.  $\xi_n$  is constant for every  $n \in \mathbb{N}$ . The genealogical tree is a Galton Watson tree, whose offspring distribution is specified by  $\{p_k : k \geq 0\}$ . Moreover, we assume that

$$\mathbb{P}_\xi(X(u1) \in dx_1, \dots, X(uk) \in dx_k | N(u) = k, X(u) = x) = p^{(k)}(x, dx_1, \dots, dx_k).$$

does not depend on  $\xi$ . Then, denoting by  $m$  the mean number of offspring per individual and

$$P^{(k,i)}(x, \cdot) = p^{(k)}(x, \mathcal{X}^{i-1} \times \cdot \times \mathcal{X}^{k-i}),$$

the auxiliary process  $Y$  is a time homogeneous Markov chain whose transition kernel is given by

$$Q(x, \cdot) = \frac{1}{m} \sum_{k=0}^{\infty} p_k \sum_{i=1}^k P^{(k,i)}(x, \cdot).$$

Thus a law of large number in probability is obtained as soon as the ergodicity of the Markov chain  $Y$  is proved. Indeed, it ensures that there exists a probability measure  $\mu$  such that for each  $x \in \mathcal{X}$  and a measurable set  $A$  such that  $\mu(\partial A^1) = 0$ , and  $Q^n(x, A) = \mathbb{P}_x(Y_n \in A) \rightarrow \mu(A)$ , so that we can apply Theorem 3.3.

We recall that sufficient conditions for the ergodicity of a Markov chain are known in the literatures, see e.g. [29]. This result is a simple generalization of law of large numbers on the binary tree in [21] and that on Galton Watson trees with at most two offsprings given in [15]. A continuous time analogous result can be found in [8].

**Example 3.2. Symmetric homogeneous Markov chains along branching trees in random environment.** We consider a Branching Markov Chain on  $\mathcal{X}$  in the stationary and ergodic environment  $\xi$ . Given  $\xi$ , for each  $u$  of generation  $n$ , the number of its offspring  $N(u)$  is determined by distribution  $p(\xi_n) = \{p_k(\xi_n) : k \geq 0\}$ . The offspring positions  $\{X(ui)\}$ , independent of each other conditioned on the position of  $u$ , are determined by

$$\mathbb{P}_\xi(X(ui) \in dy | X(u) = x) = p(x, dy),$$

where  $p(x, \cdot)$  is a probability on  $\mathcal{X}$ . We note that this example is a particular case of Example 2.1. We can see that

$$p_{\xi_n}^{(k)}(x, dx_1, \dots, dx_k) = \prod_{i=1}^k \mathbb{P}_\xi(X(ui) \in dx_i | X(u) = x) = \prod_{i=1}^k p(x, dx_i).$$

---

<sup>1</sup> $\partial A$  is the boundary of  $A$

Therefore,  $Q(x, dy) = Q_n(x, dy) = p(x, dy)$  and  $Y_n$  is a time-homogeneous Markov chain with transition probability  $p$ . As in the previous example, the problem is reduced to the ergodicity of  $Y$ . Such a class may be relevant to model the dispersion of plants in spaces where the reproduction is homogeneous and the time environment only influences the genealogy.

The convergence of the Markov chain (3.5) is difficult to get under general assumptions. Indeed, soon as the auxiliary kernel  $Q$  depends on  $\xi$ , the auxiliary process is time non-homogeneous and the forward convergence in distribution (3.5) won't hold in general<sup>2</sup>. It is the case for most of the models we have in mind but the two previous examples. For such an ergodic convergence, the backward convergence is much more adapted (see [32]). Moreover, general sufficient conditions can be found for ergodic (Birkhoff's) theorems and we can use in particular [32, 34]. Thus, we derive in the next subsection a backward law of large numbers and then one on the whole tree.

### 3.2 Backward law of large numbers in generation $n$

Now we consider the environment  $\xi$  time reversed. Thus, for each  $n \in \mathbb{N}$ , we define

$$\xi^{(n)} = (\xi_{n-1}, \dots, \xi_0).$$

For each individual  $u$  of generation  $r$  ( $0 \leq r \leq n$ ), the number of its offspring  $N(u)$  is determined by the distribution  $p(\xi_{n-r-1})$ , and the positions of its offspring  $\{X(ui)\}$  are determined by the transition kernel  $p_{\xi_{n-r-1}}$  defined previously.

To distinguish from the forward case, we denote the counting measure of generation  $k$  by  $Z_k^{(n)}(\cdot)$ , the population size of generation  $k$  by  $N_k^{(n)}$ , and its normalization by

$$W_k^{(n)} = \frac{N_k^{(n)}}{\mathbb{E}_{\xi^{(n)}}(N_k^{(n)})}$$

for  $0 \leq k \leq n$ . We remark that unlike the forward case, here the normalized population  $W_n^{(n)}$  is not a martingale, hence the existence of its limit is not ensured. But when the environment is reversible in law, i.e.

$$(\xi_0, \dots, \xi_{n-1}) \stackrel{d}{=} (\xi_{n-1}, \dots, \xi_0),$$

then  $W_n^{(n)}$  has the same distribution as  $W_n$  under the total probability  $\mathbb{P}$ .

The following theorem is a law of large numbers in generation  $n$  for the backward case, whose proof is deferred to next section.

**Corollary 3.4.** *Let  $A \in B_{\mathcal{X}}$ . We assume that for almost all  $\xi$ ,*

$$m_0 \geq a > 1, \quad \mathbb{E}_{\xi} N^p \leq b$$

*for some constants  $a, b$  and  $p > 2$ . We also assume that there exists  $\mu_{\xi}(A) \in \mathbb{R}$  such that*

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\xi^{(n)}, x}(Y_n \in A) = \mu_{\xi}(A) \quad \text{for every } x \in \mathcal{X}. \quad (3.8)$$

*Then we have for almost all  $\xi$ ,*

$$\frac{Z_n^{(n)}(A) - \mu_{\xi}(A)N_n^{(n)}}{P_n} \rightarrow 0 \quad \text{in } \mathbb{P}_{\xi}\text{-}L^2. \quad (3.9)$$

*Moreover, if the environment is reversible in law then*

$$\mathbf{1}_{\{N_n^{(n)} > 0\}} \left[ \frac{Z_n^{(n)}(A)}{N_n^{(n)}} - \mu_{\xi}(A) \right] \rightarrow 0 \quad \text{in } \mathbb{P}\text{-probability}. \quad (3.10)$$

---

<sup>2</sup>the reader could consider for example the case of an environment containing only two components whose associated transition matrices have different stationary probability.

Thanks to [32] (see in particular Theorem 5.5.), we can check when the Assumption (3.8) is satisfied. It requires that the environment is ergodic, which holds here. As expected, the author also need the uniqueness of the invariant probability and some irreducibility assumptions. He also requires that some  $\delta^*$  is equal to 0, which is much more technical to get. As a simple case where such assumptions hold, the author gives the case of finite state space, which yields the following example.

**Example 3.3. Multitype branching processes in random environment.** In the case when  $X(u)$  belongs to a finite state space  $\mathcal{X}$  for every  $u$ , the process considered here is a multitype branching processes in random environment, where the reproduction law of each individual does not depend on its type, but the offspring distribution does. If the kernel  $Q$  is irreducible, we get the quenched convergence in probability of the proportion of each type. This asymptotic proportion is identified as the stationary measure of the auxiliary chain  $Y$ , and can be thus easily simulated. For the growth rate of the whole population in the case when the reproduction may depend on the type, we refer to [38] for such results under stability assumptions.

To get weaker assumptions (of the Doeblin type) which can be satisfied for Markov chains in stationary random environment, we focus now on limit theorems the whole tree.

### 3.3 Law of large numbers on the whole tree

In the framework of Markov chains with stationary and ergodic environments, quenched ergodic theorems are known (see e.g. [32, 34]). They ensure the convergence (for every  $x \in \mathcal{X}$ ) of

$$\frac{1}{n} \sum_{k=1}^n Q_0 \cdots Q_{k-1}(x, \cdot) \quad \text{as } n \rightarrow \infty.$$

It leads us to consider the following limit theorems on the whole tree, where each generation of the tree has the same mean weight in the limit. Such an approach is both adapted to the branching (forward) genealogy and the convergence of the underlying auxiliary time non-homogeneous Markov chain  $Y$ , whose transition are stationary and ergodic. It defers from the usual limit theorem on the whole tree [21, 15] where each cell has the same weight, but not each generation.

**Theorem 3.5.** *Let  $A \in B_{\mathcal{X}}$ . We assume that there exists  $\mu(A) \in \mathbb{R}$  such that for almost all  $\xi$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{P}_{\xi, x}(Y_k \in A) = \mu(A) \quad \text{for every } x \in \mathcal{X}. \quad (3.11)$$

*Then we have for almost all  $\xi$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \frac{Z_k(A)}{P_k} = \mu(A)W \quad \text{in } \mathbb{P}_{\xi}\text{-}L^2. \quad (3.12)$$

*and, on the non extinction event,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \frac{Z_k(A)}{N_k} = \mu(A) \quad \text{in } \mathbb{P}_{\xi}\text{-probability}. \quad (3.13)$$

The proof is also deferred to next section.

**Example 3.4. Doeblin conditions for the auxiliary Markov chain  $Y$ .** Assume that there exist a positive integer  $b$  and a measurable function  $M(\xi) : \Omega \rightarrow [1, \infty)$  such that  $\mathbb{E}|\log M(\xi)| < \infty$ , and for almost all  $\xi$ ,

$$\mathbb{P}_{\xi,x}(Y_b \in A) \leq M(\xi)\mathbb{P}_{\xi,y}(Y_b \in A) \quad \text{for all } x, y \in \mathcal{X},$$

which means that Assumption (A) of Seppäläinen [34] is satisfied. According to Theorem 2.8 (i) and (iii) of Seppäläinen [34] (with  $f = \mathbf{1}_{A \times \Omega}$ ), there exists a probability  $\Phi$  on  $\mathcal{X} \times \Omega$  such that for almost all  $\xi$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbf{1}_A(Y_k) = \Phi(A \times \Omega) \quad \mathbb{P}_{\xi,x}\text{-a.s. for every } x \in \mathcal{X}.$$

By the dominate convergence theorem, we have for almost all  $\xi$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \mathbb{P}_{\xi,x}(Y_k \in A) = \Phi(A \times \Omega) \quad \text{for every } x \in \mathcal{X}.$$

Thus (3.11) holds with  $\mu(A) = \Phi(A \times \Omega)$ , so we can use Theorem 3.5 to get (3.12) and (3.13).

### 3.4 Central limit theorem

When the auxiliary Markov chain  $Y$  is a classical random walk on  $\mathcal{X} \subset \mathbb{R}$ , we know that  $Y_n$  satisfies a central limit theorem. Such results have been extended to the framework of random walk in random environment (see e.g. [1]) and some more general Markov chains (see e.g. [19]). It leads us to state the convergence of proportions in the case when  $Y_n$  satisfies a central limit theorem.

**Theorem 3.6.** *Let  $\mathcal{X} \subset \mathbb{R}$ . We assume that for almost all  $\xi$ ,  $Y_n$  satisfies a central limit theorem: there exists a sequence of random variables  $\{(a_n(\xi), b_n(\xi))\}$  satisfying  $b_n(\xi) > 0$  such that*

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\xi,x} \left( \frac{Y_n - a_n(\xi)}{b_n(\xi)} \leq y \right) = \Phi(y) \quad \text{for every } x \in \mathcal{X}, \quad (3.14)$$

where  $\Phi$  is a continuous function on  $\mathbb{R}$ . If for each  $r \in \mathbb{N}$  fixed,

$$\lim_{n \rightarrow \infty} \frac{b_n(\xi)}{b_{n-r}(T^r \xi)} = 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{a_n(\xi) - a_{n-r}(T^r \xi)}{b_{n-r}(T^r \xi)} = 0 \quad \text{a.s.}, \quad (3.15)$$

then we have for almost all  $\xi$ ,

$$\frac{Z_n(-\infty, b_n(\xi)y + a_n(\xi))}{P_n} \rightarrow \Phi(y)W \quad \text{in } \mathbb{P}_{\xi}\text{-}L^2, \quad (3.16)$$

and conditionally on the non-extinction event,

$$\frac{Z_n(-\infty, b_n(\xi)y + a_n(\xi))}{N_n} \rightarrow \Phi(y) \quad \text{in } \mathbb{P}_{\xi}\text{-probability.} \quad (3.17)$$

**Example 3.5. Branching random walk on  $\mathbb{R}$  with random environment in time.** This model is considered in Huang & Liu [24]. The environment  $\xi = (\xi_n)_{n \in \mathbb{N}}$  is a stationary and ergodic process indexed by time  $n \in \mathbb{N}$ . Each realization of  $\xi_n$  corresponds to a distribution  $\eta_n = \eta_{\xi_n}$  on  $\mathbb{N} \otimes \mathbb{R}^{\mathbb{N}}$ . Given the environment  $\xi$ , the process is formed as follows: at time  $n$ , each particle  $u$  of generation  $n$ , located at  $X(u) \in \mathbb{R}$ , is replaced by  $N(u)$  new particles of generation  $n+1$  which scattered on  $\mathbb{R}$  with positions determined by  $X(ui) = X(u) + Li(u)$ , where the point

process  $(N(u); L_1(u), L_2(u), \dots)$  has distribution  $\eta_n$ . To fit with the notations of this paper, we can see that

$$\begin{aligned} p_k(\xi_n) &= \eta_n(k, \mathbb{R} \times \mathbb{R} \times \dots), \\ p_{\xi_n}^{(k,i)}(x, y) &= \eta_n(k, \mathbb{R}^{i-1} \times \{y-x\} \times \mathbb{R}^{k-i}) =: q_{\xi_n}^{(k,i)}(y-x), \\ Q_n(x, y) &= \frac{1}{m_n} \sum_{k=0}^{\infty} p_k(\xi_n) \sum_{i=1}^k q_{\xi_n}^{(k,i)}(y-x) =: q_n(y-x). \end{aligned}$$

We note that for any measurable function  $f$  on  $\mathbb{R}$ ,

$$\int f(t) q_n(dt) = \frac{1}{m_n} \mathbb{E}_{\xi} \sum_{i=1}^{N(u)} f(L_i(u)) \quad (u \in \mathbb{T}_n).$$

Hence  $q_n$  is the normalized intensity measure of the point process  $(N(u); L_1(u), L_2(u), \dots)$  for  $u \in \mathbb{T}_n$ . We define

$$Y_n = \zeta_0 + \zeta_1 + \dots + \zeta_n,$$

where  $\zeta_j$  is independent of each other under  $\mathbb{P}_{\xi}$  and the distribution of  $\zeta_j$  for  $j \geq 1$  is  $q_j$ . Then  $Y_n$  is a non-homogeneous Markov chain, whose transition kernel satisfies

$$\mathbb{P}_{\xi}(Y_{n+1} = y | Y_n = x) = q_n(y-x) = Q_n(x, y).$$

Let  $\mu_n = \int_{\mathbb{R}} t q_n(dt)$  and  $\sigma_n^2 = \int_{\mathbb{R}} (t - \mu_n)^2 q_n(dt)$ . If  $|\mu_0| < \infty$  a.s. and  $\mathbb{E}(\sigma_0^2) \in (0, \infty)$ , according to Huang & Liu [24], the sequence  $(q_n)$  satisfies a central limit theorem:

$$q_1 * \dots * q_n(b_n(\xi)y + a_n(\xi)) \rightarrow \Phi(y) \quad a.s.,$$

where

$$a_n(\xi) = \sum_{i=0}^{n-1} \mu_n, \quad b_n(\xi) = \left( \sum_{i=0}^{n-1} \sigma_n^2 \right)^{1/2}$$

and  $\Phi$  is the distribution function of the standard normal distribution. It follows that (3.14) holds for almost all  $\xi$ . Moreover, by the ergodic theorem, (3.15) can be verified. Thus we can apply Theorem 3.6 to this model and obtain (3.17) under the hypothesis given above. This result can also be deduced from [24], where the almost sure convergence of (3.17) is shown though some tedious calculations.

**Example 3.6. Branching random walk on  $\mathbb{Z}$  with random environment in time and in locations.**

This model is considered in Liu [27]. Let  $\xi = (\xi_n)_{n \in \mathbb{N}}$  be a stationary and ergodic process denoting the environment in time, and  $\omega = (\omega_x)_{x \in \mathbb{Z}}$ , which denotes the environment in locations, be another stationary and ergodic process taking values in  $[0, 1]$ . The two sequences  $\xi, \omega$  are supposed to be independent of each other. Given the environment  $(\xi, \omega)$ , each  $u$  of generation  $n$ , located at  $X(u) \in \mathbb{Z}$ , is replaced at time  $n+1$  by  $k$  new particles with probability  $p_k(\xi_n)$ , which move immediately and independently to  $x+1$  with probability  $\omega_x$  and to  $x-1$  with probability  $1 - \omega_x$ . Namely, the position of  $ui$  is determined by

$$\mathbb{P}_{(\xi, \omega)}(X(ui) = y | X(u) = x) = Q(x, y) := \begin{cases} \omega_x & \text{if } y = x + 1; \\ 1 - \omega_x & \text{if } y = x - 1, \end{cases}$$

where  $\mathbb{P}_{(\xi, \omega)}$  denotes the conditional probability given the environment  $(\xi, \omega)$ . Notice that when the environment in locations  $\omega$  is fixed, this process is the just one considered in Example 3.2 with the state space  $\mathcal{X} = \mathbb{Z}$  and  $p(x, y) = Q(x, y)$ . So the transition probability of the Markov chain  $Y_n$  is  $Q$ , which only depends on the environment in locations  $\omega$  and is independent of

the environment in time  $\xi$ . We can regard  $Y_n$  as a random walk on  $\mathbb{Z}$  in random environment which is studied in Alili [1]. By Theorem 6.3 of Alili [1] and the continuity of  $\Phi$ , under some hypothesis, we have for every  $\omega$ ,

$$\lim_{n \rightarrow \infty} \mathbb{P}_{\omega, x} \left( \frac{Y_n - n\gamma}{\sqrt{n}} \leq y \right) = \Phi(y) \quad \text{for every } x \in \mathbb{Z},$$

where  $\Phi$  is the distribution function of the normal distribution  $\mathcal{N}(0, D)$ , and  $\gamma, D$  are two explicit constants (see [1] for details). Therefore, we can apply Theorem 3.6 and obtain (3.16) and (3.17) under the probability  $\mathbb{P}_{(\xi, \omega)}$ .

## 4 Proof of Propositions 2.1 and 2.2

Proposition 2.1 is a consequence of the following result with  $f(x) = \mathbf{1}_A(x) - \mu(A)$ . It is extended hereafter to some class of unbounded functions  $f$ .

**Lemma 4.1.** *Let  $f$  be a bounded measurable function on  $\mathcal{X}$ . We assume that*

- (i)  $N_n \rightarrow \infty$  as  $n \rightarrow \infty$ ;
- (ii)  $\limsup_{n \rightarrow \infty} \mathbb{P}(|U_n \wedge V_n| \geq K) \rightarrow 0$  as  $K \rightarrow \infty$ ;
- (iii) for all  $u \in \mathbb{T}$  and  $x \in \mathcal{X}$ ,

$$\lim_{n \rightarrow \infty} R_u(n, x) = 0,$$

where

$$R_u(n, x) = \mathbb{E} \left[ f(X(U_n^{(u)})) \middle| X(u) = x \right].$$

Then

$$\frac{\sum_{u \in \mathbb{T}_n} f(X(u))}{N_n} \rightarrow 0 \quad \text{in } L^2.$$

*Proof.* We first notice that

$$\mathbb{E} \left( \frac{\sum_{u \in \mathbb{T}_n} f(X(u))}{N_n} \right)^2 = \frac{1}{N_n^2} \mathbb{E} \left[ \sum_{u \in \mathbb{T}_n} f^2(X(u)) \right] + \frac{1}{N_n^2} \mathbb{E} \left[ \sum_{\substack{u, v \in \mathbb{T}_n \\ u \neq v}} f(X(u)) f(X(v)) \right].$$

We need to prove that both terms in the right side go to 0 as  $n \rightarrow \infty$ . For the first term, since  $f$  is bounded, there exists a constant  $C$  such that  $|f| \leq C$ . By (i),

$$\frac{1}{N_n^2} \mathbb{E} \left[ \sum_{u \in \mathbb{T}_n} f^2(X(u)) \right] = \frac{\mathbb{E} f^2(U_n)}{N_n} \leq \frac{C^2}{N_n} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

The second term can be decomposed as

$$\begin{aligned} & \frac{1}{N_n^2} \mathbb{E} \left[ \sum_{\substack{u, v \in \mathbb{T}_n \\ u \neq v}} f(X(u)) f(X(v)) \right] \\ &= \frac{1}{N_n^2} \sum_{r=0}^{n-1} \mathbb{E} \left[ \sum_{\substack{w \in \mathbb{T}_r \\ i \neq j}} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ wi\tilde{u} \in \mathbb{T}_{n-r-1}(wi) \\ wj\tilde{v} \in \mathbb{T}_{n-r-1}(wj)}} f(X(wi\tilde{u})) f(X(wj\tilde{v})) \right] \\ &= \sum_{r=0}^K \mathbb{E} A_{n,r} + \sum_{r=K+1}^{n-1} \mathbb{E} A_{n,r}, \end{aligned}$$

where  $K$  is a fixed integer suitable large, and

$$A_{n,r} = \frac{1}{N_n^2} \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} \sum_{\substack{wi\tilde{u} \in \mathbb{T}_{n-r-1}(wi) \\ wj\tilde{v} \in \mathbb{T}_{n-r-1}(wj)}} f(X(wi\tilde{u}))f(X(wj\tilde{v})).$$

It is clear that

$$\begin{aligned} \mathbb{E}A_{n,r} &= \frac{1}{N_n^2} \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} \sum_{\substack{wi\tilde{u} \in \mathbb{T}_{n-r-1}(wi) \\ wj\tilde{v} \in \mathbb{T}_{n-r-1}(wj)}} \mathbb{E}f(X(wi\tilde{u}))f(X(wj\tilde{v})) \\ &= \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} a_{n,r}(wi, wj)R_{n,r}(wi, wj), \end{aligned}$$

where

$$a_{n,r}(wi, wj) = \frac{N_{n-r-1}(wi)N_{n-r-1}(wj)}{N_n^2}$$

and

$$R_{n,r}(wi, wj) = \mathbb{E}[R_{wi}(n-r-1, X(wi))R_{wj}(n-r-1, X(wj))].$$

As  $|R_u(n, x)| \leq C$  for all  $u, n$  and  $x$ ,

$$\begin{aligned} \sum_{r=K+1}^{n-1} \mathbb{E}A_{n,r} &\leq C^2 \sum_{r=K+1}^{n-1} \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} a_{n,r}(wi, wj) \\ &= C^2 \sum_{r=K+1}^{n-1} \mathbb{P}(|U_n \wedge V_n| = r) \\ &\leq C^2 \mathbb{P}(|U_n \wedge V_n| \geq K+1). \end{aligned}$$

By (ii),  $\limsup_{n \rightarrow \infty} \mathbb{P}(|U_n \wedge V_n| \geq K+1) \rightarrow 0$  as  $K \rightarrow \infty$ . Thus  $\limsup_{n \rightarrow \infty} \sum_{r=K+1}^{n-1} \mathbb{E}A_{n,r}$  is negligible for  $K$  large enough. For  $0 \leq r \leq K$ , the fact that  $R_u(n-r-1, x)$  goes to zero for a.e.  $x$  and is bounded by  $C$  with respect to  $x$  enables us to apply the dominate convergence theorem and get

$$R_{n,r}(wi, wj) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Adding that  $a_{n,r}(wi, wj)$  is bounded by 1 yields

$$\sum_{r=0}^K \mathbb{E}A_{n,r} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

This completes the proof.  $\square$

We give here an extension of the previous result, to get asymptotic results on unbounded functions (such as  $f(x) = x^\alpha$ ).

**Lemma 4.2.** *Let  $f$  be a measurable function on  $\mathcal{X}$ . We assume that*

- (i)  $\mathbb{E}f^2(U_n)/N_n \rightarrow 0$  as  $n \rightarrow \infty$ ;
- (ii) there exists a function  $g$  such that for all  $n$ ,  $u \in \mathbb{T}$  and  $x \in \mathcal{X}$ ,  $|R_u(n, x)| \leq g(x)$ ;
- (iii)  $\mathbb{E}g(X(ui)g(X(uj))) \leq \beta_{|u|}$  for  $ui, uj \in \mathbb{T}_1(u)$  and  $i \neq j$ , and

$$\limsup_{K \rightarrow \infty} \limsup_{n \rightarrow \infty} C_{n,K} = 0,$$

where  $C_{n,K} = \sum_{r=K}^{n-1} \mathbb{P}(|U_n \wedge V_n| = r)\beta_r$ ;

(iv) for all  $u \in \mathbb{T}$  and  $x \in \mathcal{X}$ ,  $\lim_{n \rightarrow \infty} R_u(n, x) = 0$ .

Then

$$\frac{\sum_{u \in \mathbb{T}_n} f(X(u))}{N_n} \rightarrow 0 \quad \text{in } L^2.$$

*Proof.* According to the proof of Proposition 4.1, here we only need to show that  $\limsup_n \sum_{r=K+1}^{n-1} \mathbb{E}A_{n,r} \rightarrow 0$  as  $K \rightarrow \infty$ . By (ii) and (iii), for  $w \in \mathbb{T}_r$ ,

$$|R_{n,r}(wi, wj)| \leq \mathbb{E}g(X(wi)g(X(wj))) \leq \beta_r,$$

so that

$$\sum_{r=K+1}^{n-1} \mathbb{E}A_{n,r} \leq \sum_{r=K+1}^{n-1} \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} a_{n,r}(wi, wj) \beta_r = C_{n,K}.$$

Letting successively  $n$  and  $K$  go to  $\infty$  yields the result.  $\square$

Proposition 2.2 is a result of Lemma 4.3 below, with  $f(x) = \mathbf{1}_A(x) - \mu(A)$ , which also can be extended to a result similar to Lemma 4.2 for unbounded functions  $f$ , but here we omit to state it for technical convenience.

**Lemma 4.3.** *Let  $f$  be a bounded measurable function on  $\mathcal{X}$ . We assume that*

(i)  $N_n \rightarrow \infty$  as  $n \rightarrow \infty$ ;

(ii)  $\limsup_{n \rightarrow \infty} \mathbb{P}(|U_n \wedge V_n| \geq n - K) \rightarrow 0$  as  $K \rightarrow \infty$ ;

(iii)  $\lim_{n \rightarrow \infty} \sup_{u \in \mathbb{T}} |R_u(n, X(u))| = 0$ .

Then

$$\frac{\sum_{u \in \mathbb{T}_n} f(X(u))}{N_n} \rightarrow 0 \quad \text{in } L^2.$$

*Proof.* Similar to the proof of Lemma 4.1, but we split

$$\sum_{r=0}^{n-1} \mathbb{E}A_{n,r} = \sum_{r=0}^{n-K-1} \mathbb{E}A_{n,r} + \sum_{r=n-K}^{n-1} \mathbb{E}A_{n,r},$$

and show the negligibility of the two terms respectively. By (ii), we first have

$$\sum_{r=n-K}^{n-1} \mathbb{E}A_{n,r} \leq C^2 \limsup_{n \rightarrow \infty} \mathbb{P}(|U_n \wedge V_n| \geq n - K) \rightarrow 0 \quad \text{as } K \rightarrow \infty.$$

Now we consider  $\sum_{r=0}^{n-K-1} \mathbb{E}A_{n,r}$ . For  $r \leq n - K - 1$  and  $wi, wj \in \mathbb{T}_1(w)$  ( $w \in \mathbb{T}_r$ ,  $i \neq j$ ),

$$\begin{aligned} |R_{n,r}(wi, wj)| &\leq \mathbb{E}|R_{wi}(n-r-1, X(wi))R_{wj}(n-r-1, X(wj))| \\ &\leq \sup_{k \geq K} \mathbb{E} \left( \sup_{u \in \mathbb{T}} |R_u(k, X(u))|^2 \right). \end{aligned}$$

It follows that

$$\begin{aligned} \sum_{r=0}^{n-K-1} \mathbb{E}A_{n,r} &\leq \sum_{r=0}^{n-K-1} \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} a_{n,r}(wi, wj) \sup_{k \geq K} \mathbb{E} \left( \sup_{u \in \mathbb{T}} |R_u(k, X(u))|^2 \right) \\ &= \mathbb{P}(|U_n \wedge V_n| \leq n - K - 1) \sup_{k \geq K} \mathbb{E} \left( \sup_{u \in \mathbb{T}} |R_u(k, X(u))|^2 \right) \\ &\leq \sup_{k \geq K} \mathbb{E} \left( \sup_{u \in \mathbb{T}} |R_u(k, X(u))|^2 \right). \end{aligned}$$



By (iii) and the dominate convergence theorem, we have

$$\lim_{n \rightarrow \infty} \mathbb{E} \left( \sup_{u \in \mathbb{T}} |R_u(n, X(u))|^2 \right) = 0.$$

Thus  $\sup_n \sum_{r=0}^{n-K-1} \mathbb{E} A_{n,r} \rightarrow 0$  as  $K \rightarrow \infty$ .  $\square$

## 5 Proofs for branching Markov chains in random environments

In this section, we focus on the process in random environments and present proofs for the theorems stated in Section 3.

### 5.1 Many to one formula

**Lemma 5.1.** *For each  $x \in \mathbb{R}$ ,  $u \in \mathbb{T}$  and any measurable function  $f_{\xi,n}$  on  $\mathcal{X}$ ,*

$$\mathbb{E}_{\xi} [f_{\xi,n}(Y_n(u)) \mid Y_0(u) = x] = \frac{\mathbb{E}_{\xi} \left[ \sum_{uv \in \mathbb{T}_n(u)} f_{\xi,n}(X(uv)) \mid X(u) = x \right]}{\mathbb{E}_{\xi} N(u)}. \quad (5.1)$$

*Proof.* By the definition of  $\{Y_n(u)\}$ , it is easy to see that

$$\begin{aligned} & \mathbb{E}_{\xi} [f_{\xi,n}(Y_n(u)) \mid Y_0(u) = x] \\ &= \mathbb{Q}_{|u|} \cdots \mathbb{Q}_{|u|+n-1} f_{\xi,n}(x) \\ &= (m_{|u|} \cdots m_{|u|+n-1})^{-1} \times \\ & \quad \sum_{k_0, \dots, k_{n-1}=0}^{\infty} \sum_{i_0=1}^{k_0} \cdots \sum_{i_{n-1}=1}^{k_{n-1}} p_{k_0}(\xi_{|u|}) \cdots p_{k_{n-1}}(\xi_{|u|+n-1}) P_{\xi_0}^{(k_0, i_0)} \cdots P_{\xi_{n-1}}^{(k_{n-1}, i_{n-1})} f_{\xi,n}(x). \end{aligned}$$

On the other hand, we notice that

$$\begin{aligned} & \mathbb{E}_{\xi, x} \left[ \sum_{ui \in \mathbb{T}_1(u)} f_{\xi,n}(X(ui)) \right] \quad (5.2) \\ &= \mathbb{E}_{\xi} \left[ \sum_{ui \in \mathbb{T}_1(u)} f_{\xi,n}(X(ui)) \mid X(u) = x \right] \\ &= \sum_{k=0}^{\infty} \mathbb{E}_{\xi} \left[ \sum_{i=1}^k f_{\xi,n}(X(ui)) \mid N(u) = k, X(u) = x \right] \mathbb{P}_{\xi}(N(u) = k) \\ &= \sum_{k=0}^{\infty} p_k(\xi_{|u|}) \sum_{i=1}^k P_{\xi_{|u|}}^{(k, i)} f_{\xi,n}(x). \quad (5.3) \end{aligned}$$

Thus,

$$\begin{aligned}
& \mathbb{E}_{\xi,x} \left[ \sum_{uv \in \mathbb{T}_n} f_{\xi,n}(X(uv)) \right] \\
&= \mathbb{E}_{\xi,x} \left[ \sum_{uv \in \mathbb{T}_{n-1}(u)} \sum_{uvi \in \mathbb{T}_1(uv)} f_{\xi,n}(X(uvi)) \right] \\
&= \mathbb{E}_{\xi,x} \left[ \sum_{uv \in \mathbb{T}_{n-1}(u)} \mathbb{E}_{\xi} \left[ \sum_{uvi \in \mathbb{T}_1(uv)} f_{\xi,n}(X(uvi)) \middle| \mathcal{F}_{n-1}, X(uv) \right] \right] \\
&= \mathbb{E}_{\xi,x} \left[ \sum_{uv \in \mathbb{T}_{n-1}(u)} \mathbb{E}_{\xi, X(uv)} \left[ \sum_{uvi \in \mathbb{T}_1(uv)} f_{\xi,n}(X(uvi)) \right] \right] \\
&= \sum_{k=0}^{\infty} \sum_{i=1}^k p_k(\xi_{|u|+n-1}) \mathbb{E}_{\xi,x} \left[ \sum_{uv \in \mathbb{T}_{n-1}(u)} P_{\xi_{|u|+n-1}}^{(k,i)} f_{\xi,n}(X(uv)) \right].
\end{aligned}$$

By iteration, we obtain

$$\begin{aligned}
\mathbb{E}_{\xi,x} \left[ \sum_{uv \in \mathbb{T}_n} f_{\xi,n}(X(uv)) \right] &= \sum_{k_0, \dots, k_{n-1}=0}^{\infty} \sum_{i_0=1}^{k_0} \cdots \sum_{i_{n-1}=1}^{k_{n-1}} p_{k_0}(\xi_{|u|}) \cdots p_{k_{n-1}}(\xi_{|u|+n-1}) \\
&\quad \times P_{\xi_{|u|}}^{(k_0, i_0)} \cdots P_{\xi_{|u|+n-1}}^{(k_{n-1}, i_{n-1})} f_{\xi,n}(x),
\end{aligned}$$

so that

$$\frac{\mathbb{E}_{\xi,x} \left[ \sum_{uv \in \mathbb{T}_n(u)} f_{\xi,n}(X(uv)) \right]}{\mathbb{E}_{\xi} N_n(u)} = \mathbb{E}_{\xi,x} [f_{\xi,n}(Y_n(u))].$$

It ends up the proof.  $\square$

## 5.2 Proof of the law of large numbers in generation $n$

Following the definition of  $Y$  in Section 3.1, for each  $u \in \mathbb{T}$ , we define the Markov chain  $Y(u)$  associated to  $u$  by

$$\mathbb{P}_{\xi}(Y_{j+1}(u) = y | Y_j(u) = x) = Q(T^{|u|+j}\xi; x, y).$$

In particular,  $Y_n = Y_n(\emptyset)$ . Moreover, for any measurable function  $f$  on  $\mathcal{X}$ , we denote throughout

$$P_{\xi_n}^{(k,i)} f(x) := \int f(y) P_{\xi_n}^{(k,i)}(x, dy) \quad \text{and} \quad Q_j f(x) := \int f(y) Q_j(x, dy).$$

First, Lemma 5.1 below reveals the mean relation between  $Y_n(u)$  and the tree  $\mathbb{T}(u)$ , in the same vein as [21, 15, 22].

**Proposition 5.1.** *Let  $\nu_{\xi}$  be the distribution of  $X(\emptyset)$ . We assume that for almost all  $\xi$ , there exist a function  $g$ , an integer  $n_0 = n_0(\xi)$  and non negative numbers  $(\alpha_n, \beta_n) = (\alpha_n(\xi), \beta_n(\xi))$  such that*

(H1) *for all  $n \geq n_0$  and  $x \in \mathcal{X}$ ,*

$$\sup_{0 \leq r < n} Q_r \cdots Q_{n-1} |f_{\xi,n}|(x) \leq g(x).$$

(H2) *for every  $n \geq n_0$ ,*

$$\nu_{\xi} Q_0 \cdots Q_{n-1} f_{\xi,n}^2 \leq \alpha_n, \quad \alpha_n / P_n \rightarrow 0 \quad (n \rightarrow \infty);$$

(H3) for every  $n \in \mathbb{N}$ ,

$$\nu_\xi Q_0 \cdots Q_{n-1} J_{T^n \xi}(g \otimes g) \leq \beta_n, \quad \sum_n \frac{\beta_n}{P_n m_n^2} < \infty,$$

where  $J_\xi(f_\xi \otimes g_\xi)(x) := \mathbb{E}_{\xi, x} \sum_{\substack{i, j \in \mathbb{T}_1 \\ i \neq j}} f_\xi(X(i)) g_\xi(X(j))$ ;

(H4) for each  $r$  fixed,  $Q_r \cdots Q_{n-1} f_{\xi, n}(x) \rightarrow 0$  as  $n \rightarrow \infty$  for every  $x \in \mathcal{X}$ .

Then we have for almost all  $\xi$ ,

$$\frac{\sum_{u \in \mathbb{T}_n} f_{\xi, n}(X(u))}{P_n} \rightarrow 0 \quad \text{in } \mathbb{P}_\xi\text{-}L^2.$$

*Proof.* We need to show that

$$\mathbb{E}_\xi \left( \frac{\sum_{u \in \mathbb{T}_n} f_{\xi, n}(X(u))}{P_n} \right)^2 \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty.$$

Similar to the proof of Proposition 4.1, we write

$$\begin{aligned} \mathbb{E}_\xi \left( \frac{\sum_{u \in \mathbb{T}_n} f_{\xi, n}(X(u))}{P_n} \right)^2 &= \frac{1}{P_n^2} \mathbb{E}_\xi \left[ \sum_{u \in \mathbb{T}_n} f_{\xi, n}^2(X(u)) \right] \\ &\quad + \sum_{r=0}^K \mathbb{E}_\xi A_{n, r} + \sum_{r=K+1}^{n-1} \mathbb{E}_\xi A_{n, r}, \end{aligned} \quad (5.4)$$

where

$$A_{n, r} := \frac{1}{P_n^2} \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} \sum_{wi\tilde{u} \in \mathbb{T}_{n-r-1}(wi)} \sum_{wj\tilde{v} \in \mathbb{T}_{n-r-1}(wj)} f_{\xi, n}(X(wi\tilde{u})) f_{\xi, n}(X(wj\tilde{v})),$$

and  $K = K(\xi) (\geq n_0)$  is suitable large. By Lemma 5.1 and condition (H2), for  $n \geq n_0$ ,

$$\begin{aligned} \frac{1}{P_n^2} \mathbb{E}_\xi \left[ \sum_{u \in \mathbb{T}_n} f_{\xi, n}^2(X(u)) \right] &= \frac{1}{P_n} \mathbb{E}_\xi f_{\xi, n}^2(Y_n) \\ &= \frac{1}{P_n} \nu_\xi Q_0 \cdots Q_{n-1} f_{\xi, n}^2 \\ &\leq \frac{\alpha_n}{P_n} \rightarrow 0 \quad \text{a.s. as } n \rightarrow \infty. \end{aligned}$$

Again by Lemma 5.1, for every  $w \in \mathbb{T}_r$ ,  $i \neq j \in \mathbb{N}$ , denoting  $n_r := n - r - 1$  in the computation below,

$$\begin{aligned} &B(wi, wj) \\ &:= \mathbb{E}_\xi \left[ \sum_{wi\tilde{u} \in \mathbb{T}_{n_r}(wi)} f_{\xi, n}(X(wi\tilde{u})) \sum_{wj\tilde{v} \in \mathbb{T}_{n_r}(wj)} f_{\xi, n}(X(wj\tilde{v})) \middle| \mathcal{F}_{r+1}, X(wi), X(wj) \right] \\ &= \mathbb{E}_{\xi, X(wi)} \left[ \sum_{wi\tilde{u} \in \mathbb{T}_{n_r}(wi)} f_{\xi, n}(X(wi\tilde{u})) \right] \mathbb{E}_{\xi, X(wj)} \left[ \sum_{wj\tilde{v} \in \mathbb{T}_{n_r}(wj)} f_{\xi, n}(X(wj\tilde{v})) \right] \\ &= \mathbb{E}_{\xi, X(wi)} [f_{\xi, n}(Y_{n_r}(wi))] \mathbb{E}_\xi [N_{n_r}(wi)] \mathbb{E}_{\xi, X(wj)} [f_{\xi, n}(Y_{n_r}(wj))] \mathbb{E}_\xi [N_{n_r}(wj)] \\ &= m_{r+1}^2 \cdots m_{n-1}^2 Q_{r+1, n} f_{\xi, n}(X(wi)) Q_{r+1, n} f_{\xi, n}(X(wj)), \end{aligned}$$

with the notation  $Q_{r,n} = Q_r \cdots Q_{n-1}$ . Thus,

$$\begin{aligned}
& \mathbb{E}_\xi A_{n,r} \\
&= \frac{1}{P_n^2} \mathbb{E}_\xi \left[ \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} B(wi, wj) \right] \\
&= \frac{1}{P_{r+1}^2} \mathbb{E}_\xi \left[ \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} Q_{r+1,n} f_{\xi,n}(X(wi)) Q_{r+1,n} f_{\xi,n}(X(wj)) \right] \\
&= \frac{1}{P_r m_r^2} \nu_\xi Q_0 \cdots Q_{r-1} J_{T^r \xi}(Q_{r+1,n} f_{\xi,n} \otimes Q_{r+1,n} f_{\xi,n}).
\end{aligned}$$

For  $r \geq K + 1$ , by conditions (H1) and (H3),

$$\begin{aligned}
\sum_{r=K+1}^{n-1} \mathbb{E}_\xi A_{n,r} &\leq \sum_{r=K+1}^{\infty} \frac{1}{P_r m_r^2} \nu_\xi Q_0 \cdots Q_{r-1} J_{T^r \xi}(g \otimes g) \\
&\leq \sum_{r=K+1}^{\infty} \frac{\beta_r}{P_r m_r^2} \rightarrow 0 \text{ a.s. as } K \rightarrow \infty.
\end{aligned}$$

It remains to consider  $0 \leq r \leq K$ . For almost all  $\xi$ , for each  $r$  fixed, by (H4),

$$Q_{r+1,n} f_{\xi,n} \otimes Q_{r+1,n} f_{\xi,n}(y, z) \xrightarrow{n \rightarrow \infty} 0 \text{ for each } (y, z) \in \mathcal{X}^2.$$

Notice that by (H1) and (H3), for  $n \geq n_0(\xi)$ ,

$$\begin{aligned}
\nu_\xi Q_0 \cdots Q_{r-1} J_{T^r \xi}(Q_{r+1,n} f_{\xi,n} \otimes Q_{r+1,n} f_{\xi,n}) \\
\leq \nu_\xi Q_0 \cdots Q_{r-1} J_{T^r \xi}(g \otimes g) \leq \beta_r \quad \text{a.s.}
\end{aligned}$$

By the dominated convergence theorem,  $\sum_{r=0}^K \mathbb{E}_\xi A_{n,r} \rightarrow 0$  a.s. as  $n \rightarrow \infty$ . The proof is completed.  $\square$

In particular, applying Proposition 5.1 with  $f_{\xi,n} = f_\xi - \pi(f_\xi)$ , we obtain the following result. We denote

$$F_{\xi,n} = \nu_\xi Q_0 \cdots Q_{n-1}.$$

**Proposition 5.2.** *Let  $\nu_\xi$  be the distribution of  $X(\emptyset)$ . We assume that for almost all  $\xi$ , there exist a function  $g$ , an integer  $n_0 = n_0(\xi)$  and non negative numbers  $(\alpha_n, \beta_n) = (\alpha_n(\xi), \beta_n(\xi))$  such that*

(H1) *for all  $n \geq n_0$  and  $x \in \mathcal{X}$ ,*

$$\sup_{0 \leq r < n} Q_r \cdots Q_{n-1} |f_\xi|(x) \leq g(x);$$

(H2) *for every  $n \geq n_0$ ,  $F_{\xi,n}(f_\xi^2) \leq \alpha_n$ , and  $\alpha_n/P_n \rightarrow 0$  ( $n \rightarrow \infty$ ).*

(H3) *for every  $n \in \mathbb{N}$ ,*

$$\max\{F_{\xi,n}(J_{T^n \xi}(g \otimes g)), F_{\xi,n}(J_{T^n \xi}(g \otimes \mathbf{1})), F_{\xi,n}(J_{T^n \xi}(\mathbf{1} \otimes \mathbf{1}))\} \leq \beta_n, \quad \sum_n \frac{\beta_n}{P_n m_n^2} < \infty;$$

(H4) there exists  $\pi(f_\xi) \in \mathbb{R}$  bounded by some constant  $M$  such that for each  $r$  fixed,

$$\lim_{n \rightarrow \infty} Q_r \cdots Q_{n-1} f_\xi(x) = \pi(f_\xi) \quad \text{for every } x \in \mathcal{X}.$$

Under (H1)-(H4), if additionally  $W_n \rightarrow W$  in  $\mathbb{P}_\xi$ - $L^2$ , then we have for almost all  $\xi$ ,

$$\frac{\sum_{u \in \mathbb{T}_n} f_\xi(X(u))}{P_n} \rightarrow \pi(f_\xi) W \quad \text{in } \mathbb{P}_\xi$$
- $L^2$ .

In particular, if  $f_\xi = f$ , then conditions (H1) and (H4) can be simplified to

(H1')  $Q_0 \cdots Q_{n-1} |f|(x) \leq g(x)$  for all  $n \in \mathbb{N}$  and  $x \in \mathcal{X}$ ;

(H4') there exists  $\pi(f) \in \mathbb{R}$  such that  $Q_0 \cdots Q_{n-1} f(x) \rightarrow \pi(f)$  as  $n \rightarrow \infty$  for every  $x \in \mathcal{X}$ .

*Proof of Theorem 3.2.* We apply Proposition 5.1 with  $f_{\xi,n}(x) = \mathbf{1}_A(x) - \mu_{\xi,n}(A)$ . Indeed, for  $n$  large enough,  $f_{\xi,n}$  is bounded by some constant  $M$ . We can let  $g = M$ ,  $\alpha_n = M^2$  and  $\beta_n = \mathbb{E}_{T^n \xi} N^2$ . The assumptions (3.1) ensure that a.s.,

$$\frac{\alpha_n}{P_n} \rightarrow 0, \quad \sum_n \frac{\beta_n}{P_n m_n^2} < \infty \quad \text{and} \quad W_n \rightarrow W \quad \text{in } \mathbb{P}_\xi$$
- $L^2$ .

Then Proposition 5.1 yields (3.3). Moreover  $\mathbb{E}_\xi W = 1$  implies that  $q(\xi) = \mathbb{P}_\xi(W = 0) < 1$ , so  $\{W > 0\} = \{N_n \rightarrow \infty\}$   $\mathbb{P}_\xi$ -a.s.. Finally, (3.4) comes from (3.3) and the fact that  $W_n \rightarrow W > 0$  a.s. on the non-extinction event  $\{N_n \rightarrow \infty\}$ .  $\square$

*Proof of Corollary 3.3.* We apply Proposition 5.2 with  $f(x) = \mathbf{1}_A(x)$ . Take  $g = 1$ ,  $\alpha_n = 1$  and  $\beta_n = \mathbb{E}_{T^n \xi} N^2$ .  $\square$

### 5.3 Adaptation of the proof for backward law of large numbers

*Proof of Corollary 3.4 (3.9).* Let  $f_\xi(x) = \mathbf{1}_A(x) - \mu_\xi(A)$ . Clearly,  $f_\xi$  is bounded by 1. Like the proof of Proposition 5.1, we still have (5.4) with  $f_\xi$  in place of  $f_{\xi,n}$ , but here and throughout this proof,  $\mathbb{T}_k(u)$  ( $|u| + k \leq n$ ) denotes the set of individuals in  $k$ th generation of a tree rooted at  $u$  in the environment  $\xi^{(n)}$ . Firstly, we have a.s.,

$$\frac{1}{P_n^2} \mathbb{E}_\xi \left[ \sum_{u \in \mathbb{T}_n} f_\xi^2(X(u)) \right] \leq \frac{1}{P_n} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

By Lemma 5.1, for  $|u| < n$  and  $k \leq n - |u|$ ,

$$\frac{\mathbb{E}_{\xi,x} \sum_{uv \in \mathbb{T}_k(u)} f_\xi(X(uv))}{\mathbb{E}_{\xi^{(n)}} N_k(u)} = Q_{n-|u|-1} \cdots Q_{n-|u|-k} f_\xi(x). \quad (5.5)$$

Therefore,

$$\mathbb{E}_\xi A_{n,r} = (m_{n-1} \cdots m_{n-r})^{-2} \mathbb{E}_\xi \left[ \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} g_{\xi, n_r}(X(wi)) g_{\xi, n_r}(X(wj)) \right],$$

where  $g_{\xi,n}(x) = Q_{n-1} \cdots Q_0 f_{\xi}(x)$  and  $n_r := n - r - 1$ . For  $r \geq K + 1$ , since  $m_0 > a$  and  $\mathbb{E}_{\xi} N^2 \leq (\mathbb{E}_{\xi} N^p)^{2/p} \leq b^{2/p}$ , we get

$$\begin{aligned} \sum_{r=K+1}^{n-1} \mathbb{E}_{\xi} A_{n,r} &\leq \sum_{r=K+1}^{n-1} (m_{n-1} \cdots m_{n-r} m_{n_r}^2)^{-1} \mathbb{E}_{T^{n_r} \xi} N^2 \\ &\leq \sum_{r=K+1}^{\infty} a^{-(r+2)} b^{2/p} \rightarrow 0 \text{ as } K \rightarrow \infty. \end{aligned}$$

Now consider  $0 \leq r \leq K$ . The fact that  $m_0 > a$  and  $g_{\xi,n}$  is bounded by 1 yields

$$\begin{aligned} \mathbb{E}_{\xi} A_{n,r} &\leq a^{-2r} \mathbb{E}_{\xi} \left[ \sum_{w \in \mathbb{T}_r} N(w) \sum_{wi \in \mathbb{T}_1(w)} |g_{\xi,n_r}(X(wi))| \right] \\ &\leq a^{-2r} \mathbb{E}_{\xi} \left[ N_{r+1}^{(n)} \sum_{w \in \mathbb{T}_{r+1}} |g_{\xi,n_r}(X(w))| \right] \\ &= S_{n,r} + T_{n,r}, \end{aligned}$$

where

$$S_{n,r} = a^{-2r} \mathbb{E}_{\xi} \left[ N_{r+1}^{(n)} \mathbf{1}_{\{N_{r+1}^{(n)} \leq K\}} \sum_{w \in \mathbb{T}_{r+1}} |g_{\xi,n_r}(X(w))| \right]$$

and

$$T_{n,r} = a^{-2r} \mathbb{E}_{\xi} \left[ N_{r+1}^{(n)} \mathbf{1}_{\{N_{r+1}^{(n)} > K\}} \sum_{w \in \mathbb{T}_{r+1}} |g_{\xi,n_r}(X(w))| \right].$$

We first deal with  $S_{n,r}$  and fix  $r \geq 0$ . Let us label the individuals in generation  $r + 1$  as  $1, 2, \dots$ . For  $j > N_{r+1}^{(n)}$ , the value of  $g_{\xi,n_r}(X(j))$  will be 0 by convention. So

$$S_{n,r} \leq K a^{-2r} \sum_{j=1}^K \mathbb{E}_{\xi} |g_{\xi,n_r}(X(j))|.$$

By (3.5), for almost all  $\xi$ ,

$$g_{\xi,n}(x) = \mathbb{P}_{\xi^{(n)},x}(Y_n \in A) - \mu_{\xi}(A) \xrightarrow{n \rightarrow \infty} 0 \text{ for every } x \in \mathcal{X}.$$

By the bounded convergence theorem, we have for each  $r$  fixed,  $\lim_n S_{n,r} = 0$  a.s. On the other hand, following arguments in Huang & Liu [23], the fact that

$$\mathbb{E}_{\xi} \left( \frac{N}{m_0} \right)^p < b/a^p$$

for some  $p > 2$  implies that for all  $n$ ,

$$\sup_{0 \leq r \leq n} \mathbb{E}_{\xi^{(n)}} \left[ W_{r+1}^{(n)} \right]^p \leq C_p$$

for some constant  $C_p$  depending on  $p$ . Thus for each  $r$  fixed and  $\delta \in (0, p - 2)$ ,

$$\begin{aligned} T_{n,r} &\leq a^{-2r} \mathbb{E}_{\xi^{(n)}} \left[ \left( N_{r+1}^{(n)} \right)^2 \mathbf{1}_{\{N_{r+1}^{(n)} > K\}} \right] \\ &= K^{-\delta} a^{-2r} \mathbb{E}_{\xi^{(n)}} \left[ N_{r+1}^{(n)} \right]^{2+\delta} \\ &\leq C_{r,\delta} K^{-\delta} \rightarrow 0 \end{aligned}$$

as  $K \rightarrow \infty$ , where  $C_{r,\delta}$  is a constant depending on  $r$  and  $\delta$ . Therefore, a.s.,  $\lim_n \mathbb{E}_{\xi} A_{n,r} = 0$  for each  $r$  fixed, so that  $\sum_{0 \leq r \leq K} \mathbb{E}_{\xi} A_{n,r} \rightarrow 0$  as  $n \rightarrow \infty$ . So (3.9) is proved.  $\square$

*Proof of Corollary 3.4 (3.10).* Since  $\sup_n \mathbb{E}_{\xi^{(n)}} [W_n^{(n)}]^2 \leq C$  for some constant  $C$ , by the bounded convergence theorem, (3.9) implies that

$$\frac{Z_n^{(n)}(A) - \mu_\xi(A)N_n^{(n)}}{P_n} \rightarrow 0 \text{ in } L^2 \text{ and in } \mathbb{P}\text{-probability.} \quad (5.6)$$

For any  $\varepsilon > 0$  and  $\alpha > 0$ ,

$$\begin{aligned} & \mathbb{P} \left( \left| \frac{Z_n^{(n)}(A)}{N_n^{(n)}} - \mu_\xi(A) \right| > \varepsilon, N_n^{(n)} > 0 \right) \\ & \leq \mathbb{P} \left( \left| \frac{Z_n^{(n)}(A) - \mu_\xi(A)N_n^{(n)}}{P_n} \right| > \alpha\varepsilon \right) + \mathbb{P}(0 < W_n^{(n)} < \alpha). \end{aligned} \quad (5.7)$$

By (5.6), it is clear that the first term in the right side of the inequality above tends to 0 as  $n \rightarrow \infty$ . When the environment is reversible in law, under  $\mathbb{P}$ ,  $W_n^{(n)}$  has the same distribution as  $W_n$  of the forward case. Thus we have

$$\mathbb{P}(0 < W_n^{(n)} < \alpha) = \mathbb{P}(0 < W_n < \alpha).$$

In the forward case, it is known that  $W_n$  tends to a limit  $W$  a.s., and

$$\lim_{n \rightarrow \infty} \mathbb{P}(N_n = 0) = \mathbb{P}(N_n = 0 \text{ for some } n) = \mathbb{P}(W = 0) \text{ a.s.}$$

when  $\mathbb{E} \log m_0 > 0$  and  $\mathbb{E} \frac{N}{m_0} \log^+ N < \infty$ . Therefore we have

$$\limsup_{n \rightarrow \infty} \mathbb{P}(0 < W_n < \alpha) = \mathbb{P}(0 < W \leq \alpha).$$

Since  $\lim_{\alpha \downarrow 0} \mathbb{P}(0 < W \leq \alpha) = 0$ , for any  $\eta > 0$ , there exists  $\alpha > 0$  small enough such that  $\mathbb{P}(0 < W \leq \alpha) \leq \eta$ . Taking the superior limit in (5.7), we obtain

$$\limsup_{n \rightarrow \infty} \mathbb{P} \left( \left| \frac{Z_n^{(n)}(A)}{N_n^{(n)}} - \mu_\xi(A) \right| > \varepsilon, N_n^{(n)} > 0 \right) \leq \mathbb{P}(0 < W \leq \alpha) \leq \eta.$$

Letting  $\eta \rightarrow 0$  completes the proof. □

## 5.4 Proof of the law of large numbers on the whole tree

Proposition 5.3 below is in the same vein as Proposition 5.2. We recall the notation  $F_{\xi,n} = \nu_\xi Q_0 \cdots Q_{n-1}$ .

**Proposition 5.3.** *Let  $\nu_\xi$  be the distribution of  $X(\emptyset)$ . We assume that for almost all  $\xi$ , there exist non negative numbers  $(\alpha_n, \beta_n) = (\alpha_n(\xi), \beta_n(\xi))$  such that*

(H1) *for all  $n \in \mathbb{N}$  and  $x \in \mathcal{X}$ ,*

$$\sup_{0 \leq r < n} Q_r \cdots Q_{n-1} |f_\xi|(x) \leq g(x);$$

(H2) *for every  $n \in \mathbb{N}$ ,*

$$\max\{F_{\xi,n}(f_\xi^2), F_{\xi,n}(|f_\xi|g), F_{\xi,n}(g)\} \leq \alpha_n, \quad \sum_n \frac{\alpha_n}{P_n} < \infty;$$

(H3) for every  $n \in \mathbb{N}$ ,

$$\max\{F_{\xi,n}(J_{T^n\xi}(g \otimes g)), F_{\xi,n}(J_{T^n\xi}(g \otimes \mathbf{1})), F_{\xi,n}(J_{T^n\xi}(\mathbf{1} \otimes \mathbf{1}))\} \leq \beta_n, \quad \sum_n \frac{\beta_n}{P_n m_n^2} < \infty;$$

(H4) there exists  $\pi(f_\xi) \in \mathbb{R}$  bounded by some constant  $M$  such that for each  $r$  fixed,

$$\lim_{n \rightarrow \infty} \frac{1}{n-r} \sum_{k=r+1}^n Q_r \cdots Q_{k-1} f_\xi(x) = \pi(f_\xi) \quad \text{for every } x \in \mathcal{X}.$$

In particular, if  $f_\xi = f$ , then conditions (H1) and (H4) can be simplified to

(H1')  $Q_0 \cdots Q_{n-1} |f|(x) \leq g(x)$  for all  $n \in \mathbb{N}$  and  $x \in \mathcal{X}$ ;

(H4') there exists  $\pi(f) \in \mathbb{R}$  such that  $\frac{1}{n} \sum_{k=1}^n Q_0 \cdots Q_{k-1} f(x) \rightarrow \pi(f)$  as  $n \rightarrow \infty$  for every  $x \in \mathcal{X}$ .

*Proof of Proposition 5.3.* We only prove the case where  $\pi(f_\xi) = 0$ . For general case, it suffices to consider  $f_\xi - \pi(f_\xi)$  in place of  $f_\xi$ . We shall prove that under the hypothesis (H1)-(H4), a.s.,

$$\mathbb{E}_\xi \left( \frac{1}{n} \sum_{u \in \Gamma_n} \frac{f_\xi(X(u))}{P_{|u|}} \right)^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (5.8)$$

Notice that  $\Gamma_n = \bigcup_{k=1}^n \mathbb{T}_k$  is the set of all individuals in the first  $n$  generation. For  $u, v \in \Gamma_n$ , we discuss for two cases: (i)  $u$  and  $v$  in the same life line, which means that one is an ancestor of the other, i.e.  $u \wedge v = u$  or  $v$ ; (ii) the contrary case, i.e.  $u \wedge v \neq u, v$ . So we can write

$$\begin{aligned} \mathbb{E}_\xi \left( \frac{1}{n} \sum_{u \in \Gamma_n} \frac{f_\xi(X(u))}{P_{|u|}} \right)^2 &= \frac{1}{n^2} \mathbb{E}_\xi \left[ \sum_{u, v \in \Gamma_n} \frac{f_\xi(X(u)) f_\xi(X(v))}{P_{|u|} P_{|v|}} \right] \\ &= S_{n,\xi} + T_{n,\xi}, \end{aligned}$$

where

$$S_{n,\xi} = \frac{1}{n^2} \mathbb{E}_\xi \left[ \sum_{\substack{u, v \in \Gamma_n \\ u \wedge v = u \text{ or } v}} \frac{f_\xi(X(u)) f_\xi(X(v))}{P_{|u|} P_{|v|}} \right],$$

and

$$T_{n,\xi} = \frac{1}{n^2} \mathbb{E}_\xi \left[ \sum_{\substack{u, v \in \Gamma_n \\ u \wedge v \neq u, v}} \frac{f_\xi(X(u)) f_\xi(X(v))}{P_{|u|} P_{|v|}} \right].$$

We need to prove that  $S_{n,\xi}, T_{n,\xi}$  tend to 0 a.s. as  $n \rightarrow \infty$ .

At first,  $S_{n,\xi}$  can be decomposed as

$$S_{n,\xi} = \frac{2}{n^2} \mathbb{E}_\xi \left[ \sum_{\substack{u, v \in \Gamma_n \\ u < v}} \frac{f_\xi(X(u)) f_\xi(X(v))}{P_{|u|} P_{|v|}} \right] + \frac{1}{n^2} \mathbb{E}_\xi \left[ \sum_{u \in \Gamma_n} \frac{f_\xi^2(X(u))}{P_{|u|}^2} \right]. \quad (5.9)$$



For the second term in the right side of the equality above, we use (H2) and get

$$\begin{aligned}
\frac{1}{n^2} \mathbb{E}_\xi \left[ \sum_{u \in \Gamma_n} \frac{f_\xi^2(X(u))}{P_{|u|}^2} \right] &= \frac{1}{n^2} \sum_{r=1}^n \mathbb{E}_\xi \left[ \sum_{u \in \mathbb{T}_r} \frac{f_\xi^2(X(u))}{P_{|r|}^2} \right] \\
&= \frac{1}{n^2} \sum_{r=1}^n P_r^{-1} \nu_\xi Q_0 \cdots Q_{r-1} f_\xi^2 \\
&\leq \frac{1}{n^2} \sum_{r=1}^{\infty} P_r^{-1} \alpha_r \rightarrow 0 \text{ as } n \rightarrow \infty,
\end{aligned}$$

since  $\sum_n P_n^{-1} \alpha_n < \infty$  a.s.. For  $u, v \in \Gamma_n$  with  $u < v$ , we write  $v = u\tilde{v}$  with  $|\tilde{v}| = k - r$  if  $u \in \mathbb{T}_r$  and  $v \in \mathbb{T}_k$  ( $k > r$ ). Then, using (H2) and the a.s. convergence of  $\sum_n P_n^{-1} \alpha_n$  again, we have a.s.,

$$\begin{aligned}
&\frac{2}{n^2} \mathbb{E}_\xi \left[ \sum_{\substack{u, v \in \Gamma_n \\ u < v}} \frac{f_\xi(X(u)) f_\xi(X(v))}{P_{|u|} P_{|v|}} \right] \\
&= \frac{2}{n^2} \sum_{r=1}^n \sum_{k=r+1}^n \mathbb{E}_\xi \left[ \sum_{u \in \mathbb{T}_r} \sum_{u\tilde{v} \in \mathbb{T}_{k-r}(u)} \frac{f_\xi(X(u)) f_\xi(X(u\tilde{v}))}{P_{|r|} P_{|k|}} \right] \\
&= \frac{2}{n^2} \sum_{r=1}^n \sum_{k=r+1}^n P_r^{-1} P_k^{-1} \mathbb{E}_\xi \left[ \sum_{u \in \mathbb{T}_r} f_\xi(X(u)) \mathbb{E}_{\xi, X(u)} \left[ \sum_{u\tilde{v} \in \mathbb{T}_{k-r}(u)} f_\xi(X(u\tilde{v})) \right] \right] \\
&= \frac{2}{n^2} \sum_{r=1}^n \sum_{k=r+1}^n P_r^{-2} \mathbb{E}_\xi \left[ \sum_{u \in \mathbb{T}_r} f_\xi(X(u)) Q_r \cdots Q_{k-1} f_\xi(X(u)) \right] \\
&\leq \frac{2(n-r)}{n^2} \sum_{r=1}^n P_r^{-1} \nu_\xi Q_0 \cdots Q_{r-1} |f_\xi| g \\
&\leq \frac{2}{n} \sum_{r=1}^{\infty} P_r^{-1} \alpha_r \rightarrow 0 \text{ as } n \rightarrow \infty.
\end{aligned}$$

Hence we have  $S_{n,\xi} \rightarrow 0$  a.s. as  $n \rightarrow \infty$ .

Now we consider  $T_{n,\xi}$ . For  $K = K(\xi)$  fixed suitable large,

$$T_{n,\xi} = \sum_{r=1}^K \mathbb{E}_\xi A_{n,r} + \sum_{r=K+1}^{n-1} \mathbb{E}_\xi A_{n,r},$$

where

$$A_{n,r} = \frac{1}{n^2} \sum_{k,l=r+1}^n \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} \sum_{\substack{wi\tilde{u} \in \mathbb{T}_{k-r-1}(wi) \\ wj\tilde{v} \in \mathbb{T}_{l-r-1}(wj)}} \frac{f_\xi(X(wi\tilde{u})) f_\xi(X(wj\tilde{v}))}{P_k P_l}.$$

With the notation  $Q_{r,n} = Q_r \cdots Q_{n-1}$ , we compute

$$\begin{aligned}
& \mathbb{E}_\xi A_{n,r} \\
&= \frac{1}{n^2} \sum_{k,l=r+1}^n (P_k P_l)^{-1} \mathbb{E}_\xi \left[ \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} \mathbb{E}_{\xi, X(wi)} \left[ \sum_{wi\tilde{u} \in \mathbb{T}_{k-r-1}(wi)} f_\xi(X(wi\tilde{u})) \right] \right. \\
&\quad \left. \times \mathbb{E}_{\xi, X(wj)} \left[ \sum_{wj\tilde{v} \in \mathbb{T}_{l-r-1}(wj)} f_\xi(X(wj\tilde{v})) \right] \right] \\
&= \frac{1}{n^2} \sum_{k,l=r+1}^n P_{r+1}^{-2} \mathbb{E}_\xi \left[ \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} Q_{r+1,k} f_\xi(X(wi)) Q_{r+1,l} f_\xi(X(wj)) \right] \\
&= \frac{(n-r)^2}{n^2} P_{r+1}^{-2} \mathbb{E}_\xi \left[ \sum_{w \in \mathbb{T}_r} \sum_{\substack{wi, wj \in \mathbb{T}_1(w) \\ i \neq j}} R_{n,r}(X(wi)) R_{n,r}(X(wj)) \right] \\
&= \frac{(n-r)^2}{n^2} \frac{1}{P_r m_r^2} \nu_\xi Q_0 \cdots Q_{r-1} J_{T^{r+1}\xi}(R_{r,n} \otimes R_{r,n}),
\end{aligned}$$

where

$$R_{n,r}(x) = \frac{1}{n-r} \sum_{k=r+1}^n Q_{r+1,k} f_\xi(x).$$

we observe that thanks to (H1),  $\sup_{0 \leq r < n} |R_{r,n}| \leq g$  for every  $n$ . And by (H4), for almost all  $\xi$ , for each  $r$  fixed,

$$R_{n,r} \otimes R_{n,r}(y, z) \xrightarrow{n \rightarrow \infty} 0 \text{ for each } (y, z) \in \mathcal{X}^2.$$

Following similar arguments in the proof of Proposition 5.1, we see  $T_{n,\xi} \rightarrow 0$  a.s. as  $n \rightarrow \infty$ . The proof is completed.  $\square$

**Lemma 5.2.** *Let  $p > 1$ . If  $\mathbb{E}(\log m_0) > 0$  and  $\mathbb{E} \left( \log \mathbb{E}_\xi \left( \frac{N}{m_0} \right)^p \right) < \infty$ , then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n W_k = W \quad \text{in } \mathbb{P}_\xi\text{-}L^p.$$

*Proof.* By Theorems 2.1 and 2.2 in Huang & Liu [23], the integrability assumptions ensure that a.s.,

$$0 < \mathbb{E}_\xi W^p < \infty \quad \text{and} \quad \lim_{k \rightarrow \infty} \rho^k \mathbb{E}_\xi |W_k - W|^p = 0,$$

for some  $\rho > 1$ . It implies that

$$\mathbb{E}_\xi |W_k - W|^p \leq C_\xi \rho^k \text{ a.s.} \tag{5.10}$$

for some  $C_\xi < \infty$ . Therefore, a.s.,

$$\begin{aligned}
\mathbb{E}_\xi \left| \frac{1}{n} \sum_{k=1}^n W_k - W \right|^p &= \mathbb{E}_\xi \left| \frac{1}{n} \sum_{k=1}^n (W_k - W) \right|^p \\
&\leq \frac{1}{n} \sum_{k=1}^n \mathbb{E}_\xi |W_k - W|^p \\
&\leq C_\xi \frac{1}{n} \sum_{k=1}^{\infty} \rho^{-k} \rightarrow 0 \text{ as } n \rightarrow \infty,
\end{aligned}$$

since  $\sum_{k=1}^{\infty} \rho^{-k} < \infty$ . □

*Proof of Theorem 3.5.* By Lemma 5.2,  $\frac{1}{n} \sum_{k=1}^n W_k \rightarrow W$  in  $\mathbb{P}_{\xi}\text{-}L^2$ . Applying Proposition 5.3 with  $f(x) = \mathbf{1}_A(x)$ ,  $g = 1$ ,  $\alpha_n = 1$  and  $\beta_n = \mathbb{E}_{T^n \xi} N^2$ , we obtain (3.12).

Now we prove (3.13). Since  $W_n \rightarrow W$  a.s., for any  $\delta > 0$ , there exists  $n_0$  such that  $\forall n \geq n_0$ ,

$$|W_n - W| < \delta \quad \text{and} \quad \left| \frac{1}{W_n} - \frac{1}{W} \right| < \delta \quad (5.11)$$

a.s. on the non-extinction event. We write

$$\begin{aligned} & \frac{1}{n} \sum_{k=1}^n \frac{Z_k(A)}{N_k} - \mu(A) \\ &= \frac{1}{n} \sum_{k=1}^{n_0} \frac{Z_k(A)}{P_k} \left( \frac{1}{W_k} - \frac{1}{W} \right) + \frac{1}{n} \sum_{k=n_0+1}^n \frac{Z_k(A)}{P_k} \left( \frac{1}{W_k} - \frac{1}{W} \right) \\ & \quad + \frac{1}{W} \left( \frac{1}{n} \sum_{k=1}^n \frac{Z_k(A)}{P_k} - \mu(A)W \right). \end{aligned} \quad (5.12)$$

Obviously, the first term in the right hand side of (5.12) tends to 0 a.s. on the non-extinction event as  $n$  goes to infinity. And the convergence of the third term is from (3.12). Therefore, to prove (3.13), we only need to show that the second term

$$\frac{1}{n} \sum_{k=n_0+1}^n \frac{Z_k(A)}{P_k} \left( \frac{1}{W_k} - \frac{1}{W} \right) \xrightarrow{n \rightarrow \infty} 0 \quad \text{in } \mathbb{P}_{\xi}\text{-probability} \quad (5.13)$$

on the non-extinction event for almost all  $\xi$ . In fact, by (5.11), we have

$$\left| \frac{1}{n} \sum_{k=n_0+1}^n \frac{Z_k(A)}{P_k} \left( \frac{1}{W_k} - \frac{1}{W} \right) \right| < \delta(W + \delta)$$

a.s. on the non-extinction event. The arbitrariness of  $\delta$  yields (5.13). □

## 5.5 Proof of the central limit theorem

*Proof of Theorem 3.6.* We shall apply Proposition 5.1 with  $f_{\xi,n}(x) = \mathbf{1}_{A_n}(x) - \Phi(y)$ , where  $A_n = (-\infty, b_n(\xi)y + a_n(\xi)]$ . By (3.14) and Dini's Theorem, we have a.s.,

$$\lim_{n \rightarrow \infty} \sup_{y \in \mathbb{R}} \left| \mathbb{P}_{\xi,x} \left( \frac{Y_n - a_n(\xi)}{b_n(\xi)} \leq y \right) - \Phi(y) \right| = 0 \quad \text{for every } x \in \mathcal{X}. \quad (5.14)$$

Notice that  $|f_{\xi,n}| \leq 1$ . Take  $g = 1$ ,  $\alpha_n = 1$  and  $\beta_n = \mathbb{E}_{T^n \xi} N^2$ . It is easy to verify that (H1)-(H3) are satisfied. For (H4), by (5.14) and the continuity of  $\Phi$ , for each  $r$  fixed,

$$\begin{aligned} & Q_r \cdots Q_{n-1} f_{\xi,n}(x) \\ &= \mathbb{P}_{T^r \xi, x} (Y_{n-r} \leq b_n(\xi)y + a_n(\xi)) - \Phi(y) \\ &\leq \sup_{y \in \mathbb{R}} \left| \mathbb{P}_{T^r \xi, x} \left( \frac{Y_{n-r} - a_{n-r}(T^r \xi)}{b_{n-r}(T^r \xi)} \leq y \right) - \Phi(y) \right| \\ & \quad + \left| \Phi \left( \frac{b_n(\xi)y + a_n(\xi) - a_{n-r}(T^r \xi)}{b_{n-r}(T^r \xi)} \right) - \Phi(y) \right|, \end{aligned}$$

which goes to 0 as  $n \rightarrow \infty$  for every  $x \in \mathcal{X}$ . By Proposition 5.1, a.s.,

$$\frac{Z_n(A_n)}{P_n} - \Phi(y) \frac{N_n}{P_n} \rightarrow 0 \quad \text{in } \mathbb{P}_{\xi}\text{-}L^2.$$

Since  $W_n \rightarrow W$  a.s. and in  $\mathbb{P}_{\xi}\text{-}L^2$ , we get (3.16) and (3.17). □

**Acknowledgement.** This work was partially funded by Chair Modelisation Mathématique et Biodiversité VEOLIA-Ecole Polytechnique-MNHN-F.X., the professorial chair Jean Marjoulet, the project MANEGE ‘Modèles Aléatoires en Écologie, Génétique et Évolution’ 09-BLAN-0215 of ANR (French national research agency) and the National Natural Science Foundation of China, Grant No. 11101039.

## References

- [1] Alili S. (1999). Asymptotic behaviour for random walk in random environments. *J. Appl. Proba.* **36**(2), 334-349.
- [2] Athreya K.B., Karlin S. (1971). On branching processes with random environments, I : extinction probability *Ann. Math. Stat.* **42**, 1499-1520.
- [3] Athreya K.B., Karlin S. (1971). On branching processes with random environments, II : limit theorems. *Ann. Math. Stat.* **42**, 1843-1858.
- [4] Athreya K.B., Kang H.J. (1998). Some limit theorems for positive recurrent branching Markov chains I, II. *Adv. Appl. Prob.* **30** (3), 693-722.
- [5] Benjamini I., Peres Y. (1994). Markov chains indexed by trees. *Ann. Probab.* **22** (1), 219-243.
- [6] Bansaye V. (2008). Proliferating parasites in dividing cells : Kimmel’s branching model revisited. *Ann. Appl. Probab.* **18** (3), 967-996.
- [7] Bansaye V., Tran V.C. (2011). Branching feller diffusion for cell division with parasite infection. *ALEA* **8**, 95-127.
- [8] Bansaye V., Delmas J-F., Marsalle L., Tran C. T (2011). Limit theorems for Markov processes indexed by supercritical Galton Watson tree. *Ann. Appl. Probab.* Vol. 21, No. 6, 2263-2314.
- [9] Bercu B., De Saporta B., Gégout-Petit A. (2009). Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic Journal of Probability* **14** (87), 2492-2526.
- [10] Biggins J.D. (1977). Martingale convergence in the branching random walk. *J. Appl. Probab.* **14**, 25-37.
- [11] Biggins J.D. (1977). Chernoff’s theorem in the branching random walk. *J. Appl. Probab.* **14**, 630-636.
- [12] Biggins J.D. (1990). The central limit theorem for the supercritical branching random walk, and related results. *Stoch. Proc. Appl.* **34**, 255-274.
- [13] Comets F., Popov S. (2007). Shape and local growth for multidimensional branching random walks in random environment. *ALEA* **3**, 273-299.
- [14] Comets F., Yoshida N. (2011). Branching random walks in space-time random environment: survival probability, global and local growth rates. *Journal of Theoretical Probability* **24** (3), pp 657-687.
- [15] Delmas J-F, Marsalle L. (2010). Detection of cellular aging in a Galton-Watson process. *Stoch. Proc. Appl.* **120**, 2495-2519.

- [16] Doumic M., Hoffmann. M, Krell N., Robert L. (2012). Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Avialable via <http://arxiv.org/abs/1210.3240>*.
- [17] Evans S.N., Steinsaltz D. (2007). Damage segregation at fissioning may increase growth rates : a superprocess model. *Theoretical Population Biology* 71:4, pp. 473-90.
- [18] Fréville H., Colas B., Riba M., Caswell H., Mignot A., Imbert E., Olivieri I. (2004). Spatial and demographic variability in the endemic plant species *Centaurea Corymbosa* (Asteraceae). *Ecology*, **85** (3), 694-703.
- [19] Gallardo L. (1999). A central limit theorem for Markov chains and applications to hypergroups. *Proceedings of the American Mathematical Society*, **127** (6), 1837-1845.
- [20] Gantert N., Müller S., Popov S., Vachkovskaia M. (2010). Survival of branching random walks in random environment. *Journal of Theoretical Probability*, **23**, 1002-1014.
- [21] Guyon J. (2005). Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* **17**, 1538–1569 (2007).
- [22] Harris S. C., Roberts M. (2012). The many-to-few lemma and multiple spines. *Avialable via <http://arxiv.org/abs/1106.4761>*.
- [23] Huang C., Liu Q. (2010). Convergence in  $L^p$  and its exponential rate for a branching process in a random environment. *Avialable via <http://arxiv.org/abs/1011.0533>*.
- [24] Huang C., Liu Q. (2011). Branching random walk with a random environment in time. *Preprint*.
- [25] Kimmel M. (1997). Quasistationarity in a branching model of division-within-division. *Classical and modern branching processes, IMA Vol. Math. Appl.* **84**, 157-164.
- [26] Lartillot N., Poujol R. (2010). A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.* **28** (1), 729–744.
- [27] Liu Q. (2007). Branching random walks in random environment. *Proceedings of the 4th International Congress of Chinese Mathematicians, 2007 (ICCM 2007)*, Vol. II, 702-719. Eds: L. Ji, K. Liu, L. Yang, S.-T. Yau.
- [28] Nakashima M. (2011). Almost sure central limit theorem for branching random walks in random environment. *Ann. of Appl. Probab.*, **21**(1), 351-373.
- [29] Meyn S.P., Tweedie R.L. (1992). Stability of Markovian processes I: criteria for discrete-time chains. *Adv. Appl. Probab.* **24** (3), 542-574.
- [30] Müller S (2008). A criterion for transience of multidimensional branching random walk in random environment. *Electr. Jour. Probab.* **13**, 1189-1202.
- [31] Mukhamedov F. (2011). On  $L_1$ -weak ergodicity of nonhomogeneous discrete Markov processes and its applications. *Avialable via <http://arxiv.org/abs/1105.0478>*.
- [32] Orey S. (1991). Markov chains with stationary Markov transition probabilities. *Ann. Prob.* **19** (3), 907-928.
- [33] Pan H. (2010). Law of large number for even-odd Markov chain fields and a three times circulation Markov chain indexed by a tree. *PhD*, <http://globethesis.com/?t=2120360302966543>.

- [34] Seppäläinen T. (1994). Large deviations for Markov chains with Random Transitions. *Ann. Prob.* **22** (2), 713-748.
- [35] Sinclair D-A., Guarente L. (1997). Extrachromosomal rDNA circles– a cause of aging in yeast. *Cell* **91** (7), 1033-1042.
- [36] Stadler T. (2011). Inferring speciation and extinction processes from extant species data. *PNAS*. **108** (39), 16145-16146.
- [37] Stewart E.J., Madden R., Paul G., Taddei F. (2005). Aging and death in a organism that reproduces by morphologically symmetric division. *PLoS Biol*, **3** (2) : e45.
- [38] Tanny D. (1981). On multitype branching processes in a random environment. *Adv. Appl. Prob.* **13** ,464-497.
- [39] Yoshida N. (2008). Central limit theorem for random walk in random environment. *Ann. Appl. Probab.* **18** (4), 1619-1635.