Available online at www.sciencedirect.com
SCIENCE DIRECT®
Stochastic Processes and their Applications 116 (2006) 1127–1160
stochastic processes and their applications
www.elsevier.com/locate/spa

A microscopic interpretation for adaptive dynamics trait substitution sequence models

Nicolas Champagnat*

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, 10117 Berlin, Germany

Received 14 January 2005; received in revised form 2 December 2005; accepted 13 January 2006
Available online 8 February 2006

Abstract

We consider an interacting particle Markov process for Darwinian evolution in an asexual population with non-constant population size, involving a linear birth rate, a density-dependent logistic death rate, and a probability μ of mutation at each birth event. We introduce a renormalization parameter K scaling the size of the population, which leads, when $K \to +\infty$, to a deterministic dynamics for the density of individuals holding a given trait. By combining in a non-standard way the limits of large population ($K \to +\infty$) and of small mutations ($\mu \to 0$), we prove that a timescale separation between the birth and death events and the mutation events occurs and that the interacting particle microscopic process converges for finite dimensional distributions to the biological model of evolution known as the "monomorphic trait substitution sequence" model of adaptive dynamics, which describes the Darwinian evolution in an asexual population as a Markov jump process in the trait space.

MSC: 60F99; 60K35; 92D15

Keywords: Measure-valued process; Interacting particle system; Mutation-selection processes; Darwinian evolution; Trait substitution sequence; Adaptive dynamics; Finite dimensional distributions convergence; Timescale separation; Stochastic domination; Branching processes; Large deviations

1. Introduction

We will study in this article the link between two biological models of Darwinian evolution in an asexual population. The first one is a system of interacting particles modeling evolution at

* Tel.: +49 (0)30 20372474; fax: +49 (0)30 2044975.
E-mail address: champagn@wias-berlin.de.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2006.01.004

the *individual* level, referred below as the *microscopic model*, which has been already proposed and studied in [3,4,9,18,13] either as a model of Darwinian evolution or as a model of dispersal in a spatially structured population. This model involves a finite population with non-constant population size, in which each individual's birth and death events are described. Each individual's ability to survive and reproduce is characterized by a finite number of phenotypic traits (e.g. body size, rate of food intake, age at maturity), or simply *traits*. The birth rate of an individual depends on its phenotype, and its death rate depends on the distribution of phenotypes in the population and involves a competition kernel of logistic type. A mutation may occur at each birth event.

The second model describes the evolution at the *population* level as a jump Markov process in the space of phenotypic traits characterizing individuals. It is called the "trait substitution sequence" [21], and referred below as the *TSS model*. In this model, the population is *monomorphic* at each time (i.e. composed of individuals holding the *same* trait value), and the evolution proceeds by a sequence of appearances of new mutant traits, which invade the population and replace, after a short competition, the previous dominant trait. The TSS model belongs to the recent biological theory of evolution called *adaptive dynamics* [15,19,20], and has been introduced by Metz et al. [21] and Dieckmann and Law [8] and mathematically studied in [6]. The theory of adaptive dynamics investigates the effects of the ecological aspects of population dynamics on the evolutionary process, and thus describes the population on the phenotypic level, instead of the genotypic level. The TSS model is one of the fundamental models of this theory. It has revealed a powerful tool for understanding various evolutionary phenomena, such as polymorphism (stable coexistence of different traits, cf. [21]) and evolutionary branching (evolution of a monomorphic population to a polymorphic one that may lead to speciation, [7]) and is the basis of other biological models, such as the "canonical equation of adaptive dynamics" [8,6].

The heuristics leading to the TSS model (cf. [21,8]) are based on the biological assumptions of large population and rare mutations, and on another assumption stating that no two different types of individuals can coexist on a long timescale: the competition eliminates one of them. In spite of this heuristic, this model still lacks a firm mathematical basis.

We propose to prove in this article a convergence result of the microscopic model to the TSS model when the parameters are normalized in a non-standard way, leading to a *timescale separation*. Our limit combines a *large population* asymptotic with a *rare mutations* asymptotic. It will appear that this convergence holds only for finite dimensional distributions, and not for the Skorohod topology, for reasons that are linked to the timescale separation. For these reasons, and because we have to combine two limits simultaneously (large population and rare mutations), this result is different from classical timescale separation results (averaging principle, cf. [14]). The proof requires original methods, based on comparison, convergence and large deviation results on branching processes and logistic Markov birth and death processes. Our convergence result provides a mathematical justification of the TSS model and of the biological heuristic on which it is based, and gives precise conditions on the scalings of the biological parameters in the microscopic model required for the timescale separation to hold.

In Section 2, we describe precisely the microscopic model and the TSS model, and we state our main results. Our proof is based on a careful study of the behavior of the population before the first mutation, and of the phase of competition between the mutant trait and the original trait, taking place just after the first mutation. We will give an outline of the proof and of the methods in Section 3, as well as some notation used throughout the paper. Section 4 gives comparison results and large deviation results on birth and death processes (Sections 4.1 and 4.2), and several results on branching processes (Section 4.3). Based on these properties, the proof of the convergence of the microscopic model to the TSS model is given in Section 5.

2. Models and main results

Let us first describe the microscopic model. In a population, Darwinian evolution acts on a set of phenotypes, or *traits*, characterizing each individual's ability to survive and reproduce. We consider a finite number of quantitative traits in an asexual population (clonal reproduction), and we assume that the trait space $\mathcal{X}$ is a compact subset of $\mathbb{R}^l$ ($l \geq 1$).

The microscopic model involves the three basic mechanisms of Darwinian evolution: *heredity*, which transmits traits to new offsprings, *mutation*, driving a variation in the trait values in the population, and *selection* between these different trait values. The selection process, and thus a proper definition of the selective ability of a trait, or *fitness* (cf. [20]), should (and will) be the consequence of interactions between individuals in the population and of the competition for limited resources or area, modeled as follows.

For any $x, y \in \mathcal{X}$, we introduce the following biological parameters

$b(x) \in \mathbb{R}_+$ is the rate of birth from an individual holding trait x.

$d(x) \in \mathbb{R}_+$ is the rate of "natural" death for an individual holding trait x.

$\alpha(x, y) \in \mathbb{R}_+$ is competition kernel representing the pressure felt by an individual holding trait x from an individual holding trait y.

$\mu(x) \in [0, 1]$ is the probability that a mutation occurs in a birth from an individual with trait x.

$m(x, dh)$ is the law of $h = y - x$, where the mutant trait y is born from an individual with trait x. It is a probability measure on $\mathbb{R}^l$, and since y must belong to the trait space $\mathcal{X}$, the support of $m(x, \cdot)$ is a subset of

$$\mathcal{X} - x = \{y - x : y \in \mathcal{X}\}.$$

$K \in \mathbb{N}$ is a parameter rescaling the competition kernel $\alpha(\cdot, \cdot)$. Biologically, K can be interpreted as scaling the resources or area available, and is related to the biological concept of "carrying capacity". It is also called the "system size" by Metz et al. [21]. As will appear later, this parameter is linked to the size of the population: large K means a large population (provided that the initial condition is proportional to K).

$u_K \in [0, 1]$ is a parameter depending on K rescaling the probability of mutation $\mu(\cdot)$. Small u_K means rare mutations.

Let us also introduce the following notation, used throughout this paper:

$$\bar{n}_x = \frac{b(x) - d(x)}{\alpha(x, x)}, \tag{1}$$

$$\beta(x) = \mu(x) b(x) \bar{n}_x \quad \text{and} \tag{2}$$

$$f(y, x) = b(y) - d(y) - \alpha(y, x) \bar{n}_x. \tag{3}$$

As will appear below, $\bar{n}_x$ can be interpreted as the equilibrium density of a monomorphic population when there is no mutation, $\beta(x)$ as the mutation rate in this population, and $f(y, x)$ as the fitness of a mutant individual with trait y in this population.

We consider, at any time $t \geq 0$, a finite number N_t of individuals, each of them holding a trait value in $\mathcal{X}$. Let us denote by $x_1, \dots, x_{N_t}$ the trait values of these individuals. The state of the population at time $t \geq 0$, rescaled by K, can be described by the finite point measure on $\mathcal{X}$

$$\nu_t^K = \frac{1}{K} \sum_{i=1}^{N_t} \delta_{x_i}, \tag{4}$$

where δ_x is the Dirac measure at x. Let $\mathcal{M}_F$ denote the set of finite non-negative measures on $\mathcal{X}$, and define

$$\mathcal{M}^K = \left\{ \frac{1}{K} \sum_{i=1}^{n} \delta_{x_i} : n \geq 0, x_1, \dots, x_n \in \mathcal{X} \right\}.$$

An individual holding trait x in the population ν_t^K gives birth to another individual with rate $b(x)$ and dies with rate

$$d(x) + \int \alpha(x, y) \nu_t^K(\mathrm{d}y) = d(x) + \frac{1}{K} \sum_{i=1}^{N_t} \alpha(x, x_i).$$

The parameter K scales the strength of competition, thus allowing the coexistence of more individuals in the population.

A newborn holds the same trait value as its progenitor's with probability $1 - u_K \mu(x)$, and with probability $u_K \mu(x)$, the newborn is a mutant whose trait value y is chosen according to $y = x + h$, where h is a random variable with law $m(x, dh)$.

In other words, the process $(\nu_t^K, t \geq 0)$ is an $\mathcal{M}^K$-valued Markov process with infinitesimal generator defined for any bounded measurable functions ϕ from $\mathcal{M}^K$ to $\mathbb{R}$ by

$$\begin{aligned} L^K \phi(\nu) = & \int_{\mathcal{X}} \left(\phi\left(\nu + \frac{\delta_x}{K}\right) - \phi(\nu) \right) (1 - u_K \mu(x)) b(x) K \nu(\mathrm{d}x) \\ & + \int_{\mathcal{X}} \int_{\mathbb{R}^l} \left(\phi\left(\nu + \frac{\delta_{x+h}}{K}\right) - \phi(\nu) \right) u_K \mu(x) b(x) m(x, \mathrm{d}h) K \nu(\mathrm{d}x) \\ & + \int_{\mathcal{X}} \left(\phi\left(\nu - \frac{\delta_x}{K}\right) - \phi(\nu) \right) \left(d(x) + \int_{\mathcal{X}} \alpha(x, y) \nu(\mathrm{d}y) \right) K \nu(\mathrm{d}x). \end{aligned} \tag{5}$$

When the measure ν has the form (4), the integrals with respect to $K\nu(\mathrm{d}x)$ in (5) correspond to sums over all individuals in the population. The first term (linear) describes the births without mutation, the second term (linear) describes the births with mutation, and the third term (non-linear) describes the deaths by age or competition. This logistic density dependence models the competition in the population, and hence drives the selection process.

Let us denote by (A) the following three assumptions:

(A1) b, d and α are measurable functions, and there exist $\bar{b}, \bar{d}, \bar{\alpha} < +\infty$ such that

$$b(\cdot) \leq \bar{b}, \quad d(\cdot) \leq \bar{d} \quad \text{and} \quad \alpha(\cdot, \cdot) \leq \bar{\alpha}.$$

(A2) $m(x, dh)$ is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^l$ with density $m(x, h)$, and there exists a function $\bar{m} : \mathbb{R}^l \to \mathbb{R}_+$ such that $m(x, h) \leq \bar{m}(h)$ for any $x \in \mathcal{X}$ and $h \in \mathbb{R}^l$, and $\int \bar{m}(h) \mathrm{d}h < \infty$.

(A3) $\mu(x) > 0$ and $b(x) - d(x) > 0$ for any $x \in \mathcal{X}$, and there exists $\underline{\alpha} > 0$ such that

$$\underline{\alpha} \leq \alpha(\cdot, \cdot).$$

For fixed K, under (A1) and (A2) and assuming that $\mathbf{E}(\langle \nu_0^K, \mathbf{1}\rangle) < \infty$ (where $\langle \nu, f\rangle$ denotes the integral of the measurable function f with respect to the measure ν), the existence and uniqueness in law of a process with infinitesimal generator L^K have been proved by Fournier and Méléard [13]. When $K \to +\infty$, they also proved, under more restrictive assumptions and assuming the convergence of the initial condition, the convergence on $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F)$ of the process ν^K to a deterministic process solution to a non-linear integro-differential equation. We will only use particular cases of their result, stated in the next section, that can be proved under assumptions (A1) and (A2).

The biological assumption of large population corresponds to the limit $K \to +\infty$, and the assumption of rare mutations to $u_K \to 0$. As mentioned in the introduction, the biological heuristics suggest another assumption: the impossibility of coexistence of two different traits on a long timescale. As will appear in Proposition 3 in the next section, this assumption can be stated mathematically as follows:

(B) Given any $x \in \mathcal{X}$, Lebesgue almost any $y \in \mathcal{X}$ satisfies one of the two following conditions:

$$\text{either} \quad (b(y) - d(y))\alpha(x, x) - (b(x) - d(x))\alpha(y, x) < 0, \tag{6}$$

$$\text{or} \quad \begin{cases} (b(y) - d(y))\alpha(x, x) - (b(x) - d(x))\alpha(y, x) > 0, \\ (b(x) - d(x))\alpha(y, y) - (b(y) - d(y))\alpha(x, y) < 0. \end{cases} \tag{7}$$

Before coming back to this assumption in the next section, let us just observe that condition (6) is equivalent to $f(y, x) < 0$ and condition (7) to $f(y, x) > 0$ and $f(x, y) < 0$.

The TSS model of evolution that we obtain from the microscopic model is a Markov jump process in the trait space $\mathcal{X}$ with infinitesimal generator given, for any bounded measurable function φ from $\mathcal{X}$ to $\mathbb{R}$, by

$$A\varphi(x) = \int_{\mathbb{R}^l} (\varphi(x + h) - \varphi(x))\beta(x)\frac{[f(x + h, x)]_+}{b(x + h)} m(x, h)\mathrm{d}h, \tag{8}$$

where $[a]_+$ denotes the positive part of $a \in \mathbb{R}$, and where $\beta(x)$ and $f(y, x)$ are defined in (2) and (3). The existence and uniqueness in law of a process generated by A hold as soon as $\beta(x)[f(y, x)]_+/b(y)$ is bounded (see e.g. [12]), which is true under assumption (A) ($[f(y, x)]_+/b(y) \leq 1$). The biological interpretation of the function f as a *fitness* function becomes natural in view of this generator: because of the positive part function $[\cdot]_+$ in (8), the TSS process can only jump from a trait x to the traits $x + h$ such that $f(x + h, x) > 0$. Therefore, the function $f(y, x)$ measures the selective ability of trait y in a population made up of individuals with trait x (see [20,21]).

Our main result is:

Theorem 1. *Assume* (A) *and* (B). *Fix a sequence* $(u_K)_{K\in\mathbb{N}}$ *in* $[0, 1]^{\mathbb{N}}$ *such that*

$$\forall V > 0, \quad \exp(-VK) \ll u_K \ll \frac{1}{K \log K} \tag{9}$$

(where $f(K) \ll g(K)$ means that $f(K)/g(K) \to 0$ when $K \to \infty$). Fix also $x \in \mathcal{X}$, $\gamma > 0$ and a sequence of $\mathbb{N}$-valued random variables $(\gamma_K)_{K\in\mathbb{N}}$, such that $(\gamma_K/K)_{K\in\mathbb{N}}$ converges in law to γ and is bounded in $\mathbb{L}^1$. Consider the process $(\nu_t^K, t \geq 0)$ generated by (5) with initial state $(\gamma_K/K)\delta_x$. Then, for any $n \geq 1$, $\varepsilon > 0$ and $0 < t_1 < t_2 < \cdots < t_n < \infty$, and for any measurable subsets $\Gamma_1, \ldots, \Gamma_n$ of $\mathcal{X}$,

$$\lim_{K\to+\infty} \mathbf{P}\Big(\forall i \in \{1,\dots,n\}, \exists x_i \in \Gamma_i : \mathrm{Supp}(\nu^K_{t_i/Ku_K}) = \{x_i\}$$
$$\textit{and } |\langle \nu^K_{t_i/Ku_K}, \mathbf{1}\rangle - \bar{n}_{x_i}| < \varepsilon\Big) = \mathbf{P}(\forall i \in \{1,\dots,n\},\ X_{t_i} \in \Gamma_i) \tag{10}$$

where for any $\nu \in \mathcal{M}_F$, $\mathrm{Supp}(\nu)$ *is the support of ν and $(X_t, t \geq 0)$ is the TSS process generated by (8) with initial state x.*

Remark 1. The timescale $1/Ku_K$ of Theorem 1 is the timescale of the mutation events for the process ν^K (the population size is proportional to K and the individual mutation rate is proportional to u_K). Assumption (9) is the condition leading to the correct timescale separation between the mutation events and the birth and death events. The limit (10) means that, when this timescale separation occurs, the population is monomorphic at any time with high probability, and that the transition periods corresponding to the invasion of a mutant trait in the resident population and the ensuing competition are infinitesimal on this mutation timescale. Observe also that this convergence result holds only for monomorphic initial conditions. We will make some comments on more general initial conditions in the next section.

Corollary 1. *Assume additionally in Theorem 1 that $(\gamma_K/K)_{K\in\mathbb{N}}$ is bounded in $\mathbb{L}^p$ for some $p > 1$. Then the process $(\nu^K_{t/Ku_K}, t \geq 0)$ converges when $K \to +\infty$, in the sense of the finite dimensional distributions for the topology on $\mathcal{M}_F$ induced by the functions $\nu \mapsto \langle \nu, f\rangle$ with f bounded and measurable on $\mathcal{X}$, to the process $(Y_t, t \geq 0)$ defined by*

$$Y_t = \begin{cases} \gamma\delta_x & \text{if } t = 0 \\ \bar{n}_{X_t}\delta_{X_t} & \text{if } t > 0. \end{cases}$$

This corollary follows from the following long time moment estimates, which are a consequence of the stochastic domination results of Section 4.1 and will be proved therein.

Lemma 1. *Assume* (A) *and that* $\sup_{K\geq 1} \mathbf{E}(\langle \nu^K_0, 1\rangle^p) < +\infty$ *for some $p \geq 1$, then*

$$\sup_{K\geq 1}\sup_{t\geq 0} \mathbf{E}(\langle \nu^K_t, \mathbf{1}\rangle^p) < +\infty,$$

and therefore, if $p > 1$, the family of random variables $\{\langle \nu^K_t, \mathbf{1}\rangle\}_{\{K\geq 1, t\geq 0\}}$ is uniformly integrable.

Proof of Corollary 1. Let Γ be a measurable subset of $\mathcal{X}$. Let us prove that

$$\lim_{K\to+\infty} \mathbf{E}(\langle \nu^K_{t/Ku_K}, \mathbf{1}_\Gamma\rangle) = \mathbf{E}(\bar{n}_{X_t}\mathbf{1}_{\{X_t\in\Gamma\}}). \tag{11}$$

Fix $\varepsilon > 0$, and observe that $\bar{n}_x \in [0, \bar{b}/\underline{\alpha}]$. Write $[0, \bar{b}/\underline{\alpha}] \subset \cup_{i=1}^q I_i$, where q is the first integer greater than $\bar{b}/\varepsilon\underline{\alpha}$, and $I_i = [(i-1)\varepsilon, i\varepsilon[$. Define $\Gamma_i = \{x \in \mathcal{X} : \bar{n}_x \in I_i\}$ for $1 \leq i \leq q$, and apply (10) to the sets $\Gamma \cap \Gamma_1, \dots, \Gamma \cap \Gamma_q$ with $n = 1$, $t_1 = t$ and the constant ε above. Then, by Lemma 1, there exists a constant $C > 0$ such that

$$\begin{aligned}\limsup_{K\to+\infty} \mathbf{E}(\langle \nu^K_{t/Ku_K}, \mathbf{1}_\Gamma\rangle) &\leq \limsup_{K\to+\infty} \mathbf{E}(\langle \nu^K_{t/Ku_K}, \mathbf{1}_\Gamma\rangle \mathbf{1}_{\{\langle \nu^K_{t/Ku_K}, \mathbf{1}\rangle \leq C\}}) + \varepsilon \\ &\leq \sum_{i=1}^q \limsup_{K\to+\infty} \mathbf{E}(\langle \nu^K_{t/Ku_K}, \mathbf{1}_{\Gamma\cap\Gamma_i}\rangle \mathbf{1}_{\{\langle \nu^K_{t/Ku_K}, \mathbf{1}\rangle \leq C\}}) + \varepsilon\end{aligned}$$

$$\leq \sum_{i=1}^{q} (i+1)\varepsilon \mathbf{P}(X_t \in \Gamma \cap \Gamma_i) + \varepsilon$$

$$\leq \sum_{i=1}^{q} \left(\mathbf{E}(\bar{n}_{X_t} \mathbf{1}_{\{X_t \in \Gamma \cap \Gamma_i\}}) + 2\varepsilon \mathbf{P}(X_t \in \Gamma_i) \right) + \varepsilon$$

$$\leq \mathbf{E}(\bar{n}_{X_t} \mathbf{1}_{\{X_t \in \Gamma\}}) + 3\varepsilon.$$

A similar estimate for the *lim inf* ends the proof of (11), which implies the convergence of one-dimensional laws for the required topology.

The same method gives easily the required limit when we consider a finite number of times $t_1, \dots, t_n$. □

As suggested by the fact that the limit process Y is not continuous at 0^+, it is not possible to obtain the convergence in law for the Skorohod topology on $\mathbb{D}([0,T], \mathcal{M}_F)$. More generally, we can prove:

Proposition 1. *For any $s < t$, the convergence of $\nu^K_{\cdot/Ku_K}$ to Y in Corollary 1 does not hold for the Skorohod topology on $\mathbb{D}([s,t], \mathcal{M}_F)$, for any topology on $\mathcal{M}_F$ such that the total mass function $\nu \mapsto \langle \nu, \mathbf{1} \rangle$ is continuous.*

Proof of Proposition 1. Assume the converse. Then, for some $s < t$, the total mass $N^K_t = \langle \nu^K_{t/Ku_K}, \mathbf{1} \rangle$ converges for the Skorohod topology on $\mathbb{D}([s,t], \mathbb{R}_+)$ to the total mass of the process Y. In particular, by Ascoli's theorem for càdlàg processes (cf. [2]), for any $\varepsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\limsup_{K \to +\infty} \mathbf{P}(\omega'(N^K, \delta) > \eta) \leq \varepsilon,$$

where the modulus of continuity ω' is defined by

$$\omega'(\varphi, \delta) := \inf \left\{ \max_{i=0,\dots,r-1} \omega(\varphi, [t_i, t_{i+1})) \right\}$$

where the infimum is taken over all $r \in \mathbb{N}$ and all the finite partitions $s = t_0 < t_1 < \dots < t_r = t$ of $[s,t]$ such that $t_{i+1} - t_i > \delta$ for any $i \in \{0, \dots, r-1\}$, and where $\omega(\varphi, I) := \sup_{x,y \in I} |\varphi(x) - \varphi(y)|$ for any interval I.

Now, for any function $\varphi \in \mathbb{D}([s,t], \mathbb{R})$, $\omega(\varphi, \delta) \leq 2\omega'(\varphi, \delta) + \sup_{x \in [s,t]} |\varphi(x) - \varphi(x-)|$ (cf. [2]), where $\omega(\varphi, \delta) := \sup_{x,y \in [s,t], |x-y| \leq \delta} |\varphi(x) - \varphi(y)|$, and for any $K \geq 1$, $\sup_{x \in [s,t]} |N^K_x - N^K_{x-}| = 1/K$. Therefore, for any $\varepsilon > 0$ and $\eta > 0$, there exists $\delta > 0$ such that

$$\limsup_{K \geq 1} \mathbf{P}(\omega(N^K, \delta) > \eta) \leq \varepsilon.$$

This implies that the sequence $(N^K)_K$ is actually C-tight (cf. [2]) and that its limit is necessarily continuous, which is not true for $\langle Y_t, \mathbf{1} \rangle$. □

3. Notation and outline of the proof of Theorem 1

We start with some definitions needed to explain the idea of the proof of Theorem 1 and the precise meaning of assumption (B).

Definition 1. (a) For any $K \geq 1$, $b, d, \alpha \geq 0$ and for any $\mathbb{N}/K$-valued random variable z, we will denote by $\mathbf{P}^K(b, d, \alpha, z)$ the law of the $\mathbb{N}/K$-valued Markov birth and death process with initial state z and with transition rates

$$ib \quad \text{from } i/K \text{ to } (i+1)/K,$$
$$i(d + \alpha i/K) \quad \text{from } i/K \text{ to } (i-1)/K.$$

(b) For any $K \geq 1$, $b_k, d_k, \alpha_{kl} \geq 0$ with $k, l \in \{1, 2\}$, and for any $\mathbb{N}/K$-valued random variables z_1 and z_2, we will denote by

$$\mathbf{Q}^K(b_1, b_2, d_1, d_2, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, z_1, z_2)$$

the law of the $(\mathbb{N}/K)^2$-valued Markov birth and death process with initial state (z_1, z_2) and with transition rates

$$ib_1 \quad \text{from } (i/K, j/K) \text{ to } ((i+1)/K, j/K),$$
$$jb_2 \quad \text{from } (i/K, j/K) \text{ to } (i/K, (j+1)/K),$$
$$i(d_1 + \alpha_{11} i/K + \alpha_{12} j/K) \quad \text{from } (i/K, j/K) \text{ to } ((i-1)/K, j/K),$$
$$j(d_2 + \alpha_{21} i/K + \alpha_{22} j/K) \quad \text{from } (i/K, j/K) \text{ to } (i/K, (j-1)/K).$$

These two Markov processes have absorbing states at 0 and (0, 0), respectively. Observe also that, when $\alpha = 0$, the Markov process of point (a) is a continuous-time binary branching process divided by K.

Fix x and y in $\mathcal{X}$. The proof of the following two results can be found in Chap. 11 of [12].

Proposition 2. (a) *Assume $\mu \equiv 0$ and $\nu_0^K = N_x^K(0)\delta_x$. Then, for any $t \geq 0$, $\nu_t^K = N_x^K(t)\delta_x$, where N_x^K has the law $\mathbf{P}^K(b(x), d(x), \alpha(x, x), N_x^K(0))$. Assume $N_x^K(0) \to n_x(0)$ in probability when $K \to +\infty$. Then, the sequence (N_x^K) converges in probability on $[0, T]$ for the uniform norm to the deterministic solution n_x to*

$$\dot{n}_x = (b(x) - d(x) - \alpha(x, x)n_x)n_x \quad \textit{with initial condition } n_x(0). \tag{12}$$

(b) *Assume $\mu \equiv 0$ and $\nu_0^K = N_x^K(0)\delta_x + N_y^K(0)\delta_y$. Then, for any $t \geq 0$, $\nu_t^K = N_x^K(t)\delta_x + N_y^K(t)\delta_y$, where (N_x^K, N_y^K) has the law*

$$\mathbf{Q}^K(b(x), b(y), d(x), d(y), \alpha(x, x), \alpha(x, y), \alpha(y, x), \alpha(y, y), N_x^K(0), N_y^K(0)).$$

Assume $N_x^K(0) \to n_x(0)$ and $N_y^K(0) \to n_y(0)$ in probability when $K \to +\infty$. Then, (N_x^K, N_y^K) converges in probability when $K \to +\infty$ on $[0, T]$ for the uniform norm to the deterministic solution (n_x, n_y) to

$$\begin{cases} \dot{n}_x = (b(x) - d(x) - \alpha(x, x)n_x - \alpha(x, y)n_y)n_x \\ \dot{n}_y = (b(y) - d(y) - \alpha(y, x)n_x - \alpha(y, y)n_y)n_y \\ \qquad\qquad \textit{with initial condition } (n_x(0), n_y(0)). \end{cases} \tag{13}$$

Note that, under assumption (A3), the logistic equation (12) has two steady states, 0, unstable, and $\bar{n}_x$, defined in (1), stable. The system (13) has at least three steady states, (0, 0), unstable, $(\bar{n}_x, 0)$ and $(0, \bar{n}_y)$.

The assumption (B) of Section 2 is the mathematical formulation of the impossibility of coexistence of two different traits, in the sense that, starting in the neighborhood of the equilibrium $(\bar{n}_x, 0)$ of system (13), either its solution converges to this equilibrium or to the equilibrium $(0, \bar{n}_y)$. More precisely, the following proposition follows from an elementary analysis of system (13) (cf. e.g. [16, pp. 25–27]):

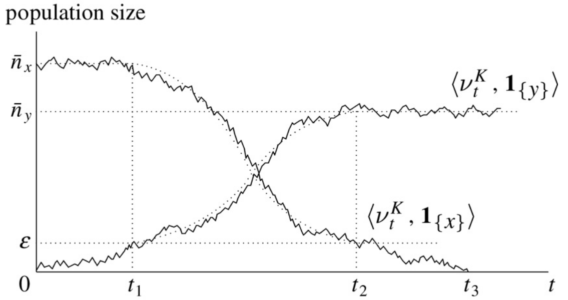

Fig. 1. The three steps of the invasion of a mutant trait y in a monomorphic population with trait x.

Proposition 3. *If x and y satisfy (6), then $(\bar{n}_x, 0)$ is a stable steady state of (13). If x and y satisfy (7), then $(\bar{n}_x, 0)$ is an unstable steady state, $(0, \bar{n}_y)$ is stable, and any solution to (13) with initial state in $(\mathbb{R}_+^*)^2$ converges to $(0, \bar{n}_y)$ when $t \to +\infty$.*

Let us now give the main ideas of the proof of Theorem 1. It is based on two main ingredients: first, when $\mu \equiv 0$ and ν_0^K is monomorphic with trait x, we have seen in Proposition 2(a) the convergence of ν^K to $n(t)\delta_x$, where $n(t)$ is solution to (12). Any solution to this equation with positive initial condition converges for large time to $\bar{n}_x$. The large deviations estimates for this convergence will allow us to show that the time during which the stochastic process stays in a neighborhood of its limit (the problem of exit from a domain, [14]) is of the order of $\exp(KV)$ with $V > 0$. Now, when u_K is small, the process ν^K with a monomorphic initial condition with trait x is close to the same process with $\mu \equiv 0$, as long as no mutation occurs. Therefore, the left inequality in (9) will allow us to prove that, with high probability, the first mutation event (occurring on the timescale t/Ku_K) occurs before the total density drifts away from $\bar{n}_x$.

The second ingredient of our proof is the study of the invasion of a mutant trait y that has just appeared in a monomorphic population with trait x. This invasion can be divided in three steps (Fig. 1), in a similar way to what is done classically by population geneticists dealing with selective sweeps [17,11]:

- Firstly, as long as the mutant population size $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ (initially equal to $1/K$) is smaller than a fixed small $\varepsilon > 0$ (before t_1 in Fig. 1), the resident dynamics is very close to what it was before the mutation, so $\langle \nu_t^K, \mathbf{1}_{\{x\}} \rangle$ stays close to $\bar{n}_x$. Then, the death rate of a mutant individual is close to the constant $d(y) + \alpha(y, x)\bar{n}_x$. Since its birth rate is constant, equal to $b(y)$, we can approximate the mutant dynamics by a binary branching process. Therefore, the probability that $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ reaches ε is approximately equal to the probability that this branching process reaches εK, which converges when $K \to +\infty$ to its probability of non-extinction $[f(y, x)]_+/b(y)$.
- Secondly, once $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ has reached ε, by Proposition 2(b), for large K, ν^K is close to the solution to (13) with initial state $(\bar{n}_x, \varepsilon)$ (represented with dotted lines in Fig. 1) with high probability. By Proposition 3, this solution will be shown to reach the ε-neighborhood of $(0, \bar{n}_y)$ in finite time (t_2 in Fig. 1).
- Finally, once $\langle \nu_t^K, \mathbf{1}_{\{y\}} \rangle$ is close to $\bar{n}_y$ and $\langle \nu_t^K, \mathbf{1}_{\{x\}} \rangle$ is small, $K\langle \nu_t^K, \mathbf{1}_{\{x\}} \rangle$ can be approximated, in a similar way to in the first step, by a binary branching process, which is sub-critical and hence becomes extinct a.s. in finite time (t_3 in Fig. 1).

We will see in Sections 4.2 and 4.3 that the time needed to complete the first and third steps is proportional to $\log K$, whereas the time needed for the second step is bounded. Therefore, since the time between two mutations is of the order of $1/Ku_K$, the right inequality in (9) will allow us to prove that, with high probability, the three steps above are completed before a new mutation occurs.

Remark 2. As observed by Metz et al. [21], the biological heuristics leading to the TSS model extend to the case of polymorphic initial condition, where the population is composed of a finite number of distinct traits (see also [5]). Our mathematical method can also be extended easily to n-morphic initial conditions, except for one difficulty: one has to replace assumption (B) by another assumption stating that, for any n, any solution to the n-morphic logistic systems generalizing (13) converges to an equilibrium (as in Proposition 3), and that the equilibria of these systems are non-degenerate, in the sense that the branching processes in the first and third steps above are not critical, or, equivalently, that a first-order linear analysis of these equilibria allows one to determine their stability. Then, one could construct a polymorphic TSS model in which the number of coexisting traits is not fixed. However, the asymptotic analysis of n-dimensional logistic systems is non-trivial and may exhibit cycles or chaos, except when $n = 1$ or 2, and analytical assumptions ensuring the condition above are difficult to find.

Section 4 will provide the large deviations and branching process results needed to make formal the previous heuristics. We will also prove several comparison results for $\langle \nu_t^K, \mathbf{1}\rangle$ and the birth and death processes of Definition 1. In Section 5, the proof of Theorem 1 is achieved by computing, for any t, the limit law of ν^K_{t/Ku_K} according to the random number of mutations having occurred between 0 and t/Ku_K.

Notation.

- $\lceil a\rceil$ denotes the first integer greater or equal to a, and $\lfloor a\rfloor$ denotes the integer part of a.
- For any $K \geq 1$ and $\nu \in \mathcal{M}^K$, we will denote by $\mathbf{P}_\nu^K$ the law of the process ν^K generated by (5) with initial state ν, and by $\mathbf{E}_\nu^K$ the expectation with respect to $\mathbf{P}_\nu^K$.
- The convergence in probability of finite dimensional random variables will be denoted by $\xrightarrow{\mathcal{P}}$.
- We will denote by $\mathcal{L}(Z)$ the law of the stochastic process $(Z_t, t \geq 0)$.
- We will denote by $\preceq$ the following stochastic domination relation: if $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are the laws of $\mathbb{R}$-valued processes, we will write $\mathbf{Q}_1 \preceq \mathbf{Q}_2$ if we can construct on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ two processes X^1 and X^2 such that $\mathcal{L}(X^i) = \mathbf{Q}_i$ $(i = 1, 2)$ and $\forall t \geq 0$, $\forall \omega \in \Omega$, $X_t^1(\omega) \leq X_t^2(\omega)$.
- Finally, if X^1 and X^2 are two random processes and T is a random time constructed on the same probability space as X^1, we will write $X_t^1 \preceq X_t^2$ for $t \leq T$ (resp. $X_t^2 \preceq X_t^1$ for $t \leq T$) if we can construct a process $\hat{X}^2$ on the same probability space as X^1, such that $\mathcal{L}(\hat{X}^2) = \mathcal{L}(X^2)$ and $\forall t \leq T$, $\forall \omega \in \Omega$, $X_t^1(\omega) \leq \hat{X}_t^2(\omega)$ (resp. $\hat{X}_t^2(\omega) \leq X_t^1(\omega)$).

4. Birth and death processes

We will collect in this section various results on the birth and death processes that appeared in Definition 1.

4.1. Comparison results

The following theorem gives various stochastic domination results.

Theorem 2. (a) *Assume* (A). *For any $K \geq 1$ and any $\mathbb{L}^1$ initial condition ν_0^K of the process ν^K,*

$$\mathcal{L}(\langle \nu^K, \mathbf{1}\rangle) \preceq \mathbf{P}^K(2\bar{b}, 0, \underline{\alpha}, \langle \nu_0^K, \mathbf{1}\rangle).$$

(b) *With the same assumptions as in* (a), *let A_t^K denote the number of mutations occurring in ν^K between times 0 and t, and let $a, a_1, a_2 \geq 0$. Then, for $t \leq \inf\{s \geq 0 : \langle \nu_s^K, \mathbf{1}\rangle \geq a\}$,*

$$A_t^K \preceq B_t^K,$$

where B^K is a Poisson process with parameters $K u_K a \bar{b}$.

If moreover $\nu_0^K = \langle \nu_0^K, \mathbf{1}\rangle \delta_x$, define $\tau_1 = \inf\{t \geq 0 : A_t^K = 1\}$ (the first mutation time). Then, for $t \leq \tau_1 \wedge \inf\{s \geq 0 : \langle \nu_s^K, \mathbf{1}\rangle \notin [a_1, a_2]\}$,

$$B_t^K \preceq A_t^K \preceq C_t^K, \tag{14}$$

where B^K and C^K are Poisson processes with respective parameter $K u_K a_1 \mu(x) b(x)$ and $K u_K a_2 \mu(x) b(x)$.

(c) *Fix $K \geq 1$ and take b, d, α, z as in Definition 1*(a). *Then, for any $\varepsilon_1, \varepsilon_2, \varepsilon_3 \geq 0$ and any $\mathbb{N}/K$-valued random variable ε_4,*

$$\mathbf{P}^K(b, d + \varepsilon_2, \alpha + \varepsilon_3, z) \preceq \mathbf{P}^K(b + \varepsilon_1, d, \alpha, z + \varepsilon_4).$$

(d) *Let (Z^1, Z^2) be a stochastic process with law*

$$\mathbf{Q}^K(b_1, b_2, d_1, d_2, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, z_1, z_2)$$

where the parameters are as in Definition 1(b). *Fix $a > 0$ and define $T = \inf\{t \geq 0, Z^2 \geq a\}$. Then, for $t \leq T$,*

$$M_t^1 \preceq Z_t^1 \preceq M_t^2, \quad \textit{where}$$
$$\mathcal{L}(M^1) = \mathbf{P}^K(b_1, d_1 + a\alpha_{12}, \alpha_{11}, z_1) \quad \textit{and}$$
$$\mathcal{L}(M^2) = \mathbf{P}^K(b_1, d_1, \alpha_{11}, z_1).$$

(e) *Take (Z^1, Z^2) as above, fix $0 \leq a_1 < a_2$ and $a > 0$, and define $T = \inf\{t \geq 0, Z^1 \notin [a_1, a_2]$ or $Z^2 \geq a\}$. Then, for $t \leq T$,*

$$M_t^1 \preceq Z_t^2 \preceq M_t^2, \quad \textit{where}$$
$$\mathcal{L}(M^1) = \mathbf{P}^K(b_2, d_2 + a_2\alpha_{21} + a\alpha_{22}, 0, z_2) \quad \textit{and}$$
$$\mathcal{L}(M^2) = \mathbf{P}^K(b_2, d_2 + a_1\alpha_{21}, 0, z_2).$$

Remark 3. Point (a) explains why it is necessary to combine simultaneously the limits $K \to +\infty$ and $u_K \to 0$ in order to obtain the TSS process in Theorem 1. The limit $K \to +\infty$ taken alone leads to a deterministic dynamics [13], so making the rare mutations limit afterwards cannot lead to a stochastic process. Conversely, taking the limit of rare mutations without making the population larger would lead to an immediate extinction of the population in the mutations timescale, because the stochastic domination of Theorem 2(a) is independent of u_K and $\mu(\cdot)$, and because a process Z with law $\mathbf{P}^K(2\bar{b}, 0, \underline{\alpha}, \gamma_K/K)$ gets a.s. extinct in finite time.

Before proving Theorem 2, let us deduce from Point (a) the Lemma 1 stated in Section 2.

Proof of Lemma 1. By Theorem 2(a), it suffices to prove that

$$\sup_{K \geq 1} \sup_{t \geq 0} \mathbf{E}((Z_t^K)^p) < +\infty,$$

where $\mathcal{L}(Z^K) = \mathbf{P}^K(2\bar{b}, 0, \underline{\alpha}, z_0^K)$ when $\sup_{K \geq 1} \mathbf{E}((z_0^K)^p) < +\infty$.

Let us define $v_t^k = \mathbf{P}(Z_t^K = k/K)$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}((Z_t^K)^p) = \sum_{k\geq 1}\left(\frac{k}{K}\right)^p \frac{\mathrm{d}v_t^k}{\mathrm{d}t}$$
$$= \frac{1}{K^p}\sum_{k\geq 1} k^p \left[2\bar{b}(k-1)v_t^{k-1} + \underline{\alpha}\frac{(k+1)^2}{K}v_t^{k+1} - k\left(2\bar{b} + \underline{\alpha}\frac{k}{K}\right)v_t^k\right]$$
$$= \frac{1}{K^p}\sum_{k\geq 1}\left[2\bar{b}\left(\left(1+\frac{1}{k}\right)^p - 1\right) + \underline{\alpha}\frac{k}{K}\left(\left(1-\frac{1}{k}\right)^p - 1\right)\right]k^{p+1}v_t^k.$$

Now, for $k/K > 4\bar{b}/\underline{\alpha}$, the quantity inside the square brackets in the last expression can be upper bounded by $-2\bar{b}[3 - 2(1-1/k)^p - (1+1/k)^p]$, which is equivalent to $-2\bar{b}p/k$ when $k \to +\infty$. Therefore, there exists a constant k_0 that can be assumed bigger than $4\bar{b}/\underline{\alpha}$ such that, for any $k \geq k_0$, $-2\bar{b}[3 - 2(1-1/k)^p - (1+1/k)^p] \leq -\bar{b}p/k$. Then, using the fact that $(1+x)^p - 1 \leq x(2^p - 1)$ for any $x \in [0, 1]$, we can write

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{E}((Z_t^K)^p) \leq \sum_{k=1}^{Kk_0-1} 2\bar{b}(2^p-1)\left(\frac{k}{K}\right)^p v_t^k - \sum_{k\geq Kk_0}\bar{b}p\left(\frac{k}{K}\right)^p v_t^k$$
$$\leq 2\bar{b}(2^p-1)k_0^p + \bar{b}pk_0^p - \bar{b}p\mathbf{E}((Z_t^K)^p).$$

Writing $C = (2(2^p-1)+p)k_0^p/p$, this differential inequality is solved giving

$$\mathbf{E}((Z_t^K)^p) \leq C + [\mathbf{E}((z_0^K)^p) - C]\mathrm{e}^{-\bar{b}pt},$$

which gives the required uniform bound. □

Proof of Theorem 2. The proof is essentially intuitive if one computes upper and lower bounds of the birth and death rates for each process considered in the statement of the theorem. We will simply give the explicit construction of the process ν^K, and the proof of (14) as an example. We leave the remaining comparison results to the reader.

We will use the construction of the process ν^K given by Fournier and Méléard [13]: let $(\Omega, \mathcal{F}, \mathbf{P})$ be a sufficiently large probability space, and consider on this space the following five independent random objects:

(i) an $\mathcal{M}^K$-valued random variable ν_0^K (the initial distribution),
(ii) a Poisson point measure $N_1(ds, di, dv)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measure $q_1(ds, di, dv) = \bar{b}ds\sum_{k\geq 1}\delta_k(di)dv$ (the birth without mutation Poisson point measure),
(iii) a Poisson point measure $N_2(ds, di, dh, dv)$ on $[0,\infty[\times\mathbb{N}\times\mathbb{R}^l\times[0,1]$ with intensity measure $q_2(ds, di, dh, dv) = \bar{b}ds\sum_{k\geq 1}\delta_k(di)\bar{m}(h)dhdv$ (the birth with mutation Poisson point measure),
(iv) a Poisson point measure $N_3(ds, di, dv)$ on $[0,\infty[\times\mathbb{N}\times[0,1]$ with intensity measure $q_3(ds, di, dv) = \bar{d}ds\sum_{k\geq 1}\delta_k(di)dv$ (the natural death Poisson point measure),
(v) a Poisson point measure $N_4(ds, di, dj, dv)$ on $[0,\infty[\times\mathbb{N}\times\mathbb{N}\times[0,1]$ with intensity measure $q_4(ds, di, dj, dv) = (\bar{\alpha}/K)ds\sum_{k\geq 1}\delta_k(di)\sum_{m\geq 1}\delta_m(dj)dv$ (the competition death Poisson point measure).

We will also need the following function, solving the purely notational problem of associating a number with each individual in the population: for any $K \geq 1$, let $H = (H^1, \ldots, H^k, \ldots)$ be the map from $\mathcal{M}^K$ into $(\mathbb{R}^l)^{\mathbb{N}}$ defined by

$$H\left(\frac{1}{K}\sum_{i=1}^{n}\delta_{x_i}\right) = (x_{\sigma(1)}, \dots, x_{\sigma(n)}, 0, \dots, 0, \dots),$$

where $x_{\sigma(1)} \preccurlyeq \dots \preccurlyeq x_{\sigma(n)}$ for the lexicographic order $\preccurlyeq$ on $\mathbb{R}^l$. For convenience, we have omitted in our notation the dependence of H and H^i on K.

Then a process ν^K with generator L^K and initial state ν_0^K can be constructed as follows: for any $t \geq 0$,

$$\begin{aligned}
\nu_t^K &= \nu_0^K + \int_0^t \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{i \leq K\langle \nu_{s-}^K, \mathbf{1}\rangle\}} \frac{\delta_{H^i(\nu_{s-}^K)}}{K} \mathbf{1}_{\left\{v \leq \frac{[1-u_K\mu(H^i(\nu_{s-}^K))]b(H^i(\nu_{s-}^K))}{\bar{b}}\right\}} N_1(\mathrm{d}s, \mathrm{d}i, \mathrm{d}v) \\
&\quad + \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq K\langle \nu_{s-}^K, \mathbf{1}\rangle\}} \frac{\delta_{H^i(\nu_{s-}^K)+h}}{K} \\
&\quad \times \mathbf{1}_{\left\{v \leq \frac{u_K\mu(H^i(\nu_{s-}^K))b(H^i(\nu_{s-}^K))}{\bar{b}} \frac{m(H^i(\nu_{s-}^K),h)}{\bar{m}(h)}\right\}} N_2(\mathrm{d}s, \mathrm{d}i, \mathrm{d}h, \mathrm{d}v) \\
&\quad - \int_0^t \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{i \leq K\langle \nu_{s-}^K, \mathbf{1}\rangle\}} \frac{\delta_{H^i(\nu_{s-}^K)}}{K} \mathbf{1}_{\left\{v \leq \frac{d(H^i(\nu_{s-}^K))}{\bar{d}}\right\}} N_3(\mathrm{d}s, \mathrm{d}i, \mathrm{d}v) \\
&\quad - \int_0^t \int_{\mathbb{N}} \int_{\mathbb{N}} \int_0^1 \mathbf{1}_{\{i \leq K\langle \nu_{s-}^K, \mathbf{1}\rangle\}} \mathbf{1}_{\{j \leq K\langle \nu_{s-}^K, \mathbf{1}\rangle\}} \frac{\delta_{H^i(\nu_{s-}^K)}}{K} \\
&\quad \times \mathbf{1}_{\left\{v \leq \frac{\alpha(H^i(\nu_{s-}^K), H^j(\nu_{s-}^K))}{\bar{\alpha}}\right\}} N_4(\mathrm{d}s, \mathrm{d}i, \mathrm{d}j, \mathrm{d}v). \qquad (15)
\end{aligned}$$

Although this formula is quite complicated, the principle is simple: for each type of event, the corresponding Poisson point process jumps faster than ν^K has to. We decide whether a jump of the process ν^K occurs by comparing v to a quantity related to the rates of the various events. The indicator functions involving i and j ensure that the ith and jth individuals are alive in the population (because $K\langle \nu_t^K, \mathbf{1}\rangle$ is the number of individuals in the population at time t).

Under (A1), (A2) and the assumption that $\mathbf{E}(\langle \nu_0^K, \mathbf{1}\rangle) < \infty$, [13] proves the existence and uniqueness of the solution to (15), and that this solution is a Markov process with infinitesimal generator (5).

Now, let us come to the proof of (14). The process A^K can be written as

$$A_t^K := \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq K\langle \nu_{s-}^K, \mathbf{1}\rangle\}} \mathbf{1}_{\left\{v \leq \frac{u_K\mu(H^i(\nu_{s-}^K))b(H^i(\nu_{s-}^K))}{\bar{b}} \frac{m(H^i(\nu_{s-}^K),h)}{\bar{m}(h)}\right\}} N_2(\mathrm{d}s, \mathrm{d}i, \mathrm{d}h, \mathrm{d}v).$$

In the case where $\nu_0^K = \langle \nu_0^K, \mathbf{1}\rangle\delta_x$, as long as $t < \tau_1$, $\nu_t^K = \langle \nu_t^K, \mathbf{1}\rangle\delta_x$. Therefore, for $t \leq \tau_1 \wedge \inf\{s \geq 0 : \langle \nu_s^K, \mathbf{1}\rangle \notin [a_1, a_2]\}$,

$$\begin{aligned}
&\int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq K a_1\}} \mathbf{1}_{\left\{v \leq \frac{u_K\mu(x)b(x)}{\bar{b}} \frac{m(x,h)}{\bar{m}(h)}\right\}} N_2(\mathrm{d}s, \mathrm{d}i, \mathrm{d}h, \mathrm{d}v) \leq A_t^K \\
&\quad \leq \int_0^t \int_{\mathbb{N}} \int_{\mathbb{R}^l} \int_0^1 \mathbf{1}_{\{i \leq K a_2\}} \mathbf{1}_{\left\{v \leq \frac{u_K\mu(x)b(x)}{\bar{b}} \frac{m(x,h)}{\bar{m}(h)}\right\}} N_2(\mathrm{d}s, \mathrm{d}i, \mathrm{d}h, \mathrm{d}v). \qquad (16)
\end{aligned}$$

Since the intensity measure of N_2 is

$$q_2(ds, di, dh, dv) = \bar{b}\, ds \sum_{k\geq 1} \delta_k(di)\bar{m}(h)dhdv,$$

the left-hand side and the right-hand side of (16) are Poisson processes with parameters $Ku_K a_1\mu(x)b(x)$ and $Ku_K a_2\mu(x)b(x)$, respectively. □

4.2. Problem of exit from a domain

Let us give some results on $\mathbf{P}^K(b, d, \alpha, z)$ when $\alpha > 0$. Points (a) and (b) of the following theorem strengthen Proposition 2, and point (c) studies the problem of exit from a domain.

Theorem 3. (a) *Let $\alpha, T > 0$ and $b, d \geq 0$, let C be a compact subset of $\mathbb{R}_+^*$, and write $\mathbf{P}_z^K = \mathbf{P}^K(b, d, \alpha, z)$ for $z \in \mathbb{N}/K$. Let ϕ_z denote the solution to*

$$\dot{\phi} = (b - d - \alpha\phi)\phi \tag{17}$$

with initial condition $\phi_z(0) = z$. Then

$$r := \inf_{z\in C} \inf_{0\leq t\leq T} |\phi_z(t)| > 0 \quad \textit{and} \quad R := \sup_{z\in C} \sup_{0\leq t\leq T} |\phi_z(t)| < +\infty.$$

Moreover, for any $\delta < r$,

$$\lim_{K\to+\infty} \sup_{z\in C} \mathbf{P}_z^K \left(\sup_{0\leq t\leq T} |w_t - \phi_z(t)| \geq \delta \right) = 0, \tag{18}$$

where w_t is the canonical process on $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$.

(b) *Let $T, \alpha_{ij} > 0$ and $b_i, d_i \geq 0$ $(i, j \in \{1, 2\})$, let C be a compact subset of $(\mathbb{R}_+^*)^2$, and write $\mathbf{Q}_{z_1,z_2}^K = \mathbf{Q}^K(b_1, b_2, d_1, d_2, \alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, z_1, z_2)$ for z_1 and z_2 in $\mathbb{N}/K$. Let $\phi_{z_1,z_2} = (\phi_{z_1,z_2}^1, \phi_{z_1,z_2}^2)$ denote the solution to*

$$\begin{cases} \dot{\phi}^1 = (b_1 - d_1 - \alpha_{11}\phi^1 - \alpha_{12}\phi^2)\phi^1 \\ \dot{\phi}^2 = (b_2 - d_2 - \alpha_{21}\phi^1 - \alpha_{22}\phi^2)\phi^2 \end{cases}$$

with initial conditions $\phi_{z_1,z_2}^1(0) = z_1$ and $\phi_{z_1,z_2}^2(0) = z_2$. Then

$$r := \inf_{z\in C} \inf_{0\leq t\leq T} \|\phi_{z_1,z_2}(t)\| > 0 \quad \textit{and} \quad \sup_{z\in C} \sup_{0\leq t\leq T} \|\phi_{z_1,z_2}(t)\| < +\infty. \tag{19}$$

Moreover, for any $\delta < r$,

$$\lim_{K\to+\infty} \sup_{z\in C} \mathbf{Q}_{z_1,z_2}^K \left(\sup_{0\leq t\leq T} \|\hat{w}_t - \phi_{z_1,z_2}(t)\| \geq \delta \right) = 0,$$

where $\hat{w}_t = (\hat{w}_t^1, \hat{w}_t^2)$ is the canonical process on $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^2)$.

(c) *Let $b, \alpha > 0$ and $0 \leq d < b$. Observe that $(b - d)/\alpha$ is the unique stable steady state of (17). Fix $0 < \eta_1 < (b - d)/\alpha$ and $\eta_2 > 0$, and define on $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$*

$$T^K = \inf\left\{ t \geq 0 : w_t \notin \left[\frac{b-d}{\alpha} - \eta_1, \frac{b-d}{\alpha} + \eta_2 \right] \right\}.$$

Then, there exists $V > 0$ such that, for any compact subset C of $](b-d)/\alpha - \eta_1, (b-d)/\alpha + \eta_2[$,

$$\lim_{K\to+\infty} \sup_{z\in C} \mathbf{P}_z^K (T^K < \mathrm{e}^{KV}) = 0. \tag{20}$$

Proof of (a) and (b). Observe that any solution to (17) with positive initial condition is bounded ($\dot{\phi} < 0$ as soon as $\phi > (b-d)/\alpha$). This implies that $R < \infty$. Moreover, a solution to (17) can be written as

$$\phi(t) = \phi(0)\exp\left(\int_0^t (b - d - \alpha\phi(s))\mathrm{d}s\right) \geq \phi(0)\exp((b - d - \alpha R)t),$$

which implies that $r > 0$.

Eq. (18) is a consequence of large deviations estimates for the sequence of laws $(\mathbf{P}_z^K)_{K\geq 1}$. As can be seen in Theorem 10.2.6 in Chap. 10 of [10], a large deviations principle on $[0, T]$ with a good rate function I_T holds for $\mathbb{Z}/K$-valued Markov jump processes with transition rates

$$\begin{aligned} &Kp(i/K) \quad \text{from } i/K \text{ to } (i+1)/K,\\ &Kq(i/K) \quad \text{from } i/K \text{ to } (i-1)/K, \end{aligned}$$

where p and q are functions defined on $\mathbb{R}$ and with positive values, bounded, Lipschitz and uniformly bounded away from 0. The rate function I_T can be written as

$$I_T(\phi) = \begin{cases} \int_0^T L(\phi(t), \dot{\phi}(t))\,\mathrm{d}t & \text{if } \phi \text{ is absol. cont. on } [0, T] \\ +\infty & \text{otherwise} \end{cases} \tag{21}$$

for some function $L : \mathbb{R}^2 \to \mathbb{R}_+$ such that $L(y, z) = 0$ if and only if $z = p(y) - q(y)$. Therefore, $I_T(\phi) = 0$ if and only if ϕ is absolutely continuous and

$$\dot{\phi} = p(\phi) - q(\phi). \tag{22}$$

Moreover, this large deviation is uniform with respect to the initial condition. This means that, if $\mathbf{R}_z^K$ denotes the law of this process with initial condition z, for any compact set $C \subset \mathbb{R}$, for any closed set F and any open set G of $\mathbb{D}([0, T], \mathbb{R})$,

$$\liminf_{K\to+\infty} \frac{1}{K} \log \inf_{z\in C} \mathbf{R}_z^K(G) \geq -\sup_{z\in C} \inf_{\psi\in G, \psi(0)=z} I_T(\psi) \quad \text{and} \tag{23}$$

$$\limsup_{K\to+\infty} \frac{1}{K} \log \sup_{z\in C} \mathbf{R}_z^K(F) \leq -\inf_{\psi\in F, \psi(0)\in C} I_T(\psi). \tag{24}$$

Our birth and death process does not satisfy these assumptions. However, if we define

$$\begin{aligned} &p(z) = b\chi(z) \quad \text{and} \quad q(z) = d\chi(z) + \alpha\chi(z)^2, \quad \text{where} \\ &\chi(z) = z \text{ if } z \in [r-\delta, R+\delta]; \quad r-\delta \text{ if } z < r-\delta; \quad R+\delta \text{ if } z > R+\delta, \end{aligned}$$

then $\mathbf{R}_z^K = \mathbf{P}_z^K$ on the time interval $[0, \tau]$, where $\tau = \inf\{t \geq 0, w_t \notin [r-\delta, R+\delta]\}$, and p and q satisfy the assumptions above. Therefore, by (24),

$$\begin{aligned} &\limsup_{K\to+\infty} \frac{1}{K} \log \sup_{z\in C} \mathbf{P}_z^K\left(\sup_{0\leq t\leq T} |w_t - \phi_z(t)| \geq \delta\right) \leq -\inf_{\psi\in F^\delta} I_T(\psi), \quad \text{where} \\ &F^\delta := \left\{\psi \in \mathbb{D}([0, T], \mathbb{R}) : \psi(0) \in C \text{ and } \exists t \in [0, T], |\psi(t) - \phi_{\psi(0)}(t)| \geq \delta\right\}. \end{aligned}$$

By the continuity of the flow of (22) (which is a classical consequence of the fact that $z \mapsto p(z) - q(z)$ is Lipschitz and of Gronwall's lemma), the set F^δ is closed. Since I_T is a good rate function, the infimum of I_T over this set is attained at some function belonging to F^δ, which cannot be a solution to (22), and thus is non-zero. This ends the proof of (18).

The proof of (b) can be constructed in a very similar way. □

Proof of (c). Define the function χ on $\mathbb{R}$ by $\chi(z) = z$ if $z \in [(b-d)/\alpha - \eta_1, (b-d)/\alpha + \eta_2]$, $\chi(z) = (b-d)/\alpha - \eta_1$ for $z < (b-d)/\alpha - \eta_1$ and $\chi(z) = (b-d)/\alpha + \eta_2$ for $z > (b-d)/\alpha - \eta_2$. As in the proof of (a), we can construct from the functions $p(z) = b\chi(z)$ and $q(z) = d\chi(z) + \alpha\chi(z)^2$ a family of laws $(\mathbf{R}_z^K)$ such that $\mathbf{R}_z^K = \mathbf{P}_z^K$ on the time interval $[0, T^K]$, and such that (23) and (24) hold for the good rate function I_T defined in (21).

Observe that all solutions to (22) are monotonous and converge to $(b-d)/\alpha$ when $t \to +\infty$. Therefore, the following estimates for the time of exit from an attracting domain are classical [14, Chap. 5, Section 4]: there exists $\bar{V} \geq 0$ such that, for any $\delta > 0$,

$$\lim_{K\to+\infty} \inf_{z\in C} \mathbf{R}_z^K \left(e^{K(\bar{V}-\delta)} < T^K < e^{K(\bar{V}+\delta)}\right) = 1,$$

which implies (20) if we can prove that $\bar{V} > 0$.

The constant $\bar{V}$ is obtained as follows (see [14, pp. 108–109]): for any $y, z \in \mathbb{R}$, define

$$V(y, z) := \inf_{t>0, \varphi(0)=y, \varphi(t)=z} I_t(\varphi).$$

Then

$$\bar{V} := V\left(\frac{b-d}{\alpha}, \frac{b-d}{\alpha} - \eta_1\right) \wedge V\left(\frac{b-d}{\alpha}, \frac{b-d}{\alpha} + \eta_2\right).$$

Now, Theorem 5.4.3 of [14] states that, for any $y, z \in \mathbb{R}$, the infimum defining $V(y, z)$ is attained at some function ϕ linking y to z, in the sense that, either there exists an absolutely continuous function ϕ defined on $[0, T]$ for some $T > 0$ such that $\phi(0) = y$, $\phi(T) = z$ and $V(y, z) = I_T(\phi) = \int_0^T L(\phi(t), \dot{\phi}(t))\,dt$, or there exists an absolutely continuous function ϕ defined on $]-\infty, T]$ for some $T > -\infty$ such that $\lim_{t\to-\infty} \phi(t) = y$, $\phi(T) = z$ and $V(y, z) = \int_{-\infty}^T L(\phi(t), \dot{\phi}(t))\,dt$.

Since any solution to (22) is decreasing as long as it stays in $[(b-d)/\alpha, +\infty[$, a function ϕ defined on $[0, T]$ or $]-\infty, T]$ linking $(b-d)/\alpha$ to $(b-d)/\alpha + \eta_2$ cannot be a solution to (22), and thus $V((b-d)/\alpha, (b-d)/\alpha + \eta_2) > 0$. Similarly, $V((b-d)/\alpha, (b-d)/\alpha - \eta_1) > 0$, and so $\bar{V} > 0$, which concludes the proof of Theorem 3. □

4.3. Some results on branching processes

When $\alpha = 0$, $\mathbf{P}^K(b, d, 0, z)$ is the law of a binary branching process divided by K. Let us give some results on these processes.

Theorem 4. *Let $b, d > 0$. As in Theorem 3, define, for any $K \geq 1$ and any $z \in \mathbb{N}/K$, $\mathbf{P}_z^K = \mathbf{P}^K(b, d, 0, z)$. Define also, for any $\rho \in \mathbb{R}$, on $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$, the stopping time*

$$T_\rho = \inf\{t \geq 0 : w_t = \rho\}.$$

Finally, let $(t_K)_{K\geq1}$ be a sequence of positive numbers such that $\log K \ll t_K$.

(a) *If $b < d$ (sub-critical case), for any $\varepsilon > 0$,*

$$\lim_{K\to+\infty} \mathbf{P}_{1/K}^K(T_0 \leq t_K \wedge T_{\lceil\varepsilon K\rceil/K}) = 1, \quad \textit{and} \tag{25}$$

$$\lim_{K\to+\infty} \mathbf{P}_{\lfloor\varepsilon K\rfloor/K}^K(T_0 \leq t_K) = 1. \tag{26}$$

Moreover, for any $K \geq 1$, $k \geq 1$ *and* $n \geq 1$,

$$\mathbf{P}^K_{n/K}(T_{kn/K} \leq T_0) \leq \frac{1}{k}. \tag{27}$$

(b) *If* $b > d$ *(super-critical case), for any* $\varepsilon > 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{1/K}(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil/K}) = \frac{d}{b} \quad \textit{and} \tag{28}$$

$$\lim_{K\to+\infty} \mathbf{P}^K_{1/K}(T_{\lceil \varepsilon K\rceil/K} \leq t_K) = 1 - \frac{d}{b}. \tag{29}$$

Proof. Let us denote by $\mathbf{Q}_n$ the law of the binary branching process with initial state $n \in \mathbb{N}$, with individual birth rate b and individual death rate d. Then (25)–(29) can be rewritten respectively as

$$\lim_{K\to+\infty} \mathbf{Q}_1(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil}) = 1, \tag{30}$$

$$\lim_{K\to+\infty} \mathbf{Q}_{\lfloor \varepsilon K\rfloor}(T_0 \leq t_K) = 1, \tag{31}$$

$$\mathbf{Q}_n(T_{kn} \leq T_0) \leq \frac{1}{k}, \tag{32}$$

$$\lim_{K\to+\infty} \mathbf{Q}_1(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil}) = \frac{d}{b} \quad \text{and} \tag{33}$$

$$\lim_{K\to+\infty} \mathbf{Q}_1(T_{\lceil \varepsilon K\rceil} \leq t_K) = 1 - \frac{d}{b}. \tag{34}$$

The limit (31) follows easily from the distribution of the extinction time for binary branching processes when $b \neq d$ (cf. [1, p. 109]): for any $t \geq 0$ and $n \in \mathbb{N}$,

$$\mathbf{Q}_n(T_0 \leq t) = \left(\frac{d\left(1 - \mathrm{e}^{-(b-d)t}\right)}{b - d\mathrm{e}^{-(b-d)t}}\right)^n. \tag{35}$$

Since $t_K \to +\infty$, $\mathbf{Q}_1(T_0 \leq t_K \wedge T_{\lceil \varepsilon K\rceil}) \to \mathbf{Q}_1(T_0 < \infty)$, which gives (30) and (33) (the probability of extinction of a binary branching process can be recovered easily from (35)).

The inequality (32) follows from the fact that, if $(Z_t, t \geq 0)$ is a process with law $\mathbf{Q}_n$, $(Z_t \exp(-(b-d)t), t \geq 0)$ is a martingale (cf. [1, p. 111]). Then, Doob's stopping theorem applied to the stopping time $T_0 \wedge T_{kn}$ yields

$$\mathbf{E}_n(kn\mathrm{e}^{(d-b)T_{kn}}\mathbf{1}_{\{T_{kn}<T_0\}}) = n,$$

where $\mathbf{E}_n$ is the expectation with respect to $\mathbf{Q}_n$. Therefore, when $b < d$, $kn\mathbf{Q}_n(T_{kn} < T_0) \leq n$, and the proof of (32) is completed.

The limit (34) follows from the fact that, if $(Z_t, t \geq 0)$ is a branching process with law $\mathbf{Q}_1$, the martingale $(Z_t \exp(-(b-d)t), t \geq 0)$ converges a.s. when $t \to +\infty$ to a random variable W, where $W = 0$ on the event $\{T_0 < \infty\}$ and $W > 0$ on the event $\{T_0 = \infty\}$ (cf. [1, p. 112]). Hence, on the event $\{T_0 = \infty\}$, when $b > d$,

$$(\log Z_t)/t \to b - d > 0.$$

Therefore, since $\log K \ll t_K$, for any $\varepsilon > 0$, $\mathbf{Q}_1(T_0 = \infty, T_{\lceil \varepsilon K\rceil} \geq t_K) \to 0$ when $K \to +\infty$. Then, (34) follows from the fact that $\mathbf{Q}_1(T_0 = \infty) = 1 - d/b$. $\square$

5. Proof of Theorem 1

Let us assume, without loss of generality, that ν^K is constructed by (15) on a sufficiently large probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

We introduce the following sequences of stopping times: for all $n \geq 1$, let τ_n be the first mutation time after time τ_{n-1}, with $\tau_0 = 0$ (i.e. τ_n is the nth mutation time), and for any $n \geq 0$, let θ_n be the first time after τ_n when the population becomes monomorphic. Observe that $\theta_0 = 0$ if the initial population is monomorphic. For any $n \geq 1$, define the random variable U_n as the new trait value appearing at the mutation time τ_n, and, when $\theta_n < \infty$, define V_n by $\mathrm{Supp}(\nu^K_{\theta_n}) = \{V_n\}$. When $\theta_n = +\infty$, define $V_n = +\infty$.

Our proof of Theorem 1 is based on the following two lemmas. The first lemma proves that there is no accumulation of mutations on the timescale of Theorem 1, and studies the asymptotic behavior of τ_1 starting from a monomorphic population, when $K \to +\infty$.

Lemma 2. (a) *Assume that the initial condition of ν^K satisfies $\sup_K \mathbf{E}(\langle \nu^K_0, \mathbf{1}\rangle) < +\infty$. Then, for any $\eta > 0$, there exists $\varepsilon > 0$ such that, for any $t > 0$,*

$$\limsup_{K\to+\infty} \mathbf{P}^K_{\nu^K_0}\left(\exists n \geq 0 : \frac{t}{Ku_K} \leq \tau_n \leq \frac{t+\varepsilon}{Ku_K}\right) < \eta. \tag{36}$$

Let $x \in \mathcal{X}$ and let $(z_K)_{K\geq 1}$ be a sequence of integers such that $z_K/K \to z > 0$.

(b) *For any $\varepsilon > 0$,*

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \log K, \sup_{t\in[\log K,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| > \varepsilon\right) = 0. \tag{37}$$

Since $\log K \ll 1/Ku_K$, *by* (a) *with* $t = 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}(\tau_1 < \log K) = 0.$$

In particular, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$, $\nu^K_{\log K} \xrightarrow{\mathcal{P}} \bar{n}_x\delta_x$ and $\nu^K_{\tau_1 -} \xrightarrow{\mathcal{P}} \bar{n}_x\delta_x$.

If, moreover, $z = \bar{n}_x$, then, for any $\varepsilon > 0$,

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\sup_{t\in[0,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| > \varepsilon\right) = 0. \tag{38}$$

(c) *For any $t > 0$,*

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}\right) = \exp(-\beta(x)t),$$

where $\beta(\cdot)$ has been defined in (2).

The second lemma studies the asymptotic behavior of θ_0 and V_0 starting from a dimorphic population, when $K \to +\infty$.

Lemma 3. *Fix $x, y \in \mathcal{X}$ satisfying* (6) *or* (7), *and let $(z_K)_{K\geq 1}$ be a sequence of integers such that $z_K/K \to \bar{n}_x$. Then,*

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x + \frac{1}{K}\delta_y}(V_0 = y) = \frac{[f(y,x)]_+}{b(y)}, \tag{39}$$

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}(V_0 = x) = 1 - \frac{[f(y,x)]_+}{b(y)}, \tag{40}$$

$$\forall \eta > 0, \qquad \lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}\left(\theta_0 > \frac{\eta}{K u_K} \wedge \tau_1\right) = 0 \quad \textit{and} \tag{41}$$

$$\forall \varepsilon > 0, \qquad \lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}\left(|\langle \nu^K_{\theta_0}, \mathbf{1}\rangle - \bar{n}_{V_0}| < \varepsilon\right) = 1, \tag{42}$$

where $f(y,x)$ has been defined in (3).

Observe that (41) implies in particular that

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}(\theta_0 < \tau_1) = 1.$$

The proofs of these lemmas are postponed at the end of this section.

Proof of Theorem 1. Observe that the generator A, defined in (8), of the TSS process $(X_t, t \geq 0)$ of Theorem 1 can be written as

$$A\varphi(x) = \int_{\mathbb{R}^l} (\varphi(x+h) - \varphi(x))\beta(x)\kappa(x, \mathrm{d}h), \tag{43}$$

where the probability measure $\kappa(x, dh)$ is defined by

$$\kappa(x, dh) = \left(1 - \int_{\mathbb{R}^l} \frac{[f(x+v,x)]_+}{b(x+v)} m(x,v)\mathrm{d}v\right)\delta_0(dh) + \frac{[f(x+h,x)]_+}{b(x+h)} m(x,h)dh. \tag{44}$$

This means that the TSS model X with initial state x can be constructed as follows: let $(Z(k), k = 0, 1, 2, \ldots)$ be a Markov chain in $\mathcal{X}$ with initial state x and with transition kernel $\kappa(x, dh)$, and let $(N(t), t \geq 0)$ be an independent standard Poisson process. Then, the process $(X_t, t \geq 0)$ defined by

$$X_t := Z\left(N\left(\int_0^t \beta(X_s)\mathrm{d}s\right)\right)$$

is a Markov process with infinitesimal generator (43) (cf. [12, Chap. 6]). Let $\mathbf{P}_x$ denote its law, let $(T_n)_{n\geq 1}$ denote the sequence of jump times of the Poisson process N and define $(S_n)_{n\geq 1}$ by $T_n = \int_0^{S_n} \beta(X_s)\mathrm{d}s$. By (A1) and (A3), $\beta(\cdot) > 0$, and so S_n is finite for any $n \geq 1$. Observe that any jump of the process X occurs at some time S_n, but that not all S_n have to be effective jump times for X, because of the Dirac mass at 0 appearing in (44).

Fix $t > 0$, $x \in \mathcal{X}$ and a measurable subset Γ of $\mathcal{X}$. Under $\mathbf{P}_x$, S_1 and X_{S_1} are independent, S_1 is an exponential random variable with parameter $\beta(x)$, and $X_{S_1} - x$ has law $\kappa(x, \cdot)$. Therefore, for any $n \geq 1$, the strong Markov property applied to X at time S_1 yields

$$\begin{aligned}
&\mathbf{P}_x(S_n \leq t < S_{n+1},\ X_t \in \Gamma) \\
&\quad = \int_0^t \beta(x)\mathrm{e}^{-\beta(x)s} \int_{\mathbb{R}^l} \mathbf{P}_{x+h}(S_{n-1} \leq t-s < S_n,\ X_{t-s} \in \Gamma)\kappa(x, \mathrm{d}h)\mathrm{d}s.
\end{aligned} \tag{45}$$

Moreover,

$$\mathbf{P}_x(0 \leq t < S_1,\ X_t \in \Gamma) = \mathbf{1}_{\{x\in\Gamma\}}\mathrm{e}^{-\beta(x)t}. \tag{46}$$

The idea of our proof of Theorem 1 is to show that the same relations hold when we replace S_n by τ_n and X_t by the support of ν^K_{t/Ku_K} (when it is a singleton) and when $K \to +\infty$.

More precisely, fix $x \in \mathcal{X}$, $t > 0$ and a measurable subset Γ of $\mathcal{X}$, and observe that

$$\left\{\exists y \in \Gamma : \mathrm{Supp}(\nu^K_{t/Ku_K}) = \{y\},\ |\langle \nu^K_{t/Ku_K}, \mathbf{1}\rangle - \bar{n}_y| < \varepsilon\right\} = \bigcup_{n\geq 0} A^K_n(t, \Gamma, \varepsilon), \tag{47}$$

where

$$A^K_n(t,\Gamma,\varepsilon) := \left\{\theta_n \leq \frac{t}{Ku_K} < \tau_{n+1},\ V_n \in \Gamma,\ |\langle \nu^K_{t/Ku_K}, \mathbf{1}\rangle - \bar{n}_{V_n}| < \varepsilon\right\}.$$

Let us define, for any $z \in \mathbb{N}$ and $n \geq 0$,

$$p^K_n(t,x,\Gamma,\varepsilon,z) := \mathbf{P}^K_{\frac{z}{K}\delta_x}\left(\theta_n \leq \frac{t}{Ku_K} < \tau_{n+1}, V_n \in \Gamma, \sup_{s\in[\theta_n,\tau_{n+1}]} |\langle \nu^K_s, \mathbf{1}\rangle - \bar{n}_{V_n}| < \varepsilon\right)$$

and

$$\begin{aligned} q^K_0(t,x,\Gamma,\varepsilon,z) &:= \mathbf{P}^K_{\frac{z}{K}\delta_x}\left(\frac{t}{Ku_K} < \tau_1, V_0 \in \Gamma, \sup_{s\in[\log K,\tau_1]} |\langle \nu^K_s, \mathbf{1}\rangle - \bar{n}_{V_0}| < \varepsilon\right) \\ &= \mathbf{1}_{\{x\in\Gamma\}}\mathbf{P}^K_{\frac{z}{K}\delta_x}\left(\frac{t}{Ku_K} < \tau_1, \sup_{s\in[\log K,\tau_1]} |\langle \nu^K_s, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right). \end{aligned}$$

Let us also extend these definitions to $\varepsilon = \infty$ by suppressing the condition involving the supremum of $|\langle \nu^K, \mathbf{1}\rangle - \bar{n}_{V_n}|$.

Then

Lemma 4. (a) *For any $x \in \mathcal{X}$, $n \geq 1$, $t > 0$, $\varepsilon \in]0,\infty]$ and for any sequence of integers (z_K) such that $z_K/K \to z > 0$, $p_n(t,x,\Gamma) := \lim_{K\to+\infty} p^K_n(t,x,\Gamma,\varepsilon,z_K)$ exists, and is independent of (z_K), z and ε.*

Similarly, $p_0(t,x,\Gamma) := \lim_{K\to+\infty} q^K_0(t,x,\Gamma,\varepsilon,z_K)$ exists, and is independent of (z_K), z and ε, and, if $z = \bar{n}_x$, $\lim_{K\to+\infty} p^K_0(t,x,\Gamma,\varepsilon,z_K)$ exists and is also equal to $p_0(t,x,\Gamma)$.

Finally, if we assume that (z_K) is a sequence of $\mathbb{N}$-valued random variables such that z_K/K converge in probability to a deterministic $z > 0$, then the limits above hold in probability (with the same restriction that z has to be equal to $\bar{n}_x$ for p^K_0).

(b) *The functions $p_n(t,x,\Gamma)$ are continuous with respect to t and measurable with respect to x, and satisfy*

$$p_0(t,x,\Gamma) = \mathbf{1}_{\{x\in\Gamma\}}\mathrm{e}^{-\beta(x)t} \quad \textit{and} \quad \forall n \geq 0,$$

$$p_{n+1}(t,x,\Gamma) = \int_0^t \beta(x)\mathrm{e}^{-\beta(x)s}\int_{\mathbb{R}^l} p_n(t-s,x+h,\Gamma)\kappa(x,\mathrm{d}h)\mathrm{d}s. \tag{48}$$

Let us postpone the proof of this lemma to after the proof of Theorem 1.

Observe that, because of (45) and (46), Lemma 4(b) implies that $\mathbf{P}_x(S_n \leq t < S_{n+1},\ X_t \in \Gamma) = p_n(t,x,\Gamma)$.

Now, let $\tilde{\mathbf{P}}^K_\nu$ denote the law of the process ν^K with random initial state ν. Since ν^K is Markov, $\tilde{\mathbf{P}}^K_{\gamma_K/K\delta_x} = \mathbf{E}[\mathbf{P}^K_{\gamma_K(\omega)/K\delta_x}]$. By (47),

$$\begin{aligned}\tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}&\left(\exists y \in \Gamma : \mathrm{Supp}(\nu^K_{t/Ku_K}) = \{y\}, |\langle \nu^K_{t/Ku_K}, \mathbf{1}\rangle - \bar{n}_y| < \varepsilon\right)\\ &= \sum_{n\geq 0} \tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}(A^K_n(t,\Gamma,\varepsilon)),\end{aligned}$$

where (γ_K) is the sequence of $\mathbb{N}$-valued random variables of Theorem 1.

For any $K \geq 1$ and $n \geq 1$,

$$\begin{aligned}&p^K_n(t,x,\Gamma,\varepsilon,\gamma_K) \leq \mathbf{P}^K_{\frac{\gamma_K}{K}\delta_x}(A^K_n(t,\Gamma,\varepsilon)) \leq p^K_n(t,x,\Gamma,\infty,\gamma_K), \quad \text{and}\\ &q^K_0(t,x,\Gamma,\varepsilon,\gamma_K) \leq \mathbf{P}^K_{\frac{\gamma_K}{K}\delta_x}(A^K_n(t,\Gamma,\varepsilon)) \leq p^K_n(t,x,\Gamma,\infty,\gamma_K),\end{aligned}$$

so, by Lemma 4(a), for any $n \geq 0$, $\mathbf{P}^K_{(\gamma_K/K)\delta_x}(A^K_n(t,\Gamma,\varepsilon)) \xrightarrow{\mathcal{P}} p_n(t,x,\Gamma)$, and therefore,

$$\lim_{K\to+\infty} \tilde{\mathbf{P}}^K_{(\gamma_K/K)\delta_x}(A^K_n(t,\Gamma,\varepsilon)) = p_n(t,x,\Gamma). \tag{49}$$

Now, by (47), for any $K \geq 1$,

$$\sum_{n=0}^{+\infty}\left[\tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}(A^K_n(t,\Gamma,\varepsilon)) + \tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}(A^K_n(t,\Gamma^c,\varepsilon))\right] \leq 1, \tag{50}$$

where Γ^c denotes the complement of Γ. Moreover, $\sum_{n=0}^{+\infty}[p_n(t,x,\Gamma) + p_n(t,x,\Gamma^c)] = 1$. Therefore, for any $\eta > 0$, there exists n_0 such that

$$\sum_{n=0}^{n_0}[p_n(t,x,\Gamma) + p_n(t,x,\Gamma^c)] \geq 1 - \eta.$$

Then, one can easily deduce from (49) and (50) that

$$\limsup_{K\to+\infty} \sum_{n\geq n_0}\left[\tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}(A^K_n(t,\Gamma,\varepsilon)) + \tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}(A^K_n(t,\Gamma^c,\varepsilon))\right] \leq \eta,$$

from which it follows, by (47), that

$$\begin{aligned}\lim_{K\to+\infty} &\tilde{\mathbf{P}}^K_{\frac{\gamma_K}{K}\delta_x}\left(\exists y \in \Gamma : \mathrm{Supp}(\nu^K_{t/Ku_K}) = \{y\},\ y \in \Gamma,\ |\langle \nu^K_{t/Ku_K}, \mathbf{1}\rangle - \bar{n}_y| < \varepsilon\right)\\ &= \sum_{n\geq 0} p_n(t,x,\Gamma) = \mathbf{P}_x(X_t \in \Gamma),\end{aligned}$$

which is (10) in the case of a single time t.

In order to complete the proof of Theorem 1, we have to generalize this limit to any sequence of times $0 < t_1 < \cdots < t_n$.

We will specify the method only in the case of two times $0 < t_1 < t_2$. It can be easily generalized to a sequence of n times. We introduce for any integers $0 \leq n_1 \leq n_2$ the probabilities

$$p^K_{n_1,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z)$$
$$:= \mathbf{P}^K_{\frac{z}{K}\delta_x}\left(\theta_{n_1} \leq \frac{t_1}{Ku_K} < \tau_{n_1+1},\ V_{n_1}\in\Gamma_1,\ \sup_{s\in[\theta_{n_1},\tau_{n_1+1}]}|\langle\nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_{n_1}}|<\varepsilon,\right.$$
$$\left.\theta_{n_2} \leq \frac{t_2}{Ku_K} < \tau_{n_2+1},\ V_{n_2}\in\Gamma_2 \text{ and } \sup_{s\in[\theta_{n_2},\tau_{n_2+1}]}|\langle\nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_{n_2}}|<\varepsilon\right),$$

and

$$q^K_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z)$$
$$:= \mathbf{1}_{\{x\in\Gamma_1\}}\mathbf{P}^K_{\frac{z}{K}\delta_x}\left(\frac{t_1}{Ku_K} < \tau_1,\ \sup_{s\in[\log K,\tau_1]}|\langle\nu^K_s,\mathbf{1}\rangle - \bar{n}_x|<\varepsilon,\right.$$
$$\left.\theta_{n_2} \leq \frac{t_2}{Ku_K} < \tau_{n_2+1},\ V_{n_2}\in\Gamma_2 \text{ and } \sup_{s\in[\theta_{n_2},\tau_{n_2+1}]}|\langle\nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_{n_2}}|<\varepsilon\right).$$

Then, we can use a calculation very similar to the proof of Lemma 4 to prove that, as $K\to+\infty$, $p^K_{n_1,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z_K)$ converges to a limit $p_{n_1,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2)$ independent of $\varepsilon\in]0,\infty]$, z_K and the limit $z>0$ of z_K/K (with the restriction that z has to be equal to $\bar{n}_x$ if $n_1=0$), and that $\lim q^K_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2,\varepsilon,z) = p_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2)$, where

$$\begin{cases} p_{0,n_2}(t_1,t_2,x,\Gamma_1,\Gamma_2) = \mathbf{1}_{\{x\in\Gamma_1\}}\mathrm{e}^{-\beta(x)t_1}p_{n_2}(t_2-t_1,x,\Gamma_2); \\ p_{n_1+1,n_2+1}(t_1,t_2,x,\Gamma_1,\Gamma_2) \\ \quad = \int_0^{t_1}\beta(x)\mathrm{e}^{-\beta(x)s}\int_{\mathbb{R}^l}p_{n_1,n_2}(t_1-s,t_2-s,x+h,\Gamma_1,\Gamma_2)\kappa(x,\mathrm{d}h)\mathrm{d}s. \end{cases}$$

As above, we obtain Eq. (10) for $n=2$ by observing that the same relation holds for the TSS process X.

This completes the proof of Theorem 1. □

Proof of Lemma 4. First, let us prove that the convergence of $p^K_n(t,x,\Gamma,\varepsilon,z_K)$ when $z_K\in\mathbb{N}$ in Lemma 4(a) implies the convergence in probability of these quantities when z_K are random variables: if (z_K) is a sequence of random variables such that $z_K/K\overset{\mathcal{P}}{\to}z$, by Skorohod's Theorem, we can construct on an auxiliary probability space $\hat{\Omega}$ a sequence of random variables $(\hat{z}_K)$ such that $\mathcal{L}(\hat{z}_K)=\mathcal{L}(z_K)$ and $\hat{z}_K(\hat{\omega})/K\to z$ for any $\hat{\omega}\in\hat{\Omega}$. Then, $\lim p^K_n(t,x,\Gamma,\varepsilon,\hat{z}_K(\hat{\omega})) = p_n(t,x,\Gamma)$ for any $\hat{\omega}\in\hat{\Omega}$, which implies that $p^K_n(t,x,\Gamma,\varepsilon,z_K)\overset{\mathcal{P}}{\to}p_n(t,x,\Gamma)$. The same method applies to $q^K_0(t,x,\Gamma,\varepsilon,z_k)$.

We will prove Lemma 4(a) and (b) by induction over $n\geq 0$.

First, when $t>0$, it follows from the fact that $t/Ku_K>\log K$ for sufficiently large K, and from Lemma 2(b) and (c), that

$$\lim_{K\to+\infty}q^K_0(t,x,\Gamma,\varepsilon,z_K) = \mathbf{1}_{\{x\in\Gamma\}}\mathrm{e}^{-\beta(x)t},$$

and that, if $z=\bar{n}_x$,

$$\lim_{K\to+\infty}p^K_0(t,x,\Gamma,\varepsilon,z_K) = \mathbf{1}_{\{x\in\Gamma\}}\mathrm{e}^{-\beta(x)t}.$$

Then, fix $n\geq 0$ and assume that Lemma 4(a) holds for n. We intend to prove the convergence of $p^K_{n+1}(t,x,\Gamma,\varepsilon,z_K)$ to $p_{n+1}(t,x,\Gamma)$ satisfying (48) by applying the strong Markov property at time τ_1, in a similar way to how we obtained (45). However, the convergence of $p_n^K(t,x,\Gamma,\varepsilon,z_K)$ to $p_n(t,x,\Gamma)$ only holds for *non-random* t. Therefore, we will divide the time interval $[0,t]$ into a finite number of small intervals and use the Markov property at time τ_1 when τ_1 is in each of these intervals. Moreover, we will also use the Markov property at time θ_1 and we will use the fact that U_1 is independent of τ_1 and $\nu^K_{\tau_1-}$ and that U_1-x is a random variable with law $m(x,h)dh$.

Following this program, we can bound $p_{n+1}^K(t,x,\Gamma,\varepsilon,z_K)$ from above as follows: fix $\eta>0$; using Lemma 2(a) in the first inequality, for sufficiently large $k\geq 0$ and $K\geq 1$,

$$\begin{aligned}
p_{n+1}^K(t,x,\Gamma,\varepsilon,z_K) &\leq \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\theta_{n+1}\leq\frac{t}{Ku_K},\tau_{n+2}>\frac{t+2/2^k}{Ku_K},V_{n+1}\in\Gamma\right)+\eta\\
&\leq \sum_{i=0}^{\lceil t2^k\rceil-1}\mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\frac{i}{2^kKu_K}\leq\tau_1\leq\frac{i+1}{2^kKu_K},\theta_{n+1}\leq\frac{t}{Ku_K},\right.\\
&\qquad\left.\tau_{n+2}>\frac{t+2/2^k}{Ku_K}\text{ and }V_{n+1}\in\Gamma\right)+\eta\\
&\leq \sum_{i=0}^{\lceil t2^k\rceil-1}\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\leq\tau_1\leq\frac{i+1}{2^kKu_K}\right\}}\mathbf{P}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{U_1}}\left(\theta_n\leq\frac{t-i/2^k}{Ku_K},\right.\right.\\
&\qquad\left.\left.\tau_{n+1}>\frac{t-(i-1)/2^k}{Ku_K}\text{ and }V_n\in\Gamma\right)\right]+\eta\\
&\leq \sum_{i=0}^{\lceil t2^k\rceil-1}\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\leq\tau_1\leq\frac{i+1}{2^kKu_K}\right\}}\int_{\mathbb{R}^l}\mathbf{E}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}\left(\mathbf{1}_{\left\{\theta_0\geq\frac{1}{2^kKu_K}\wedge\tau_1\right\}}\right.\right.\\
&\qquad\left.\left.+\mathbf{1}_{\left\{\theta_0<\frac{1}{2^kKu_K}\wedge\tau_1\right\}}\mathbf{P}^K_{\nu^K_{\theta_0}}\left(\theta_n\leq\frac{t-i/2^k}{Ku_K}<\tau_{n+1},\ V_n\in\Gamma\right)\right)m(x,h)dh\right]+\eta\\
&\leq \sum_{i=0}^{\lceil t2^k\rceil-1}\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^kKu_K}\leq\tau_1\leq\frac{i+1}{2^kKu_K}\right\}}\int_{\mathbb{R}^l}\mathbf{E}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}\left(\mathbf{1}_{\left\{\theta_0\geq\frac{1}{2^kKu_K}\wedge\tau_1\right\}}\right.\right.\\
&\qquad\left.\left.+\mathbf{1}_{\left\{\theta_0<\frac{1}{2^kKu_K}\wedge\tau_1\right\}}p_n^K(t-i/2^k,V_0,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)\right)m(x,h)\mathrm{d}h\right]+\eta. \qquad (51)
\end{aligned}$$

Now, since $\nu^K_{\tau_1-}=\langle\nu^K_{\tau_1-},\mathbf{1}\rangle\delta_x$, under $\mathbf{P}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}$, on the event $\{\theta_0<\tau_1\}$,

$$\begin{aligned}
p_n^K(t-i/2^k,V_0,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle) &= \mathbf{1}_{\{V_0=x\}}p_n^K(t-i/2^k,x,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)\\
&\quad+\mathbf{1}_{\{V_0=x+h\}}p_n^K(t-i/2^k,x+h,\Gamma,\infty,K\langle\nu^K_{\theta_0},\mathbf{1}\rangle). \qquad (52)
\end{aligned}$$

By Lemma 2(b), $\nu^K_{\tau_1-}\xrightarrow{\mathcal{P}}\bar{n}_x\delta_x$ under $\mathbf{P}_{\frac{z_K}{K}\delta_x}$, so we can use Skorohod's Theorem to construct random variables $\hat{N}_K$ on an auxiliary probability space $\hat{\Omega}$ with the same law of $\langle\nu^K_{\tau_1-},\mathbf{1}\rangle$ and converging to $\bar{n}_x$ for any $\hat{\omega}\in\hat{\Omega}$.

Fix $\hat{\omega} \in \hat{\Omega}$. Under $\mathbf{P}^K_{\hat{N}_K(\hat{\omega})\delta_x + \frac{1}{K}\delta_{x+h}}$, define

$$Z_1^K = \langle \nu^K_{\theta_0}, \mathbf{1} \rangle \mathbf{1}_{\{V_0 = x, \theta_0 < \tau_1\}} + \frac{\lceil K \bar{n}_x \rceil}{K} \mathbf{1}_{\{V_0 \neq x\} \cup \{\theta_0 \geq \tau_1\}}.$$

It follows from Lemma 3 (41) and (42), and from assumption (B) that, for Lebesgue almost every h, $Z_1^K \xrightarrow{\mathcal{P}} \bar{n}_x$, so, by the induction assumption, under $\mathbf{P}^K_{\hat{N}_K(\hat{\omega})\delta_x + \frac{1}{K}\delta_{x+h}}$,

$$p_n^K(t - i/2^k, x, \Gamma, \infty, K Z_1^K) \xrightarrow{\mathcal{P}} p_n(t - i/2^k, x, \Gamma).$$

Now, given two sequences of uniformly bounded random variables $(X_K)_{K \geq 1}$ and $(Y_K)_{K \geq 1}$ such that X_K and Y_K are defined on the same probability space for any $K \geq 1$, and such that, when $K \to +\infty$, X_K converges in probability to a constant C and $\lim_K \mathbf{E}(Y_K)$ exists, it is standard to prove that

$$\lim_{K \to +\infty} \mathbf{E}(X_K Y_K) = C \lim_{K \to +\infty} \mathbf{E}(Y_K). \tag{53}$$

Applying this with $X_K = p_n^K(t - i/2^k, x, \Gamma, \infty, K Z_1^K)$ and $Y_K = \mathbf{1}_{\{V_0 = x,\ \theta_0 < \tau_1\}}$, by Lemma 3's (40) and (41) and assumption (B), for Lebesgue almost any h, and for any $\hat{\omega} \in \hat{\Omega}$,

$$\begin{aligned} &\lim_{K \to +\infty} \mathbf{E}^K_{\hat{N}_K(\hat{\omega})\delta_x + \frac{1}{K}\delta_{x+h}} \left(\mathbf{1}_{\{V_0 = x,\ \theta_0 < \tau_1\}} p_n^K(t - i/2^k, x, \Gamma, \infty, K \langle \nu^K_{\theta_0}, \mathbf{1} \rangle) \right) \\ &\quad = \left(1 - \frac{[f(x+h, x)]_+}{b(x+h)} \right) p_n(t - i/2^k, x, \Gamma). \end{aligned}$$

Finally, we obtain that, for Lebesgue almost any h, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$,

$$\begin{aligned} &\mathbf{E}^K_{\nu^K_{\tau_1 -} + \frac{1}{K}\delta_{x+h}} \left(\mathbf{1}_{\{V_0 = x, \theta_0 < \tau_1\}} p_n^K(t - i/2^k, x, \Gamma, \infty, K \langle \nu^K_{\theta_0}, \mathbf{1} \rangle) \right) \\ &\quad \xrightarrow{\mathcal{P}} \left(1 - \frac{[f(x+h, x)]_+}{b(x+h)} \right) p_n(t - i/2^k, x, \Gamma). \end{aligned} \tag{54}$$

Similarly, we can use Lemma 3's (39) and the random variable

$$Z_2^K = \langle \nu^K_{\theta_0}, \mathbf{1} \rangle \mathbf{1}_{\{V_0 = x+h, \theta_0 < \tau_1\}} + \bar{n}_{x+h} \mathbf{1}_{\{V_0 \neq x+h\} \cup \{\theta_0 \geq \tau_1\}}$$

to prove that, for Lebesgue almost any h, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$,

$$\begin{aligned} &\mathbf{E}^K_{\nu^K_{\tau_1 -} + \frac{1}{K}\delta_{x+h}} \left(\mathbf{1}_{\{V_0 = x+h,\ \theta_0 < \tau_1\}} p_n^K(t - i/2^k, x+h, \Gamma, \infty, K \langle \nu^K_{\theta_0}, \mathbf{1} \rangle) \right) \\ &\quad \xrightarrow{\mathcal{P}} \frac{[f(x+h, x)]_+}{b(x+h)} p_n(t - i/2^k, x+h, \Gamma). \end{aligned} \tag{55}$$

Moreover, by Lemma 3's (41), for Lebesgue almost any h, under $\mathbf{P}^K_{(z_K/K)\delta_x}$,

$$\mathbf{P}^K_{\nu^K_{\tau_1 -} + \frac{1}{K}\delta_{x+h}} \left(\theta_0 \geq \frac{1}{2^k K u_K} \wedge \tau_1 \right) \xrightarrow{\mathcal{P}} 0. \tag{56}$$

Collecting these results together, applying (53) again, it follows from Lemma 2(c) and (52) that, for Lebesgue almost any h,

$$\lim_{K\to+\infty} \mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^k K u_K}\le\tau_1\le\frac{i+1}{2^k K u_K}\right\}}\mathbf{E}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}\left(\mathbf{1}_{\left\{\theta_0\ge\frac{1}{2^k K u_K}\wedge\tau_1\right\}}\right.\right.$$

$$\left.\left.+\mathbf{1}_{\left\{\theta_0<\frac{1}{2^k K u_K}\wedge\tau_1\right\}} p^K_n(t-i/2^k, V_0, \Gamma, \infty, K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)\right)\right]$$

$$= \left(\mathrm{e}^{-\beta(x)\frac{i}{2^k}} - \mathrm{e}^{-\beta(x)\frac{i+1}{2^k}}\right)\left[\frac{[f(x+h,x)]_+}{b(x+h)} p_n(t-i/2^k, x+h, \Gamma)\right.$$

$$\left. + \left(1-\frac{[f(x+h,x)]_+}{b(x+h)}\right) p_n(t-i/2^k, x, \Gamma)\right].$$

Finally, taking the integral of both sides with respect to $m(x,h)dh$, the dominated convergence theorem and (51) yield

$$\limsup_{K\to+\infty} p^K_{n+1}(x,t,\Gamma,\varepsilon,z_K)$$

$$\le \sum_{i=0}^{\lceil t2^k\rceil-1}\left(\mathrm{e}^{-\beta(x)\frac{i}{2^k}} - \mathrm{e}^{-\beta(x)\frac{i+1}{2^k}}\right)\int_{\mathbb{R}^l} p_n(t-i/2^k, x+h, \Gamma)\kappa(x,\mathrm{d}h)+\eta.$$

Taking the limit $k\to+\infty$ first and then $\eta\to 0$, it follows from the fact that

$$\mathrm{e}^{-\beta(x)i/2^k} - \mathrm{e}^{-\beta(x)(i+1)/2^k} = \mathrm{e}^{-\beta(x)i/2^k}(\beta(x)/2^k + O(1/2^{2k}))$$

and from the convergence of Riemann sums that

$$\limsup_{K\to+\infty} p^K_{n+1}(x,t,\Gamma,\varepsilon,z_K) \le \int_0^t \beta(x)\mathrm{e}^{-\beta(x)s}\int_{\mathbb{R}^l} p_n(t-s, x+h, \Gamma)\kappa(x,\mathrm{d}h)\mathrm{d}s.$$

Using the same method as for (51), we can give a lower bound for p^K_n as follows: for any $\eta>0$, for sufficiently large $k\ge 0$ and $K\ge 1$,

$$p^K_{n+1}(t,x,\Gamma,\varepsilon,z_K) \ge \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\theta_{n+1}\le\frac{t}{Ku_K}, \tau_{n+2}>\frac{t-2/2^k}{Ku_K}, V_{n+1}\in\Gamma\right.$$

$$\left.\text{and}\quad \sup_{s\in[\theta_{n+1},\tau_{n+2}]}|\langle\nu^K_s,\mathbf{1}\rangle - \bar{n}_{V_{n+1}}|<\varepsilon\right)-\eta$$

$$\ge \sum_{i=0}^{\lfloor t2^k\rfloor-3}\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^k K u_K}\le\tau_1\le\frac{i+1}{2^k K u_K}\right\}}\mathbf{P}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{U_1}}\left(\theta_n\le\frac{t-(i+1)/2^k}{Ku_K},\right.\right.$$

$$\left.\left.\tau_{n+1}>\frac{t-(i+2)/2^k}{Ku_K}, V_n\in\Gamma \text{ and } \sup_{s\in[\theta_n,\tau_{n+1}]}|\langle\nu^K_s,\mathbf{1}\rangle-\bar{n}_{V_n}|<\varepsilon\right)\right]-\eta$$

$$\ge \sum_{i=0}^{\lfloor t2^k\rfloor-3}\mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\left\{\frac{i}{2^k K u_K}\le\tau_1\le\frac{i+1}{2^k K u_K}\right\}}\int_{\mathbb{R}^l}\mathbf{E}^K_{\nu^K_{\tau_1-}+\frac{1}{K}\delta_{x+h}}\left(\mathbf{1}_{\left\{\theta_0<\frac{1}{2^k K u_K}\wedge\tau_1\right\}}\right.\right.$$

$$\left.\left.p^K_n(t-(i+2)/2^k, V_0, \Gamma, \varepsilon, K\langle\nu^K_{\theta_0},\mathbf{1}\rangle)\right)m(x,h)\mathrm{d}h\right]-\eta.$$

Then, as above, letting $K \to +\infty$, then $k \to +\infty$ and finally $\eta \to 0$, we obtain

$$\liminf_{K\to+\infty} p^K_{n+1}(x, t, \Gamma, \varepsilon, z_K) \geq \int_0^t \beta(x)\mathrm{e}^{-\beta(x)s} \int_{\mathbb{R}^l} p_n(t-s, x+h, \Gamma)\kappa(x, \mathrm{d}h)\mathrm{d}s,$$

which completes the proof of Lemma 4 by induction. □

Proof of Lemma 2(a). Fix $\eta > 0$. By Theorem 2(a) and (c), for any $K \geq 1$,

$$\langle \nu^K, \mathbf{1}\rangle \preceq Z^K, \quad \text{where}$$
$$\mathcal{L}(Z^K) = \mathbf{P}^K(2\bar{b}, 0, \underline{\alpha}, \langle \nu^K_0, \mathbf{1}\rangle + 1).$$

Since $\sup_K \mathbf{E}(\langle \nu^K_0, \mathbf{1}\rangle) < +\infty$, we can choose $M < +\infty$ such that

$$\sup_{K\geq 1} \mathbf{P}(\langle \nu^K_0, \mathbf{1}\rangle + 1 > M) < \eta/3.$$

Then, apply Theorem 3(c) to $\mathbf{P}^K(2\bar{b}, 0, \underline{\alpha}, \langle \nu^K_0, \mathbf{1}\rangle + 1)$ with $C = [1, M]$, $\eta_2 = M$ and η_1 such that $0 < 2\bar{b}/\underline{\alpha} - \eta_1 < 1/2$: there exists a $V > 0$ such that

$$\limsup_{K\to+\infty} \mathbf{P}(T^K < \mathrm{e}^{KV}) < \eta/3, \quad \text{where}$$
$$T^K = \inf\{t \geq 0, Z^K_t \notin [1/2, M + 2\bar{b}/\underline{\alpha}]\}. \tag{57}$$

Fix $t, \varepsilon > 0$. Since, for $s \leq T^K$, $\langle \nu^K_s, \mathbf{1}\rangle \leq M + 2\bar{b}/\underline{\alpha}$, if we apply Theorem 2(b) to the process $(\nu^K_{s+(t/Ku_K)} - \nu^K_{t/Ku_K},\ s \geq 0)$, we obtain, for $s \leq T^K - t/Ku_K$,

$$A^K_{s+(t/Ku_K)} - A^K_{t/Ku_K} \preceq B^K_s,$$

where A^K_s is the number of mutations occurring between 0 and s, and where B^K is a Poisson process with parameter $Ku_K\bar{b}(M + 2\bar{b}/\underline{\alpha})$. Therefore, combining (57) with the fact that $1/Ku_K \ll \mathrm{e}^{KV}$, we obtain that, for sufficiently large K

$$\begin{aligned}\mathbf{P}(A^K_{(t+\varepsilon)/Ku_K} - A^K_{t/Ku_K} \geq 1) &\leq \mathbf{P}(B^K_{\varepsilon/Ku_K} \geq 1) + 2\eta/3\\ &= 1 - \exp(-\bar{b}(M + 2\bar{b}/\underline{\alpha})\varepsilon) + 2\eta/3,\end{aligned}$$

which can be made smaller than η if ε is sufficiently small. This ends the proof of (36). □

Proof of Lemma 2(b). Fix $\varepsilon > 0$. It follows from the construction (15) of ν^K that, for $t < \tau_1$, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x}$,

$$\nu^K_t = Z^K_t \delta_x, \quad \text{where}$$
$$\mathcal{L}(Z^K) = \mathbf{P}^K((1 - u_K\mu(x))b(x), d(x), \alpha(x, x), z_K/K).$$

Therefore, by Theorem 2(c), for K such that $u_K < \varepsilon$ and for $t \leq \tau_1$,

$$Z^{K,1}_t \preceq \langle \nu^K_t, \mathbf{1}\rangle \preceq Z^{K,2}_t, \quad \text{where}$$
$$\mathcal{L}(Z^{K,1}) = \mathbf{P}^K((1-\varepsilon)b(x), d(x), \alpha(x, x), z_K/K) \quad \text{and}$$
$$\mathcal{L}(Z^{K,2}) = \mathbf{P}^K(b(x), d(x), \alpha(x, x), z_K/K). \tag{58}$$

Now, let ϕ^1_y, resp. ϕ^2_y, be the solution to

$$\dot{\phi} = ((1-\varepsilon)b(x) - d(x) - \alpha(x, x)\phi)\phi, \quad \text{resp.}$$
$$\dot{\phi} = (b(x) - d(x) - \alpha(x, x)\phi)\phi,$$

with initial state y, and observe that, for any $y > 0$, when $t \to +\infty$, $\phi^1_y(t) \to \mathrm{e}^1 := \bar{n}_x - \varepsilon b(x)/\alpha(x,x)$ and $\phi^2_y(t) \to \mathrm{e}^2 := \bar{n}_x$.

Define, for any $y > 0$, $t^{i,y}_\varepsilon$ the first time such that $\forall s \geq t^{i,y}_\varepsilon$, $\phi^i_y(s) \in [\mathrm{e}^i - \varepsilon, \mathrm{e}^i + \varepsilon]$ $(i = 1, 2)$. Because of the continuity of the flows of these ODEs,

$$t^i_\varepsilon := \sup_{y \in [z/2, 2z]} t^{i,y}_\varepsilon < +\infty.$$

Let us apply Theorem 3(a) to $Z^{K,1}$ and $Z^{K,2}$ on $[0, t_\varepsilon]$, where $t_\varepsilon = t^1_\varepsilon \vee t^2_\varepsilon$: since $z_K/K \to z$, for sufficiently small $\delta > 0$, and for $i = 1, 2$,

$$\lim_{K \to +\infty} \mathbf{P}\left(\sup_{0 \leq t \leq t_\varepsilon} |Z^{K,i}_t - \phi^i_{z_K/K}(t)| > \delta \right) = 0.$$

If we choose $\delta < \varepsilon$, we obtain, for $i = 1, 2$,

$$\lim_{K \to +\infty} \mathbf{P}(|Z^{K,i}_{t_\varepsilon} - \mathrm{e}^i| < 2\varepsilon) = 1,$$

and so, for $i = 1, 2$,

$$\lim_{K \to +\infty} \mathbf{P}(|Z^{K,i}_{t_\varepsilon} - \bar{n}_x| < M\varepsilon) = 1, \tag{59}$$

where $M = 2 + b(x)/\alpha(x,x)$.

Now, assuming ε sufficiently small for $(M+1)\varepsilon < \bar{n}_x$, define the stopping times

$$T^{K,i}_\varepsilon = \inf\{t \geq t_\varepsilon : |Z^{K,i}_t - \bar{n}_x| > (M+1)\varepsilon\}$$

for $i = 1, 2$, and $T^K_\varepsilon = T^{K,1}_\varepsilon \wedge T^{K,2}_\varepsilon$.

For any $z \in \mathbb{N}/K$, define also

$$\mathbf{P}^{K,1}_z := \mathbf{P}^K((1-\varepsilon)b(x), d(x), \alpha(x,x), z).$$

Then, applying Theorem 3(c) to $\mathbf{P}^{K,1}_z$ with $C = [\bar{n}_x - M\varepsilon, \bar{n}_x + M\varepsilon]$, there exists a $V_1 > 0$ such that

$$\lim_{K \to +\infty} \inf_{z \in C} \mathbf{P}^{K,1}_z(\hat{T}_\varepsilon > \mathrm{e}^{KV_1}) = 1, \quad \text{where}$$

$$\hat{T}_\varepsilon = \inf\{t \geq 0 : |w_t - \bar{n}_x| > (M+1)\varepsilon\}. \tag{60}$$

Therefore, applying the Markov property at time t_ε, it follows from (59) that

$$\lim_{K \to +\infty} \mathbf{P}(T^{K,1}_\varepsilon > \mathrm{e}^{KV_1} + t_\varepsilon) = 1.$$

Similarly, there exists $V_2 > 0$ such that

$$\lim_{K \to +\infty} \mathbf{P}(T^{K,2}_\varepsilon > \mathrm{e}^{KV_2} + t_\varepsilon) = 1,$$

and thus

$$\lim_{K \to +\infty} \mathbf{P}(T^K_\varepsilon > \mathrm{e}^{KV}) = 1, \tag{61}$$

where $V := V_1 \wedge V_2$.

Now, because of (58),

$$\forall t \in [t_\varepsilon, T^K_\varepsilon \wedge \tau_1], \quad |\langle \nu^K_s, \mathbf{1} \rangle - \bar{n}_x| < (M+1)\varepsilon. \tag{62}$$

Therefore, since $\log K > t_\varepsilon$ for sufficiently large K, in order to complete the proof of (37), it suffices to show that

$$\lim_{K\to+\infty} \mathbf{P}(\tau_1 < T^K_\varepsilon) = 1. \tag{63}$$

If we denote by A^K_t the number of mutations occurring between t_ε and $t+t_\varepsilon$, by Theorem 2(b), for t such that $t_\varepsilon + t \le T^K_\varepsilon \wedge \tau_1$,

$$B^K \preceq A^K,$$

where B^K is a Poisson process with parameter $Ku_K(\bar{n}_x - (M+1)\varepsilon)\mu(x)b(x)$.

Therefore, if we denote by S^K the first time when $B^K_t = 1$, on the event $\{t_\varepsilon + S^K < T^K_\varepsilon\}$,

$$\tau_1 \le t_\varepsilon + S^K.$$

Since $\exp(-KV) \ll Ku_K$, $\lim_K \mathbf{P}(t_\varepsilon + S^K < e^{KV}) = 1$, and hence, by (61),

$$\lim_{K\to+\infty} \mathbf{P}(t_\varepsilon + S^K < T^K_\varepsilon) = 1,$$

which implies (63).

In the case where $z_K/K \to \bar{n}_x$, using (60) as above, we obtain easily

$$\lim_{K\to+\infty} \mathbf{P}(S^K_\varepsilon > e^{KV}) = 1, \quad \text{where}$$
$$S^K_\varepsilon = \inf\{t \ge 0 : |Z^{K,i}_t - \bar{n}_x| > (M+1)\varepsilon, i = 1, 2\}.$$

Then, the proof of (38) can be completed using the same method as we used above. $\square$

Proof of Lemma 2(c). Fix $t > 0$ and $\varepsilon > 0$. Take K large enough for $\log K < t/Ku_K$. The Markov property at time $\log K$ for ν^K yields

$$\mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}, \sup_{t\in[\log K,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right)$$
$$= \mathbf{E}^K_{\frac{z_K}{K}\delta_x}\left[\mathbf{1}_{\{\tau_1>\log K\}}\mathbf{P}^K_{\nu^K_{\log K}}\left(\tau_1 > \frac{t}{Ku_K} - \log K, \sup_{t\in[0,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right)\right]. \tag{64}$$

For any initial condition $\nu^K_0 = \langle \nu^K_0, \mathbf{1}\rangle\delta_x$ of ν^K, by Theorem 2(b), the number A^K_t of mutations of ν^K between 0 and t satisfies, for any $t \le \tau_1$ such that $\sup_{s\in[0,t]} |\langle \nu^K_s, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon$,

$$B^K_t \preceq A^K_t \preceq C^K_t,$$

where B^K_t and C^K_t are Poisson processes with respective parameters $Ku_K(\bar{n}_x - \varepsilon)\mu(x)b(x)$ and $Ku_K(\bar{n}_x + \varepsilon)\mu(x)b(x)$.

Therefore, on the event $\{\sup_{s\in[0,\tau_1]} |\langle \nu^K_s, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\}$, $S^K \le \tau_1 \le T^K$, where T^K is the first time when $B^K_t = 1$, and S^K the first time when $C^K_t = 1$.

Now, by Lemma 2(b), under $\mathbf{P}^K_{(z_K/K)\delta_x}$, $\nu^K_{\log K} \xrightarrow{\mathcal{P}} \bar{n}_x\delta_x$, so, by Skorohod's Theorem, we can construct $\hat{N}^K$ with the same law as $\langle \nu^K_{\log K}, \mathbf{1}\rangle$ on an auxiliary probability space $\hat{\Omega}$ such that

$\hat{N}^K(\hat{\omega}) \to \bar{n}_x$ for any $\hat{\omega} \in \hat{\Omega}$. Fix $\hat{\omega} \in \hat{\Omega}$. Then, by Lemma 2(b),

$$\lim_{K\to+\infty} \mathbf{P}^K_{\hat{N}(\hat{\omega})\delta_x}\left(\sup_{t\in[0,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right) = 1,$$

and so,

$$\begin{aligned}\limsup_{K\to+\infty} &\mathbf{P}^K_{\hat{N}(\hat{\omega})\delta_x}\left(\tau_1 > \frac{t}{Ku_K} - \log K, \sup_{t\in[0,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right) \\ &\le \limsup_{K\to+\infty} \mathbf{P}^K_{\hat{N}(\hat{\omega})\delta_x}\left(T^K > \frac{t}{Ku_K} - \log K\right) = \exp(-t(\bar{n}_x - \varepsilon)\mu(x)b(x)).\end{aligned}$$

Therefore, under $\mathbf{P}^K_{(z_K/K)\delta_x}$,

$$\begin{aligned}\limsup_{K\to+\infty} &\mathbf{P}^K_{\nu^K_{\log K}}\left(\tau_1 > \frac{t}{Ku_K} - \log K, \sup_{t\in[0,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right) \\ &\le \exp(-t(\bar{n}_x - \varepsilon)\mu(x)b(x))\end{aligned}$$

in probability (where $\limsup X_n \le a$ in probability means that, for any $\eta > 0$, $\mathbf{P}(X_n > a+\eta) \to 0$).

Similarly, under $\mathbf{P}^K_{(z_K/K)\delta_x}$,

$$\begin{aligned}\liminf_{K\to+\infty} &\mathbf{P}^K_{\nu^K_{\log K}}\left(\tau_1 > \frac{t}{Ku_K} - \log K, \sup_{t\in[0,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right) \\ &\ge \exp(-t(\bar{n}_x + \varepsilon)\mu(x)b(x))\end{aligned}$$

in probability.

Now, by Lemma 2(a) and (b),

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}(\tau_1 > \log K) = 1 \quad \text{and}$$

$$\lim_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\sup_{t\in[\log K,\tau_1]} |\langle \nu^K_t, \mathbf{1}\rangle - \bar{n}_x| < \varepsilon\right) = 1.$$

So, using property (53), it follows from (64) that

$$\begin{aligned}&\limsup_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}\right) \le \exp(-t(\bar{n}_x - \varepsilon)\mu(x)b(x)) \quad \text{and} \\ &\liminf_{K\to+\infty} \mathbf{P}^K_{\frac{z_K}{K}\delta_x}\left(\tau_1 > \frac{t}{Ku_K}\right) \ge \exp(-t(\bar{n}_x + \varepsilon)\mu(x)b(x)).\end{aligned}$$

Since this holds for any $\varepsilon > 0$, we have completed the proof of Lemma 2(c). □

Proof of Lemma 3. The proof of this lemma follows the three steps of the invasion of a mutant described in Section 3 (cf. Fig. 1).

Fix $\eta > 0$, $\varepsilon_0 > 0$ and $0 < \varepsilon < \varepsilon_0$. By Lemma 2(a), there exists a constant $\rho > 0$ that we can assume smaller than η, such that, for sufficiently large K,

$$\mathbf{P}^K_{\frac{z_K}{K}\delta_x + \frac{1}{K}\delta_y}\left(\tau_1 < \frac{\rho}{Ku_K}\right) < \varepsilon. \tag{65}$$

Observe that, under $\mathbf{P}^K_{\frac{z_K}{K}\delta_x+\frac{1}{K}\delta_y}$, for $t \leq \tau_1$,

$$\mathcal{L}((\langle \nu^K, \mathbf{1}_{\{x\}}\rangle, \langle \nu^K, \mathbf{1}_{\{y\}}\rangle)) = \mathbf{Q}^K((1-u_K\mu(x))b(x), (1-u_K\mu(y))b(y), d(x), d(y), \alpha(x,x), \alpha(x,y), \alpha(y,x), \alpha(y,y), z_K/K, 1/K).$$

Fix K large enough for $u_K < \varepsilon$. Define

$$S^K_\varepsilon := \inf\{s \geq 0 : \langle \nu^K_s, \mathbf{1}_{\{y\}}\rangle \geq \varepsilon\}.$$

By Theorem 2(c) and (d), for $t < \tau_1 \wedge S^K_\varepsilon$,

$$\begin{aligned} &Z^{K,1}_t \preceq \langle \nu^K_t, \mathbf{1}_{\{x\}}\rangle \preceq Z^{K,2}_t, \quad \text{where} \\ &\mathcal{L}(Z^{K,1}) = \mathbf{P}^K((1-\varepsilon)b(x), d(x)+\varepsilon\alpha(x,y), \alpha(x,x), z_K/K) \quad \text{and} \\ &\mathcal{L}(Z^{K,2}) = \mathbf{P}^K(b(x), d(x), \alpha(x,x), z_K/K). \end{aligned} \tag{66}$$

Using the method that led us to (61), we can deduce from Theorem 3(c) that there exists $V > 0$ such that

$$\begin{aligned} &\lim_{K\to+\infty} \mathbf{P}(R^K_\varepsilon > \mathrm{e}^{KV}) = 1, \quad \text{where} \\ &R^K_\varepsilon = \inf\{t \geq 0 : |Z^{K,i}_t - \bar{n}_x| > M\varepsilon, i = 1, 2\}, \end{aligned} \tag{67}$$

with $M = 3 + (b(x)+\alpha(x,y))/\alpha(x,x)$.

Now, observe that, by (66),

$$\forall t \leq \tau_1 \wedge S^K_\varepsilon \wedge R^K_\varepsilon, \quad \langle \nu^K_t, \mathbf{1}_{\{x\}}\rangle \in [\bar{n}_x - M\varepsilon, \bar{n}_x + M\varepsilon].$$

Therefore, by Theorem 2(c) and (e), for $t \leq \tau_1 \wedge S^K_\varepsilon \wedge R^K_\varepsilon$

$$\begin{aligned} &Z^{K,3}_t \preceq \langle \nu^K_t, \mathbf{1}_{\{y\}}\rangle \preceq Z^{K,4}_t, \quad \text{where} \\ &\mathcal{L}(Z^{K,3}) = \mathbf{P}^K((1-\varepsilon)b(y), d(y)+(\bar{n}_x+M\varepsilon)\alpha(y,x)+\varepsilon\alpha(y,y), 0, 1/K) \quad \text{and} \\ &\mathcal{L}(Z^{K,4}) = \mathbf{P}^K(b(y), d(y)+(\bar{n}_x-M\varepsilon)\alpha(y,x), 0, 1/K). \end{aligned} \tag{68}$$

Define, for any $K \geq 1$, $n \in \mathbb{N}$ and $i \in \{3, 4\}$, the stopping time

$$T^{K,i}_{n/K} = \inf\{t \geq 0 : Z^{K,i}_t = n/K\}.$$

Observe that, if $S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon$,

$$T^{K,4}_{\lceil \varepsilon K\rceil/K} \leq S^K_\varepsilon \leq T^{K,3}_{\lceil \varepsilon K\rceil/K} \tag{69}$$

and that, if $T^{K,4}_0 < T^{K,4}_{\lceil \varepsilon K\rceil/K} \wedge \tau_1 \wedge R^K_\varepsilon$,

$$\theta_0 \leq T^{K,4}_0.$$

If $Z^{K,4}$ is sub-critical, apply Theorem 4 (25), and if $Z^{K,4}$ is super-critical, apply Theorem 4 (28) (the critical case can be excluded by slightly changing the value of ε). Since $\log K \ll 1/Ku_K$, we obtain

$$\begin{aligned} &\lim_{K\to+\infty} \mathbf{P}\left(T^{K,4}_0 \leq \frac{\rho}{Ku_K} \wedge T^{K,4}_{\lceil \varepsilon K\rceil/K}\right) \\ &\quad = \frac{d(y)+(\bar{n}_x-M\varepsilon)\alpha(y,x)}{b(y)} \wedge 1 \geq 1 - \frac{[f(y,x)]_+}{b(y)} - \frac{\alpha(y,x)}{b(y)}M\varepsilon. \end{aligned} \tag{70}$$

Combining (65), (67), (68) and (70), and using the facts that $\rho < \eta$, $\varepsilon < \varepsilon_0$ and $\exp(KV) > \rho/Ku_K$ for sufficiently large K, we obtain, taking K larger if necessary,

$$\begin{aligned}
&\mathbf{P}\left(\theta_0 < \tau_1 \wedge \frac{\eta}{Ku_K}, V_0 = x \text{ and } |\langle \nu^K_{\theta_0}, \mathbf{1}\rangle - \bar{n}_x| < M\varepsilon_0\right) \\
&\quad \geq \mathbf{P}\left(\theta_0 < \tau_1 \wedge S^K_\varepsilon \wedge R^K_\varepsilon \wedge \frac{\rho}{Ku_K} \text{ and } V_0 = x\right) \\
&\quad \geq \mathbf{P}\left(T^{K,4}_0 < \tau_1 \wedge T^{K,4}_{\lceil \varepsilon K\rceil/K} \wedge R^K_\varepsilon \wedge \frac{\rho}{Ku_K}\right) \\
&\quad \geq 1 - \frac{[f(y,x)]_+}{b(y)} - \left(\frac{\alpha(y,x)}{b(y)}M + 3\right)\varepsilon.
\end{aligned} \tag{71}$$

This ends the proof of Lemma 3 in the case where $f(y,x) \leq 0$.

Let us assume that $f(y,x) > 0$, i.e. that $b(y) - d(y) - \bar{n}_x\alpha(y,x) > 0$. If we choose $\varepsilon > 0$ sufficiently small, then $Z^{K,3}$ is super-critical. By Theorem 4 (29),

$$\begin{aligned}
\lim_{K\to+\infty} \mathbf{P}\left(T^{K,3}_{\lceil \varepsilon K\rceil/K} < \frac{\rho}{3Ku_K}\right) &= \frac{(1-\varepsilon)b(y) - d(y) - (\bar{n}_x + M\varepsilon)\alpha(y,x) - \varepsilon\alpha(y,y)}{(1-\varepsilon)b(y)} \\
&\geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - \varepsilon\frac{b(y) + M\alpha(y,x) + \alpha(y,y)}{(1-\varepsilon)b(y)}.
\end{aligned}$$

Therefore, by (65) and (67), assuming (without loss of generality) that $\varepsilon < 1/2$, for sufficiently large K,

$$\mathbf{P}\left(T^{K,3}_{\lceil \varepsilon K\rceil/K} < \tau_1 \wedge R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}\right) \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - M'\varepsilon,$$

where $M' := 2(b(y) + M\alpha(y,x) + \alpha(y,y))/b(y) + 3$. Then, it follows from (69) that

$$\mathbf{P}\left(S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}\right) \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - M'\varepsilon. \tag{72}$$

Observe that, on the event $\{S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon \wedge (\rho/3Ku_K)\}$,

$$\langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{y\}}\rangle = \lceil \varepsilon K\rceil/K \quad \text{and} \quad |\langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{x\}}\rangle - \bar{n}_x| < M\varepsilon. \tag{73}$$

Now, since we have assumed $f(y,x) > 0$, x and y satisfy (7) and, by Proposition 3, any solution to (13) with initial state in the compact set $[\bar{n}_x - M\varepsilon, \bar{n}_x + M\varepsilon] \times [\varepsilon/2, 2\varepsilon]$ converges to $(0, \bar{n}_y)$ when $t \to +\infty$. As in the proof of Lemma 2(b), because of the continuity of the flow of system (13), we can find $t_\varepsilon < +\infty$ large enough that none of these solutions leave the set $[0, \varepsilon^2/2] \times [\bar{n}_y - \varepsilon/2, \bar{n}_y + \varepsilon/2]$ after time t_ε.

Apply Theorem 3(b) on $[0, t_\varepsilon]$, with $C = [\bar{n}_x - M\varepsilon, \bar{n}_x + M\varepsilon] \times [\varepsilon/2, 2\varepsilon]$ and with a constant $\delta < \varepsilon^2/2 \wedge r$, where r is defined in (19) (with $T = t_\varepsilon$). Then, with the notation of Theorem 3(b), because of (72) and (73), the Markov property at time S^K_ε yields

$$\liminf_{K\to+\infty} \mathbf{P}\Bigg(S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon \wedge \frac{\rho}{3Ku_K},$$
$$\sup_{S^K_\varepsilon \le s \le S^K_\varepsilon + t_\varepsilon} \left\| \Big(\langle \nu^K_s, \mathbf{1}_{\{x\}}\rangle, \langle \nu^K_s, \mathbf{1}_{\{y\}}\rangle \Big) - \phi_{\langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{x\}}\rangle, \langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{y\}}\rangle}(s) \right\| \le \delta \Bigg)$$
$$\ge \frac{f(y,x)}{(1-\varepsilon)b(y)} - M'\varepsilon. \tag{74}$$

Now, observe that, since $\delta < r$, on the event

$$\left\{ S^K_\varepsilon < \tau_1 \wedge R^K_\varepsilon, \sup_{S^K_\varepsilon \le s \le S^K_\varepsilon + t_\varepsilon} \left\| \Big(\langle \nu^K_s, \mathbf{1}_{\{x\}}\rangle, \langle \nu^K_s, \mathbf{1}_{\{y\}}\rangle \Big) - \phi_{\langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{x\}}\rangle, \langle \nu^K_{S^K_\varepsilon}, \mathbf{1}_{\{y\}}\rangle}(s) \right\| \le \delta \right\},$$

for any $t \in [S^K_\varepsilon, S^K_\varepsilon + t_\varepsilon]$, $\langle \nu^K_t, \mathbf{1}_{\{x\}}\rangle \ge r - \delta > 0$ and $\langle \nu^K_t, \mathbf{1}_{\{y\}}\rangle \ge r - \delta > 0$, and thus

$$\theta_0 > S^K_\varepsilon + t_\varepsilon.$$

Therefore, since $\delta < \varepsilon^2/2 < \varepsilon/2$, by (65) and (74), for sufficiently large K,

$$\mathbf{P}\Bigg(S^K_\varepsilon < R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}, \tau_1 > \frac{\rho}{3Ku_K} + t_\varepsilon, \theta_0 > S^K_\varepsilon + t_\varepsilon,$$
$$\langle \nu^K_{S^K_\varepsilon + t_\varepsilon}, \mathbf{1}_{\{x\}}\rangle < \varepsilon^2 \text{ and } \langle \nu^K_{S^K_\varepsilon + t_\varepsilon}, \mathbf{1}_{\{y\}}\rangle \in [\bar{n}_y - \varepsilon, \bar{n}_y + \varepsilon] \Bigg)$$
$$\ge \frac{f(y,x)}{(1-\varepsilon)b(y)} - (M'+2)\varepsilon. \tag{75}$$

Now, we will compare $\langle \nu^K, \mathbf{1}_{\{x\}}\rangle$ with a branching process after time $S^K_\varepsilon + t_\varepsilon$ in order to prove that trait x becomes extinct with a very high probability. We will use a method very similar to the one we used in the beginning of this proof. First, on the event inside the probability in (75), $\langle \nu^K_{S^K_\varepsilon + t_\varepsilon}, \mathbf{1}_{\{x\}}\rangle < \varepsilon^2$. In order to prove that the population with trait x stays small after $S^K_\varepsilon + t_\varepsilon$, let us define the stopping time

$$\hat{S}^K_\varepsilon = \inf\{t \ge S^K_\varepsilon + t_\varepsilon : \langle \nu^K_t, \mathbf{1}_{\{x\}}\rangle > \varepsilon\}$$

(recall that $\varepsilon^2 < \varepsilon$ since $\varepsilon < 1/2$). Using Theorem 2(c) and (d) again, we see that, on the event

$$F^{K,\varepsilon} := \{\langle \nu^K_{S^K_\varepsilon + t_\varepsilon}, \mathbf{1}_{\{x\}}\rangle < \varepsilon^2, \langle \nu^K_{S^K_\varepsilon + t_\varepsilon}, \mathbf{1}_{\{y\}}\rangle \in [\bar{n}_y - \varepsilon, \bar{n}_y + \varepsilon]\},$$

for any $t \ge 0$ such that $S^K_\varepsilon + t_\varepsilon + t \le \hat{S}^K_\varepsilon \wedge \tau_1$,

$$Z^{K,5}_t \preceq \langle \nu^K_{S^K_\varepsilon + t_\varepsilon + t}, \mathbf{1}_{\{y\}}\rangle \preceq Z^{K,6}_t, \quad \text{where}$$
$$\mathcal{L}(Z^{K,5}) = \mathbf{P}^K((1-\varepsilon)b(y), d(y) + \varepsilon\alpha(y,x), \alpha(y,y), \lfloor(\bar{n}_y - \varepsilon)K\rfloor/K) \quad \text{and}$$
$$\mathcal{L}(Z^{K,6}) = \mathbf{P}^K(b(y), d(y), \alpha(y,y), \lceil(\bar{n}_y + \varepsilon)K\rceil/K).$$

We can apply Theorem 3(c) to $Z^{K,5}$ and $Z^{K,6}$ as above to obtain a constant $V' > 0$ such that

$$\lim_{K\to+\infty} \mathbf{P}(\hat{R}^K_\varepsilon > \mathrm{e}^{KV'}) = 1, \quad \text{where}$$
$$\hat{R}^K_\varepsilon = \inf\{t \geq 0 : |Z^{K,i}_t - \bar{n}_y| > M''\varepsilon, i = 5, 6\}, \tag{76}$$

with $M'' = 3 + (b(y) + \alpha(y, x))/\alpha(y, y)$.

Observe that, on the event $F^{K,\varepsilon}$, for any $t \leq \hat{R}^K_\varepsilon$ such that $S^K_\varepsilon + t_\varepsilon + t \leq \hat{S}^K_\varepsilon \wedge \tau_1$,

$$|\langle \nu^K_{S^K_\varepsilon+t_\varepsilon+t}, \mathbf{1}_{\{y\}}\rangle - \bar{n}_y| \leq M''\varepsilon,$$

and so, by Theorem 2(c) and (e), on $F^{K,\varepsilon}$ and for t as above,

$$\langle \nu^K_{S^K_\varepsilon+t_\varepsilon+t}, \mathbf{1}_{\{x\}}\rangle \preceq Z^{K,7}_t \quad \text{where}$$
$$\mathcal{L}(Z^{K,7}) = \mathbf{P}^K(b(x), d(x) + (\bar{n}_y - M''\varepsilon)\alpha(x, y), 0, \lceil \varepsilon^2 K\rceil/K).$$

Now, since x and y satisfy (7), $Z^{K,7}$ is sub-critical for sufficiently small ε. Fix such an $\varepsilon > 0$ and define for any $n \geq 0$

$$\hat{T}^K_{n/K} = \inf\{t \geq 0 : Z^{K,7}_t = n/K\}.$$

If $\hat{T}^K_{\lceil\varepsilon K\rceil/K} \leq \hat{R}^K_\varepsilon$ and $S^K_\varepsilon + t_\varepsilon + \hat{T}^K_{\lceil\varepsilon K\rceil/K} \leq \tau_1$, then

$$\hat{S}^K_\varepsilon \geq S^K_\varepsilon + t_\varepsilon + \hat{T}^K_{\lceil\varepsilon K\rceil/K}$$

and if $\hat{T}^K_0 \leq \hat{R}^K_\varepsilon$ and $S^K_\varepsilon + t_\varepsilon + \hat{T}^K_0 \leq \hat{S}^K_\varepsilon \wedge \tau_1$, then

$$\theta_0 \leq \hat{T}^K_0.$$

Moreover, by Theorem 4's (26) and (27), for sufficiently large K,

$$\mathbf{P}\left(\hat{T}^K_0 \leq \frac{\rho}{3Ku_K}\right) \geq 1 - \varepsilon \quad \text{and}$$
$$\mathbf{P}(\hat{T}^K_{\lceil K\varepsilon\rceil/K} \leq \hat{T}^K_0) \leq 2\varepsilon.$$

Combining the last two inequalities with (65), (75) and (76), and recalling that $\rho < \eta$ and $\varepsilon < \varepsilon_0$, we finally obtain, for sufficiently large K,

$$\begin{aligned}
&\mathbf{P}\left(\theta_0 < \tau_1 \wedge \frac{\eta}{Ku_K}, V_0 = y \text{ and } |\langle \nu^K_{\theta_0}, \mathbf{1}\rangle - \bar{n}_y| < M''\varepsilon_0\right) \\
&\quad \geq \mathbf{P}\Bigg(S^K_\varepsilon < R^K_\varepsilon \wedge \frac{\rho}{3Ku_K}, \theta_0 > S^K_\varepsilon + t_\varepsilon, \tau_1 > \frac{2\rho}{3Ku_K} + t_\varepsilon, \langle \nu^K_{S^K_\varepsilon+t_\varepsilon}, \mathbf{1}_{\{x\}}\rangle < \varepsilon^2, \\
&\qquad \langle \nu^K_{S^K_\varepsilon+t_\varepsilon}, \mathbf{1}_{\{y\}}\rangle \in [\bar{n}_y - \varepsilon, \bar{n}_y + \varepsilon], \hat{T}^K_0 < \frac{\rho}{3Ku_K} \wedge \hat{T}^K_{\lceil K\varepsilon\rceil/K} \text{ and } \hat{R}^K_\varepsilon > \frac{\rho}{Ku_K}\Bigg) \\
&\quad \geq \frac{f(y, x)}{(1-\varepsilon)b(y)} - (M' + 7)\varepsilon.
\end{aligned}$$

Adding this inequality to (71), we obtain

$$\mathbf{P}\left(\theta_0 < \tau_1 \wedge \frac{\eta}{Ku_K}\right) \geq 1 - \frac{\varepsilon}{1-\varepsilon}\frac{f(y, x)}{b(y)} - \left(M\frac{\alpha(y, x)}{b(y)} + M' + 10\right)\varepsilon \geq 1 - M'''\varepsilon,$$

where $M''' = 2f(y, x)/b(y) + M\alpha(y, x)/b(y) + M' + 10$, which implies (41), and

$$\mathbf{P}\left(|\langle \nu^K_{\theta_0}, \mathbf{1}\rangle - \bar{n}_{V_0}| < (M \vee M'')\varepsilon_0\right) \geq 1 - M'''\varepsilon,$$

which implies (41).

Therefore,

$$\mathbf{P}(V_0 = x) \geq 1 - \frac{f(y,x)}{b(y)} - 2M'''\varepsilon \quad \text{and} \quad \mathbf{P}(V_0 = y) \geq \frac{f(y,x)}{(1-\varepsilon)b(y)} - 2M'''\varepsilon.$$

Since $\mathbf{P}(V_0 = x) \leq 1 - \mathbf{P}(V_0 = y)$, we finally obtain (39) and (40). □

Acknowledgments

I would like to thank S. Méléard and R. Ferrière for their continual guidance during my work. I also thank the referees for useful comments and suggestions.

References

[1] K.B. Athreya, P.E. Ney, Branching Processes, Springer, New York, 1972.
[2] P. Billingsley, Convergence of Probability Measures, Wiley, New York, 1968.
[3] B.M. Bolker, S.W. Pacala, Using moment equations to understand stochastically driven spatial pattern formation in ecological systems, Theor. Popul. Biol. 52 (1997) 179–197.
[4] B.M. Bolker, S.W. Pacala, Spatial moment equations for plant competition: Understanding spatial strategies and the advantages of short dispersal, Am. Nat. 153 (1999) 575–602.
[5] N. Champagnat, Étude mathématique de modèles stochastiques d'évolution issus de la théorie écologique des dynamiques adaptatives, Ph.D. Thesis, University of Paris X, Nanterre, 2004.
[6] N. Champagnat, R. Ferrière, G. Ben Arous, The canonical equation of adaptive dynamics: A mathematical view, Selection 2 (2001) 73–83.
[7] U. Dieckmann, M. Doebeli, On the origin of species by sympatric speciation, Nature 400 (1999) 354–357.
[8] U. Dieckmann, R. Law, The dynamical theory of coevolution: A derivation from stochastic ecological processes, J. Math. Biol. 34 (1996) 579–612.
[9] U. Dieckmann, R. Law, Relaxation projections and the method of moments, in: U. Dieckmann, R. Law, J.A.J. Metz (Eds.), The Geometry of Ecological Interactions: Simplifying Spatial Complexity, Cambridge University Press, Cambridge, 2000, pp. 412–455.
[10] P. Dupuis, R.S. Ellis, A Weak Convergence Approach to the Theory of Large Deviations, Wiley, New York, 1997.
[11] R. Durrett, J. Schweinsberg, Approximating selective sweeps, Theor. Pop. Biol. 66 (2004) 129–138.
[12] S.N. Ethier, T.G. Kurtz, Markov Processes, Characterization and Convergence, Wiley, New York, 1986.
[13] N. Fournier, S. Méléard, A microscopic probabilistic description of a locally regulated population and macroscopic approximations, Ann. Appl. Probab. 14 (2004) 1880–1919.
[14] M.I. Freidlin, A.D. Wentzell, Random Perturbations of Dynamical Systems, Springer, New York, 1984.
[15] J. Hofbauer, R. Sigmund, Adaptive dynamics and evolutionary stability, Appl. Math. Lett. 3 (1990) 75–79.
[16] J. Istas, Introduction aux modélisations mathématiques pour les sciences du vivant, Springer, New York, 2000.
[17] N.L. Kaplan, R.R. Hudson, C.H. Langley, The hitch-hiking effect revisited, Genetics 123 (1989) 887–899.
[18] R. Law, C.J. Murrel, U. Dieckmann, Population growth in space and time: Spatial logistic equations, Ecology 84 (2003) 252–262.
[19] P. Marrow, R. Law, C. Cannings, The coevolution of predator–prey interactions: ESSs and Red Queen dynamics, Proc. Roy. Soc. Lond. B 250 (1992) 133–141.
[20] J.A.J. Metz, R.M. Nisbet, S.A.H. Geritz, How should we define 'fitness' for general ecological scenarios? Trends in Ecology and Evolution 7 (1992) 198–202.
[21] J.A.J. Metz, S.A.H. Geritz, G. Meszéna, F.A.J. Jacobs, J.S. van Heerwaarden, Adaptive Dynamics, a geometrical study of the consequences of nearly faithful reproduction, in: S.J. van Strien, S.M. Verduyn Lunel (Eds.), Stochastic and Spatial Structures of Dynamical Systems, North Holland, Amsterdam, 1996, pp. 183–231.