The Annals of Applied Probability
2020, Vol. 30, No. 5, 2393–2438
https://doi.org/10.1214/20-AAP1561

STOCHASTIC APPROXIMATION ON NONCOMPACT MEASURE SPACES AND APPLICATION TO MEASURE-VALUED PÓLYA PROCESSES

BY CÉCILE MAILLER[1] AND DENIS VILLEMONAIS[2]

[1]*Department of Mathematical Sciences, University of Bath, c.mailler@bath.ac.uk*
[2]*Université de Lorraine, CNRS, Inria, IECL, denis.villemonais@univ-lorraine.fr*

Our main result is to prove almost-sure convergence of a stochastic-approximation algorithm defined on the space of measures on a noncompact space. Our motivation is to apply this result to measure-valued Pólya processes (MVPPs, also known as infinitely-many Pólya urns). Our main idea is to use Foster–Lyapunov type criteria in a novel way to generalize stochastic-approximation methods to measure-valued Markov processes with a noncompact underlying space, overcoming in a fairly general context one of the major difficulties of existing studies on this subject.

From the MVPPs point of view, our result implies almost-sure convergence of a large class of MVPPs; this convergence was only obtained until now for specific examples, with only convergence in probability established for general classes. Furthermore, our approach allows us to extend the definition of MVPPs by adding "weights" to the different colors of the infinitely-many-color urn. We also exhibit a link between non-"balanced" MVPPs and quasi-stationary distributions of Markovian processes, which allows us to treat, for the first time in the literature, the nonbalanced case.

Finally, we show how our result can be applied to designing stochastic-approximation algorithms for the approximation of quasi-stationary distributions of discrete- and continuous-time Markov processes on noncompact spaces.

1. Introduction. Measure-valued Pólya processes (MVPPs) are a generalization of Pólya urns to the infinitely-many-color case. Pólya urns date back to Pólya and Eggenberger [29], and have been thoroughly studied since then; highlights include, for example, the seminal works of Athreya and Karlin [3] and Janson [37]. Although the question of generalizing Pólya urns to infinitely-many colors was posed in 2004 in [37], MVPPs were only introduced recently by Bandyopadhyay and Thacker [5] and Mailler and Marckert [45]. In both papers, MVPPs are coupled with branching Markov chains on the random recursive tree.

The main idea of this article is to use stochastic-approximation methods (in the spirit of Duflo [28] and Benaïm [7]) to prove almost-sure convergence of a class of MVPPs; the main difficulty comes from the fact that the stochastic-approximation algorithm that we consider is defined on the space of measures on a *noncompact* space.

The stochastic-approximation approach is a classical method for the study of Pólya urn processes when the color-set is finite. For instance, in Section 2.2 of Benaïm [7], the author introduces the reformulation of the classical Pólya urn model in terms of stochastic approximations and provide some ideas for generalizations; in Laruelle and Pagès [40], the authors reformulate the study of several urn models in the setting of stochastic approximations, with applications to clinical trials based on randomized urn models (see also Laruelle and Pagès [41] with applications to optimal asset allocation in finance and Zhang [62] with applications to adaptive designs); we also refer the reader to Pemantle [52], which provides a survey of random processes with reinforcement using stochastic-approximation methods. Since

Received September 2018; revised July 2019.
MSC2020 subject classifications. 62L20, 60J25, 60J80.
Key words and phrases. Stochastic approximation, infinitely-many colors Pólya urns, Markov processes, quasi-stationary distributions.

stochastic approximation naturally applies to processes in general state spaces, it is natural to extend the above methods to the case of MVPPs.

Our main contribution from the stochastic-approximation point of view is to prove convergence of a stochastic-approximation algorithm defined on a noncompact space, namely the set of probability measures on the color-space (being an arbitrary Polish space). To our knowledge, very little is known for measure valued stochastic-approximation algorithm on noncompact spaces, with some exceptions such as [38] and [44]. In the first reference, Janson deals with the compactness issue by proving that the considered model can be restricted to finite subspaces; in the second one, Maillard and Paquette prove that a specific stochastic approximation on the set of measures on $[0, \infty)$ converges almost surely, using an ad hoc coupling with the Kakutani and the uniform process. Our generalization of measure-valued stochastic-approximation methods to noncompact state spaces is made by using abstract FosterLyapunov type criteria in an original way, yielding the tightness of the stochastic-approximation algorithm.

Our main contribution to the theory of MVPPs is to prove almost-sure convergence for a large class of MVPPs (instead of the convergence in probability shown by Mailler and Marckert [45]). Furthermore, we generalize the definition of measure-valued Pólya processes to allow different colors to have different "weights", and to allow the so-called "replacement rule" to be random (two features that are classical in the context of Pólya urns). We are also able to treat the "nonbalanced" case, which was not treated at all by Bandyopadhyay and Thacker [5] or Mailler and Marckert [45].

We believe that the applications of our results go beyond the field of MVPPs: in particular, we detail an application to the approximation of quasi-stationary distributions. Consider a Markov process that gets absorbed when it reaches a state ∂. A quasi-stationary distribution (QSD), if it exists, is the limiting distribution of this Markov process conditioned on not reaching ∂ (we refer the reader to [21, 48, 56] for general introductions to quasi-stationary distributions). Given an absorbed Markov process, it is in general a hard question to prove existence and uniqueness of a QSD; an even harder question is to find an explicit formula for it. With many applications, including the study of interacting particle systems [23, 24], of population dynamics [15, 58], of the simulation of metastable systems [27] and of Monte-Carlo methods [60], numerical approximation methods for quasi-stationary distributions have attracted a lot of interest during the last decades (see for instance, [32, 34, 35, 46, 50]). A recent method introduced independently by Benaïm and Cloez [9] and by Blanchet, Glynn and Zheng [13] makes use of a stochastic-approximation algorithm for computing quasi-stationary distributions on finite state spaces. This method has been recently extended to compact state space cases by Benaïm, Cloez and Panloup [10] and Wang, Roberts and Steinsaltz [61]. We show (see Section 2.3.3) that our result can be applied to prove almost-sure convergence of such QSD-approximation algorithms for absorbed Markov processes taking values on a noncompact space.

1.1. *Definition of the model and main result.* Throughout the article, E is a Polish space endowed with its Borel sigma-field. A measure-valued Pólya process (MVPP) is a Markov chain $(m_n)_{n\geq 0}$ taking values in the set of measures on a Polish space E. It depends on three parameters: its *initial composition* m_0 a nonzero nonnegative measure on E, a sequence of i.i.d. *replacement* kernels[1] $(R^{(n)})_{n\geq 1}$ on E, and a nonnegative *weight* kernel P on E. We assume that:

($\mathrm{T}_{>0}$) almost surely, for all $x \in E$, $R_x^{(n)}$ is a nonnegative measure.

[1] A kernel (resp. a nonnegative kernel) on E is, by definition, a function from E into the set of measures (resp. nonnegative measures) on E. In particular, for all $x \in E$, $R_x^{(n)}$ is a measure on E almost surely.

Given m_n, we define m_{n+1} as follows: pick a random element Y_{n+1} of E according to the probability distribution proportional to $m_n P$, that is, for all Borel set A of E,

$$\mathbb{P}(Y_{n+1} \in A \mid m_n) = \frac{\int_E P_x(A)\,\mathrm{d}m_n(x)}{\int_E P_x(E)\,\mathrm{d}m_n(x)}; \tag{1}$$

and then set

$$m_{n+1} = m_n + R^{(n+1)}_{Y_{n+1}}.$$

Measure-valued Pólya processes were originally introduced by [5] and [45], as a generalization of d-color Pólya urns, although they did not consider "weighted" MVPPs (they always had $P_x = \delta_x$ for all $x \in E$). Let us recall the definition of a Pólya urn and show why MVPPs generalize this model: A d-color Pólya urn is a Markov process $(U(n))_{n\geq 0}$ on $\mathbb{N}^d$ that depends on three parameters: the initial composition vector $U(0)$, the replacement matrix M, and weights $w_1, \dots, w_d \in (0, \infty)$. The vector $U(n)$ represents the content of an urn that contains balls of d different colors; balls of color i all have weight w_i. Given $U(n)$, one defines $U(n+1)$ by picking a ball at random in the urn with probability proportional to its weight, denoting the color of this random ball ξ_{n+1}, and setting $U(n+1) = U(n) + M_{\xi_n}$, where $M_1, \dots, M_d$ are the lines of M.

If we let $E = \{1, \dots, d\}$ and $m_n = \sum_{i=1}^d U_i(n)\delta_i$ for all $n \geq 0$, then m_n is a measure-valued Pólya process with replacement kernel

$$R^{(n)}_x = \sum_{i=1}^d M_{x,i}\delta_i \quad \text{(almost surely for all } n \geq 0,\ 1 \leq x \leq d),$$

and weight kernel $P_x = w_x \delta_x$ for all $1 \leq x \leq d$.

Therefore, the MVPP process $(m_n)_{n\geq 0}$ can be thought of as a composition measure on a set E of colors, and the random variable Y_{n+1} can be seen as the color of the "ball" drawn at time $n+1$. The main advantage of this wider model is that one can consider Pólya urns defined on an infinite, and even uncountable, set.

Our main result is to prove almost-sure convergence of the sequence $(m_n/m_n(E))_{n\geq 0}$ to a deterministic measure under the following assumptions: We denote by R the common expectation of the $R^{(n)}$'s and set $Q^{(n)} = R^{(n)}P$ for all $n \geq 1$, and $Q = RP$, meaning that, for all $x \in E$ and all Borel set $A \subseteq E$,

$$Q^{(n)}_x(A) = \int_E P_y(A)\,\mathrm{d}R^{(n)}_x(y) \quad \text{and} \quad Q_x(A) = \int_E P_y(A)\,\mathrm{d}R_x(y).$$

We assume that:

(A1) for all $x \in E$, $Q_x(E) \leq 1$, and there exists a probability measure μ on $\mathbb{R}$ with positive mean such that, for all $x \in E$, the law of $Q^{(i)}_x(E)$ stochastically dominates μ. In particular, setting $c_1 = \int_0^\infty x\,\mathrm{d}\mu(x)$,

$$0 < c_1 \leq \inf_{x\in E} Q_x(E) \leq \sup_{x\in E} Q_x(E) \leq 1;$$

(A2) there exists a locally-bounded lower semicontinuous function $V : E \to [1, +\infty)$ such that:

- (i) for all $N \geq 0$, the set $\{x \in E : V(x) \leq N\}$ is relatively compact;
- (ii) there exist two constants $\theta \in (0, c_1)$ and $K \geq 0$ such that

$$Q_x \cdot V \leq \theta V(x) + K \quad (\forall x \in E),$$

(iii) and that there exist three constants $r > 1$, $p > \frac{\ln\theta}{\ln(\theta/c_1)} \vee 2$, $A > 0$ such that

$$\mathbb{E}\big[R_x^{(1)}(E)^r\big] \vee \mathbb{E}\big[Q_x^{(1)}(E)^p\big] \leq AV(x) \quad (\forall x \in E).$$

Under Assumption (A1), Q is a nonnegative kernel such that $\sup_x Q_x(E) \leq 1$, so that $Q - I$ is the jump kernel (or infinitesimal generator) of a unique sub-Markovian transition kernel $(P_t)_{t\geq 0}$ on E. We consider the continuous-time pure-jump Markov process $(X_t)_{t\geq 0}$ on $E \cup \{\partial\}$, where $\partial \notin E$ is an absorbing state, with Markovian transition kernel $P_t + (1 - P_t(E))\delta_\partial$. A probability distribution ν is a *quasi-stationary distribution* of $(X_t)_{t\geq 0}$ if, and only if, there exists a probability measure α on E such that, for all Borel sets $A \subseteq E$,

$$\mathbb{P}_\alpha(X_t \in A \mid X_t \neq \partial) \xrightarrow[t\to+\infty]{} \nu(A),$$

where $\mathbb{P}_\alpha$ is the law of X with initial distribution α.

(A3) the continuous-time pure jump Markov process X with sub-Markovian jump kernel $Q - I$ admits a quasi-stationary distribution $\nu \in \mathcal{P}(E)$. We further assume that the convergence of $\mathbb{P}_\alpha(X_t \in \cdot \mid X_t \neq \partial)$ holds uniformly with respect to the total variation norm on $\{\alpha \in \mathcal{P}(E) \mid \alpha \cdot V^{1/q} \leq C\}$, for each $C > 0$, where $q = p/(p-1)$.

Finally, we need the following technical assumption:

(A4) for all bounded continuous functions $f : E \to \mathbb{R}$, $x \in E \mapsto R_x f$ and $x \in E \mapsto Q_x f$ are continuous.

Under these assumptions, we are able to prove almost-sure convergence of the renormalized MVPP $\tilde{m}_n := m_n/m_n(E)$:

THEOREM 1. *Under Assumptions* ($\mathrm{T}_{>0}$) *and* (A1)–(A4), *if* $m_0 \cdot V < \infty$ *and* $m_0 P \cdot V < \infty$, *then the sequence of random measures* $(m_n/n)_{n\geq 0}$ *converges almost surely to* νR *with respect to the topology of weak convergence. Moreover,* $\sup_n\{m_n P \cdot V^{1/q}/n\} < +\infty$ *almost surely, where* $q = p/(p-1)$.

Furthermore, if $\nu R(E) > 0$, *then* $(\tilde{m}_n)_{n\in\mathbb{N}}$ *converges almost surely to* $\nu R/\nu R(E)$ *with respect to the topology of weak convergence.*

REMARK 1. If $R = Q$, then the quasi-stationary distribution ν is a left eigenfunction for R, with associated eigenvalue $\theta_0 \in (0, 1]$. In particular, Theorem 1 implies that the average mass of m_n, that is, $m_n(E)/n$, converges almost surely to θ_0.

REMARK 2. The main result holds under a weaker versions of Assumption 3: namely, the total variation distance can be replaced by any metric inducing the topology of weak convergence (or a stronger one).

REMARK 3. To illustrate how this theorem applies, let us first consider the simple case of a classical d-color Pólya urn of random replacement matrix $M^{(n)}$ with no weights, where $(M^{(n)})_n$ is a sequence of i.i.d. random matrices with nonnegative entries and mean M. We assume that $\sum_{i=1}^d M_{x,i} > 0$ for all $1 \leq x \leq d$ and that M is irreducible. Let $S = \max_{x=1}^d \sum_{i=1}^d M_{x,i}$, and let $m_n = \frac{1}{S}\sum_{i=1}^d U_i(n)\delta_i$, where $U_i(n)$ is the number of balls of color i in the urn at time n. One can check that $(m_n)_{n\geq 0}$ is an MVPP on $E = \{1, \dots, d\}$ with replacement kernel $R_x^{(n)} = \frac{1}{S}\sum_{i=1}^d M_{x,i}^{(n)}\delta_i$, for all $n \geq 0$ and $1 \leq x \leq d$, such that $R = M/S$.

Note that, since we have no weights, $R = Q$. Let μ be the distribution of $\min_{x\in\{1,\dots,d\}} X_x$, where $X_1, \dots, X_d$ are independent random variables respectively distributed as $Q_1^{(1)}(E), \dots, Q_d^{(1)}(E)$. Assumption (A1) is satisfied since μ has positive mean $c_1 \leq Q_x(E) \leq 1$ for all

$1 \le x \le d$. Assumption (A2) is automatically satisfied since the color space E is compact. Consider the process X on $E \cup \{\partial\}$ absorbed at ∂ and whose jump matrix restricted to E is given by $M/S - I$. Then, since M/S is irreducible, the process X conditioned on not hitting ∂ has a unique quasi-stationary distribution $\nu = \sum_{i=1}^n v_i \delta_i$, which is given by the unique nonnegative left eigenvector v of $M/S - I$ and hence of M. It is also known (see, e.g., Darroch and Seneta [22]) that there exists $C, \delta > 0$ such that $\|\mathbb{P}_\alpha(X_t \in \cdot \mid X_t \notin \partial) - \nu\|_{\mathrm{TV}} \le C e^{-\delta t}$ for all $\alpha \in \mathcal{P}(E)$, which thus implies (A3). Finally, Assumption (A4) is trivially satisfied since E is discrete.

Thus, Theorem 1 applies, and we get that, almost surely when n tends to infinity, $\tilde{m}_n \to \nu R/\nu R(E) = \nu$ (with respect to the topology of weak convergence), and thus, $U(n)/n \to v$, a result that dates back to Athreya and Karlin's work on generalized Pólya urns [3].

REMARK 4. In the original Pólya urn model, the replacement matrix is the identity and is not irreducible. In this case, there are several quasi-stationary distributions and thus Assumption (A3) fails. We may thus say that the equivalent of the irreducible assumption in Athreya and Karlin's result is our Assumption (A3).

In Section 2 we apply our result to many more examples, and, in particular, to examples where the color space E is infinite, and even noncompact. Before that, in the rest of this introduction, we discuss our result and its assumptions.

1.2. *Discussion of the result in view of the existing literature on MVPPs.* Our definition of a measure-valued Pólya process is more general than the definition of Bandyopadhyay and Thacker [5] and Mailler and Marckert [45]; indeed, their model can be obtained from ours by taking $R^{(i)} = R$ almost surely for all $i \ge 1$ (deterministic replacement rule), and $P_x = \delta_x$ for all $x \in E$ (no weights). [5] and [45] also make the following assumptions:

(I) $0 < m_0(E) < +\infty$;

(B) for all $x \in E$, $R_x(E) = 1$;

(E) there exist two sequences $(a_n)_{n\ge 0}$ and $(b_n)_{n\ge 0}$ such that the Markov chain $(W_n)_{n\ge 0}$ on E of transition kernel $(R_x)_{x\in E}$ satisfies

$$\frac{W_n - b_n}{a_n} \Rightarrow \nu,$$

in distribution when n goes to infinity, independently from the initial distribution of W_0.

(R) the sequences $(a_n)_{n\ge 0}$ and $(b_n)_{n\ge 0}$ are such that, for all $\varepsilon_n = o(\sqrt{n})$, for all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \frac{b_{n+x\sqrt{n}+\varepsilon_n} - b_n}{a_n} = f(x) \quad \text{and} \quad \lim_{n\to\infty} \frac{a_{n+x\sqrt{n}+\varepsilon_n}}{a_n} = g(x),$$

where f and g are two measurable functions.

The names of the assumptions are (I) for *initial composition*, (B) for *balance*, (E) for *ergodicity* and (R) for *regularity*. Under these assumptions Mailler and Marckert [45] prove that (a slightly weaker version of this result is proved by [5]):

THEOREM 2 (Mailler and Marckert [45]). *If $(m_n)_{n\ge 0}$ is a MVPP that satisfies assumptions* (I), (B), (E) *and* (R), *then*

$$n^{-1} m_n(a_{\log n} \cdot + b_{\log n}) \to \mu, \tag{2}$$

in probability when n goes to infinity, for the topology of weak convergence, where μ is the distribution of $f(\Lambda) + g(\Lambda)\Phi$, where $\Lambda \sim \mathcal{N}(0,1)$ and $\Phi \sim \nu$ are independent.

Note that Theorem 1 applies under (I), (B), (E) and (R) if we assume additionally that $a_n \equiv 1$ and $b_n \equiv 0$, and it gives that

$$\frac{m_n}{n} \to \nu \quad \text{almost surely,}$$

which improves the convergence in probability of Theorem 2. Our theorem though does not cover the cases of more general renormalization sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$.

In summary, our main contributions to the theory of MVPPs are to:

(α) remove the balance hypothesis (B) and replace it by the weaker (A1);
(β) prove convergence almost sure in equation (2) when $a_n \equiv 1$ and $b_n \equiv 0$;
(γ) allow the weighting of the different elements of E, and to
(δ) allow the re-sampling of the replacement measures at each time-step in an i.i.d. way.

Our result was motivated by the classical Pólya urn theory (see, e.g., [37]), in which all these features are standard. Since this paper was submitted, Janson [39] generalised Theorem 2 to the random replacement case, thus treating (γ) in that case. Also, Bandyopadyhay, Janson and Thacker [4] prove almost sure convergence of a class of balanced MVPPs for which the set of colors is countable and under a condition of strong ergodicity for the underlying Markov chain, thus treating (β) in that case.

REMARK 5. A standard generalization of finitely-many-color Pólya urns is indeed to add weights (or activities): each color x is given a weight $w(x)$, and, at every time-step, one picks a ball in the urn with probability proportional to the weights (vs. uniformly at random in the nonweighted model) and then applies the replacement rule associated to this color (see, e.g., [37]). In our model, if $P_x = w(x)\delta_x$, where $w(x)$ is nonnegative, then

$$\mathbb{P}(Y_{n+1} \in A \mid m_n) = \frac{\int_A w(x)\,\mathrm{d}m_n(x)}{\int_E w(x)\,\mathrm{d}m_n(x)},$$

which corresponds to weighting the color x by a weight $w(x)$. The introduction of a weight kernel is a generalization of the weight concept: one can for example see P as a noise on the color drawn at random.

REMARK 6. Our model, assumptions and result can be easily adapted to the situation where $R^{(1)}$ is a kernel from E to an other Polish state space F and P is a nonnegative kernel from F to E. The main point of this extension is to check that the proof of Theorem 1 mainly makes use of the properties of the composed kernel $Q^{(1)}$. For instance, in the d-color Pólya urn model (see the end of Section 1.1), if $\sum_{j=1}^d M_{i,j} > 0$ for all $i \in \{1,\dots,d-1\}$ and if $\sum_{j=1}^d M_{d,j} = 0$, then one can choose $E = \{1,\dots,d-1\}$ and $F = \{1,\dots,d\}$ together with the kernels $R^{(i)}_{i,j} = R_{i,j} = M_{i,j}/S$ for all $(i,j) \in E \times F$ and $P_{ij} = \mathbf{1}_{i\neq d}\delta_i$ for all $i \in F$. In this case, we thus have $Q^{(1)}_{i,j} = Q_{i,j} = M_{i,j}/S$ for all $(i,j) \in E \times E$. If M restricted to $E \times E$ is irreducible, we get that there exists a unique quasi-stationary distribution ν on E for the continuous time Markov process X with infinitesimal generator $Q - I$ (see [22]). Hence, using our approach to MVPPs in this slightly more general context, we get that the d-color Pólya urn converges almost surely, when $n \to +\infty$, to $\nu R/\nu R(E)$ (which is a probability measure on F), a result that can be found, for example, in [37].

REMARK 7. The main idea in [5] and [45] is to show a link between the MVPP of replacement kernel R and the Markov chain of kernel R. This relationship breaks down if the balance assumption is not satisfied since R is no longer a probability kernel but a sub-Markovian kernel (we can assume without loss of generality that the upper bound of

$\sup_x R_x(E)$ is 1). Our main idea to relax the balance assumption is to add an absorbing state ∂ that "makes" the transition kernel Markovian; note that this idea is similar to adding "dummy" balls in the finitely-many-color case (see [37]). The ergodicity assumption (E) then naturally becomes Assumption (A3) that the Markov chain has a quasi-stationary distribution.

The link between Pólya urns and quasi-stationary distributions already exists in the literature; for example, Aldous, Flannery and Palacios [2] apply the convergence results of Athreya and Karlin [3] to approximating quasi-stationary distributions on a finite state space. Our main result generalizes this work to the case of measure-valued Pólya processes.

REMARK 8. Another difference with [5] and [45] is that Theorem 1 naturally covers periodic transition kernels since we consider the continuous time process associated to it, which is never periodic.

1.3. *Discussion of the assumptions.* In Assumption (A1), we assume that $Q_x(E)$ is uniformly bounded from above by 1. If the supremum $\kappa = \sup_{x\in E} Q_x(E)$ is finite (but larger than 1), one can consider the process defined by $\hat{m}_n := m_n/\kappa$ for all $n \geq 0$. One can easily check that $\hat{m}_n$ is an MVPP with parameters $\hat{R}^{(i)} = R^{(i)}/\kappa$, $\hat{P} = P$, and $\hat{Q} = \hat{R}\hat{P}$, and such that $\hat{m}_0 = m_0/\kappa$. Also, it satisfies $\hat{Q}_x(E) \leq 1$ as in Assumption (A1).

For the lower bound, we assume that the random value $Q_x^{(i)}(E)$ stochastically dominates an integrable probability measure μ on $\mathbb{R}$ with mean $c_1 > 0$. This is used to prove that, for any fixed $c' \in (\theta, c_1)$

$$\liminf_{n\to+\infty} \frac{m_n P(E)}{n} \geq c',$$

almost surely; this is done by a coupling argument (see Lemma 3). An alternative assumption, which may be particularly useful when $Q_x^{(i)}(E)$ can take negative values as in Section 1.4 below, is that there exist $c_1 > 0$ and $\beta > 1$ such that

$$c_1 \leq \inf_{x\in E} Q_x(E) \leq \sup_{x\in E} Q_x(E) \leq 1 \quad \text{and} \quad \sup_{x\in E} \mathbb{E}\left|Q_x^{(i)}(E) - Q_x(E)\right|^\beta < +\infty. \tag{3}$$

For instance, in the example developed in Remark 3, take $E = \{1, 2\}$ and

$$M^{(n)} = \varepsilon_n \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} + (1 - \varepsilon_n) \begin{pmatrix} 1 & 2 \\ 1 & 0 \end{pmatrix},$$

where $(\varepsilon_n)_{n\geq 1}$ is a sequence of i.i.d. Bernoulli random variables with parameter $1/2$. Then any probability measure μ on E as in Assumption (A1) has nonpositive mean, so that this assumption is not satisfied. However, Assumption (3) is satisfied with $c_1 = 1$.

Assumption (A2) is a Lyapunov assumption and is standard in the study of the ergodicity of Markov processes. In Section 2, we show how to apply our main result to examples, and therefore give examples of such Lyapunov functions. There is no general method to find Lyapunov functions, except testing functions from classical families (polynomials, exponentials, etc). For instance, for processes in $\mathbb{Z}$, $\mathbb{R}$ or $\mathbb{R}^d$ with a drift towards 0, exponential or power functionals of the distance to 0 often prove to be useful. Sometimes, probabilistic arguments can help find a Lyapunov function; indeed, if, for some $\theta \in (0, 1)$, $\mathbb{E}_x[\theta^{\tau_K}]$ is finite for all $x \in E$ (where τ_K denotes here the first entry time in a set K of a discrete-time Markov chain with transition probability given by Q), then $V : x \mapsto \mathbb{E}_x[\theta^{\tau_K}]$ satisfies $Q_x \cdot V \leq \theta V(x)$ for all $x \in E \setminus K$.

When $Q_x(E) = 1$ for all $x \in E$, the existence of a Lyapunov function for Q can be used to prove the ergodicity of the Markov process X. More precisely, if compact subsets of E are *petite sets* for X, then the existence of a Lyapunov function entails the ergodicity of X (see Meyn and Tweedie [49], for the definition of a petite set and for the deduction that X

is ergodic) and hence Assumption (A3). Note that our proof does not seem to generalize to the case of a weaker form of Lyapunov function (satisfying, for instance, $Q_x(V) \leq V(x) - V^{1/2}(x) + C$ for all $x \in E$), although those weaker forms are generally sufficient to prove the ergodicity of the process.

When Q is a sub-Markovian kernel, it has been recently proved in Champagnat and Villemonais [16] that the Lyapunov condition (A2-ii), with additional suitable assumptions, can be used to prove the existence of a quasi-stationary distribution ν and to prove that the domain of attraction of ν contains $\{\alpha \in \mathcal{P}(E) \mid \alpha \cdot V^{1/q} < \infty\}$. These criteria will be used extensively in our examples. Note that this result, when applicable, entails the existence of a quasi-stationary distribution ν and the uniform convergence of Assumption (A3) in total variation norm.

For conditions implying Assumption (A3), we also refer the reader to Villemonais [59] where the case of birth and death processes is considered, to Gosselin [33], and Ferrari, Kesten and Martínez [31] for population processes and the utility of the theory of R-positive matrices in this matter. This is also implied by the general results provided in Champagnat and Villemonais [17].

1.4. *Removing balls from the urn.* In the finitely-many-color case, it is often allowed to remove balls from the urn, that is, the coefficients of the replacement matrix can be negative. In Theorem 1, we have assumed that the measures $(R_x)_{x\in E}$ are positive, but we can in fact consider situations where $(R_x)_{x\in E}$ are signed kernels as soon as they satisfy additional assumptions (which are already implied by conditions (A1)–(A4) when $(R_x)_{x\in E}$ are positive measures). In Section 2, we give examples that fall into this special framework.

In this section, we assume that $(R_x^{(i)})_{x\in E}$ is almost surely a signed kernel such that, for all $x \in E$, Q_x restricted to $E \setminus \{x\}$ is a positive measure and $Q_x(\{x\}) \in \mathbb{R}$. We assume that:

(T) for all $n \geq 0$, m_n is almost surely a positive measure.

In the finitely-many-color case, this assumption is called tenability. It is clearly satisfied when Assumption ($\mathrm{T}_{>0}$) holds true. We refer the reader to [53], Definition 1.1-(iii), for a sufficient condition for tenability in the finite state space case. As will appear in the examples section, tenability is often naturally satisfied.

In the case when $(R_x^{(i)})_{x\in E}$ is allowed to be a signed kernel, we need to replace Assumption (A2) by:

(A′2) there exist a locally bounded function $V : E \to [1, +\infty)$ and some constants $r > 1$, $p > 2$, $q' > q := p/(p-1)$, $\theta \in (0, c_1)$, $K > 0$, $A \geq 1$, and $B \geq 1$, such that:

(i) for all $N \geq 0$, the set $\{x \in E\colon\ V(x) \leq N\}$ is relatively compact.

(ii) for all $x \in E$,

$$Q_x \cdot V \leq \theta V(x) + K \quad \text{and} \quad Q_x \cdot V^{1/q} \leq \theta V^{1/q}(x) + K \quad (\forall x \in E).$$

(iii) for all continuous functions $f : E \to \mathbb{R}$ bounded by 1 and all $x \in E$,

$$|Q_x \cdot f|^{q'} \vee \mathbb{E}\big[\big|R_x^{(i)} \cdot f - R_x \cdot f\big|^r\big] \vee \mathbb{E}\big[\big|Q_x^{(i)} \cdot f - Q_x \cdot f\big|^p\big] \leq AV(x),$$

(iv) and

$$\big|Q_x \cdot V^{1/q}\big|^q \vee |Q_x \cdot V| \vee \mathbb{E}\big[\big|Q_x^{(i)} \cdot V^{1/q} - Q_x \cdot V^{1/q}\big|^r\big] \leq BV(x).$$

Assuming in addition that Assumptions (A1), (A3) and (A4) are satisfied, the conclusions of Theorem 1 hold true. Since the set of assumptions (T), (A1), (A′2), (A3), (A4) is actually implied (see Lemma 1 below) by the assumptions of Theorem 1, we prove this result in the more general situation of the present subsection.

LEMMA 1. *Assumptions* ($T_{>0}$), (A1)–(A4) *imply Assumptions* (T), (A1), (A′2), (A3), (A4).

PROOF. The fact that Assumption ($T_{>0}$) implies Assumption (T) is straightforward. Fix $q = p/(p-1)$; using Hölder's inequality ($q \geq 1$) and Assumption (A2-ii), we get, for all $x \in E$,

$$(Q_x \cdot V^{1/q})^q \leq Q_x(E)^{q/p} Q_x \cdot V \leq \theta V(x) + K.$$

Using the fact that, by concavity, for all $a \leq 1$ and $u \geq 0$, $(1+u)^a \leq 1 + u^a$, we thus get

$$Q_x \cdot V^{1/q} \leq \theta^{1/q} V^{1/q}(x) + K^{1/q}.$$

To prove (A′2-ii), it is thus enough to show that $\theta^{1/q} < c_1$. This follows since, by assumption on p,

$$\frac{1}{q} \ln \theta = (1 - 1/p) \ln \theta < \left(1 - \frac{\ln(\theta/c_1)}{\ln \theta}\right) \ln \theta = \ln c_1.$$

Now we prove (A2-iii); first note that, since $q' := p > q > 1$, we have, by Jensen's inequality, for all continuous function bounded by 1,

$$|Q_x \cdot f|^{q'} \leq \mathbb{E}\big[|Q_x^{(1)} \cdot f|^{q'}\big] \leq \mathbb{E}\big[Q_x^{(1)}(E)^{q'}\big] \leq AV(x),$$

where we have used (A2-iii). Similarly, for all $r' \in (1, r]$, using the convexity of $u \mapsto u^{r'}$ and Jensen's inequality, we get that,

$$\begin{aligned}\mathbb{E}\big[|R_x^{(1)} \cdot f - R_x \cdot f|^{r'}\big] &\leq 2^{r'-1}\mathbb{E}\big[|R_x^{(1)} \cdot f|^{r'} + |R_x \cdot f|^{r'}\big] \\ &\leq 2^{r'}\mathbb{E}\big[|R_x^{(1)} \cdot f|^{r'}\big] \leq A2^{r'-1}V(x),\end{aligned}$$

and similarly for $\mathbb{E}[|Q_x^{(1)} \cdot f - Q_x \cdot f|^p]$.

It only remains to prove (A2-iv). We have, using Hölder's inequality, the fact that Q_x is nonnegative and the fact that $V(x) \geq 1$,

$$\begin{aligned}|Q_x \cdot V^{1/q}|^q = (Q_x \cdot V^{1/q})^q &\leq Q_x(E)^{q/p} Q_x \cdot V \leq Q_x \cdot V \\ &\leq \theta V(x) + K \leq (\theta + K)V(x).\end{aligned}$$

Then, using the convexity of $u \mapsto u^{r'}$ and Jensen's inequality, we get that

$$\begin{aligned}\mathbb{E}\big[|Q_x^{(1)} \cdot V^{1/q} - Q_x \cdot V^{1/q}|^{r'}\big] &\leq 2^{r'-1}\mathbb{E}\big[|Q_x^{(1)} \cdot V^{1/q}|^{r'} + |Q_x \cdot V^{1/q}|^{r'}\big] \\ &\leq 2^{r'}\mathbb{E}\big[(Q_x^{(1)} \cdot V^{1/q})^{r'}\big].\end{aligned}$$

Now, using Hölder's inequality, we obtain

$$Q_x^{(1)} \cdot V^{1/q} \leq (Q_x^{(1)} \cdot V)^{1/q} Q_x^{(1)}(E)^{1/p}.$$

Using again Hölder's inequality, we have, setting $\varpi = q/r'$,

$$\begin{aligned}\mathbb{E}\big[(Q_x^{(1)} \cdot V^{1/q})^{r'}\big] &\leq \mathbb{E}\big[(Q_x^{(1)} \cdot V)^{\frac{\varpi r'}{q}}\big]^{1/\varpi}\mathbb{E}\big[Q_x^{(1)}(E)^{\frac{r'\varpi}{p(\varpi-1)}}\big]^{(\varpi-1)/\varpi} \\ &\leq \mathbb{E}\big[Q_x^{(1)} \cdot V\big]^{1/\varpi}\mathbb{E}\big[1 + Q_x^{(1)}(E)^p\big]^{(\varpi-1)/\varpi},\end{aligned}$$

where we used that $\frac{r'\varpi}{p(\varpi-1)} = \frac{r'(q-1)}{q-r'} \leq p$ for r' small enough in $(1, r]$. Using Assumption (A2-ii), we get $\mathbb{E}[Q_x^{(1)} \cdot V] = Q_x(V) \leq (\theta + K)V(x)$ and, using Assumption (A2-iii), $\mathbb{E}[Q_x^{(1)}(E)^p] \leq AV(x)$. We finally deduce that

$$\mathbb{E}\big[(Q_x^{(1)} \cdot V^{1/q})^{r'}\big] \leq (\theta + K + 1 + A)V(x),$$

where we have used that $1/\varpi + (\varpi - 1)/\varpi = 1$. This concludes the proof. □

REMARK 9. When $Q_x(\{x\})$ is not bounded uniformly in x, the infinitesimal generator $Q - I$ may not define a unique sub-Markovian transition kernel $(P_t)_{t\geq 0}$, and hence a unique pure jump Markov process X (in distribution). The problem of existence and uniqueness of such a transition kernel has been considered in great generality by Feller in [30] and is also studied in details in [19], Chapter 2. In our case, Assumption (A2-ii) and Theorem [19], Theorem 2.25, imply that $Q - I$ uniquely determines a sub-Markovian semigroup $(P_t)_{t\in[0,+\infty)}$ and hence a unique jump-process X (in distribution). As a consequence, Assumption (A3) remains unambiguous when Assumptions ($T_{>0}$), (A2) are replaced by Assumptions (T), (A′2).

Plan of the paper. In Section 2, we apply Theorem 1 to several examples. In particular, in Section 2.2, we look at examples that come from studying different characteristics (degree distribution, protected nodes) in random recursive trees or forests. In Section 2.3, we detail the case when the replacement kernels are the occupation measures of Markov processes, in discrete and continuous time, and show how one can apply these results to the numerical approximation of QSDs on a noncompact space (see Section 2.3.3). Finally, Section 3 contains the proof of Theorem 1.

2. Examples.

2.1. *Markov chains.*

2.1.1. *Ergodic Markov chains.* In [45], the following example is treated: take $E = \mathbb{N} := \{0, 1, 2, \dots\}$, fix $0 < \lambda < \mu$, and set

$$R_x = \frac{\lambda}{x\mu + \lambda}\delta_{x+1} + \frac{x\mu}{x\mu + \lambda}\delta_{x-1},$$

for all $x \neq 0$, and $R_0 = \delta_1$. This example is not weighted, meaning that $P_x = \delta_x$ for all $x \in E$, and balanced since $R_x(E) = 1$ for all $x \in E$. Note that the Markov chain of transition kernel R is the $M/M/\infty$ queue. Theorem 2 implies that this MVPP satisfies

$$n^{-1}m_n \to \gamma \quad \text{in probability},$$

where γ is the stationary measure of the $M/M/\infty$ queue, that is,

$$\gamma(x) = \left(\frac{\lambda}{\mu}\right)^x \frac{e^{-\lambda/\mu}}{x!} \quad (\forall x \in \mathbb{N}).$$

Let us show how our result implies almost-sure convergence of this MVPP. Note that, in this example, the $R^{(i)}$ are deterministic and equal to R, $P_x = \delta_x$; therefore, $Q^{(i)} = Q = R$ ($\forall i \geq 1$). Since $R_x(E) = 1$ for all $x \in \mathbb{N}$, then (A1) is satisfied (we can take $\mu = \delta_1$, and thus, $c_1 = 1$). Assumption (A2) also holds: one can take $V(x) = e^x$, implying that

$$R_x \cdot V = \frac{\lambda e^{x+1} + \mu x e^{x-1}}{\lambda + \mu x} = \frac{\lambda e^2 + \mu x}{\lambda + \mu x} e^{x-1} = \frac{\lambda e^2 + \mu x}{e(\lambda + \mu x)} V(x).$$

Note that

$$\frac{\lambda e^2 + \mu x}{e(\lambda + \mu x)} < \frac{2}{e} \quad \Leftrightarrow \quad x > \frac{\lambda(e^2 - 2)}{\mu},$$

therefore,

$$R_x \cdot V \leq \theta V(x) + K,$$

where $\theta = \frac{2}{\mathrm{e}} \in (0, c_1)$ and $K = \sup_{x \le \lambda(\mathrm{e}^2-2)/\mu} R_x \cdot V$. Also note that, for all $r, p > 1$, we have

$$\mathbb{E} R_x^{(1)}(E)^r \vee \mathbb{E} Q_x^{(1)}(E)^p = R_x(E)^r \vee R_x(E)^p = 1,$$

implying that (A2-iii) holds. Since the queue $M/M/\infty$ is ergodic with stationary distribution γ, we can infer that the continuous-time Markov process of generator $R - I$ is also ergodic and the domain of attraction of γ is $\mathcal{P}(\mathbb{N})$. Moreover, the same procedure as in the proof of Lemma 1 shows that, for any $q > 1$, $Q_x \cdot V^{1/q} \le \theta^{1/q} V(x) + K^{1/q}$, where $\theta^{1/q} < 1$. This and the Foster–Lypanuov type criteria of [49] provide the uniform convergence to ν required in Assumption (A3). Finally, since $\mathbb{N}$ is discrete, (A4) is trivially satisfied. Thus, Theorem 1 applies and we can conclude that if $\sum_{k \ge 0} \mathrm{e}^k m_0(k)$ is finite, then

$$n^{-1} m_n \to \gamma \quad \text{almost surely when } n \to \infty.$$

2.1.2. *Quasi-ergodic Markov chains.* Let us now consider the more general case where $E = \mathbb{N}$ and, for all $x \in E$,

$$R_x = \lambda_x \delta_{x+1} + \mu_x \delta_{x-1},$$

where $(\lambda_x)_x$ and $(\mu_x)_x$ are families of positive numbers such that $\mu_0 = 0$, $\lambda_0 > 0$, $\inf_{x \ge 1} \mu_x > 0$, $\sup_x \mu_x < \infty$ and $\lambda_x = o(\mu_x)$ when $x \to +\infty$. In this situation, the MVPP is not weighted, so that $P_x = \delta_x$ and $Q_x = R_x$ for all $x \in E$, and it is not balanced (hence Theorem 2 does not apply).

We assume, without loss of generality, that $\sup_x(\lambda_x + \mu_x) = 1$, so that $Q_x(E) \le 1$ for all $x \in E$. Let μ be the Dirac mass at $\inf_x(\lambda_x + \mu_x)$, which is positive. Assumption (A1) is satisfied with this choice of μ, and $c_1 = \inf_x(\lambda_x + \mu_x)$. Let

$$V(x) = \mathrm{e}^{ax} \quad \text{with } a > 0 \text{ such that } e^{-a} \le c_1/4.$$

Assumption (A2-i) is clearly satisfied, and (A2-ii) can be checked easily: for all $x \in E$,

$$\begin{aligned} Q_x \cdot V &= \lambda_x \mathrm{e}^{a(x+1)} + \mu_x \mathrm{e}^{a(x-1)} = V(x)\big(\lambda_x \mathrm{e}^a + \mu_x \mathrm{e}^{-a}\big) \\ &\le V(x) \sup_y \mu_y \left(\frac{\lambda_x}{\mu_x} \mathrm{e}^a + \mathrm{e}^{-a}\right) \le V(x)\left(\frac{\lambda_x}{\mu_x} \mathrm{e}^a + \frac{c_1}{4}\right) \\ &\le \theta V(x) + K, \end{aligned}$$

where $\theta = \frac{c_1}{2}$ and $K = \max\{V(y)(\frac{\lambda_y}{\mu_y} \mathrm{e}^a + \frac{c_1}{4})$, with y s.t. $\frac{\lambda_y}{\mu_y} \mathrm{e}^a + \frac{c_1}{4} \ge \frac{c_1}{2}\}$ (note that this last set is finite by assumption and hence that $K < \infty$). Since $R_x(E) = Q_x(E)$ is uniformly bounded from above, (A2-iii) is trivial for any fixed $p > 2 \vee \frac{\ln \theta}{\ln \theta - \ln c_1}$. Assumption (A4) is also clearly satisfied in this case since E is discrete.

The same procedure as in the proof of Lemma 1 shows that $Q_x \cdot V^{1/q} \le \theta^{1/q} V(x) + K^{1/q}$, where $\theta^{1/q} < c_1$ since we fixed $p > \frac{\ln \theta}{\ln \theta - \ln c_1}$. Now, using Theorem 5.1 and Remark 11 in [16] for the irreducible process X with infinitesimal generator $Q - I$, we deduce that there exist a quasi-stationary distribution ν_{QSD} for X and two positive constants Cst, $\delta > 0$ such that, for all probability measure $\alpha \in E$, satisfying $\alpha \cdot V^{1/q} < +\infty$,

$$\big\| \mathbb{P}_\alpha(X_t \in \cdot \mid t < \tau_\partial) - \nu_{\mathrm{QSD}} \big\|_{\mathrm{TV}} \le \mathrm{Cst}\, \alpha \cdot V^{1/q} \mathrm{e}^{-\delta t},$$

which entails Assumption (A3) and provides a candidate for the long time behavior of the MVPP $m_n / m_n(E)$.

Finally, using the fact that $\nu_{\mathrm{QSD}}(Q - I) = -\lambda_0 \nu_{\mathrm{QSD}}$ for some $\lambda_0 > 0$ (this is a classical property of quasi-stationary distributions, see for instance [56]) and hence that $\nu_{\mathrm{QSD}} R$ is proportional to ν_{QSD}, Theorem 1 entails that if $\sum_{k\geq 0} \mathrm{e}^{ak} m_0(k)$ is finite, then

$$\frac{m_n}{m_n(E)} \xrightarrow[n\to+\infty]{\text{a.s.}} \frac{\nu_{\mathrm{QSD}} R}{\nu_{\mathrm{QSD}} R(E)} = \nu_{\mathrm{QSD}},$$

with respect to the topology of weak convergence.

2.2. *Random trees.* As discussed in Janson [37], Examples 7.5 and 7.6, infinitely-many-color urns are particularly useful for the study of some functionals of random trees; we give below two examples where our main result applies, and gives stronger convergence results.

2.2.1. *Outdegree profiles.*

DEFINITION 1. We define the out-degree profile of a rooted tree τ as

$$\mathrm{Out}(\tau) = \sum_{\nu\in\tau} \delta_{\mathrm{outdeg}(\nu)},$$

where for all node ν in τ, $\mathrm{outdeg}(\nu)$ is the out-degree of ν (i.e., its number of children).

Out-degree profile in the random recursive tree. The random recursive tree $(\mathrm{RRT}_n)_{n\geq 1}$ is a sequence of random rooted trees defined recursively as follows:

- RRT_1 has one node (the root);
- we build RRT_{n+1} from RRT_n by choosing a node of RRT_n uniformly at random, and adding a child to this node.

It is straightforward to see that the sequence $(\mathrm{Out}(\mathrm{RRT}_n))_{n\geq 1}$ of the out-degree profile of the random recursive tree is a MVPP on $\mathbb{N}$ of initial composition $m_1 = \delta_0$, and replacement kernel

$$R_x = -\delta_x + \delta_0 + \delta_{x+1} \quad (\forall x \geq 0).$$

Note that the replacement measures R_x are not positive, but the process satisfies Assumption (T) by definition and thus this MVPP falls into the framework of Section 1.4. In this case, $P_x = \delta_x$, and $R^{(i)} = R = Q$ almost surely for all $i \geq 1$. Note that $Q_x(\mathbb{N}) = 1$ for all $x \in \mathbb{N}$, and, therefore, Assumption (A1) holds with $\mu = \delta_1$ and $c_1 = 1$.

Fix $\varepsilon \in (0, 1/2)$ and let $V(x) = (2-\varepsilon)^x$ for all $x \geq 0$; Assumption (A′2-i) holds, and we have

$$Q_x \cdot V = -(2-\varepsilon)^x + 1 + (2-\varepsilon)^{x+1} = 1 + (1-\varepsilon)V(x),$$

for all $q \in (1, 2]$,

$$\begin{aligned} Q_x \cdot V^{1/q} &= -(2-\varepsilon)^{x/q} + 1 + (2-\varepsilon)^{(x+1)/q} = 1 + \big((2-\varepsilon)^{1/q} - 1\big)V(x)^{1/q} \\ &\leq 1 + (1-\varepsilon)V(x)^{1/q}, \end{aligned}$$

since $1/q < 1$ and $2 - \varepsilon > 1$. Therefore, Assumption (A′2-ii) is satisfied with $\theta = 1 - \varepsilon$ and $K = 1$. Note that, for all continuous function $f : \mathbb{N} \to \mathbb{R}$ bounded by 1, we have, for all $q' \in (1, 3]$

$$|Q_x \cdot f|^{q'} \leq \big|1 - f(x) + f(x+1)\big|^{q'} \leq 3^{q'} \leq 27V(x),$$

since $1 \leq V(x)$ for all $x \in \mathbb{N}$. Therefore, since $Q^{(i)} = R^{(i)} = R = Q$ almost surely for all $i \geq 1$, Assumption (A′2-iii) holds with $A = 27$. Using again that $V(x) \geq 1$ for all $x \in \mathbb{N}$, we have

$$|Q_x \cdot V| = 1 + (1-\varepsilon)V(x) \leq (2-\varepsilon)V(x),$$

and, for all $q \in (1, 2]$,

$$|Q_x \cdot V^{1/q}|^q \leq \bigl(1 + (1-\varepsilon)V(x)^{1/q}\bigr)^q \leq 2^q V(x),$$

since $2 - \varepsilon < 2$. Therefore, Assumption (A′2-iv) holds and so does (A′2); note that p can be arbitrary in $(2, \infty)$, making q arbitrary in $(1, 2)$. Note that q' is restricted to be in $(q, 3]$.

One can check that the Markov chain of kernel $(R_x)_{x\in\mathbb{N}}$ is ergodic, with unique stationary distribution $\nu_x = 2^{-x-1}$ ($\forall x \geq 0$). By [49], we obtain the uniform convergence to ν required in Assumption (A3). Finally, (A4) holds since $E = \mathbb{N}$ is discrete.

Therefore, Theorem 1 applies and gives that

$$n^{-1}\,\mathrm{Out}(\mathrm{RRT}_n) \to \nu \quad \text{weakly, almost surely when } n \to \infty, \tag{4}$$

since $\nu R = \nu$. Different versions of this result can be found in the literature: Bergeron, Flajolet and Salvy [12], Corollary 4, prove it using generating functions, Mahmoud and Smythe [42] prove a joint central limit theorem for the number of nodes of out-degree 0, 1 and 2, Janson [37], Example 7.5, extends this result by considering out-degrees $0, 1, \ldots, M$ for all $M \geq 0$, which implies (4). The approach of [42] and [37] relies on the remarkable fact that, in that particular example, one can reduce the problem to finitely many types.

Our main contribution for this example is to prove the convergence in a stronger sense, and thus answer a question of Janson (see Remark 1.2 [38]). Indeed, Theorem 1 also gives that, for all $q \in (1, 2)$,

$$\sup_n \frac{\mathrm{Out}(\mathrm{RRT}_n)}{n} \cdot V^{1/q} < +\infty,$$

since $P_x = \delta_x$ for all x, in this example. Therefore,

PROPOSITION 1. *For all $\varepsilon \in (0, 1/2)$, for all $q \in (1, 2)$, for all functions $f : \mathbb{N} \to \mathbb{R}$ such that $f(x) = o((2-\varepsilon)^{x/q})$ when $x \to \infty$, we have*

$$\frac{1}{n}\int f \,\mathrm{d}\,\mathrm{Out}(\mathrm{RRT}_n) \to \sum_{x=0}^{\infty} 2^{-x-1} f(x) \quad \textit{almost surely when } n \to \infty.$$

Our approach also has the advantage of providing a framework that can be easily generalized, as, for example, in the next application to which Janson's finitely-many-types approach wouldn't apply.

Out-degree profile in a random recursive forest with multiple children. Let us now consider the following generalization of the random recursive tree studied above. The random recursive forest $(\mathrm{RRF}_n)_{n\geq 1}$ with multiple children is defined as a sequence of random rooted forests defined recursively as follows: consider a probability measure α on $\{-1\} \cup \{1, 2, \ldots\}$ (with $0 < \alpha_{-1} < 1$) and a probability measure β on $\{1, 2, \ldots\}$;

- RRF_1 has one node (the root);
- we build RRF_{n+1} from RRF_n by choosing a node of RRF_n uniformly at random, and, if this node has at least one child,
  - with probability α_{-1}, remove the edge between the node and one of his children (hence forming an other tree in the forest),

– with probability α_k ($k \geq 1$), add k children to this node,
while, if this node has 0 child, with probability β_k ($k \geq 1$), add k children to this node.

We define $\mathrm{Out}(\mathrm{RRF}_n)$ as the sum of the out-degree profiles (see Definition 1) of the trees composing the forest RRF_n.

PROPOSITION 2. *Assume that α and β both admit an exponential moment of order λ, for some fixed $\lambda > 0$. There exists a probability distribution ν_{QSD} such that, for all $q \in (1,2)$, for all $a > 0$ satisfying*

$$\sum_{k=1}^{+\infty} \alpha_k e^{ak} < 2 \sum_{k=1}^{\infty} \alpha_k,$$

and for all function $f : E = \mathbb{N} \to \mathbb{R}$ such that $f(x) = o(e^{ax/q})$ when $x \to \infty$, we have

$$\int f \frac{\mathrm{d}\,\mathrm{Out}(\mathrm{RRF}_n)}{\mathrm{Out}(\mathrm{RRF}_n)(E)} \to \int f \,\mathrm{d}\nu_{\mathrm{QSD}} \quad \textit{almost surely when } n \to \infty. \tag{5}$$

PROOF. It is straightforward to see that the sequence $(\mathrm{Out}(\mathrm{RRF}_n))_{n\geq 1}$ of the out-degree profile of the random recursive forest is a MVPP on $\mathbb{N}$ of initial composition $m_0 = \delta_0$, and random replacement kernel given, for all $x \geq 1$ by

$$R_x^{(i)} = \begin{cases} -\delta_x + \delta_{x-1} & \text{with probability } \alpha_{-1}, \\ -\delta_x + k\delta_0 + \delta_{x+k} & \text{with probability } \alpha_k, \text{for all } k \geq 1, \end{cases}$$

and

$$R_0^{(i)} = (k-1)\delta_0 + \delta_k \quad \text{with probability } \beta_k, \text{for all } k \geq 1.$$

In particular, for all $x \geq 1$,

$$R_x = -\delta_x + \sum_{k=1}^{\infty} k\alpha_k \delta_0 + \alpha_{-1}\delta_{x-1} + \sum_{k=1}^{\infty} \alpha_k \delta_{x+k},$$

and

$$R_0 = \sum_{k=1}^{\infty} (k-1)\beta_k \delta_0 + \sum_{k=1}^{\infty} \beta_k \delta_k.$$

We deduce that, for all $x \geq 1$, $R_x(E) = M_\alpha := \sum_{k\in\mathbb{N}\cup\{-1\}} |k|\alpha_k$ (the first absolute moment of α) and $R_0(E) = M_\beta := \sum_{k\in\mathbb{N}} k\beta_k$ (the mean of β). From now on, we consider the MVPP m_n^M with replacement kernel $\bar{R}_x^{(i)} := \frac{1}{M} R_x^{(i)}$, where $M = M_\alpha \vee M_\beta$. Although the replacement measures $\bar{R}^{(i)}$ are not positive, the process satisfies Assumption (T) by definition and thus this MVPP falls into the framework of Section 1.4, with weight kernel $\bar{P}_x = \delta_x$ and $\bar{Q}_x^{(i)} = \frac{1}{M} R_x^{(i)}$ for all $x \geq 0$.

For any fixed $p > 2$ and $q = \frac{p}{p-1} \in (1,2)$, we have, for all $i \geq 1$, for all $x \geq 1$,

$$\bar{Q}_x^{(i)}(E) = \begin{cases} \dfrac{k}{M} & \text{with probability } \alpha_k, \text{for all } k \geq 1, \\ 0 & \text{with probability } \alpha_{-1}, \end{cases}$$

and $\bar{Q}_0^{(i)}(E) = k/M$ with probability β_k for all $k \geq 1$. Thus, if we set

$$\mu = \alpha_{-1}\delta_0 + \left(\sum_{k=1}^{\infty} \alpha_k\right)\delta_{1/M},$$

we get that $c_1 = \int x \,\mathrm{d}\mu(x) = \sum_{k\geq 1} \alpha_k / M > 0$, and thus Assumption (A1) holds.

Let us now check that (A′2) holds with $V(x) = \mathrm{e}^{ax}$, where $a \in (0, \lambda)$ satisfies

$$\sum_{k=1}^{+\infty} \alpha_k e^{ak} < 2 \sum_{k=1}^{\infty} \alpha_k.$$

Assumption (A′2-i) is straightforward. Moreover, we have, for all $x \geq 1$,

$$\begin{aligned}\bar{Q}_x \cdot V &= \frac{-1}{M} V(x) + \frac{\sum_{k=1}^{\infty} k\alpha_k}{M} V(0) + \frac{1}{M}\alpha_{-1} V(x-1) + \frac{1}{M}\sum_{k=1}^{+\infty} \alpha_k V(x+k) \\ &\leq \frac{1}{M}\left(-1 + \alpha_{-1}\mathrm{e}^{-a} + \sum_{k=1}^{+\infty} \alpha_k \mathrm{e}^{ak}\right) V(x) + 1 \\ &\leq \frac{\sum_{k\geq 1} \alpha_k e^{ak} - \sum_{k\geq 1} \alpha_k}{\sum_{k\geq 1} \alpha_k} c_1 V(x) + 1,\end{aligned}$$

where $\frac{\sum_{k\geq 1} \alpha_k e^{ak} - \sum_{k\geq 1} \alpha_k}{\sum_{k\geq 1} \alpha_k} < 1$ by assumption. Similarly,

$$\begin{aligned}\bar{Q}_x \cdot V^{1/q} &\leq \frac{1}{M}\left(-1 + \alpha_{-1}\mathrm{e}^{-a/q} + \sum_{k=1}^{+\infty} \alpha_k \mathrm{e}^{ak/q}\right) V^{1/q}(x) + 1 \\ &\leq \frac{\sum_{k\geq 1} \alpha_k e^{ak} - \sum_{k\geq 1} \alpha_k}{\sum_{k\geq 1} \alpha_k} c_1 V^{1/q}(x) + 1,\end{aligned}$$

so that (A′2-ii) is satisfied with $\theta = \frac{\sum_{k\geq 1} \alpha_k e^{ak} - \sum_{k\geq 1} \alpha_k}{\sum_{k\geq 1} \alpha_k} c_1 \in (0, c_1)$ and $K = 1 \vee \bar{Q}_0 \cdot V$. For all $x \geq 1$, for all $q' > 1$, and for all function $f : \mathbb{N} \to \mathbb{R}$ continuous and bounded by 1, we have

$$\begin{aligned}|\bar{Q}_x \cdot f|^{q'} &= \frac{1}{M^{q'}}\left|-f(x) + \alpha_{-1} f(x-1) + \left(\sum_{k\geq 1} k\alpha_k\right) f(0) + \sum_{k\geq 1} \alpha_k f(x+k)\right|^{q'} \\ &\leq A_1^{q'} V(x),\end{aligned} \tag{6}$$

where $A_1 = 1 \vee (3 + M_\alpha/M)$, since $V(x) \geq 1$ for all $x \geq 0$; we also have

$$\begin{aligned}|\bar{Q}_0 \cdot f|^{q'} &= \frac{1}{M^{q'}}\left|\left(\sum_{k\geq 1} (k-1)\beta_k\right) f(0) + \sum_{k\geq 1} \beta_k f(k)\right| \\ &\leq \left(\frac{M_\beta}{M}\right)^{q'} \leq 1 \leq A_1 V(0),\end{aligned} \tag{7}$$

since $V(0) \geq 1$ and $A_1 \geq 1$ by definition. We also have that, for all $r > 1$,

$$\begin{aligned}\mathbb{E}\left[\left|\bar{R}_x^{(i)} \cdot f - \bar{R}_x \cdot f\right|^r\right] &\leq \mathbb{P}\left(\left|\bar{R}_x^{(i)} \cdot f - \bar{R}_x \cdot f\right| \leq 1\right) + 2^{r-1}\mathbb{E}\left[\left|\bar{R}_x^{(i)} \cdot f\right|^r + |\bar{R}_x \cdot f|^r\right] \\ &\leq 1 + 2^{r-1}\mathbb{E}\left[\left|\bar{R}_x^{(i)} \cdot f\right|^r\right] + 2^{r-1} A_1^r V(x),\end{aligned}$$

because of equations (6) and (7) applied to the special case $q' = r$. Note that

$$\begin{aligned}&\mathbb{E}\left[\left|\bar{R}_x^{(i)} \cdot f\right|^r\right] \\ &\quad= \frac{\alpha_{-1}}{M}\left|-f(x) + f(x-1)\right|^r + \sum_{k\geq 1} \frac{\alpha_k}{M}\left|-f(x) + kf(0) + f(x+k)\right|^r \\ &\quad\leq \frac{2^r \alpha_{-1} + \sum_{k\geq 1} (2+k)^r \alpha_k}{M} =: A_{2,r} < +\infty,\end{aligned}$$

since α admits an exponential moment, and therefore has finite polynomial moments. Therefore, using again that V is bounded from below by 1, we get that

$$\mathbb{E}\big[\big|\bar{R}_x^{(i)}\cdot f-\bar{R}_x\cdot f\big|^r\big]\le\big(1+2^{r-1}A_1+2^{r-1}A_{2,r}\big)V(x),$$

for all $x\ge 1$. A similar reasoning, using that β also has exponential moments, implies that

$$\mathbb{E}\big[\big|\bar{R}_0^{(i)}\cdot f-\bar{R}_0\cdot f\big|^r\big]\le\big(1+2^{r-1}A_1+2^{r-1}A_{3,r}\big)V(0),$$

where $A_3=\sum_{k\ge 1}\beta_k k^r$. Since $\bar{R}^{(i)}=\bar{Q}^{(i)}$ almost surely, we obtain

$$\mathbb{E}\big[\big|\bar{Q}_0^{(i)}\cdot f-\bar{Q}_0\cdot f\big|^p\big]\le\big(1+2^{p-1}A_1+2^{p-1}A_{3,p}\big)V(0).$$

Therefore, setting $A=1+2^{p-1}A_1+2^{p-1}(A_{2,r}\vee A_{3,r}\vee A_{3,p})$, we can conclude that Assumption (A′2-iii) holds.

Finally, let us check Assumption (A′2-iv): for all $x\ge 1$, for all $\ell>1$ and $s\le 2$, we have

$$\begin{aligned}\big|\bar{Q}_x\cdot V^{1/\ell}\big|^s&=\frac{1}{M^s}\left|-V(x)^{1/\ell}+V(0)^{1/\ell}+\alpha_{-1}V(x-1)^{1/\ell}+\sum_{k=1}^{+\infty}\alpha_k V(x+k)^{1/\ell}\right|^s\\ &\le\frac{1}{M^s}\left(2+\alpha_{-1}+\sum_{k=1}^{+\infty}\alpha_k e^{\lambda k/\ell}\right)^s V(x)^{s/\ell}\qquad(8)\\ &\le\left(\frac{3+\sum_{k\ge 1}\alpha_k e^{\lambda k}}{M}\right)^s V(x)^{s/\ell},\end{aligned}$$

and, for all $r\in(1,q)$,

$$\begin{aligned}\mathbb{E}\big[\big|\bar{Q}_x^{(i)}\cdot V^{1/q}\big|^r\big]&=\frac{1}{M^r}\bigg(\alpha_{-1}\big|-e^{ax/q}+1+e^{a(x-1)/q}\big|^r\\ &\qquad+\sum_{k=1}^{+\infty}\alpha_k\big|-e^{ax/q}+k+e^{a(x+k)/q}\big|^r\bigg)\qquad(9)\\ &\le\frac{V(x)^{1/q}}{M^r}\bigg(\alpha_{-1}3^r+3^{r-1}\sum_{k\ge 1}\alpha_k\big(1+k^q+e^{ak}\big)\bigg)\\ &\le B_1V(x)^{1/q},\end{aligned}$$

where $B_1=3^2(\alpha_{-1}+\sum_{k\ge 1}\alpha_k(1+k^2+e^{\lambda k})/M<+\infty$. Similar calculations hold for $x=0$; we thus now reason as if equations (8) and (9) also hold for $x=0$. Applying equation (8) to $\ell=s=1$ gives that $|Q_x\cdot V|\le B_2V(x)$ for all $x\ge 0$, where $B_2=(3+\sum_{k\ge 1}\alpha_k e^{\lambda k})/M$. Applying equation (8) to $\ell=s=q$ gives that $|Q_x\cdot V^{1/q}|^q\le B_3V(x)^{1/q}$ for all $x\ge 0$, where $B_3=((2+\sum_{k\ge 1}\alpha_k e^{\lambda k})/M)^q$. Finally, applying equation (8) to $\ell=q$ and $s=r$, and using equation (9), we get that

$$\begin{aligned}&\mathbb{E}\big[\big|\bar{Q}_x^{(i)}\cdot V^{1/q}-\bar{Q}_x\cdot V^{1/q}\big|^r\big]\\ &\quad\le 2^{r-1}\big(\mathbb{E}\big[\big|\bar{Q}_x^{(i)}\cdot V^{1/q}\big|^r\big]+\mathbb{E}\big[\big|\bar{Q}_x\cdot V^{1/q}\big|^r\big]\big)\\ &\quad\le 2^{r-1}\bigg(B_1V(x)^{r/q}+\bigg(2+\sum_{k\ge 1}\alpha_k e^{\lambda k}/M\bigg)^r V(x)^{r/q}\bigg)\\ &\quad\le B_4V(x),\end{aligned}$$

with $B_4=2^{r-1}(B_1+(2+\sum_{k\ge 1}\alpha_k e^{\lambda k}/M)^r)$, because $r/q<1$, and $V(x)\ge 1$ for all $x\ge 0$. Therefore, taking $B=B_2\vee B_3\vee B_4$, we conclude that Assumption (A′2-iv) holds.

The continuous-time pure jump Markov process X with sub-Markovian jump matrix $Q-I$ is irreducible and clearly satisfies the assumptions of Theorem 5.1 and Remark 11 in [16]. Therefore, there exist a quasi-stationary distribution ν_{QSD} for X and two positive constants Cst, $\delta > 0$ such that, for all probability measure $\alpha \in E$ satisfying $\alpha \cdot V^{1/q} < +\infty$, for all $t \geq 0$,

$$\left\|\mathbb{P}_\alpha(X_t \in \cdot \mid t < \tau_\partial) - \nu_{\mathrm{QSD}}\right\|_{\mathrm{TV}} \leq \mathrm{Cst}\,\alpha \cdot V^{1/q} \mathrm{e}^{-\delta t},$$

which entails Assumption (A3). Since Assumption (A4) is clearly satisfied, Theorem 1 applies and hence

$$\frac{\mathrm{Out}(\mathrm{RRF}_n)}{\mathrm{Out}(\mathrm{RRF}_n)(E)} \to \frac{\nu_{\mathrm{QSD}} R}{\nu_{\mathrm{QSD}} R(E)} \quad \text{weakly, almost surely when } n \to \infty. \tag{10}$$

Since $\nu_{\mathrm{QSD}} R$ is proportional to ν_{QSD}, and since we also have, again by Theorem 1,

$$\sup_n \frac{\mathrm{Out}(\mathrm{RRF}_n)}{\mathrm{Out}(\mathrm{RRF}_n)(E)} \cdot V^{1/q} < +\infty,$$

this concludes the proof of Proposition 2. □

2.2.2. *Protected nodes.* A node ν of a tree τ is 2-protected if the closest leaf is at distance at least 2 from ν; in a social network, 2-protected nodes can be users who used to invite new users to the network but have not done so recently. The proportion of such nodes in different models of random trees have been studied in the literature: Motzkin trees in Cheon and Shapiro [20], random binary search tree in Bóna [14], and more recently in the m-ary search tree in Holmgren, Janson and Šileikis [36]. Devroye and Janson [25] show how results of Aldous [1] about fringe trees can be used to study this question with a unified approach for different models of random trees, including simply generating trees and the random recursive tree. We show here how our main result allows to get information about protected nodes in random trees.

Protected nodes in the random recursive tree. For all $n \geq 1$ and $x \geq 0$, let us denote by $X_{n,x}$ the number of internal nodes in RRT_n having exactly x leaf-children. The random measure

$$m_n = \sum_{x \in \mathbb{N}} X_{n,x} \delta_x$$

is a MVPP of initial composition $m_0 = \delta_1$. The replacement kernel of $(m_n)_{n \geq 0}$ is (for all $i \geq 1$ and $x \geq 1$)

$$R_0^{(i)} = -\delta_0 + \delta_1 \quad \text{and} \quad R_x^{(i)} = B_{1/x+1}^{(i)} \delta_{x+1} + \left(1 - B_{1/x+1}^{(i)}\right)(\delta_{x-1} + \delta_1) - \delta_x,$$

where $(B_{1/x+1}^{(i)})$ is a sequence of i.i.d. random Bernoulli-distributed variables of parameters $1/x+1$ for all $x \geq 1$. The weight kernel of $(m_n)_{n \geq 0}$ is $P_x = (x+1)\delta_x$ (for all $x \in \mathbb{N}$). We therefore have

$$R_0 = -\delta_0 + \delta_1 \quad \text{and} \quad R_x = \frac{1}{x+1}\delta_{x+1} + \frac{x}{x+1}(\delta_{x-1} + \delta_1) - \delta_x,$$

and

$$Q_x = \frac{x+2}{x+1}\delta_{x+1} + \frac{x}{x+1}(x\delta_{x-1} + 2\delta_1) - (x+1)\delta_x,$$

for all $x \geq 0$. Note that $Q_x(\mathbb{N}) = 1$ for all $x \geq 0$. Let us check the assumptions of Theorem 1; (T) is satisfied by construction of the model, (A1) is satisfied with $\mu = \delta_1$ and thus $c_1 = 1$. Fix

$\varepsilon > 0$, $V(0) = V(1) = 1$, and $V(x) = \prod_{i=2}^{x}(i-\varepsilon)$ for all $x \geq 2$; (A′2-i) is clearly satisfied, and for all $x \in \mathbb{N}$,

$$\begin{aligned} Q_x \cdot V &= \frac{x+2}{x+1}V(x+1) + \frac{x}{x+1}\big(xV(x-1)+2V(1)\big) - (x+1)V(x) \\ &= V(x)\left(\frac{x+2}{x+1}(x+1-\varepsilon) + \frac{x^2}{(x+1)(x-\varepsilon)} + \frac{2x}{x+1}\frac{1}{V(x)} - x - 1\right). \end{aligned}$$

Note that, when $x \to \infty$,

$$\frac{x+2}{x+1}(x+1-\varepsilon) + \frac{x^2}{(x+1)(x-\varepsilon)} + \frac{2x}{x+1}\frac{1}{V(x)} - x - 1 = 1 - \varepsilon + o(1),$$

implying that there exists x_0 such that, for all $x \geq x_0$, $Q_x \cdot V \leq 1 - \varepsilon/2$, and thus, for all $x \geq 0$,

$$Q_x \cdot V \leq (1-\varepsilon/2)V(x) + \sup_{z \leq x_0} Q_z \cdot V.$$

The same reasoning gives that, for all $p > 2$, $q = p/(p-1) \in (1,2)$,

$$\begin{aligned} &Q_x \cdot V^{1/q} \\ &\quad = V(x)^{1/q}\left(\frac{x+2}{x+1}(x+1-\varepsilon)^{1/q} + \frac{x^2}{(x+1)(x-\varepsilon)^{1/q}} + \frac{2x}{x+1}\frac{1}{V(x)^{1/q}} - x - 1\right) \\ &\quad = V(x)^{1/q}\big(x^{1/q} + x^{1-1/q} - x + \mathcal{O}(1)\big) = -V(x)^{1/q}\big(x + o(x)\big), \end{aligned}$$

and there exists x_1 such that for all $z \geq x_1$, $Q_x \cdot V^{1/q} \leq 0$. Thus, (A′2-ii) is satisfied with $\theta = 1 - \varepsilon/2$ and $K = \sup_{z \leq x_0} Q_z \cdot V + \sup_{z \leq x_1} Q_z \cdot V^{1/q}$. Let f be a function from $\{0, 1, \ldots\}$ to $\mathbb{R}$ continuous and bounded by 1, and $r \in (1,2)$; we have

$$\begin{aligned} |Q_x \cdot f|^r &= \left|\frac{x+2}{x+1}f(x+1) + \frac{x^2}{x+1}f(x-1) + \frac{2x}{x+1}f(1) - (x+1)f(x)\right|^r \\ &\leq 4^{r-1}\left(\left(\frac{x+2}{x+1}\right)^r + \left(\frac{x^2}{x+1}\right)^r + \left(\frac{2x}{x+1}\right)^r + (x+1)^r\right). \end{aligned}$$

When $x \to \infty$, we have

$$\left(\frac{x+2}{x+1}\right)^r + \left(\frac{x^2}{x+1}\right)^r + \left(\frac{2x}{x+1}\right)^r + (x+1)^r = \big(2 + o(1)\big)x^r.$$

Note that, when $x \to \infty$, $x^r = o(x^2) = o(V(x))$, which implies that there exists a constant A such that, for all $x \geq 0$, $|Q_x \cdot f|^r \leq AV(x)$. One can check that, $R_0^{(i)} = R_0$, and, for all $i \geq 1$,

$$\big|R_x^{(i)} \cdot f - R_x \cdot f\big|^r \leq 3,$$

because a Bernoulli random variable is at most at distance 1 from its mean, almost surely. We also have

$$\begin{aligned} \mathbb{E}\big|Q_x^{(i)} \cdot f - Q_x \cdot f\big|^p = \bigg|&(x+2)f(x+1)\left(B_{1/x+1}^{(i)} - \frac{1}{x+1}\right) \\ &+ x\big(xf(x-1) + 2f(1)\big)\left(\frac{1}{x+1} - B_{1/x+1}^{(i)}\right)\bigg|^p \\ &= \big((x+2)^r + x^r(x+2)^r\big) \leq AV(x), \end{aligned}$$

for A large enough, since $x^{2r} = o(V(x))$ when $x \to \infty$. We have thus checked that (A′2-iii) holds. Assumption (A′2-iv) can be checked in the same way; we leave the details to the reader. Note that $p > 2$, and thus $q \in (1, 2)$ are arbitrary.

Set

$$\nu_0 = \frac{\mathrm{e}-2}{1+2\mathrm{e}}, \qquad \nu_1 = \frac{4(\mathrm{e}-2)}{1+2\mathrm{e}} \quad \text{and} \quad \nu_i = \frac{2(i+1)}{1+2\mathrm{e}} \sum_{j \geq i+1} \frac{1}{j!} \quad (\forall i \geq 2).$$

One can check that the Markov process with jump measure $Q - I$ is ergodic, that $\nu = (\nu_i)_{i\geq 0}$ is the unique stationary distribution of $Q - I$. Using (A2) and [49], we get that (A3) is satisfied. Therefore, our main result applies ((A4) is immediate since $E = \mathbb{N}$ is discrete) and we get that $\tilde{m}_n$ converges almost surely to $\pi := \nu R/\nu R(\mathbb{N})$. Let us denote by $\hat{\pi} = \nu R$; it is straightforward to check that

$$\hat{\pi}_0 = \frac{\mathrm{e}-2}{1+2\mathrm{e}}, \qquad \hat{\pi}_1 = \frac{2\mathrm{e}-4}{1+2\mathrm{e}} \quad \text{and} \quad \hat{\pi}_x = \frac{2}{1+2\mathrm{e}} \sum_{i \geq x+1} \frac{1}{i!},$$

and thus that $\nu R(\mathbb{N}) = \mathrm{e}/(1+2\mathrm{e})$, implying that

$$\pi_0 = 1 - \frac{2}{\mathrm{e}}, \qquad \pi_1 = 2 - \frac{4}{\mathrm{e}} \quad \text{and} \quad \pi_x = \frac{2}{\mathrm{e}} \sum_{i \geq x+1} \frac{1}{i!}.$$

We have thus proved the following:

PROPOSITION 3. *For all $x \geq 1$, the proportion $p_{n,x}$ of internal nodes having exactly x leaf-children in the n-node random recursive tree converges almost surely to*

$$\frac{2}{\mathrm{e}} \sum_{i \geq x+1} \frac{1}{i!}.$$

The proportion $p_{n,0}$ of protected internal nodes converges almost surely to $1-2/\mathrm{e}$. Moreover, for all $q \in (1, 2)$ and all function $f : \{0, 1, \dots\} \to \mathbb{R}$ such that $f(x) = o(\prod_{i=2}^{x}(i-\varepsilon)^{1/q})$ for some $\varepsilon > 0$ when $x \to \infty$, we have

$$\sum_{i \geq 0} p_{n,i} f(i) \to (1-2/\mathrm{e}) f(0) + \frac{2}{\mathrm{e}} \sum_{i \geq 1} f(i) \sum_{j \geq i+1} \frac{1}{j!}$$

almost surely when $n \to \infty$.

Note that, in the proposition above, the proportions are calculated among internal nodes only. To translate this result in terms of proportion among all nodes, we need one last calculation to take into account the leaf-nodes. Note that the limit proportion of leaves in the random recursive tree is given by

$$\frac{\sum_{i\geq 0} i\pi_i}{1 + \sum_{i \geq 0} i\pi_i} = 1/2,$$

because $\sum_{i\geq 0} i\pi_i = 1$ (this result is folklore and was already discussed in Section 2.2.1). Therefore, the proportion of nodes having exactly i leaf-children in the n-node random recursive tree converges almost surely to $\pi_i/2$: We get that, for all $i \geq 1$, the proportion of nodes having exactly i leaf-children in the n-node random recursive tree converges almost surely to

$$\frac{1}{\mathrm{e}} \sum_{j \geq i+1} \frac{1}{j!}.$$

The proportion of protected internal nodes converges almost surely to $1/2 - 1/e$. Note that the convergence in probability of the proportion of protected nodes in the random recursive tree was already proved by Ward and Mahmoud [43]; we have shown how our main result implies almost-sure convergence.

2.3. *"Sample paths" Pólya urns.* In this section we consider the case where the replacement measures are the empirical occupation measures of sample paths of Markov processes. The section is divided into three subsections: the first one is devoted to the discrete-time setting, the second to the continuous-time setting, the third one to an application to stochastic-approximation algorithms for the computation of quasi-stationary distributions.

2.3.1. *Discrete-time sample paths Pólya urns.* Let $(X_n)_{n\in\{0,1,2,\dots\}}$ be a Markov chain evolving in a Polish locally-compact state space $E \cup \{\partial\}$, where $\partial \notin E$ is an absorbing point: $X_n = \partial$ for all $n \geq \tau_\partial := \min\{k \geq 0, X_k \in \partial\}$ almost surely. We denote by $\mathbb{P}_x$ and $\mathbb{E}_x$ the law of the process X starting from $x \in E \cup \partial$ and its associated expectation. Also fix $\mathcal{T}$ a probability distribution on $\mathbb{N} \cup \{+\infty\}$ such that $\mathcal{T}(\{0\}) < 1$ and such that, if (T, X) is distributed according to $\mathcal{T} \otimes \mathbb{P}_x$, then $\tau_\partial \wedge T$ admits an exponential moment uniformly bounded with respect to $x \in E$; in other words, there exists $\lambda > 0$ such that

$$\sup_{x\in E} \mathbb{E}_x\big[\exp\big(\lambda(T \wedge \tau_\partial)\big)\big] < \infty$$

(with a slight abuse of notation, since we also denote by $\mathbb{E}_x$ the expectation under $\mathcal{T} \otimes \mathbb{P}_x$).

We consider the MVPP on E with random replacement measures $(R_x^{(i)})_{x\in E, i\geq 1}$ being i.i.d. copies of

$$R_x^{(1)} = \sum_{n=0}^{T\wedge(\tau_\partial - 1)} \delta_{X_n},$$

for all $x \in E$ and all $i \geq 0$, where (T, X) is a random variable of distribution $\mathcal{T} \otimes \mathbb{P}_x$. This means that, at each time, we add to the urn the empirical measure of a sample path of length $T \wedge (\tau_\partial - 1)$ of X. For simplicity, we consider the case without weights, that is, $P_x = \delta_x$ for all $x \in E$, so that $Q^{(i)} = R^{(i)}$. Note that the mass of $R_x^{(i)}$ is random, equal in law to $(T + 1) \wedge \tau_\partial$ under $\mathcal{T} \otimes \mathbb{P}_x$, and is not uniformly bounded in general (although its expectation is, by assumption, uniformly bounded with respect to x). In particular, the considered MVPP is unbalanced.

To ensure the convergence of this MVPP, we assume that the following particular instance of the assumptions of Theorem 2.1 in [16] is satisfied. This abstract criterion ensures the existence of a quasi-stationary distribution for X; we will show later many examples that fall into this framework.

ASSUMPTION (E). There exist a positive integer n_1, positive real constants $\alpha_0, \alpha_1, \alpha_2, \alpha_3$, a locally bounded function with compact level sets $V : E \to [1, +\infty)$ and a probability measure π on a compact subset $K \subset E$ such that:

(E1) (*Local Dobrushin coefficient*). For all $x \in K$,

$$\mathbb{P}_x(X_{n_1} \in \cdot) \geq \alpha_0 \pi(\cdot \cap K).$$

(E2) (*Global Lyapunov criterion*). We have $\alpha_1 < \alpha_2$ and, for all $x \in E$,

$$\mathbb{E}_x V(X_1) \leq \alpha_1 V(x) + \alpha_3 \mathbf{1}_K(x) \text{ and } \mathbb{P}_x(1 < \tau_\partial) \geq \alpha_2.$$

(E3) (*Local Harnack inequality*). We have

$$\sup_{n\in\mathbb{Z}_+} \frac{\sup_{y\in K} \mathbb{P}_y(n < \tau_\partial)}{\inf_{y\in K} \mathbb{P}_y(n < \tau_\partial)} \leq \alpha_3.$$

(E4) (*Aperiodicity/irreducibility*). For all $x \in E$, there exists $n_4(x)$ such that, for all $n \geq n_4(x)$,

$$\mathbb{P}_x(X_n \in K) > 0.$$

Under Assumption (E), it is proved in [16] that X admits one and only one quasi-stationary distribution ν_{QSD} such that $\nu_{\mathrm{QSD}} \cdot V < +\infty$ and which corresponds to the so-called minimal quasi-stationary distribution (or Yaglom limit). It is also proved in [16] that there exist two positive constants $C > 0$, $\delta > 0$ such that, for all $t \geq 0$,

$$\left\|\mathbb{P}_\alpha(X_t \in \cdot \mid X_t \notin \partial) - \nu_{\mathrm{QSD}}\right\|_{\mathrm{TV}} \leq C\alpha \cdot V \mathrm{e}^{-\delta t}.$$

PROPOSITION 4. *Under Assumption (E), if $x \mapsto \mathbb{E}_x f(X_1)$ is continuous on E for all continuous bounded function $f : E \to \mathbb{R}$ and if $m_0 \cdot V < \infty$, then the normalized sequence of probability measures $(\tilde{m}_n)_{n\in\mathbb{N}}$ associated to the MVPP with random replacement kernel $(R^{(i)})_{i\geq 1}$ converges almost surely to the quasi-stationary distribution ν_{QSD} of X in $\mathcal{P}(E)$.*

Before turning to the proof of Proposition 4, we provide typical examples that satisfy Assumption (E) and consequently fall into the framework of Proposition 4.

EXAMPLE 1. If E is finite and X is irreducible in E (i.e., $\exists n \geq 1$ s.t. $\mathbb{P}_x(X_n = y) > 0$ for all $x, y \in E$) and $\mathbb{P}_x(\tau_\partial < +\infty) = 1$ for all $x \in E$, then Assumption (E) is satisfied for any probability distribution $\mathcal{T}$ (one simply chooses $K = E$ and $V = 1$).

EXAMPLE 2. Consider the case $E = \mathbb{N}$ and X is a discrete-time birth-and-death process with transition probabilities given by

$$\mathbb{P}_x(X_1 = y) = \begin{cases} b_x & \text{if } y = x + 1, \\ d_x & \text{if } y = x - 1, \\ \kappa_x & \text{if } y = \partial, \end{cases}$$

where $(b_x)_{x\in\mathbb{N}}$, $(d_x)_{x\in\mathbb{N}}$, $(\kappa_x)_{x\in\mathbb{N}}$ are families of nonnegative numbers such that $b_x + d_x + \kappa_x = 1$ for all $x \in \mathbb{N}$, $d_0 = 0$ and $\inf_{x\geq 1} d_x > 0$ for all $x \geq 1$. If

$$b_x \to 0 \quad \text{when } x \to +\infty,$$

then Assumption (E) is satisfied for any probability distribution $\mathcal{T}$ such that there exists $\lambda > 0$ satisfying $\mathbb{E}\mathrm{e}^{\lambda T} < +\infty$ (where the random variable T has distribution $\mathcal{T}$). To see this, one simply chooses K large enough and $V(x) = \mathrm{e}^{ax}$ with $a > 0$ large enough.

EXAMPLE 3. Assume that $(X_n)_{n\geq 0}$ is a d-type Galton–Watson process. We recall that such a process X evolves in $\mathbb{N}^d = E \cup \{\partial\}$ and is absorbed at $\partial = (0, \dots, 0)$. Also, for all $n \geq 0$ and $i \in \{1, \dots, d\}$, we have

$$X_{n+1}^i = \sum_{k=1}^{d} \sum_{\ell=1}^{X_n^k} \zeta_{k,i}^{(n,\ell)},$$

where $(\zeta_{k,1}^{(n,\ell)}, \dots, \zeta_{k,d}^{(n,\ell)})_{n,\ell,k}$ is a family of independent random variables in $\mathbb{N}^d$ such that, for all $k \in \{1, \dots, d\}$, $(\zeta_{k,1}^{(n,\ell)}, \dots, \zeta_{k,d}^{(n,\ell)})_{n,\ell}$ is an independent and identically distributed family. We assume that the matrix of mean offspring denoted by $M = (M_{k,i})_{1 \le k,i \le d}$ and defined by

$$M_{k,i} = \mathbb{E}\zeta_{k,i}^{(n,\ell)} \quad \forall k, i \in \{1, \dots, d\},$$

is finite and that there exists $n \ge 1$ such that $M_{k,i}^n > 0$ for all $k, i \in \{1, \dots, d\}$. Let v be a positive right eigenvector of the matrix M and denote by $\rho(M)$ its spectral radius.

We assume that X is subcritical (i.e., $\rho(M) < 1$), aperiodic, and irreducible. Then, if there exists $\alpha > 0$ such that $\mathbb{E}[\exp(\alpha|X_1|) \mid X_0 = (1, \dots, 1)] < \infty$, then X satisfies Assumption (E). To check this, one simply observes that $\inf_{x \in E} \mathbb{P}_x(1 < \tau_\partial) > 0$ and carefully checks that there exists $\varepsilon > 0$ small enough and K large enough so that Assumption (E) is satisfied with $V : x \in E \mapsto e^{\varepsilon\langle v, x\rangle}$.

EXAMPLE 4. Assume that X evolves in $E = \mathbb{R}^d$ according to the following perturbed dynamical systems

$$X_{n+1} = f(X_n) + \xi_n,$$

where $f : \mathbb{R}^d \to \mathbb{R}^d$ is a measurable function such that $|x| - |f(x)| \to +\infty$ when $|x| \to +\infty$, $(\xi_n)_{n \in \mathbb{N}}$ is an i.i.d. sequence of Gaussian random variables with positive density in $\mathbb{R}^d$. We assume that the process evolves in a measurable set E of $\mathbb{R}^d$: it is immediately sent to $\partial \notin \mathbb{R}^d$ as soon as $X_n \notin E$. If E is such that

$$\inf_{x \in E} \mathbb{P}\big(f(x) + \xi_1 \in E\big) > 0,$$

then Assumption (E) is satisfied. This result is obtained by observing that $\inf_{x \in E} \mathbb{P}_x(1 < \tau_\partial) > 0$, by choosing K a large enough ball and $V(x) = e^{|x|}$ (see [16], Example 9, for more details).

PROOF OF PROPOSITION 4. For all $n \ge 0$, let $\hat{m}_n = m_n / \sup_{x \in E} R_x(E)$. First note that $\hat{m}_n$ is well defined since $\sup_{x \in E} R_x(E) \le \sup_{x \in E} \mathbb{E}_x[T \wedge \tau_\partial] < +\infty$, by assumption on the existence of a uniform exponential moment for $T \wedge \tau_\partial$. Moreover, $(\hat{m}_n)_{n \ge 0}$ is an MVPP of replacement kernel $\hat{R}^{(i)} = R^{(i)} / \sup_{x \in E} R_x(E)$ and weight kernel $\hat{P}_x = \delta_x$ (for all $x \in E$). Let us check that Assumption (A) is satisfied by $(\hat{m}_n)_{n \ge 0}$. Note that, for all $x \in E$ and all bounded measurable function $f : E \to \mathbb{R}$,

$$R_x \cdot f := \mathbb{E}\big[R_x^{(i)}\big] \cdot f = \mathbb{E}\left[\sum_{n=0}^{T} G_n \cdot f(x)\right],$$

where $G_n \cdot f(x) = \mathbb{E}_x[f(X_n)\mathbf{1}_{n < \tau_\partial}]$ is the sub-Markovian semigroup of the absorbed process X.

Moreover, we have that

$$\hat{R}_x(E) \ge \frac{\mathbb{E}[\sum_{n=0}^{T} \alpha_2^n]}{\sup_{y \in E} R_y(E)} = \frac{1 - \mathbb{E}[\alpha_2^{T+1}]}{(1 - \alpha_2)\sup_{y \in E} R_y(E)} =: c_1 > 0,$$

so that Assumption (A1) is satisfied (take μ the law of $\frac{(T+1) \wedge \Delta}{\sup_{y \in E} R_y(E)}$, where T and Δ are independent and Δ is distributed with respect to a geometric law with parameter $1 - \alpha_2$ on $\{1, 2, \dots\}$). Moreover, we deduce from (E) that, for some constant $C > 0$,

$$\begin{aligned}
&G_n \cdot V(x) \\
&\quad \le \alpha_1^n V(x) + C\big(G_{n-1} \cdot \mathbf{1}_E(x) + \alpha_1 G_{n-2} \cdot \mathbf{1}_E(x) + \cdots + \alpha_1^{n-1}\mathbf{1}_E(x)\big) \\
&\quad \le \alpha_1^n V(x) + \frac{C G_n \cdot \mathbf{1}_E(x)}{\alpha_2}\left(1 + \frac{\alpha_1}{\alpha_2} + \cdots + \frac{\alpha_1^{n-1}}{\alpha_2^{n-1}}\right),
\end{aligned}$$

where we used (E2) and Markov's property for the second inequality. Since $\alpha_1 < \alpha_2$ by assumption, then there exists some constant C' such that

$$R_x \cdot V = \mathbb{E}\left[\sum_{n=0}^{T} G_n \cdot V(x)\right] \leq \mathbb{E}\left[\sum_{n=0}^{T} \alpha_1^n\right] V(x) + C'\mathbb{E}\left[\sum_{n=0}^{T} G_n \cdot \mathbf{1}_E(x)\right]$$
$$= \mathbb{E}\left[\sum_{n=0}^{T} \alpha_1^n\right] V(x) + C' \sup_{y\in E} \mathbb{E}_y[T \wedge \tau_\partial].$$

We thus get

$$\hat{R}_x \cdot V \leq \theta V(x) + \frac{C' \sup_{y\in E} \mathbb{E}_y[T \wedge \tau_\partial]}{\sup_{y\in E} R_y(E)},$$

where

$$\theta := \frac{1 - \mathbb{E}[\alpha_1^{T+1}]}{(1-\alpha_1)\sup_{y\in E} R_y(E)} < c_1.$$

Assumption (A2-ii) is thus satisfied by $\hat{R}$. Assumption (A2-iii) is satisfied for any $p > 2 \vee \frac{\ln\theta}{\ln\theta - \ln c_1}$ since $R_x^{(1)}(E) \leq T \wedge \tau_\partial$, which admits a uniformly bounded exponential moment by assumption. Since (A2-i) is assumed to be true under (E), we deduce that Assumption (A2) is implied by Assumption (E).

To prove that (A3) holds true, it is sufficient, by Theorem 2.1 in [16], to prove that $\hat{R}$ satisfies Assumption (E) with Lyapunov function $V^{1/q}$. Since $T \geq 1$ with positive probability, and since X satisfies Assumption (E1), we get that $\hat{R}$ also satisfies Assumption (E1). We have already proved that $\hat{R}$ satisfies Assumptions (A1-2) with Lyapunov function V and hence with Lyapunov function $V^{1/q}$ (see the proof of Lemma 1), which implies that $\hat{R}$ satisfies Assumption (E2) with Lyapunov function $V^{1/q}$. Moreover, for all $n \geq 0$ and all $x, y \in K$, we have

$$R_x^n(E) = \mathbb{E}\left[\sum_{\ell=1}^{n} \sum_{i_\ell=0}^{T_\ell} G_{i_1+\dots+i_n} \cdot \mathbf{1}_E(x)\right]$$
$$\leq a_3\mathbb{E}\left[\sum_{\ell=1}^{n} \sum_{i_\ell=0}^{T_\ell} G_{i_1+\dots+i_n} \cdot \mathbf{1}_E(y)\right] = a_3 R_y^n(E),$$

where $T_1, \dots, T_n$ are i.i.d. random variables with distribution $\mathcal{T}$ and where we used Assumption (E3) for X; this implies that Assumption (E3) is satisfied by $\hat{R}$. The fact that $\hat{R}$ satisfies Assumption (E4) is an immediate consequence of (E4) for X, since $T \geq 1$ with positive probability. By Theorem 2.1 in [16], this implies that the discrete-time Markov process with transition probabilities given by $\hat{R}$ admits a unique quasi-stationary distribution ν such that $\nu \cdot V^{1/q} < +\infty$. More precisely, it implies that there exist $\alpha \in (0, 1)$ and $C > 0$ such that, for any probability measure μ on E such that $\mu \cdot V^{1/q} < +\infty$,

$$\left\| \frac{\mu\hat{R}^n}{\mu\hat{R}^n(E)} - \nu \right\|_{\mathrm{TV}} \leq C\alpha^n \mu \cdot V^{1/q}.$$

In particular, for all measurable set $A \subset E$,

$$\left|\mu\hat{R}^n(A) - \mu\hat{R}^n(E)\nu(A)\right| \leq C\mu\hat{R}^n(E)\alpha^n \mu \cdot V^{1/q},$$

and hence that for all $t \geq 0$,

$$\left|\mu e^{t\hat{R}}(A) - \mu e^{t\hat{R}}(E)\nu(A)\right| \leq C\mu e^{t\alpha\hat{R}}(E)\mu \cdot V^{1/q}.$$

Since $\alpha \in (0,1)$, $\sum_{n=0}^{+\infty} \frac{t^n \alpha^n}{n!} \mu \hat{R}^n(E)$ is negligible in front of $\sum_{n=0}^{+\infty} \frac{t^n}{n!} \mu \hat{R}^n(E)$ when $t \to +\infty$, so that

$$\left| \frac{\mu e^{t\hat{R}}(A)}{\mu e^{t\hat{R}}(E)} - \nu(A) \right| \leq C \frac{\mu e^{t\alpha \hat{R}}(E)}{\mu e^{t\hat{R}}(E)} \mu \cdot V^{1/q} \to 0 \quad \text{when } t \to +\infty.$$

Note that $\mu e^{t\hat{R}}(A)/\mu e^{t\hat{R}}(E)$ is the law of the continuous-time process with sub-Markovian jump kernel $\hat{R} - \mathrm{Id}$ at time t conditioned not to be absorbed at time t. Therefore, we can conclude that (A3) is satisfied by $\hat{R}$.

Finally, the continuity of $\hat{R}_x$ with respect to x directly derives from the continuity of $\delta_x G_1$ with respect to x and from the uniform boundedness of $\mathbb{E}_x[e^{\lambda T \wedge \tau_\partial}]$ with respect to x. Therefore, Theorem 1 applies and gives that $\hat{m}_n / \hat{m}_n(E) = \tilde{m}_n$ converge almost surely (for the topology of weak convergence) to a probability measure ν. This distribution ν is the unique quasi-stationary distribution of the process of sub-Markovian jump kernel $\hat{R} - I$ such that $\nu \cdot V^{1/q} < +\infty$.

It only remain to show that ν is indeed equal to ν_{QSD}, the unique quasi-stationary distribution of X such that $\nu_{\mathrm{QSD}} \cdot V < +\infty$. Since ν_{QSD} is a quasi-stationary distribution for X, we have

$$\nu_{\mathrm{QSD}} R \cdot f = \mathbb{E} \sum_{n=0}^{T} \nu_{\mathrm{QSD}} G_n \cdot f = \mathbb{E} \sum_{n=0}^{T} \theta_0^n \nu_{\mathrm{QSD}} \cdot f = \mathbb{E}\left[\sum_{n=0}^{T} \theta_0^n \right] \nu_{\mathrm{QSD}} \cdot f,$$

where $\theta_0 := \nu_{\mathrm{QSD}} G_1(E)$. This implies that ν_{QSD} is a quasi-stationary distribution of the discrete-time sub-Markov process of transitions $\hat{R}$. Moreover, since $\nu_{\mathrm{QSD}} \cdot V < +\infty$, $V \geq 1$ and $1/q < 1$, we have $\nu_{\mathrm{QSD}} \cdot V^{1/q} < +\infty$, implying that $\nu = \nu_{\mathrm{QSD}}$, by uniqueness of ν. □

2.3.2. *Continuous-time sample paths Pólya urns.* Let $(X_t)_{t \in [0,+\infty)}$ be the solution in $E = \mathbb{R}^d$ to the stochastic differential equation

$$dX_t = dB_t + b(X_t)\, dt,$$

where B is a standard d-dimensional Brownian motion and $b : \mathbb{R}^d \mapsto \mathbb{R}^d$ is locally Hölder-continuous in $\mathbb{R}^d$. We assume that X is subject to an additional soft killing $\kappa : \mathbb{R}^d \mapsto [0, +\infty)$, which is continuous and uniformly bounded: the process is sent to a cemetery point $\partial \notin \mathbb{R}^d$ at rate $\kappa(X_t)$ and we denote by τ_∂ the hitting time of ∂ by X. As in the discrete-time case, we denote by $\mathbb{P}_x$ and $\mathbb{E}_x$ the law of the process X starting from $x \in E \cup \partial$ and its associated expectation, and we consider $\mathcal{T}$ a probability distribution on $[0, +\infty]$ such that $\tau_\partial \wedge T$ admits under $\mathcal{T} \otimes \mathbb{P}_x$ an exponential moment uniformly bounded with respect to $x \in E$.

We consider the unbalanced MVPP on E without weights and with random replacement kernels $(R^{(i)})_{i \geq 1}$ being i.i.d. copies of

$$R_x^{(1)} = \int_0^{T \wedge \tau_\partial} \delta_{X_t}\, dt \quad (\forall x \in E),$$

where (T, X) is distributed according to $\mathcal{T} \otimes \mathbb{P}_x$.

PROPOSITION 5. *If*

$$\limsup_{|x| \to +\infty} \frac{\langle b(x), x \rangle}{|x|} < -\frac{3}{2} \|\kappa\|_\infty^{1/2},$$

then Theorem 1 applies with $V : x \in \mathbb{R}^d \mapsto \exp(\|\kappa\|_\infty^{1/2} |x|)$. *In particular, if* $m_0 \cdot V < \infty$, *the normalized sequence of probability measures* $(\tilde{m}_n)_{n \in \mathbb{N}}$ *associated to the MVPP with random replacement kernels* $(R^{(i)})_{i \geq 1}$ *converges almost surely to the unique quasi-stationary distribution* ν_{QSD} *of* X *such that* $\nu_{\mathrm{QSD}} \cdot V < +\infty$.

REMARK 10. The fact that X admits a unique quasi-stationary distribution ν_{QSD} such that $\nu_{\text{QSD}} \cdot V < +\infty$ is proved in [16]. Proposition 5 could be generalized to diffusion processes with a nonconstant diffusion coefficient; the proof would be very similar. More generally, Condition (F) of [16] can be used to show that Theorem 1 applies to other continuous-time processes. We do not develop these generalizations further, but provide two simple examples that fall into the framework of the proof of Proposition 5:

EXAMPLE 1. If E is finite and X is regular and irreducible in E (i.e., $\mathbb{P}_x(\exists t \geq 0, \text{s.t. } X_t = y) > 0$ for all $x, y \in E$), and if $\mathbb{P}_x(\tau_\partial < +\infty) = 1$ for all $x \in E$, then Theorem 1 applies for any probability distribution $\mathcal{T}$. (One can take $V = 1$.)

EXAMPLE 2. Let X be a continuous-time multitype birth and death process, taking values in $E \cup \{\partial\} = \mathbb{N}^d$ for some $d \geq 1$, with transition rates

$$q_{x,y} = \begin{cases} b_i(x) & \text{if } y = x + e_i, \\ d_i(x) & \text{if } y = x - e_i, \\ 0 & \text{otherwise,} \end{cases}$$

where $(e_1, \dots, e_d)$ is the canonical basis of $\mathbb{N}^d$, and $\partial = (0, \dots, 0)$. We assume that $b_i(x) > 0$ and $d_i(x) > 0$ for all $1 \leq i \leq d$ and $x \in E$.

If

$$\frac{1}{|x|} \sum_{i=1}^{d} \big(d_i(x) - b_i(x)\big) \to +\infty \quad \text{when } |x| \to +\infty, \tag{11}$$

or if there exists $\delta > 1$ such that

$$\sum_{i=1}^{d} \big(d_i(x) - \delta b_i(x)\big) \to +\infty \quad \text{when } |x| \to +\infty, \tag{12}$$

then Theorem 1 applies for any probability distribution $\mathcal{T}$ admitting an exponential moment. One can choose $V(x) = |x| = x_1 + \cdots + x_d$ if (11) is satisfied, and $V(x) = \exp(\varepsilon x_1 + \cdots + \varepsilon x_d)$ with $\varepsilon > 0$ small enough if (12) is satisfied. To prove this, one would simply use the same approach as in the proof of Proposition 5 together with the results of [16], Example 7, and the fact that the killing rate is bounded by $d_1(e_1) + \cdots + d_d(e_d)$.

If moreover the birth and death process comes back from infinity (see for instance [47] for the one-dimensional case), then τ_∂ admits a uniformly bounded exponential moment and hence the conclusion of Proposition 5 applies for any probability distribution $\mathcal{T}$.

PROOF OF PROPOSITION 5. For all $n \geq 0$, we let $\hat{m}_n = m_n / \sup_{x \in E} R_x(E)$; note that $(\hat{m}_n)_{n \geq 0}$ is well defined since $\sup_{x \in E} R_x(E) \leq \sup_{x \in E} \mathbb{E}_x[T \wedge \tau_\partial] < +\infty$, by assumption on the existence of a uniform exponential moment for $T \wedge \tau_\partial$. One can check that $(\hat{m}_n)_{n \geq 0}$ is an MVPP of replacement kernel $\hat{R}^{(i)} = R^{(i)} / \sup_{x \in E} R_x(E)$ and weight kernel $\hat{P}_x = \delta_x$ (for all $x \in E$); note that we have $\hat{Q} = \hat{R}\hat{P} = \hat{R}$. Let us check that Assumption (A) is satisfied by $(\hat{m}_n)_{n \geq 0}$. Note that, for all $x \in E$ and all bounded measurable function $f : E \to \mathbb{R}$,

$$R_x \cdot f := \mathbb{E}\big[R^{(i)}\big] \cdot f = \mathbb{E}\bigg[\int_0^T G_t \cdot f(x)\, \mathrm{d}t\bigg],$$

where $G_t \cdot f(x) = \mathbb{E}_x[f(X_t)\mathbf{1}_{t < \tau_\partial}]$ is the sub-Markovian semigroup of the absorbed process X.

We have

$$\hat{R}_x(E) \geq c_1 := \frac{\mathbb{E}[\int_0^T \exp(-\|\kappa\|_\infty t)\,\mathrm{d}t]}{\sup_{y\in E} R_y(E)},$$

implying that Assumption (A1) is satisfied (take $\mu = \delta_{c_1}$).

Let us now check Assumption (A2). The function V clearly satisfies (A2-i). Moreover, one easily checks that

$$\frac{1}{2}\sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} V(x) + \sum_{i=1}^d b_i(x)\frac{\partial}{\partial x_i} V(x) \leq -\big(\|\kappa\|_\infty + \varepsilon\big)V(x) + C,$$

for some positive constants ε and C. Setting $V(\partial) = 0$, using Dynkin's formula for the killed process and a localization argument, we get that

$$\mathbb{E}_x\big[\mathrm{e}^{(\|\kappa\|_\infty+\varepsilon)t\wedge\tau_\partial} V(X_{t\wedge\tau_\partial})\big] \leq V(x) + C\mathbb{E}_x\bigg[\int_0^{t\wedge\tau_\partial} \mathrm{e}^{(\|\kappa\|_\infty+\varepsilon)s}\,\mathrm{d}s\bigg],$$

and hence that

$$\begin{aligned} G_t V(x) &= \mathbb{E}_x\big[V(X_t)\mathbf{1}_{t<\tau_\partial}\big] \\ &\leq \mathrm{e}^{-(\|\kappa\|_\infty+\varepsilon)t} V(x) + C\int_0^t \mathrm{e}^{-(\|\kappa\|_\infty+\varepsilon)(t-s)}\mathbb{P}_x(s<\tau_\partial)\,\mathrm{d}s. \end{aligned}$$

As a consequence, we have

$$\begin{aligned} R_x V &= \mathbb{E}\bigg[\int_0^T G_t V(x)\,\mathrm{d}t\bigg] \\ &\leq \mathbb{E}\bigg[\int_0^T \mathrm{e}^{-(\|\kappa\|_\infty+\varepsilon)t}\,\mathrm{d}t\bigg]V(x) + \frac{C}{\|\kappa\|_\infty+\varepsilon}\mathbb{E}_x[T\wedge\tau_\partial] \qquad (13) \\ &\leq \theta \sup_{y\in E} R_y(E)V(x) + \frac{C}{(\|\kappa\|_\infty+\varepsilon)\lambda}\sup_{y\in E}\mathbb{E}_y\big[\mathrm{e}^{\lambda(T\wedge\tau_\partial)}\big], \end{aligned}$$

where $\theta := \mathbb{E}[\int_0^T \exp(-\lambda_1 t)\,\mathrm{d}t]/\sup_{y\in E} R_y(E) < c_1$, and where $\sup_{y\in E}\mathbb{E}_y[\mathrm{e}^{\lambda(T\wedge\tau_\partial)}] < +\infty$ by assumption. Dividing the above inequality by $\sup_{x\in E} R_x(E)$ entails that Assumption (A2-ii) is satisfied. Finally, Assumption (A2-iii) is implied by the fact that $R_x^{(i)}(E)$ is stochastically dominated by $T\wedge\tau_\partial$ under $\mathbb{P}_x$, which admits a uniformly bounded exponential moment by assumption. As a consequence, we deduce that Assumption (A2) is satisfied by $\hat{R}$.

To prove that (A3) holds true, we first prove that $\hat{R}$ satisfies Assumption (E) above.

Using the same approach as in [16], Proposition 12.1, we deduce that there exist a probability measure π on K and two positive constants b and t_π such that

$$\mathbb{P}_x(X_{t_\pi}\in\cdot) \geq b\pi(\cdot) \quad \forall x\in K.$$

Since X is an elliptic diffusion process in $\mathbb{R}^d$, it satisfies, for any $t>0$, $\inf_{x\in K}\mathbb{P}_x(X_t\in K) > 0$. Using Markov's property, we deduce that, for any $t > t_\pi$, there exists a constant $b_t > 0$ such that $\mathbb{P}_x(X_t\in\cdot) \geq b_t\pi(\cdot)$, for all $x\in K$. In particular, we obtain, for any integer $n\geq 1$ and any measurable set $A\subset K$, that, for all $x\in K$,

$$\begin{aligned} R_x^n\cdot\mathbf{1}_A &= \mathbb{E}\bigg[\int_0^{T_1}\cdots\int_0^{T_n} G_{t_1+\cdots+t_n}\cdot\mathbf{1}_A(x)\,\mathrm{d}t_1\cdots\mathrm{d}t_n\bigg] \\ &\geq \mathbb{E}\bigg[\int_0^{T_1}\cdots\int_0^{T_n} b_{t_1+\cdots+t_n}\mathbf{1}_{t_1+\cdots+t_n\geq t_\pi}\,\mathrm{d}t_1\cdots\mathrm{d}t_n\bigg]\pi(A), \end{aligned}$$

where $T_1, \dots, T_n$ are i.i.d. random variables distributed with respect to $\mathcal{T}$. Since $\mathbb{P}(T_1 > 0) > 0$, we deduce that there exists n_1 large enough such that $\mathbb{P}(t_1 + \cdots + t_{n_1} \geq t_\pi) > 0$ and hence such that

$$\mathbb{E}\bigg[\int_0^{T_1} \cdots \int_0^{T_{n_1}} b_{t_1+\cdots+t_{n_1}} \mathbf{1}_{t_1+\cdots+t_{n_1} \geq t_\pi} \, dt_1 \cdots dt_{n_1}\bigg] > 0.$$

In particular, there exists a constant $\alpha_0 > 0$ such that

$$\hat{R}_x^{n_1} \cdot \mathbf{1}_A \geq \alpha_0 \pi(A \cap K). \tag{14}$$

This entails that Condition (E1) is satisfied.

We already proved that $\hat{R}_x(E) \geq c_1$ for all $x \in E$. Now, for any fixed $\alpha_1 \in (\theta^{1/q}, c_1)$ and $\rho > 0$ large enough, we deduce from (13) and as in the proof of Lemma 1 that

$$\hat{R}_x \cdot V^{1/q} \leq \alpha_1 V^{1/q}(x) + \alpha_3 \mathbf{1}_{|x| \leq \rho} \quad \forall x \in \mathbb{R}^d.$$

Setting $K = \{x \in \mathbb{R}^d, |x| \leq \rho\}$, we deduce that Condition (E2) holds true with α_1, $\alpha_2 = c_1$ and α_3 large enough, with Lyapunov function $V^{1/q}$.

We also deduce from [16], Proposition 12.1, that

$$\alpha_3 := \inf_{t \geq 0} \frac{\inf_{x \in K} G_t \cdot \mathbf{1}_E(x)}{\sup_{x \in K} G_t \cdot \mathbf{1}_E(x)} = \inf_{t \geq 0} \frac{\inf_{x \in K} \mathbb{P}_x(t < \tau_\partial)}{\sup_{x \in K} \mathbb{P}_x(t < \tau_\partial)} > 0.$$

Since $R_x(E) = \mathbb{E}[\int_0^T G_t \cdot \mathbf{1}_E(x) \, dt]$, we get that

$$\inf_{t \geq 0} \frac{\inf_{x \in K} \hat{R}_x(E)}{\sup_{x \in K} \hat{R}_x(E)}) = \inf_{t \geq 0} \frac{\inf_{x \in K} R_x(E)}{\sup_{x \in K} R_x(E)} = \alpha_3 > 0.$$

This implies that Condition (E3) holds true.

Finally, using similar calculations as in the derivation of (14), we deduce that Condition (E4) also holds true. This concludes the proof of Condition (E) with Lyapunov function $V^{1/q}$.

By Theorem 2.1 in [16], this implies that the discrete-time Markov process with transition probabilities given by $\hat{R}$ admits a unique quasi-stationary distribution ν_{QSD} such that $\nu_{\mathrm{QSD}} \cdot V^{1/q} < +\infty$. Using the same argument as in the proof of (A3) in the proof of Proposition 4, we can show that this implies that (A3) is satisfied by $\hat{R}$.

The continuity of $x \mapsto R_x$ (and thus of $x \mapsto \hat{R}_x$) is a consequence of the continuity of $x \mapsto \mathbb{E}_x[f(X_t)\mathbf{1}_{t<\tau_\partial}]$ for all continuous bounded function $f : E \to \mathbb{R}$ and all $t \geq 0$ (see, e.g., [55], Theorem 7.2.4); therefore, Assumption (A4) is also satisfied.

We have proved that Assumption (A) holds true for the MVPP of replacement kernels $(\hat{R}^{(i)})$; therefore, Theorem 1 applies. To conclude the proof, note that the continuous-time process X also admits a unique quasi-stationary distribution μ_{QSD} such that $\mu_{\mathrm{QSD}} \cdot V^{1/q} < +\infty$ (see [16], Example 2), that is, a probability measure such that $\mu_{\mathrm{QSD}} \cdot G_t = \mu_{\mathrm{QSD}} \cdot G_t(E)\mu_{\mathrm{QSD}}$ for all $t > 0$. The definition of $\hat{R}$ implies that μ_{QSD} is also a quasi-stationary distribution for $\hat{R}$; because $\mu_{\mathrm{QSD}} \cdot V^{1/q} < +\infty$ and by uniqueness, we get that $\nu_{\mathrm{QSD}} = \mu_{\mathrm{QSD}}$, which concludes the proof. □

2.3.3. *Application to stochastic-approximation algorithms for the computation of quasi-stationary distributions*. It is a difficult question to give an explicit formula for the quasi-stationary distribution of a sub-Markovian process, even when one can prove that this distribution exists and is unique. Stochastic approximation provides algorithms that allow to numerically approximate the quasi-stationary distribution of a given sub-Markovian process.

The recent papers [9, 10, 13] introduce such stochastic approximation algorithms for discrete-time sub-Markovian processes evolving in compact spaces and [61] studies these algorithms for diffusion processes in compact manifolds. Our results allow to extend these convergence results to discrete- and continuous-time processes in compact and noncompact spaces. We illustrate this approach with the case of the approximation of the quasi-stationary distribution of a diffusion process satisfying the conditions of Proposition 5 by a stochastic-approximation algorithm. This particular example was not covered by the previous literature since it is a continuous-time process and its state space is not compact.

As in the previous section, let $(X_t)_{t\in[0,+\infty)}$ be the solution in $E=\mathbb{R}^d$ to the stochastic differential equation

$$\mathrm{d}X_t = \mathrm{d}B_t + b(X_t)\,\mathrm{d}t,$$

where B is a standard d-dimensional Brownian motion and $b:\mathbb{R}^d \mapsto \mathbb{R}^d$ is locally Hölder continuous in $\mathbb{R}^d$. We assume that X is subject to an additional soft killing $\kappa : x \mapsto [0,+\infty)$, which is continuous, uniformly bounded and such that $\kappa \geq 1$. Note that the quasi-stationary distribution of X with killing rate κ is the same as the quasi-stationary distribution of X with a killing rate $\kappa - 1$.

We also assume that

$$\limsup_{|x|\to+\infty} \frac{\langle b(x), x\rangle}{|x|} < -\frac{3}{2}\|\kappa\|_\infty^{1/2},$$

so that the process X admits a unique quasi-stationary distribution ν_{QSD} such that $\nu_{\mathrm{QSD}} \cdot V < +\infty$, where $V : x \in \mathbb{R}^d \mapsto \exp(\|\kappa\|_\infty^{1/2}|x|)$ (see the previous subsection for details).

We consider the self-interacting process $(Y_t)_{t\geq 0}$ evolving with the same dynamic of X but, at rate κ, instead of being killed, it jumps to a new position chosen accordingly to its empirical occupation measure $\frac{1}{t}\int_0^t \delta_{Y_s}\,\mathrm{d}s$. More formally, it evolves following the dynamic

$$\mathrm{d}Y_t = \mathrm{d}B_t + b(Y_t)\,\mathrm{d}t + \mathrm{d}N_t, \qquad Y_0 = y \in \mathbb{R}^d,$$

where $(N_t)_{t\geq 0}$ is a time inhomogeneous pure jump process with jump measure given by

$$\frac{\kappa(Y_{t-})}{t}\int_0^t \delta_{Y_s - Y_{t-}}\,\mathrm{d}s.$$

PROPOSITION 6. *The empirical occupation measure* $\frac{1}{t}\int_0^t \delta_{Y_s}\,\mathrm{d}s$ *converges almost-surely when* $t\to+\infty$, *with respect to the topology of weak convergence*, *to the quasi-stationary distribution* ν_{QSD} *of* X.

PROOF. Denote by $0<\tau_1<\tau_2<\cdots$ the jump times of Y and set $\tau_0 = 0$. Then, for all $n\geq 0$ and conditionally on Y_{τ_n},

$$\int_{\tau_n}^{\tau_{n+1}} \delta_{Y_s}\,\mathrm{d}s = R^{(n+1)}_{Y_{\tau_n}},$$

where $R^{(n+1)}$ is defined as in the proof of Proposition 5. Moreover, Y_{τ_n} is distributed according to the probability measure $\frac{1}{\tau_n}\int_0^{\tau_n}\delta_{Y_s}\,\mathrm{d}s$. As a consequence, setting $m_0 = \int_0^{\tau_1}\delta_{Y_s}\,\mathrm{d}s$ (which satisfies $m_0\cdot V<+\infty$ almost surely) and $m_n := \frac{1}{\tau_{n+1}}\int_0^{\tau_{n+1}}\delta_{Y_s}\,\mathrm{d}s$, the sequence $(m_n)_{n\in\mathbb{N}}$ has the law of the MVPP of Proposition 5. Applying this proposition with $T=+\infty$ almost surely (note that $\kappa\geq 1$ implies that $\tau_\partial \wedge \infty = \tau_\partial$ admits a uniformly bounded exponential moment), we obtain that

$$\frac{1}{\tau_n}\int_0^{\tau_n}\delta_{Y_s}\,\mathrm{d}s \xrightarrow[n\to+\infty]{\text{a.s.}} \nu_{\mathrm{QSD}} \tag{15}$$

with respect to the topology of weak convergence.

Since $\kappa \geq 1$, one can couple the sequence $(\tau_{n+1} - \tau_n)_{n\geq 0}$ with a sequence of i.i.d. random variables $(D_n)_{n\geq 0}$ with exponential law of parameter 1 such that $0 \leq \tau_{n+1} - \tau_n \leq D_n$ almost surely for all $n \geq 0$. Moreover $\tau_n \to +\infty$ almost surely when $n \to +\infty$ (this is due to the fact that κ is uniformly bounded). Hence, using (15), we get

$$\frac{1}{\tau_n}\int_0^{\tau_{n+1}} \delta_{Y_s}\,\mathrm{d}s \xrightarrow[n\to+\infty]{\text{a.s.}} \nu_{\text{QSD}}$$

and

$$\frac{1}{\tau_{n+1}}\int_0^{\tau_n} \delta_{Y_s}\,\mathrm{d}s \xrightarrow[n\to+\infty]{\text{a.s.}} \nu_{\text{QSD}}.$$

For all $t \geq 0$, we define $\alpha(t) := \sup\{n \geq 0, \tau_n \leq t\}$. In particular, for all $t \geq 0$, $\alpha(t) < +\infty$, $\tau_{\alpha(t)} \leq t < \tau_{\alpha(t)+1}$ and $\alpha(t) \to +\infty$ almost surely when $t \to +\infty$. As a consequence, for all bounded continuous function $f : \mathbb{R}^d \to [0, +\infty)$,

$$\frac{1}{\tau_{\alpha(t)+1}}\int_0^{\tau_{\alpha(t)}} f(Y_s)\,\mathrm{d}s \leq \frac{1}{t}\int_0^t f(Y_s)\,\mathrm{d}s \leq \frac{1}{\tau_{\alpha(t)}}\int_0^{\tau_{\alpha(t)+1}} f(Y_s)\,\mathrm{d}s.$$

This and the above convergence results allow us to conclude the proof. □

REMARK 11. Since the submission of this paper, Benaïm, Champagnat and Villemonais [8] proved almost sure convergence of a similar stochastic approximation algorithm, where the diffusion process is resampled according to its empirical occupation measure when it hits the boundary of a bounded domain. On the one hand, their result do not apply to the model studied in this section where the state space is not bounded; on the other hand, our result do not apply to their situation, since Assumption (A1) would fail in that case.

3. Proof of Theorem 1. Let us define an auxiliary sequence of random distributions: let $\eta_0 = 0$, and, for all $n \geq 1$,

$$\eta_n = \eta_{n-1} + \delta_{Y_n} = \sum_{i=1}^n \delta_{Y_i}.$$

Recall that, by definition,

$$m_n = m_0 + \sum_{i=1}^n R^{(i)}_{Y_i} = m_0 + \sum_{i=1}^n \delta_{Y_i} R^{(i)}$$

and that, conditionally on the sigma-algebra $\mathcal{F}_n$ generated by $\{m_i\}_{0\leq i\leq n} \cup \{Y_i\}_{1\leq i\leq n}$, the random variable Y_{n+1} is distributed according to $m_n P/m_n P(E)$ and $R^{(n+1)}$ is chosen independently of $\mathcal{F}_n$ and Y_{n+1}.

We set $\tilde{\eta}_0 = 0$, and, for all $n \geq 1$,

$$\tilde{\eta}_n = \frac{\eta_n}{\eta_n(E)} = \frac{\eta_n}{n}.$$

We first prove that $\tilde{\eta}_n$ converges almost surely weakly to ν when n goes to infinity and then deduce almost-sure convergence of $\tilde{m}_n$ to $\nu R/\nu R(E)$:

PROPOSITION 7. *Under the Assumptions* (T), (A1), (A′2), (A3), (A4), *the sequence* $(\tilde{\eta}_n)_{n\geq 0}$ *converges weakly almost surely to* ν *when* n *goes to infinity. Said differently,*

$$\frac{1}{n}\sum_{i=1}^n \delta_{Y_i} \to \nu \quad \textit{almost surely when } n \to \infty.$$

3.1. *Proof of Proposition* 7. We consider the dynamical system defined by

$$\frac{d\mu_t \cdot f}{dt} = \mu_t Q \cdot f - \mu_t Q(E)\mu_t \cdot f, \tag{16}$$

for all bounded continuous functions $f : E \to \mathbb{R}$, where $(\mu_t)_{t\geq 0}$ shall not depend on f. Existence, uniqueness and continuity properties of the flow induced by this dynamical system are stated and proved in Lemma 7.

To prove almost-sure convergence of $\tilde{\eta}_n$ to ν (i.e., Proposition 7), we prove that a linearization of it is a *pseudo-asymptotic trajectory* (see Section 3 of [7]) of the semiflow induced by the dynamical system (16). To do so, we need to prove several intermediate results: In Lemma 2, we write down the studied stochastic algorithm. In Lemma 4, we prove that the expectation of V with respect to the measure-valued process remains bounded. In Lemma 5, we prove almost-sure convergence of the quantity introduced in Proposition 4.1 of [7] to control the error term between the dynamical system (16) and its linearized counterpart (the almost-sure convergence of this error to zero is sometimes called the Kushner and Clark's condition). In Lemma 6, we prove that the sequence $(\tilde{\eta}_n)_n$ is relatively compact for the topology of weak convergence on $\mathcal{P}(E)$. All these elements allow us to conclude the proof of Proposition 7 using standard stochastic-approximation methods, as developed in [11].

From now on, we assume that all the hypotheses of Proposition 7 hold.

LEMMA 2. *For all* $n \geq 1$, *we have*

$$\tilde{\eta}_{n+1} - \tilde{\eta}_n = \gamma_{n+1}\big(F(\tilde{\eta}_n) + U_{n+1}\big),$$

where

$$\gamma_{n+1} = \frac{1}{\eta_{n+1}(E)\tilde{\eta}_n Q(E)},$$

and

$$F(\tilde{\eta}_n) = \tilde{\eta}_n Q - \tilde{\eta}_n Q(E)\tilde{\eta}_n,$$
$$U_{n+1} = \tilde{\eta}_n Q(E)\delta_{Y_{n+1}} - \tilde{\eta}_n Q.$$

The term γ_{n+1} may be interpreted as the step size of a stochastic Euler scheme approximation of equation (16) and it decreases to 0 when $n \to +\infty$. For instance, in the simple case where $Q(E) = 1$, γ_{n+1} equals $1/(n+1)$.

PROOF. The result directly follows from

$$\tilde{\eta}_{n+1} = \left(1 - \frac{1}{n+1}\right)\tilde{\eta}_n + \frac{1}{n+1}\delta_{Y_{n+1}} = \tilde{\eta}_n + \frac{1}{n+1}(\delta_{Y_{n+1}} - \tilde{\eta}_n). \qquad \square$$

LEMMA 3. *Fix* $c' \in (\theta, c_1)$, *for all* $k \geq 1$, *we let*

$$\sigma_k = \inf\{n \geq k : m_n P(E) < c'n\}. \tag{17}$$

We have $\mathbb{P}(\bigcup_{k\geq 1}\{\sigma_k = \infty\}) = 1$.

PROOF. Recall that $m_n P(E) = m_0 P(E) + \sum_{i=1}^n R_{Y_i}^{(i)} P(E)$, and, therefore,

$$m_n P(E) = m_0 P(E) + \sum_{i=1}^n Q_{Y_i}^{(i)}(E).$$

Assumption (A1) and, conditionally on $Y_1, \dots, Y_n, \dots$, the independence of the random variables $Q^{(i)}_{Y_i}(E)$ entails (by coupling) that there exists a sequence of independent random variables $Z_1, \dots, Z_n, \dots$ with law μ such that, conditionally on $Y_1, \dots, Y_n, \dots$, we have $Q^{(i)}_{Y_i}(E) \geq Z_i$ for all $i \geq 1$. The law of large numbers hence implies that

$$\liminf_{n\to+\infty} \frac{m_n P(E)}{n} \geq c_1 \quad \text{almost surely,}$$

which concludes the proof. □

We claimed that Assumption (A1) can be replaced by equation (3) in Theorem 1, to prove this claim, we need to prove Lemma 3 in this alternative setting:

PROOF OF LEMMA 3 WITH ASSUMPTION (A1) REPLACED BY (3). Recall that

$$m_n P(E) = m_0 P(E) + \sum_{i=1}^{n} R^{(i)}_{Y_i} P(E),$$

and, therefore,

$$m_n P(E) = m_0 P(E) + \sum_{i=1}^{n} \mathbb{E}_{i-1} Q^{(i)}_{Y_i}(E) + \sum_{i=1}^{n} \bigl(Q^{(i)}_{Y_i}(E) - \mathbb{E}_{i-1} Q^{(i)}_{Y_i}(E)\bigr),$$

where $\mathbb{E}_{i-1}$ denotes the expectation conditionally on $(m_1, \dots, m_{i-1})$. Note that, since $Q^{(i)}$ is independent from $\mathcal{F}_{i-1}$ and Y_i, we have

$$\sum_{i=1}^{n} \mathbb{E}_{i-1} Q^{(i)}_{Y_i}(E) = \sum_{i=1}^{n} \mathbb{E}_{i-1} Q_{Y_i}(E) \geq c_1 n, \tag{18}$$

by Assumption (A1). Also note that

$$M_n := \sum_{i=1}^{n} \bigl(Q^{(i)}_{Y_i}(E) - \mathbb{E}_{i-1} Q^{(i)}_{Y_i}(E)\bigr)$$

is a martingale. Using Lemma 1 in [18] (without loss of generality, we assume that $\beta \in (1, 2]$), one deduces from Assumption (3) that

$$\begin{aligned} \mathbb{E}|M_n|^\beta &\leq 2 \sum_{i=1}^{n} \mathbb{E}_{i-1} \bigl| Q^{(i)}_{Y_i}(E) - \mathbb{E}_{i-1} Q^{(i)}_{Y_i}(E)\bigr|^\beta \\ &\leq 2n \sup_{x\in E} \mathbb{E}\bigl|Q^{(i)}_x(E) - Q_x(E)\bigr|^\beta. \end{aligned}$$

Hence, using (3), we get that the sequence $(n^{-1}\mathbb{E}|M_n|^\beta)_{n\geq 1}$ is bounded. This implies, by an immediate adaptation of Theorem 1.3.17 in [28] (the main point is to use Doob's inequality instead of Kolmogorov's inequality), that $n^{-1} M_n$ goes almost surely to zero when n goes to infinity.

Therefore, we have that, almost surely when $n \to +\infty$,

$$m_n P(E) = \sum_{i=1}^{n} \mathbb{E}_{i-1} Q^{(i)}_{Y_i}(E) + o(n),$$

and, using equation (18), we get

$$m_n P(E) \geq c_1 n + o(n) \quad \text{almost surely,}$$

which concludes the proof because $c' < c_1$. □

LEMMA 4. *For all $k \geq 1$, there exists a constant $C_k > 0$ such that*

$$\sup_{n\geq 1}\mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right] \vee \sup_{n\geq 1}\mathbb{E}\left[\frac{m_{n\wedge\sigma_k}P\cdot V}{n}\right] \vee \sup_{n\geq 1}\mathbb{E}\left[V(Y_{n+1})\mathbf{1}_{n<\sigma_k}\right] \leq C_k.$$

PROOF. Fix $n \geq k+1$, we have

$$\mathbb{E}\left[\frac{\eta_{(n+1)\wedge\sigma_k}\cdot V}{n+1}\right] = \left(1-\frac{1}{n+1}\right)\mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right] + \frac{\mathbb{E}[V(Y_{n+1})\mathbf{1}_{n<\sigma_k}]}{n+1}. \tag{19}$$

Note that, by definition of σ_k (see equation (17)), we have, almost surely and for all $n \in \{k+1,\dots,\sigma_k - 1\}$,

$$m_n P(E) \geq c' n.$$

Hence, by definition of Y_{n+1}, we have (recall that m_n, and thus $m_n P$, is assumed to be a positive measure almost surely), for all $n \geq k+1$,

$$\begin{aligned}\mathbb{E}\left[V(Y_{n+1})\mathbf{1}_{n<\sigma_k}\right] &= \mathbb{E}\left[\frac{m_n P\cdot V}{m_n P(E)}\mathbf{1}_{n<\sigma_k}\right] \leq \frac{1}{c'n}\mathbb{E}[m_{n\wedge\sigma_k}P\cdot V]\\ &= \frac{1}{c'n}\mathbb{E}\left[m_0 P\cdot V + \sum_{i=1}^{n\wedge\sigma_k} Q^{(i)}_{Y_i}\cdot V\right]\\ &\leq \frac{1}{c'n}\mathbb{E}\left[m_0 P\cdot V + \sum_{i=1}^{n} Q^{(i)}_{Y_i}\cdot V\mathbf{1}_{i\leq\sigma_k}\right]\\ &= \frac{1}{c'n}\mathbb{E}\left[m_0 P\cdot V + \sum_{i=1}^{n} Q_{Y_i}\cdot V\mathbf{1}_{i\leq\sigma_k}\right],\end{aligned}$$

where the last equality is obtained by conditioning on $\mathcal{F}_{i-1}$ and Y_i, and using the fact that $\mathbf{1}_{i\leq\sigma_k}$ is measurable with respect to $\mathcal{F}_i \cup \sigma(Y_i)$ and that $Q^{(i)}$ is independent of $\mathcal{F}_i \cup \sigma(Y_i)$. We thus get, using the Lyapunov assumption (A′2-i) in the second inequality,

$$\begin{aligned}\mathbb{E}\left[V(Y_{n+1})\mathbf{1}_{n<\sigma_k}\right] &\leq \frac{1}{c'n}\mathbb{E}[m_{n\wedge\sigma_k}P\cdot V] &(20)\\ &\leq \frac{m_0 P\cdot V + \mathbb{E}[\eta_{n\wedge\sigma_k}Q\cdot V]}{c'n}\\ &\leq \frac{m_0 P\cdot V + nK + \theta\mathbb{E}[\eta_{n\wedge\sigma_k}\cdot V]}{c'n}\\ &\leq \frac{m_0 P\cdot V + nK}{c'n} + \frac{\theta}{c'}\mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right]. &(21)\end{aligned}$$

Thus, using equation (19), we get, for all $n \geq k+1$,

$$\mathbb{E}\left[\frac{\eta_{(n+1)\wedge\sigma_k}\cdot V}{n+1}\right] \leq \left(1-\frac{1-\theta/c'}{n+1}\right)\mathbb{E}\left[\frac{\eta_{n\wedge\sigma_k}\cdot V}{n}\right] + \frac{m_0 P\cdot V + nK}{c'n(n+1)}.$$

One easily checks that $\mathbb{E}[\tilde{\eta}_{n\wedge\sigma_k}\cdot V] < +\infty$ for all $n \leq k$ and, since we assumed that $m_0 P\cdot V < +\infty$ and since $\theta < c' < 1$, we can infer that $\mathbb{E}[\eta_{n\wedge\sigma_k}\cdot V/n]$ is uniformly bounded in n. Finally, the inequality between (20) and (21) implies that both $\mathbb{E}[m_{n\wedge\sigma_k}P\cdot V/n]$ and $\mathbb{E}[V(Y_{n+1})\mathbf{1}_{n<\sigma_k}]$ are also uniformly bounded in n. □

LEMMA 5 (Kushner and Clark's condition). *Set $W = V^{1/q}$. Almost surely*

$$\lim_{n\to+\infty}\sum_{\ell=1}^{n}\gamma_\ell U_\ell\cdot W$$

exists and is finite.

PROOF. Fix $k \geq 1$. Following [54], Lemma 1, we let $Z_\ell = \gamma_\ell U_\ell \cdot W$ and $M_n = \sum_{\ell=1}^{n\wedge\sigma_k}(Z_\ell - \mathbb{E}_{\ell-1}Z_\ell)$, where $\mathbb{E}_{\ell-1}$ denotes the conditional expectation conditionally on $\mathcal{F}_{\ell-1}$. The rest of the proof is done into two steps: first, we prove that the martingale $(M_n)_{n\geq 0}$ is uniformly bounded in L^r, implying that it converges almost surely, second, we prove that $\sum_{\ell=1}^{n\wedge\sigma_k}\mathbb{E}_{\ell-1}Z_\ell$ converges almost surely when n tends to infinity.

Step 1: Using Jensen's inequality, we get that the constant r can be assumed to be arbitrarily small as long as it is larger than 1; in particular, we can assume that $r < 2$. Using this together with Lemma 1 in [18], we get

$$\mathbb{E}|M_n|^r \leq 2\sum_{\ell=1}^n \mathbb{E}\big[|Z_\ell - \mathbb{E}_{\ell-1}Z_\ell|^r \mathbf{1}_{\ell\leq\sigma_k}\big] \leq 8\sum_{\ell=1}^n \mathbb{E}\big[|Z_\ell|^r \mathbf{1}_{\ell\leq\sigma_k}\big]. \tag{22}$$

Recall that, by definition, $U_\ell = \tilde{\eta}_{\ell-1}Q(E)\delta_{Y_\ell} - \tilde{\eta}_{\ell-1}Q$ and $\gamma_\ell = (\eta_\ell(E)\tilde{\eta}_{\ell-1}Q(E))^{-1}$ (see Lemma 2); therefore, we have

$$\begin{aligned}\mathbb{E}\big[|Z_\ell|^r\mathbf{1}_{\ell\leq\sigma_k}\big] &= \mathbb{E}\bigg[\frac{|\tilde{\eta}_{\ell-1}Q(E)W(Y_\ell) - \tilde{\eta}_{\ell-1}Q\cdot W|^r}{|\eta_\ell(E)\tilde{\eta}_{\ell-1}Q(E)|^r}\mathbf{1}_{\ell\leq\sigma_k}\bigg] \\ &\leq 2\mathbb{E}\bigg[\frac{V(Y_\ell)}{\eta_\ell(E)^r}\mathbf{1}_{\ell\leq\sigma_k} + \frac{1}{\eta_\ell(E)^r}\bigg|\frac{\tilde{\eta}_{\ell-1}Q}{\tilde{\eta}_{\ell-1}Q(E)}\cdot W\bigg|^r\mathbf{1}_{\ell\leq\sigma_k}\bigg],\end{aligned}$$

where we recall that $W = V^{1/q}$. Using Assumption (A′2-iv) and the fact that $\eta_\ell(E) = \ell$, $\tilde{\eta}_{\ell-1}Q(E) \geq c_1$ (see Assumption (A1)) and $\mathbb{E}[V(Y_\ell)\mathbf{1}_{\ell\leq\sigma_k}] \leq C_k$ (see Lemma 4), we get

$$\mathbb{E}\big[|Z_\ell|^r\mathbf{1}_{\ell\leq\sigma_k}\big] \leq \frac{2C_k}{\ell^r} + \frac{2\mathbb{E}[\tilde{\eta}_{\ell-1}|Q\cdot W|^r\mathbf{1}_{\ell\leq\sigma_k}]}{c_1^r\ell^r} \leq \frac{2}{\ell^r}\bigg(C_k + \frac{BC_k}{c_1^r}\bigg),$$

where we used Lemma 4 and Assumption (A′2-iv) for the last inequality (recall that, by Jensen's inequality, r can be assumed to be arbitrarily close to one, and thus smaller than q, in particular). Using equation (22), this implies that the martingale $(M_n)_{n\geq 0}$ is uniformly bounded in L^r and hence that it converges almost surely.

Step 2: Using the fact that $\eta_\ell(E) = \ell$, we also have

$$\begin{aligned}\mathbb{E}\big|\mathbb{E}_{\ell-1}[Z_\ell]\mathbf{1}_{\ell\leq\sigma_k}\big| &= \mathbb{E}\bigg|\mathbb{E}_{\ell-1}\bigg[\frac{\tilde{\eta}_{\ell-1}Q(E)W(Y_\ell) - \tilde{\eta}_{\ell-1}Q\cdot W}{\eta_\ell(E)\tilde{\eta}_{\ell-1}Q(E)}\bigg]\mathbf{1}_{\ell\leq\sigma_k}\bigg| \\ &= \frac{1}{\ell}\mathbb{E}\bigg|\mathbb{E}_{\ell-1}\bigg[W(Y_\ell) - \frac{\tilde{\eta}_{\ell-1}Q\cdot W}{\tilde{\eta}_{\ell-1}Q(E)}\bigg]\mathbf{1}_{\ell\leq\sigma_k}\bigg| \\ &= \frac{1}{\ell}\mathbb{E}\bigg|\frac{m_{\ell-1}P\cdot W}{m_{\ell-1}P(E)}\mathbf{1}_{\ell\leq\sigma_k} - \frac{\eta_{\ell-1}Q\cdot W}{\eta_{\ell-1}Q(E)}\mathbf{1}_{\ell\leq\sigma_k}\bigg|,\end{aligned}$$

where we used for the last equality that the conditional distribution of Y_ℓ given $\mathcal{F}_{\ell-1}$ is $m_{\ell-1}P/m_{\ell-1}P(E)$. By the triangular inequality, and using the fact that $\eta_{\ell-1}Q(E) \geq c_1(\ell-1)$ almost surely (see Assumption (A1)), we get

$$\begin{aligned}\mathbb{E}\big|\mathbb{E}_{\ell-1}[Z_\ell]\mathbf{1}_{\ell\leq\sigma_k}\big| &\leq \frac{1}{c_1\ell(\ell-1)}\mathbb{E}\big[|m_{\ell-1}P\cdot W - \eta_{\ell-1}Q\cdot W|\mathbf{1}_{\ell\leq\sigma_k}\big] \\ &\quad + \frac{1}{\ell}\mathbb{E}\bigg[\bigg|\frac{1}{m_{\ell-1}P(E)} - \frac{1}{\eta_{\ell-1}Q(E)}\bigg|m_{\ell-1}P\cdot W\mathbf{1}_{\ell\leq\sigma_k}\bigg].\end{aligned} \tag{23}$$

Let us first bound the first term of the above sum. Using Jensen's inequality and Lemma 1 in [18] (note that $(m_{\ell\wedge\sigma_k}P - \eta_{\ell\wedge\sigma_k}Q)_{\ell\geq 0}$ is a martingale), we get

$$\begin{aligned}&\mathbb{E}\big[|m_{\ell-1}P\cdot W - \eta_{\ell-1}Q\cdot W|\mathbf{1}_{\ell\leq\sigma_k}\big]^r \\ &\quad\leq \mathbb{E}\big[|m_{\ell-1}P\cdot W - \eta_{\ell-1}Q\cdot W|\mathbf{1}_{\ell-1\leq\sigma_k}\big]^r\end{aligned}$$

$$\le \mathbb{E}|m_{(\ell-1)\wedge\sigma_k} P \cdot W - \eta_{(\ell-1)\wedge\sigma_k} Q \cdot W|^r$$

$$\le 2(m_0 P \cdot W)^r + 2\sum_{i=1}^{\ell-1} \mathbb{E}\big[\big|Q^{(i)} \cdot W(Y_i) - Q \cdot W(Y_i)\big|^r \mathbf{1}_{i\le\sigma_k}\big]$$

$$\le 2(m_0 P \cdot W)^r + 2\sum_{i=1}^{\ell-1} B\mathbb{E}\big[V(Y_i)\mathbf{1}_{i\le\sigma_k}\big],$$

where we used the fact that $\mathbf{1}_{\ell\le\sigma_k} \le \mathbf{1}_{\ell-1\le\sigma_k}$ almost surely, that $\mathbf{1}_{i\le\sigma_k}$ is measurable with respect to $\mathcal{F}_{i-1} \cup \sigma(Y_i)$, and Assumption (A′2-iv). Finally, Lemma 4 implies that there exists a constant $C'_k > 0$ such that

$$\mathbb{E}\big[|m_{\ell-1} P \cdot W - \eta_{\ell-1} Q \cdot W|\mathbf{1}_{\ell\le\sigma_k}\big] \le C'_k\big((m_0 P \cdot W)^r + \ell - 1\big)^{1/r}. \tag{24}$$

Let us now look at the second term in the right-hand side of equation (23); using Assumption (A1), we have that

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg|\frac{1}{m_{\ell-1}P(E)} - \frac{1}{\eta_{\ell-1}Q(E)}\bigg| m_{\ell-1}P \cdot W\mathbf{1}_{\ell\le\sigma_k}\bigg] \\
&\quad = \mathbb{E}\bigg[\frac{|\eta_{\ell-1}Q(E) - m_{\ell-1}P(E)|}{\eta_{\ell-1}Q(E)} \frac{m_{\ell-1}P \cdot W}{m_{\ell-1}P(E)}\mathbf{1}_{\ell\le\sigma_k}\bigg] \\
&\quad \le \frac{1}{c_1(\ell-1)}\mathbb{E}\bigg[\big|\eta_{\ell-1}Q(E) - m_{\ell-1}P(E)\big|\frac{m_{\ell-1}P \cdot W}{m_{\ell-1}P(E)}\mathbf{1}_{\ell\le\sigma_k}\bigg] \\
&\quad \le \frac{1}{c_1(\ell-1)}\mathbb{E}\big[\big|\eta_{\ell-1}Q(E) - m_{\ell-1}P(E)\big|^p\mathbf{1}_{\ell\le\sigma_k}\big]^{1/p}\mathbb{E}\bigg[\bigg(\frac{m_{\ell-1}P \cdot W}{m_{\ell-1}P(E)}\bigg)^q\mathbf{1}_{\ell\le\sigma_k}\bigg]^{1/q} \\
&\quad \le \frac{C_k^{1/q}}{c_1(\ell-1)}\mathbb{E}\big[\big|\eta_{\ell-1}Q(E) - m_{\ell-1}P(E)\big|^p\mathbf{1}_{\ell\le\sigma_k}\big]^{1/p},
\end{aligned}$$

where we used Hölder's inequality (in the second inequality), Jensen's inequality and Lemma 4 (in the last inequality). Now, using the main result of [26], we obtain that, for some constant $d_p > 0$,

$$\begin{aligned}
&\mathbb{E}\big[\big|\eta_{\ell-1}Q(E) - m_{\ell-1}P(E)\big|^p\mathbf{1}_{\ell\le\sigma_k}\big] \\
&\quad \le \mathbb{E}\big[\big|\eta_{\ell\wedge\sigma_k-1}Q(E) - m_{\ell\wedge\sigma_k-1}P(E)\big|^p\big] \\
&\quad \le 2^{p-1}\bigg[m_0P(E)^p + d_p(\ell-1)^{p/2-1}\sum_{i=1}^{\ell-1}\mathbb{E}\big[\big|Q_{Y_i}(E) - Q_{Y_i}^{(i)}(E)\big|^p\mathbf{1}_{i<\sigma_k}\big]\bigg] \\
&\quad \le 2^{p-1}\bigg[m_0P(E)^p + d_p(\ell-1)^{p/2-1}\sum_{i=1}^{\ell-1}A\mathbb{E}\big[V(Y_i)\mathbf{1}_{i\le\sigma_k}\big]\bigg],
\end{aligned}$$

where we used Assumption (A′2-iii). Hence, using Lemma 4, we deduce that

$$\begin{aligned}
&\mathbb{E}\bigg[\bigg|\frac{1}{m_{\ell-1}P(E)} - \frac{1}{\eta_{\ell-1}Q(E)}\bigg| m_{\ell-1}P \cdot W\bigg]^p \\
&\quad \le \frac{C_k^{p/q}2^{p-1}}{c_1^p(\ell-1)^p}\big(m_0P(E)^p + d_p(\ell-1)^{p/2}AC_k\big).
\end{aligned} \tag{25}$$

Finally, from inequalities (23), (24) and (25), we deduce that $\sum_{\ell=1}^{\infty}\mathbb{E}|\mathbb{E}_{\ell-1}Z_\ell\mathbf{1}_{\ell\le\sigma_k}| < \infty$. As a consequence, $\sum_{\ell=1}^{\sigma_k}|\mathbb{E}_{\ell-1}Z_\ell| < \infty$ almost surely, implying that $\sum_{\ell=1}^{n\wedge\sigma_k}\mathbb{E}_{\ell-1}Z_\ell$ converges almost surely when $n \to \infty$. Recall that we have proved that $M_n = \sum_{\ell=1}^{n\wedge\sigma_k}(Z_\ell - \mathbb{E}_{\ell-1}Z_\ell)$ converges almost surely when n goes to infinity (we showed earlier that it was uniformly bounded in L^r). Therefore, we can imply that $\sum_{\ell=1}^{n\wedge\sigma_k} Z_k$ converges almost surely. Since $\mathbb{P}(\bigcup_{k\geq 1}\{\sigma_k = +\infty\}) = 1$ (see Lemma 3), we get that $\sum_{\ell=1}^{n} Z_k$ converges almost surely, which concludes the proof. □

From now on, for all $C > 0$, we set

$$\mathcal{P}_C(E) := \{\mu : \mu \text{ is a probability on } E \text{ such that } \mu \cdot W \leq C\},$$

where we recall that $W = V^{1/q}$. Note that $\mathcal{P}_C(E)$ is a compact subset of $\mathcal{P}(E)$ (the set of Borel probability measures on E) with respect to the topology of weak convergence.

LEMMA 6. *The sequence $(\tilde{\eta}_n)_{n\geq 0}$ is almost surely relatively compact in $\mathcal{P}(E)$ with respect to the topology of weak convergence. More precisely, there exists a random value $C > 0$ such that, almost surely, $\tilde{\eta}_n \in \mathcal{P}_C(E)$ for all $n \in \mathbb{N}$.*

PROOF. Using Lemma 2, we have that, for all $n \geq 0$ (recall that $W = V^{1/q}$),

$$\tilde{\eta}_{n+1} \cdot W = \tilde{\eta}_n \cdot W + \gamma_{n+1}\bigl(U_{n+1} \cdot W + F(\tilde{\eta}_n) \cdot W\bigr),$$

where

$$F(\tilde{\eta}_n) \cdot W = \tilde{\eta}_n Q \cdot W - \tilde{\eta}_n Q(E)\tilde{\eta}_n \cdot W \leq \theta\tilde{\eta}_n \cdot W + K - c_1\tilde{\eta}_n \cdot W,$$

where we have used Assumptions (A1) and (A′2-ii). Therefore, we get

$$\tilde{\eta}_{n+1} \cdot W \leq \tilde{\eta}_n \cdot W + \gamma_{n+1}\bigl(U_{n+1} \cdot W + K + (\theta - c_1)\tilde{\eta}_n \cdot W\bigr). \tag{26}$$

We define the random variable

$$M = \sup_{m\geq n\geq 1}\left|\sum_{k=n}^{m} \gamma_{n+1}U_{k+1} \cdot W\right|$$

which is finite almost surely (by Lemma 5). Let us prove by induction that

$$\tilde{\eta}_n \cdot W \leq 2M + \frac{1 + c_1 - \theta}{c_1 - \theta}\hat{K}, \tag{27}$$

where $\hat{K} = K/c_1 \vee (\tilde{\eta}_1 \cdot W)$ (note that $\hat{K}$ is random and that $\hat{K} \geq K/c_1 \geq K$). The result is immediate for $n = 1$. Assume now that the result holds true for $n \geq 1$. If $\tilde{\eta}_n \cdot W \leq \frac{\hat{K}}{c_1-\theta}$, then (26) entails that

$$\tilde{\eta}_{n+1} \cdot W \leq \frac{\hat{K}}{c_1 - \theta} + M + \hat{K} \leq M + \frac{1 + c_1 - \theta}{c_1 - \theta}\hat{K},$$

because $\gamma_{n+1} \leq 1/c_1$ almost surely by Assumption (A1). If $\tilde{\eta}_n \cdot W > \frac{\hat{K}}{c_1-\theta}$, then we define the (random) integer n_0 by

$$n_0 = \sup\left\{k \in \{1, \dots, n\} \text{ such that } \tilde{\eta}_k \cdot W > \frac{\hat{K}}{c_1 - \theta} \text{ and } \tilde{\eta}_{k-1} \cdot W \leq \frac{\hat{K}}{c_1 - \theta}\right\},$$

which is well defined since $\tilde{\eta}_1 \cdot W \leq \hat{K}$ by definition of $\hat{K}$. We can thus deduce as above that $\tilde{\eta}_{n_0} \cdot W \leq M + \frac{1+c_1-\theta}{c_1-\theta}\hat{K}$ and hence

$$\tilde{\eta}_{n+1} \cdot W \leq \tilde{\eta}_{n_0} \cdot W + \sum_{k=n_0}^{n} \gamma_{k+1}U_{k+1} \cdot W \leq M + \frac{1 + c_1 - \theta}{c_1 - \theta}\hat{K} + M.$$

Finally, we deduce by induction that (27) holds true for all $n \geq 1$.

Since the right-hand side of (27) does not depend on n and since $W = V^{1/q}$ has relatively compact level sets by Assumption (A′2), we deduce that $(\tilde{\eta}_n)_{n\in\mathbb{N}}$ is almost surely relatively compact for the topology of weak convergence on $\mathcal{P}(E)$ (see for instance [51], Theorem 6.7, Chapter II). □

LEMMA 7. *For any $C \geq \frac{K}{c_1-\theta}$ and any $\mu_0 \in \mathcal{P}_C(E)$, $t \mapsto \nu_t := \mathbb{P}_{\mu_0}(X_t \in \cdot \mid X_t \neq \partial)$ is the unique solution to the dynamical system* (16) *with values in $\mathcal{P}_C(E)$ and it is continuous with respect to $(\mu_0, t) \in \mathcal{P}_C(E) \times [0, +\infty)$.*

PROOF. *Step 1. Existence*. Fix $C > 0$ and $\mu_0 \in \mathcal{P}_C(E)$. We consider the weak forward-Kolmogorov equation defined as

$$\frac{\mathrm{d}\mu_t \cdot f}{\mathrm{d}t} = \mu_t(Q - I) \cdot f, \tag{28}$$

for all bounded continuous functions $f : E \to \mathbb{R}$. If μ_0 is a Dirac measure δ_x, then, by [19], Theorem 2.21, $t \mapsto \mathbb{P}_x(X_t \in \cdot)$ is a solution of this equation. Recall that $W = V^{1/q}$; equation (2.29) in [19] states that if there exists a constant $c > 0$ such that $(Q - I)W \leq cW$, then, for all $x \in E$, for all $s \geq 0$, $\mathbb{E}_x[W(X_s)] \leq W(x)\mathrm{e}^{cs}$ (here and below, we always assume that the considered functions vanish on ∂, so that $\mathbb{E}_x[W(X_s)] = \mathbb{E}_x[W(X_s)\mathbf{1}_{X_s\neq\partial}]$). Using Assumption (A′2-iv), we get that $|Q_x W| \leq B^{1/q} W$, which thus implies that

$$\mathbb{E}_x W(X_s) \leq e^{(B^{1/q}+1)s} W(x) \quad \text{for all } s \geq 0. \tag{29}$$

If μ_0 is not a Dirac mass, we get, from equation (29) and from Assumption (A′2-iii), that $(s, x) \mapsto \mathbb{E}_x[(Q - I) f(X_s)]$ is integrable with respect to $\mathrm{d}s\,\mu(\mathrm{d}x)$ on $[0, t] \times E$. Therefore, we can use Fubini's theorem and get that, for all $t \geq 0$,

$$\mathbb{E}_{\mu_0} f(X_t) = \mu_0 \cdot f + \int_0^t \mathbb{E}_{\mu_0}\big[(Q - I) f(X_s)\big]\,\mathrm{d}s, \tag{30}$$

which means that $t \mapsto \mathbb{P}_{\mu_0}(X_t \in \cdot)$ is a solution of (28).

In both cases (μ_0 being a Dirac mass or not), $t \mapsto \mathbb{P}_{\mu_0}(X_t \in \cdot)$ is a solution of (28), and, thus, ν_t is a solution of (16). Since, by Assumption (A1), $\mathbb{P}_{\mu_0}(X_t \in E) \geq e^{-(1-c_1)t}$ for all $t \geq 0$, we get that

$$\nu_t \cdot W \leq \mathrm{e}^{(B^{1/q}+2-c_1)t} \nu_0 \cdot W \quad \text{for all } t \geq 0. \tag{31}$$

Step 2. Compactness. Let us now prove that $\nu_t \in \mathcal{P}_C(E)$ for all $t \geq 0$. We denote by T_N the first hitting time of $\{W \geq N\}$, that is,

$$T_N = \inf\{t \geq 0, W(X_t) \geq N\}.$$

Note that T_N is a stopping time for the natural filtration of the process (see for instance Theorem 2.4 in [6]). Using the fact that $(Q - I) \cdot W \leq (\theta - 1)W + K$ and Dynkin's formula, we obtain that, for all $x \in E$ and all $0 \leq s < t$,

$$\begin{aligned} &\mathbb{E}_x\big[\mathrm{e}^{(1-c_1)[(t-s)\wedge T_N]} W(X_{(t-s)\wedge T_N})\mathbf{1}_{(t-s)\wedge T_N<\tau_\partial}\big] \\ &\quad = W(x) + \mathbb{E}_x\bigg[\int_s^{t\wedge(s+T_N)} \mathrm{e}^{(1-c_1)(u-s)}\big((\theta - c_1)W(X_{u-s})\mathbf{1}_{u-s<\tau_\partial} + K\big)\,\mathrm{d}u\bigg]. \end{aligned} \tag{32}$$

The same computation with c_1 replaced by θ and $s = 0$ shows that, for any fixed $t \geq 0$, $\mathbb{E}_x[W(X_{t\wedge T_N})\mathbf{1}_{t\wedge T_N<\tau_\partial}]$ is uniformly bounded over $N \geq 1$, so that,

$$\mathbb{P}_x(T_N \leq t) \leq \mathbb{E}_x\bigg[\frac{W(X_{t\wedge T_N})}{N}\mathbf{1}_{t\wedge T_N<\tau_\partial}\bigg] \xrightarrow[N\to+\infty]{} 0,$$

where we have used Markov's inequality. This implies in particular that the almost surely nondecreasing sequence $(T_N)_{N\geq 0}$ converges to $+\infty$ almost surely. Using in addition Fatou's Lemma in the left-hand side of (32) and the monotone convergence theorem in the right-hand side (separating the W term and the K term and using the fact that $\theta < c_1$ and that T_N is almost surely nondecreasing), we obtain

$$\begin{aligned}&\mathbb{E}_x\big[\mathrm{e}^{(1-c_1)(t-s)}W(X_{t-s})\mathbf{1}_{t-s<\tau_\partial}\big]\\&\quad\leq W(x)+\int_s^t \mathrm{e}^{(1-c_1)(u-s)}\big((\theta-c_1)\mathbb{E}_x\big[W(X_{u-s})\mathbf{1}_{u-s<\tau_\partial}\big]+K\big)\,\mathrm{d}u.\end{aligned}$$

Integrating with respect to the law of X_s under $\mathbb{P}_{\mu_0}$ and using Fubini's theorem, we thus get that

$$\begin{aligned}\mathbb{E}_{\mu_0}\big[\mathrm{e}^{(1-c_1)t}W(X_t)\mathbf{1}_{t<\tau_\partial}\big]&\leq \mathbb{E}_{\mu_0}\big[\mathrm{e}^{(1-c_1)s}W(X_s)\mathbf{1}_{s<\tau_\partial}\big]\\&\quad+\int_s^t \mathrm{e}^{(1-c_1)u}\big((\theta-c_1)\mathbb{E}_{\mu_0}\big[W(X_u)\mathbf{1}_{u<\tau_\partial}\big]+K\big)\,\mathrm{d}u.\end{aligned}$$

This implies that $\mathbb{E}_{\mu_0}[\mathrm{e}^{(1-c_1)t}W(X_t)\mathbf{1}_{t<\tau_\partial}]\leq \mu_0\cdot W\vee\frac{K}{c_1-\theta}$ (we detail the proof of this implication in Lemma 8 below) and, since $\mathbb{P}_{\mu_0}(t<\tau_\partial)\geq e^{-(1-c_1)t}$, that $\nu_t\cdot W\leq \nu_0\cdot W\vee\frac{K}{c_1-\theta}$, for all $t\geq 0$, that is, that $\nu_t\in\mathcal{P}_{C\vee\frac{K}{c_1-\theta}}$ for all $t\geq 0$.

Step 3. Weak continuity of the semigroup. Our aim is to prove the continuity of $(\mu_0,t)\mapsto \mathbb{E}_{\mu_0}f(X_t)$ for any bounded continuous functions $f:E\to\mathbb{R}$. We prove first the continuity of the application

$$(x,t)\in E\times[0,+\infty)\mapsto \mathbb{E}_x f(X_t).$$

Recall that T_N is the first hitting time of $\{W\geq N\}$ and is a stopping time for the natural filtration of the process. We have, for all $x\in E$ and $t\geq 0$,

$$\begin{aligned}\big|\mathbb{E}_x f(X_t)-\mathbb{E}_x\big[f(X_{t\wedge T_N})\big]\big|&\leq 2\|f\|_\infty\mathbb{P}_x(T_N<t)\leq 2\|f\|_\infty\mathbb{E}_x\bigg[\frac{W(X_{t\wedge T_N})}{N}\bigg]\\&\leq 2\|f\|_\infty e^{(B^{1/q}+1)t}W(x)/N,\end{aligned}$$

where the last inequality is a consequence of Assumption (A′2-iv) and (29). In particular, since V is locally bounded, $(x,t)\mapsto\mathbb{E}_x f(X_t)$ is the locally-uniform limit (when $N\to+\infty$) of $(x,t)\mapsto\mathbb{E}_x[f(X_{t\wedge T_N})]$, which is continuous with respect to (x,t) since it is the expectation of a pure jump Markov process with uniformly-bounded continuous jump measure. As a consequence, the application $(x,t)\mapsto\mathbb{E}_x f(X_t)$ is continuous (and bounded).

Let us now prove that, for any bounded continuous function $f:E\to\mathbb{R}$, the function

$$(\mu_0,t)\mapsto\mathbb{E}_{\mu_0}f(X_t)$$

is continuous on $\mathcal{P}_C(E)\times[0,+\infty)$, for all $C\geq 0$. Let $\mu_n\in\mathcal{P}_C(E)\to\mu$ and $t_n\to t$ when $n\to+\infty$ (note that $\mu\in\mathcal{P}_C(E)$ since this set is closed for the topology of weak convergence). Then, we have

$$\begin{aligned}\big|\mathbb{E}_{\mu_n}f(X_{t_n})-\mathbb{E}_\mu f(X_t)\big|&\leq\big|\mathbb{E}_{\mu_n}\big[f(X_{t_n})-f(X_t)\big]\big|+\big|\mathbb{E}_{\mu_n}f(X_t)-\mathbb{E}_\mu f(X_t)\big|\\&\to 0\quad\text{when } n\to+\infty,\end{aligned}$$

where we used (for the first term in the right-hand side) the almost-sure continuity of $s\mapsto X_s$ at time t and the dominated convergence theorem, and (for the second term in the right-hand side) the continuity of $x\mapsto\mathbb{E}_x f(X_t)$ and the weak convergence of μ_n toward μ.

Step 4. Uniqueness. Let $t \mapsto \mu_t$ be a solution to (16) in $\mathcal{P}_C(E)$ for some $C \geq 0$ and let us consider

$$\theta_t := \exp\biggl(\int_0^t \mu_s(Q-I)(E)\,\mathrm{d}s\biggr)\mu_t.$$

By Assumption (A′2-iii), $|\mu_s(Q-I)(E)| \leq A^{1/q}\mu_s \cdot W + 1 \leq A^{1/q}C + 1$, so that θ_t is well defined for all $t \geq 0$. Moreover, for all bounded continuous functions $f : E \to \mathbb{R}$, $\theta_t \cdot f$ is differentiable and we have

$$\frac{\partial \theta_t \cdot f}{\partial t} = \mu_t(Q-I)(E)\theta_t \cdot f + \theta_t Q \cdot f - \mu_t Q(E)\theta_t \cdot f = \theta_t(Q-I) \cdot f.$$

Said differently, θ_t is solution to (28). Hence, for any continuous function f, we have

$$\frac{\mathrm{d}\mathbb{E}_{\theta_s} f(X_{t-s})}{\mathrm{d}s} = \theta_s(Q-I) \cdot \mathbb{E}_{\cdot} f(X_{t-s}) - \theta_s(Q-I) \cdot \mathbb{E}_{\cdot} f(X_{t-s}) = 0,$$

where we used (28) for $(\theta_t)_t$ to handle the first right-hand-side term (recall that $x \mapsto \mathbb{E}_x f(X_{t-s})$ is bounded continuous) and the backward Kolmogorov equation for the second right-hand-side term (see for instance Theorem 2.21 in [19]). This implies that $\theta_t \cdot f = \mathbb{E}_{\theta_0} f(X_t)$ and hence that

$$\mu_0 \cdot f = \frac{\theta_t \cdot f}{\theta_t(E)} = \frac{\mathbb{E}_{\mu_0} f(X_t)}{\mathbb{P}_{\mu_0}(X_t \neq \partial)} = \nu_t \cdot f$$

for all $t \geq 0$ and all bounded continuous functions $f : E \to \mathbb{R}$. This implies that $\mu = \nu$, which is thus the unique solution of (16). □

In Step 2 of the proof above, we used the following technical lemma:

LEMMA 8. *Let $g : [0, +\infty) \to \mathbb{R}$ and $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be two measurable functions such that $t \in [0, +\infty) \mapsto f(t, g(t)) \in \mathbb{R}$ is locally integrable. If*

$$g(t) - g(s) \leq \int_s^t f\bigl(u, g(u)\bigr)\,\mathrm{d}u \quad \forall 0 \leq s \leq t$$

and if there exists $M \in \mathbb{R}$ such that $f(u, g(u)) \leq 0$ for all $u \in [0, +\infty)$ such that $g(u) \geq M$. Then

$$g(t) \leq g(0) \vee M \quad \forall t \geq 0.$$

PROOF. We assume without loss of generality that $M \geq g(0)$ and proceed by contradiction: assume that there exist $\varepsilon > 0$ and $t \geq 0$ such that $g(t) \geq M + \varepsilon$ and let $t_0 = \inf\{t \geq 0 \text{ s.t. } g(t) \geq M + \varepsilon\}$. Note that, for all $t \geq t_0$,

$$g(t) \leq g(t_0) + \int_{t_0}^t f\bigl(u, g(u)\bigr)\,\mathrm{d}u \xrightarrow[t \downarrow t_0]{} g(t_0),$$

and hence $g(t_0) \geq M + \varepsilon$. Now, let $s_0 = \sup\{s \leq t_0 \text{ s.t. } g(s) \leq M\}$, and note that

$$g(s_0) \leq \liminf_{s \uparrow s_0} \biggl\{g(s) + \int_s^{s_0} f\bigl(u, g(u)\bigr)\,\mathrm{d}u\biggr\} = \liminf_{s \uparrow s_0} g(s),$$

implying that $g(s_0) \leq M$. Finally, since $g(s) \in [M, M + \varepsilon]$ for all $s \in [s_0, t_0]$, we have

$$M + \varepsilon \leq g(t_0) \leq g(s_0) + \int_{s_0}^{t_0} f\bigl(u, g(u)\bigr)\,\mathrm{d}u \leq g(s_0) \leq M.$$

□

We are now ready to prove Proposition 7:

PROOF OF PROPOSITION 7. Our approach is based on [7] (see also [11] for an application of this theorem on a set of probability measures on a compact space). In view of [57], since E is separable by assumption, there exists a metrization of the topology of E such that E is totally bounded (this distance is imposed on E from now on). Also, still by [57], there exists a family of bounded uniformly continuous functions $(g_k)_{k\geq 1}$ that is dense in $U(E,\mathbb{R})$, the set of all bounded uniformly-continuous functions from E to $\mathbb{R}$. Finally, [57], also states that a sequence $(\mu_n)_{n\in\mathbb{N}}$ of nonnegative measures converges weakly to μ if and only if $\mu_n \cdot g_k \to \mu \cdot g_k$ when $n \to +\infty$, for all $k \in \mathbb{N}$. We also consider the function $g_0 : x \in E \mapsto Q_x(E)$, which is continuous by Assumption (A4) and bounded by Assumption (A1), and the family of functions indexed by $k, M \in \mathbb{N}$ defined by

$$g_k^M(x) = -M \vee \big(Q \cdot g_k(x) \wedge M\big)$$

and which are continuous (by Assumption (A4)) and bounded. In particular, the distance

$$d(\mu_1,\mu_2) = \big|\mu_1 Q(E) - \mu_2 Q(E)\big| + \sum_{k=1}^{\infty} \frac{|\mu_1 \cdot g_k - \mu_2 \cdot g_k| \wedge 1}{2^k(1+\|g_k\|_\infty)} + \sum_{k=1,M=1}^{\infty} \frac{|\mu_1 \cdot g_k^M - \mu_2 \cdot g_k^M| \wedge 1}{2^{k+M}(1+\|g_k^M\|_\infty)}$$

is a metric for the weak convergence in the set of nonnegative measures on E.

We introduce the increasing sequence $(\tau_n)_{n\geq 1}$ defined as

$$\tau_n = \gamma_1 + \gamma_2 + \cdots + \gamma_n$$

(see Lemma 2 for the definition of γ_n) and we consider the time-changed and linearized versions $(\bar{\mu}_t)_{t\in[1,+\infty)}$ and $(\mu_t)_{t\in[1,+\infty)}$ of $(\tilde{\eta}_n)_{n\in\mathbb{N}}$ defined, for all $n \geq 1$ and all $t \in [\tau_n, \tau_{n+1}]$, by

$$\bar{\mu}_t = \tilde{\eta}_n \quad \text{and} \quad \mu_t = \tilde{\eta}_n + \frac{t-\tau_n}{\tau_{n+1}-\tau_n}(\tilde{\eta}_{n+1} - \tilde{\eta}_n).$$

Similarly, we define $\bar{U}_t = U_{n+1}$ for all $t \in [\tau_n, \tau_{n+1}]$ (see Lemma 2 for the definition of U_n).

To prove that $(\mu_t)_{t\geq 0}$ is an asymptotic pseudo-trajectory of the semiflow induced by (16), we apply [7], Theorem 3.2, (and refer the reader to [7] for the definition of an asymptotic pseudo-trajectory).

Note that $\mu_t \in \mathcal{P}_C(E)$ for all $t \geq 0$, and hence $(\mu_t)_{t\geq 0}$ has compact closure in $\mathcal{P}_C(E)$ (since this set is itself compact). Also, by construction, $t \mapsto \mu_t$ is uniformly continuous (and even Lipschitz) with respect to the distance d on $\mathcal{P}_C(E)$. Indeed, for all $s, t \in [\tau_n, \tau_{n+1}]$,

$$\begin{aligned} d(\mu_s,\mu_t) &= \frac{t-s}{\tau_{n+1}-\tau_n} d(\tilde{\eta}_{n+1}, \tilde{\eta}_n) = \frac{t-s}{\gamma_{n+1}} d(\tilde{\eta}_{n+1}, \tilde{\eta}_n) \\ &\leq (t-s)\big(2\|Q(E)\|_\infty + 4\big), \end{aligned}$$

where we have used the fact (see Lemma 2) that, for all bounded measurable function $g : E \to \mathbb{R}_+$,

$$\left|\frac{\tilde{\eta}_{n+1} \cdot g - \tilde{\eta}_n \cdot g}{\gamma_{n+1}}\right| = \big|\tilde{\eta}_n Q(E)\big(g(Y_{n+1}) - \tilde{\eta}_n \cdot g\big)\big| \leq 2\|g\|_\infty.$$

Therefore, to apply [7], Theorem 3.2, it only remains to prove that all limit points of $(\Theta_t(\mu))_{t\geq 0}$ in $C(\mathbb{R}_+, \mathcal{P}_C(E))$ endowed with the topology of uniform convergence on compact sets are solutions of (16), where $\Theta_t(\mu) := (\mu_{t+s})_{s\geq 0}$. Let $\mu^\infty \in C(\mathbb{R}_+, \mathcal{P}_C(E))$ be such a limit point: in other words, we assume that there exists an increasing sequence of positive numbers $(t_n)_{n\geq 0}$ converging to $+\infty$ such that $(\Theta_{t_n}(\mu))_{n\geq 0}$ converges to μ^∞ in $C(\mathbb{R}_+, \mathcal{P}_C(E))$.

For all $t \in [\tau_n, \tau_{n+1})$ and all $s \geq 0$ such that $t+s \in [\tau_m, \tau_{m+1})$, we deduce from Lemma 2 that

$$\begin{aligned}\int_t^{t+s} F(\bar{\mu}_u) + \bar{U}_u \,\mathrm{d}u &\\ &= (\tau_{n+1} - t)\big(F(\tilde{\eta}_n) + U_{n+1}\big)\\ &\quad + \sum_{k=n+1}^{m-1} \gamma_{k+1}\big(F(\tilde{\eta}_k) + U_{k+1}\big) + (t+s-\tau_m)\big(F(\tilde{\eta}_m) + U_{m+1}\big)\\ &= \frac{\tau_{n+1}-t}{\tau_{n+1}-\tau_n}(\tilde{\eta}_{n+1} - \tilde{\eta}_n) + \tilde{\eta}_m - \tilde{\eta}_{n+1} + \frac{t+s-\tau_m}{\tau_{m+1}-\tau_m}(\tilde{\eta}_{m+1} - \tilde{\eta}_m)\\ &= -\mu_t + \mu_{t+s}.\end{aligned} \tag{33}$$

For all $k \in \mathbb{N}$, we define $L_F^k : C(\mathbb{R}_+, \mathcal{P}_C(E)) \to \mathbb{R}^{[0,+\infty)}$ by

$$L_F^k(\nu)(t) = \nu_0 + \int_0^t F(\nu_s) \cdot g_k \,\mathrm{d}s,$$

for any $\nu \in C(\mathbb{R}_+, \mathcal{P}_C(E))$ (see Lemma 2 for the definition of the function F), so that, by equation (33),

$$\Theta_t(\mu) \cdot g_k = L_F^k\big(\Theta_t(\mu)\big) + A_t^k + B_t^k, \tag{34}$$

where, for all $s \geq 0$,

$$A_t^k(s) = \int_t^{t+s} F(\bar{\mu}_u) \cdot g_k - F(\mu_u) \cdot g_k \,\mathrm{d}u \quad \text{and} \quad B_t^k(s) = \int_t^{t+s} \bar{U}_u \cdot g_k \,\mathrm{d}u.$$

The rest of the proof is divided into four steps: The first two steps are devoted to prove that A_t^k and, respectively, B_t^k converge uniformly to 0 on compact sets when $t \to +\infty$. In the third step, we prove that $L_F^k(\Theta_{t_n}(\mu))$ converges to $L_F^k(\mu^\infty)$ for all subsequence $t_n \to +\infty$ such that $(\Theta_{t_n}(\mu))_{n\geq 0}$ converges to μ^∞ in $C(\mathbb{R}_+, \mathcal{P}_C(E))$. Finally, in the fourth step, we conclude the proof of Proposition 7.

Step 1: A_t^k *converges to* 0. For all $u \in [\tau_n, \tau_{n+1})$, we have

$$\begin{aligned}&\big|F(\bar{\mu}_u) \cdot g_k - F(\mu_u) \cdot g_k\big|\\ &\quad\leq |\bar{\mu}_u Q \cdot g_k - \mu_u Q \cdot g_k| + \big|\bar{\mu}_u Q(E)\bar{\mu}_u \cdot g_k - \mu_u Q(E)\mu_u \cdot g_k\big|\\ &\quad\leq |\tilde{\eta}_{n+1} Q \cdot g_k - \tilde{\eta}_n Q \cdot g_k| + \|g_k\|_\infty \big|\bar{\mu}_u Q(E) - \mu_u Q(E)\big| + |\bar{\mu}_u \cdot g_k - \mu_u \cdot g_k|\\ &\quad\leq |\tilde{\eta}_{n+1} Q \cdot g_k - \tilde{\eta}_n Q \cdot g_k| + \|g_k\|_\infty \big|\tilde{\eta}_{n+1} Q(E) - \tilde{\eta}_n Q(E)\big|\\ &\qquad + |\tilde{\eta}_{n+1} \cdot g_k - \tilde{\eta}_n \cdot g_k|\\ &\quad\leq \frac{1}{n+1} |Q_{Y_{n+1}} \cdot g_k - \tilde{\eta}_n Q \cdot g_k| + \frac{\|g_k\|_\infty}{n+1} \big|Q_{Y_{n+1}}(E) - \tilde{\eta}_n Q(E)\big|\\ &\qquad + \frac{1}{n+1} \big|g_k(Y_{n+1}) - \tilde{\eta}_n \cdot g_k\big|\\ &\quad\leq \frac{\|g_k\|_\infty}{n+1} \big(B^{1/q} V(Y_{n+1})^{1/q} + B^{1/q} C + 1 + 2\big),\end{aligned}$$

where we used Assumptions (A′2-iii) and (A1) and the fact that, almost surely, $\eta_m \in \mathcal{P}_C(E)$ for all $n \geq 0$. Hence, if we denote by n_t the unique integer such that $t \in [\tau_{n_t}, \tau_{n_t+1})$, for any $t \geq 0$ (such an integer exists since $\tau_n \to +\infty$ when $n \to +\infty$), we have, for all $s \geq 0$,

$$\begin{aligned} A_t^k(s) &\leq \frac{\|g\|_k}{n_t+1} \sum_{k=n_t}^{n_{t+s}} \gamma_{k+1} B^{1/q} V(Y_{k+1})^{1/q} + \frac{\|g\|_k (B^{1/q}C+3)s}{n_t+1} \\ &\leq \frac{\|g\|_k}{n_t+1} \frac{B^{1/q}}{c_1} \tilde{\eta}_{n_{t+s}+1} \cdot V^{1/q} + \frac{\|g\|_k (B^{1/q}C+3)s}{n_t+1}, \end{aligned}$$

where we used that $\gamma_n \leq 1/(c_1 n)$, for all $n \geq 1$, by Assumption (A1). Finally, for all $T \geq 0$, we have

$$\sup_{s\in[0,T]} \left|A_t^k(s)\right| \leq \frac{T\|g_k\|_\infty (B^{1/q}C + B^{1/q}C/c_1 + 3)}{n_t+1} \to 0 \quad \text{when } t \to +\infty.$$

Step 2: *B_t^k converges to* 0. We have, for all $t \in [\tau_n, \tau_{n+1})$ and $t+s \in [\tau_{n+m}, \tau_{n+m+1})$,

$$\begin{aligned} \left|B_t^k(s)\right| &\leq (\tau_{n+1} - t)|U_{n+1} \cdot g_k| + \left|\sum_{\ell=n+1}^{n+m-1} \gamma_{\ell+1} U_{\ell+1} \cdot g_k\right| + (s - \tau_{n+m})|U_{n+m+1} \cdot g_k| \\ &\leq \gamma_{n+1}|U_{n+1} \cdot g_k| + \left|\sum_{\ell=n+1}^{n+m-1} \gamma_{\ell+1} U_{\ell+1} \cdot g_k\right| + \gamma_{n+m+1}|U_{n+m+1} \cdot g_k|. \end{aligned}$$

Using a similar approach as in the proof of Lemma 5, one easily obtains that, for any bounded continuous function $f : E \to \mathbb{R}$, $\sum_{\ell=0}^n \gamma_{\ell+1} U_{\ell+1} \cdot f$ converges almost surely when $n \to +\infty$. Hence, we have that, almost surely,

$$\lim_{n\to+\infty} \sup_{m\geq 1} \left\{ \gamma_{n+1}|U_{n+1} \cdot g_k| + \left|\sum_{\ell=n+1}^{n+m-1} \gamma_{\ell+1} U_{\ell+1} \cdot g_k\right| + \gamma_{n+m+1}|U_{n+m+1} \cdot g_k| \right\} = 0.$$

In particular, we have that, for all $T \geq 0$,

$$\sup_{s\in[0,T]} \left|B_t^k(s)\right| \to 0 \quad \text{when } t \to +\infty.$$

Step 3: *$L_F^k(\Theta_{t_n}(\mu))$ converges to $L_F^k(\mu^\infty)$ for all subsequence $t_n \to +\infty$ such that $(\Theta_{t_n}(\mu))_{n\geq 0}$ converges to μ^∞ in $C(\mathbb{R}_+, \mathcal{P}_C(E))$*. To prove this, it is enough to show that L_F^k is sequentially continuous in $C(\mathbb{R}_+, \mathcal{P}_C(E))$. Let $(\nu^n)_{n\geq 0}$ be a sequence of elements of $C(\mathbb{R}_+, \mathcal{P}_C(E))$ which converges to $\nu \in C(\mathbb{R}_+, \mathcal{P}_C(E))$. For all $n \geq 0$ and all $t \geq 0$, we have

$$(35) \quad \left|L_F^k(\nu^n)(t) - L_F^k(\nu)(t)\right| \leq \left|\nu_0^n \cdot g_k - \nu_0 \cdot g_k\right| + \int_0^t \left|F(\nu_s^n) \cdot g_k - F(\nu_s) \cdot g_k\right| \mathrm{d}s.$$

The first term of the right-hand side converges to 0 because of the weak convergence of $(\nu_0^n)_{n\geq 0}$ to ν. Let us now focus on the second term of the right-hand side; we have

$$\begin{aligned} &\left|F(\nu_s^n) \cdot g_k - F(\nu_s) \cdot g_k\right| \\ &\quad \leq \left|\nu_s^n Q \cdot g_k - \nu_s Q \cdot g_k\right| + \left|\nu_s^n Q(E) \nu_s^n \cdot g_k - \nu_s Q(E) \nu_s g_k\right|. \end{aligned}$$

Since ν^n converges uniformly on compact sets toward ν, we deduce that the term $s \mapsto |\nu_s^n Q(E) \nu_s^n \cdot g_k - \nu_s Q(E) \nu_s g_k|$ converges uniformly to 0 on compact sets when $n \to +\infty$ (we use here the fact that $g_0 = Q.(E)$ appears in the distance d). Moreover, since $\nu_s^n \in \mathcal{P}_C(E)$

and since $|Q \cdot g_k| \le B^{1/q'} \|g_k\|_\infty W^{q/q'}$ by Assumption (A2-iii) (recall that $W := V^{1/q}$), we deduce that, for all $M \ge 1$,

$$\begin{aligned}
|\nu_s^n Q \cdot g_k - \nu_s Q \cdot g_k| &\le |(\nu_s^n - \nu_s) g_k^M| + (\nu_s^n + \nu_s)|Q \cdot g_k - g_k^M| \\
&\le |(\nu_s^n - \nu_s) g_k^M| + (\nu_s^n + \nu_s)|Q \cdot g_k \mathbf{1}_{|Q \cdot g_k| > M}| \\
&\le |(\nu_s^n - \nu_s) g_k^M| + B^{1/q'} \|g_k\|_\infty (\nu_s^n + \nu_s) |W^{q/q'} \mathbf{1}_{B^{1/q} \|g_k\|_\infty^{q'/q} W > M^{q'/q}}| \\
&\le |(\nu_s^n - \nu_s) g_k^M| + \frac{B^{1/q} \|g_k\|_\infty^{q'/q}}{M^{q'/q-1}} (\nu_s^n + \nu_s)(W) \\
&\le |(\nu_s^n - \nu_s) g_k^M| + \frac{B^{1/q} \|g_k\|_\infty^{q'/q} 2C}{M^{q'/q-1}},
\end{aligned}$$

where we have used the fact that $\nu_s^n \in \mathcal{P}_C(E)$ for all $n \in \mathbb{N}$ and all $s \ge 0$. The term $\frac{B^{1/q} \|g_k\|_\infty^{q'/q} 2C}{M^{q'/q-1}}$ goes to 0 when $M \to +\infty$ uniformly in $s \ge 0$ and the term $|(\nu_s^n - \nu_s) g_k^M|$ converges to 0 uniformly in s in compact sets. As a consequence, we deduce that $|\nu_s^n Q \cdot g_k - \nu_s Q \cdot g_k|$ converges to 0 uniformly in s in compact sets. This allows us to conclude that the second term of the right hand side of (35) converges to 0 when $n \to +\infty$, which was the aim of Step 3.

Step 4: *conclusion*. Steps 1 to 3 above entail that any limit point μ^∞ of $(\Theta_t(\mu))_{t \ge 0}$ satisfies

$$\mu_t^\infty \cdot g_k = \mu_0^\infty \cdot g_k + \int_0^t F(\mu_s^\infty) \cdot g_k \,\mathrm{d}s \quad (\forall k \ge 1).$$

Since $(g_k)_{k \ge 1}$ is dense in the set $U(E, \mathbb{R})$, we conclude (see for instance [57], Lemma 2.3) that

$$\mu_t^\infty = \mu_0^\infty + \int_0^t F(\mu_s^\infty) \,\mathrm{d}s.$$

As a consequence, μ^∞ is solution to the dynamical system (16). Using [7], Theorem 3.2, we deduce that $(\mu_t)_{t \ge 0}$ is a pseudo asymptotic trajectory in $\mathcal{P}_C(E)$ for the semiflow induced by the well-posed dynamical system (16) in $\mathcal{P}_C(E)$. Therefore, Assumption (A3) entails that the set of limit points of $(\mu_t)_{t \ge 0}$ is included in the uniformly attracting set $\{\nu\}$ of the semiflow generated by (16). In particular, the only limit point of the compact sequence $(\tilde{\eta}_n)_{n \ge 1}$ is ν. This concludes the proof of Proposition 7. □

REMARK 12. Without Assumption (A3), we still get that $(\mu_t)_{t \ge 0}$ is a pseudo asymptotic trajectory in $\mathcal{P}_C(E)$ for the semiflow induced by the well-posed dynamical system (16) in $\mathcal{P}_C(E)$. In particular, the set of limit points of $(\mu_t)_{t \ge 0}$ is included in the limit sets of the flow (see [7], Section 5.2).

3.2. *Proof of Theorem 1 from Proposition 7*. Fix $c' \in (\theta, c_1)$. For all $k \ge 1$, we define

$$\sigma_k := \inf\{n \ge k, m_n P(E) < c' n\}.$$

For all $n \ge 1$ and any bounded continuous function $f : E \to \mathbb{R}$, we set $\Psi_n = m_{n \wedge \sigma_k} \cdot f - \eta_{n \wedge \sigma_k} R \cdot f$, so that $(\Psi_n)_{n \ge 1}$ is a martingale and

$$\Psi_n = m_0 \cdot f + \sum_{i=1}^{n \wedge \sigma_k} (R_{Y_i}^{(i)} \cdot f - R_{Y_i} \cdot f).$$

An immediate adaptation of Theorem 1.3.17 in [28] tells us that if the sequence $(n^{-1}\mathbb{E}[|\Psi_n|^r])_{n\geq 1}$ is bounded, then $n^{-1}\Psi_n$ goes almost surely to zero when n goes to infinity. We have, using Lemma 1 in [18],

$$\begin{aligned}\frac{\mathbb{E}[|\Psi_n|^r]}{n} &\leq \frac{2(m_0\cdot f)^r}{n} + \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\big[\big|R^{(i)}_{Y_i}\cdot f - R_{Y_i}\cdot f\big|^r\mathbf{1}_{i\leq\sigma_k}\big] \\ &\leq \frac{2(m_0\cdot f)^r}{n} + \frac{2\|f\|^r_\infty}{n}\sum_{i=1}^{n}A\mathbb{E}\big[V(Y_i)\mathbf{1}_{i\leq\sigma_k}\big],\end{aligned}$$

where we used the fact that $\mathbf{1}_{i\leq\sigma_k}$ is $\mathcal{F}_{i-1}$-measurable and independent of Y_i and Assumption (A$'$2-iii).

Using Lemma 4, we deduce that the sequence $(n^{-1}\mathbb{E}[|\Psi_n|^r])_n$ is uniformly bounded and hence that $n^{-1}\Psi_n$ goes almost surely to zero when n goes to infinity (since we have assumed, in particular, that $m_0\cdot V<+\infty$, which entails $m_0(E)<\infty$).

Since this is true for any $k\geq 1$ and since $\mathbb{P}(\bigcup_{k=1}^{\infty}\{\sigma_k=+\infty\})=1$ (see Lemma 3), we deduce that, almost surely, $m_n(f)=\eta_n R(f)+o(n)$ when n goes to infinity. In view of Proposition 7, and by Assumption (A4) (namely continuity of R), we get that $(\eta_n R\cdot f/n)_{n\geq 1}$ and $(\eta_n R(E)/n)_{n\geq 1}$ converge almost surely to $\nu R\cdot f$ and $\nu R(E)$ respectively, which concludes the proof of the first part and the last part of Theorem 1.

To get the almost-sure boundedness of $m_n P\cdot V^{1/q}/n$, recall that, by definition, $m_n = m_0+\sum_{i=1}^{n}R^{(i)}_{Y_i}$, implying that, for all $n\geq 0$,

$$m_n P\cdot V^{1/q} = m_0 P\cdot V^{1/q} + \sum_{i=1}^{n} Q^{(i)}_{Y_i}\cdot V^{1/q}.$$

As above, we let

$$\Phi_n = m_0 P\cdot V^{1/q} + \sum_{i=1}^{n\wedge\sigma_k}\big(Q^{(i)}_{Y_i}\cdot V^{1/q} - Q_{Y_i}\cdot V^{1/q}\big).$$

The sequence $(\Phi_n)_{n\geq 0}$ is a martingale, and, similarly as above, we get that

$$\begin{aligned}\frac{\mathbb{E}|\Phi_n|^r}{n} &\leq \frac{2|m_0 P\cdot V^{1/q}|^r}{n} + \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}\big[\big|Q^{(i)}_{Y_i}\cdot V^{1/q} - Q_{Y_i}\cdot V^{1/q}\big|^r\mathbf{1}_{i\leq\sigma_k}\big] \\ &\leq \frac{2|m_0 P\cdot V^{1/q}|^r}{n} + \frac{2B}{n}\sum_{i=1}^{n}\mathbb{E}\big[V(Y_i)\mathbf{1}_{i\leq\sigma_k}\big].\end{aligned}$$

Using Lemma 4, we imply that $(\mathbb{E}|\Phi_n|^r/n)_{n\geq 0}$ is uniformly bounded, and thus that Φ_n/n converges almost surely to 0 when $n\to\infty$. Therefore, we have that, almost surely when $n\to\infty$,

$$\frac{m_n P\cdot V^{1/q}}{n} = \frac{1}{n}\sum_{i=1}^{n} Q_{Y_i}\cdot V^{1/q} + o(1) = \tilde{\eta}_n Q\cdot V^{1/q} + o(1).$$

Note that, by Assumption (A$'$2-iv), we have

$$\big|\tilde{\eta}_n Q\cdot V^{1/q}\big| \leq B^{1/q}\tilde{\eta}_n\cdot V^{1/q},$$

and recall that, by equation (27), $\tilde{\eta}_n\cdot V^{1/q}$ is almost surely uniformly bounded. We can thus conclude that $m_n P\cdot V^{1/q}/n$ is almost surely uniformly bounded, as claimed.

Acknowledgments. The authors would like to thank Michel Benaïm, Pascal Maillard and Andi Q. Wang for their useful comments and suggestions on this paper.

Cécile Mailler is grateful to EPSRC for support through the Fellowship EP/R022186/1.

REFERENCES

[1] ALDOUS, D. (1991). Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** 228–266. MR1102319

[2] ALDOUS, D., FLANNERY, B. and PALACIOS, J. L. (1988). Two applications of urn processes the fringe analysis of search trees and the simulation of quasi-stationary distributions of Markov chains. *Probab. Engrg. Inform. Sci.* **2** 293–307.

[3] ATHREYA, K. B. and KARLIN, S. (1968). Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. *Ann. Math. Stat.* **39** 1801–1817. MR0232455 https://doi.org/10.1214/aoms/1177698013

[4] BANDYOPADHYAY, A., JANSON, S. and THACKER, D. (2019). Strong convergence of infinite color balanced urns under uniform ergodicity. Preprint. Available at arXiv:1904.06144.

[5] BANDYOPADHYAY, A. and THACKER, D. (2016). A new approach to Pólya urn schemes and its infinite color generalization. Preprint. Available at arxiv:1606.05317.

[6] BASS, R. F. (2010). The measurability of hitting times. *Electron. Commun. Probab.* **15** 99–105. MR2606507 https://doi.org/10.1214/ECP.v15-1535

[7] BENAÏM, M. (1999). Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII. Lecture Notes in Math.* **1709** 1–68. Springer, Berlin. MR1767993 https://doi.org/10.1007/BFb0096509

[8] BENAÏM, M., CHAMPAGNAT, N. and VILLEMONAIS, D. (2019). Stochastic approximation of quasi-stationary distributions for diffusion processes in a bounded domain. Preprint. Available at arXiv:1904.08620.

[9] BENAÏM, M. and CLOEZ, B. (2015). A stochastic approximation approach to quasi-stationary distributions on finite spaces. *Electron. Commun. Probab.* **20** no. 37, 14. MR3352332 https://doi.org/10.1214/ECP.v20-3956

[10] BENAIM, M., CLOEZ, B. and PANLOUP, F. (2018). Stochastic approximation of quasi-stationary distributions on compact spaces and applications. *Ann. Appl. Probab.* **28** 2370–2416. MR3843832 https://doi.org/10.1214/17-AAP1360

[11] BENAÏM, M., LEDOUX, M. and RAIMOND, O. (2002). Self-interacting diffusions. *Probab. Theory Related Fields* **122** 1–41. MR1883716 https://doi.org/10.1007/s004400100161

[12] BERGERON, F., FLAJOLET, P. and SALVY, B. (1992). Varieties of increasing trees. In *Colloquium on Trees in Algebra and Programming. Lecture Notes in Computer Science* **581** 24–48. Springer, Berlin. MR1251994 https://doi.org/10.1007/3-540-55251-0_2

[13] BLANCHET, J., GLYNN, P. and ZHENG, S. (2016). Analysis of a stochastic approximation algorithm for computing quasi-stationary distributions. *Adv. in Appl. Probab.* **48** 792–811. MR3568892 https://doi.org/10.1017/apr.2016.28

[14] BÓNA, M. (2014). k-protected vertices in binary search trees. *Adv. in Appl. Math.* **53** 1–11. MR3149690 https://doi.org/10.1016/j.aam.2013.09.003

[15] CATTIAUX, P., COLLET, P., LAMBERT, A., MARTÍNEZ, S., MÉLÉARD, S. and SAN MARTÍN, J. (2009). Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Probab.* **37** 1926–1969. MR2561437 https://doi.org/10.1214/09-AOP451

[16] CHAMPAGNAT, N. and VILLEMONAIS, D. (2017). General criteria for the study of quasi-stationarity. Preprint. Available at arXiv:1712.08092.

[17] CHAMPAGNAT, N. and VILLEMONAIS, D. (2017). Uniform convergence to the Q-process. *Electron. Commun. Probab.* **22** Paper No. 33, 7. MR3663104 https://doi.org/10.1214/17-ECP63

[18] CHATTERJI, S. D. (1969). An L^p-convergence theorem. *Ann. Math. Stat.* **40** 1068–1070. MR0251769 https://doi.org/10.1214/aoms/1177697609

[19] CHEN, M.-F. (2004). *From Markov Chains to Non-equilibrium Particle Systems*, 2nd ed. World Scientific Co., Inc., River Edge, NJ. MR2091955 https://doi.org/10.1142/9789812562456

[20] CHEON, G.-S. and SHAPIRO, L. W. (2008). Protected points in ordered trees. *Appl. Math. Lett.* **21** 516–520. MR2402845 https://doi.org/10.1016/j.aml.2007.07.001

[21] COLLET, P., MARTÍNEZ, S. and SAN MARTÍN, J. (2013). *Quasi-Stationary Distributions: Markov Chains, Diffusions and Dynamical Systems. Probability and Its Applications (New York)*. Springer, Heidelberg. MR2986807 https://doi.org/10.1007/978-3-642-33131-2

[22] DARROCH, J. N. and SENETA, E. (1967). On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *J. Appl. Probab.* **4** 192–196. MR0212866 https://doi.org/10.2307/3212311

[23] DE OLIVEIRA, M. M. and DICKMAN, R. (2006). Quasi-stationary simulation: The subcritical contact process. *Braz. J. Phys.* **36** 685–689.

[24] DESHAYES, A. and ROLLA, L. T. (2017). Scaling limit of subcritical contact process. *Stochastic Process. Appl.* **127** 2630–2649. MR3660885 https://doi.org/10.1016/j.spa.2016.12.005

[25] DEVROYE, L. and JANSON, S. (2014). Protected nodes and fringe subtrees in some random trees. *Electron. Commun. Probab.* **19** no. 6, 10. MR3164753 https://doi.org/10.1214/ECP.v19-3048

[26] DHARMADHIKARI, S. W., FABIAN, V. and JOGDEO, K. (1968). Bounds on the moments of martingales. *Ann. Math. Stat.* **39** 1719–1723. MR0230363 https://doi.org/10.1214/aoms/1177698339

[27] DI GESÙ, G., LELIÈVRE, T., LE PEUTREC, D. and NECTOUX, B. (2016). Jump Markov models and transition state theory: The quasi-stationary distribution approach. *Faraday Discuss.* **195** 469–495.

[28] DUFLO, M. (1997). *Random Iterative Models. Applications of Mathematics (New York)* **34**. Springer, Berlin. MR1485774 https://doi.org/10.1007/978-3-662-12880-0

[29] EGGENBERGER, F. and PÓLYA, G. (1923). Über die statistik verketetter vorgäge. *ZAMM Z. Angew. Math. Mech.* **1** 279–289.

[30] FELLER, W. (1940). On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. Math. Soc.* **48** 488–515. MR0002697 https://doi.org/10.2307/1990095

[31] FERRARI, P. A., KESTEN, H. and MARTÍNEZ, S. (1996). R-positivity, quasi-stationary distributions and ratio limit theorems for a class of probabilistic automata. *Ann. Appl. Probab.* **6** 577–616. MR1398060 https://doi.org/10.1214/aoap/1034968146

[32] FERRARI, P. A. and MARIĆ, N. (2007). Quasi stationary distributions and Fleming–Viot processes in countable spaces. *Electron. J. Probab.* **12** 684–702. MR2318407 https://doi.org/10.1214/EJP.v12-415

[33] GOSSELIN, F. (2001). Asymptotic behavior of absorbing Markov chains conditional on nonabsorption for applications in conservation biology. *Ann. Appl. Probab.* **11** 261–284. MR1825466 https://doi.org/10.1214/aoap/998926993

[34] GRIGORESCU, I. and KANG, M. (2004). Hydrodynamic limit for a Fleming–Viot type system. *Stochastic Process. Appl.* **110** 111–143. MR2052139 https://doi.org/10.1016/j.spa.2003.10.010

[35] GROISMAN, P. and JONCKHEERE, M. (2013). Simulation of quasi-stationary distributions on countable spaces. *Markov Process. Related Fields* **19** 521–542. MR3156964

[36] HOLMGREN, C., JANSON, S. and ŠILEIKIS, M. (2017). Multivariate normal limit laws for the numbers of fringe subtrees in m-ary search trees and preferential attachment trees. *Electron. J. Combin.* **24** Paper 2.51, 49. MR3690263 https://doi.org/10.37236/6374

[37] JANSON, S. (2004). Functional limit theorems for multitype branching processes and generalized Pólya urns. *Stochastic Process. Appl.* **110** 177–245. MR2040966 https://doi.org/10.1016/j.spa.2003.12.002

[38] JANSON, S. (2005). Asymptotic degree distribution in random recursive trees. *Random Structures Algorithms* **26** 69–83. MR2116576 https://doi.org/10.1002/rsa.20046

[39] JANSON, S. (2019). Random replacements in Pólya urns with infinitely many colours. *Electron. Commun. Probab.* **24** Paper No. 23, 11. MR3953824 https://doi.org/10.1214/19-ECP226

[40] LARUELLE, S. and PAGÈS, G. (2013). Randomized urn models revisited using stochastic approximation. *Ann. Appl. Probab.* **23** 1409–1436. MR3098437 https://doi.org/10.1214/12-aap875

[41] LARUELLE, S. and PAGÈS, G. (2019). Nonlinear randomized urn models: A stochastic approximation viewpoint. *Electron. J. Probab.* **24** Paper No. 98, 47. MR4017116 https://doi.org/10.1214/19-ejp312

[42] MAHMOUD, H. M. and SMYTHE, R. T. (1992). Asymptotic joint normality of outdegrees of nodes in random recursive trees. *Random Structures Algorithms* **3** 255–266. MR1164839 https://doi.org/10.1002/rsa.3240030305

[43] MAHMOUD, H. M. and WARD, M. D. (2015). Asymptotic properties of protected nodes in random recursive trees. *J. Appl. Probab.* **52** 290–297. MR3336863 https://doi.org/10.1239/jap/1429282623

[44] MAILLARD, P. and PAQUETTE, E. (2016). Choices and intervals. *Israel J. Math.* **212** 337–384. MR3504330 https://doi.org/10.1007/s11856-016-1289-6

[45] MAILLER, C. and MARCKERT, J.-F. (2017). Measure-valued Pólya urn processes. *Electron. J. Probab.* **22** Paper No. 26, 33. MR3629870 https://doi.org/10.1214/17-EJP47

[46] MARIĆ, N. (2015). Fleming–Viot particle system driven by a random walk on $\mathbb{N}$. *J. Stat. Phys.* **160** 548–560. MR3366092 https://doi.org/10.1007/s10955-015-1275-0

[47] MARTÍNEZ, S., SAN MARTÍN, J. and VILLEMONAIS, D. (2014). Existence and uniqueness of a quasistationary distribution for Markov processes with fast return from infinity. *J. Appl. Probab.* **51** 756–768. MR3256225 https://doi.org/10.1239/jap/1409932672

[48] MÉLÉARD, S. and VILLEMONAIS, D. (2012). Quasi-stationary distributions and population processes. *Probab. Surv.* **9** 340–410. MR2994898 https://doi.org/10.1214/11-PS191

[49] MEYN, S. P. and TWEEDIE, R. L. (1993). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295 https://doi.org/10.2307/1427522

[50] OÇAFRAIN, W. and VILLEMONAIS, D. (2017). Convergence of a non-failable mean-field particle system. *Stoch. Anal. Appl.* **35** 587–603. MR3651131 https://doi.org/10.1080/07362994.2017.1288136

[51] PARTHASARATHY, K. R. (1967). *Probability Measures on Metric Spaces*. *Probability and Mathematical Statistics* **3**. Academic Press, New York. MR0226684

[52] PEMANTLE, R. (2007). A survey of random processes with reinforcement. *Probab. Surv.* **4** 1–79. MR2282181 https://doi.org/10.1214/07-PS094

[53] POUYANNE, N. (2008). An algebraic approach to Pólya processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 293–323. MR2446325 https://doi.org/10.1214/07-AIHP130

[54] RENLUND, H. (2010). Generalized Pólya urns via stochastic approximation. Preprint. Available at arXiv:1002.3716.

[55] STROOCK, D. W. and VARADHAN, S. R. S. (2007). *Multidimensional Diffusion Processes*. *Classics in Mathematics*. Springer, Berlin. MR2190038

[56] VAN DOORN, E. A. and POLLETT, P. K. (2013). Quasi-stationary distributions for discrete-state models. *European J. Oper. Res.* **230** 1–14. MR3063313 https://doi.org/10.1016/j.ejor.2013.01.032

[57] VARADARAJAN, V. S. (1958). Weak convergence of measures on separable metric spaces. *Sankhyā* **19** 15–22. MR0094838

[58] VERBOOM, J., LANKESTER, K. and METZ, J. A. J. (1991). Linking local and regional dynamics in stochastic metapopulation models. *Biol. J. Linn. Soc.* **42** 39–55.

[59] VILLEMONAIS, D. (2015). Minimal quasi-stationary distribution approximation for a birth and death process. *Electron. J. Probab.* **20** no. 30, 18. MR3325100 https://doi.org/10.1214/EJP.v20-3482

[60] WANG, A. Q., KOLB, M., ROBERTS, G. O. and STEINSALTZ, D. (2019). Theoretical properties of quasi-stationary Monte Carlo methods. *Ann. Appl. Probab.* **29** 434–457. MR3910008 https://doi.org/10.1214/18-AAP1422

[61] WANG, A. Q., ROBERTS, G. O. and STEINSALTZ, D. (2018). An approximation scheme for quasi-stationary distributions of killed diffusions. Preprint. Available at arXiv:1808.07086.

[62] ZHANG, L.-X. (2016). Central limit theorems of a recursive stochastic algorithm with applications to adaptive designs. *Ann. Appl. Probab.* **26** 3630–3658. MR3582813 https://doi.org/10.1214/16-AAP1187