

## Some limit results for Markov chains indexed by trees\*

Peter Czappon<sup>†</sup>      Peter Pfaffelhuber\*

### Abstract

We consider a sequence of Markov chains  $(\mathcal{X}^n)_{n=1,2,\dots}$  with  $\mathcal{X}^n = (X_\sigma^n)_{\sigma \in \mathcal{T}}$ , indexed by the full binary tree  $\mathcal{T} = \mathcal{T}_0 \cup \mathcal{T}_1 \cup \dots$ , where  $\mathcal{T}_k$  is the  $k$ -th generation of  $\mathcal{T}$ . In addition, let  $(\Sigma_k)_{k=0,1,2,\dots}$  be a random walk on  $\mathcal{T}$  with  $\Sigma_k \in \mathcal{T}_k$  and  $\tilde{\mathcal{R}}^n = (\tilde{R}_t^n)_{t \geq 0}$  with  $\tilde{R}_t^n := X_{\Sigma_{[tn]}}$ , arising by observing the Markov chain  $\mathcal{X}^n$  along the random walk. We present a law of large numbers concerning the empirical measure process  $\tilde{\mathcal{Z}}^n = (\tilde{Z}_t^n)_{t \geq 0}$  where  $\tilde{Z}_t^n = \sum_{\sigma \in \mathcal{T}_{[tn]}} \delta_{X_\sigma^n}$  as  $n \rightarrow \infty$ . Precisely, we show that if  $\tilde{\mathcal{R}}^n \xrightarrow{n \rightarrow \infty} \mathcal{R}$  for some Feller process  $\mathcal{R} = (R_t)_{t \geq 0}$  with deterministic initial condition, then  $\tilde{\mathcal{Z}}^n \xrightarrow{n \rightarrow \infty} \mathcal{Z}$  with  $Z_t = \delta_{\mathcal{L}(R_t)}$ .

**Keywords:** Tree-indexed Markov chain; weak convergence; tightness; random measure; empirical measure.

**AMS MSC 2010:** AMS Subject classification: 60F15; 60F05.

Submitted to ECP on June 14, 2014, final version accepted on October 31, 2014.

Supersedes arXiv:1406.3768.

## 1 Introduction

In [2], Benjamini and Peres introduced the notion of a tree-indexed Markov chain. Since then, a lot of effort has been spent in studying weak and strong laws of large numbers for very general types of and even possibly random trees [3, 5, 6, 9, 12, 13, 18, 19, 20].

Our work is motivated by an observation in microbiology, where a population of bacteria is growing (along a binary tree, say), and every individual bacterial cell is in a certain *state* (e.g. some gene expression profile), which can be – at least partially – inherited. It has been observed for a long time that such populations tend to be heterogeneous although all cells carry the same genome; see [16] for an early reference.

The question which has arisen is about the mechanisms which are responsible for such phenotypic heterogeneity. Two competing views exist: either, random fluctuations lead to heterogeneity [8, 14] or social interactions of cells together with a regulatory mechanism are key drivers for heterogeneity [15, 17]. Several examples are today known to fall in one of the two categories; see the review of Avery [1].

In this manuscript, we analyse one consequence of the first view, i.e. a law of large numbers. This results entails that the dynamics of single cells can be stochastic while the behavior of the whole population becomes deterministic. We will define a Markov kernel dependent on some scaling parameter  $n$  (which will tend to infinity) and look at

---

\*This research was supported by the DFG through grant Pf-672/5-1.

<sup>†</sup>University of Freiburg, Germany. E-mail: pfaffelhuber@stochastik.uni-freiburg.de

the empirical measure process in the  $[nt]$ -th generation of the population,  $t \geq 0$ , which corresponds to a time-scaling of the process of empirical measures. We will prove the weak convergence of the empirical measure process, which will be a deterministic limit (if the initial distribution is deterministic).

After presenting the general setup in Section 2, we present our main result in Theorem 3.3 in Section 3, together with two simple examples. Then, we give the proof of Theorem 3.3 in Section 4.

## 2 Setup

Let

$$\mathcal{T} = \bigcup_{k=0}^{\infty} \mathcal{T}_k, \quad \mathcal{T}_0 = \{\emptyset\}, \quad \mathcal{T}_k = \{0, 1\}^k, k = 1, 2, \dots$$

be a *complete binary tree*, rooted at  $\emptyset \in \mathcal{T}_0$ , where  $\sigma 0, \sigma 1 \in \mathcal{T}_{k+1}$  are the two children of  $\sigma \in \mathcal{T}_k$ ,  $k = 0, 1, 2, \dots$ . For  $\sigma \in \mathcal{T}_k$  and  $j \leq k$ , we denote by  $\pi_j \sigma$  the prefix of  $\sigma$  of length  $j$ . On  $\mathcal{T}$ , we set  $|\sigma| = k$  iff  $\sigma \in \mathcal{T}_k$  and in addition, set  $\pi_{-1} \sigma := \pi_{|\sigma|-1} \sigma$ , the immediate ancestor of  $\sigma$ . Define the  $\leq$ -relation by writing

$$\sigma \leq \tau \quad \text{iff} \quad \text{there is } j \text{ such that } \pi_j \tau = \sigma$$

and

$$\tau \wedge \tau' := \sup\{\sigma : \sigma \leq \tau, \sigma \leq \tau'\}$$

as the *most recent common ancestor* of  $\tau$  and  $\tau'$ .

Let  $(E, r)$  be a complete and separable metric space, and denote by  $\mathcal{B}(E)$  the Borel- $\sigma$ -field, or the set of bounded measurable functions (with an abuse of notations). A stochastic process  $\mathcal{X} = (X_\sigma)_{\sigma \in \mathcal{T}}$  is called a *time-homogeneous, tree-indexed Markov chain* (extending a notion introduced in [2]), if there is a Markov transition kernel  $p$  from  $E$  to  $\mathcal{B}(E^2)$  (the Borel- $\sigma$ -field on  $E^2$ ) such that for all  $\sigma \in \mathcal{T}$  and  $A_0, A_1 \in \mathcal{B}(E)$ ,

$$\begin{aligned} \mathbf{P}(X_{\sigma 0} \in A_0, X_{\sigma 1} \in A_1 \mid X_\sigma = x_\sigma, X_\tau = x_\tau \text{ for } \tau \in \mathcal{T} \text{ with } \tau \wedge \sigma < \sigma) \\ = \mathbf{P}(X_{\sigma 0} \in A_0, X_{\sigma 1} \in A_1 \mid X_\sigma = x_\sigma) = p(x_\sigma, A_0 \times A_1). \end{aligned}$$

With  $\mathcal{X}$ , we connect the Markov chain  $\mathcal{R} = (R_n)_{n=0,1,2,\dots}$ , with transition kernel

$$p_{\mathcal{R}}(x, A) := \frac{1}{2}(p(x, A \times E) + p(x, E \times A)).$$

Here,  $\mathcal{R}$  arises from observing the state of  $\mathcal{X}$  when walking along  $\mathcal{T}$  starting from the root from  $\sigma$  to  $\sigma 0$  and  $\sigma 1$  purely at random. Another representation of  $\mathcal{R}$  is as follows: Let  $(\Sigma_k)_{k=0,1,2,\dots}$  be a symmetric forward random walk on  $\mathcal{T}$  (independent of  $\mathcal{X}$ ), i.e.  $\Sigma_k \in \mathcal{T}_k$  almost surely and  $\mathbf{P}(\Sigma_{k+1} = \sigma 0 \mid \Sigma_k = \sigma) = \mathbf{P}(\Sigma_{k+1} = \sigma 1 \mid \Sigma_k = \sigma) = \frac{1}{2}$ . Then,  $\mathcal{R} \stackrel{d}{=} (X_{\Sigma_k})_{k=0,1,2,\dots}$ .

If  $(X_\sigma)_{\sigma \in \mathcal{T}}$  is a (time-homogeneous) Markov chain, we then define the *process of empirical measures*  $\mathcal{Z} = (Z_k)_{k=0,1,2,\dots}$  through

$$Z_k := 2^{-k} \sum_{\sigma \in \mathcal{T}_k} \delta_{X_\sigma}.$$

Note that  $\mathcal{Z}$  takes values in  $\mathcal{P}(E)$ , the set of probability measures on  $\mathcal{B}(E)$  and that  $\mathcal{Z}$  is a non-homogeneous Markov chain (indexed by  $k = 0, 1, 2, \dots$ ).

**Remark 2.1** (Symmetric, tree-indexed Markov chains). The idea to consider different transition mechanisms to the two different children comes from the work of [9]. A special, classical case is that of a symmetric tree-indexed Markov chain as follows:

We call a time-homogeneous, (tree-indexed) Markov chain with transition kernel  $p$  (from  $E$  to  $\mathcal{B}(E^2)$ ) *symmetric*, if there is a Markov transition kernel  $q$  (from  $E$  to  $\mathcal{B}(E)$ ) such that for all  $x \in E$ ,  $A_0, A_1 \in \mathcal{B}(E)$

$$p(x, A_0 \times A_1) = q(x, A_0) \cdot q(x, A_1).$$

In other words, the transitions from  $X_\sigma$  to  $X_{\sigma_0}$  and to  $X_{\sigma_1}$  are independent. In this case, we have that  $R_k \stackrel{d}{=} X_\sigma$  for all  $\sigma \in \Sigma_k$ .

In the next section, we will deal with a sequence  $(\mathcal{X}^n)_{n=1,2,\dots}$  of tree-indexed Markov chains.

### 3 Results

Now, we state our main limit theorem for the setup given in the last section. Therefore, let  $(\mathcal{X}^n)_{n=1,2,\dots}$  be a sequence of tree-indexed Markov chains with complete and separable metric state spaces  $(E^n, r^n)_{n=1,2,\dots}$ . As a limiting state space, we have a complete separable metric space  $(E, r)$  and Borel-measurable maps  $\eta^n : E^n \rightarrow E$ .

Let  $\mathcal{R}^n$  be the process of observing  $\mathcal{X}^n$  when moving randomly along the tree. We denote the corresponding transition kernel by  $p_n$  (for  $\mathcal{X}^n$ ) and  $p_{\mathcal{R}^n}$ , respectively. Moreover, let  $\mathcal{Z}^n$  be the process of empirical measures based on  $\mathcal{X}^n$ , which has state space  $\mathcal{P}(E^n)$ ,  $n = 1, 2, \dots$ . Our goal is to find sufficient conditions for  $\mathcal{X}^n$  (via  $\mathcal{R}^n$ ), such that the process of empirical measures  $\mathcal{Z}^n$  converges, and to characterize the limit process. We first recall some basic notation.

**Remark 3.1** (Notation). Throughout the manuscript, we will consider a complete and separable metric space  $(E, r)$ . The space of (continuous,) real-valued, bounded functions on  $E$  are denoted by  $\mathcal{B}(E)(\mathcal{C}_b(E))$ . Weak convergence is denoted by  $\Rightarrow$ . If  $f : [0, \infty) \rightarrow E_1$  and  $\eta : E_1 \rightarrow E_2$ , we write, abusing notation,  $\eta \circ f = \eta(f)$  for the function  $\eta \circ f : t \mapsto \eta(f(t))$ . If  $(E_1, r_1), (E_2, r_2)$  are two metric spaces,  $\eta : E_1 \rightarrow E_2$  is measurable, and  $\nu \in \mathcal{P}(E_1)$ , we define the image measure of  $\nu$  under  $\eta$  by  $\eta_*\nu \in \mathcal{P}(E_2)$ , i.e.  $\eta_*\nu(A_2) = \nu(\eta^{-1}(A_2))$ . Sometimes, we write  $\langle z, \varphi \rangle := \int \varphi dz$  for  $z \in \mathcal{P}(E)$  and  $\varphi \in \mathcal{B}(E)$ . For  $f \in \mathcal{C}_b(E)$  we write  $\|f\| := \sup_{x \in E} |f(x)|$ .

We need two more notions.

**Definition 3.2** (Feller property, compact containment condition). Recall that  $(E, r)$  is complete and separable.

1. A Markov process  $\mathcal{X} = (X_t)_{t \geq 0}$  with state space  $E$  and càdlàg paths satisfies the Feller property, iff (i)  $X_t \xrightarrow{t \rightarrow 0} X_0$  and (ii) the map  $x \mapsto \mathbf{E}[f(X_t) | X_0 = x]$  is continuous for all  $f \in \mathcal{C}_b(E)$ ,  $t \geq 0$  and all  $x \in E$ . Equivalently, let  $(S_t)_{t \geq 0}$  be the semigroup of  $\mathcal{X}$ , i.e.  $S_t f(x) = \mathbf{E}[f(X_t) | X_0 = x]$ . Then,  $\mathcal{X}$  is a Feller-process iff  $(S_t)_{t \geq 0}$  is a Feller semigroup.
2. A Feller semigroup is a family of operators  $(S_t)_{t \geq 0}$ ,  $S_t : \mathcal{C}_b(E) \rightarrow \mathcal{C}_b(E)$  with (i)  $S_t f(x) \xrightarrow{t \rightarrow 0} f(x)$  for all  $x \in E$  and  $f \in \mathcal{C}_b(E)$  and (ii)  $S_t f \in \mathcal{C}_b(E)$  if  $f \in \mathcal{C}_b(E)$ . We say that (iii)  $(S_t)_{t \geq 0}$  is a contraction iff  $\|S_t f\| \leq \|f\|$  and (iv)  $(S_t)_{t \geq 0}$  is strongly continuous iff  $\|S_t f - f\| \xrightarrow{t \rightarrow 0} 0$ . We say that an operator  $G : \mathcal{D}(G) \subseteq \mathcal{C}_b(E) \rightarrow \mathcal{C}_b(E)$  generates a strongly continuous semigroup  $(S_t)_{t \geq 0}$  if

$$Gf(x) := \lim_{t \rightarrow 0} \frac{1}{t} (S_t f(x) - f(x)) \tag{3.1}$$

for all  $f \in C_b(E)$  for which the limit in (3.1) exists.

Recall that for the generator  $G$  of a Feller semigroup, the domain  $\mathcal{D}(G)$  is dense in  $C_b(E)$  (with respect to the topology of uniform convergence on compacta) ([7, Corollary 1.1.6]), and if  $(E, r)$  is locally compact, every Feller semigroup is a strongly continuous contraction semigroup ([11, Theorem 17.6]) and is uniquely characterized by its generator ([11, Lemma 17.5])

3. For a sequence  $(X_t^1)_{t \geq 0}, (X_t^2)_{t \geq 0}, \dots$  of  $E$ -valued stochastic processes, we say that the compact containment condition (in  $E$ ) holds, if for every  $\varepsilon > 0$  and  $T \geq 0$  there is a compact set  $K_{\varepsilon, T} \subseteq E$  such that

$$\sup_{n=1,2,\dots} \mathbf{P}(X_t^n \in K_{\varepsilon, T}^c \text{ for all } 0 \leq t \leq T) < \varepsilon.$$

Now we can formulate our main result.

**Theorem 3.3** (Convergence of  $\mathcal{Z}^n$ ). *Let  $\mathcal{X}^n, \mathcal{R}^n, \mathcal{Z}^n$  be as above,  $n = 1, 2, \dots$ . Moreover, let  $\tilde{\mathcal{R}}^n := (\tilde{R}_t^n)_{t \geq 0} := (R_{[nt]}^n)_{t \geq 0}$ , and  $\tilde{\mathcal{Z}}^n := (\tilde{Z}_t^n)_{t \geq 0} := (Z_{[nt]}^n)_{t \geq 0}$ ,  $n = 1, 2, \dots$ . Assume that  $\eta^n(X_0^n) \xrightarrow{n \rightarrow \infty} \nu \in \mathcal{P}(E)$  and that the compact containment condition holds for  $\eta^1(\tilde{\mathcal{R}}^1), \eta^2(\tilde{\mathcal{R}}^1), \dots$*

*In addition, assume that there is a linear operator  $G_{\mathcal{R}} : \mathcal{D}(G_{\mathcal{R}}) \subseteq C_b(E) \rightarrow C_b(E)$ , which generates a strongly continuous contraction semigroup, and such that  $\mathcal{D}(G_{\mathcal{R}})$  contains an algebra  $\Pi$  that separates points. For each  $\varphi \in \mathcal{D}(G_{\mathcal{R}})$ , there is a sequence  $\varphi_1 \in \mathcal{B}(E^1), \varphi_2 \in \mathcal{B}(E^2), \dots$  such that  $\sup_{n=1,2,\dots} \|\varphi_n\| < \infty$  and*

$$\lim_{n \rightarrow \infty} \sup_{x \in E^n} |\varphi \circ \eta^n(x) - \varphi_n(x)| = 0, \tag{3.2}$$

$$\lim_{n \rightarrow \infty} \sup_{x \in E^n} |(G_{\mathcal{R}}\varphi) \circ \eta^n(x) - G_{\tilde{\mathcal{R}}^n} \varphi_n(x)| = 0, \tag{3.3}$$

where

$$G_{\tilde{\mathcal{R}}^n} \varphi(x) := n \cdot \mathbf{E}[\varphi(\tilde{R}_{1/n}^n) - \varphi(\tilde{R}_0^n) \mid \tilde{R}_0^n = x].$$

Then, there is an  $E$ -valued Feller process  $\mathcal{R} = (R_t)_{t \geq 0}$  with  $R_0 \sim \nu$  and generator  $G_{\mathcal{R}}$  with  $\eta^n(\tilde{\mathcal{R}}^n) \xrightarrow{n \rightarrow \infty} \mathcal{R}$ , and a  $\mathcal{P}(E)$ -valued stochastic process  $\mathcal{Z} = (Z_t)_{t \geq 0}$  such that  $\eta_*^n \tilde{\mathcal{Z}}^n \xrightarrow{n \rightarrow \infty} \mathcal{Z}$  with  $Z_0 \sim \delta_{R_0} \in \mathcal{P}(\mathcal{P}(E))$ . Moreover, if  $\nu = \delta_x$  for  $x \in E$ , then  $Z_t = \delta_{\mathcal{L}(R_t)}$ .

**Remark 3.4** (Convergence, Deterministic limit, CLT). 1. Actually, the convergence  $\tilde{\mathcal{R}}^n \circ \eta^n \xrightarrow{n \rightarrow \infty} \mathcal{R}$  was shown in [7, Corollary 4.8.9], under the assumptions given above.

2. As the Theorem shows, the limiting process of empirical measures  $\mathcal{Z}$  is deterministic (if the initial distribution is a Dirac-measure). The heuristics behind this result is that two distinct values  $X_\sigma^n, X_\tau^n$  with  $\sigma, \tau \in \mathcal{T}_{[nt]}$  have already evolved independently for  $O(n)$  steps. Hence,  $\tilde{Z}_t^n$  is approximately given by the empirical measure of  $2^{[nt]}$  independent processes, which leads to a deterministic limit. This argument will be made precise below.
3. Having obtained a law of large numbers, it would be interesting to see a central limit theorem, as well. In the present context, this would require a fine analysis of the error terms  $\varepsilon_n$  appearing in (4.7). We devote this study to future research.

We now give two simple examples for normal and Poisson convergence.

**Example 3.5.** 1. Let  $(Y_\sigma)_{\sigma \in \mathcal{T}}$  be a family of independent, identically distributed random real-valued variables with  $\mathbf{E}[Y_\sigma] = 0, \mathbf{Var}[Y_\sigma] = 1$ . Moreover, let  $X_0^n := 0$  and  $(X_{\sigma_0}^n, X_{\sigma_1}^n) := (X_\sigma^n + \frac{1}{\sqrt{n}} Y_\sigma, X_\sigma^n - \frac{1}{\sqrt{n}} Y_\sigma)$ . (In other words, the states of the two children of  $\sigma$  are a pair of dependent random variables.) Then, the process  $\mathcal{R}^n = (R_t^n)_{t=0,1,2,\dots}$  can be written as  $R_t^n \stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{k=0}^{t-1} \tilde{Y}_k$ , where  $(\tilde{Y}_k)_{k=0,1,2,\dots}$  are

independent and identically distributed with  $(\tilde{Y}_k)_* \mathbf{P} = \frac{1}{2}(Y_\sigma)_* \mathbf{P} + \frac{1}{2}(-Y_\sigma)_* \mathbf{P}$ , a mixture of the distributions of  $Y_\sigma$  and  $-Y_\sigma$ . Donsker's Theorem yields the convergence  $\tilde{\mathcal{R}}^n \xrightarrow{n \rightarrow \infty} \mathcal{B}$  to the standard Brownian motion  $\mathcal{B}$ . Our theorem now says that the limiting process  $\mathcal{Z}$  is the law of  $\mathcal{B}$ , so we find that  $\mathcal{Z} = (N(0, t))_{t \geq 0}$ , where  $N(0, t)$  is the normal distribution with mean 0 and variance  $t$ .

- Let  $(Y_\sigma^n)_{\sigma \in \mathcal{T}}$  be a family of independent, identically distributed random variables with values in  $\mathbb{Z}_+$  and  $\mathbf{P}(Y_\sigma^n > 0) = 2\lambda/n + o(1/n)$ ,  $\mathbf{P}[Y_\sigma^n > 1] = o(1/n)$ . Moreover, let  $X_0^n := 0$  and  $(X_{\sigma_0}^n, X_{\sigma_1}^n) := (X_\sigma, X_\sigma + Y_\sigma^n)$ . (In other words, the state of the left child equals the state of its parent while the state of the right child has a small probability of having increased by 1.) Then, the process  $\mathcal{R}^n = (R_t^n)_{t=0,1,2,\dots}$  can be written as  $R_t^n \stackrel{d}{=} \sum_{k=0}^{t-1} \tilde{Y}_k^n$ , where  $(\tilde{Y}_k^n)_{k=0,1,2,\dots}$  are independent and identically distributed with  $(\tilde{Y}_k^n)_* \mathbf{P} = \frac{1}{2}\delta_0 + \frac{1}{2}(Y_\sigma^n)_* \mathbf{P}$ , i.e.  $\mathbf{P}[\tilde{Y}_k^n > 0] = \lambda/n + o(1/n)$ ,  $\mathbf{P}[\tilde{Y}_k^n > 1] = o(1/n)$ . Classical convergence results (see e.g. [11, Theorem 5.7]) then show that  $\tilde{\mathcal{R}}^n$  converges weakly to a Poisson process with rate  $\lambda$ . Consequently, we then have by the above theorem that  $\mathcal{Z} = (Z_t)_{t \geq 0}$  with  $Z_t = \text{Poi}(\lambda t)$ .

### 4 Proof of Theorem 3.3

Throughout this section, we build on the same assumptions as in Theorem 3.3. We will replace  $\eta^n(\tilde{\mathcal{R}}^n)$  by  $\tilde{\mathcal{R}}^n$  and  $\eta_*^n \tilde{\mathcal{Z}}^n$  by  $\tilde{\mathcal{Z}}^n$  in the sequel (and similarly for the processes without  $\sim$ ). This should not cause confusion and increase readability.

Before we start, we give basic relationships between the processes  $\tilde{\mathcal{R}}^n$  and  $\tilde{\mathcal{Z}}^n$ , which we will frequently use. (Some more refined relationships will be given in the proof of Lemma 4.2. Let  $\varphi \in \mathcal{C}_b(E)$ . Then,

$$\begin{aligned} \mathbf{E}[\langle \tilde{\mathcal{Z}}_t^n, \varphi \rangle] &= \mathbf{E}[\langle Z_{[nt]}^n, \varphi \rangle] = \mathbf{E}\left[\frac{1}{2^{[nt]}} \sum_{\sigma \in \mathcal{T}_{[nt]}} \langle \delta_{X_\sigma^n}, \varphi \rangle\right] = \mathbf{E}\left[\frac{1}{2^{[nt]}} \sum_{\sigma \in \mathcal{T}_{[nt]}} \varphi(X_\sigma^n)\right] \\ &= \mathbf{E}[\varphi(R_{[nt]}^n)] = \mathbf{E}[\varphi(\tilde{R}_t^n)]. \end{aligned} \tag{4.1}$$

Similarly, we write

$$\langle Z_k^n, \varphi \rangle = 2^{-k} \sum_{\sigma \in \mathcal{T}_k} \varphi(X_\sigma^n) = \mathbf{E}[\varphi(R_k^n) | Z_k^n], \tag{4.2}$$

and  $\langle Z_k^n, \varphi \rangle = \mathbf{E}[\varphi(R_k^n) | Z_k^n, Z_{k-1}^n]$  holds by the same argument, such that

$$\mathbf{E}[\langle Z_k^n, \varphi \rangle | Z_{k-1}^n] = \mathbf{E}[\mathbf{E}[\varphi(R_k^n) | Z_k^n, Z_{k-1}^n] | Z_{k-1}^n] = \mathbf{E}[\varphi(R_k^n) | Z_{k-1}^n]. \tag{4.3}$$

In the proof of Theorem 3.3, it suffices to assume that  $\nu = \delta_x$ , i.e. deterministic starting conditions. (The general case then follows by mixing over the initial condition.) We need to show two assertions:

- The sequence  $(\tilde{\mathcal{Z}}^n)_{n=1,2,\dots}$  is tight.
- The finite-dimensional distributions of  $(\tilde{\mathcal{Z}}^n)_{n=1,2,\dots}$  converge, such that  $\tilde{\mathcal{Z}}_t^n \xrightarrow{n \rightarrow \infty} \delta_{\mathcal{L}(R_t)}$ .

For 2., we will show in Lemma 4.2 that  $\tilde{\mathcal{Z}}_t^n \xrightarrow{n \rightarrow \infty} \delta_{\mathcal{L}(R_t)}$  holds for all  $t \geq 0$ . Since the right hand side is deterministic, we have already shown convergence of finite dimensional distribution and we are left with showing 1. Here, we use Jakubowski's tightness criterion, which is recalled in Proposition A.3 in the appendix. For this criterion, we have to show that (i)  $\tilde{\mathcal{Z}}^n$  satisfies the compact containment condition in  $\mathcal{P}(E)$  (see Definition 3.2) and (ii) that the sequence  $(\langle \tilde{\mathcal{Z}}_t^1, \varphi \rangle)_{t \geq 0}, (\langle \tilde{\mathcal{Z}}_t^2, \varphi \rangle)_{t \geq 0}, \dots$  is tight for all

$\varphi \in \Pi'$  (a vector space which separates points). (i) will be resolved in Lemma 4.3, while (ii) is a result in Lemma 4.4. Hence, we are done once we have shown Lemmas 4.2, 4.3 and 4.4.

We start with a fundamental fact, which is based on the fact that two random leaves from  $\mathcal{T}_n$  have a most recent common ancestor node which is close to the root.

Recall that by [7, Corollary 4.8.9] we already have that  $\tilde{\mathcal{R}}^n \xrightarrow{n \rightarrow \infty} \mathcal{R}$  for a Feller- (hence càdlàg-)process  $\mathcal{R}$ .

**Lemma 4.1** (Convergence to the initial state). *Assume that  $\tilde{\mathcal{R}}^n \xrightarrow{n \rightarrow \infty} \mathcal{R}$  for a càdlàg-process  $\mathcal{R} = (R_t)_{t \geq 0}$  with state space  $E$ . Then, the following holds:*

1. Let  $\sigma_1, \dots, \sigma_k \in \mathcal{T}$  be fixed. Then,

$$(X_{\sigma_i}^n)_{i=1, \dots, k} \xrightarrow{n \rightarrow \infty} (R_0)_{i=1, \dots, k}$$

in probability.

2. Let  $\Sigma_1^n, \Sigma_2^n$  be two vertices, chosen uniformly at random from  $\mathcal{T}_{[nt]}$ . Then,

$$(X_{\Sigma_1^n \wedge \Sigma_2^n}^n, X_{(\Sigma_1^n \wedge \Sigma_2^n)_0}^n, X_{(\Sigma_1^n \wedge \Sigma_2^n)_1}^n) \xrightarrow{n \rightarrow \infty} (R_0, R_0, R_0)$$

in probability.

*Proof.* Recall that for the (independent) random walk  $(\Sigma_k)_{k=0,1,\dots}$  on  $\mathcal{T}$  we have that  $R_k^n = X_{\Sigma_k}^n$ . It suffices to prove the result for deterministic  $R_0 \in E$ . By assumption, for all  $m \in \mathbb{N}$ ,

$$\mathbf{P}(r(R_m^n, R_0) > \varepsilon) = \mathbf{P}(r(\tilde{R}_{m/n}^n, R_0) > \varepsilon) \xrightarrow{n \rightarrow \infty} 0, \tag{4.4}$$

since  $\mathcal{R}$  has càdlàg paths.

1. Let  $\sigma \in \mathcal{T}$  and  $|\sigma| = m$ . We have for all  $\varepsilon > 0$  that

$$\mathbf{P}(r(R_m^n, X_\sigma^n) \leq \varepsilon) \geq \mathbf{P}(r(R_m^n, X_\sigma^n) \leq \varepsilon, R_m^n = X_\sigma^n) = \mathbf{P}(R_m^n = X_\sigma^n) \geq 2^{-m}$$

for all  $\varepsilon > 0$ , since the random walk  $(\Sigma_m)_{m=0,1,2,\dots}$  along we read off  $R^n$  has a chance of  $2^{-m}$  to pass through vertex  $\sigma$ . Assume that the assertion does not hold, i.e.  $X_\sigma^n$  does not converge weakly to  $R_0$ . Let  $\varepsilon > 0$  such that  $\mathbf{P}(r(X_\sigma^n, R_0) > \varepsilon) > \varepsilon$  for all  $n$ . Then, for  $\varepsilon > 0$  as above

$$\begin{aligned} \mathbf{P}(r(R_m^n, R_0) > \varepsilon) &\geq \mathbf{P}(r(R_m^n, R_0) > \varepsilon, R_m^n = X_\sigma^n) \geq \mathbf{P}(r(X_\sigma^n, R_0) > \varepsilon, \Sigma_m = \sigma) \\ &= \mathbf{P}(r(X_\sigma^n, R_0) > \varepsilon) \cdot \mathbf{P}(\Sigma_m = \sigma) \geq \varepsilon 2^{-m} \end{aligned}$$

in contradiction to (4.4). Hence, 1. follows.

2. Let  $\varepsilon > 0$  and  $m$  be large enough for  $2^{-m} < 2\varepsilon$ . From 1., we have that  $(X_\sigma^n)_{\sigma \in \mathcal{T}_m} \xrightarrow{n \rightarrow \infty} (R_0)_{\sigma \in \mathcal{T}_m}$ . Moreover, for  $n > m$ ,  $\mathbf{P}(\Sigma_1^n \wedge \Sigma_2^n \in \mathcal{T}_m) = \sum_{k=0}^m 2^{-(k+1)} = 1 - 2^{-(m+1)} > 1 - \varepsilon$ . Hence, we can write

$$\begin{aligned} \mathbf{P}(r(X_{\Sigma_1^n \wedge \Sigma_2^n}^n, R_0) > \varepsilon) &\leq \mathbf{P}(r(X_{\Sigma_1^n \wedge \Sigma_2^n}^n, R_0) > \varepsilon, \Sigma_1^n \wedge \Sigma_2^n \in \mathcal{T}_m) + \mathbf{P}(\Sigma_1^n \wedge \Sigma_2^n \notin \mathcal{T}_m) \\ &\leq \mathbf{P}(\sup_{\sigma \in \mathcal{T}_m} r(X_\sigma^n, R_0) > \varepsilon) + \mathbf{P}(\Sigma_1^n \wedge \Sigma_2^n \notin \mathcal{T}_m) \\ &\xrightarrow{n \rightarrow \infty} 2^{-(m+1)} < \varepsilon \end{aligned}$$

by 1. and we have shown that  $X_{\Sigma_1^n \wedge \Sigma_2^n}^n \xrightarrow{n \rightarrow \infty} R_0$  in probability. By the same arguments, we also find that  $X_{(\Sigma_1^n \wedge \Sigma_2^n)_i}^n \xrightarrow{n \rightarrow \infty} R_0$  in probability for  $i = 0, 1$  and we are done.  $\square$

**Lemma 4.2** (Convergence of  $\tilde{Z}^n$  at fixed times). *Consider the same situation as in Theorem 3.3 and let  $t \geq 0$ . If  $\nu = \delta_x$  for some  $x \in E$ , we have that  $\tilde{Z}_t^n \xrightarrow{n \rightarrow \infty} \delta_{\mathcal{L}(R_t)}$ .*

*Proof.* Note that the assertion holds once we show that

$$\langle Z_t^n, \varphi \rangle \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi(R_t)] \tag{4.5}$$

for all  $\varphi \in \mathcal{C}_b(E)$ . (Indeed, the family  $(\langle Z_t^n, \varphi \rangle)_{n=1,2,\dots}$  is tight by the boundedness of  $\varphi$  and any subsequent limit point is deterministic by Lemma A.2.) For this, we already know from (4.1) that  $\mathbf{E}[\langle Z_t^n, \varphi \rangle] = \mathbf{E}[\varphi(\tilde{R}_t^n)] \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi(R_t)]$ . Further we will show that

$$\mathbf{Var}[\langle Z_t^n, \varphi \rangle] \xrightarrow{n \rightarrow \infty} 0 \tag{4.6}$$

which then implies (4.5). For this, consider two randomly picked vertices  $\Sigma_1, \Sigma_2 \in \mathcal{T}_{[nt]}$  with  $\Sigma_1 \neq \Sigma_2$ . Then, without loss of generality we assume that  $\pi_{|\Sigma_1 \wedge \Sigma_2|+1} X_{\Sigma_1} = X_{(\Sigma_1 \wedge \Sigma_2)0}$  and  $\pi_{|\Sigma_1 \wedge \Sigma_2|+1} X_{\Sigma_2} = X_{(\Sigma_1 \wedge \Sigma_2)1}$  such that (setting  $\delta_n := 2^{-[nt]} (\mathbf{E}[\varphi^2(X_{\Sigma_1}^n) - \varphi(X_{\Sigma_1}^n)\varphi(X_{\Sigma_2}^n)])$ )

$$\begin{aligned} \mathbf{E}[\langle \tilde{Z}_t^n, \varphi \rangle^2] &= \frac{1}{2^{2[nt]}} \sum_{\sigma_1, \sigma_2 \in \mathcal{T}_{[nt]}} \mathbf{E}[\varphi(X_{\sigma_1}^n)\varphi(X_{\sigma_2}^n)] \\ &= \mathbf{E}[\varphi(X_{\Sigma_1}^n)\varphi(X_{\Sigma_2}^n)] + \delta_n \\ &= \mathbf{E}[\mathbf{E}[\varphi(X_{\Sigma_1}^n)\varphi(X_{\Sigma_2}^n) \mid X_{(\Sigma_1^n \wedge \Sigma_2^n)0}^n, X_{(\Sigma_1^n \wedge \Sigma_2^n)1}^n]] + \delta_n \\ &= \mathbf{E}[\mathbf{E}[\varphi(X_{\Sigma_1}^n) \mid X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_1}^n] \cdot \mathbf{E}[\varphi(X_{\Sigma_2}^n) \mid X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_2}^n]] + \delta_n \\ &= \mathbf{E}[\mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_1}^n] \\ &\quad \cdot \mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_2}^n]] + \delta_n \\ &= \mathbf{E}[\varphi(\tilde{R}_t^n)]^2 + \varepsilon_t^n = \mathbf{E}[\langle \tilde{Z}_t^n, \varphi \rangle]^2 + \varepsilon_t^n \end{aligned}$$

for

$$\begin{aligned} \varepsilon_t^n &:= \mathbf{E}[\mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) - \mathbf{E}[\varphi(R_{[nt]}^n)] \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_1}^n] \\ &\quad \cdot \mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) - \mathbf{E}[\varphi(R_{[nt]}^n)] \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_2}^n]] \\ &\quad + \mathbf{E}[\varphi(R_{[nt]}^n)] \cdot \mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) - \mathbf{E}[\varphi(R_{[nt]}^n)] \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_1}^n]] \\ &\quad + \mathbf{E}[\varphi(R_{[nt]}^n)] \cdot \mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) - \mathbf{E}[\varphi(R_{[nt]}^n)] \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_2}^n]] + \delta_n \\ &= \mathbf{COV}[\mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_1}^n], \\ &\quad \mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_2}^n]] \\ &\quad + 2 \cdot \mathbf{E}[\varphi(R_{[nt]}^n)] \cdot \mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) - \mathbf{E}[\varphi(R_{[nt]}^n)] \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_2}^n]] + \delta_n. \end{aligned} \tag{4.7}$$

Hence, we must show  $\varepsilon_t^n \xrightarrow{n \rightarrow \infty} 0$  for (4.6), which is implied by the boundedness of  $\varphi$  (showing convergence to 0 of the last term of (4.7)), by the Cauchy-Schwartz inequality once we show that

$$\mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) \mid R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_1}^n] \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi(R_t) \mid R_0 = x] \tag{4.8}$$

in probability. Indeed, we already know from Lemma 4.1 that  $X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}\Sigma_1}^n \xrightarrow{n \rightarrow \infty} x$  in probability, such that, since  $\mathcal{R}$  has càdlàg paths, convergence of semigroups [7, Theorem

1.6.1 (see also Remark 4.8.8)] and the strong continuity of the semigroup for  $\mathcal{R}$ ,

$$\begin{aligned} & |\mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) | R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}^n \Sigma_1}^n] - \mathbf{E}[\varphi(R_t) | R_0 = x]| \\ & \leq |\mathbf{E}[\varphi(R_{[nt]-|\Sigma_1 \wedge \Sigma_2|-1}^n) | R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}^n \Sigma_1}^n] - \mathbf{E}[\varphi(R_{[nt]}) | R_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}^n \Sigma_1}^n]| \\ & \quad + |\mathbf{E}[\varphi(\tilde{R}_t^n) | \tilde{R}_0^n = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}^n \Sigma_1}^n] - \mathbf{E}[\varphi(R_t) | R_0 = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}^n \Sigma_1}^n]| \\ & \quad + |\mathbf{E}[\varphi(R_t) | R_0 = X_{\pi_{|\Sigma_1^n \wedge \Sigma_2^n|+1}^n \Sigma_1}^n] - \mathbf{E}[\varphi(R_t) | R_0 = x]| \\ & \xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

in probability, which shows (4.8). This completes the proof.  $\square$

Now, we come to the proof of the compact containment condition for  $(\tilde{Z}^n)_{n=1,2,\dots}$ .

**Lemma 4.3** (Compact containment condition for  $\tilde{\mathcal{R}}^n$ ). *If  $(\tilde{\mathcal{R}}^n)_{n=1,2,\dots}$  satisfies the compact containment condition (in  $E$ ), then  $(\tilde{Z}^n)_{n=1,2,\dots}$  satisfies the compact containment condition (in  $\mathcal{P}(E)$ ) as well.*

*Proof.* For all  $\delta > 0$ , let  $K_\delta \subseteq E$  be compact and such that

$$\sup_{n=1,2,\dots} \mathbf{P}(\tilde{R}_t^n \notin K_\delta \text{ for some } 0 \leq t \leq T) < \delta.$$

This is the same as saying that for the set  $\mathcal{T}_n$  of  $\{0, \frac{1}{n}, \dots, \frac{[Tn]}{n}\}$ -valued random times,

$$\sup_{n=1,2,\dots} \sup_{\tau_n \in \mathcal{T}_n} \mathbf{P}(\tilde{R}_{\tau_n}^n \notin K_\delta) < \delta.$$

For  $\delta, \varepsilon > 0$ , set

$$L_\delta := \{\mu \in \mathcal{P}(E) : \mu(K_{\delta^2}^c) < \delta\}, \quad L := \bigcap_{n=1}^{\infty} L_{\varepsilon 2^{-n}}.$$

Then, the closure of  $L$  is a compact subset of  $\mathcal{P}(E)$  by Prohorov's Theorem. Moreover,

$$\begin{aligned} & \sup_{n=1,2,\dots} \sup_{\tau_n \in \mathcal{T}_n} \mathbf{P}(\tilde{Z}_{\tau_n}^n(K_{\varepsilon 2^{4-\ell}}^c) > \varepsilon 2^{-\ell}) \leq \sup_{n=1,2,\dots} \sup_{\tau_n \in \mathcal{T}_n} \frac{1}{\varepsilon 2^{-\ell}} \mathbf{E}[\langle \tilde{Z}_{\tau_n}^n, 1_{K_{\varepsilon 2^{4-\ell}}^c} \rangle] \\ & = \sup_{n=1,2,\dots} \sup_{\tau_n \in \mathcal{T}_n} \frac{1}{\varepsilon 2^{-\ell}} \mathbf{E}[1_{K_{\varepsilon 2^{4-\ell}}^c}(\tilde{R}_{\tau_n}^n)] = \sup_{n=1,2,\dots} \sup_{\tau_n \in \mathcal{T}_n} \frac{1}{\varepsilon 2^{-\ell}} \mathbf{P}(\tilde{R}_{\tau_n}^n \notin K_{\varepsilon 2^{4-\ell}}) < \varepsilon 2^{-\ell}, \end{aligned}$$

hence,

$$\begin{aligned} \sup_{n=1,2,\dots} \mathbf{P}(\tilde{Z}_t^n \notin L \text{ for some } 0 \leq t \leq T) &= \sup_{n=1,2,\dots} \sup_{\tau_n \in \mathcal{T}_n} \mathbf{P}(\tilde{Z}_{\tau_n}^n \notin L) \\ &\leq \sup_{n=1,2,\dots} \sup_{\tau_n \in \mathcal{T}_n} \sum_{\ell=1}^{\infty} \mathbf{P}(\tilde{Z}_{\tau_n}^n(K_{\varepsilon 2^{4-\ell}}^c) > \varepsilon 2^{-\ell}) < \varepsilon. \end{aligned}$$

$\square$

**Lemma 4.4** (Martingale convergence). *Consider the same situation as in Theorem 3.3 with  $\nu = \delta_x$  for some  $x \in E$ . Let  $\varphi \in \Pi$  and  $\varphi_n \in \mathcal{B}(E)$  such that  $\|\varphi_n - \varphi\| \xrightarrow{n \rightarrow \infty} 0$ . For  $f_n(z) := \langle z, \varphi_n \rangle$ , consider the mean-zero martingale  $\mathcal{M}^{n,\varphi_n} = (M_t^{n,\varphi_n})_{t \geq 0}$ , given by*

$$M_t^{n,\varphi_n} := f_n(Z_{[nt]}^n) - f_n(Z_0^n) - \sum_{k=1}^{[nt]} \mathbf{E}[f_n(Z_k^n) - f_n(Z_{k-1}^n) | Z_{k-1}^n].$$

Then,  $\mathcal{M}^{n,\varphi_n} \xrightarrow{n \rightarrow \infty} 0$  and  $f_n(\tilde{Z}^n) \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi(\mathcal{R})]$ . In particular,  $(f_n(\tilde{Z}^n))_{n=1,2,\dots}$  is tight.

*Proof.* We start by reformulating, using (4.2),

$$\begin{aligned}
 M_t^{n,\varphi_n} &= \langle Z_{[nt]}^n, \varphi_n \rangle - \langle Z_0^n, \varphi_n \rangle - \sum_{k=1}^{[nt]} \mathbf{E}[\langle Z_k^n, \varphi_n \rangle - \langle Z_{k-1}^n, \varphi_n \rangle \mid Z_{k-1}^n] \\
 &= \mathbf{E}[\varphi_n(R_{[nt]}^n) \mid Z_{[nt]}^n] - \mathbf{E}[\varphi_n(R_0^n) \mid Z_0^n] - \int_0^{t-1/n} n \cdot \mathbf{E}[\mathbf{E}[\varphi_n(R_{[ns]+1}^n) - \varphi_n(R_{[ns]}^n) \mid Z_{[ns]}^n]] ds \\
 &= \mathbf{E}[\varphi_n(\tilde{R}_t^n) \mid \tilde{Z}_t^n] - \mathbf{E}[\varphi_n(\tilde{R}_0^n) \mid \tilde{Z}_0^n] - \int_0^{t-1/n} n \cdot \mathbf{E}[\varphi_n(\tilde{R}_{s+1/n}^n) - \varphi_n(\tilde{R}_s^n) \mid \tilde{Z}_s^n] ds. \quad (4.9)
 \end{aligned}$$

We now show that  $M_t^{n,\varphi_n} \xrightarrow{n \rightarrow \infty} 0$ . From Lemma 4.2, we already know that  $\tilde{Z}_t^n \xrightarrow{n \rightarrow \infty} \mathcal{L}(R_t)$ . We complement this by showing that (note that the right hand side is deterministic) for all  $s \geq 0$

$$n \cdot \mathbf{E}[\varphi_n(\tilde{R}_{s+1/n}^n) - \varphi_n(\tilde{R}_s^n) \mid \tilde{Z}_s^n] \xrightarrow{n \rightarrow \infty} \mathbf{E}[G_{\mathcal{R}}\varphi(R_s)].$$

Indeed,

$$\begin{aligned}
 &\mathbf{E}\left[ \left| n \cdot \mathbf{E}[\varphi_n(\tilde{R}_{s+1/n}^n) - \varphi_n(\tilde{R}_s^n) \mid \tilde{Z}_s^n] - \mathbf{E}[G_{\mathcal{R}}\varphi(R_s)] \right| \right] \\
 &\leq \mathbf{E}\left[ \left| n \cdot \mathbf{E}[\varphi_n(\tilde{R}_{s+1/n}^n) - \varphi_n(\tilde{R}_s^n) \mid \tilde{Z}_s^n] - \mathbf{E}[G_{\mathcal{R}}\varphi(\tilde{R}_s^n) \mid \tilde{Z}_s^n] \right| \right] \quad (4.10) \\
 &\quad + \mathbf{E}\left[ \left| \mathbf{E}[G_{\mathcal{R}}\varphi(\tilde{R}_s^n) \mid \tilde{Z}_s^n] - \mathbf{E}[G_{\mathcal{R}}\varphi(R_s)] \right| \right] \xrightarrow{n \rightarrow \infty} 0
 \end{aligned}$$

in probability, by (3.3) and Lemma 4.2, which shows that – using (4.2) –

$$\mathbf{E}[G_{\mathcal{R}}\varphi(\tilde{R}_s^n) \mid \tilde{Z}_s^n] = \langle \tilde{Z}_s^n, G_{\mathcal{R}}\varphi \rangle \xrightarrow{n \rightarrow \infty} \mathbf{E}[G_{\mathcal{R}}\varphi(R_s)].$$

For every  $t \geq 0$ , we now have that

$$M_t^{n,\varphi_n} \xrightarrow{n \rightarrow \infty} \mathbf{E}[\varphi(R_t)] - \int_0^t \mathbf{E}[G_{\mathcal{R}}\varphi(R_s)] ds = 0. \quad (4.11)$$

Hence, we can write by Doob’s inequality

$$\mathbf{P}\left( \sup_{0 \leq s \leq t} |M_s^{n,\varphi_n}| > \varepsilon \right) \leq \frac{1}{\varepsilon} \mathbf{E}[|M_t^{n,\varphi_n}|] \xrightarrow{n \rightarrow \infty} 0, \quad (4.12)$$

since  $M_t^{n,\varphi}$  is bounded in  $n$  and convergence in (4.11) also holds in probability. Then, using (4.9),

$$\begin{aligned}
 &\mathbf{P}\left( \sup_{0 \leq s \leq t} |\langle \tilde{Z}_s^n, \varphi_n \rangle - \mathbf{E}[\varphi(R_s)]| > 3\varepsilon \right) \\
 &\leq \mathbf{P}\left( \sup_{0 \leq s \leq t} |M_s^{n,\varphi_n}| > \varepsilon \right) \\
 &\quad + \mathbf{P}\left( \int_0^{t-1/n} \left| n \cdot \mathbf{E}[\varphi_n(\tilde{R}_{s+1/n}^n) - \varphi_n(\tilde{R}_s^n) \mid \tilde{Z}_s^n] ds - \mathbf{E}[G_{\mathcal{R}}\varphi(R_s)] \right| ds > \varepsilon \right) \\
 &\quad + \mathbf{P}\left( \int_{t-1/n}^t \left| \mathbf{E}[G_{\mathcal{R}}\varphi(R_s)] \right| ds > \varepsilon \right) \\
 &\xrightarrow{n \rightarrow \infty} 0
 \end{aligned}$$

by (4.12) and (4.10). □

## A Random probability measures

In the following,  $(E, r)$  is a complete and separable metric space and  $\mathcal{P}(E)$  is the set of probability measures on (the Borel  $\sigma$ -algebra of)  $E$ , equipped with the topology of weak convergence. We will state some results about random measures.

**Definition A.1** (First two moment measures). *For a random variable  $Z$ , taking values in  $\mathcal{P}(E)$ , and  $k = 1, 2, \dots$ , there is a uniquely determined measure  $\mu^{(k)}$  on  $\mathcal{B}(E^k)$  such that*

$$\mathbf{E}[Z(A_1) \cdots Z(A_k)] = \mu^{(k)}(A_1 \times \cdots \times A_k)$$

for  $A_1, \dots, A_k \in \mathcal{B}(E)$ . This is called the  $k$ -th moment measure. Equivalently,  $\mu^{(k)}$  is the unique measure such that  $\mathbf{E}[\langle Z, \varphi_1 \rangle \cdots \langle Z, \varphi_k \rangle] = \langle \mu^{(k)}, \varphi_1 \cdots \varphi_k \rangle$ , where  $\langle \cdot, \cdot \rangle$  denotes integration.

**Lemma A.2** (Characterisation of deterministic random measures). *Let  $Z$  be a random variable taking values in  $\mathcal{P}(E)$  with the first two moment measures  $\mu := \mu^{(1)}$  and  $\mu^{(2)}$ . Then the following assertions are equivalent:*

1. *There is  $\nu \in \mathcal{P}(E)$  with  $Z = \nu$ , almost surely.*
2. *The second moment measure has product-form, i.e.  $\mu^{(2)} = \mu \otimes \mu$  (which is equivalent to*

$$\mathbf{E}[\langle Z, \varphi_1 \rangle \cdot \langle Z, \varphi_2 \rangle] = \langle \mu, \varphi_1 \rangle \cdot \langle \mu, \varphi_2 \rangle$$

for all  $\varphi_1, \varphi_2 \in \mathcal{C}_b(E)$ ). (This is in fact equivalent to  $\mathbf{E}[\langle Z, \varphi \rangle^2] = \langle \mu, \varphi \rangle^2$  for all  $\varphi \in \mathcal{C}_b(E)$ ).

In either case,  $\mu = \nu$ .

*Proof.* 1. $\Rightarrow$ 2.: This is clear since we have  $\mathbf{E}[Z(A)] = \nu(A)$ , i.e.  $\mu = \nu$ . Moreover,  $\mathbf{E}[Z(A_1)Z(A_2)] = \nu(A_1)\nu(A_2) = \mu(A_1)\mu(A_2) = \mu \otimes \mu(A_1 \times A_2)$ .

2. $\Rightarrow$ 1.: Since the second moment-measure has product form, for any measurable  $A \subseteq E$ ,  $\mathbf{Var}[Z(A)] = \mathbf{E}[Z(A)Z(A)] - \mathbf{E}[Z(A)]^2 = \mu^{(2)}(A \times A) - (\mu(A))^2 = 0$ , i.e. the random variable  $Z(A)$  has zero variance and therefore is deterministic. In particular,  $Z(A) = \mathbf{E}[Z(A)] = \mu(A)$  and the assertions follows with  $\nu = \mu$ .  $\square$

We end this appendix by recalling Jakubowski's tightness criterion from [10]; see also [4, Theorem 3.6.4].

**Proposition A.3** (Jakubowski's tightness criterion). *Assume the family  $\Pi \subseteq \mathcal{C}_b(E)$  is a vector space that separates points. A sequence  $\mathcal{Z}^1 = (Z_t^1)_{t \geq 0}$ ,  $\mathcal{Z}^2 = (Z_t^2)_{t \geq 0}, \dots$  of  $\mathcal{P}(E)$ -valued processes with càdlàg-paths is tight if the following holds:*

1.  $(\mathcal{Z}^n)_{n=1,2,\dots}$  satisfies the compact containment condition.
2. For every  $f \in \Pi$ , the sequence  $(f(\mathcal{Z}^n))_{n=1,2,\dots}$  with  $f(\mathcal{Z}^n) = (\langle f, Z_t^n \rangle)_{t \geq 0}$  is tight.

**Acknowledgments.** We thank an anonymous referee for careful reading which improved the manuscript.

## References

- [1] S.V. Avery. Microbial cell individuality and the underlying sources of heterogeneity. *Nat. Rev. Microbiol.*, 4:577–587, 2006.
- [2] I. Benjamini and Y. Peres. Markov chains indexed by trees. *Ann. Probab.*, 22(1):219–243, 1994. MR-1258875
- [3] B. Bercu, B. de Saporta and A. Gégout-Petit. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.*, 14(87):2492–2526, 2009.
- [4] D.A. Dawson. Measure-valued Markov processes. In P.L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XXI–1991*, volume 1541 of *Lecture Notes in Mathematics*, pages 1–260, Berlin, 1993. Springer. MR-1242575
- [5] J.-F. Delmas and L. Marsalle. Detection of cellular aging in a Galton–Watson process. *Stoch. Process. and Appl.*, 120(12):2495–2519, 2010. MR-2728175

- [6] B. de Saporta, A. Gégout-Petit and L. Marsalle. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Elec. J. Statistics* 5:1313–1353, 2011.
- [7] S.N. Ethier and T.G. Kurtz. *Markov Processes. Characterization and Convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. MR-0838085
- [8] M.B. Elowitz, A.J. Levine, E.D. Siggia, and P. Swain. Stochastic gene expression in a single cell. *Science Signalling*, 297:1183–1186, 2002.
- [9] J. Guyon. Limit theorems for bifurcating markov chains. application to the detection of cellular aging. *Ann. Appl. Probab.*, 17:1538–1569, 2007. MR-2358633
- [10] A. Jakubowski. On the Skorohod topology. *Ann. Inst. H. Poincar'e*, B22:263–285, 1986. MR-0871083
- [11] O. Kallenberg. *Foundations of Modern Probability*. Springer, New York, 2nd edition, 2002. MR-1876169
- [12] W. Liu and L. Wang. The Markov approximation of the random fields on Cayley trees and a class of small deviation theorems. *Statistics & Probability Letters*, 63(2):113–121, 2003. MR-1986681
- [13] W. Liu and W. Yang. Some strong limit theorems for Markov chain fields on trees. *Probab. Eng. Inf. Sci.*, 18(3):411–422, 2004. MR-2082475
- [14] H.H. McAdams and A. Arkin. It's a noisy business! genetic regulation at the nanomolar scale. *Trends Genet.*, 15:65–69, 1999.
- [15] L. Pelkmans. Using cell-to-cell variability – a new era in molecular biology. *Science*, 336:425–426, 2012.
- [16] J.L. Spudich and D.E. Koshland. Non-genetic individuality: chance in the single cell. *Nature*, 262:467–471, 1976.
- [17] B. Snijder and L. Pelkmans. Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell. Biol.*, 12:119–125, 2011.
- [18] C. Takacs. On the fundamental matrix of finite state Markov chains, its eigensystem and its relation to hitting times. *Mathematica Pannonica*, 17(2):183–193, 2006. MR-2272894
- [19] W. Yang. Some limit properties for markov chains indexed by a homogeneous tree. *Statistics & Probability Letters*, 65(3):241–250, 2003. MR-2018036
- [20] W. Yang and W. Liu. Strong law of large numbers and Shannon-McMillan theorem for Markov chain fields on trees. *IEEE Trans. Inf. Theor.*, 48(1):313–318, 2006. MR-1872187

---

# Electronic Journal of Probability

## Electronic Communications in Probability

---

### Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)

### Economical model of EJP-ECP

- Low cost, based on free software (OJS<sup>1</sup>)
- Non profit, sponsored by IMS<sup>2</sup>, BS<sup>3</sup>, PKP<sup>4</sup>
- Purely electronic and secure (LOCKSS<sup>5</sup>)

### Help keep the journal free and vigorous

- Donate to the IMS open access fund<sup>6</sup> (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

---

<sup>1</sup>OJS: Open Journal Systems <http://pkp.sfu.ca/ojs/>

<sup>2</sup>IMS: Institute of Mathematical Statistics <http://www.imstat.org/>

<sup>3</sup>BS: Bernoulli Society <http://www.bernoulli-society.org/>

<sup>4</sup>PK: Public Knowledge Project <http://pkp.sfu.ca/>

<sup>5</sup>LOCKSS: Lots of Copies Keep Stuff Safe <http://www.lockss.org/>

<sup>6</sup>IMS Open Access Fund: <http://www.imstat.org/publications/open.htm>