



**HAL**  
open science

# Sharp approximation and hitting times for stochastic invasion processes

Vincent Bansaye, Xavier Erny, Sylvie Méléard

► **To cite this version:**

Vincent Bansaye, Xavier Erny, Sylvie Méléard. Sharp approximation and hitting times for stochastic invasion processes. 2022. hal-03915479v1

**HAL Id: hal-03915479**

**<https://hal.science/hal-03915479v1>**

Preprint submitted on 29 Dec 2022 (v1), last revised 11 Dec 2023 (v3)

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

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

# Sharp approximation and hitting times for stochastic invasion processes

Vincent Bansaye\*, Xavier Erny\* and Sylvie Méléard\*

\* *CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France*

**Abstract:** We are interested in the invasion phase for stochastic processes with interactions when a single mutant with positive fitness arrives in a resident population at equilibrium. By a now classic approach, the first stage of the invasion is well approximated by a branching process. The macroscopic phase, when the mutant population is of the same order of the resident population, is described by the limiting dynamical system. We obtain sharper estimates and capture the intermediate mesoscopic phase for the invasive population. It allows us to characterize the hitting times of thresholds, which inherit a large variance from the first stages. These issues are motivated in particular by quantifying times to reach critical values for cancer population or epidemics.

**MSC2020 subject classifications:** 37N25, 60F15, 60G55, 60H10.

**Keywords and phrases:** Branching process, Dynamical system, Stochastic processes approximation, Hitting times approximation.

## 1. Introduction and main results

We aim to finely quantify a mutant invasion in a resident population at equilibrium. Such situation is standard in eco-evolution, in cancer emergence or in epidemiology when a single individual particularly well adapted can develop its own subpopulation and invade the global population. After a certain amount of time, the number of mutants becomes non negligible with respect to the resident population size, allowing to summarize the dynamics of the stochastic population processes by their deterministic approximations. In Champagnat [14], a systematic approach was introduced to quantify the mutant invasion success, based on the fine properties of the mutant birth and death process and its coupling with branching processes for which the survival probability was easily computed. That allowed to characterize the probability for the mutant process to attain a certain fixed threshold. This approach has been used by many authors [15], [7], [2], [8], [12].

To go in details in this invasion process, it is necessary to introduce some scaling parameter  $K$  characterizing the macroscopic population sizes and allowing to quantify each invasion step. We consider a bi-type birth and death process modeling the population size dynamics of resident and mutant individuals which are interacting. We denote by  $N_R^K(t)$  (resp.  $N_M^K(t)$ ) the number of resident individuals (resp. mutant individuals) at time  $t$ . The interaction is modeled through density dependence of birth and death rates. For that purpose, we introduce the scaled processes

$$X^K(t) = (X_R^K(t), X_M^K(t)) := (N_R^K(t)/K, N_M^K(t)/K).$$

The processes  $(N_R^K, N_M^K)$  are defined on a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , as solutions of the stochastic differential systems driven by Poisson point measures  $\mathcal{N}_R^b, \mathcal{N}_R^d, \mathcal{N}_M^b, \mathcal{N}_M^d$  with Lebesgue measure intensity on  $\mathbb{R}_+ \times \mathbb{R}_+$ : for  $\bullet \in \{R, M\}$ , for any  $K$  and time  $t$ ,

$$\begin{aligned} N_\bullet^K(t) = & N_\bullet^K(0) + \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq N_\bullet^K(s-) b_\bullet(X^K(s-))\}} \mathcal{N}_\bullet^b(ds, su) \\ & - \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq N_\bullet^K(s-) d_\bullet(X^K(s-))\}} \mathcal{N}_\bullet^d(ds, du). \end{aligned} \quad (1.1)$$

We assume that, for  $\bullet \in \{R, M\}$ ,  $\mathcal{N}_\bullet^b$  and  $\mathcal{N}_\bullet^d$  are independent, but  $(\mathcal{N}_R^b, \mathcal{N}_R^d)$  and  $(\mathcal{N}_M^b, \mathcal{N}_M^d)$  are not necessarily independent. We refer to Section 5.1 for an application where independence holds and 5.2

for an example where it does not.

The individual birth and death rates of resident and mutant individuals denoted by  $b_\bullet, d_\bullet$  are functions of the scaled resident and mutant population sizes  $X_R^K$  and  $X_M^K$ . This dependence models the interaction between both populations, as competition for resources, which for example may impact the death rates. The individual growth rates are defined on  $\mathbb{R}_+ \times \mathbb{R}_+$  by

$$F_R(x_R, x_M) = b_R(x_R, x_M) - d_R(x_R, x_M); \quad F_M(x_R, x_M) = b_M(x_R, x_M) - d_M(x_R, x_M).$$

It is well known (cf [19]) that under suitable assumptions on the parameters and when initial conditions of the branching processes are of order  $K$ , the stochastic process  $((X_R^K(t), X_M^K(t)) : t \geq 0)$  converges in probability, on any finite time interval, to the dynamical system

$$\begin{cases} x'_R &= x_R F_R(x_R, x_M) = G_R(x_R, x_M); \\ x'_M &= x_M F_M(x_R, x_M) = G_M(x_R, x_M). \end{cases} \quad (1.2)$$

The dynamics of the macroscopic resident population  $(N_R^K(t), t \geq 0)$  is then approximated by  $(K x_R(t), t \geq 0)$ . We assume the existence of an equilibrium  $x_R^* > 0$  satisfying

$$F_R(x_R^*, 0) = b_R(x_R^*, 0) - d_R(x_R^*, 0) = 0.$$

When a mutation appears in the population at the equilibrium, the mutant process  $(N_M^K(t), t \geq 0)$  will firstly be close to a branching process with birth and death rates respectively

$$b_\star = b_M(x_R^*, 0); \quad d_\star = d_M(x_R^*, 0).$$

We assume that the initial growth rate of this process is positive

$$r_\star = b_\star - d_\star > 0. \quad (1.3)$$

This ensures that with positive probability the mutant population invade.

The invasion success can be characterized by the fact that the mutant process attains a threshold of order of magnitude  $K$  with a positive probability when  $K$  tends to infinity. Let us introduce the hitting times

$$\tau_M^K(n) = \inf\{t \geq 0 : N_M^K(t) \geq n\} = \inf\{t \geq 0 : X_M^K(t) \geq n/K\}. \quad (1.4)$$

When such levels are reached, Champagnat [14] proved that the time  $\tau_M^K(\varepsilon K)$  needed is of order of magnitude  $\log K$ . Its proof is based on coupling between the process of interest and two branching processes, respectively  $Z_\varepsilon^+$  and  $Z_\varepsilon^-$  above and below:

$$Z_\varepsilon^-(t) \leq N_M^K(t) \leq Z_\varepsilon^+(t) \quad \text{a.s.}$$

as long as  $N_M^K$  is smaller than  $\varepsilon K$ . The Malthusian parameters of these branching processes are respectively  $\lambda_-(\varepsilon)$  and  $\lambda_+(\varepsilon)$  satisfying  $\lambda_-(\varepsilon) \leq r_\star \leq \lambda_+(\varepsilon)$  and converging both to  $r_\star$  when  $\varepsilon$  tends to 0. For branching processes, it is known that the population size grows a.s. exponentially on the survival event, more precisely

$$Z_\varepsilon^-(t) \sim_{t \rightarrow \infty} W_\varepsilon^- e^{\lambda_-(\varepsilon)t}, \quad Z_\varepsilon^+(t) \sim_{t \rightarrow \infty} W_\varepsilon^+ e^{\lambda_+(\varepsilon)t} \quad \text{a.s.},$$

Thus, for any  $\varepsilon > 0$  and  $(t_K)_K$  a sequence such that  $t_K \gg \log K$ ,

$$\lim_{K \rightarrow \infty} \mathbb{P}(\tau_M^K(\varepsilon K) \leq t_K) = 1 - \frac{d_\star}{b_\star},$$

where the last hand side is the survival probability of a birth and death process with individual birth rate  $b_\star$  and death rate  $d_\star$ . From Champagnat's work, we can further deduce that for any  $\eta > 0$ , for  $\varepsilon$  small enough,

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(\left|\frac{\tau_M^K(\varepsilon K)}{\log K} - \frac{1}{r_\star}\right| \geq \eta; \forall t > 0, N_M^K(t) > 0\right) = 0.$$

The time for reaching macroscopic levels of order  $K$  is thus approximated, at the first order, by  $\log K/r_*$ . In the situations when  $K$  is very large, as for huge cohorts of microorganisms (bacteria, cells), this result may be precise enough. But in many other cases,  $K$  is not so large. We can refer to the modeling of hematopoiesis and leukemias [10, 11] where  $K$  is the number of stem cells, of order  $10^5$ . In that case, we need to go further and to quantify  $\tau_M^K(\varepsilon K)$  as function of  $K$ , for large  $K$ . In particular, we were interested by the variance of this time and its dependence on  $\varepsilon$ . Our ambition has become more generally to capture the dynamics of the invasion process until macroscopic levels, in particular to approximate the process when it is large but negligible compared to  $K$  (mesoscopic scale).

The first step consists in evaluating the random time until which the mutation process  $N_M^K$  and the coupled binary branching process  $Z$  with birth and death rates  $b_*$  and  $d_*$  coincide, when  $K$  becomes large. Such exact coupling are known and used in particular in epidemiology to determine outbreak criterion and characterize the growth of epidemics in the beginning. For SIR model in continuous time, we refer in particular to Ball and Donnelly [4] for a proof of exact coupling during a time interval whose length grows as  $c \log K$  for some  $c > 0$ . But this exact coupling does not hold any longer when the population becomes too large. Thus, in Ball and Donnelly, it holds until the population reaches order of magnitude  $\sqrt{K}$ . After this time window where processes coincide, the error between the processes grows but the two processes may stay comparable, see in particular Barbour and Utev [6] for comparison before the infected population reaches  $K^{2/3}$  in a discrete counterpart of SIR model. We prove that the branching approximation is still valid in the sense of an a.s. equivalence for any submacroscopic level (up to  $\log K$ ). More precisely, for any  $\xi_K \ll K/\log(K)^a$  and  $\eta > 0$ ,

$$\begin{aligned} & \lim_{K \rightarrow +\infty} \mathbb{P} \left( \sup_{t \leq \tau_M^K(\xi_K)} \left| \frac{N_M^K(t)}{Z(t)} - 1 \right| > \eta; Z(\tau_M^K(\xi_K)) > 0 \right) \\ &= \lim_{T \rightarrow \infty} \sup_{K \in \mathbb{N}} \mathbb{P} \left( \sup_{T \leq t \leq \tau_M^K(\xi_K)} \left| \frac{N_M^K(t)}{W e^{r_* t}} - 1 \right| > \eta; W > 0 \right) = 0, \end{aligned}$$

where

$$W = \lim_{t \rightarrow \infty} Z(t) e^{-r_* t} \in [0, \infty) \quad a.s.$$

is an exponential variable with an additional atom in 0 corresponding to the extinction event. The value of  $a$  can be chosen to be 1 or 2 depending on the fact that the equilibrium of the resident population alone is exponentially stable or not, see forthcoming Theorem 2.3 for a precise statement. For such results, we need to control the gap between the two coupled populations (the original process and the branching process). We use that individuals which belong to one of two populations but not to the other in the coupling can be seen as branching structure with an additional immigration whose intensity increases with the value of  $Z$  or  $N_M$ . In this vein, we mention Barbour and Reinert [5] for the description of the epidemic curve which describes the outbreak and the link with initial branching approximation by coupling.

When the population of mutants becomes large but is not yet macroscopic, we prove that the deterministic approximation by the dynamical system is already valid in the following sense:

$$\sup_{t \leq \tau^M(vK)} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| \xrightarrow{K \rightarrow \infty} 0, \text{ in probability,}$$

where  $v > 0$ ,  $1 \ll N_M^K(0) = K X_M^K(0) \ll K$  is the initial number of mutant and  $x_M^K$  is the solution of (1.2) with the corresponding initial condition  $1/K \ll x_M^K(0) = X_M^K(0) \ll 1$ . The boundary being absorbing for the dynamical system but unstable for invasion population makes the problem delicate when controlling the stochastic flow. In the time window  $[0, \tau^M(vK)]$  whose size is of order of magnitude  $\log(K)$ ,  $X^K$  goes from  $X_M^K(0) \ll 1/K$  to  $v$ . We refer to Section 3.2 for a precise statement

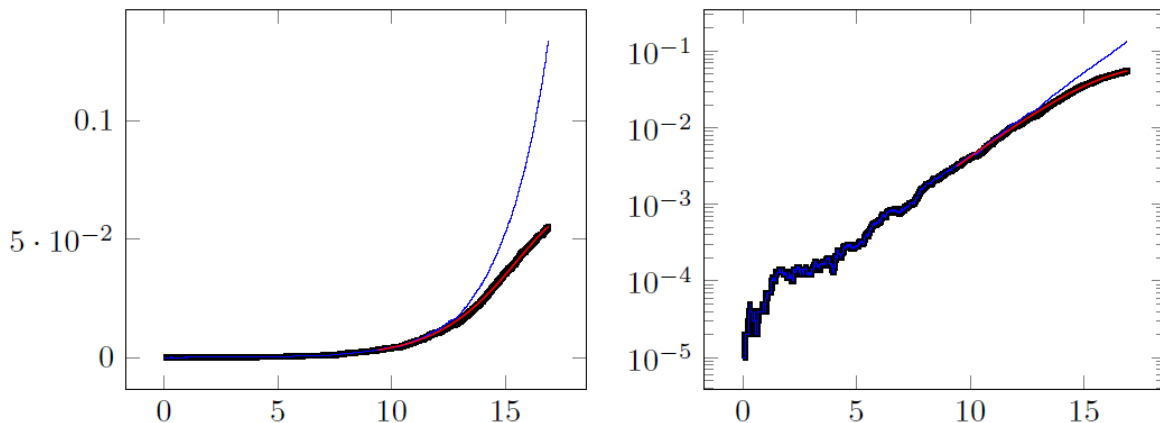


FIG 1. The black curve represents a trajectory of the process  $X_M^K(t) = N_M^K(t)/K$ , the blue curve a trajectory of  $Z(t)/K$ , and the red curve is the deterministic function  $x_M^K(t)$  starting at time  $\tau_M^K(\xi^K)$ . Both figures represent the simulations of the same trajectory in different scales (logarithmic scale at the right picture, and standard scale at the left one). These simulations have been done with the following parameters:  $K = 100000$ ,  $F_R(x_R, x_M) = -1.5x_M$ ,  $F_M(x_R, x_M) = 1.5x_R - 1$ ,  $x_R^* = 1$ ,  $\xi_K = 1/\sqrt{K}$  and  $v = 0.055$ .

which gives the admissible values of  $v$  and quantitative estimates for the convergence. The simulations of Figure 1 illustrates the different approximations corresponding to the SIR example of Section 5.2, first by the branching process and then by the dynamical system.

As a byproduct, we obtain in Theorem 4.1 the following convergence in law of hitting times when  $K$  tends to infinity. For  $(\xi_K)_K$  going to infinity such that  $\xi_K/K \rightarrow v \geq 0$ , we show that on the survival event

$$\tau_M^K(\xi_K) - \frac{\log(\xi_K)}{r_\star} \Longrightarrow \tau(v) - \frac{\log(W^\star)}{r_\star},$$

where the function  $\tau$  is continuous and  $\tau(0) = 0$  is and  $W^\star$  is an exponential random variable with parameter  $r_\star/b_\star$ . Thus, the hitting times of mesoscopic or macroscopic levels are shifted Gumbel laws. We refer to Section 3.1 for the definition  $\tau$  the characterization of admissible values  $[0, v_\star]$  of the level  $v$ . This result is illustrated by the simulations of Figure 2 corresponding to the SIR example of Section 5.2.

**Assumptions.** We consider the following assumptions on the birth and death rates. They ensure in particular that the resident population alone has an equilibrium point (which may be unstable) and mutant population can invade.

**Assumption 1.1.**

- (R) *Regularity:* the functions  $b_R, d_R, b_M, d_M$  are  $C^2(\mathbb{R}_+ \times \mathbb{R}_+)$ .
- (E) *Equilibrium:* there exists  $x_R^* > 0$  such that  $F_R(x_R^*, 0) = 0$  and  $\partial_R F_R(x_R^*, 0) \leq 0$ .
- (I) *Invasion:*  $F_M(x_R^*, 0) > 0$ .

Under Assumption (R), the bi-type birth and death process  $X^K$  is well defined until its explosion time, which may be finite with positive probability. In our framework, the process is studied before hitting times of fixed levels, clearly smaller than the explosion time.

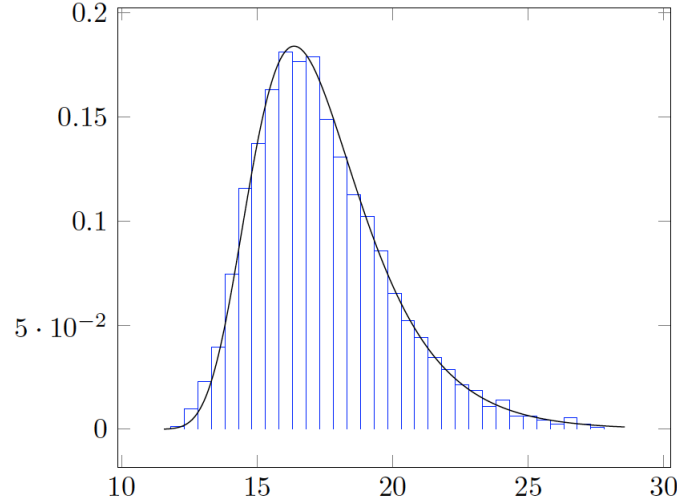


FIG 2. The histogram represents the values of  $\tau_M^K(vK)$ . The simulations have been done with the following parameters:  $K = 100000$ ,  $F_R(x_R, x_M) = -1.5x_M$ ,  $F_M(x_R, x_M) = 1.5x_R - 1$ ,  $x_R^* = 1$  and  $v = 0.055$ . The size of the sample to draw the histogram is 10000. The curve is the density  $f$  of the random variable  $\log(vK)/r_* + \tau(v) - \log(W^*)/r_*$ , where the density of  $-\log(W^*)/r_*$  is equal to  $r_*^2/b_* \exp(-r_*t) \exp(-r_*/b_* e^{-r_*t})$ .

In the eco-evolutionary framework,  $F_M(x^*)$  is called invasion fitness. It quantifies the individual growth rate for a small mutant population in a resident population at the equilibrium.

### Notation.

- We write  $x^* = (x_R^*, 0)$ .
- For  $f : (x_R, x_M) \in \mathbb{R}^2 \mapsto f(x_R, x_M)$ , we denote  $\partial_R f$  (resp.  $\partial_M f$ ) the derivative w.r.t. the first (resp. second) variable.
- if  $\mathcal{N}(dx)$  is a Poisson measure with intensity  $\lambda(dx)$ , we denote  $\tilde{\mathcal{N}}(dx) := \mathcal{N}(dx) - \lambda(dx)$  the compensated measure.
- Throughout the paper,  $C$  denotes any positive constant depending only on the model parameters. The value of  $C$  can change from line to line in an equation. If a constant depends on some (non-model) parameter  $\theta$  we write instead  $C_\theta$ .

## 2. Approximation by branching processes

In this section, we assume that at the initial time, the resident population is close to its equilibrium and a single mutant individual appears :

**Assumption 2.1.** *There exists  $C > 0$  such that for any  $K > 0$ ,*

$$\left| \frac{N_R^K(0)}{K} - x_R^* \right| \leq \frac{C}{K} \quad \text{and} \quad N_M^K(0) = 1 \quad a.s.$$

We introduce the branching process  $(Z(t) : t \geq 0)$  with individual birth rate  $b_* = b_M(x^*)$  and (individual) death rate  $d_* = d_M(x^*)$ , defined on the same probability space as  $N_M^K$  and coupled with

$N_M^K$  in the following way :

$$Z(t) = 1 + \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq b_* Z(s-)\}} \mathcal{N}_M^b(ds, du) - \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq d_* Z(s-)\}} \mathcal{N}_M^d(ds, du),$$

where the Poisson point measures  $\mathcal{N}_M^b$  and  $\mathcal{N}_M^d$  have been introduced in (1.1).

We recall from (1.3) that  $r_* = b_* - d_* > 0$  and the branching process  $Z$  is supercritical. The survival event is positive probability and as  $t$  tends to infinity, the martingale  $W(t) = Z(t) \exp(-r_* t)$  converges a.s. to a finite random variable  $W$  which is positive on the survival event :

$$\{W > 0\} = \{\forall t > 0 : Z(t) > 0\}.$$

The random variable  $W$  is exponentially distributed with an additional atom in 0 :

$$W \stackrel{d}{=} \frac{b_*}{d_*} \delta_0 + \frac{r_*}{b_*} \text{Exp}\left(\frac{r_*}{b_*}\right),$$

where  $\delta_0$  is a Dirac mass in 0 and  $\text{Exp}(\lambda)$  is an exponential random variable with parameter  $\lambda$ . We refer to [1] Chapter 3 or [18]-Theorem 1 for this result.

Let us study the coupling between the processes  $Z$  and  $N_M^K$ . First, when the branching process  $Z$  becomes extinct, so does the mutant process  $N_M$  and we have the following result.

**Proposition 2.2.** *Under Assumptions 1.1 and 2.1,*

$$\lim_{K \rightarrow \infty} \mathbb{P}(\forall t \geq 0 : N_M^K(t) = Z(t) | W = 0) = 1.$$

Indeed, the two processes coincide by coupling on a time window whose size goes to infinity with probability 1 as  $K$  to infinity. Besides,  $Z$  is a.s. absorbed at 0 in finite time on the event  $\{W = 0\}$ . Such arguments are classical for epidemiological models and can be adapted to our context. We refer to [4].

Let us turn our attention to the survival event and state the main result of the section. It compares the mutant process and its branching approximation as long as the process is not too close from macroscopic scales. We prove that the ratio of the two processes converges to 1 in probability, with an explicit speed, before time  $\tau_M^K(\xi_K)$  when mutants reach a certain threshold  $\xi_K$  (see (1.4) for definition).

**Theorem 2.3.** *Under Assumptions 1.1 and 2.1, there exists  $C > 0$  such that the following result holds for any  $\eta > 0$  and  $K \geq 1$ .*

*i) if  $\partial_R F_R(x^*) < 0$ , for any sequence  $\xi$  such that  $\sqrt{K \log K} \ll \xi_K \ll K / \log K$ , we have*

$$\mathbb{P}\left(\sup_{t \leq \tau_M^K(\xi_K)} \left| \frac{N_M^K(t)}{Z(t)} - 1 \right| > \eta ; W > 0\right) \leq C \left( \frac{\sqrt{K \log K}}{\xi_K} + \frac{1}{K^{1/3}} + \frac{1}{\eta^{1/4}} \left( \frac{\xi_K \log K}{K} \right)^{1/10} \right),$$

*ii) if  $\partial_R F_R(x^*) = 0$ , for any sequence  $\xi$  such that  $\log K \ll \xi_K \ll K / (\log K)^2$ , we have*

$$\mathbb{P}\left(\sup_{t \leq \tau_M^K(\xi_K)} \left| \frac{N_M^K(t)}{Z(t)} - 1 \right| > \eta ; W > 0\right) \leq C \left( \sqrt{\frac{\log K}{\xi_K}} + \frac{1}{K^{1/3}} + \frac{1}{\eta^{1/4}} \left( \log K \sqrt{\frac{\xi_K}{K}} \right)^{1/10} \right).$$

We observe that the lower bound on  $\xi_K$  is needed only for the speed of convergence. Thus  $N_M^K/Z - 1$  converges to 1 in probability for any sequence  $\xi_K$  such that  $1 \ll \xi_K \ll K / (\log K)$ . We refer to Section 4 for the asymptotic behavior of  $\tau_M^K(\xi_K)$ .

There are two statements depending on the value of  $\partial_R F_R(x^*)$ . Indeed, the proof is based on a comparison between  $\tau_M^K(\xi_K)$  and the first time where the resident population exits a neighborhood of its equilibrium. The choice of this neighborhood depends  $\partial_R F_R(x^*)$ . If  $\partial_R F_R(x^*) < 0$ , the manifold  $\{x_R = 0\}$  is stable around the fixed point  $(x_R^*, 0)$ . This stability helps for the control of the resident population and guarantees that the distance of the resident population process to its equilibrium remains smaller than  $\xi_K$  (up to some well chosen constant). If  $\partial_R F_R(x^*) = 0$ , exponential (local) stability is lost and we can (only) prove that the distance is smaller than  $\sqrt{K\xi_K}$ . We refer to Lemma 2.8 for details.

The proof of Theorem 2.3 will be obtained in several steps that we develop below. We first consider a classical coupling construction between  $N_M^K$  and  $Z$ . Let us associate a color to individuals: the black ones realize the coupling and the red ones belong to one population but not to the other. The two subpopulations will be indexed respectively by  $n$  and  $r$ .

We introduce

$$m^K(s) = \min(N_{M,n}^K(s)b_M(X^K(s)), Z_n(s)b_\star), \quad p^K(s) = \max(N_{M,n}^K(s)d_M(X^K(s)), Z_n(s)d_\star)$$

where  $(N_{M,n}^K, N_{M,r}^K)$  and  $(Z_n, Z_r)$  are defined as follows:

$$\begin{aligned} N_{M,n}^K(t) &= 1 + \int_0^t \int_{\mathbb{R}_+} 1_{u \leq m^K(s-)} \mathcal{N}_M^b(ds, du) - \int_0^t \int_{\mathbb{R}_+} 1_{u \leq p^K(s-)} \mathcal{N}_M^d(ds, du) \\ N_{M,r}^K(t) &= \int_0^t \int_{\mathbb{R}_+} 1_{m^K(s-) < u \leq N_{M,n}^K(s-)b_M(X^K(s-))} \mathcal{N}_M^b(ds, du) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} 1_{N_{M,n}^K(s-)d_M(X^K(s-)) < u \leq p^K(s-)} \mathcal{N}_M^d(ds, du) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} 1_{0 \leq u - N_{M,n}^K(s-)b_M(X^K(s-)) \leq N_{M,r}^K(s-)b_M(X^K(s-))} \mathcal{N}_M^b(ds, du) \\ &\quad - \int_0^t \int_{\mathbb{R}_+} 1_{0 \leq u - p^K(s-) \leq N_{M,r}^K(s-)d_M(X^K(s-))} \mathcal{N}_M^d(ds, du) \end{aligned}$$

and in a similar way,

$$\begin{aligned} Z_n^K(t) &= 1 + \int_0^t \int_{\mathbb{R}_+} 1_{u \leq m^K(s-)} \mathcal{N}_M^b(ds, du) - \int_0^t \int_{\mathbb{R}_+} 1_{u \leq p^K(s-)} \mathcal{N}_M^d(ds, du) \\ Z_r^K(t) &= \int_0^t \int_{\mathbb{R}_+} 1_{m^K(s-) < u \leq Z_n^K(s-)b_\star} \mathcal{N}_M^b(ds, du) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} 1_{Z_n^K(s-)d_\star < u \leq p^K(s-)} \mathcal{N}_M^d(ds, du) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} 1_{0 \leq u - Z_n^K(s-)b_\star \leq Z_r^K(s-)b_\star} \mathcal{N}_M^b(ds, du) \\ &\quad - \int_0^t \int_{\mathbb{R}_+} 1_{0 \leq u - p^K(s-) \leq Z_r^K(s-)d_\star} \mathcal{N}_M^d(ds, du). \end{aligned}$$

We can easily check that

**Lemma 2.4.** *The following equalities of processes hold, almost surely:*

- i)  $N_M^K = N_{M,n}^K + N_{M,r}^K$  and  $Z = Z_n^K + Z_r^K$ .
- ii)  $N_{M,n}^K = Z_n^K$ .



*Proof.* Item i) results from the fact that  $N_{M,n}^K + N_{M,r}^K$  and  $N_M^K$  are solutions of the same stochastic differential equation for which pathwise uniqueness holds. A similar argument also holds for  $Z$ . Item ii) results from the construction of  $N_{M,n}^K$  and  $Z_n^K$ .  $\square$

Let us introduce the time when the gap to the equilibrium for resident population goes beyond some level  $n \geq 0$ :

$$\tau_R^K(n) = \inf \{t > 0 : |N_R^K(t) - x_R^* K| > n\} = \inf \{t > 0 : |X_R^K(t) - x_R^*| > n/K\}$$

We introduce the stopping time

$$T^K(m, n) = \tau_M^K(m) \wedge \tau_R^K(n) \wedge \frac{4 \log K}{3r_*}. \quad (2.1)$$

**Remark 2.5.** *In our proof, we need to bound the stopping time  $T^K(m, n)$  by some deterministic value. For the need of our proof, it is possible to consider any time of the form  $\lambda(\log K)/r_*$  (with  $\lambda > 1$ ). We choose  $\lambda := 4/3$  in definition (2.1) to optimize the convergence speed of Theorem 2.3 (this particular choice only matters in the proof of Lemma 2.7).*

The residual processes (red individuals) will be compared to an inhomogeneous branching process and proved to be negligible until  $T^K(m, n)$ , with  $m$  and  $n$  suitably chosen.

**Lemma 2.6.** *Under Assumptions 1.1 and 2.1,*

*i) For any  $L > 0$ , there exists  $C_L > 0$  such that for any  $\xi_K \ll K/\log K$ ,*

$$\mathbb{E} \left[ \sup_{t \leq T^K(\xi_K, L\xi_K)} (Z_r^K(t) + N_{M,r}^K(t)) e^{-r_* t} \right] \leq C_L \sqrt{\log K} \frac{\xi_K}{K}.$$

*ii) For any  $\xi_K \ll K/(\log K)^2$ ,*

$$\mathbb{E} \left[ \sup_{t \leq T^K(\xi_K, \sqrt{K\xi_K})} (Z_r^K(t) + N_{M,r}^K(t)) e^{-r_* t} \right] \leq C \left( \log K \sqrt{\frac{\xi_K}{K}} \right)^{1/2}.$$

*Proof.* In case i), we consider  $\sqrt{K \log K} \ll \xi_K \ll K/\log K$ . Let us denote  $T^K = T^K(\xi_K, L\xi_K)$  for simplicity. Since  $b_M$  and  $d_M$  are locally Lipschitz continuous, then for  $s \leq T^K$ , almost surely,

$$Z_n^K(s) |b_M(X^K(s)) - b_*| + Z_n^K(s) |d_M(X^K(s)) - d_*| \leq C Z_n^K(s) \frac{\xi_K}{K}, \quad (2.2)$$

and then

$$Z_n^K(s) b_* - m^K(s) + p^K(s) - Z_n^K(s) d_* \leq C Z_n^K(s) \frac{\xi_K}{K},$$

for  $C$  a positive constant number. Therefore, the process  $Z_r^K$  can be stochastically dominated on the time window  $[0, T^K]$  by an branching process with inhomogeneous immigration  $Y$ , solution of

$$\begin{aligned} Y(t) &= \int_0^t 1_{u \leq C Z_n^K(s-) \frac{\xi_K}{K}} \mathcal{N}^I(ds) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} 1_{0 \leq u \leq Y(s-) b_*} \mathcal{N}_M^b(ds, du) - \int_0^t \int_{\mathbb{R}_+} 1_{0 \leq u \leq Y(s-) d_*} \mathcal{N}_M^d(ds, du), \end{aligned}$$

where  $\mathcal{N}^I$  is a Poisson point measure with Lebesgue intensity on  $\mathbb{R}_+$ , independent of  $\mathcal{N}_M^b$  and  $\mathcal{N}_M^d$ . We can easily note that

$$\mathbb{E} [Z_n^K(t)] \leq \mathbb{E} [Z(t)] = e^{r_* t}$$

and straightforward computation leads to

$$\mathbb{E} [Y(t)] \leq C \frac{\xi_K}{K} e^{2r_* t}.$$

Let  $M(t)$  be the martingale part of the semimartingale  $Y(t)$ . Then

$$\langle M \rangle(t) = \int_0^t \left( CZ_n^K(s) \frac{\xi_K}{K} + Y(s)(b_* + d_*) \right) ds$$

and Doob's and Cauchy Schwarz inequality yields

$$\mathbb{E} \left( \sup_{t \leq T^K} \left| \int_0^t e^{-r_* s} dM(s) \right| \right) \leq C \mathbb{E} \left( \int_0^{T^K} e^{-2r_* s} d\langle M \rangle(s) \right)^{1/2} \leq C \sqrt{\frac{2T^K \xi_K}{K}}.$$

We compute

$$\begin{aligned} Y(t)e^{-r_* t} &= \int_0^t e^{-r_* s} dM(s) + \int_0^t e^{-r_* s} \left( CZ_n^K(s) \frac{\xi_K}{K} + Y(s)(b_* - d_*) - Y(s)(b_* - d_*) \right) ds \\ &= \int_0^t e^{-r_* s} dM(s) + \int_0^t e^{-r_* s} CZ_n^K(s) \frac{\xi_K}{K} ds. \end{aligned}$$

Combining the previous estimates gives

$$\mathbb{E} \left( \sup_{t \leq T^K} Y(t)e^{-r_* t} \right) \leq C \sqrt{\frac{2T^K \xi_K}{K}} + \frac{CT^K \xi_K}{K}.$$

By (2.1) and the domination of  $Z_r^K$  by  $Y$ , we obtain

$$\mathbb{E} \left( \sup_{t \leq T^K} Z_r^K(t)e^{-r_* t} \right) \leq C \sqrt{\log K \frac{\xi_K}{K}},$$

for  $C > 0$  suitably chosen. The proof follows if  $\log K \xi_K \ll K$ .

The assertion *i*) for  $N_{M,r}^K$  can be proved in a similar way. Indeed the individual birth and death rates are respectively upper-bounded by  $b_* + C\xi_K/K$  and lower-bounded by  $d_* + C\xi_K/K$  for  $t \leq T^K$ . So  $N_{M,r}^K$  is dominated on the time window  $[0, T^K]$  by the process  $Y$  defined by

$$\begin{aligned} Y(t) &= \int_0^t \mathbf{1}_{u \leq CZ(s-)\xi_K/K} \mathcal{N}^I(ds) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{0 \leq u \leq Y(s-)(b_* + C\xi_K/K)} \mathcal{N}_M^b(ds, du) - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{0 \leq u \leq Y(s-)(d_* - C\xi_K/K)} \mathcal{N}_M^d(ds, du). \end{aligned}$$

The conclusion follows as above by considering  $Y(t) \exp(-(r_* + 2C\xi_K/K)t)$  since, for  $t \leq T^K$ ,  $(2C\xi_K/K)t \leq 1$  for  $K$  large enough. recalling that  $\xi_K \ll K/\log K$ .

For the second case *ii*), we consider  $\xi_K \ll K/(\log K)^2$ . The difference with the first case is that (2.2) becomes  $T^K = T^K(\xi_K, \sqrt{K\xi_K})$  and for  $s \leq T^K$ , almost surely,

$$Z_n^K(s) |b_M(X^K(s)) - b_*| + Z_n^K(s) |d_M(X^K(s)) - d_*| \leq CZ_n^K(s) \sqrt{\frac{\xi_K}{K}},$$

since  $\xi_K/K \ll \sqrt{\xi_K/K} \ll 1$ . Then, similar computation as in Step 1 gives

$$\mathbb{E} \left( \sup_{t \leq T^K} Z_r^K(t) e^{-r_* t} \right) \leq C \left( \log K \sqrt{\frac{\xi_K}{K}} \right)^{1/2}.$$

Proceeding similarly for  $N_{M,r}^K$  ends the proof.  $\square$

**Lemma 2.7.** *Under Assumptions 1.1 and 2.1, we have that for any  $\eta > 0$ ,*

*i) For any  $L > 0$ , there exists  $C_L > 0$  such that for any  $\sqrt{K \log K} \ll \xi_K \ll K/(\log K)$ ,*

$$\mathbb{P} \left( \sup_{t \leq T^K(\xi_K, L\xi_K)} \frac{|N_M^K(t) - Z(t)|}{Z(t)} > \eta ; W > 0 \right) \leq \frac{C_L}{\eta^{1/4}} \left( \frac{\xi_K \log K}{K} \right)^{1/10}. \quad (2.3)$$

*ii) For  $\log K \ll \xi_K \ll K/(\log K)^2$ ,*

$$\mathbb{P} \left( \sup_{t \leq T^K(\xi_K, \sqrt{K\xi_K})} \frac{|N_M^K(t) - Z(t)|}{Z(t)} > \eta ; W > 0 \right) \leq \frac{C}{\eta^{1/4}} \left( \log K \sqrt{\frac{\xi_K}{K}} \right)^{1/10}. \quad (2.4)$$

*Proof.* Let us prove *i)* and focus on the survival event  $\{W > 0\}$  where the process  $Z$  stays positive. On this event, using Lemma 2.4, we write

$$\frac{|N_M^K(t) - Z(t)|}{Z(t)} \leq \frac{|N_{M,n}^K(t) - Z_n^K(t)|}{Z(t)} + \frac{|N_{M,r}^K(t) - Z_r^K(t)|}{Z(t)},$$

and the first term is null. Moreover

$$\frac{|N_{M,r}^K(t) - Z_r^K(t)|}{Z(t)} = \frac{|N_{M,r}^K(t) - Z_r^K(t)| e^{-r_* t}}{W(t)},$$

where  $W(t)$  is the classical martingale associated with  $Z$ :  $W(t) = Z(t)e^{-r_* t}$ .

Using Lemma 2.6 *i)* and the fact that  $W(t)$  converges to  $W$  a.s., we deduce that for  $t \leq T^K(\xi_K, L\xi_K)$ ,  $|N_{M,r}^K(t) - Z_r^K(t)|/Z(t)$  tends to 0 in probability on the event  $W > 0$ , as  $K$  tends to infinity.

Let us now obtain an explicit convergence speed. For convenience we define

$$T^K = T^K(\xi_K, L\xi_K), \quad U^K = \sup_{t \leq T^K} Z_r(t) e^{-r_* t} + \sup_{t \leq T^K} N_{M,r}^K(t) e^{-r_* t}.$$

The previous computations ensure that

$$\frac{|N_M^K(t) - Z(t)|}{Z(t)} \leq \frac{U^K}{\inf_{t \geq 0} W(t)}$$

a.s. on the event  $W > 0$ . Moreover, writing  $\varepsilon_K = \xi_K \log K/K$ , for any  $\eta > 0$

$$\mathbb{P} \left( U^K \geq \eta \inf_{t \geq 0} W(t) ; W > 0 \right) \leq \mathbb{P} \left( U^K \geq \varepsilon_K^{2/5} \right) + \mathbb{P} \left( \inf_{t \geq 0} W(t) \leq \frac{1}{\eta} \varepsilon_K^{2/5} ; W > 0 \right).$$

By Markov's inequality and Lemma 2.6,

$$\mathbb{P} \left( U^K \geq \varepsilon_K^{2/5} \right) \leq C_L \varepsilon_K^{1/2-2/5} = C_L \varepsilon_K^{1/10}.$$

On the other hand, by Lemma 6.2,

$$\mathbb{P}\left(\inf_{t \geq 0} W(t) \leq \frac{1}{\eta} \varepsilon_K^{2/5}; W > 0\right) \leq C \frac{1}{\eta^{1/4}} \varepsilon_K^{2/5 \times 1/4} = C \frac{1}{\eta^{1/4}} \varepsilon_K^{1/10}.$$

This proves the result case *i*). Case *ii*) is proven similarly.  $\square$

Let us now prove that the mutant population will reach their threshold before the resident population. We quantify the speed at which the probability of the event  $\tau_R^K \leq T^K$  goes to 0 as  $K$  goes to infinity:

**Lemma 2.8.** *Under Assumption 1.1,*

*i) If  $\partial_R F_R(x^*) < 0$ , we set*

$$L := \frac{7|\partial_M G_R(x^*, 0)|(1 + |\partial_R F_R(x^*, 0)|)}{1 - e^{\partial_R F_R(x^*)}},$$

*then for any  $(\xi_K)_K$  such that  $\sqrt{K \log K} \ll \xi_K \ll K/\log K$ ,*

$$\mathbb{P}\left(\tau_M^K(\xi_K) \geq \tau_R^K(L\xi_K) \wedge \frac{4 \log K}{3r_\star}; W > 0\right) \leq C \left( \frac{\sqrt{K \log K}}{\xi_K} + \frac{1}{K^{1/3}} + \left(\frac{\xi_K \log K}{K}\right)^{1/10} \right).$$

*ii) If  $\partial_R F_R(x^*) = 0$ , then for any  $(\xi_K)_K$  such that  $\log K \ll \xi_K \ll K/(\log K)^2$ ,*

$$\mathbb{P}\left(\tau_M^K(\xi_K) \geq \tau_R^K(\sqrt{K\xi_K}) \wedge \frac{4 \log K}{3r_\star}; W > 0\right) \leq C \left( \sqrt{\frac{\log K}{\xi_K}} + \frac{1}{K^{1/3}} + \left(\log K \sqrt{\frac{\xi_K}{K}}\right)^{1/10} \right).$$

This result will guarantee that the probability of the event  $\{\tau_M^K = T^K\}$  goes to one as  $K$  goes to infinity.

*Proof of case i).* The first STEP consists in controlling the resident population and we prove that

$$\mathbb{P}\left(\sup_{t \leq T^K} |X_R^K(t) - x_R^*| \geq L\xi_K/K\right) \leq C \frac{\sqrt{K \ln K}}{\xi_K}, \quad (2.5)$$

with

$$T^K := T^K(\xi_K, L\xi_K) := \tau_M^K(\xi_K) \wedge \tau_R^K(L\xi_K) \wedge \frac{4 \log K}{3r_\star}.$$

For that purpose, we notice that, for any  $t \leq T^K$  and  $K$  large enough,

$$X_M^K(t) \leq \frac{\xi_K}{K} \leq 1 \text{ and } X_R^K(t) \leq x_R^* + L \frac{\xi_K}{K} \leq x_R^* + 1.$$

Hence, the processes  $X_M^K$  and  $X_R^K$  are bounded up to time  $T^K$ . We write

$$\begin{aligned} X_R^K(t) &= X_R^K(0) + \frac{1}{K} \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{z \leq N_R^K(s-) b_R(X^K(s-))\}} \tilde{\mathcal{N}}_R^b(ds, dz) \\ &\quad - \frac{1}{K} \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{z \leq N_R^K(s-) d_R(X^K(s-))\}} \tilde{\mathcal{N}}_R^d(ds, dz) \\ &\quad + \int_0^t G_R(X^K(s-)) ds \\ &=: X_R^K(0) + M^K(t) + \int_0^t G_R(X^K(s-)) ds, \end{aligned}$$

where  $\tilde{\mathcal{N}}_R^b$  and  $\tilde{\mathcal{N}}_R^d$  are the compensated Poisson martingale measures of  $\mathcal{N}_R^b$  and  $\mathcal{N}_R^d$ , and

$$G_R(x_R, x_M) := x_R(b_R - d_R)(x_R, x_M) = x_R F_R(x_R, x_M).$$

Note that

$$\langle M^K \rangle(t) = \frac{1}{K} \int_0^t X_R^K(s) \left( b_R(X^K(s-)) + d_R(X^K(s-)) \right) ds. \quad (2.6)$$

Observing that  $G_R(x^*) = 0$  by definition of  $x^*$ , we rewrite the dynamics above as

$$X_R^K(t) = X_R^K(0) + M^K(t) + H_1^K(t) + H_2^K(t) + \int_0^t (X_R^K(s) - x_R^*) \partial_R G_R(x^*) ds,$$

where

$$H_1^K(t) = \int_0^t X_M^K(s) \partial_M G_R(x^*) ds$$

and

$$H_2^K(t) = \int_0^t \left( G_R(X^K(s-)) - G_R(x^*) - (X_R^K(s) - x_R^*) \partial_R G_R(x^*) - X_M^K(s) \partial_M G_R(x^*) \right) ds.$$

Denoting

$$Y_R^K(t) := X_R^K(t) - X_R^K(0),$$

we have

$$Y_R^K(t) = M^K(t) + H_1^K(t) + H_2^K(t) + \int_0^t Y_R^K(s) \partial_R G_R(x^*) ds. \quad (2.7)$$

Recall that  $\partial_R F_R(x^*) < 0$  and so  $\partial_R G_R(x^*) < 0$ . We define

$$H^K := M^K + H_1^K + H_2^K.$$

Thanks to Lemma 6.3, for  $K$  large enough, for any  $t \leq T^K$ ,

$$\sup_{0 \leq s \leq t} |Y_R^K(s)| \leq \Gamma \sup_{0 \leq s \leq t} |H^K(s) - H^K(\lfloor s \rfloor)|,$$

with

$$\Gamma := (1 + |\partial_R F_R(x^*)|) \frac{1}{1 - e^{\partial_R F_R(x^*)}}.$$

Then, for  $K$  large enough,

$$\begin{aligned} & \mathbb{P} \left( \sup_{t \leq T^K} |X_R^K(t) - x_R^*| > L\xi_K/K \right) \\ & \leq \mathbb{P} \left( \sup_{t \leq T^K} |Y_R^K(t)| > \frac{1}{2} L\xi_K/K \right) \\ & \leq \mathbb{P} \left( \sup_{t \leq T^K} |H^K(t) - H^K(\lfloor t \rfloor)| > \frac{L}{2\Gamma} \xi_K/K \right) \\ & \leq \mathbb{P} \left( \sup_{t \leq T^K} |M^K(t) - M^K(\lfloor t \rfloor)| > \frac{L}{6\Gamma} \xi_K/K \right) \\ & \quad + \sum_{i=1}^2 \mathbb{P} \left( \sup_{t \leq T^K} |H_i^K(t) - H_i^K(\lfloor t \rfloor)| > \frac{L}{6\Gamma} \xi_K/K \right). \end{aligned}$$

Using (2.6) and the definition of  $T_K$  for an upper bound  $X_R^K$  on  $[0, T^K]$ , we get  $\mathbb{E} [\langle M^K \rangle_{T^K}] \leq C(\log K)/K$ . By Markov's and Burkholder-Davis-Gundy's inequalities, we then obtain

$$\begin{aligned} \mathbb{P} \left( \sup_{t \leq T^K} |M^K(t) - M^K(\lfloor t \rfloor)| > \frac{L}{6\Gamma} \xi_K / K \right) &\leq C \frac{K}{\xi_K} \mathbb{E} \left[ \sup_{t \leq T^K} |M^K(t)| \right] \\ &\leq C \frac{K}{\xi_K} \mathbb{E} [\langle M^K \rangle_{T^K}]^{1/2} \\ &\leq C \frac{K}{\xi_K} \left( \frac{\log K}{K} \right)^{1/2} \leq C \frac{\sqrt{K \log K}}{\xi_K}. \end{aligned}$$

Besides, almost surely for  $K$  large enough,

$$\sup_{t \leq T^K} |H_1^K(t) - H_1^K(\lfloor t \rfloor)| \leq \frac{\xi_K}{K} |\partial_M G_R(x^*)| < \frac{\xi_K}{K} \cdot \frac{L}{6\Gamma},$$

since  $L := 7|\partial_M G_R(x^*)|\Gamma$ , and thus for  $K$  large enough

$$\sup_{t \leq T^K} |H_2^K(t) - H_2^K(\lfloor t \rfloor)| \leq C \left( \frac{\xi_K}{K} \right)^2 < \frac{L}{6\Gamma} \cdot \frac{\xi_K}{K}.$$

which proves (2.5) and ends the first STEP.

We turn to STEP 2 and prove that

$$\begin{aligned} \mathbb{P} \left( \tau_M^K(\xi_K) \wedge \tau_R^K(L\xi_K) \geq \frac{4 \log K}{3r_\star}; W > 0 \right) \\ \leq CK^{-1/3} + \frac{C}{\eta^{1/4}} \left( \log K \sqrt{\frac{\xi_K}{K}} \right)^{1/10}. \end{aligned} \quad (2.8)$$

For convenience, we write  $S^K := \frac{4}{3r_\star} \log K$  and we recall that  $W(t) := Z(t)e^{-r_\star t}$ . Let us notice that, a.s. the event  $\{W > 0\}$ ,

$$\begin{aligned} &\{ \tau_M^K(\xi_K) \wedge \tau_R^K(L\xi_K) \geq S^K \} \\ &\subseteq \left( \{ N_M^K(S^K) \leq \xi_K + 1 \leq K/2 \} \cap \left\{ \frac{Z(S^K)}{N_M^K(S^K)} \leq 2 \right\} \right) \cup \left\{ \sup_{t \leq T^K} \left| \frac{N_M^K(t)}{Z(t)} - 1 \right| > \frac{1}{2} \right\} \\ &\subseteq \{ Z(S^K) \leq K \} \cup \left\{ \sup_{t \leq T^K} \left| \frac{N_M^K(t)}{Z(t)} - 1 \right| > \frac{1}{2} \right\}. \end{aligned}$$

Whence

$$\mathbb{P} \left( \tau_M^K(\xi_K) \wedge \tau_R^K(L\xi_K) \geq S^K; W > 0 \right) \leq \mathbb{P} \left( W(S^K) \leq K^{-1/3}; W > 0 \right) + \mathbb{P} \left( \sup_{t \leq T^K} \left| \frac{N_M^K(t)}{Z(t)} - 1 \right| > \frac{1}{2}; W > 0 \right).$$

The first term involves an estimate on the infimum of a single type branching process which we state and prove in Lemma 6.1 in forthcoming Appendix. The second term of the sum in the RHS above has been controlled in Lemma 2.7. Combining these two estimates yields (2.8).

We can now conclude the case *i*) combining the two STEPs. Using the fact that, for any  $x, y, z \in \mathbb{R}$ ,  $(x \geq y \wedge z) \iff (x \wedge y \geq z \text{ or } x \wedge z \geq y)$ , we can write

$$\begin{aligned} \mathbb{P} \left( \tau_M^K(\xi_K) \geq \tau_R^K(L\xi_K) \wedge \frac{4 \log K}{3r_\star}; W > 0 \right) \\ \leq \mathbb{P} \left( \tau_M^K(\xi_K) \wedge \tau_R^K(L\xi_K) \geq \frac{4 \log K}{3r_\star}; W > 0 \right) + \mathbb{P} \left( \tau_M^K(\xi_K) \wedge \frac{4 \log K}{3r_\star} \geq \tau_R^K(L\xi_K) \right). \end{aligned}$$

We use (2.8) to estimate the first term in the right hand side. We can use (2.5) for the second one since

$$\left\{ \tau_R^K(L\xi_K) \leq \tau_M^K(\xi_K) \wedge \frac{4 \log K}{3r_\star} \right\} \subset \left\{ \sup_{t \leq T^K} |N_R^K(t) - x_R^\star K| \geq L\xi_K \right\}.$$

It ends the proof.  $\square$

*Proof of the case ii).* We prove a similar result under the hypothesis  $\partial_R F_R(x^\star) = 0$ . Now we define

$$T^K := T^K(\xi_K, \sqrt{\xi_K K}),$$

and we want to prove that

$$\mathbb{P} \left( \sup_{t \leq T^K} |X_R^K(t) - x_R^\star| > \sqrt{\xi_K/K} \right) \leq C \sqrt{\frac{\log K}{\xi_K}}.$$

Here (2.7) gives trivially

$$\sup_{s \leq t} |Y_R^K(s)| \leq \sup_{s \leq t} |M^K(s)| + \sup_{s \leq t} |H_1^K(s)| + \sup_{s \leq t} |H_2^K(s)|,$$

whence,

$$\begin{aligned} & \mathbb{P} \left( \sup_{t \leq T^K} |X_R^K(t) - x_R^\star| > \sqrt{\xi_K/K} \right) \\ & \leq \mathbb{P} \left( \sup_{t \leq T^K} |M^K(t)| > C\sqrt{\xi_K/K} \right) + \sum_{i=1}^2 \mathbb{P} \left( \sup_{t \leq T^K} |H_i^K(t)| > C\sqrt{\xi_K/K} \right). \end{aligned}$$

We have the following estimates

$$\begin{aligned} \mathbb{P} \left( \sup_{t \leq T^K} |M^K(t)| > C\sqrt{\xi_K/K} \right) & \leq C \sqrt{\frac{K}{\xi_K}} \cdot \sqrt{\frac{\log K}{K}} = C \sqrt{\frac{\log K}{\xi_K}}, \\ \sup_{t \leq T^K} |H_1^K(t)| & \leq C \frac{\xi_K}{K} T^K \leq C \frac{\xi_K}{K} \log K \ll \sqrt{\frac{\xi_K}{K}}, \\ \sup_{t \leq T^K} |H_2^K(t)| & \leq C \frac{\xi_K}{K} T^K \leq C \frac{\xi_K}{K} \log K \ll \sqrt{\frac{\xi_K}{K}}, \end{aligned}$$

which proves that, for  $K$  large enough,

$$\mathbb{P} \left( \sup_{t \leq T^K} |X_R^K(t) - x_R^\star| > \sqrt{\xi_K/K} \right) \leq C \sqrt{\frac{\log K}{\xi_K}}.$$

It ends the proof.  $\square$

Theorem 2.3 becomes now a consequence of the two previous lemmas.

*Proof of Theorem 2.3.* We use

$$\begin{aligned} \mathbb{P} \left( \sup_{t \leq \tau_M^K(\xi_K)} \frac{|N_M^K(t) - Z(t)|}{Z(t)} > \eta; W > 0 \right) & \leq \mathbb{P} \left( \tau_M^K(\xi_K) > \tau_R^K \wedge \frac{4}{3r_\star} \log K \right) \\ & \quad + \mathbb{P} \left( \sup_{t \leq T^K} \frac{|N_M^K(t) - Z(t)|}{Z(t)} > \eta; W > 0 \right). \end{aligned}$$

The two terms of the right hand side can be controlled using respectively Lemmas 2.8 and 2.7. This completes the proof.  $\square$

### 3. Approximation by dynamical systems

The goal of this section is to approximate the process  $X^K = (X_R^K, X_M^K)$  defined in Section 1 by the dynamical system defined by (1.2) under degenerate initial conditions. Usually such an approximation is proved when the initial conditions are of order of magnitude one (cf for example [19]). Here, we focus on a small deterministic initial density of mutants,  $1/K \ll x_M = N_M^K(0)/K \ll 1$  and an initial density of residents  $x_R = N_R^K(0)/K$  close to the equilibrium value  $x_R^*$ . The initial mutant density may be very small regarding  $K$  but we still enjoy a large number of mutants allowing to compare the stochastic process to its expected deterministic behavior. Starting from such sub-macroscopic level of mutants, the time for them to reach macroscopic levels will be of order  $\log K$ . We will compare the density process and its deterministic approximation on this time scale of invasion. They are both small at the beginning and we will show that their ratio remains close to 1 on the full time window allowing to reach macroscopic density of mutants.

This section emphasizes the role of (small) initial conditions and we introduce the flow notation for convenience. The process  $X^K(x, \cdot)$  started from  $x$  is defined for any  $K$  and time  $t$  by

$$\begin{aligned} X_{\bullet}^K(x, t) = & x_{\bullet} + \frac{1}{K} \int_{[0, t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq K X_{\bullet}^K(x, s-) b_{\bullet}(X^K(x, s-))\}} \mathcal{N}_{\bullet}^b(ds, su) \\ & - \frac{1}{K} \int_{[0, t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq K X_{\bullet}^K(x, s-) d_{\bullet}(X^K(x, s-))\}} \mathcal{N}_{\bullet}^d(ds, du). \end{aligned} \quad (3.1)$$

for  $\bullet \in \{R, M\}$ . The flow  $\phi(x, t) = (\phi_R(x, t), \phi_M(x, t))$  is the solution of (1.2) starting from initial value  $x$ , i.e. the unique solution of

$$\phi(x, 0) = x, \quad \frac{\partial}{\partial t} \phi(x, t) = G(\phi(x, t)),$$

where we recall that

$$G_{\bullet}(x) = x_{\bullet}(b_{\bullet}(x) - d_{\bullet}(x)).$$

In this section we assume that the flow  $\phi$  satisfies a boundedness assumption:

**Assumption 3.1.** *There exists a compact domain  $\mathcal{D}$  of  $\mathbb{R}_+^2$  such that*

$$\exists \varepsilon > 0, (x_R^* - \varepsilon, x_R^* + \varepsilon) \times [0, \varepsilon) \subset \mathcal{D} \quad \text{and} \quad \{\phi(x, t) : x \in \mathcal{D}, t \geq 0\} \text{ is bounded.}$$

#### 3.1. Properties of the dynamical system and hitting times

Let us consider the hitting time for the mutant population starting from  $x$

$$\tau_v^M(x) = \inf\{t \geq 0 : \phi_M(x, t) = v\},$$

with by convention  $\inf \emptyset = \infty$  and the hitting time is then infinite. The following result quantifies the hitting time  $\tau_v^M(x)$  in function of the two variables  $x$  and  $v$ , starting close to the equilibrium  $x^* = (x_R^*, 0)$ . Let us define, for  $\eta > 0$ ,

$$\mathcal{C}(x, \eta) = \sup\{\phi_M(x, t) : t \leq \bar{\tau}_{\eta}(x)\}$$

where

$$\bar{\tau}_{\eta}(x) = \inf\{t \geq 0 : F_M(\phi(x, t)) \leq \eta\}.$$



We set

$$v_\eta = \lim_{r \downarrow 0} \inf_{x \in B_r(x^*) \cap (\mathbb{R}_+^*)^2} \mathcal{C}(x, \eta).$$

Observing that it decreasing with  $\eta$ , we let  $\eta$  go to zero and define  $v_\star$  as the upper-bound of the values of mutants which can be reached starting the neighborhood of the equilibrium:

$$v_\star := \lim_{\eta \downarrow 0} v_\eta.$$

The next result guarantees that  $v_\star$  is positive and finite. It also quantifies the hitting times  $\tau_v^M(x)$  when  $x$  goes close to  $x^\star$  and  $v$  is smaller than  $v_\star$ .

**Proposition 3.2.** *Under Assumptions 1.1 and 3.1, the value  $v_\star$  is positive and finite.*

*Besides there exists a continuous increasing function  $\tau : [0, v_\star) \rightarrow \mathbb{R}_+$  such that  $\tau(0) = 0$  and for any  $\bar{v} \in (0, v_\star)$ ,*

$$\lim_{x \rightarrow x^\star} \sup_{v \in (x_M, \bar{v}] } \left| \tau_v^M(x) - \frac{1}{r_\star} \log(v/x_M) - \tau(v) \right| = 0,$$

where the limit is taken for  $x \in (\mathbb{R}_+^*)^2$  if  $\partial_R F_R(x^\star) < 0$ , and for

$$x \in D(\eta) = \{ (x_R, x_M) \in (\mathbb{R}_+^*)^2 : |x_R - x_R^\star| \log(1/x_M) \leq \eta \}$$

for some  $\eta > 0$  if  $\partial_R F_R(x^\star) = 0$ .

The proof of this proposition necessitates the three next lemmas, which guarantee that the resident process to remain close to its equilibrium value as long as the mutant process remains small. The proofs of the two first ones are a bit technical and postponed in Appendix.

**Lemma 3.3.** *i) If  $\partial_R F_R(x^\star) < 0$ , then there exists  $0 < u_0 < 1$  such that for any  $u \leq u_0$  and  $x = (x_R, x_M) \in B(x^\star, u)$ ,*

$$\sup_{t \leq \tau_u^M(x)} |\phi_R(x, t) - x_R^\star| \leq C\sqrt{u}.$$

*ii) If  $\partial_R F_R(x^\star) = 0$ , then there exist some  $0 < u_0 < 1$  and  $\eta_0 > 0$  such that, for all  $u \leq u_0$  and all  $x \in \mathcal{B}(x^\star, u) \cap D(\eta_0)$ ,*

$$\sup_{t \leq \tau_u^M(x)} |\phi_R(x, t) - x_R^\star| \leq C\sqrt{u}.$$

In both cases with the same notation as previously,

$$\tau_u^M(x) \leq \frac{2}{r_\star} \log(1/x_M).$$

We also need an estimate of the cumulated gap :

**Lemma 3.4.** *i) Assume that  $\partial_R F_R(x^\star) < 0$ . Then there exist some  $0 < u_0 < 1$  and  $\eta_0 > 0$  such that for all  $u \leq u_0, \eta \leq \eta_0$  and  $x \in \mathcal{B}(x^\star, u) \cap D(\eta)$ ,*

$$\int_0^{\tau_u^M(x)} \sup_{s \leq t} |\phi_R(x, s) - x_R^\star| dt \leq C(u + \eta).$$

*ii) Assume that  $\partial_R F_R(x^\star) = 0$ . Then there exist some  $u_0 \in (0, 1)$  and  $\eta_0 > 0$  such that for all  $u \leq u_0$ , and  $\eta \leq \eta_0$  and  $x \in \mathcal{B}(x^\star, u) \cap D(\eta)$ ,*

$$\int_0^{\tau_u^M(x)} \sup_{s \leq t} |\phi_R(x, s) - x_R^\star| dt \leq C\sqrt{u + \eta}.$$

**Lemma 3.5.** *For all  $v \in (0, v_\star)$ , there exists some  $\eta \in (0, r_\star)$  and  $r > 0$  such that, for all  $x \in B_r(x^\star) \cap (\mathbb{R}_+^*)^2$ , for all  $t \leq \tau_v^M(x)$ ,  $F_M(\phi(x, t)) \geq \eta$ .*

*Proof.* Since  $v < v_*$ , there exists  $\eta \in (0, r_*)$  such that  $v < v_\eta$ . Hence there exists some  $r > 0$  such that

$$v < \inf_{x \in B_r((x_R^*, 0)) \cap (\mathbb{R}_+^*)^2} \mathcal{C}(x, \eta).$$

Let  $x$  belong to  $B_r((x_R^*, 0)) \cap (\mathbb{R}_+^*)^2$ . Then

$$v < \mathcal{C}(x, \eta) = \sup\{\phi_M(x, t) : t \leq \bar{\tau}_\eta(x)\}$$

and we get  $F_M(\phi(x, t)) \geq \eta$  for  $t \leq \tau_v^M(x)$  by continuity of  $t \rightarrow \phi(x, t)$ .  $\square$

Let us now give the proof of Proposition 3.2.

*Proof of Proposition 3.2.* We consider the case  $\partial_R F_R(x^*) < 0$  and the other can be treated similarly. Thanks to Lemma 3.3, we can choose  $u_0 > 0$  and  $\eta_0 > 0$  such that, for all  $u \leq u_0$ , and  $x \in B_u(x^*)$ ,

$$|\phi_R(x, \tau_u^M(x)) - x_R^*| \leq \sup_{t \leq \tau_u^M(x)} |\phi_R(x, t) - x_R^*| \leq C\sqrt{u}. \quad (3.2)$$

Let  $0 < \varepsilon_0 \leq u_0$  small enough such that

$$\inf F_M([x_R^* - C\sqrt{\varepsilon_0}, x_R^* + C\sqrt{\varepsilon_0}] \times [0, \varepsilon_0]) \geq 3r_*/4.$$

We have, for any  $x \in B_0 = B_{C\sqrt{\varepsilon_0}}(x^*) \cap (\mathbb{R}_+^*)^2$ ,

$$\tau_{\varepsilon_0}^M(x) < \inf \{s \geq 0 : F_M(\phi(x, s)) \leq r_*/2\}.$$

As a consequence, for all  $\eta \leq r_*/2$  and  $x \in B_0$ ,  $\mathcal{C}(x, \eta) \geq \varepsilon_0$  and  $v_* \geq \varepsilon_0$ . This proves that  $v_*$  is positive. It is also finite owing to boundedness Assumption 3.1.

Let us now prove the second part of the proposition. We start by proving the result for small values of  $v$  using the first part which guarantees a lowerbound for the growth rate of mutants. For any  $v \in (0, \varepsilon_0]$  and  $x \in B_0$  and  $t \leq \tau_v(x)$ ,

$$F_M(\phi(x, t)) \geq r_*/2.$$

Furthemore

$$\tau_{\phi_M(x, t)}^M(x) = t$$

and

$$\partial_t \phi_M(t, x) = \phi_M(t, x) F_M(\phi_R(x, \tau_{\phi_M(x, t)}^M(x)), \phi_M(x, t)).$$

By separation of variables (put  $u = \phi_M(x, t)$ ) and integration on the time interval  $[0, \tau_v^M(x)]$ , we get

$$\tau_v^M(x) = \int_{x_M}^v \frac{1}{u F_M(\phi_R(x, \tau_u^M(x)), u)} du = \frac{1}{r_*} \log(v/x_M) + \mathcal{R}_v(x),$$

for any  $v \in (0, \varepsilon_0]$  and  $x \in B_0$  and  $t \leq \tau_v(x)$ , where

$$\mathcal{R}_v(x) = \int_{x_M}^v \frac{1}{u} \left( \frac{1}{F_M(\phi_R(x, \tau_u^M(x)), u)} - \frac{1}{r_*} \right) du.$$

To get the expected estimate, we need to evaluate the term inside the integral for  $u$  close to 0. Using that  $F_M$  is locally Lipschitz and  $r_* = F_M(x^*)$  and (3.2), and we obtain that for  $u \leq \varepsilon_0$  and  $x \in B_0$ ,

$$\frac{1}{u} \left( \frac{1}{F_M(\phi_R(x, \tau_u^M(x)), u)} - \frac{1}{r_*} \right) \leq \frac{C}{\sqrt{u}}.$$

The integral of the right hand side is convergent in 0. Besides,  $(x, u) \rightarrow F_M(\phi_R(x, \tau_u^M(x)), u)$  is uniformly continuous for  $u \leq \varepsilon_0$  and  $x \in B_0$  such that  $x_M \leq u$ . It ensures that

$$\sup \{ |\mathcal{R}_v(x) - \mathcal{R}_v(x')| : x, x' \in B_r(x_*) \cap (\mathbb{R}_+^*)^2, v \in [\max(x_M, x'_M), \varepsilon_0] \} \xrightarrow{r \rightarrow 0} 0.$$

The uniform Cauchy criterion ensures then that for any  $v \in (0, \varepsilon_0)$ ,  $\mathcal{R}_v(x)$  converges, uniformly in  $v$ , to a real number  $\tau(v)$  as  $x$  goes to  $x^*$ . Therefore,

$$\tau(v) = \lim_{x \rightarrow x^*} \int_{x_M}^v \frac{1}{u} \left( \frac{1}{F_M(\phi_R(x, \tau_u^M(x)), u)} - \frac{1}{r_*} \right) du. \quad (3.3)$$

and the previous estimates also guarantee that  $\tau$  is continuous on  $(0, \varepsilon_0]$  and that  $\tau(v)$  goes to 0 as  $v$  goes to 0. This ends the proof for small  $v \in (0, \varepsilon_0]$ .

To deal deal with  $v \in [\varepsilon_0, v_*)$ , we can use :

$$\tau_v^M(x) = \tau_{v_0}^M(x) + \tau_v^M(\phi_R(x, \tau_{v_0}^M(x)), v_0).$$

It remains to check that the last term converges as  $x$  goes to  $x^*$ . This can be achieved as above with uniform continuity of  $(x, u) \rightarrow F_M(\phi_R(x, \tau_u^M(x)), u)$ .  $\square$

### 3.2. Quantitative approximation of the stochastic process and speed of convergence

The main result of this section can now be stated.

**Theorem 3.6.** *Under Assumptions 1.1 and 3.1, for all  $x_K = (x_R^K, x_M^K) \in (\mathbb{R}_+^*)^2$ , satisfying*

$$1/K \ll x_M^K \ll 1 \quad \text{and} \quad |x_R^K - x_R^*| \log(1/x_M^K) \ll 1,$$

*we have for any  $v \in (0, v_*)$ ,*

$$\sup_{t \leq \tau_v^M(x_K)} \left| \frac{X_M^K(x_K, t)}{\phi_M(x_K, t)} - 1 \right| \xrightarrow{K \rightarrow \infty} 0, \text{ in probability.}$$

Note that the most interesting assumption is  $1/K \ll x_M^K$ , since initial conditions of order one have been extensively studied. Our result stops at hitting times  $\tau_v^M(x)$  (which imposes  $x_M^K \ll 1$ ) but could be easily extended on finite time interval and with initial conditions of order one.

In all this subsection, the hypotheses of Theorem 3.6 are in force, and we note, for  $\bullet \in \{R, M\}$ ,

$$X_\bullet^K(t) := X_\bullet^K(x_K, t), \quad x_\bullet^K(t) := \phi_\bullet(x_K, t)$$

and

$$Y_\bullet^K(t) := X_\bullet^K(t) - x_\bullet^K(t).$$

In addition, we introduce the following stopping times: for  $\bullet \in \{R, M\}$ ,

$$\begin{aligned} \theta_\bullet^K &:= \inf \left\{ t > 0 : \frac{X_\bullet^K(t)}{x_\bullet^K(t)} \geq 2 \right\}, \quad \theta^K := \theta_R^K \wedge \theta_M^K, \\ \sigma_\eta^K &:= \inf \left\{ t > 0 : \int_0^t \sup_{r \leq s} |Y_R^K(r)| ds \geq \eta \right\} \text{ for any } \eta > 0, \\ \tilde{\sigma}_\varepsilon^K &:= \inf \left\{ t > 0 : \sup_{s \leq t} |Y_R^K(s)| \geq \varepsilon^{1/2} \right\} \text{ for any } \varepsilon > 0. \end{aligned}$$

In order to prove Theorem 3.6, we need the following result providing a control in probability when  $K$  goes to infinity of the stopping times  $\theta_R^K$ ,  $\sigma_\eta^K$  and  $\tilde{\sigma}_\varepsilon^K$ .

**Lemma 3.7.** *There exist  $L, \eta_0 > 0$  such that for all  $0 < \eta \leq \eta_0$ ,*

$$\mathbb{P}(\sigma_\eta^K \leq \tau_{L\eta}^M(x^K) \wedge \theta^K) \leq \frac{C}{\eta} K^{-1/2} (\log(1/x_M^K))^{3/2}, \quad (3.4)$$

$$\mathbb{P}(\tilde{\sigma}_{L\eta}^K \leq \tau_{L\eta}^M(x^K) \wedge \theta^K \wedge \sigma_\eta^K) \leq C\eta^{-1/2} K^{-1/2} (\log(1/x_M^K))^{1/2}, \quad (3.5)$$

$$\mathbb{P}(\theta_R^K \leq \tau_{L\eta}^M(x^K) \wedge \theta_M^K) \leq \frac{C}{\eta} K^{-1/2} (\log(1/x_M^K))^{3/2}. \quad (3.6)$$

*Proof.* To begin with, let us prove (3.4). For the sake of notation, let us denote

$$S^K := \tau_{L\eta}^M(x^K) \wedge \theta^K \wedge \sigma_\eta^K.$$

By definition of  $\sigma_\eta^K$ ,

$$\{\sigma_\eta^K \leq \tau_{L\eta}^M(x^K) \wedge \theta^K\} = \left\{ \int_0^{S^K} \sup_{r \leq s} |Y_R^K(r)| ds \geq \eta \right\}.$$

To control the probability of the event above, let us define

$$\begin{aligned} H^K(t) &:= M^K(t) + R^K(t) + \int_0^t Y_M^K(r) \partial_M G_R(x^K(r)) dr, \\ M^K(t) &:= \frac{1}{K} \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{z \leq K X_R^K(r-) b_R(X^K(r-))\}} d\tilde{\mathcal{N}}_R^b(r, z) \\ &\quad - \frac{1}{K} \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{z \leq K X_R^K(r-) d_R(X^K(r-))\}} d\tilde{\mathcal{N}}_R^d(r, z), \\ R^K(t) &:= \int_0^t [G_R(X^K(r)) - G_R(x^K(r)) - Y_M^K(r) \partial_M G_R(x^K(r)) - Y_R^K(r) \partial_R G_R(x^K(r))] dr. \end{aligned}$$

such that for all  $t > 0$ ,

$$Y_R^K(t) = H^K(t) + \int_0^t Y_R^K(s) \partial_R G_R(x^K(r)) dr.$$

In a first time, let us treat the case  $\partial_R G_R(x^*) = x_R^* \partial_R F_R(x^*) < 0$ . Thanks to Lemma 6.3 we have

$$\sup_{s \leq t} |Y_R^K(s)| \leq \sup_{s \leq t} |H^K(s) - H^K(\lfloor s \rfloor)| \leq 2 \sup_{s \leq t} |H^K(t)|.$$

On the other hand, if  $\partial_R G_R(x^*) = x_R^* \partial_R F_R(x^*) = 0$ , then

$$|\partial_R G_R(x^K(r))| \leq C x_M^K(r) + C |x_R^K(r) - x_R^*|,$$

and, by Grönwall's lemma,

$$|Y_R^K(t)| \leq |H^K(t)| \exp\left(C \int_0^t x_M^K(r) dr + C \int_0^t |x_R^K(r) - x_R^*| dr\right).$$

Consequently, Lemma 3.3 implies that

$$|Y_R^K(t)| \leq |H^K(t)| e^{C\sqrt{L\eta}} \leq 2|H^K(t)|,$$

for  $\eta \leq \eta_0$  fixed to some small enough value.

Gathering the two cases, we have proved that if  $\partial_R F_R(x^*) \leq 0$ , for  $t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K$ ,

$$\begin{aligned} \sup_{s \leq t} |Y_R^K(s)| &\leq 2 \sup_{s \leq t} |H^K(s)| \\ &\leq 2 \sup_{s \leq t} |M^K(s)| + 2 \sup_{s \leq t} |R^K(s)| + 2 \int_0^t |Y_M^K(s)| \cdot |\partial_M G_R(x^K(s))| ds. \end{aligned} \quad (3.7)$$

In addition, for  $t \leq \tau_{L\eta}^M(x^K)$ ,

$$\begin{aligned} \int_0^t x_M^K(s) ds &= \int_0^t \frac{1}{F_M(x^K(s))} (x_M^K)'(s) ds \\ &\leq C \int_0^t (x_M^K)'(s) ds = C(x_M^K(t) - x_M^K(0)) \leq Cx_M^K(t). \end{aligned} \quad (3.8)$$

Using twice this inequality

$$\begin{aligned} &\int_0^{S^K} \int_0^s |Y_M^K(r)| \cdot |\partial_M G_R(x^K(r))| dr ds \\ &\leq C \int_0^{S^K} \int_0^s \left( \frac{X_M^K(r)}{x_M^K(r)} + 1 \right) x_M^K(r) dr ds \leq C \int_0^{\tau_{L\eta}^M(x^K)} \int_0^s x_M^K(r) dr ds \\ &\leq C \int_0^{\tau_{L\eta}^M(x^K)} x_M^K(s) ds \leq Cx_M^K(\tau_{L\eta}^M(x^K)) \leq CL\eta. \end{aligned} \quad (3.9)$$

By Taylor-Lagrange's inequality and Lemma 3.3 (which guarantees that the process  $x^K$  is bounded up to time  $S^K$ ), we also have

$$\int_0^{S^K} \sup_{r \leq s} |R^K(r)| ds \leq C \int_0^{S^K} \int_0^s (|Y_M^K(r)|^2 + |Y_M^K(r)| \cdot |Y_R^K(r)| + |Y_R^K(r)|^2) dr ds.$$

Since  $\tau_{L\eta}^M(x^K) \leq S^K$ , and using the same computation as (3.9),

$$\int_0^{S^K} \int_0^s |Y_M^K(r)|^2 dr ds \leq L\eta \int_0^{S^K} \int_0^s |Y_M^K(r)| dr ds \leq C(L\eta)^2.$$

Besides  $|Y_M^K(r)| \cdot |Y_R^K(r)| \leq |Y_M^K(r)| \sup_{r' \leq s} |Y_R^K(r')|$  for  $r \leq s$ ,

$$\int_0^{S^K} \int_0^s |Y_M^K(r)| \cdot |Y_R^K(r)| dr ds \leq CL\eta \int_0^{S^K} \sup_{r' \leq s} |Y_R^K(r')| ds \leq CL\eta^2.$$

Similarly,  $\int_0^{S^K} |Y_R^K(r)|^2 dr ds \leq C\eta \int_0^{S^K} \sup_{r' \leq s} |Y_R^K(r')| ds \leq C\eta^2$ . Combining these estimates yields

$$\int_0^{S^K} \sup_{r \leq s} |R^K(r)| ds \leq C\eta^2. \quad (3.10)$$

Plugging (3.9) and (3.10) in (3.7), we can choose  $L > 0$  small enough and get

$$\mathbb{P} \left( \int_0^{\tau_{L\eta}^M(x^K) \wedge \theta^K} |Y_R^K(s)| ds > \eta \right) \leq \mathbb{P} \left( \int_0^{\tau_{L\eta}^M(x^K) \wedge \theta^K} |M^K(s)| ds > \frac{\eta}{4} \right)$$

Using again Doob and Cauchy-Schwarz inequalities

$$\begin{aligned} \mathbb{E} \left[ \sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K} |M^K(t)| \right] &\leq C \mathbb{E} \left[ \langle M^K \rangle_{\tau_{L\eta}^M(x^K) \wedge \theta^K} \right]^{1/2} \\ &\leq CK^{-1/2} \mathbb{E} \left[ \int_0^{\tau_{L\eta}^M(x^K) \wedge \theta^K} X_R^K(s) (b_R + d_R) (X^K(s)) ds \right]^{1/2} \\ &\leq CK^{-1/2} \tau_{L\eta}^M(x^K). \end{aligned}$$

Using Markov inequality and the comparison of  $\tau_{L\eta}^M(x^K)$  with  $\log(1/x_M^K)$  (see Section 3.1), we get

$$\mathbb{P}\left(\int_0^{\tau_{L\eta}^M(x^K) \wedge \theta^K} |Y_R^K(s)| ds > \eta\right) \leq \frac{C}{\eta} K^{-1/2} (\log(1/x_M^K))^{3/2}.$$

This proves (3.4). Now we prove (3.5). By definition of  $\tilde{\sigma}_{L\eta}^K$ ,

$$\left\{ \tilde{\sigma}_{L\eta}^K \leq \tau_{L\eta}^M(x^K) \wedge \theta^K \wedge \sigma_\eta^K \right\} = \left\{ \sup_{s \leq \tau_{L\eta}^M(x^K) \wedge \theta^K \wedge \sigma_\eta^K \wedge \tilde{\sigma}_{L\eta}^K} |Y_R^K(s)| > (L\eta)^{1/2} \right\}.$$

For the sake of readability, we note

$$T^K := \tau_{L\eta}^M(x^K) \wedge \theta^K \wedge \sigma_\eta^K \wedge \tilde{\sigma}_{L\eta}^K.$$

To control the probability of the event above, we use (3.7) again. We have that

$$\begin{aligned} & \int_0^{T^K} |Y_M^K(s)| \cdot |\partial_M G_R(x^K(s))| ds \\ & \leq C \int_0^{T^K} x_M^K(s) \frac{|Y_M^K(s)|}{x_M^K(s)} ds \leq C \int_0^{\tau_\varepsilon^M(x^K)} x_M^K(s) ds \\ & \leq C \int_0^{\tau_{L\eta}^M(x^K)} \frac{1}{\psi_M(x^K(s))} (x_M^K)'(s) ds \leq C(x_M^K(\tau_{L\eta}^M(x^K)) - x_M^K(0)) \leq CL\eta, \end{aligned}$$

and

$$\begin{aligned} \sup_{s \leq T^K} |R^K(s)| & \leq C \int_0^{T^K} |X_M^K(s) - x_M^K(s)|^2 ds + C \int_0^{T^K} |Y_R^K(s)|^2 ds + C \int_0^{T^K} |Y_R^K(s)| \cdot |Y_M^K(s)| ds \\ & \leq C(L\eta)^2 + C\eta(L\eta)^{1/2} + C(L\eta)^{3/2} \leq C\eta^{3/2}. \end{aligned}$$

Since  $\eta$  is arbitrary small, we have

$$\begin{aligned} \mathbb{P}\left(\sup_{s \leq T^K} |Y_R^K(s)| > (L\eta)^{1/2}\right) & \leq \mathbb{P}\left(\sup_{s \leq T^K} |M_s^K| > (L\eta)^{1/2}/4\right) \\ & \leq C\eta^{-1/2} \mathbb{E}[\langle M_s^K \rangle_{T^K}]^{1/2} \leq C\eta^{-1/2} K^{-1/2} (\log(1/x_M^K))^{1/2}, \end{aligned}$$

which proves (3.5).

Finally, we prove (3.6) using (3.4) and (3.5). Notice that

$$\left\{ \theta_R^K \leq \tau_{L\eta}^M(x^K) \wedge \theta_M^K \right\} \subseteq \left\{ \sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K} \frac{|Y_R^K(t)|}{x_R^K(t)} \geq 1 \right\}.$$

Then, by Lemma 3.3, we know that for all  $t \leq \tau_{L\eta}^M(x^K)$ ,

$$x_R^K(t) \geq x_R^* - C\sqrt{L\eta} \geq x_R^*/2,$$

if  $\eta > 0$  is small enough. Hence, for  $\eta > 0$  small enough,

$$\begin{aligned} \left\{ \theta_R^K \leq \tau_{L\eta}^M(x^K) \wedge \theta_M^K \right\} & \subseteq \left\{ \sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K} |Y_R^K(t)| \geq x_R^*/2 \right\} \\ & \subseteq \left\{ \tilde{\sigma}_{L\eta}^K \leq \tau_{L\eta}^M(x^K) \wedge \theta^K \right\} \\ & \subseteq \left\{ \sigma_\eta^K \leq \tau_{L\eta}^M(x^K) \wedge \theta^K \right\} \cup \left\{ \tilde{\sigma}_{L\eta}^K \leq \tau_{L\eta}^M(x^K) \wedge \theta^K \wedge \sigma_\eta^K \right\}. \end{aligned}$$

Since the probabilities of the two last events above are controlled respectively by (3.4) and (3.5), the result is proved.  $\square$

*Proof of Theorem 3.6.* We decompose the proof of Theorem 3.6 in three steps. The first step is the approximation of  $X_M^K$  by  $x_M^K$  until the level of mutants reaches some small macroscopic value. The second step is more classical and consists in showing that the approximation still holds true during an additional time length  $T > 0$ . In the last step, we show that have reached levels of mutants  $v_0 < v^*$ .

*Step 1.* In this first step, we prove that there exist  $C, L, \eta_0 > 0$  such that for any  $\eta \in (0, \eta_0)$ ,

$$\mathbb{P} \left( \sup_{t \leq \tau_{L\eta}^M(x^K)} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| > \eta \right) \leq C \cdot \frac{1}{\eta} \cdot K^{-1/2} \left( \ln \left( \frac{1}{x_M^K(0)} \right) \right)^{3/2}. \quad (3.11)$$

By Itô's formula,

$$\begin{aligned} \frac{X_M^K(t)}{x_M^K(t)} &= 1 - \int_0^t \frac{X_M^K(s)}{x_M^K(s)} (b_M - d_M)(x^K(s)) ds \\ &\quad + \int_{[0,t] \times \mathbb{R}_+} \frac{1}{Kx_M^K(s)} \mathbf{1}_{\{u \leq K \cdot X_M^K(s-) \cdot b_M(X^K(s-))\}} \tilde{\mathcal{N}}_M^b(ds, du) \\ &\quad - \int_{[0,t] \times \mathbb{R}_+} \frac{1}{Kx_M^K(s)} \mathbf{1}_{\{u \leq K \cdot X_M^K(s-) \cdot d_M(X^K(s-))\}} \tilde{\mathcal{N}}_M^d(ds, du). \end{aligned}$$

Writing  $\tilde{\mathcal{N}}_M^b$  and  $\tilde{\mathcal{N}}_M^d$  the compensated measure of  $\mathcal{N}_M^b$  and  $\mathcal{N}_M^d$ , and introducing

$$\begin{aligned} E_M^K(t) &:= \int_{[0,t] \times \mathbb{R}_+} \frac{1}{Kx_M^K(s)} \mathbf{1}_{\{u \leq K \cdot X_M^K(s-) \cdot b_M(X^K(s-))\}} \tilde{\mathcal{N}}_M^b(ds, du) \\ &\quad - \int_{[0,t] \times \mathbb{R}_+} \frac{1}{Kx_M^K(s)} \mathbf{1}_{\{u \leq K \cdot X_M^K(s-) \cdot d_M(X^K(s-))\}} \tilde{\mathcal{N}}_M^d(ds, du), \end{aligned}$$

we have

$$\frac{X_M^K(t)}{x_M^K(t)} = 1 + E_M^K(t) + \int_0^t \frac{X_M^K(s)}{x_M^K(s)} (F_M(X^K(s)) - F_M(x^K(s))) ds,$$

and, for  $t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K$ ,

$$\langle E_M^K \rangle_t = \int_0^t \frac{1}{Kx_M^K(s)} \cdot \frac{X_M^K(s)}{x_M^K(s)} \cdot (b_M + d_M)(X^K(s)) ds \leq \frac{2C}{K} \int_0^t \frac{1}{x_M^K(s)} ds.$$

Consequently, for any  $T \leq \tau_{L\eta}^M(x^K) \wedge \theta^K$ ,

$$\begin{aligned} \sup_{t \leq T} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| &\leq \sup_{t \leq T} |E_M^K(t)| + C \int_0^T (|X_M^K(s) - x_M^K(s)| + |X_R^K(s) - x_R^K(s)|) ds \\ &\leq \sup_{t \leq T} |E_M^K(t)| + C \int_0^T x_M^K(s) \cdot \sup_{r \leq s} \left| \frac{X_M^K(r)}{x_M^K(r)} - 1 \right| ds + \int_0^T |X_R^K(s) - x_R^K(s)| ds. \end{aligned}$$

So, by Gronwall's lemma, for any  $T \leq \tau_{L\eta}^M(x^K) \wedge \theta^K$ ,

$$\sup_{t \leq T} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| \leq \left( \sup_{t \leq T} |E_M^K(t)| + C \int_0^T |X_R^K(s) - x_R^K(s)| ds \right) e^{C \int_0^T x_M^K(s) ds}. \quad (3.12)$$

Besides, by Burkholder-Davis-Gundy's inequality, for all stopping time  $T \leq \tau_{L\eta}^M(x^K) \wedge \theta^K$ ,

$$\mathbb{E} \left[ \sup_{t \leq T} |E_M^K(t)| \right] \leq \mathbb{E} \left[ \langle E_M^K \rangle_T^{1/2} \right] \leq CK^{-1/2} \mathbb{E} \left[ \int_0^T \frac{1}{x_M^K(s)} ds \right]^{1/2}.$$

Then, as  $F_M(x^K(t))$  is lower-bounded on  $[0, t_{L\eta}^M(x^K)]$  by some  $\kappa > 0$  (thanks to Lemma 3.3 since  $F_M(x^*) > 0$  if  $\eta$  is chosen small enough), we have

$$\mathbb{E} \left[ \sup_{t \leq T} |E_M^K(t)| \right] \leq C \cdot \frac{1}{\sqrt{Kx_M^K(0)}} \mathbb{E} \left[ \int_0^T e^{-\kappa s} ds \right]^{1/2} \leq C \cdot \frac{1}{\kappa} \cdot \frac{1}{\sqrt{Kx_M^K(0)}}, \quad (3.13)$$

where the value of  $C$  has changed in the last inequality, but is still independent of  $K$ . Recalling that, thanks to (3.8), we have

$$\int_0^{\tau_{L\eta}^M(x^K)} x_M^K(s) ds \leq CL\eta.$$

So, inequality (3.12) can be rewritten as

$$\sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| \leq C \sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K} |E_M^K(t)| + C \int_0^{\tau_{L\eta}^M(x^K) \wedge \theta^K} |Y_R^K(s)| ds.$$

By Markov's inequality and (3.13),

$$\mathbb{P} \left( \sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K} |E_M^K(t)| > \frac{\eta}{2C} \right) \leq \frac{C}{\eta \sqrt{Kx_M^K(0)}}.$$

Besides, by (3.4) from Lemma 3.7,

$$\mathbb{P} \left( \int_0^{\tau_{L\eta}^M(x^K) \wedge \theta^K} |Y_R^K(s)| ds > \frac{\eta}{2C} \right) \leq \frac{C}{\eta} K^{-1/2} \left( \log \left( \frac{1}{x_M^K(0)} \right) \right)^{3/2}.$$

We obtain from the three last inequalities that

$$\mathbb{P} \left( \sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta^K} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| > \eta \right) \leq \frac{C}{\eta} \cdot K^{-1/2} \left( \ln \left( \frac{1}{x_M^K(0)} \right) \right)^{3/2}.$$

Using now that for  $\eta \in (0, 1)$ ,

$$\left\{ \theta_M^K < \infty, \theta_M^K \leq \tau_{L\eta}^M(x^K) \right\} \subseteq \left\{ \sup_{t \leq \tau_{L\eta}^M(x^K) \wedge \theta_M^K} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| > \eta \right\},$$

and (3.6) proves (3.11).

*Step 2.* We fix  $T > 0$  and we prove now that the approximation of  $X_M^K$  by  $x_M^K$  is still valid on the time interval  $[\tau_{L\eta}^M(x^K), \tau_{L\eta}^M(x^K) + T]$ . The fact that the mutant population has not left the neighborhood of 0 makes that such estimates are more classical. More precisely, we denote

$$\tilde{X}_\bullet^K(t) := X_\bullet^K(t + \tau_{L\eta}^M(x^K)), \quad \tilde{x}_\bullet^K(t) := x_\bullet^K(t + \tau_{L\eta}^M(x^K))$$

for  $\bullet \in \{R, M\}$ . First, we observe that

$$\gamma_{T,\eta} = \inf_{t \leq T, K \geq 1} \tilde{x}_M^K(t) > 0,$$

since the boundary  $\mathbb{R}_+ \times \{0\}$  is stable for the flow  $\phi$ , which is continuous, and the initial value  $\tilde{x}^K(0)$  belongs to a compact set of  $(\mathbb{R}_+^*)^2$  (thanks to Lemma 3.3)). Then

$$\mathbb{P} \left( \sup_{t \leq T} \left| \frac{\tilde{X}_M^K(t) - \tilde{x}_M^K(t)}{\tilde{x}_M^K(t)} \right| > \eta \right) \leq \mathbb{P} \left( \sup_{t \leq T} \left| \tilde{X}_M^K(t) - \tilde{x}_M^K(t) \right| > \eta \gamma_{T,\eta} \right),$$



and it remains to prove that for any  $\eta > 0$ ,

$$\mathbb{P} \left( \sup_{t \leq T} \left| \tilde{X}_M^K(t) - \tilde{x}_M^K(t) \right| > \eta \right) \xrightarrow{K \rightarrow \infty} 0. \quad (3.14)$$

This result is a classical approximation of population processes by dynamical systems on finite time intervals, see Ethier-Kurtz [19]. For sake of completeness and give explicit bounds, let us prove the result here. We use again a stopping time which ensures boundedness :

$$\tilde{\theta}^K := \inf \left\{ t > 0 : \tilde{X}_M^K(t) > 2\tilde{x}_M^K(t) \text{ or } \tilde{X}_R^K(t) > 2\tilde{x}_R^K(t) \right\}.$$

We have, for any  $t \leq \tilde{\theta}^K$ ,

$$\left| \tilde{X}_{\bullet}^K(t) - \tilde{x}_{\bullet}^K(t) \right| \leq |M_{\bullet}^K(t)| + C \int_0^t \left( |\tilde{X}_R^K(s) - \tilde{x}_R^K(s)| + |\tilde{X}_M^K(s) - \tilde{x}_M^K(s)| \right) ds,$$

where  $M_R^K, M_M^K$  are local martingales satisfying

$$\langle M_{\bullet}^K \rangle_{t \wedge \tilde{\theta}^K} \leq CtK^{-1}.$$

Then, Gronwall's lemma implies that

$$\sup_{t \leq T \wedge \tilde{\theta}^K} \left( \left| \tilde{X}_M^K(t) - \tilde{x}_M^K(t) \right| + \left| \tilde{X}_R^K(t) - \tilde{x}_R^K(t) \right| \right) \leq \left( \sup_{t \leq T \wedge \tilde{\theta}^K} |M_R^K(t)| + \sup_{t \leq T \wedge \tilde{\theta}^K} |M_M^K(t)| \right) e^{CT}.$$

Hence, using Markov's and Burkholder-Davis-Gundy's inequalities, we obtain

$$\mathbb{P} \left( \sup_{t \leq T \wedge \tilde{\theta}^K} \left( \left| \tilde{X}_M^K(t) - \tilde{x}_M^K(t) \right| + \left| \tilde{X}_R^K(t) - \tilde{x}_R^K(t) \right| \right) > \eta' \right) \leq \frac{1}{\eta'} C_T K^{-1/2}. \quad (3.15)$$

Finally, let us now dismiss the stopping time  $\tilde{\theta}^K$ . Let  $A_{\eta'}^K(T)$  be the event

$$A_{\eta'}^K(T) := \left\{ \sup_{t \leq T} \left( \left| \tilde{X}_M^K(t) - \tilde{x}_M^K(t) \right| + \left| \tilde{X}_R^K(t) - \tilde{x}_R^K(t) \right| \right) > \eta' \right\}.$$

We have

$$\begin{aligned} \mathbb{P} (A_{\eta'}^K(T)) &\leq \mathbb{P} \left( A_{\eta'}^K(T) \cap \{ \tilde{\theta}^K \geq T \} \right) + \mathbb{P} \left( A_{\eta'}^K(T) \cap \{ \tilde{\theta}^K < T \} \right) \\ &\leq \mathbb{P} \left( A_{\eta'}^K(T \wedge \tilde{\theta}^K) \right) + \mathbb{P} \left( A_1^K(T \wedge \tilde{\theta}^K) \right) \\ &\leq 2\mathbb{P} \left( A_{\eta'}^K(T \wedge \tilde{\theta}^K) \right). \end{aligned}$$

Recalling that (3.15) provides a control of  $\mathbb{P} \left( A_{\eta'}^K(T \wedge \tilde{\theta}^K) \right)$ , (3.14) is proved.

*Step 3.* To conclude the proof of the theorem, it only remains to show that, for all  $L\eta < v < v_*$ , there exists some  $T_v(\eta) > 0$  independent of  $K$  such that

$$\tau_v^M(x^K) \leq \tau_{L\eta}^M(x^K) + T_v(\eta),$$

with the same  $L > 0$  as in the two first steps of the proof. This property relies on Proposition 3.2. With the same notation, let us denote

$$\delta(x, v') := \tau_{v'}^M(x) - \frac{1}{r_{\star}} \log(v'/x_M^K) - \tau(v') \text{ and } \Delta(x) := \sup_{0 < v' < v_{\star}} |\delta(x, v')|.$$

By Proposition 3.2, we know that  $\Delta(x^K)$  vanishes as  $K$  goes to infinity. Recall that  $x_M^K < L\eta < v$ . We have, by definition of  $\Delta$ ,

$$\tau_v^M(x^K) \leq \frac{1}{r_\star} \log(v/x_M^K) + |\tau(v)| + \Delta(x^K),$$

and

$$\tau_{L\eta}^M(x^K) \geq \frac{1}{r_\star} \log(L\eta/x_M^K) - |\tau(L\eta)| - \Delta(x^K).$$

As a consequence,

$$\begin{aligned} \tau_v^M(x^K) &\leq \frac{1}{r_\star} \log(v/(L\eta)) + \frac{1}{r_\star} \log(L\eta/x_M^K) + |\tau(v)| + \Delta(x^K) \\ &\leq \tau_{L\eta}^M(x^K) + \frac{1}{r_\star} \log(v/(L\eta)) + |\tau(v)| + |\tau(L\eta)| + 2\Delta(x^K). \end{aligned}$$

Since the sequence  $\Delta(x^K)$  vanishes, it is bounded w.r.t.  $K$ . Then, choosing

$$T_v(\eta) := \frac{1}{r_\star} \log(v/(L\eta)) + |\tau(v)| + |\tau(L\eta)| + 2 \sup_{K \in \mathbb{N}^*} \Delta(x^K),$$

the result is proved.  $\square$

#### 4. Thresholds hitting times

In this section, we use the results of the two previous ones to approximate the hitting times of the process  $(N_M^K(t))_t$  by some simple and explicit random variables. Let us recall that, for any  $n \geq 1$ , we have defined in (1.4) the stopping time

$$\tau_M^K(n) := \inf \{t > 0 : N_M^K(t) \geq n\}.$$

Besides  $\tau$  is the deterministic function defined in (3.3) and  $W$  is the a.s. limit of  $Z(t) \exp(-r_\star t)$ , which is distributed as

$$\frac{d_\star}{b_\star} \delta_0 + \frac{b_\star - d_\star}{b_\star} \text{Exp} \left( \frac{b_\star - d_\star}{b_\star} \right).$$

The main result of this section is the following, recalling that  $v_\star$  is defined at the beginning of Section 3.1. It gives a quantitative information on the law of hitting times of thresholds of order  $K$  but also for thresholds of size less than  $K$ .

**Theorem 4.1.** *Grant Assumptions 1.1 and 3.1. Let  $(\zeta_K)_K$  be sequence which tends to infinity and satisfies  $\zeta_K/K \rightarrow v$  as  $K \rightarrow \infty$ , for some  $v \in [0, v_\star)$ . Then, for any  $\varepsilon > 0$ ,*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \left| \tau_M^K(\zeta_K) - \frac{\log(\zeta_K/W)}{r_\star} - \tau(v) \right| \geq \varepsilon; W > 0 \right) = 0,$$

where  $\tau$  is the increasing continuous function such that  $\tau(0) = 0$  defined in Proposition 3.2.

Let us recall that  $(Z(t) : t \geq 0)$  is birth-death process with individual birth rate  $b_\star := b_M(x^\star)$  and individual death rate  $d_\star = d_M(x^\star)$  with initial condition  $Z(0) = 1$  that is coupled with  $(N_M^K(t) : t \geq 0)$  using the same Poisson measures (see the beginning of Section 2).

In all the section, we work on the surviving event of  $Z$ . With the notation of Section 2, it corresponds to the event  $\{W > 0\}$ . Since a.s.  $Z(t) \sim W \exp(r_\star t)$ , as  $t$  tends to infinity, we know that, on this event,  $\tau_M^K(\xi_K) < \infty$ . This allows us to define, for  $\bullet \in \{R, M\}$ ,

$$X_\bullet^K(t) := \frac{1}{K} N_\bullet^K(t + \tau_M^K(\xi_K)) = X_\bullet^K \left( \frac{1}{K} N_\bullet^K(\tau_M^K(\xi_K)), t \right),$$

using the stochastic flow defined in (3.1). By definition of  $\tau_M^K(\xi_K)$ , we have  $X_M^K(0) = \lceil \xi_K \rceil / K$ .

The proof of Theorem 4.1 is decomposed in three steps. The first one consists in approximating  $\tau_M^K(\xi_K)$  by

$$\tau_Z(\xi_K) := \inf \{t > 0 : Z(t) \geq \xi_K\},$$

and the second one in approximating  $\tau_M^K(Kv) - \tau_M^K(\xi_K)$  by

$$\tau_v^M(X^K(0)) := \inf \{t > 0 : X_M^K(t) \geq v\}.$$

The third steps merely consists in combining the two previous approximation. To establish our approximations, we need the two following lemmas. Roughly speaking, they allow to obtain an approximation of hitting-times of two processes provided a suitable approximation (in the sense of Theorems 2.3 and 3.6) of the processes themselves.

The first results link estimates for processes and estimates for hitting times.

**Lemma 4.2.** *Let  $X, Y$  be two càdlàg processes such that  $Y$  does not hit 0. We set*

$$R_t := \frac{X_t}{Y_t} - 1.$$

*Let  $T_v^X$  and  $T_v^Y$  be the hitting times of the value  $v$  for respectively  $X$  and  $Y$ . Let  $v, \rho > 0$ . Assume that there exists a.s. some  $u := u(v, \rho) > 0$  such that*

$$T_v^Y - \rho \leq T_{v-u}^Y \text{ and } T_v^Y + \rho \geq T_{v+u}^Y.$$

*Then,*

$$\{|T_v^X - T_v^Y| > \rho\} \subseteq \left\{ \sup_{t \leq T_{v+u}^Y \wedge T_v^X} |R_t| > \frac{u(v, \rho)}{v + u(v, \rho)} \right\} \cup \left\{ \sup_{t \leq T_{v-u}^Y \wedge T_v^X} |R_t| > \frac{u(v, \rho)}{v} \right\}.$$

*Proof.* To begin with

$$\{|T_v^X - T_v^Y| > \rho\} = \{T_v^X \notin [T_v^Y - \rho, T_v^Y + \rho]\} \subseteq \{T_v^X \notin [T_{v-u}^Y, T_{v+u}^Y]\}.$$

Then, since

$$\begin{aligned} X_{T_{v+u}^Y} &= \left( \frac{X_{T_{v+u}^Y}}{Y_{T_{v+u}^Y}} - 1 \right) Y_{T_{v+u}^Y} + Y_{T_{v+u}^Y} \\ &= Y_{T_{v+u}^Y} \left( 1 + R_{T_{v+u}^Y} \right) \geq (v + u) \left( 1 + R_{T_{v+u}^Y} \right) \geq v + u + (v + u) R_{T_{v+u}^Y} \geq v + \left( u - (v + u) |R_{T_{v+u}^Y}| \right), \end{aligned}$$

we have

$$\left\{ |R_{T_{v+u}^Y}| \leq \frac{u}{v + u} \right\} \subseteq \{T_v^X \leq T_{v+u}^Y\},$$

implying

$$\{T_v^X > T_{v+u}^Y\} \subseteq \left\{ \sup_{t \leq T_{v+u}^Y \wedge T_v^X} |R_t| > \frac{u}{v + u} \right\}.$$

On the other hand, by symmetry,

$$\{T_v^X < T_{v-u}^Y\} \subseteq \left\{ \sup_{t \leq T_{v-u}^Y \wedge T_v^X} |R_t| > \frac{u}{v} \right\}.$$

Finally, as

$$\{T_v^X \notin [T_{v-u}^Y, T_{v+u}^Y]\} = \{T_v^X > T_{v+u}^Y\} \cup \{T_v^X < T_{v-u}^Y\},$$

the result is proved.  $\square$

We use Lemma 4.2 to compare the hitting times of  $N^K$  with those of the branching process  $Z$  on the survival event. This relies on the results of Section 2.

**Lemma 4.3.** *On the event  $\{W > 0\}$ ,  $|\tau_M^K(\xi_K) - \tau_Z(\xi_K)|$  tends to 0 as  $K$  tends to infinity, for any sequence  $\xi$  satisfying the conditions of Theorem 2.3.*

*Proof.* Let use the notation of Lemma 4.2, with

$$X_t := N_M^K(t); Y_t := Z(t) \text{ and } v := \xi_K.$$

In particular

$$\tau_M^K(\xi_K) = T_v^X \text{ and } \tau_Z(v') = T_{v'}^Y \text{ for any } v' > 1.$$

In all the proof, we work on the surviving event of  $Z$ . To apply Lemma 4.2, we have to find, for any fixed  $\rho > 0$ , some  $u_K := u(\xi_K, \rho)$  such that

$$T_{\xi_K}^Y - \rho \leq T_{\xi_K - u_K}^Y \text{ and } T_{\xi_K}^Y + \rho \geq T_{\xi_K + u_K}^Y.$$

We will choose  $u_K$  such that  $u_K \leq \xi_K/2$ .

By definition, for all  $v' > 0$ ,  $Z_{T_{v'}^Y} = \lceil v' \rceil$ , and recall that, on the surviving event of  $Z$ ,

$$Z(t)e^{-r_*t}$$

converges almost surely to some random variable  $W > 0$ . It implies that the sequences  $(W_K)_K$ ,  $(W_K^1)_K$  and  $(W_K^2)_K$  defined as

$$\begin{aligned} W_K &:= \lceil \xi_K \rceil e^{-r_*T_{\xi_K}^Y}, \\ W_K^1 &:= \lceil \xi_K + u_K \rceil e^{-r_*T_{\xi_K + u_K}^Y}, \\ W_K^2 &:= \lceil \xi_K - u_K \rceil e^{-r_*T_{\xi_K - u_K}^Y} \end{aligned}$$

converge almost surely to  $W$  as  $K$  goes to infinity (on the surviving event of  $Z$ ), since  $\xi_K - u_K \geq \xi_K/2$  tends to infinity almost surely (thanks to (4.5)). As a consequence, we can write explicitly

$$\begin{aligned} T_{\xi_K}^Y &= \frac{1}{r_*} \ln(\lceil \xi_K \rceil / W_K), \\ T_{\xi_K + u_K}^Y &= \frac{1}{r_*} \ln(\lceil \xi_K + u_K \rceil / W_K^1), \\ T_{\xi_K - u_K}^Y &= \frac{1}{r_*} \ln(\lceil \xi_K - u_K \rceil / W_K^2). \end{aligned}$$

Consequently, the condition

$$T_{\xi_K}^Y + \rho \geq T_{\xi_K + u_K}^Y$$

is equivalent to

$$\frac{1}{r_*} \ln(\lceil \xi_K \rceil / W_K) + \rho \geq \frac{1}{r_*} \ln(\lceil \xi_K + u_K \rceil / W_K^1),$$

which is weaker than

$$\frac{1}{r_*} \ln(\xi_K / W_K) + \rho \geq \frac{1}{r_*} \ln((\xi_K + u_K + 1) / W_K^1).$$

This gives the condition

$$u_K \leq \xi_K e^{r_*\rho} (W_K^1 / W_K) - 1. \quad (4.1)$$

Similarly, to guarantee the other condition

$$T_{\xi_K}^Y - \rho \leq T_{\xi_K - u_K}^Y,$$

it is sufficient to have

$$u_K \leq \xi_K (1 - e^{-r_*\rho}(W_K^2/W_K)) - e^{-r_*\rho}(W_K^2/W_K). \quad (4.2)$$

According to the conditions (4.1) and (4.2), it is possible to choose

$$u_K := \xi_K \min(1/2; e^{r_*\rho}/2; (1 - e^{-r_*\rho})/2) =: \xi_K \lambda.$$

Indeed, with this choice of  $u_K$ , the probability that (4.1) and (4.2) are satisfied goes to one as  $K$  goes to infinity since

$$\mathbb{P}\left(\xi_K e^{r_*\rho}/2 > \xi_K e^{r_*\rho} \frac{W_K^1}{W_K} - 1\right) = \mathbb{P}\left(\frac{W_K^1}{W_K} < \frac{1}{2} + \frac{1}{\xi_K e^{r_*\rho}}\right) \leq \mathbb{P}\left(\frac{W_K^1}{W_K} < \frac{3}{4}\right) \xrightarrow{K \rightarrow \infty} 0,$$

and

$$\begin{aligned} & \mathbb{P}\left(\xi_K(1 - e^{-r_*\rho})/2 > \xi_K(1 - e^{-r_*\rho}(W_K^2/W_K)) - e^{-r_*\rho}(W_K^2/W_K)\right) \\ &= \mathbb{P}\left(-\xi_K e^{-r_*\rho}/2 > \xi_K/2 - (\xi_K + 1)e^{-r_*\rho}(W_K^2/W_K)\right) \\ &= \mathbb{P}\left(W_K^2/W_K > \frac{\xi_K}{2(\xi_K + 1)} + \frac{\xi_K}{2(\xi_K + 1)}e^{r_*\rho}\right) \leq \mathbb{P}\left(W_K^2/W_K > (e^{r_*\rho} + 1)/4\right) \xrightarrow{K \rightarrow \infty} 0. \end{aligned}$$

On the event where (4.1) and (4.2) are satisfied, we have, by Lemma 4.2,

$$\begin{aligned} & \mathbb{P}\left(|\tau_M^K(\xi_K) - \tau_Z(\xi_K)| > \rho\right) \leq \\ & \mathbb{P}\left(\sup_{t \leq \tau_M^K(\xi_K)} \left|\frac{N_M^K(t)}{Z(t)} - 1\right| > \frac{\lambda}{1 + \lambda}\right) + \mathbb{P}\left(\sup_{t \leq \tau_M^K(\xi_K)} \left|\frac{N_M^K(t)}{Z(t)} - 1\right| > \lambda\right), \end{aligned}$$

which vanishes by Theorem 2.3. That concludes the proof.  $\square$

**Proposition 4.4.** *On the surviving event of  $Z$ ,  $|\tau_M^K(\xi_K) - \log(\xi_K/W)/r_*|$  vanishes in probability, for any sequence  $\xi$  satisfying the conditions of Theorem 2.3.*

*Proof.* We show that

$$\mathbb{P}\left(\left|\tau_M^K(\xi_K) - \frac{1}{r_*} \log(\xi_K/W)\right| > \eta; W > 0\right) \xrightarrow{K \rightarrow \infty} 0$$

in two steps. Defining

$$\tau_Z(n) = \inf\{t > 0 : Z(t) > n\},$$

we will prove that

$$\mathbb{P}\left(|\tau_M^K(\xi_K) - \tau_Z(\xi_K)| > \eta; W > 0\right) \xrightarrow{K \rightarrow \infty} 0, \quad (4.3)$$

$$\mathbb{P}\left(\left|\tau_Z(\xi_K) - \frac{1}{r_*} \log(\xi_K/W)\right| > \eta; W > 0\right) \xrightarrow{K \rightarrow \infty} 0. \quad (4.4)$$

Property (4.3) is proved by Lemma 4.3. So it only remains to prove (4.4).

To begin with, let us show that, for any  $\zeta_K$  tending to infinity with  $K$ ,

$$\mathbb{P}\left(\tau_Z(\zeta_K) \geq \frac{1}{2r_*} \log \zeta_K\right) \xrightarrow{K \rightarrow \infty} 1. \quad (4.5)$$

Indeed, we have

$$\begin{aligned} \mathbb{P}\left(\tau_Z(\zeta_K) < \frac{1}{2r_\star} \log \zeta_K\right) &\leq \mathbb{P}\left(W(\tau_Z(\zeta_K)) > \lceil \zeta_K \rceil e^{-r_\star \frac{1}{2r_\star} \log \zeta_K}\right) \\ &\leq \mathbb{P}\left(W(\tau_Z(\zeta_K)) > \sqrt{\zeta_K}\right) \leq \mathbb{P}\left(\sup_{t \geq 0} W(t) > \sqrt{\zeta_K}\right), \end{aligned}$$

which vanishes as  $K$  goes to infinity since  $\zeta_K$  tends to infinity and  $W(t)$  converges almost surely to the finite random variable  $W$ . So (4.5) is proved.

Besides

$$\begin{aligned} \mathbb{P}\left(\tau_Z(\xi_K) < \frac{\log(\xi_K/W)}{r_\star} - \eta; W > 0\right) &\leq \mathbb{P}\left(W(\tau_Z(\xi_K)) > \lceil \xi_K \rceil \frac{W}{\xi_K} e^{r_\star \eta}\right) \\ &\leq \mathbb{P}\left(W(\tau_Z(\xi_K)) > W e^{r_\star \eta}\right) \\ &\leq \mathbb{P}\left(\tau_Z(\xi_K) < \frac{\log \xi_K}{2r_\star}\right) + \mathbb{P}\left(\sup_{t \geq (\log \xi_K)/2r_\star} W(t) > W e^{r_\star \eta}\right), \end{aligned}$$

where the first term of the sum above vanishes by (4.5), and the second term vanishes since  $W(t)$  converges a.s. to  $W$  as  $t$  goes to infinity. Similarly we also have that

$$\begin{aligned} \mathbb{P}\left(\tau_Z(\xi_K) > \frac{1}{r_\star} \log \xi_K - \frac{1}{r_\star} \log W + \eta; W > 0\right) \\ \leq \mathbb{P}\left(\tau_Z(\xi_K) < \frac{1}{2r_\star} \log \xi_K\right) + \mathbb{P}\left(\inf_{t \geq (\log \xi_K)/2r_\star} W(t) < W e^{-r_\star \eta}\right), \end{aligned}$$

which vanishes as  $K$  goes to infinity. This proves (4.4).  $\square$

Afterwards, we use the results of Section 3 to approximate the hitting times of the normalized process  $X^K$  with the deterministic hitting times of the dynamical system  $x^K$ .

**Lemma 4.5.** *With the notation of Section 3, recall that*

$$\tau_v^M(X^K(0)) := \inf\{t > 0 : x_M^K(t) = v\},$$

and define

$$T_v^M(X^K(0)) := \inf\{t > 0 : X_M^K(t) \geq v\}.$$

Then for all  $v < v_\star$ ,  $|\tau_v^M(X^K(0)) - T_v^M(X^K(0))|$  vanishes in probability as  $K$  goes to infinity.

*Proof.* This proof is similar to the one of Proposition 4.3, using Lemma 4.2. The proof consists in choosing some  $u := u(v, \rho)$  such that

$$\tau_v^M(X^K(0)) + \rho \geq \tau_{v+u}^M(X^K(0)) \text{ and } \tau_v^M(X^K(0)) - \rho \leq \tau_{v-u}^M(X^K(0)). \quad (4.6)$$

Let us fix some  $\mu > 1$  such that  $\mu v < v_\star$ . We will choose  $u$  such that  $u \leq (\mu - 1)v$  (i.e.  $v + u \leq \mu v$ ) and  $u \leq v$ . Let us recall that

$$\tau_v^M(X^K(0)) = \frac{1}{r_\star} \log v + \frac{1}{r_\star} \log(K/\xi_K) + \mathcal{R}_v(X^K(0)),$$

with

$$\mathcal{R}_v(x) = \int_{x_M}^v \frac{1}{u'} \left( \frac{1}{F_M(\phi_R(x, \tau_{u'}^M(x)), u')} - \frac{1}{r_\star} \right) du'.$$

We have  $\mu v < v_*$  and the probability that  $X^K(0)$  converges to  $x^*$  goes to one (since  $X_M^K(0) = \xi_K/K$  and  $|X_R^K(0) - x_R^*| \leq C\sqrt{\xi_K/K}$  thanks to Lemma 2.8).

We introduce the event

$$A^K(\eta) = \left\{ \inf_{t \leq \tau_{\mu v}^M(X^K(0))} F_M(x^K(t)) \geq \eta \right\}.$$

By Lemma 3.5, there exists some  $0 < \eta_v < r_*$  such that  $\mathbb{P}(A^K(\eta_v))$  tends to 1 as  $K$  tends to infinity. This property allows us to work now on  $A^K(\eta_v)$ . On this event, for all  $v_1 \leq v_2 \leq \mu v$ ,

$$|\mathcal{R}_{v_2}(X^K(0)) - \mathcal{R}_{v_1}(X^K(0))| \leq \int_{v_1}^{v_2} \frac{1}{u'} \left| \frac{1}{F_M(\phi_R(x, \tau_{u'}^M(x)), u')} - \frac{1}{r_*} \right| du' \leq \frac{v_2 - v_1}{v_1 \eta_v}$$

since  $F_M(\phi_R(x, \tau_{u'}^M(x)), u') \geq \eta_v$  for  $u \in [v_1, v_2]$ . Thus,

$$\tau_{v+u}^M(X^K(0)) - \tau_v^M(X^K(0)) \leq \frac{1}{r_*} \ln(1 + u/v) + \frac{u}{v \eta_v} \leq \left( \frac{1}{r_*} + \frac{1}{\eta_v} \right) \frac{u}{v},$$

and similarly

$$\tau_v^M - \tau_{v-u}^M \leq \left( \frac{1}{r_*} + \frac{1}{\eta_v} \right) \frac{u}{v-u}.$$

We choose  $u := \lambda v$  with

$$\lambda := \min \left( 1; \mu - 1; \frac{\rho}{1/r_* + 1/\eta_v}; \frac{1}{1 + (r_* + \eta_v)/\rho} \right) > 0.$$

One can check that Conditions (4.6) are satisfied. Then Lemma 4.2 implies that

$$\begin{aligned} & \mathbb{P}(|T_v^M(X^K(0)) - \tau_v^M(X^K(0))| > \rho) \\ & \leq \mathbb{P} \left( \sup_{\tau_{v+u}^M} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| > \frac{\lambda}{\lambda + 1} \right) + \mathbb{P} \left( \sup_{\tau_{v-u}^M} \left| \frac{X_M^K(t)}{x_M^K(t)} - 1 \right| > \lambda \right), \end{aligned}$$

and (3.11) allows to conclude. □

**Proposition 4.6.** *With the same notation as in the previous lemma, for all  $v < v_*$ ,*

$$T_v^M(X^K(0)) - \frac{1}{r_*} \log(v/X_M^K(0)) - \tau(v)$$

*goes to 0 in probability as  $K$  goes to infinity.*

*Proof.* Writing

$$\begin{aligned} & \left| T_v^M(X^K(0)) - \frac{1}{r_*} \log(v/X_M^K(0)) - \tau(v) \right| \\ & \leq |T_v^M(X^K(0)) - \tau_v^M(X^K(0))| + \left| \tau_v^M(X^K(0)) - \frac{1}{r_*} \log(v/X_M^K(0)) - \tau(v) \right|, \end{aligned}$$

the result is a direct consequence of Proposition 3.2 and Lemma 4.5. □

Finally, we end this section with the proof of its main result.

*Proof of Theorem 4.1.* Let us begin with the proof when  $v > 0$ , and to simplify, let us assume that  $\zeta_K = Kv$ . With the notation introduced in this section (and in particular with  $T_v^M$  introduced in Lemma 4.5), we have

$$\tau_M^K(Kv) = \tau_M^K(\xi_K) + (\tau_M^K(Kv) - \tau_M^K(\xi_K)) = \tau_M^K(\xi_K) + T_v^M(X^K(0)).$$

Recalling that  $X_M^K(0) = \xi_K/K$  and using that

$$\log(Kv) = \log(v/X_M^K(0)) + \log \xi_K,$$

we have

$$\begin{aligned} & \left| \tau_M^K(Kv) - \left( \frac{1}{r_\star} \log(Kv) + \tau(v) + \frac{1}{r_\star} \log(1/W) \right) \right| \\ & \leq \left| T_v^M(X^K(0)) - \frac{1}{r_\star} \log(v/X_M^K(0)) - \tau(v) \right| + \left| \tau_M^K(\xi_K) - \frac{1}{r_\star} \log \xi_K - \frac{1}{r_\star} \log(1/W) \right|, \end{aligned}$$

which vanishes in probability as  $K$  goes to infinity, on the surviving event of  $Z$  according to Propositions 4.4 and 4.6. That concludes the case where  $v > 0$  and  $\zeta_K = Kv$ . To treat the case where  $v > 0$  and  $\zeta_K \neq Kv$  (but still  $\zeta_K \sim Kv$ ), it is sufficient to fix some  $\eta_\varepsilon > 0$  small enough and to use the fact that, for  $K$  large enough,  $K(v - \eta_\varepsilon) < \zeta_K < K(v + \eta_\varepsilon)$ , implying that  $\tau_M^K(K(v + \eta_\varepsilon)) \leq \tau_M^K(\zeta_K) < \tau_M^K(K(v - \eta_\varepsilon))$ . The conclusion follows from the estimates we have for those two times and from the continuity of the function  $\tau$ .

Now we treat the case  $v = 0$ . Firstly, if the sequence  $(\zeta_K)_K$  satisfies the same conditions as  $(\xi_K)_K$  in Theorem 2.3, the result has been shown in Propositions 4.4. To extend the result to any  $\zeta_K \ll K$ , we introduce some  $\xi_K \ll \zeta_K$  satisfying the assumptions of Theorem 2.3 and consider the above approximation by the dynamical system starting from  $X_M^K(0) = \xi_K/K$ .  $\square$

## 5. Examples : competitive and epidemiological models

### 5.1. Lotka Volterra competitive model

We consider the dynamics of a two-dimensional population composed of resident and mutant individuals, which have their own demographic parameters and interact in a competitive way, for example in sharing resources or niche areas. This model corresponds to the general framework of the paper, with

$$b_\bullet(X^K(s-)) = b_\bullet$$

and

$$d_R(X^K(s)) = d_R + c_{1,1}X_R^K(s) + c_{1,2}X_M^K(s)$$

and

$$d_M(X^K(s)) = d_M + c_{2,1}X_R^K(s) + c_{2,2}X_M^K(s),$$

with the individual birth and death rates  $b_R, b_M \geq 0$  and  $d_R, d_M \geq 0$  and the competition matrix  $(c_{i,j})_{1 \leq i, j \leq 2}$  with non-negative coefficients. We assume that the coefficients  $b_R - d_R, b_M - d_M, c_{1,1}$  and  $c_{2,2}$  are positive.

The corresponding dynamical system is the competitive Lotka-Volterra system. The limiting ODE is (1.2) with

$$F_R(x_R, x_M) = b_R - d_R - c_{1,1}x_R - c_{1,2}x_M, \quad F_M(x_R, x_M) = b_M - d_M - c_{2,1}x_R - c_{2,2}x_M.$$

Then

$$x_R^\star = \frac{c_{1,1}}{b_R - d_R}.$$



Let us check that we can apply our results. First, we observe that regularity of Assumption 1.1 is satisfied. Second,  $(x_R^*, 0)$  is an equilibrium and

$$\partial_R F_R(x_R^*, 0) = -c_{2,2} < 0.$$

This means that the first part of Assumption 1.1 is satisfied for  $(x_R^*, 0)$ . Moreover,  $x_R^*$  is a stable equilibrium for the resident population alone, which corresponds to the first case considered in the branching process approximation.

Finally, we start from a population with only one mutant,  $N_R^K(0) = K - 1, N_M^K(0) = 1$  and we consider  $x_R^* = 1$ . So the invasion condition  $F_M(x_R^*, 0) > 0$  reads  $b_M - d_M - (b_R - d_R) \frac{c_{2,1}^2}{c_{1,1}} > 0$ . It is then easy to

compute  $v^*$ , which is either  $x_M^* = \frac{b_M - d_M}{c_{2,2}}$  when  $F_R(0, x_M^*) < 0$  or  $v^* = \frac{(b_M - d_M)c_{1,1} - (b_R - d_R)c_{2,1}}{c_{1,1}c_{2,2} - c_{1,2}c_{2,1}}$  when  $F_R(0, x_M^*) > 0$ .

The branching process  $Z$  is a birth and death process with individual birth rate  $b_M$  and individual death rate  $d_M + c_{2,1}Kx_R^*$ .

## 5.2. SIR model

Let  $\beta > 0$  be the infection rate (per pair of individuals) in a mixed population and  $\gamma > 0$  the individual recovery rate. Let  $K \geq 1$  be the total population size. At time  $t$ , the number of susceptible individuals is denoted by  $S^K(t)$ , the number of infected individuals is denoted by  $I^K(t)$  and  $K - S^K(t) - I^K(t)$  yields the number of recovered individuals. Each susceptible becomes infected at time  $t$  with rate  $\beta I^K(t)/K$ . Starting from one infected individual, the process  $N^K = (N_R^K, N_M^K) = (S^K, I^K)$  is the unique strong solution of

$$\begin{aligned} S^K(t) &= K - 1 - \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq \beta S^K(s-) I^K(s-)/K\}} \mathcal{N}_I(ds, du), \\ I^K(t) &= 1 + \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq \beta S^K(s-) I^K(s-)/K\}} \mathcal{N}_I(ds, du) - \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq \gamma I^K(s-)\}} \mathcal{N}_G(ds, du), \end{aligned}$$

where  $\mathcal{N}_I$  and  $\mathcal{N}_G$  are Poisson point measure on  $\mathbb{R}_+^2$  with intensity  $dsdu$ .

The process  $N^K$  is a bitype birth and death process as considered in Introduction with  $\mathcal{N}_I = \mathcal{N}_M^b = \mathcal{N}_R^d$  and birth and death rates defined by

$$\begin{aligned} b_R(x_R; x_M) &= 0, & d_R(x_R; x_M) &= \beta x_M \\ b_M(x_R; x_M) &= \beta x_R, & d_M(x_R; x_M) &= \gamma. \end{aligned}$$

The ODE describing the limit of  $N^K/K$  is given by (1.2) with

$$F_R(x_R, x_M) = -\beta x_M, \quad F_M(x_R, x_M) = \beta x_R - \gamma.$$

Let us check that we can apply our results and that it corresponds to the case where the first derivative of  $F_M$  with respect to  $x_M$  cancels. First, we observe that regularity of Assumption 1.1 is satisfied. Second, for any  $x_R > 0$ ,  $(x_R, 0)$  is an equilibrium and

$$\partial_R F_R(x_R, 0) = 0.$$

This means that the second part of Assumption 1.1 is satisfied for any  $x_R > 0$ . Moreover, no geometric stability for the resident population alone, which correspond to the second case considered in the branching process approximation.

Finally, we start from a population with only one mutant,  $S^K(0) = N_R^K(0) = K - 1, I^K(0) = N_M^K(0) = 1$

and we consider  $x_R^* = 1$ . So the invasion condition  $F_M(x_R^*, 0) > 0$  reads  $\beta > \gamma$ . the branching process  $Z$  is a birth and death process with individual birth rate  $\beta$  and individual death rate  $\gamma$ .

In this example, we can also determine the maximal value that the invasive population can reach, which is the pic of the epidemic :

$$v_* = 1 - \frac{\gamma}{\beta} + \frac{\gamma}{\beta} \log\left(\frac{\gamma}{\beta}\right).$$

We obtain the following results when there is an outbreak of the epidemics, where the hitting times for the infected population is defined by

$$\tau_I^K(n) = \inf\{t \geq 0, I^K(t) \geq n\}.$$

**Proposition 5.1.** *If  $\beta > \gamma$ , then*

*i) for any  $\xi_K \ll \log(K)/K^2$  and  $\eta > 0$ ,*

$$\lim_{K \rightarrow +\infty} \mathbb{P}\left(\sup_{t \leq \tau_M^K(\xi_K)} \left| \frac{I^K(t)}{I(t)} - 1 \right| > \eta; I(\tau_I^K(\xi_K)) > 0\right) = 0,$$

where  $I$  is the unique strong solution of

$$I(t) = 1 + \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq \beta I(s-)\}} \mathcal{N}_I(ds, du) - \int_{[0,t] \times \mathbb{R}_+} \mathbf{1}_{\{u \leq \gamma I(s-)\}} \mathcal{N}_G(ds, du).$$

*ii) For any  $v < v_*$ , on the survival event  $\{\forall t > 0, I(t) > 0\}$ ,*

$$\lim_{T \rightarrow \infty} \limsup_{K \rightarrow \infty} \sup_{t \in [T, \tau_I^K(vK)]} \left| \frac{I^K(t)}{K x_I^{K,T}(t)} - 1 \right| = 0, \text{ in probability,}$$

where  $(x_S^{K,T}(t), x_I^{K,T}(t))$  is the solution of

$$\begin{cases} \frac{d}{dt} x_S^{K,T}(t) &= -\beta x_S^{K,T}(t) x_I^{K,T}(t), \\ \frac{d}{dt} x_I^{K,T}(t) &= (\beta x_S^{K,T}(t) - \gamma) x_I^{K,T}(t). \end{cases}$$

such that  $(x_S^{K,T}(T), x_I^{K,T}(T)) = (S^K(T)/K, I^K(T)/K)$ .

The point *i*) is a consequence of Theorem 2.3 and point *ii*) comes from Theorem 3.6. These two parts have an intersection : in the time window when the number of infected is large but negligible compared to  $\log(K)/K^2$ , we have informally

$$I_M^K(t) \sim I(t) \sim W e^{(\beta-\gamma)t} \sim x_M^{K,T}(t - T).$$

We obtain also an approximation of the time when the epidemics reaches a given level :

**Corollary 5.2.** *Let  $(\xi_K)_K$  be sequence which tends to infinity and satisfies  $\xi_K/K \rightarrow v$  as  $K \rightarrow \infty$ , for some  $v \in [0, v_*)$ . Then, for any  $\varepsilon > 0$ ,*

$$\lim_{K \rightarrow \infty} \mathbb{P}\left(\left| \tau_M^K(\xi_K) - \frac{\log(\xi_K/W)}{r_*} - \tau(v) \right| \geq \varepsilon, W > 0\right) = 0,$$

where  $\tau$  is an increasing continuous function such that  $\tau(0) = 0$  defined in Proposition 3.2.

## 6. Appendix

### 6.1. Minimum of the martingale of branching process

Recalling notation Section 2, we are interested in the minimal value of the martingale  $W(t) := Z(t) \exp(-r_* t)$  for positive time. Such object for simple branching processes has already attracted lot of attention. We are not aware in the literature of the estimates given here and provide the proof for completeness. This relies on classical  $L^2$  estimates to control the speed of the convergence of the martingale.

**Lemma 6.1.** *For any  $\varepsilon > 0$ , and any  $Z$ -adapted stopping time  $\tau$ ,*

$$\mathbb{P}(W(\tau) \leq \varepsilon; W > 0) \leq C \left( \varepsilon + \varepsilon^{-1} \mathbb{E} \left[ \mathbf{1}_{\{\tau < \infty\}} e^{-r_* \tau / 2} \right] \right),$$

where  $W$  is the almost sure limit of  $W(t)$  as  $t$  goes to infinity, and  $C$  is some positive constant independent of  $\varepsilon$  and  $\tau$ . We use the convention  $W(+\infty) := W$  for the inequality above to make sense.

*Proof.* To simplify the proof, we firstly treat the case where  $\tau$  is deterministic and finite. To begin with,

$$W(\tau + 1) - W(\tau) = e^{-r_* \tau} \left( \sum_{i=1}^{Z(\tau)} N_i e^{-r_*} - Z(\tau) \right),$$

where, conditionally on  $Z(\tau)$ , the  $N_i$  are i.i.d. with the same distribution as  $Z(1)$ . This implies that

$$\begin{aligned} \mathbb{E} \left[ (W(\tau + 1) - W(\tau))^2 \right] &= e^{-2r_* \tau} \mathbb{E} \left[ \mathbb{V} \left[ \sum_{i=1}^{Z(\tau)} (N_i e^{-r_*} - 1) \mid Z(\tau) \right] \right] \\ &= e^{-2r_* \tau} \mathbb{E} [Z(\tau)] \mathbb{V} [N_1 e^{-r_*} - 1] \leq C e^{-r_* \tau}. \end{aligned} \quad (6.1)$$

On the other hand, we write

$$\begin{aligned} \mathbb{P}(W(\tau) \leq \varepsilon; W > 0) &\leq \mathbb{P}(W \leq 2\varepsilon) + \mathbb{P}(W(\tau) \leq \varepsilon; W > 2\varepsilon) \\ &\leq \mathbb{P}(W \leq 2\varepsilon; W > 0) + \mathbb{P}(W - W(\tau) > \varepsilon). \end{aligned}$$

Recalling that, on the event  $\{W > 0\}$ ,  $W$  follows an exponential distribution, we can bound the first term of the sum above by  $1 - e^{-C\varepsilon} \leq C\varepsilon$ .

The second term of the sum can be handled as follows:

$$\begin{aligned} \mathbb{P}(W - W(\tau) > \varepsilon) &= \mathbb{P} \left( \sum_{n \geq 0} W(\tau + n + 1) - W(\tau + n) > \varepsilon \right) \\ &\leq \frac{1}{\varepsilon} \sum_{n \geq 0} \mathbb{E} [|W(\tau + n + 1) - W(\tau + n)|] \\ &\leq \varepsilon^{-1} \sum_{n \geq 0} \mathbb{E} \left[ |W(\tau + n + 1) - W(\tau + n)|^2 \right]^{1/2} \\ &\leq C \varepsilon^{-1} e^{-r_* \tau / 2} \sum_{n \geq 0} e^{-r_* n / 2} \leq C \varepsilon^{-1} e^{-r_* \tau / 2}. \end{aligned}$$

This ends the proof in the case where  $\tau$  is deterministic.

Now to prove the result when  $\tau$  is some almost surely finite  $Z$ -adapted stopping time, we just have to do the same computation conditionally on  $\mathcal{F}_\tau$ . Indeed, since  $Z(\tau)$  is  $\mathcal{F}_\tau$ -measurable and  $Z(\tau + 1) - Z(\tau)$  independent of  $\mathcal{F}_\tau$ , (6.1) becomes

$$\mathbb{E} \left[ (W(\tau + 1) - W(\tau))^2 \mid \mathcal{F}_\tau \right] \leq C e^{-r_* \tau},$$

and the last computation of the previous case gives

$$\mathbb{P}(W - W(\tau) > \varepsilon | \mathcal{F}_\tau) \leq C\varepsilon^{-1}e^{-r_*\tau/2}.$$

The result of the lemma is then proved by taking the expectation in the inequality above.

Finally let us treat the case where  $\tau$  is not necessary almost surely finite. We write

$$\begin{aligned} \mathbb{P}(W(\tau) \leq \varepsilon; W > 0) &= \mathbb{P}(W(\tau) \leq \varepsilon; W > 0; \tau < \infty) + \mathbb{P}(W(\tau) \leq \varepsilon; W > 0; \tau = \infty) \\ &\leq \mathbb{P}(W(\tau) \leq \varepsilon; W > 0; \tau < \infty) + \mathbb{P}(0 < W \leq \varepsilon) \end{aligned} \quad (6.2)$$

Note that, on the event  $\{\tau < \infty\}$ , we have the almost sure convergence

$$\mathbf{1}_{\{W(\tau \wedge n) \leq \varepsilon; W > 0\}} \xrightarrow{n \rightarrow \infty} \mathbf{1}_{\{W(\tau) \leq \varepsilon; W > 0\}}.$$

Whence, by Fatou's lemma,

$$\begin{aligned} \mathbb{P}(W(\tau) \leq \varepsilon; W > 0; \tau < \infty) &\leq \liminf_n \mathbb{P}(W(\tau \wedge n) \leq \varepsilon; W > 0; \tau < \infty) \\ &\leq \liminf_n \mathbb{P}(W(\tau \wedge n) \leq \varepsilon; W > 0) \end{aligned}$$

Then, applying the result of the Lemma with the almost surely finite stopping time  $\tau \wedge n$ , we obtain

$$\mathbb{P}(W(\tau) \leq \varepsilon; W > 0; \tau < \infty) \leq C \left( \varepsilon + \varepsilon^{-1} \liminf_n \mathbb{E} \left[ e^{-r_*(\tau \wedge n)/2} \right] \right).$$

Noticing that

$$e^{-r_*(\tau \wedge n)/2} \leq \mathbf{1}_{\{\tau < \infty\}} e^{-r_*\tau/2} + e^{-r_*n/2},$$

we have

$$\mathbb{P}(W(\tau) \leq \varepsilon; W > 0; \tau < \infty) \leq C \left( \varepsilon + \varepsilon^{-1} \mathbb{E} \left[ \mathbf{1}_{\{\tau < \infty\}} e^{-r_*\tau/2} \right] \right).$$

In addition, to control the second term of the sum in (6.2), let us recall that the distribution of  $W$  is

$$\frac{d_*}{b_*} \delta_0 + \lambda \mathcal{E}(\lambda),$$

with  $\lambda := (b_* - d_*)/d_*$ . Hence

$$\mathbb{P}(0 < W \leq \varepsilon) = \lambda (1 - e^{-\lambda\varepsilon}) \leq \lambda^2 \varepsilon.$$

Recalling (6.2), the result of the lemma is proved.  $\square$

**Lemma 6.2.** *With the same notation as in Lemma 6.1, for any  $0 < \varepsilon < 1$ ,*

$$\mathbb{P} \left( \inf_{t \geq 0} W(t) \leq \varepsilon; W > 0 \right) \leq C\varepsilon^{1/4},$$

for some positive constant  $C$  independent of  $\varepsilon$ .

*Proof.* Let us introduce

$$s_\varepsilon := \frac{1}{r_*} \log \left( \frac{1}{\varepsilon} \right).$$

On the event  $\{W > 0\}$ , we have  $Z(t) \geq 1$  for all  $t \geq 0$ . And, for  $t < s_\varepsilon$ ,  $e^{-r_*t} > \varepsilon$ . As a consequence

$$\left\{ \inf_{t \geq 0} W(t) \leq \varepsilon; W > 0 \right\} \subseteq \left\{ \inf_{t \geq s_\varepsilon} W(t) \leq \varepsilon; W > 0 \right\}.$$

For some  $\lambda > 1$  whose value will be fixed later, let us note

$$S_\varepsilon := \inf \left\{ t \geq s_\varepsilon : W(t) \leq \varepsilon^{1/\lambda} \right\}.$$

Since  $\lambda > 1$ , we have  $\varepsilon < \varepsilon^{1/\lambda}$ , hence

$$\left\{ \inf_{t \geq s_\varepsilon} W(t) \leq \varepsilon \right\} \subseteq \{S_\varepsilon < \infty\} \subseteq \{W(S_\varepsilon) \leq \varepsilon^{1/\lambda}\}.$$

Note that the last event above is trivially satisfied only if  $S_\varepsilon < \infty$ .

Then, thanks to Lemma 6.1,

$$\mathbb{P} \left( W(S_\varepsilon) \leq \varepsilon^{1/\lambda}; W > 0 \right) \leq C \left( \varepsilon^{1/\lambda} + \varepsilon^{-1/\lambda} \mathbb{E} \left[ \mathbf{1}_{\{S_\varepsilon < \infty\}} e^{-r_* S_\varepsilon / 2} \right] \right).$$

Recall that, by definition of  $S_\varepsilon$ ,  $S_\varepsilon \geq s_\varepsilon$ , and so

$$\mathbb{E} \left[ \mathbf{1}_{\{S_\varepsilon < \infty\}} e^{-r_* S_\varepsilon / 2} \right] \leq e^{-r_* s_\varepsilon / 2} = \varepsilon^{1/2},$$

implying that

$$\mathbb{P} \left( W(S_\varepsilon) \leq \varepsilon^{1/\lambda}; W > 0 \right) \leq C \left( \varepsilon^{1/\lambda} + \varepsilon^{1/2-1/\lambda} \right).$$

Finally, choosing  $\lambda := 4$  to optimize the bound above proves the result.  $\square$

## 6.2. Stability of pertubated dynamical system

The following lemma is strongly inspired from Lemmas 3.1 and 3.2 of [24]. We consider an exponential stable dynamical system :  $y' \leq -cy$ , with  $c > 0$ . We add a source term, which acts a perturbation given by  $h$ . The following result allows to control the value of the dynamical system in terms of fluctuations of the source term  $h$  on a unit time interval.

**Lemma 6.3.** *Assume that there are three measurable and locally bounded functions  $y, h, \phi$  with  $h(0) = 0$  and satisfying that for all  $t \geq 0$ ,*

$$y(t) = y(0) + h(t) + \int_0^t y(s)\phi(s)ds.$$

*Suppose in addition that  $\phi$  is bounded, upper-bounded by some negative and*

$$-\infty < \inf \phi \leq \sup \phi < 0.$$

*Then, for all  $t \geq 0$ ,*

$$\sup_{s \leq t} |y(s)| \leq \Gamma \left( |y(0)| \vee \sup_{s \leq t} |h(s) - h(\lfloor s \rfloor)| \right),$$

*for some positive constant  $\Gamma$  that depends only on  $\phi$ .*

*Proof.* To begin with, we introduce

$$\Phi(t, s) = e^{\int_s^t \phi(r)dr} \leq e^{-C(t-s)},$$

for some positive constant  $C := -\sup \phi > 0$ . Then

$$\begin{aligned} \Phi(t, t) &= 1; \quad \Phi(t, s)\Phi(s, u) = \Phi(t, u), \\ \partial_t \Phi(t, s) &= \Phi(t, s)\phi(t); \quad \partial_s \Phi(t, s) = -\Phi(t, s)\phi(s). \end{aligned}$$

By a standard constant variation argument, one can write

$$y(t) = \Phi(t, 0)y(0) + h(t) + \int_0^t \Phi(t, s)h(s)\phi(s)ds.$$

Then it results the following decomposition:

$$\begin{aligned} y(t) = & \Phi(t, 0)y(0) + \sum_{j=1}^{\lfloor t \rfloor} \Phi(t, j) \left( h(j) - h(j-1) + \int_{j-1}^j \Phi(j, s)\phi(s) (h(s) - h(j-1)) ds \right) \\ & + h(t) - h(\lfloor t \rfloor) + \int_{\lfloor t \rfloor}^t \Phi(t, s)\phi(s) (h(s) - h(\lfloor t \rfloor)) ds. \end{aligned}$$

Using the upperbound of  $\phi$ , we obtain

$$|y(t)| \leq e^{-Ct}|y(0)| + \sum_{j=1}^{\lfloor t \rfloor} e^{-C(t-j)} \left( \sup_{j-1 \leq s \leq j} |h(s) - h(j-1)| \Gamma' \right) + \sup_{\lfloor t \rfloor \leq s \leq t} |h(s) - h(\lfloor t \rfloor)| \Gamma',$$

with  $\Gamma' := 1 + \|\phi\|_\infty$ . The result is proved by choosing  $\Gamma = \Gamma' \sum_{j=0}^\infty e^{-Cj} = \Gamma'/(1 - e^{-C})$ .  $\square$

*Proof of Lemma 3.3.* To prove the result, it is in fact sufficient to prove that

$$\sup_{t \leq s_\varepsilon} |\phi_R(t, x) - x_R^*| \leq \sqrt{2\varepsilon}, \quad (6.3)$$

with

$$s_\varepsilon := \tau_\varepsilon^M(x) \wedge \left( \inf\{t > 0 : |\phi_R(t, x) - x_R^*| > \varepsilon^{1/4}\} \right).$$

Indeed, if the inequality (6.3) is true, it implies that

$$\tau_\varepsilon^M(x) \leq \inf\{t > 0 : |\phi_R(t, x) - x_R^*| > \varepsilon^{1/4}\},$$

whence  $s_\varepsilon = \tau_\varepsilon^M(x)$ .

For the sake of notation, we introduce

$$y_R(t) := \phi_R(t, x) - x_R.$$

*Step 1.* In this first step, we prove (6.3) assuming that  $\partial_R F_R(x^*) < 0$ . It allows to assume that

$$\sup_{t \leq s_\varepsilon} \partial_R F_R(\phi(t, x)) < 0. \quad (6.4)$$

Let us write

$$\begin{aligned} y_R(t) = & \int_0^t \phi_R(s, x) F_R(x_R, \phi_M(s, x)) ds \\ & + \int_0^t \phi_R(s, x) [F_R(\phi(s, x)) - F_R(x_R, \phi_M(s, x)) - y_R(s) \partial_R F_R(x_R, \phi_M(s, x))] ds \\ & + \int_0^t \phi_R(s, x) y_R(s) \partial_R F_R(x_R, \phi_M(s, x)) ds \\ = & x_R + H^1(t) + H^2(t) + \int_0^t \phi_R(s, x) y_R(s) \partial_R F_R(x_R, \phi_M(s, x)) ds, \end{aligned}$$

where

$$\begin{cases} H^1(t) := \int_0^t \phi_R(s, x) F_R(x_R, \phi_M(s, x)) ds, \\ H^2(t) := \int_0^t \phi_R(s, x) [F_R(\phi(s, x)) - F_R(x_R, \phi_M(s, x)) - y_R(s) \partial_R F_R(x_R, \phi_M(s, x))] ds. \end{cases} \quad (6.5)$$

Thanks to (6.4) and Lemma 6.3, we have for any  $T > 0$ ,

$$\begin{aligned} \sup_{t \leq T \wedge s_\varepsilon} |y_R(t)| &\leq C \left( \sup_{t \leq s_\varepsilon \wedge T} |H^1(t) - H^1(\lfloor t \rfloor)| + \sup_{t \leq s_\varepsilon \wedge T} |H^2(t) - H^2(\lfloor t \rfloor)| \right) \\ &\leq C |x_R(0) - x_R^*| + C\varepsilon + C \sup_{t \leq s_\varepsilon \wedge T} |y_R(t)|^2, \end{aligned} \quad (6.6)$$

for some positive constant  $C$  independent of  $x$  and  $\varepsilon$ . Let us note

$$g(T) := \sup_{t \leq T \wedge s_\varepsilon} |y_R(t)|.$$

We have, for all  $T > 0$ ,

$$g(T) \leq 2C\varepsilon + Cg(T)^2.$$

Let us now end the proof by contradiction. Assume that

$$s_1 := \inf\{T > 0 : g(T) > \sqrt{\varepsilon}\}$$

is finite. Then since  $g$  is continuous  $\sqrt{\varepsilon} \leq 2C\varepsilon + C(\sqrt{\varepsilon})^2 = 3C\varepsilon$ , which is absurd since  $\varepsilon$  can be chosen arbitrarily small. This proves (6.3).

*Step 2.* Now we prove the same result (6.3) under the assumption  $\partial_R F_R(x^*) = 0$ . The same computation as in Step 1 allows to obtain

$$\begin{aligned} y'_R(t) &= \phi_R(t, x) F_R(x_R, \phi_M(t, x)) \\ &\quad + \phi_R(t, x) [F_R(\phi(t, x)) - F_R(x_R, \phi_M(t, x)) - y_R(t) \partial_R F_R(x_R, \phi_M(t, x))] \\ &\quad + \phi_R(t, x) y_R(t) \partial_R F_R(x_R, \phi_M(t, x)). \end{aligned}$$

Hence, for  $t \leq s_\varepsilon$ ,

$$\begin{aligned} |y'_R(t)| &\leq C\phi_M(t, x) + C|x_R - x_R^*| + C(y_R(t))^2 \\ &\leq C \left[ (|x_R - x_R^*| + \phi_M(t, x)) \vee (y_R(t))^2 \right]. \end{aligned}$$

Since  $t \mapsto \phi_M(t, x)$  is increasing on  $[0, s_\varepsilon]$ , we have for  $t \leq s_\varepsilon$ ,

$$\sup_{s \leq t} |y'_R(s)| \leq C \left[ (|x_R - x_R^*| + \phi_M(t, x)) \vee \sup_{s \leq t} y_R(s)^2 \right]. \quad (6.7)$$

Now, let us introduce

$$\mathcal{T} := \left\{ 0 \leq t \leq s_\varepsilon : |x_R - x_R^*| + \phi_M(t, x) \geq \sup_{s \leq t} y_R(s)^2 \right\}.$$

If  $s_\varepsilon \in \mathcal{T}$ , then

$$\sup_{s \leq s_\varepsilon} y_R(s)^2 \leq |x_R - x_R^*| + \phi_M(s_\varepsilon, x) \leq 2\varepsilon,$$

and (6.3) is true. Otherwise, let  $t_1 := \sup \mathcal{T}$  and note that  $\mathcal{T}$  cannot be empty since  $y_R(0) = 0$ . Since  $\mathcal{T}$  is closed, we know that  $t_1 < s_\varepsilon$  and that

$$|x_R - x_R^*| + \phi_M(t_1, x) = \sup_{s \leq t_1} y_R(s)^2.$$

Let us remark that for any  $r \leq t_1 < s \leq s_\varepsilon$ ,

$$|y_R(r)| \leq \sup_{t \leq t_1} |y_R(t)| = \sqrt{|x_R - x_R^*| + \phi_M(t_1, x)} < \sqrt{|x_R - x_R^*| + \phi_M(s, x)} < \sup_{t \leq s} |y_R(t)|.$$

Using that  $y_R$  is continuous, it ensures that the maximal value of  $y_R$  on the time interval  $[s, t]$  is reached on  $[t_1, t]$ . Writing

$$g(t) := \sup_{t_1 \leq s \leq t} |y_R(s)|,$$

and using also (6.7), we get for  $t \geq t_1$ ,

$$g(t) \leq |y_R(t_1)| + \int_{t_1}^t |y_R'(s)| ds \leq |y_R(t_1)| + C \int_0^t g(s)^2 ds$$

According to Osgood's lemma (see e.g. Lemma 3.4 of [3]), for all  $t_1 \leq t \leq s_\varepsilon$ ,

$$\frac{1}{|y_R(t_1)|} - \frac{1}{g(t)} \leq C(t - t_1),$$

and

$$\frac{1}{g(t)} \geq \frac{1}{|y_R(t_1)|} - C(t - t_1).$$

We have that  $\frac{1}{|y_R(t_1)|} - C(t - t_1) > 0$ , because

$$\begin{aligned} |y_R(t_1)|C(t - t_1) &= |x_R - x_R^*|C(t - t_1) + \phi_M(t_1, x)C(t - t_1) \\ &\leq C|x_R - x_R^*|s_\varepsilon + C \int_{t_1}^t \phi_M(s, x) ds. \end{aligned} \quad (6.8)$$

Since we work on  $[0, s_\varepsilon]$ , we can choose  $\varepsilon$  small enough such that  $s \leq s_\varepsilon$ ,

$$F_M(x) \geq r_*/2 \text{ and } F_M(\phi(s, x)) \geq r_*/2.$$

Then it is possible to control the second term in the sum of the RHS (6.8) in the following way

$$\int_{t_1}^t \phi_M(s, x) ds = \int_0^{s_\varepsilon} \partial_s \phi_M(s, x) \frac{1}{F_M(\phi(s, x))} ds \leq C \int_0^{\tau_\varepsilon^M(x)} \partial_s \phi_M(s, x) ds \leq C\varepsilon.$$

And, to control the first term of the sum (6.8), we can notice that, on  $[0, s_\varepsilon]$ ,

$$\partial_t \phi_M(t, x) \geq \frac{r_*}{2} \phi_M(t, x),$$

whence

$$s_\varepsilon \leq \frac{2}{r_*} \log(1/x_M) + \frac{2}{r_*} \log(x_M(s_\varepsilon)) \leq \frac{2}{r_*} \log(1/x_M) + \frac{2}{r_*} \log \varepsilon \leq \frac{2}{r_*} \log(1/x_M), \quad (6.9)$$

such that

$$|x_R - x_R^*|s_\varepsilon \leq \frac{2}{r_*} |x_R - x_R^*| \log(1/x_M) \leq C\eta_0.$$



Hence

$$|y_R(t_1)|C(t-t_1) \leq C|x_R(0) - x_R^*|_{s_\varepsilon} + \int_0^{s_\varepsilon} \phi_M(s, x) ds \leq C\eta_0 + C\varepsilon \leq \frac{1}{2}.$$

Then, for all  $t_1 \leq t \leq s_\varepsilon$ ,

$$g(t) \leq \frac{1}{\frac{1}{|y_R(t_1)|} - C(t-t_1)} = \frac{|y_R(t_1)|}{1 - |y_R(t_1)|C(t-t_1)} \leq \frac{\sqrt{|x_R - x_R^*| + \phi_M(t_1, x)}}{1 - 1/2} \leq 2\sqrt{2\varepsilon}.$$

Recalling that  $g(t) := \sup_{t_1 \leq t} |y_R(t)|$ , we have proved (6.3).  $\square$

*Proof of Lemma 3.4.* To simplify the notation, let us write

$$z(t) := \sup_{s \leq t} |\phi_R(s, x) - x_R^*|.$$

*Step 1.* Firstly, let us prove the result under the assumption  $\partial_R F_R(x^*) < 0$ . Using the computation of (6.6) from Step 1 in the proof of Lemma 3.3 above, we have for all  $t \leq \tau_\varepsilon^M(x)$ ,

$$\begin{aligned} z(t) &\leq C|x_R(0) - x_R^*| + C \int_{[t]}^t \phi_M(x, s) ds + C \int_{[t]}^t z(s)^2 ds \\ &\leq C|x_R - x_R^*| + C\phi_M(x, t) + Cz(t)^2, \end{aligned}$$

since the function  $t \mapsto \phi_M(x, t)$  and  $t \mapsto z(t)$  are non-decreasing on  $[0, \tau_\varepsilon^M(x)]$ . As a consequence, for all  $t \leq \tau_\varepsilon^M(x)$ ,

$$\begin{aligned} \int_0^t z(s) ds &\leq Ct|x_R(0) - x_R^*| + C \int_0^t \phi_M(x(0), s) ds + C \int_0^t z(s)^2 ds \\ &\leq C|x_R(0) - x_R^*|\tau_\varepsilon^M(x) + C \int_0^{\tau_\varepsilon^M(x)} \phi_M(x(0), s) ds + Cz(t) \int_0^t z(s) ds. \end{aligned}$$

Then, by Lemma 3.3, we have  $z(\tau_\varepsilon^M(x)) \leq C\sqrt{\varepsilon}$  and  $\tau_\varepsilon^M(x) \leq 2 \log(1/x_M)/r_*$ , we have

$$\int_0^t z(s) ds \leq C\eta + C\varepsilon + C\sqrt{\varepsilon} \int_0^t z(s) ds.$$

Wence, for  $\varepsilon$  small enough (such that  $C\sqrt{\varepsilon} \leq 1/2$  for example),

$$\int_0^t z(s) ds \leq C\eta + C\varepsilon.$$

*Step 2.* Let us prove the same result under the assumption  $\partial_R F_R(x^*) = 0$ . We have proved in Step 2 of the proof of Lemma 3.3 above the following inequality: for all  $t \leq \tau_\varepsilon^M(x)$ ,

$$\begin{aligned} z(t) &\leq C|x_R - x_R^*| + C \int_0^t \phi_M(x, s) ds + C \int_0^t z(s)^2 ds \\ &\leq C|x_R - x_R^*| + C\phi_M(x, t) + C \int_0^t z(s)^2 ds. \end{aligned}$$

Hence, for all  $T \leq \tau_\varepsilon^M(x)$ ,

$$\begin{aligned} \int_0^T z(t) dt &\leq CT|x_R - x_R^*| + C \int_0^T \phi_M(x, s) ds + C \int_0^T \int_0^t z(s)^2 ds dt \\ &\leq C|x_R - x_R^*| \ln(1/x_M) + C\varepsilon + C \int_0^T z(t) \int_0^t z(s) ds dt \\ &\leq C\eta + C\varepsilon + \left( \int_0^T z(t) dt \right)^2. \end{aligned}$$

The exact same reasoning as the one used at the very end of Step 1 in the proof of Lemma 3.3 allows us to conclude that, for all  $T \leq \tau_\varepsilon^M(x)$ ,

$$\int_0^T z(t)dt \leq \sqrt{\varepsilon + \eta},$$

what ends the proof. □

**Acknowledgements:** This work has been supported by the Chair “Modélisation Mathématique et Biodiversité of Veolia Environnement-Ecole Polytechnique-Museum national d’Histoire naturelle-Fondation X and by ITMO Cancer of Aviesan within the framework of the 2021-2023 Cancer Control Strategy, on funds administrated by Inserm.

## References

- [1] K. B. Athreya, P. E. Ney (2004). *Branching processes*. Dover Publications, Inc., Mineola, NY.
- [2] M. Baar, A. Bovier, N. Champagnat. From stochastic, individual-based models to the canonical equation of adaptive dynamics in one step. *Ann. Appl. Probab.* 27 (2017), no. 2.
- [3] H. Bahouri, J. Chemin and R. Danchin. *Fourier Analysis and Nonlinear Partial Differential Equations* Springer, Grundlehren der mathematischen Wissenschaften, 2011.
- [4] F. Ball, P. Donnelly. Strong approximations for epidemic models. *Stoc. Proc. Appl.* Vol. 55, 1995, Pages 1-21
- [5] A. D. Barbour, G. Reinert. Approximating the epidemic curve. *Electron. J. Probab.* 18 (2013), no. 54, 1-30.
- [6] A.D. Barbour, S. Utev. Approximating the Reed-Frost epidemic process. *Stoc. Proc. Appl.* Vol. 113 (2004) p 173-197.
- [7] S. Billiard, R. Ferrière, S. Méléard, V.C. Tran. Stochastic dynamics of adaptive trait and neutral marker driven by eco-evolutionary feedbacks. *J. Math. Biol.* 71 (2015), no. 5.
- [8] S. Billiard, C. Smadi. The interplay of two mutations in a population of varying size: a stochastic eco-evolutionary model for clonal interference. *Stochastic Process. Appl.* 127 (2017), no. 3, 701-748.
- [9] J. Blath, T. Paul and A. Tóbiás. A Stochastic Adaptive Dynamics Model for Bacterial Populations with Mutation, Dormancy and Transfer arXiv:2105.09228 (2021).
- [10] C. Bonnet, P. Gou, S. Girel, V. Bansaye, C. Lacout, K. Bailly, M.H. Schlägelter, E. Lauret, S. Méléard, S. Giraudier. Multistage hematopoietic stem cell regulation in the mouse: a combined biological and mathematical approach. *iScience*, 2021 Nov 6; 24 (12):103399. Open access.
- [11] C. Bonnet, S. Méléard. Large fluctuations in multi-scale modeling for rest erythropoiesis. *J. Math. Biol.* 82 (2021), no. 6, Paper No. 58.
- [12] A. Bovier, L. Coquille, R. Neukirch The recovery of a recessive allele in a Mendelian diploid model. *J. Math. Biol.* 77 (2018), no. 4.
- [13] A. Bovier, L. Coquille, C. Smadi. Crossing a fitness valley as a metastable transition in a stochastic population model. *Ann. Appl. Probab.* 29 no 6 (2019), 3541–3589.
- [14] N. Champagnat. A microscopic interpretation for adaptative dynamics trait substitution sequence models. *Stochastic Processes and their Applications*, 116 (2006), 1127–1160.
- [15] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probability Theory and Related Fields*, 151 no 1-2 (2011), 45–94.
- [16] L. Coquille, A. Kraut, C. Smadi. Stochastic individual-based models with power law mutation rate on a general finite trait space. *Electron. J. Probab.* 26(123), 37 p., 2021.
- [17] O. Diekmann, P.-E. Jabin, S. Mischler, and B. Perthame. The dynamics of adaptation: an illuminating example and a Hamilton-Jacobi approach. *Theoretical Population Biology*, 67 ( 2005), 257–271.

- [18] R. Durrett. Branching process models of Cancer. Mathematical Biosciences Institute Lecture Series 1.1, Stochastics in Biological Systems, Springer, 2015.
- [19] S.N. Ethier, T.G. Kurtz. Markov processes, Characterization and Convergence. John Wiley & Sons, New York, 1986
- [20] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer-Verlag, Berlin, 1987.
- [21] A. Jakubowski. On the Skorokhod topology. *Annales de l'Institut Henri Poincaré*, 22 no 3 (1986), 263–285.
- [22] A. Joffe and M. Métivier. Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, 18 (1986), 20–65.
- [23] J.A.J. Metz, S.A.H. Geritz, G. Meszéna, F.A.J. Jacobs, and J.S. Van Heerwaarden. Adaptative dynamics, a geometrical study of the consequences of nearly faithful reproduction. *S.J. Van Strien & S.M. Verduyn Lunel (ed.), Stochastic and Spatial Structures of Dynamical Systems*, 45, 183–231, 1996.
- [24] A. Prodhomme. Strong Gaussian approximation of metastable density-dependent Markov chains on large time scales. arXiv:2010.06861 (2020).