The Annals of Applied Probability
2020, Vol. 30, No. 6, 2880–2922
https://doi.org/10.1214/20-AAP1576

EXTINCTION TIME FOR THE WEAKER OF TWO COMPETING SIS EPIDEMICS

BY FABIO LOPES[1] AND MALWINA LUCZAK[2]

[1]*Departamento de Matemáticas, Universidad Tecnológica Metropolitana, f.marcellus@utem.cl*
[2]*School of Mathematics and Statistics, University of Melbourne, mluczak@unimelb.edu.au*

We consider a simple Markov model for the spread of a disease caused by two virus strains in a closed homogeneously mixing population of size N. The spread of each strain in the absence of the other one is described by the stochastic SIS logistic epidemic process, and we assume that there is perfect cross-immunity between the two strains, that is, individuals infected by one are temporarily immune to re-infections and infections by the other. For the case where one strain is strictly stronger than the other, and the stronger strain on its own is supercritical, we derive precise asymptotic results for the distribution of the time when the weaker strain disappears from the population. We further extend our results to certain parameter values where the difference between the basic reproductive ratios of the two strains may tend to 0 as $N \to \infty$.

In our proofs, we illustrate a new approach to a fluid limit approximation for a sequence of Markov chains in the vicinity of a stable fixed point of the limit differential equation, valid over long time intervals.

1. Introduction. Mathematical models of epidemics provide important tools for understanding the spread of many diseases relevant to public health, and may help health organizations develop measures to prevent and manage epidemic outbreaks, as well as control the emergence of new infections.

The vast majority of mathematical models of epidemics view infectious diseases as caused by a single pathogen strain; such models tend to be more tractable but inappropriate for predicting the long-term evolutionary dynamics of pathogen populations, see Humplik et al. [10], or for analysing pathogen infections where host susceptibility may be altered due to infections by other pathogens. For instance, it is known that pathogen strains which are sufficiently antigenically similar may induce a (partial) cross-protective immune response, so that hosts infected by one of the strains may acquire different degrees of temporary or permanent immunity to re-infections and infections by antigenically similar strains. Thus, if a certain closed population of hosts is affected by a particular virus strain and a number of individuals infected by an antigenically similar strain are introduced, then the different pathogen strains may interact as if competing for susceptible individuals in the host population.

A long-standing principle in ecology known as the *competitive exclusion principle* (Levin [14]) predicts that, when species sharing the same ecological niche compete for limited resources, the one with even the slightest advantage will eventually outcompete the others and become dominant. In the context of infectious diseases, for instance, it is shown by Bahl et al. [3] that viral gene flow from Eurasia had led to replacement of endemic avian influenza viruses in North America; moreover, the authors argue that the most likely mechanism for that was competition for susceptible hosts.

Received April 2019; revised March 2020.
MSC2020 subject classifications. 60J27, 92D30.
Key words and phrases. Stochastic SIS logistic epidemic, competing SIS epidemics, time to extinction, near-critical epidemic.

In this work, we consider a simple Markov model for the spread of a disease with stochastic susceptible-infective-susceptible (SIS) dynamics caused by two different virus strains with a perfect cross-protective immune response, so individuals infected by one strain are temporarily immune to re-infections and infections by the other strain. We focus on the case where one of the virus strains has some advantage over its competitor (a higher basic reproductive ratio), and competitive exclusion occurs. Starting with positive but otherwise arbitrary proportions of infected individuals of each virus strain in a large host population, we track the long-term evolution of this process, so as to obtain the distribution of the time until competitive exclusion occurs, that is, the extinction time of the weaker virus strain.

The simplest stochastic model for a disease with SIS dynamics is the stochastic SIS logistic epidemic model. In that model, each individual within the population is either susceptible or infective. We assume a population of size N, and let $\lambda > 0$ denote the infection rate. Each infective individual encounters a uniformly chosen member of the population at rate λ (as is standard, contacts with oneself are allowed); if the encountered individual is susceptible, then he/she becomes infective. Also, each infective individual recovers at rate $\mu > 0$ and, once recovered, becomes susceptible again.

Let $Y_N(t)$ denote the number of infective individuals in the population at time t; then $(Y_N(t))_{t\geq 0}$ is a continuous-time Markov chain on $\{0, 1, \dots, N\}$ with transition rates from state Y given by

$$\begin{aligned} Y &\to Y+1 \quad \text{at rate } \lambda Y(1-Y/N); \\ Y &\to Y-1 \quad \text{at rate } \mu Y. \end{aligned}$$

The extinction time τ_N is defined as $\tau_N = \inf\{t \geq 0 : Y_N(t) = 0\}$, and, since the state space is finite, τ_N is a.s. finite. The following theorem summarises asymptotic results for the distribution of τ_N in the case where the initial epidemic infects a positive proportion of the population, see Andersson and Djehiche [1], Brightwell, House, and Luczak [6], and Foxall [9].

THEOREM 1. *Let $\lambda, \mu, \alpha > 0$, and suppose that $Y_N(0)/N \to \alpha$ as $N \to \infty$.*

(i) (*Supercritical case*; *Andersson and Djehiche* (1998), *Foxall* (2018).) *If $\lambda > \mu$, then, as $N \to \infty$, $\tau_N/\mathbb{E}(\tau_N) \to Z$ in distribution, where Z is an exponential random variable with parameter* 1. *Furthermore*,

$$\mathbb{E}(\tau_N) \sim \sqrt{\frac{2\pi}{N}} \frac{\lambda}{(\lambda-\mu)^2} e^{Nv},$$

as $N \to \infty$, where $v = \log(\lambda/\mu) - 1 + \frac{\mu}{\lambda}$.

(ii) (*Subcritical case*; *Brightwell*, *House and Luczak* (2018).) *If $\lambda < \mu$, then, as $N \to \infty$,*

$$(\mu-\lambda)\tau_N - \bigl\{\log\alpha + \log N + \log(1-\lambda/\mu) - \log\bigl(1+\lambda\alpha/(\mu-\lambda)\bigr)\bigr\} \to G$$

in distribution, where G is a standard Gumbel random variable.

When $\lambda = \mu$ (i.e., the basic reproductive ratio equals 1), for most starting states the time to extinction is of the order $N^{1/2}$, see Nåsell [17]. Brightwell, House and Luczak [6] also consider more general initial conditions, as well as determine the extinction time when $\lambda = \lambda(N)$, $\mu = \mu(N)$ satisfy $\mu - \lambda \to 0$ and $(\mu-\lambda)N^{1/2} \to \infty$ (the barely subcritical case). Foxall [9], Theorem 5, includes the barely supercritical case, where $\lambda = \lambda(N)$, $\mu = \mu(N)$, $\lambda - \mu \to 0$ and $(\lambda-\mu)N^{1/2} \to \infty$, extending the formula of Andersson and Djehiche [1].

The stochastic SIS logistic competition model describes the spread of a disease in a homogeneously mixing population via two different virus strains, say types 1 and 2, which are sufficiently antigenically similar to induce a cross-protective immune response. An individual infected with strain i $(i = 1, 2)$ stays infected for an exponentially distributed time with rate $\mu_i > 0$, and, during the infectious period, each such individual independently makes an infectious contact to a uniformly random individual according to a Poisson process with rate $\lambda_i > 0$; if the contacted individual is currently susceptible, then he/she becomes infected with strain i as a result. The dynamics can thus be described as a two-dimensional continuous-time Markov chain $(X_N(t))_{t\geq 0} = (X_{N,1}(t), X_{N,2}(t))_{t\geq 0}$, where $X_{N,1}(t)$ and $X_{N,2}(t)$ denote the numbers of individuals infected with strains of type 1 and 2 respectively, at time t. The state space is $\{(X_1, X_2)^T : X_1, X_2 \in \mathbb{Z}^+, 0 \leq X_1 + X_2 \leq N\}$, and the transition rates from state (X_1, X_2) can be written as follows:

$$\begin{aligned}
(X_1, X_2) &\to (X_1 + 1, X_2) \quad \text{at rate } \lambda_1 X_1(1 - X_1/N - X_2/N);\\
(X_1, X_2) &\to (X_1, X_2 + 1) \quad \text{at rate } \lambda_2 X_2(1 - X_1/N - X_2/N);\\
(X_1, X_2) &\to (X_1 - 1, X_2) \quad \text{at rate } \mu_1 X_1;\\
(X_1, X_2) &\to (X_1, X_2 - 1) \quad \text{at rate } \mu_2 X_2.
\end{aligned}$$

We note that, in the absence of one of the strains, the other strain evolves according to the basic stochastic SIS logistic epidemic model described above.

Let $R_{0,1} = \lambda_1/\mu_1$ and $R_{0,2} = \lambda_2/\mu_2$ denote the basic reproductive ratios of the two strains. We assume that $R_{0,1} > R_{0,2}$, and $R_{0,1} > 1$, and that $X_{N,1}(0)/N \to \alpha$ and $X_{N,2}(0)/N \to \beta$ as $N \to \infty$ $(0 < \alpha, \beta, \alpha + \beta \leq 1)$. Informally, the assumption $R_{0,1} > R_{0,2}$ means that strain 1 is more infectious than strain 2. Since $R_{0,1} > 1$, Theorem 1(i) implies that the stronger subtype, in the absence of its competitor, would stay endemic in the population for a time that grows exponentially in the size N of the population.

Surprisingly, competing SIS epidemic dynamics was first studied in a spatial context by Neuhauser [18], with infectious interactions occuring ‘locally’, along the edges of the lattice. The model with ‘global’ infectious interactions, defined above, was proposed by Parsons and Quince [19, 20] as an extension to the Moran model for a haploid population studied in Moran [16]. They assume that both alleles (strains) are supercritical, which in our setting translates to assuming also that $R_{0,2} > 1$. Parsons and Quince [19] consider the case where one of the alleles is weaker (the case considered in the present paper), while Parsons and Quince [20] consider the case where both types of allele have equal fitness (basic reproduction number), that is, $R_{0,1} = R_{0,2}$, the case not studied here. Parsons and Quince [19] study the probability that the weaker strain replaces the stronger one for various starting conditions, in finite populations, using approximations and numerics. They also study the extinction times of both the weaker and stronger strains numerically. Also, Humplik et al. [10] study, using approximations and numerics, the probability that the stronger strain replaces the weaker one, assuming that initially the weaker one is in quasi-equilibrium and there is one individual infected with the stronger one. They focus mainly on finite populations, as well as what happens when the basic reproduction numbers of the strains may tend to infinity, and do not consider the time taken by the strains to reach extinction.

For a closely related model with $K \geq 2$ types, Parsons, Quince and Plotkin [21] obtain analytic approximations for the expected time until competitive exclusion occurs, which turns out to be linear in the population size when all the alleles have the same (supercritical) basic reproductive ratio. These authors further conjecture that a similar result should hold for the model considered in our paper, and Kogan et al. [12] have argued, using an approximating “perturbation method”, this is indeed the case for $K = 2$ strains of equal strength.

Theorem 2 below concerns the case where there is a dominant, supercritical, strain and each of the two strains initially affects a positive fraction of the population. Under these conditions, competitive exclusion of the weaker strain by the stronger occurs with high probability (i.e., with probability tending to 1 as the population size $N \to \infty$). Our result shows that, with high probability, the extinction time for the weaker type scales logarithmically as the population size grows large, with randomness asymptotically Gumbel distributed. On the other hand, the time to extinction for the dominant strain grows exponentially with population size.

The corresponding deterministic SIS logistic competition model is among the simplest epidemic models for infections caused by multiple pathogen strains. It is represented by the pair

$$\begin{aligned} \frac{dx_1(t)}{dt} &= \lambda_1 x_1(t)\big(1 - x_1(t) - x_2(t)\big) - \mu_1 x_1(t), \\ \frac{dx_2(t)}{dt} &= \lambda_2 x_2(t)\big(1 - x_1(t) - x_2(t)\big) - \mu_2 x_2(t) \end{aligned} \tag{1}$$

of differential equations, and is thus a particular instance of the deterministic Lotka–Volterra system—see Lotka [15], Volterra [25], Zeeman [27]. In the context of epidemics, this type of deterministic Lotka–Volterra system is considered, for instance, by Kirupaharan and Allen [11] (see also references therein), generalised to allow births and deaths; these authors also consider SDE versions of such systems. When births and deaths are allowed, there are parameter values for which several strains can coexist in the deterministic system, although the authors' numerical studies suggest that this is unlikely in the corresponding SDEs.

Let $\kappa_N = \inf\{t \geq 0 : X_{N,2}(t) = 0\}$, the time when the weaker species goes extinct. Let also $\tilde{\kappa}_N = \inf\{t \geq 0 : X_{N,1}(t) = 0\}$, the time the stronger species goes extinct. We now state our main result, concerning the distribution of $\kappa_N, \tilde{\kappa}_N$ in the case when initially both strains are already established in the population, each infecting a positive proportion of its size.

THEOREM 2. *Suppose that $R_{0,1} > R_{0,2}$ and that $R_{0,1} > 1$. Suppose further that $X_{N,1}(0)/N \to \alpha$ and $X_{N,2}(0)/N \to \beta$ as $N \to \infty$, where $\alpha, \beta > 0$ and $\alpha + \beta \leq 1$. Then, as $N \to \infty$,*

$$\mu_2\left(1 - \frac{R_{0,2}}{R_{0,1}}\right)\kappa_N - \left[\log\left(N\beta\left(1 - \frac{R_{0,2}}{R_{0,1}}\right)\right) + \frac{R_{0,2}\mu_2}{R_{0,1}\mu_1}\log\left(\frac{1 - R_{0,1}^{-1}}{\alpha}\right)\right] \to G,$$

in distribution, where G is a standard Gumbel random variable.

Furthermore, as $N \to \infty$,

$$\mathbb{E}(\tilde{\kappa}_N) \sim \sqrt{\frac{2\pi}{N}} \frac{R_{0,1}}{\mu_1(R_{0,1} - 1)^2} e^{N v_1},$$

as $N \to \infty$, where $v_1 = \log R_{0,1} - 1 + R_{0,1}^{-1}$, and $\tilde{\kappa}_N / \mathbb{E}(\tilde{\kappa}_N) \to Z$ in distribution, and Z is an exponential random variable with parameter 1.

Theorem 2 thus shows that, if initially strain 1 infects around αN individuals and strain 2 infects around βN individuals, then the extinction time κ_N of strain 2 (the weaker strain) can be written as

$$\kappa_N = \frac{\log N + G_N}{\mu_2(1 - \frac{R_{0,2}}{R_{0,1}})},$$

where G_N is a random variable with a bounded mean and variance, while the scaled extinction time $\tilde{\kappa}_N$ of the stronger strain asymptotically has the same distribution as if the weaker strain was absent to begin with.

The long-term behaviour of Markov population processes is of considerable importance in applications. In epidemic models, long-term phenomena include extinction of certain pathogen strains, or replacement of a dominant pathogen strain in the host population by another more adapted pathogen strain introduced into the host population, for example, due to mutation or migration. Mathematically, these phenomena are related to the behaviour of the scaled process near fixed points of its approximating differential equation, including absorbing boundaries for one or more coordinates.

Recently, Barbour, Hamza, Kaspi and Klebaner [4] have shown that, under appropriate conditions, a density dependent Markov population process that starts near an absorbing boundary and manages to escape from it, still can be well-approximated by the deterministic solution as described by the standard theory but with a random time shift, and that the time to escape from such a boundary is random and of order $O(\log N)$, see Theorem 1.1 in Barbour et al. [4]. Also, similar to the phenomenon we investigate in the present work, they describe in a very general setting the behaviour of a class of population processes near a fixed point at which one or more coordinates of the process have value 0, that is, they are extinct, and derive the limit distribution for the extinction times for such processes as a standard Gumbel random variable, after scaling and centering, see Theorem 1.2 in Barbour et al. [4]. In both results, the randomness when the process is escaping or reaching an absorbing boundary is captured by a branching process approximation. However, at their level of generality, the formulae they obtain contain nonexplicit constants, and their bounds on the rate of convergence are too weak to investigate near-critical phenomena. Also, rigorous justification of such a general approximation, based on an abstract coupling of Thorisson (see Theorem 7.3 in Thorisson [24]), is quite involved.

In the present work, we develop a related but more direct and precise approximation to prove an explicit formula for the extinction time of the weaker virus strain in the stochastic logistic SIS competition model. Like the approach of Barbour et al. [4], our approach is based on decomposing the drifts of the process into linear and nonlinear parts, and using a variation of constants formula. However, we additionally take full advantage of the fact that the nonlinear parts are small in the neighbourhood of a fixed point, and provide more refined bounds on the deviations of the martingale transform appearing in the equations. Similar ideas were also used in a different context in discrete time by Brightwell and Luczak [7].

Unlike the approach of Barbour et al. [4], the precision of our approximation facilitates study of near-critical phenomena, and we extend Theorem 2 to a near-critical case where $R_{0,1} = R_{0,1}(N)$ and $R_{0,2} = R_{0,2}(N)$ are such that $R_{0,1} - R_{0,2} \to 0$, while $R_{0,1} - 1$ may or may not tend to 0 but satisfies $(R_{0,1} - R_{0,2})(R_{0,1} - 1)^{-1} \to 0$ as $N \to \infty$. As is argued in Brightwell et al. [6]—see Section 1.1 and Appendix A, as well as references therein—there is evidence from data that near-critical phenomena may manifest themselves in large finite populations. Also, there is evidence that many real-life epidemics are in a sense "near-critical", corresponding to scenarios where, for instance, a mutating pathogen or waning population immunity pushes the basic reproductive ratio just above 1, or where control measures such as mass vaccination push the basic reproductive ratio just below 1.

As a proof of concept, we consider the following special case where $\mu_1 = \mu_2 = 1$, $\lambda_1 = \lambda_1(N) > \lambda_2 = \lambda_2(N) > 1$, and $\lambda_1 - \lambda_2 \to 0$ as $N \to \infty$ (while λ_1 is bounded, and may or may not tend to 1). This may model a real-world scenario where a slightly more infectious strain emerges during an outbreak, for instance, via a mutation, and we want to know the time it takes to supplant the weaker one in the population. We assume that $(\lambda_1 - \lambda_2)(\lambda_1 - 1)^{-1} \to 0$; this implies that, at least for large N, $\lambda_2 > 1$, and so the weaker strain is also supercritical. Also note that we do allow $\lambda_1 \to 0$, but our assumptions imply then the separation of both $R_{0,1}$ and $R_{0,2}$ from the critical value 1 is, asymptotically as $N \to \infty$, greater than the separation between $R_{0,1}$ and $R_{0,2}$.

We further assume that

$$N(\lambda_1-\lambda_2)^3(\lambda_1-1)^{-1}/\log\log\big(N(\lambda_1-\lambda_2)^2\big)\to\infty. \tag{2}$$

The initial conditions are as in Theorem 2, that is, $X_{N,1}(0)/N\to\alpha$ and $X_{N,2}(0)/N\to\beta$ as $N\to\infty$, where $\alpha,\beta>0$ and $\alpha+\beta\le 1$.

THEOREM 3. *Under the above assumptions*,

$$\kappa_N=\frac{\lambda_1}{\lambda_1-\lambda_2}\left(\log\left(N\frac{(\lambda_1-1)(\lambda_1-\lambda_2)}{\lambda_1^2}\frac{\beta}{\alpha}\right)+G_N\right),$$

where G_N converges in distribution to a Gumbel random variable G.

Furthermore, as $N\to\infty$,

$$\mathbb{E}(\tilde\kappa_N)\sim\sqrt{\frac{2\pi}{N}}\frac{\lambda_1}{(\lambda_1-1)^2}e^{N(\log\lambda_1-1+\lambda_1^{-1})},$$

as $N\to\infty$, and $\tilde\kappa_N/\mathbb{E}(\tilde\kappa_N)\to Z$ in distribution, and Z is an exponential random variable with parameter 1.

This means that the formulae for the distribution of κ_N and $\tilde\kappa_N$ in Theorem 2 extend to this near-critical regime. The exact form of the technical condition (2) is an artefact of our proof technique, and does not define the transition to criticality. We believe that, given $\lambda_1=\lambda_1(N)$, $\lambda_2=\lambda_2(N)$ bounded and such that $(\lambda_2-\lambda_1)(\lambda_1-1)^{-1}\to 0$, imposing the additional condition $N(\lambda_1-\lambda_2)^2\to\infty$ is necessary and sufficient for Theorem 3 to hold in this case. Note that, if $N(\lambda_1-\lambda_2)^3(\lambda_1-1)^{-1}\to\infty$ and $(\lambda_1-\lambda_2)(\lambda_1-1)^{-1}\to 0$, then $N(\lambda_1-\lambda_2)^2\to\infty$. We conjecture that there is a further regime, with a different distribution of the randomness (not Gumbel), where $N(\lambda_1-\lambda_2)(\lambda_1-1)\to\infty$ but $N(\lambda_1-\lambda_2)^2$ does not tend to infinity. A discussion of this as well as of what happens when condition (2) is not satisfied is included in Section 7.1. In particular, it seems feasible to refine our proof technique by splitting the differential equation approximation phase into subphases, possibly with enough precision to go all the way to what we believe to be the critical window; however, in the interest of clarity, we do not explore such improvements in the present paper.

Clearly, we do not cover the entire spectrum of near-critical behaviours: the example considered is meant as a proof of concept, and a full investigation will be carried out systematically in future work. One challenge of such an investigation will be to understand the behaviour of the approximating deterministic process in various near-critical regimes. In particular, in future work we intend to study cases where the strengths of the two strains are even closer to identical than considered in the present paper. A further project is to rigorously study the probability that the stronger strain wins, starting with only a small number of infected individuals relative to the number of infectives of the weaker strain, in particular in near-critical scenarios, where there is likely to be a delicate interplay between initial conditions and the asymptotic differences between the strain strengths.

We will further work to extend our results to competition of more than two strains. Away from criticality, the fact that the randomness of the extinction time of the weaker strains has an approximate Gumbel distribution as the population size grows large follows from Theorem 1.2 in Barbour et al. [4], however obtaining an explicit formula does not seem straightforward. With one strain (the stochastic SIS logistic epidemic process), an explicit solution to the deterministic system is available, see Brightwell et al. [6] and references therein, as well as equation (19) in the present paper. With two strains, we are able to draw on the following formula satisfied by any solution $x(t)=(x_1(t),x_2(t))^T$ to (1):

$$\frac{(x_1(t))^{\lambda_2}}{(x_2(t))^{\lambda_1}}=\frac{(x_1(0))^{\lambda_2}}{(x_2(0))^{\lambda_1}}e^{(\mu_2\lambda_1-\mu_1\lambda_2)t}\quad\text{for all } t\ge 0.$$

This is easy to verify by differentiating $x_1(t)^{\lambda_2}/x_2(t)^{\lambda_1}$. However, we are not aware of analogous properties for systems with three or more strains, which makes it more challenging to study near-critical behaviour.

The rest of the paper is organised as follows. In Section 2, we present some preliminaries concerning the stability of fixed points of the deterministic logistic SIS competition model. Furthermore, we give an overview of the strategy used to prove Theorem 2. The idea is that the stochastic SIS logistic competition process follows closely the corresponding deterministic process for a long time, until the latter one is close to its attractive fixed point at $((\lambda_1-\mu_1)/\lambda_1, 0)^T$. From there on, the time to extinction for the second species is short, and well approximated by a linear birth-and-death chain, with the randomness captured by the Gumbel distribution. We break up the analysis of the process into phases, similarly to the approach of Brightwell, House and Luczak [6] used to prove a general version of Theorem 1(ii). We analyze each of these phases in the subsequent sections. In Section 6, we combine the results from the preceding sections to prove Theorem 2. In Section 7, we prove Theorem 3.

Throughout our proofs, we treat $X_N(t)$ and $x(t)$ as column vectors.

2. Preliminaries. In this section, we discuss the deterministic Lotka–Volterra system. For suitable choices of parameter values, this model becomes the deterministic logistic SIS competition model, and approximates the stochastic logistic SIS competition model over certain timescales.

We further outline the proof of our main result, Theorem 2.

2.1. *A deterministic version of the competition model.* The deterministic competitive Lotka–Volterra system represents a community of k mutually competing species described by equations

$$\frac{dx_i(t)}{dt} = x_i(t)\left(b_i - \sum_{j=1}^{k} a_{ij}x_j(t)\right), \qquad i=1,\dots,k, \tag{3}$$

where $x_i(t)$ denotes the population size of the ith species at time t. It is assumed that $b_i > 0$ for all i, and $a_{ij} > 0$ for all i, j. For each $i = 1, \dots, k$, species i would by itself, in the absence of all the other species, exhibit logistic growth, with its growth rate decreasing as its population increases near the carrying capacity of the environment. Mathematically,

$$\frac{dx_i(t)}{dt} = x_i(t)\big(b_i - a_{ii}x_i(t)\big), \qquad b_i, a_{ii} > 0.$$

This equation has two fixed points: 0 and b_i/a_{ii}, the latter being the *carrying capacity* of species i. Also, dx_i/dt decreases in x_i near the carrying capacity.

The following result of Zeeman [27] gives simple criteria on the parameters b_i and a_{ij} of (3) which guarantee that, for all strictly positive initial conditions of (3), all but one of the species is driven to extinction, while the one remaining species stabilizes at its own carrying capacity.

We recall that a fixed point x^* of a system of ordinary differential equations is *globally attractive* on a set U if and only if its basin of attraction is equal to U. In other words, x^* is globally attractive if every solution to the system with initial condition in U converges to x^* as $t \to \infty$.

THEOREM 4 (Zeeman 1995 [27]). *Suppose that the parameters in* (3) *satisfy*

$$\begin{aligned} &\frac{b_j}{a_{jj}} < \frac{b_i}{a_{ij}} \quad \forall i < j, \\ &\frac{b_j}{a_{jj}} > \frac{b_i}{a_{ij}} \quad \forall i > j. \end{aligned} \tag{4}$$

Then fixed point $(\frac{b_1}{a_{11}}, 0, \dots, 0)^T$ *is globally attractive on the interior of* $\mathbb{R}^k_+$.

Clearly, if conditions (4) are satisfied, and $x_i(0) = 0$ for some $i > 1$, then the solution $x(t)$ still converges to $(\frac{b_1}{a_{11}}, 0, \dots, 0)^T$ as $t \to \infty$, as this case amounts to eliminating species i from the equations.

In the case $\lambda_1/\mu_1 > \lambda_2/\mu_2 > 1$, the stochastic competition model defined in the Introduction can be naturally associated with a particular two-dimensional instance of (3) with $b_i = \lambda_i - \mu_i$ for $i = 1, 2$, and $a_{ij} = \lambda_i$ for $i, j = 1, 2$, which is precisely system (1). In epidemic modelling, $x_1(t)$ and $x_2(t)$ represent the proportions of individuals who at time t are infected by strains 1 and 2 respectively, in a closed homogeneously mixing population.

By Theorem 4, all solutions with $x_1(0) > 0$ converge as $t \to \infty$ to $(\frac{\lambda_1 - \mu_1}{\lambda_1}, 0)^T$. We claim this still holds even when $\lambda_2/\mu_2 \le 1$ (corresponding to the case $b_2 \le 0$). First of all, using (1),

$$\frac{d}{dt}\left(\frac{x_2(t)^{\mu_1}}{x_1(t)^{\mu_2}}\right) = \mu_1\mu_2 \frac{x_2(t)^{\mu_1}}{x_1(t)^{\mu_2}}\big(1 - x_1(t) - x_2(t)\big)\left(\frac{\lambda_2}{\mu_2} - \frac{\lambda_1}{\mu_1}\right) < 0,$$

so $x_2(t)^{\mu_1}/x_1(t)^{\mu_2}$ is decreasing in t. Hence, one can see that $\frac{dx_1(t)}{dt} \ge 0$ if $x_1(t) \le \varepsilon$, for $\varepsilon > 0$ small enough: we can choose ε such that $\varepsilon + (\varepsilon/x_1(0))^{\mu_2/\mu_1} \le 1 - \mu_1/\lambda_1$, and then, from (1) again,

$$\frac{dx_1(t)}{dt} \ge \lambda_1 x_1\left(1 - \frac{\mu_1}{\lambda_1} - \varepsilon - \left(\frac{\varepsilon}{x_1(0)}\right)^{\mu_2/\mu_1}\right) \ge 0.$$

Now consider the Lyapunov function

$$\phi(x_1, x_2) = \frac{1}{2}\left(x_1 + x_2 - 1 + \frac{\mu_1}{\lambda_1}\right)^2 + x_2\left(\frac{\mu_2}{\lambda_2} - \frac{\mu_1}{\lambda_1}\right).$$

This function is nonnegative, and is zero only at the fixed point $(1 - \frac{\mu_1}{\lambda_1}, 0)^T$. The derivative is given by

$$\begin{aligned}
&\frac{d}{dt}\phi\big(x_1(t), x_2(t)\big) \\
&= \left(x_1 + x_2 - 1 + \frac{\mu_1}{\lambda_1}\right)\left(\frac{dx_1}{dt} + \frac{dx_2}{dt}\right) + \left(\frac{\mu_2}{\lambda_2} - \frac{\mu_1}{\lambda_1}\right)\frac{dx_2}{dt} \\
&= -\lambda_1 x_1\left(x_1 + x_2 - 1 + \frac{\mu_1}{\lambda_1}\right)^2 + \left(x_1 + x_2 - 1 + \frac{\mu_2}{\lambda_2}\right)\frac{dx_2}{dt} \\
&= -\lambda_1 x_1\left(x_1 + x_2 - 1 + \frac{\mu_1}{\lambda_1}\right)^2 - \lambda_2 x_2\left(x_1 + x_2 - 1 + \frac{\mu_2}{\lambda_2}\right)^2,
\end{aligned}$$

and is nonpositive everywhere; furthermore, for any $0 < \varepsilon < 1 - \mu_1/\lambda_1$, it is zero in $\{(x_1, x_2)^T : x_1 \ge \varepsilon, x_2 \ge 0, x_1 + x_2 \le 1\}$ only at the fixed point $(1 - \frac{\mu_1}{\lambda_1}, 0)^T$. Since $dx_1(t)/dt \ge 0$ if $x_1(t) \le \varepsilon$, the set $\{(x_1, x_2)^T : x_1 \ge \varepsilon, x_2 \ge 0, x_1 + x_2 \le 1\}$ is invariant for the deterministic logistic SIS competition model. As ε can be taken arbitrarily small, the claim will follow from a standard Lyapunov stability argument—see, for instance, Chapter 3 of Arrowsmith and Place [2]—as follows.

Take any solution $x(t)$ with $x_1(0) > 0$. Note that there is some $\varepsilon > 0$ so that $x_1(t) > \varepsilon$ for all t. As $\phi(x_1(t), x_2(t)) \ge 0$ and decreasing, it converges to some nonnegative limit. From the mean value theorem, there is a sequence (t_k) of times such that $t_k \to \infty$ and $\frac{d}{dt}\phi(x_1(t_k), x_2(t_k))$ tends to zero. Using that and the fact that $x_1(t_k)$ is bounded away from zero, it follows that $x_1(t_k) + x_2(t_k) \to 1 - \mu_1/\lambda_1$. This then implies that $x_1(t_k) + x_2(t_k) \not\to 1 - \mu_2/\lambda_2$ and so $x_2(t_k) \to 0$ and $x_1(t_k) \to 1 - \mu_1/\lambda_1$—in other words, $x(t_k)$ approaches the fixed point. Now we can conclude that $\phi(x_1(t_k), x_2(t_k)) \to 0$. As $\phi(x_1(t), x_2(t))$ is decreasing, it tends to 0. From the formula for ϕ, it follows that $x_2(t)$ tends to 0, and then that $x_1(t) \to 1 - \mu_1/\lambda_1$—in other words, $x(t)$ approaches the fixed point.

It is easy to check that each solution $x(t) = (x_1(t), x_2(t))^T$ to (1) satisfies

$$\frac{(x_1(t))^{\lambda_2}}{(x_2(t))^{\lambda_1}} = \frac{(x_1(0))^{\lambda_2}}{(x_2(0))^{\lambda_1}} e^{(\mu_2\lambda_1 - \mu_1\lambda_2)t} \quad \text{for all } t \geq 0. \tag{5}$$

This can be used to calculate the time $t_{c\to d}$ spent by $x(t)$ to travel from a point $c = (c_1, c_2)^T$ to another point $d = (d_1, d_2)^T$ on the same trajectory:

$$t_{c\to d} = \frac{\lambda_2}{\mu_2\lambda_1 - \mu_1\lambda_2} \log(d_1/c_1) - \frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} \log(d_2/c_2). \tag{6}$$

The Jacobian of (1) at $(\frac{\lambda_1-\mu_1}{\lambda_1}, 0)^T$ is given by

$$A = \begin{pmatrix} -(\lambda_1 - \mu_1) & -(\lambda_1 - \mu_1) \\ 0 & -(\mu_2 - \lambda_2\mu_1/\lambda_1) \end{pmatrix}, \tag{7}$$

and thus has eigenvalues $-(\lambda_1 - \mu_1)$ and $-(\mu_2 - \lambda_2\mu_1/\lambda_1)$, which are real and strictly negative under the assumptions that $\lambda_1/\mu_1 > 1$ and $\lambda_1/\mu_1 > \lambda_2/\mu_2$. By standard theory, the speed of convergence is determined by $-\min\{\lambda_1 - \mu_1, \mu_2 - \lambda_2\mu_1/\lambda_1\}$. Indeed, by Chapter VII §29, Theorem VII in Walter [26], for any $0 < \sigma < \min\{\lambda_1 - \mu_1, \mu_2 - \lambda_2\mu_1/\lambda_1\}$, there exist $\eta > 0$, $C > 0$ such that, if $\|(x_1(0), x_2(0))^T - (\frac{\lambda_1-\mu_1}{\lambda_1}, 0)^T\|_2 < \eta$, then

$$\left\|(x_1(t), x_2(t))^T - \left(\frac{\lambda_1 - \mu_1}{\lambda_1}, 0\right)^T\right\|_2 \leq Ce^{-\sigma t} \quad \text{for all } t \geq 0.$$

In Section 2.2 below, we will give a stronger bound, as well as a lower bound on the speed of convergence.

2.2. *Convergence to fixed point.* We now give quantitative upper and lower bounds on the speed of convergence. Let $\eta_1 = \lambda_1 - \mu_1$ and $\eta_2 = \mu_2 - \lambda_2\mu_1/\lambda_1$, so that $-\eta_1, -\eta_2$ are the eigenvalues of A. Let

$$a = 1 - \frac{\eta_2}{\eta_1}, \tag{8}$$

and assume that $a \neq 0$, that is, $\eta_2 \neq \eta_1$. When $a = 0$, the matrix (7) has repeated eigenvalues. We will consider this case at the end of this subsection.

We introduce new coordinates $\tilde{x}_1(t) = x_1(t) - \frac{\lambda_1-\mu_1}{\lambda_1} + \frac{1}{a}x_2(t)$ and $\tilde{x}_2(t) = x_2(t)$, and let $\tilde{x}(t) = (\tilde{x}_1(t), \tilde{x}_2(t))^T$. In the new coordinates, the differential equation (1) is expressed as

$$\begin{aligned} \frac{d\tilde{x}_1(t)}{dt} &= -\eta_1\tilde{x}_1(t) - \lambda_1\tilde{x}_1(t)^2 - \frac{\eta_2(\lambda_1 - \lambda_2)}{\eta_1}\left(\frac{\tilde{x}_2(t)}{a}\right)^2 \\ &\quad + \left(\lambda_1 - \lambda_2 + \frac{\lambda_1\eta_2}{\eta_1}\right)\tilde{x}_1(t)\frac{\tilde{x}_2(t)}{a}, \\ \frac{d\tilde{x}_2(t)}{dt} &= -\eta_2\tilde{x}_2(t) - \lambda_2\tilde{x}_2(t)\tilde{x}_1(t) + \frac{\lambda_2}{a}\frac{\eta_2}{\eta_1}\tilde{x}_2(t)^2. \end{aligned} \tag{9}$$

Note the diagonal form of the linear terms in the equation, reflecting the fact that $(1, 1/a)$ and $(0, 1)$ are the left eigenvectors of the matrix A, with eigenvalues $-\eta_1$ and $-\eta_2$ respectively.

LEMMA 1. *Suppose that $a \neq 0$. Let $L = \min\{\eta_1, \eta_2\}$ and $L_1 = (\lambda_1 + |\lambda_1 - \lambda_2|)\frac{\eta_1+\eta_2}{\eta_1}$. Suppose $\tilde{x}(0)$ is such that*

$$y(0) = \max\{|\tilde{x}_1(0)|, \tilde{x}_2(0)/|a|\} \leq L/2L_1.$$

Then, for all $t \geq 0$, $|\tilde{x}_1(t)| \leq 2y(0)e^{-tL}$, and $\tilde{x}_2(t) \leq 2|a|y(0)e^{-tL}$.

PROOF. We write, as is standard, $\tilde{x}(t)=\tilde{x}(0)+\int_0^t F(\tilde{x}(s))\,ds$, where, from (9),

$$F(x)=\begin{pmatrix} -\eta_1 x_1-\lambda_1 x_1^2-\dfrac{\eta_2(\lambda_1-\lambda_2)}{\eta_1}\left(\dfrac{x_2}{a}\right)^2+\left(\lambda_1-\lambda_2+\dfrac{\lambda_1\eta_2}{\eta_1}\right)x_1\dfrac{x_2}{a} \\ -\eta_2 x_2-\lambda_2 x_2 x_1+\dfrac{\lambda_2}{a}\dfrac{\eta_2}{\eta_1}x_2^2 \end{pmatrix}. \tag{10}$$

We then decompose $F(x)=\tilde{A}\begin{pmatrix}x_1\\x_2\end{pmatrix}+\tilde{F}(x)$, where

$$\tilde{A}=\begin{pmatrix}-\eta_1 & 0\\ 0 & -\eta_2\end{pmatrix} \tag{11}$$

and

$$\tilde{F}(x)=\begin{pmatrix} -\lambda_1 x_1^2-\dfrac{\eta_2(\lambda_1-\lambda_2)}{\eta_1}\left(\dfrac{x_2}{a}\right)^2+\left(\lambda_1-\lambda_2+\dfrac{\lambda_1\eta_2}{\eta_1}\right)x_1\dfrac{x_2}{a} \\ -\lambda_2 x_2 x_1+\dfrac{\lambda_2}{a}\dfrac{\eta_2}{\eta_1}x_2^2 \end{pmatrix}. \tag{12}$$

Then, treating (9) as a perturbation of a linear system, see Chapter 6 of Pazy [22], the solution $\tilde{x}(t)$ satisfies

$$\tilde{x}(t)=e^{t\tilde{A}}\tilde{x}(0)+\int_0^t e^{(t-s)\tilde{A}}\tilde{F}(\tilde{x}(s))\,ds,$$

or, equivalently,

$$\begin{aligned}\begin{pmatrix}\tilde{x}_1(t)\\ \tilde{x}_2(t)\end{pmatrix}&=\begin{pmatrix}e^{-t\eta_1}\tilde{x}_1(0)\\ e^{-t\eta_2}x_2(0)\end{pmatrix}\\ &\quad+\int_0^t\begin{pmatrix} e^{-(t-s)\eta_1}\left[-\lambda_1\tilde{x}_1(s)^2-\dfrac{\eta_2(\lambda_1-\lambda_2)}{\eta_1}\left(\dfrac{\tilde{x}_2(s)}{a}\right)^2+\left(\lambda_1-\lambda_2+\dfrac{\lambda_1\eta_2}{\eta_1}\right)\tilde{x}_1(s)\dfrac{\tilde{x}_2(s)}{a}\right] \\ e^{-(t-s)\eta_2}\left[-\lambda_2\tilde{x}_2(s)\tilde{x}_1(s)+\dfrac{\lambda_2}{a}\dfrac{\eta_2}{\eta_1}\tilde{x}_2(s)^2\right]\end{pmatrix}ds.\end{aligned}$$

Let $y_1(t)=|\tilde{x}_1(t)|e^{Lt}$, let $y_2(t)=\frac{\tilde{x}_2(t)}{|a|}e^{Lt}$ and let $y(t)=\max\{y_1(t),y_2(t)\}$. Then, from the above,

$$\begin{aligned} y_1(t)&\le y_1(0)+\int_0^t e^{Ls}\left[\lambda_1\tilde{x}_1(s)^2+\frac{\eta_2|\lambda_1-\lambda_2|}{\eta_1}\left(\frac{\tilde{x}_2(s)}{a}\right)^2\right.\\ &\qquad\left.+\left|\lambda_1-\lambda_2+\frac{\lambda_1\eta_2}{\eta_1}\right||\tilde{x}_1(s)|\frac{\tilde{x}_2(s)}{|a|}\right]ds\\ &\le y_1(0)+\left(\lambda_1+\frac{\eta_2|\lambda_1-\lambda_2|}{\eta_1}+|\lambda_1-\lambda_2|+\frac{\lambda_1\eta_2}{\eta_1}\right)\int_0^t y(s)^2e^{-Ls}\,ds\\ &=y_1(0)+\frac{\eta_1+\eta_2}{\eta_1}(\lambda_1+|\lambda_1-\lambda_2|)\int_0^t y(s)^2e^{-Ls}\,ds\\ &=y_1(0)+L_1\int_0^t y(s)^2e^{-Ls}\,ds,\end{aligned}$$

and also

$$y_2(t)\le y_2(0)+\lambda_2\frac{\eta_1+\eta_2}{\eta_1}\int_0^t y(s)^2e^{-Ls}\,ds\le y_2(0)+L_1\int_0^t y(s)^2e^{-Ls}\,ds,$$

so that

$$y(t)\le y(0)+L_1\int_0^t y(s)^2e^{-Ls}\,ds.$$

Now, the equation $z(t) = z(0) + L_1 \int_0^t z(s)^2 e^{-Ls}\, ds$ is solved by

$$z(t) = \frac{Lz(0)}{L + L_1 z(0)(e^{-Lt} - 1)}$$

for all t, as long as $z(0) < L/L_1$, and, if $z(0) \le L/2L_1$, then $z(t) \le 2z(0)$ for all t. Now a standard argument, considering the difference $z(t) - y(t)$, shows that $y(t) \le z(t) \le 2z(0)$ provided that $y(0) \le z(0) \le L/2L_1$. Hence, $|\tilde{x}_1(t)| \le 2\max\{\tilde{x}_1(0), \frac{\tilde{x}_2(0)}{|a|}\}e^{-Lt}$ and $\tilde{x}_2(t) \le 2|a|\max\{\tilde{x}_1(0), \frac{\tilde{x}_2(0)}{|a|}\}e^{-Lt}$ for all t, as required. □

LEMMA 2. *Suppose that $a \ne 0$. Let L, L_1 be as in Lemma 1. Suppose $\tilde{x}(0)$ is such that*

$$y(0) = \max\{|\tilde{x}_1(0)|, \tilde{x}_2(0)/|a|\} \le L/8L_1.$$

Then, for all $t \ge 0$, $x_2(t) \le 2x_2(0)e^{-t\eta_2}$, and $x_2(t) \ge \frac{1}{2}x_2(0)e^{-t\eta_2}$.

PROOF. Defining $y(t)$ as in the proof of Lemma 1, and using that lemma,

$$\begin{aligned} x_2(t) &\le x_2(0)e^{-t\eta_2} + \lambda_2 \int_0^t e^{-(t-s)\eta_2} |\tilde{x}_1(s)| x_2(s)\, ds + \frac{\lambda_2}{|a|}\frac{\eta_2}{\eta_1} \int_0^t e^{-(t-s)\eta_2} \tilde{x}_2(s)^2\, ds \\ &\le x_2(0)e^{-t\eta_2} + 2\lambda_2 \frac{\eta_1 + \eta_2}{\eta_1} e^{-t\eta_2} y(0) \int_0^t x_2(s) e^{s\eta_2} e^{-sL}\, ds. \end{aligned}$$

Letting $\tilde{y}_2(t) = x_2(t)e^{t\eta_2}$, and using the fact that $\lambda_2 \le \lambda_1 + |\lambda_1 - \lambda_2|$,

$$\tilde{y}_2(t) \le \tilde{y}_2(0) + 2L_1 y(0) \int_0^t \tilde{y}_2(s) e^{-sL}\, ds,$$

so, by Grönwall's inequality,

$$\tilde{y}_2(t) \le \tilde{y}_2(0) \exp\biggl(2L_1 y(0) \int_0^t e^{-sL}\, ds\biggr) \le \tilde{y}_2(0) \exp(2L_1 y(0)/L),$$

and so $\tilde{y}_2(t) \le 2\tilde{y}_2(0)$, for all t, as required, since $y(0) \le L/8L_1$.

Furthermore, for all t, using the upper bound on $x_2(t)$ derived above,

$$\begin{aligned} x_2(t) &\ge x_2(0)e^{-t\eta_2} - \lambda_2 \int_0^t e^{-(t-s)\eta_2} |\tilde{x}_1(s)| x_2(s)\, ds - \frac{\lambda_2}{|a|}\frac{\eta_2}{\eta_1} \int_0^t e^{-(t-s)\eta_2} \tilde{x}_2(s)^2\, ds \\ &\ge x_2(0)e^{-t\eta_2} - 2L_1 e^{-t\eta_2} y(0) \int_0^t x_2(s) e^{s\eta_2} e^{-sL}\, ds \\ &\ge x_2(0)e^{-t\eta_2} - 4L_1 y(0) x_2(0) e^{-t\eta_2} \int_0^t e^{-sL}\, ds \\ &\ge x_2(0)e^{-t\eta_2} - \frac{4L_1 y(0)}{L} x_2(0) e^{-t\eta_2} \ge \frac{1}{2} x_2(0) e^{-t\eta_2}. \end{aligned}$$

□

In the case $\eta_1 = \eta_2$, we work with the original variables $x_1(t), x_2(t)$, and write

$$\begin{aligned} \begin{pmatrix} x_1(t) - \dfrac{\lambda_1 - \mu_1}{\lambda_1} \\ x_2(t) \end{pmatrix} &= e^{tA} \begin{pmatrix} x_1(0) - \dfrac{\lambda_1 - \mu_1}{\lambda_1} \\ x_2(0) \end{pmatrix} + \int_0^t e^{A(t-s)} \tilde{F}(x(s))\, ds \\ &= e^{tA} \begin{pmatrix} x_1(0) - \dfrac{\lambda_1 - \mu_1}{\lambda_1} \\ x_2(0) \end{pmatrix} \\ &\quad + \int_0^t e^{A(t-s)} \begin{pmatrix} -\lambda_1 \Bigl(x_1(s) - \dfrac{\lambda_1 - \mu_1}{\lambda_1}\Bigr)^2 - \lambda_1 \Bigl(x_1(s) - \dfrac{\lambda_1 - \mu_1}{\lambda_1}\Bigr) x_2(s) \\ -\lambda_2 \Bigl(x_1(s) - \dfrac{\lambda_1 - \mu_1}{\lambda_1}\Bigr) x_2(s) - \lambda_2 (x_2(s))^2 \end{pmatrix} ds, \end{aligned}$$

where

$$e^{tA} = \begin{pmatrix} e^{-t(\lambda_1-\mu_1)} & -(\lambda_1-\mu_1)te^{-t(\lambda_1-\mu_1)} \\ 0 & e^{-t(\lambda_1-\mu_1)} \end{pmatrix}.$$

Since, in the near-critical scenario we consider, the eigenvalues are always distinct, we settle below for a fairly weak bound on the speed of convergence. Let $y(t) = e^{t(\lambda_1-\mu_1)/2} \times \max\{|x_1(t) - (\lambda_1-\mu_1)/\lambda_1|, x_2(t)\}$.

LEMMA 3. *Suppose that* $a = 0$. *Assume that* $y(0) \le (\lambda_1-\mu_1)/32(\lambda_1+\lambda_2)$. *Then, for all* $t \ge 0$, $|x_1(t) - (\lambda_1-\mu_1)/\lambda_1| \le 4y(0)e^{-t(\lambda_1-\mu_1)/2}$ *and* $x_2(t) \le 4y(0)e^{-t(\lambda_1-\mu_1)/2}$.

PROOF. Using the inequalities $\lambda_2[(\lambda_1-\mu_1)t + \lambda_1/\lambda_2]e^{-t(\lambda_1-\mu_1)/2} \le 2(\lambda_1+\lambda_2)$ and $[(\lambda_1-\mu_1)t+1]e^{-t(\lambda_1-\mu_1)/2} \le 2$, we have

$$\begin{aligned}
&|x_1(t) - (\lambda_1-\mu_1)/\lambda_1| \\
&\quad \le 2y(0)e^{-t(\lambda_1-\mu_1)/2} + 4(\lambda_1+\lambda_2)e^{-t(\lambda_1-\mu_1)/2}\int_0^t y(s)^2 e^{-s(\lambda_1-\mu_1)/2}\,ds
\end{aligned}$$

and

$$x_2(t) \le y(0)e^{-t(\lambda_1-\mu_1)/2} + 2\lambda_2 e^{-t(\lambda_1-\mu_1)/2}\int_0^t y(s)^2 e^{-s(\lambda_1-\mu_1)/2}\,ds.$$

It follows using the same argument as in the proof of Lemma 1 that

$$y(t) \le 2y(0) + 4(\lambda_1+\lambda_2)\int_0^t y(s)^2 e^{-s(\lambda_1-\mu_1)/2}\,ds,$$

so

$$y(t) \le \frac{(\lambda_1-\mu_1)y(0)}{(\lambda_1-\mu_1)/2 + 8(\lambda_1+\lambda_2)y(0)(e^{-t(\lambda_1-\mu_1)/2}-1)}.$$

Thus if $y(0) \le (\lambda_1-\mu_1)/32(\lambda_1+\lambda_2)$, then $y(t) \le 4y(0)$, and so $|x_1(t) - (\lambda_1-\mu_1)/\lambda_1| \le 4y(0)e^{-t(\lambda_1-\mu_1)/2}$ and $x_2(t) \le 4y(0)e^{-t(\lambda_1-\mu_1)/2}$, as required. □

LEMMA 4. *Suppose that* $a = 0$. *Assume that* $y(0) \le (\lambda_1-\mu_1)/32(\lambda_1+\lambda_2)$. *Then, for all* $t \ge 0$, $x_2(t) \le 2x_2(0)e^{-t(\lambda_1-\mu_1)}$ *and* $x_2(t) \ge \frac{1}{2}x_2(0)e^{-t(\lambda_1-\mu_1)}$.

PROOF. Letting $\tilde{y}_2(t) = x_2(t)e^{(\lambda_1-\mu_1)t}$, we have, using Lemma 3,

$$\tilde{y}_2(t) \le \tilde{y}_2(0) + 4\lambda_2 y(0)\int_0^t \tilde{y}_2(s)e^{-s(\lambda_1-\mu_1)/2}\,ds,$$

so

$$\tilde{y}_2(t) \le \tilde{y}_2(0)\exp\bigl(8\lambda_2 y(0)/(\lambda_1-\mu_1)\bigr) \le 2\tilde{y}_2(0),$$

since $y(0) \le (\lambda_1-\mu_1)/32\lambda_2$, and so $x_2(t) \le 2x_2(0)e^{-t(\lambda_1-\mu_1)}$. Finally,

$$\begin{aligned}
\tilde{y}_2(t) &\ge \tilde{y}_2(0) - 4\lambda_2 y(0)\int_0^t \tilde{y}_2(s)e^{-s(\lambda_1-\mu_1)/2}\,ds \\
&\ge \tilde{y}_2(0) - 8\lambda_2 y(0)x_2(0)\int_0^t e^{-s(\lambda_1-\mu_1)/2}\,ds \\
&\ge x_2(0) - 16\lambda_2 y(0)x_2(0)/(\lambda_1-\mu_1),
\end{aligned}$$

so $y_2(t) \ge \frac{1}{2}x_2(0)e^{-t(\lambda_1-\mu_1)}$. □

2.3. *Proof strategy.* As we mentioned in the Introduction, we break up the analysis of the competition process into phases as follows.

Initial phase ("burn-in" period): By standard theory, see for instance Kurtz [13] or Darling and Norris [8], for large N, $X_N(t)/N$, is well approximated by the solution $x(t)$ of (1) starting from the same (or nearby) initial condition, at least over a fixed length (i.e., independent of N) interval $[0, t_0]$. We will choose t_0 such that $x_1(t_0)$ is close to $\frac{\lambda_1-\mu_1}{\lambda_1}$ and $x_2(t_0)$ is very small. (This is possible by Theorem 4 and the discussion following it, since the fixed point $((\lambda_1-\mu_1)/\lambda_1, 0)^T$ is stable.)

Intermediate phase: After time t_0, we linearise (1) and its stochastic analogue around $((\lambda_1-\mu_1)/\lambda_1, 0)^T$, and use this to show that $x_N(t) = X_N(t)/N$ follows the solution $x(t)$ to (1) for quite a long time after t_0. Our approach here is a variation on standard martingale techniques adapted to exploit the proximity of a stable fixed point.

We choose the time $t_{N,1}$ as the time when $x_2(t)$ drops down to $N^{-1/4}$ (so $X_{N,2}(t)$ will be around $N^{3/4}$).

Final phase: This phase starts with $X_{N,1}(t)$ near $\frac{\lambda_1-\mu_1}{\lambda_1}N$ and $X_{N,2}(t)$ near $N^{3/4}$. From then onwards, "logistic effects" can be ignored, and the path of $X_{N,2}(t)$ can be sandwiched between the paths of two subcritical linear birth and death processes also starting near $N^{3/4}$. Since the time to extinction of a linear birth and death process is well known, we obtain the distribution of the remaining time until the extinction of $X_{N,2}(t)$.

Theorem 2 follows by adding up the times spent in each phase.

3. Initial phase.

LEMMA 5. *Set $x_N(t) = X_N(t)/N$ and $\ell = 5(\lambda_1+\lambda_2+\mu_1+\mu_2)$. Let $t_0 > 0$, let $0 < \delta \le (\log 4)t_0\ell$, and let $x(t)$ be a solution to (1) such that $\|x_N(0) - x(0)\|_1 \le \delta$. Then*

$$\mathbb{P}\Big(\sup_{t\le t_0}\|x_N(t) - x(t)\|_1 > 2\delta e^{\ell t_0}\Big) \le 4e^{-\delta^2 N/(4t_0\ell)}.$$

PROOF. We use a general method described in, for instance, Darling and Norris [8]. In the next section, we will develop a variant adapted to the case where the solution $x(t)$ is in the neighbourhood of a stable fixed point.

As is standard, we write

$$x(t) = x(0) + \int_0^t F\big(x(s)\big)\,ds,$$

where $F : \mathbb{R}^2 \to \mathbb{R}^2$ is given by

$$F(x) = \begin{pmatrix} F_1(x) \\ F_2(x) \end{pmatrix} = \begin{pmatrix} \lambda_1 x_1(1-x_1-x_2) - \mu_1 x_1 \\ \lambda_2 x_2(1-x_1-x_2) - \mu_2 x_2 \end{pmatrix}.$$

Also, $x_N(t) = X_N(t)/N$ satisfies

$$x_N(t) = x_N(0) + \int_0^t F\big(x_N(s)\big)\,ds + M_N(t),$$

where $(M_N(t))$ is a zero-mean martingale.

We can take $\ell = 5(\lambda_1+\lambda_2+\mu_1+\mu_2)$ for a Lipschitz constant of F with respect to $\|\cdot\|_1$ in the subset of $\mathbb{R}^2$ given by $\{x = (x_1, x_2)^T : 0 \le x_1, x_2 \le 1\}$. Then for $t \le t_0$,

$$\begin{aligned}
\|x_N(t) - x(t)\|_1 &\le \|x_N(0) - x(0)\|_1 + \int_0^{t_0} \big\|F\big(x_N(s)\big) - F\big(x(s)\big)\big\|_1\,ds + \sup_{t\le t_0}\|M_N(t)\|_1 \\
&\le \|x_N(0) - x(0)\|_1 + \ell\int_0^{t_0} \sup_{u\le s}\|x_N(u) - x(u)\|_1\,ds + \sup_{t\le t_0}\|M_N(t)\|_1 \\
&\le \Big(\|x_N(0) - x(0)\|_1 + \sup_{t\le t_0}\|M_N(t)\|_1\Big)e^{\ell t_0},
\end{aligned}$$

so, by Grönwall's inequality,

$$\sup_{t\le t_0}\|x_N(t)-x(t)\|_1 \le \Big(\|x_N(0)-x(0)\|_1+\sup_{t\le t_0}\|M_N(t)\|_1\Big)e^{\ell t_0}. \tag{13}$$

For $\theta\in\mathbb{R}^2$, let

$$\begin{aligned} Z_N(t,\theta)&=\exp\Big(\theta^T\big(x_N(t)-x_N(0)\big)-\int_0^t ds\sum_y q_N\big(x_N(s),x_N(s)+y\big)\big(e^{\theta^T y}-1\big)\Big)\\ &=\exp\Big(\theta^T M_N(t)-\int_0^t ds\sum_y q_N\big(x_N(s),x_N(s)+y\big)\big(e^{\theta^T y}-1-\theta^T y\big)\Big), \end{aligned}$$

where $q_N(x,x+y)$ denotes the rate of y jumps (i.e., jumps resulting in an increment by a vector y) of $x_N(t)$ when in state x. Then $(Z_N(t,\theta))$ is a mean 1 martingale. Note that, for our model, the jumps y are of the form $(\pm 1/N,0)^T$ and $(0,\pm 1/N)^T$, and $F(x)=\sum_y q_N(x,x+y)y$.

Using the identity $e^z-1-z=z^2\int_{r=0}^1 e^{rz}(1-r)\,dr$, $Z_N(t,\theta)$ equals

$$\exp\Big(\theta^T M_N(t)-\int_0^t\sum_y q_N\big(x_N(s),x_N(s)+y\big)(\theta^T y)^2\Big(\int_0^1 e^{r\theta^T y}(1-r)\,dr\Big)ds\Big).$$

As the jumps y are of the form $(\pm 1/N,0)^T$ and $(0,\pm 1/N)^T$,

$$\int_0^1 e^{r\theta^T y}(1-r)\,dr\le\frac{1}{2}e^{\gamma}$$

for $\|\theta\|_2\le\gamma N$. It follows that, for all t,

$$Z_N(t,\theta)\ge\exp\Big(\theta^T M_N(t)-\frac{1}{2}e^{\gamma}\int_0^t\sum_y q_N\big(x_N(s),x_N(s)+y\big)\big(\theta^T y\big)^2\,ds\Big).$$

In particular, let $\theta_1,\theta_2\in\mathbb{R}$, and let $\theta^i=\theta_i e_i$ (where e_i is the unit vector with 1 in the ith coordinate). Then, for all t, for $i=1,2$, with $M_{N,i}(t)$ denoting the ith component of $M_N(t)$,

$$Z_N\big(t,\theta^i\big)\ge\exp\Big(\theta_i M_{N,i}(t)-\frac{1}{2}e^{\gamma}\theta_i^2 t(\lambda_i+\mu_i)\frac{1}{N}\Big).$$

For $\delta>0$, let $T_+^i(\delta)=\inf\{t\ge 0: M_{N,i}(t)>\delta\}$ and let $T_-^i(\delta)=\inf\{t\ge 0: M_{N,i}(t)<-\delta\}$. By optional stopping and Markov's inequality,

$$\mathbb{P}\big(T_+^i(\delta)\le t_0\big)\le\exp\Big(-\theta_i\delta+\frac{1}{2}e^{\gamma}\theta_i^2 t_0\ell\frac{1}{N}\Big).$$

Choosing $\gamma=\log 2$, $\theta_i=N\delta/(2t_0\ell)$, we have $0\le\theta_i\le\gamma N$, as long as $\delta\le(\log 4)t_0\ell$. We then obtain, for $i=1,2$,

$$\mathbb{P}\big(T_+^i(\delta)\le t_0\big)\le e^{-\delta^2 N/(4t_0\ell)}.$$

Arguing similarly about negative δ, we see that, for $i=1,2$,

$$\mathbb{P}\Big(\sup_{t\le t_0}\big|M_{N,i}(t)\big|>\delta\Big)\le 2e^{-\delta^2 N/(4t_0\ell)}.$$

Then the lemma follows from (13). □

4. Intermediate phase: Long-term differential equation approximation. As in the previous section, we use $x_N(t)$ to denote $X_N(t)/N$. The aim of this section is to show that $x_N(t)$ stays concentrated around the solution $x(t)$ of the deterministic system (1) for a long time, provided $x_N(0)$ and $x(0)$ are close to each other, and to the fixed point $((\lambda_1 - \mu_1)/\lambda_1, 0)^T$.

We will treat in detail only the case where the eigenvalues of matrix A are distinct, so that $a \neq 0$. By analogy with the notation in Section 2.2, we let $\tilde{x}_{N,1}(t) = x_{N,1}(t) - \frac{\lambda_1 - \mu_1}{\lambda_1} + \frac{1}{a} x_{N,2}(t)$, $\tilde{x}_{N,2}(t) = x_{N,2}(t)$, and we let $\tilde{x}_N(t)$ be the column vector with components $\tilde{x}_{N,1}(t)$ and $\tilde{x}_{N,2}(t)$.

As in Section 2.2, $L = \min\{\eta_1, \eta_2\}$ and $L_1 = (\lambda_1 + |\lambda_1 - \lambda_2|)\frac{\eta_1 + \eta_2}{\eta_1}$.

Additionally, we let $b = \frac{|a|+1}{|a|}$, $c = \frac{b^2}{2L}(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)$, and $\tilde{L} = \max\{\eta_1, \eta_2\}$.

LEMMA 6. *Suppose* $0 < \omega \leq 4(\log 2)^2 Nc/b^2$. *Let*

$$f_N(t) := \max\{|\tilde{x}_{N,1}(t) - \tilde{x}_1(t)|, |\tilde{x}_{N,2}(t) - \tilde{x}_2(t)||a|^{-1}\}, \quad t \geq 0,$$

and suppose that

$$f_N(0) \leq e^{\tilde{L}} \left(\frac{\omega c}{N}\right)^{1/2}.$$

Suppose also that $y(0) := \max\{\tilde{x}_1(0), |a|^{-1}\tilde{x}_2(0)\} \leq L/8L_1$. *Then*

$$\mathbb{P}\left(\sup_{t \leq \lceil e^{\omega/8} \rceil} f_N(t) > 8e^{\tilde{L}} \left(\frac{\omega c}{N}\right)^{1/2}\right) \leq 8e^{-\omega/8}.$$

The proof of Lemma 6 will follow shortly. By standard theory,

$$\tilde{x}_N(t) = \tilde{x}_N(0) + \int_0^t F(\tilde{x}_N(s))\, ds + M_N(t),$$

where $(M_N(t))$ is a martingale, and $F(x)$ is the drift of $(\tilde{x}_N(t))$ when in state x, given as in (10). Analogously to the deterministic process $x(t)$,

$$\tilde{x}_N(t) = e^{\tilde{A}t} \tilde{x}_N(0) + \int_0^t e^{\tilde{A}(t-s)} \tilde{F}(X_N(s))\, ds + \int_0^t e^{\tilde{A}(t-s)}\, dM_N(s), \tag{14}$$

where $\tilde{A}$ is as in (11) and $\tilde{F}$ is as in (12). (This formula is proven in the same way as the "variation of constants" formula in Lemma 4.1 in Barbour and Luczak [5].)

The following analysis of the martingale transform $\int_0^t e^{\tilde{A}(t-s)}\, dM_N(s)$ is general, and applicable to any finite-dimensional jump Markov chain. We will use it to prove Lemma 6. In what follows, the matrix norm we use is the operator norm induced by the vector $\|\cdot\|_2$ norm.

LEMMA 7. *Let* $(X(t))$ *be a Markov chain with state space* $\mathcal{X} \subseteq \mathbb{R}^k$, *where* k *is a positive integer. For* $x, y \in \mathbb{R}^k$ *such that* $x, x + y \in \mathcal{X}$, *let* $\tilde{q}(x, x + y)$ *denote the rate of jump* y *from* x, *and assume that there is a bound* $B > 0$ *such that* $\|y\|_2 \leq B$ *for each possible jump* y. *Suppose further that, for each* $x \in \mathcal{X}$, *the drift* $F(x) := \sum_y y\tilde{q}(x, x + y)$ *at* x *can be written in the form*

$$F(x) = \tilde{A}x + \tilde{F}(x),$$

where $\tilde{A}$ *is a* $k \times k$ *matrix with nonpositive eigenvalues. Let* $(M(t))$ *be the corresponding Dynkin martingale, that is,*

$$X(t) = X(0) + \int_0^t (\tilde{A}X(s) + \tilde{F}(X(s)))\, ds + M(t).$$

Given a vector $\mathbf{e} \in \mathbb{R}^k$ *with* $\|\mathbf{e}\|_2 = 1$, *let* $M_{\mathbf{e}}(t) = \int_0^t \mathbf{e}^T e^{\tilde{A}(t-s)}\, dM(s)$. *For* $\delta > 0$, *let* $T_{\mathbf{e}}^+(\delta) = \inf\{t \geq 0 : M_{\mathbf{e}}(t) > \delta\}$ *and let* $T_{\mathbf{e}}^-(\delta) = \inf\{t \geq 0 : M_{\mathbf{e}}(t) < -\delta\}$. *Let* $T_{\mathbf{e}}(\delta) = T_{\mathbf{e}}^+(\delta) \wedge T_{\mathbf{e}}^-(\delta)$, *the infimum of times t such that* $M_{\mathbf{e}}(t)$ *exceeds* δ *in absolute value.*

Further, given vector $\mathbf{e} \in \mathbb{R}^k$ *and* $u \in \mathbb{R}_+$, *let* $v_{\mathbf{e}}(x, u) = \sum_y \tilde{q}(x, x + y)(\mathbf{e}^T e^{\tilde{A}u} y)^2$, *and, for* $K > 0$, *let* $S_{\mathbf{e}}(K) = \inf\{t \geq 0 : \int_0^t v_{\mathbf{e}}(X(s), t - s)\, ds > K\}$. *Let* $S_i(K) = S_{\mathbf{e}_i}(K)$, *where* $\mathbf{e}_i$ *is a unit vector with* 1 *in the* i*th coordinate.*

There is a constant D dependent on matrix $\tilde{A}$ *such that the following holds. Suppose* $\mathbf{e}^T$ *is a unit left eigenvector of* $\tilde{A}$ *with eigenvalue* $-\eta$, *where* $\eta \geq 0$. *Then, given* $K, \sigma > 0$, *and* $0 < \omega \leq 4(\log 2)^2 K/(BD)^2$,

$$\mathbb{P}\big(T_{\mathbf{e}}\big(e^{\sigma\eta}\sqrt{\omega K}\big) \leq \sigma\lceil e^{\omega/8}\rceil \wedge S_{\mathbf{e}}(K)\big) \leq 4e^{-\omega/8}.$$

Given numbers $K_1, \dots, K_k > 0$, *let* ω *satisfy* $0 < \omega < 4(\log 2)^2 K_i/(BD)^2$ *for each* i. *Then, for a unit vector* $\mathbf{e}$,

$$\mathbb{P}\left(T_{\mathbf{e}}\left(e^{\|\tilde{A}\|_2\sigma}\sqrt{\omega\sum_i K_i}\right) \leq \sigma\lceil e^{\omega/8}\rceil \wedge \left(\bigwedge_{i=1}^k S_i(K_i)\right)\right) \leq 4ke^{-\omega/8}.$$

If $\tilde{A}$ *is diagonal, then we may take* $D = 1$; *in general,* $D = \max\{\|e^{\tilde{A}t}\|_2 : t \geq 0\}$, *which is finite, since* $\tilde{A}$ *has negative eigenvalues.*

REMARK 1. The function $v_{\mathbf{e}}(x, u)$ defined in the statement of the lemma is a measure of variance of jumps in direction $\mathbf{e}$ for a time-inhomogeneous Markov chain with transition rates $\tilde{q}(x, x + y)$, where jump y at time u gets transformed by the matrix $e^{\tilde{A}u}$.

PROOF. Fix a time $\tau > 0$, and consider M^τ given by

$$\begin{aligned} M^\tau(t) &= \int_0^{t\wedge\tau} e^{\tilde{A}(\tau-s)}\, dM(s) \\ &= \int_0^{t\wedge\tau} e^{\tilde{A}(\tau-s)}\Big(dX(s) - \sum_y y\tilde{q}\big(X(s), X(s) + y\big)\, ds\Big). \end{aligned}$$

Then $(M^\tau(t))$ is a zero mean martingale. Also, for each $t \geq 0$,

$$\int_0^t e^{\tilde{A}(t-s)}\, dM(s) = M^t(t).$$

We now define

$$\begin{aligned} Y^\tau(t) &= \int_0^{t\wedge\tau} e^{\tilde{A}(\tau-s)} F\big(X(s)\big)\, ds + M^\tau(t) \\ &= \int_0^{t\wedge\tau} \sum_y e^{\tilde{A}(\tau-s)} y\tilde{q}\big(X(s), X(s) + y\big)\, ds \\ &\quad + \int_0^{t\wedge\tau} e^{\tilde{A}(\tau-s)}\Big(dX(s) - \sum_y y\tilde{q}\big(X(s), X(s) + y\big)\, ds\Big) \\ &= \int_0^{t\wedge\tau} e^{\tilde{A}(\tau-s)} dX(s). \end{aligned}$$

The process $(Y^\tau(t))$ is a time-inhomogeneous Markov chain, stopped at τ.

For $\theta \in \mathbb{R}^k$, let $R^\tau(t,\theta)$ be defined by

$$R^\tau(t,\theta) = e^{\theta^T Y^\tau(t)} - \int_0^{t\wedge\tau} \sum_y \tilde{q}\big(X(r), X(r)+y\big)\big(e^{\theta^T[Y^\tau(r)+e^{\tilde{A}(\tau-r)}y]} - e^{\theta^T Y^\tau(r)}\big)\,dr.$$

Then $(R^\tau(t,\theta))$ is a martingale. Also, for $\theta \in \mathbb{R}^k$, $(Z^\tau(t,\theta))$ given by

$$Z^\tau(t,\theta) = e^{\theta^T Y^\tau(t)} \exp\biggl(-\int_0^{t\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-s)}y} - 1\big)\,ds\biggr)$$

is a mean 1 martingale, since, for all t, using integration by parts to obtain the first equality,

$$\begin{aligned}
Z^\tau(t,\theta) &= e^{\theta^T Y^\tau(0)} \\
&\quad + \int_0^t \exp\biggl(-\int_0^{r\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-s)}y} - 1\big)\,ds\biggr)\,de^{\theta^T Y^\tau(r)} \\
&\quad - \int_0^{t\wedge\tau} e^{\theta^T Y^\tau(r)} \sum_y \tilde{q}\big(X(r), X(r)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-r)}y} - 1\big) \\
&\quad \times \exp\biggl(-\int_0^{r\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-s)}y} - 1\big)\,ds\biggr)\,dr \\
&= 1 + \int_0^{t\wedge\tau} \exp\biggl(-\int_0^{r\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-s)}y} - 1\big)\,ds\biggr) \\
&\quad \times \biggl(de^{\theta^T Y^\tau(r)} - \sum_y \tilde{q}\big(X(r), X(r)+y\big)\big(e^{\theta^T[Y^\tau(r)+e^{\tilde{A}(\tau-r)}y]} - e^{\theta^T Y^\tau(r)}\big)\,dr\biggr) \\
&= 1 + \int_0^{t\wedge\tau} \exp\biggl(-\int_0^{r\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-s)}y} - 1\big)\,ds\biggr)\,dR^\tau(r,\theta).
\end{aligned}$$

Note that $Z^\tau(t,\theta)$ can be written as

$$\begin{aligned}
Z^\tau(t,\theta) &= \exp\biggl(\theta^T Y^\tau(t) - \int_0^{t\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\theta^T e^{\tilde{A}(\tau-s)}y\,ds\biggr) \\
&\quad \times \exp\biggl(-\int_0^{t\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-s)}y} - 1 - \theta^T e^{\tilde{A}(\tau-s)}y\big)\,ds\biggr) \\
&= \exp\biggl(\theta^T M^\tau(t) \\
&\quad - \int_0^{t\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(e^{\theta^T e^{\tilde{A}(\tau-s)}y} - 1 - \theta^T e^{\tilde{A}(\tau-s)}y\big)\,ds\biggr).
\end{aligned}$$

Since $e^z - 1 - z = z^2 \int_{r=0}^1 e^{rz}(1-r)\,dr$, we see that $Z^\tau(t,\theta)$ equals

$$\exp\biggl(\theta^T M^\tau(t) - \int_0^{t\wedge\tau} \sum_y \tilde{q}\big(X(s), X(s)+y\big)\big(\theta^T e^{\tilde{A}(\tau-s)}y\big)^2 \biggl(\int_0^1 e^{r\theta^T e^{\tilde{A}(\tau-s)}y}(1-r)\,dr\biggr)\,ds\biggr).$$

Recall that each jump y satisfies $\|y\|_2 \le B$, and set $D = \max\{\|e^{\tilde{A}u}\|_2 : u \ge 0\}$: as $\tilde{A}$ has negative eigenvalues, the bound D is always finite, and is equal to 1 if $\tilde{A}$ is diagonal. Since $s \le \tau$, it then follows that, for $\|\theta\|_2 \le \Gamma$,

$$\int_0^1 e^{r\theta^T e^{\tilde{A}(\tau-s)}y}(1-r)\,dr \le \frac{1}{2}e^{B\Gamma\|e^{\tilde{A}(\tau-s)}\|_2} \le \frac{1}{2}e^{B\Gamma D}.$$

Hence, for all t,

$$Z^\tau(t,\theta) \ge \exp\biggl(\theta^T M^\tau(t) - \frac{1}{2}e^{B\Gamma D}\int_0^{t\wedge\tau}\sum_y \tilde{q}\bigl(X(s), X(s)+y\bigr)\bigl(\theta^T e^{\tilde{A}(\tau-s)}y\bigr)^2\,ds\biggr).$$

Writing $\theta = \|\theta\|_2\mathbf{e}$, for a unit vector $\mathbf{e} \in \mathbb{R}^k$,

$$Z^\tau(t,\theta) \ge \exp\biggl(\|\theta\|_2\mathbf{e}^T M^\tau(t) - \frac{1}{2}e^{B\Gamma D}\|\theta\|_2^2\int_0^{t\wedge\tau} v_{\mathbf{e}}\bigl(X(s), \tau-s\bigr)\,ds\biggr).$$

Since $(X(t))$ is right-continuous, for $K > 0$, time $S_{\mathbf{e}}(K)$ defined by $S_{\mathbf{e}}(K) = \inf\{t \ge 0 : \int_0^t v_{\mathbf{e}}(X(s), t-s)\,ds > K\}$ is a stopping time. Given $u > 0$, let

$$S_{\mathbf{e}}^u(K) = \inf\biggl\{t \ge 0 : \int_0^{t\wedge u} v_{\mathbf{e}}\bigl(X(s), u-s\bigr)\,ds > K\biggr\}.$$

Then, using continuity of the integral, necessarily either $S_{\mathbf{e}}^u(K) < u$ or $S_{\mathbf{e}}^u(K) = \infty$. Note that, if $S_{\mathbf{e}}(K) \ge \tau$, then $\int_0^t v_{\mathbf{e}}(X(s), t-s)\,ds \le K$ for all $t \le \tau$ so $S_{\mathbf{e}}^t(K) = \infty$ for all $t \le \tau$.

For $t > 0$ and unit vector $\mathbf{e}$, let $M_{\mathbf{e}}^\tau(t) = \mathbf{e}^T M^\tau(t)$. Also given $\delta > 0$, let $T_{\mathbf{e}}^{\tau,+}(\delta) = \inf\{t \ge 0 : M_{\mathbf{e}}^\tau(t) > \delta\}$, and let $T_i^{\tau,-}(\delta) = \inf\{t \ge 0 : M_{\mathbf{e}}^\tau(t) < -\delta\}$. Then $T_{\mathbf{e}}^{\tau,\pm}(\delta) \le \tau$ or $T_{\mathbf{e}}^{\tau,\pm}(\delta) = \infty$.

Given $K > 0$, on the event $\{T_{\mathbf{e}}^{\tau,+}(\delta) \le \tau \wedge S_{\mathbf{e}}^\tau(K)\}$,

$$Z^\tau\bigl(T_{\mathbf{e}}^{\tau,+}(\delta), \|\theta\|_2\mathbf{e}\bigr) \ge \exp\biggl(\|\theta\|_2\delta - \frac{1}{2}e^{B\Gamma D}\|\theta\|_2^2 K\biggr).$$

By optional stopping and Markov inequality,

$$\mathbb{P}\bigl(T_{\mathbf{e}}^{\tau,+}(\delta) \le \tau \wedge S_{\mathbf{e}}^\tau(K)\bigr) \le \exp\biggl(-\|\theta\|_2\delta + \frac{1}{2}e^{B\Gamma D}\|\theta\|_2^2 K\biggr).$$

Choosing $\|\theta\|_2 = \delta/e^{B\Gamma D}K$, and assuming $\delta \le \Gamma K e^{B\Gamma D}$ so that $\|\theta\|_2 \le \Gamma$, we obtain

$$\mathbb{P}\bigl(T_{\mathbf{e}}^{\tau,+}(\delta) \le \tau \wedge S_{\mathbf{e}}^\tau(K)\bigr) \le e^{-\delta^2/2Ke^{B\Gamma D}},$$

and, similarly,

$$\mathbb{P}\bigl(T_{\mathbf{e}}^{\tau,-}(\delta) \le \tau \wedge S_{\mathbf{e}}^\tau(K)\bigr) \le e^{-\delta^2/2Ke^{B\Gamma D}}.$$

Letting $T_{\mathbf{e}}^\tau(\delta) = T_{\mathbf{e}}^{\tau,+}(\delta) \wedge T_{\mathbf{e}}^{\tau,-}(\delta)$, it follows that

$$\mathbb{P}\bigl(T_{\mathbf{e}}^\tau(\delta) \le \tau \wedge S_{\mathbf{e}}^\tau(K)\bigr) \le 2e^{-\delta^2/2Ke^{B\Gamma D}}.$$

Choosing $\Gamma = (BD)^{-1}\log 2$ and $\delta = \sqrt{\omega K}$, with $0 < \omega < 4(\log 2)^2 K/(BD)^2$, we have $\|\theta\|_2 \le \Gamma$, so

$$\mathbb{P}\bigl(T_{\mathbf{e}}^\tau(\sqrt{\omega K}) \le \tau \wedge S_{\mathbf{e}}^\tau(K)\bigr) \le 2e^{-\omega/4}. \tag{15}$$

Note that, for $t \le \tau$, $M^t(t) = e^{\tilde{A}(t-\tau)}M^\tau(t)$, and so $M_{\mathbf{e}}^t(t) = \mathbf{e}^T e^{\tilde{A}(t-\tau)}M^\tau(t)$. In particular, if $\mathbf{e}^T$ is a left eigenvector of $\tilde{A}$ with eigenvalue $-\eta$ ($\eta > 0$), then $M_{\mathbf{e}}(t) = M_{\mathbf{e}}^t(t) = e^{\eta(\tau-t)}M_{\mathbf{e}}^\tau(t)$. In general,

$$M_{\mathbf{e}}(t) = \mathbf{e}^T e^{\tilde{A}(t-\tau)}M^\tau(t) \le \bigl\|e^{\tilde{A}(t-\tau)}\bigr\|_2\bigl\|M^\tau(t)\bigr\|_2 \le e^{\|\tilde{A}\|_2(\tau-t)}\bigl\|M^\tau(t)\bigr\|_2.$$

Choosing $0 < \sigma < \tau$, if $\mathbf{e}^T$ is a left eigenvector of $\tilde{A}$ with eigenvalue $-\eta$, then

$$\sup_{\tau-\sigma \le t \le \tau} \left| M_{\mathbf{e}}^t(t) \right| \le e^{\sigma\eta} \sup_{\tau-\sigma \le t \le \tau} \left| M_{\mathbf{e}}^\tau(t) \right|.$$

For a general vector $\mathbf{e}$,

$$\sup_{\tau-\sigma \le t \le \tau} \left| M_{\mathbf{e}}^t(t) \right| \le e^{\sigma \|\tilde{A}\|_2} \left\| M^\tau(t) \right\|_2.$$

Applying (15) to all times $\tau_1 = \sigma,\ \tau_2 = 2\sigma,\ \ldots,\ \tau_{j_0} = j_0\sigma,\ \tau_{j_0+1} = S_{\mathbf{e}}(K) \wedge \lceil e^{\omega/8} \rceil \sigma$, where $j_0 \le \lceil e^{\omega/8} \rceil - 1$ is the largest j such that $j\sigma < S_{\mathbf{e}}(K) \wedge \lceil e^{\omega/8} \rceil \sigma$, we see that

$$\mathbb{P}\left(\exists j \le j_0 + 1 : T_{\mathbf{e}}^{\tau_j}(\sqrt{\omega K}) \le \tau_j \wedge S_{\mathbf{e}}^{\tau_j}(K)\right) \le 4e^{-\omega/8}.$$

As noted above, $t \le S_{\mathbf{e}}(K)$ implies $S_{\mathbf{e}}^u(K) = \infty$ for all $u \le t$, and so, for each $j \le j_0 + 1$, we have $S_{\mathbf{e}}^{\tau_j}(K) = \infty$.

So, for $\mathbf{e}^T$ a left eigenvector of $\tilde{A}$ with eigenvalue η, we obtain from the above that

$$\mathbb{P}\left(T_{\mathbf{e}}\left(e^{\sigma\eta}\sqrt{\omega K}\right) \le \sigma \lceil e^{\omega/8} \rceil \wedge S_{\mathbf{e}}(K)\right) \le 4e^{-\omega/8}.$$

In general, given numbers $K_1, \ldots, K_k > 0$, and a vector $\mathbf{e}$, we similarly obtain

$$\mathbb{P}\left(T_{\mathbf{e}}\left(e^{\|\tilde{A}\|_2\sigma}\sqrt{\omega \sum_i K_i}\right) \le \sigma \lceil e^{\omega/8} \rceil \wedge \left(\bigwedge_{i=1}^k S_i(K_i)\right)\right) \le 4ke^{-\omega/8}. \qquad \square$$

PROOF OF LEMMA 6. We will apply Lemma 7 to the chain $(\tilde{x}_N(t))$. The possible jumps y of $(\tilde{x}_N(t))$ are of the form $\pm(1/N, 0)^T$ and $\pm(1/(aN), 1/N)^T$, so $\|y\|_2 \le B := b/N$.

We let $v_i(\tilde{x}_N(s), u) = v_{\mathbf{e}_i}(\tilde{x}_N(s), u)$, where $v_{\mathbf{e}}$ is defined in the statement of Lemma 7. We first bound $\int_0^t v_i(\tilde{x}_N(s), t-s)\, ds = \int_0^t \sum_y \tilde{q}_N(\tilde{x}_N(s), \tilde{x}_N(s) + y)(e^{\tilde{A}(t-s)}y)_i^2\, ds$:

$$\begin{aligned}
\int_0^t v_1(\tilde{x}_N(s), t-s)\, ds &\le \frac{\lambda_1}{N} \int_0^t x_{N,1}(s)\big(1 - x_{N,1}(s) - x_{N,2}(s)\big) e^{-2\eta_1(t-s)}\, ds \\
&\quad + \frac{\mu_1}{N} \int_0^t x_{N,1}(s) e^{-2\eta_1(t-s)}\, ds + \frac{\mu_2}{Na^2} \int_0^t x_{N,2}(s) e^{-2\eta_1(t-s)}\, ds \\
&\quad + \frac{\lambda_2}{Na^2} \int_0^t x_{N,2}(s)\big(1 - x_{N,1}(s) - x_{N,2}(s)\big) e^{-2\eta_1(t-s)}\, ds \\
&\le N^{-1}(\lambda_1 + \mu_1) \int_0^t e^{-2\eta_1(t-s)}\, ds \\
&\quad + N^{-1} \frac{(\lambda_2 + \mu_2)}{a^2} \int_0^t e^{-2\eta_1(t-s)}\, ds \\
&\le N^{-1}(2\eta_1)^{-1}\big[(\lambda_1 + \mu_1) + a^{-2}(\lambda_2 + \mu_2)\big] \le \frac{c}{N}; \\
\int_0^t v_2(\tilde{x}_N(s), t-s)\, ds &\le \frac{\lambda_2}{N} \int_0^t x_{N,2}(s)\big(1 - x_{N,1}(s) - x_{N,2}(s)\big) e^{-2\eta_2(t-s)}\, ds \\
&\quad + \frac{\mu_2}{N} \int_0^t x_{N,2}(s) e^{-2\eta_2(t-s)}\, ds \\
&\le N^{-1}(\lambda_2 + \mu_2) \int_0^t e^{-2\eta_2(t-s)}\, ds \le N^{-1} \frac{\lambda_2 + \mu_2}{2\eta_2} \le \frac{c}{b^2 N}.
\end{aligned}$$

Using (14), (11), (12), and the definition of $f_N(t)$,

$$\begin{aligned}
\big|\tilde{x}_{N,2}(t)-\tilde{x}_2(t)\big| &\le e^{-t\eta_2}\big|\tilde{x}_{N,2}(0)-\tilde{x}_2(0)\big| + \left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right| \\
&\quad + \frac{\lambda_2}{|a|}\frac{\eta_2}{\eta_1}\int_0^t e^{-(t-s)\eta_2}\big|\big(\tilde{x}_{N,2}(s)\big)^2-\big(\tilde{x}_2(s)\big)^2\big|\,ds \\
&\quad + \lambda_2\int_0^t e^{-(t-s)\eta_2}\big|\tilde{x}_{N,1}(s)\tilde{x}_{N,2}(s)-\tilde{x}_1(s)\tilde{x}_2(s)\big|\,ds \\
&\le e^{-t\eta_2}|a|f_N(0) + \left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right| \\
&\quad + \frac{\lambda_2\eta_2}{\eta_1}\int_0^t e^{-(t-s)\eta_2} f_N(s)\big(2\tilde{x}_2(s)+|a|f_N(s)\big)\,ds \\
&\quad + \int_0^t e^{-(t-s)\eta_2}\lambda_2 f_N(s)\big(\tilde{x}_2(s)+|a|f_N(s)+|a|\big|\tilde{x}_1(s)\big|\big)\,ds
\end{aligned}$$

and

$$\begin{aligned}
&\big|\tilde{x}_{N,1}(t)-\tilde{x}_1(t)\big| \\
&\quad\le e^{-t\eta_1}\big|\tilde{x}_{N,1}(0)-\tilde{x}_1(0)\big| + \left|\int_0^t e^{-(t-s)\eta_1}\,dM_{N,1}(s)\right| \\
&\qquad + \lambda_1\int_0^t e^{-(t-s)\eta_1}\big|\tilde{x}_{N,1}(s)^2-\tilde{x}_1(s)^2\big|\,ds \\
&\qquad + \frac{\eta_2|\lambda_1-\lambda_2|}{\eta_1 a^2}\int_0^t e^{-(t-s)\eta_1}\big|\tilde{x}_{N,2}(s)^2-\tilde{x}_2(s)^2\big|\,ds \\
&\qquad + \frac{1}{|a|}\left(|\lambda_1-\lambda_2|+\frac{\lambda_1\eta_2}{\eta_1}\right)\int_0^t e^{-(t-s)\eta_1}\big|\tilde{x}_{N,1}(s)\tilde{x}_{N,2}(s)-\tilde{x}_1(s)\tilde{x}_2(s)\big|\,ds \\
&\quad\le e^{-t\eta_1}f_N(0) + \left|\int_0^t e^{-(t-s)\eta_1}\,dM_{N,1}(s)\right| + \lambda_1\int_0^t e^{-(t-s)\eta_1}f_N(s)\big(2\big|\tilde{x}_1(s)\big|+f_N(s)\big)\,ds \\
&\qquad + \frac{|\lambda_1-\lambda_2|\eta_2}{\eta_1}\int_0^t e^{-(t-s)\eta_1}f_N(s)\left(2\frac{\tilde{x}_2(s)}{|a|}+f_N(s)\right)ds \\
&\qquad + \left(|\lambda_1-\lambda_2|+\frac{\lambda_1\eta_2}{\eta_1}\right)\int_0^t e^{-(t-s)\eta_1}f_N(s)\left(\frac{\tilde{x}_2(s)}{|a|}+f_N(s)+\big|\tilde{x}_1(s)\big|\right)ds.
\end{aligned}$$

Then, by Lemma 1, and since $L_1=(\lambda_1+|\lambda_1-\lambda_2|)(\eta_1+\eta_2)/\eta_1$,

$$\begin{aligned}
\frac{1}{|a|}\big|\tilde{x}_{N,2}(t)-\tilde{x}_2(t)\big| &\le e^{-t\eta_2}f_N(0)+\frac{1}{|a|}\left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right| \\
&\quad + L_1\int_0^t\big(f_N(s)\big)^2 e^{-(t-s)\eta_2}\,ds + 4L_1 y(0)e^{-tL}\int_0^t f_N(s)\,ds
\end{aligned}$$

and

$$\begin{aligned}
\big|\tilde{x}_{N,1}(t)-\tilde{x}_1(t)\big| &\le e^{-t\eta_1}f_N(0)+\left|\int_0^t e^{-(t-s)\eta_1}\,dM_{N,1}(s)\right| \\
&\quad + L_1\int_0^t e^{-(t-s)\eta_1}\big(f_N(s)\big)^2\,ds + 4L_1 y(0)e^{-tL}\int_0^t f_N(s)\,ds.
\end{aligned}$$

Let T_1 be the infimum of times t such that for either $i=1$ or $i=2$

$$\left|\int_0^t e^{-(t-s)\eta_i}\,dM_{N,i}(s)\right| > \left(\frac{\omega c_i}{N}\right)^{1/2} e^{\eta_i},$$

where $c_1 = c$ and $c_2 = c/b^2$. On the event $t < T_1$,

$$\frac{|\tilde{x}_{N,2}(t) - \tilde{x}_2(t)|}{|a|} \le e^{-t\eta_2} f_N(0) + L_1 \int_0^t e^{-(t-s)\eta_2} (f_N(s))^2 \, ds + 4L_1 y(0) e^{-tL} \int_0^t f_N(s)\, ds + \frac{1}{|a|b} \left(\frac{\omega c}{N}\right)^{1/2} e^{\eta_2}$$

and

$$|\tilde{x}_{N,1}(t) - \tilde{x}_1(t)| \le e^{-t\eta_1} f_N(0) + L_1 \int_0^t e^{-(t-s)\eta_1} (f_N(s))^2 \, ds + 4L_1 y(0) e^{-tL} \int_0^t f_N(s)\, ds + \left(\frac{\omega c}{N}\right)^{1/2} e^{\eta_1}.$$

Hence, for $t < T_1$,

$$f_N(t) \le e^{-tL} f_N(0) + L_1 \int_0^t e^{-(t-s)L} (f_N(s))^2 \, ds + 4L_1 y(0) e^{-tL} \int_0^t f_N(s)\, ds + \left(\frac{\omega c}{N}\right)^{1/2} e^{\tilde{L}}.$$

Let $T_2 = \inf\{t : f_N(t) > 10 e^{\tilde{L}} (\frac{\omega c}{N})^{1/2}\}$. Then on the event $t < T_1 \wedge T_2$,

$$f_N(t) \le e^{-tL} f_N(0) + \frac{100 e^{2\tilde{L}} L_1}{L} \frac{\omega c}{N} + 4L_1 y(0) e^{-tL} \int_0^t f_N(s)\, ds + \left(\frac{\omega c}{N}\right)^{1/2} e^{\tilde{L}},$$

so, for N large enough,

$$f_N(t) \le e^{-tL} f_N(0) + 4L_1 y(0) e^{-tL} \int_0^t f_N(s)\, ds + 2\left(\frac{\omega c}{N}\right)^{1/2} e^{\tilde{L}}.$$

Letting $g_N(t) = f_N(t) e^{tL}$, we see that, for large N, on the event $t < T_1 \wedge T_2$,

$$g_N(t) \le g_N(0) + 4L_1 y(0) \int_0^t e^{-sL} g_N(s)\, ds + 2e^{tL} \left(\frac{\omega c}{N}\right)^{1/2} e^{\tilde{L}}.$$

By Grönwall's inequality, for large N, for $t < T_1 \wedge T_2$,

$$g_N(t) \le \left(g_N(0) + 2e^{tL+\tilde{L}} \left(\frac{\omega c}{N}\right)^{1/2}\right) e^{\frac{4L_1 y(0)}{L}},$$

so, if $y(0) \le L/8L_1$ and $f_N(0) \le e^{\tilde{L}} (\omega c/N)^{1/2}$, then

$$f_N(t) \le 2\left(f_N(0) + 2e^{\tilde{L}} \left(\frac{\omega c}{N}\right)^{1/2}\right) \le 6 e^{\tilde{L}} \left(\frac{\omega c}{N}\right)^{1/2}.$$

Fix $0 < t_0 \le e^{\omega/8}$. Let $T_3 = \inf\{t : f_N(t) > 8e^{\tilde{L}} (\frac{\omega c}{N})^{1/2}\}$. We now apply Lemma 7 to $(\tilde{x}_N(t))$, with matrix $\tilde{A}$ as in (11) and $\tilde{F}$ as in (12), $B = b/N$, $\sigma = 1$. We take η to be equal to η_i, $D = 1$, K to be equal to $K_i = c_i N^{-1}$, $\mathbf{e}$ to be equal to $\mathbf{e}_i$, for $i = 1, 2$, and note that we have shown that $\mathbb{P}(\int_0^t v_i(\tilde{x}_N(s), t-s)\, ds \le K_i, i = 1, 2, \forall t \le \lceil e^{\omega/8} \rceil\}) = 1$, and so $\mathbb{P}(S_i(K_i) = \infty) = 1$ for $i = 1, 2$. Lemma 7 then implies that $\mathbb{P}(T_1 \le \lceil e^{\omega/8} \rceil) \le 8e^{-\omega/8}$.

Also, since jumps are of size $O(1/N)$, $\mathbb{P}(T_2 \le T_3) = 0$ for large N. Furthermore, we showed above that $\mathbb{P}(T_3 < T_1 \wedge T_2) = 0$. Then we can only have $T_3 < T_1$ if $T_3 \ge T_2$, and hence $\mathbb{P}(T_3 < T_1) = 0$. It follows that

$$\mathbb{P}(T_3 \le \lceil e^{\omega/8} \rceil) \le \mathbb{P}(T_1 \le \lceil e^{\omega/8} \rceil) + \mathbb{P}(T_3 < T_1) \le 8e^{-\omega/8},$$

which completes the proof of Lemma 6. □

REMARK 2. In the case when $a = 0$, the matrix A has a repeated eigenvalue, and is not diagonalisable. We work with the original variables $x_{N,1}(t)$ and $x_{N,2}(t)$, and let $f_N(t) = \max_{i=1,2} |x_{N,i}(t) - x_i(t)|$. Using Lemmas 3 and 4, and the fact that, for all t, $\lambda_2 e^{-(\lambda_1-\mu_1)t/2}[(\lambda_1 - \mu_1)t + \lambda_1/\lambda_2] \le \lambda_1 + \lambda_2$, we show that

$$f_N(t) \le 2e^{-(\lambda_1-\mu_1)t/2} f_N(0) + 2(\lambda_1 + \lambda_2) \int_0^t e^{-(\lambda_1-\mu_1)(t-s)/2} f_N(s)^2 \, ds$$
$$+ 16(\lambda_1 + \lambda_2) y(0) e^{-(\lambda_1-\mu_1)t/2} \int_0^t f_N(s) \, ds + M,$$

where

$$M = \max\left\{ \left| \int_0^t \left(e^{A(t-s)} \, dM_N(s)\right)_1 \right|, \left| \int_0^t \left(e^{A(t-s)} \, dM_N(s)\right)_2 \right| \right\},$$
$$e^{uA} = \begin{pmatrix} e^{-u(\lambda_1-\mu_1)} & -(\lambda_1 - \mu_1) u e^{-u(\lambda_1-\mu_1)} \\ 0 & e^{-u(\lambda_1-\mu_1)} \end{pmatrix},$$

$y(0) = \max\{|x_1(0) - (\lambda_1 - \mu_1)/\lambda_1|, x_2(0)\}$. The remainder of the analysis can then be carried out in a way analogous to the case with distinct eigenvalues. We apply Lemma 7, taking $\mathbf{e} = \mathbf{e}_i$ for $i = 1, 2$. We further take $B = 1/N$, $K_1 = K_2 = N^{-1}(2\eta)^{-1}(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)$, where $\eta = \eta_1 = \eta_2 = \lambda_1 - \mu_1$. Also, using the fact that the operator-induced norm $\|\cdot\|_2$ of a matrix is at most its Frobenius norm, we can take $D = \sqrt{2}$.

Then we can bound the martingale deviation M by $e^{2(\lambda_1-\mu_1)} \sqrt{\frac{\omega(\lambda_1+\mu_1+\lambda_2+\mu_2)}{N(\lambda_1-\mu_1)}}$, and hence we can show that, for large N, with probability at least $1 - 8e^{-\omega/8}$,

$$f_N(t) \le 8e^{2(\lambda_1-\mu_1)} N^{-1/2} \sqrt{\omega(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)/(\lambda_1 - \mu_1)},$$

over an interval of length $\lceil e^{\omega/8} \rceil$, provided $\omega \le 4(\log 2)^2(\lambda_1 + \lambda_2 + \mu_1 + \mu_2)\eta^{-1} N$, and provided $y(0) \le (\lambda_1 - \mu_1)/64(\lambda_1 + \lambda_2)$, and

$$f_N(0) \le e^{2(\lambda_1-\mu_1)} N^{-1/2} \sqrt{\omega(\lambda_1 + \mu_1 + \lambda_2 + \mu_2)/(\lambda_1 - \mu_1)}.$$

5. Final phase.

LEMMA 8. *For $w \in \mathbb{R}$, let*

$$t_N(y, w) = (\mu_2 - \lambda_2\mu_1/\lambda_1)^{-1}\big(\log y + \log(1 - \lambda_2\mu_1/\lambda_1\mu_2) + w\big).$$

Let $0 < \varepsilon < 1/4$. Suppose that $|x_{N,1}(0) - \frac{\lambda_1-\mu_1}{\lambda_1}| \le N^{-\varepsilon}$ and $|x_{N,2}(0) - N^{-1/4}| \le N^{-1/3}$. Then, the weaker species extinction time κ_N satisfies $\mathbb{P}(\kappa_N \le t_N(N^{3/4}, w)) \to e^{-e^{-w}}$ as $N \to \infty$.

Let $(Y(t))_{t\ge 0}$ be a subcritical linear birth and death chain, with birth and death rates λ and μ, where $\mu > \lambda$. Let $T^Y = \inf\{t \ge 0 : Y(t) = 0\}$. It is a well known fact (see, e.g., Renshaw [23]) that, for $t \ge 0$,

$$\mathbb{P}(T^Y \le t) = \mathbb{P}\big(Y(t) = 0\big) = \left(1 - \frac{(\mu - \lambda)e^{-(\mu-\lambda)t}}{\mu - \lambda e^{-(\mu-\lambda)t}}\right)^{Y(0)}.$$

Assume that $Y(0) = y$, and let

$$t(y, w) = \frac{\log y + \log(\mu - \lambda) - \log \mu + w}{\mu - \lambda}$$

for y and w such that $t(y,w) \ge 0$. Then $e^{-(\mu-\lambda)t(y,w)} = \mu e^{-w}/(\mu-\lambda)y$, so

$$\mathbb{P}(T^Y \le t(y,w)) = \left(1 - \frac{e^{-w}/y}{1-\lambda e^{-w}/(\mu-\lambda)y}\right)^y.$$

Now, consider a sequence $(Y_N(t))$ of linear birth and death chains with birth rate $\lambda = \lambda(N)$ and death rate $\mu = \mu(N)$, where $\mu(N) > \lambda(N)$. Assume further that $Y_N(0) = y(N)$, where $y(N)(\mu-\lambda) \to \infty$. Then, as $N \to \infty$,

$$\begin{aligned}\mathbb{P}(T^{Y_N} \le t(y(N),w)) &= \left(1 - \frac{e^{-w}/y(N)}{1-\lambda e^{-w}/(\mu-\lambda)y(N)}\right)^{y(N)}\\ &= \left(1 - \frac{e^{-w}}{y(N)-\lambda e^{-w}/(\mu-\lambda)}\right)^{y(N)} \to e^{-e^{-w}}.\end{aligned}$$

In other words, the following holds for the asymptotic distribution of the extinction times of a sequence of subcritical linear birth and death chains.

LEMMA 9. *Let $(Y_N(t))$ be a sequence of subcritical linear birth and death chains with birth and death rates $\lambda(N)$ and $\mu(N)$, respectively, where $\mu(N) > \lambda(N)$. Suppose further that $Y_N(0) = y(N)$, where $y(N)(\mu-\lambda) \to \infty$. Let $T^{Y_N} = \inf\{t \ge 0 : Y_N(t) = 0\}$. Then, as $N \to \infty$,*

$$(\mu(N)-\lambda(N))T^{Y_N} - \bigl(\log y(N) + \log(\mu(N)-\lambda(N)) - \log \mu(N)\bigr) \to G,$$

in distribution, where G has a standard Gumbel distribution.

PROOF OF LEMMA 8. Let $x_1(0) = x_{N,1}(0)$ and $x_2(0) = x_{N,2}(0)$, so $|x_1(0) - \frac{\lambda_1-\mu_1}{\lambda_1}| \le N^{-\varepsilon}$ and $x_2(0) \le 2N^{-1/4}$.

By Lemma 2, for large enough N, for all $t \ge 0$, $x_2(t) \le 4N^{-1/4}$. Also, by Lemma 1, if N is large enough, for all $t \ge 0$, $|x_1(t) - (\lambda_1-\mu_1)/\lambda_1| \le 4N^{-\varepsilon}$.

Let Z_N be a linear birth and death process defined as follows. The death rate is μ_2, the birth rate is

$$\lambda_2\left(1 - \frac{\lambda_1-\mu_1}{\lambda_1} + 5N^{-\varepsilon}\right) = \frac{\lambda_2\mu_1}{\lambda_1} + 5\lambda_2 N^{-\varepsilon},$$

and $Z_N(0) = N^{3/4} + N^{2/3}$. By Lemma 9, as $N \to \infty$, in distribution,

$$\begin{aligned}&\left(\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1} - 5\lambda_2N^{-\varepsilon}\right)T^{Z_N}\\ &\quad - \left(\log(N^{3/4}+N^{2/3}) + \log\left(\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1} - 5\lambda_2N^{-\varepsilon}\right) - \log\mu_2\right) \to G,\end{aligned}$$

where G has a standard Gumbel distribution, and so, for $w \in \mathbb{R}$,

$$\mathbb{P}\left(T^{Z_N} \le \frac{\log(N^{3/4}+N^{2/3}) + \log(\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1} - 5\lambda_2N^{-\varepsilon}) - \log\mu_2 + w}{\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1} - 5\lambda_2N^{-\varepsilon}}\right) \to e^{-e^{-w}}.$$

This means that

$$\mathbb{P}\left(T^{Z_N} \le \frac{\log N^{3/4} + \log(\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1}) - \log\mu_2 + w + o(1)}{\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1}}\right) \to e^{-e^{-w}},$$

and so also

$$\left(\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1}\right)T^{Z_N} - \left(\log N^{3/4} + \log\left(\mu_2 - \frac{\lambda_2\mu_1}{\lambda_1}\right) - \log\mu_2\right) \to G.$$

Let W_N be a linear birth and death process defined as follows. The death rate is μ_2, the birth rate is

$$\lambda_2\left(1-\frac{\lambda_1-\mu_1}{\lambda_1}-6N^{-\varepsilon}\right)=\frac{\lambda_2\mu_1}{\lambda_1}-6\lambda_2N^{-\varepsilon},$$

and $W_N(0)=N^{3/4}-N^{2/3}$. By Lemma 9, in distribution,

$$\left(\mu_2-\frac{\lambda_2\mu_1}{\lambda_1}+6\lambda_2N^{-\varepsilon}\right)T^{W_N}-\log\bigl(N^{3/4}(1-N^{-1/12})\bigr)$$
$$-\log\left(\mu_2-\frac{\lambda_2\mu_1}{\lambda_1}+6\lambda_2N^{-\varepsilon}\right)+\log\mu_2\to G,$$

where G has a standard Gumbel distribution. As above, it follows also that

$$\left(\mu_2-\frac{\lambda_2\mu_1}{\lambda_1}\right)T^{W_N}-\left(\log N^{3/4}+\log\left(\mu_2-\frac{\lambda_2\mu_1}{\lambda_1}\right)-\log\mu_2\right)\to G.$$

Let $f_N(t)$ be as in Lemma 6, and let $\mathcal{E}_N(t)$ be the event that $f_N(s)\le N^{-1/3}$ for all $s<t$. For N large enough, on the event $\mathcal{E}_N(t)$, for all $s<t$,

$$-5N^{-\varepsilon}\le x_{N,1}(s)-\frac{\lambda_1-1}{\lambda_1}\le 5N^{-\varepsilon},$$

and, furthermore,

$$0\le x_{N,2}(s)\le 4N^{-1/4}+N^{-1/3}\le N^{-\varepsilon}.$$

Therefore, on the event $\mathcal{E}_N(t)$, we can couple Z_N, W_N and $X_{N,2}$ in such a way that, for $s\le t$,

$$W_N(s)\le X_{N,2}(s)\le Z_N(s).$$

It follows that, on the event $\mathcal{E}_N(t)$, $T^{Z_N}\le t$ implies $\kappa_N\le t$, and $\kappa_N\le t$ implies $T^{W_N}\le t$. Also, by Lemma 6 (with any $\omega=\omega(N)$ such that $\omega(N)/N^{1/3}\to 0$), $\mathbb{P}(\overline{\mathcal{E}_N(t)})\to 0$ as long as $t\le e^{\omega/8}$. So, choosing $\omega(N)=N^{1/4}$, for $t\le e^{\omega/8}$,

$$\mathbb{P}\bigl(\{T^{Z_N}\le t\}\cap\mathcal{E}_N(t)\bigr)\le\mathbb{P}\bigl(\{\kappa_N\le t\}\cap\mathcal{E}_N(t)\bigr)\le\mathbb{P}\bigl(\{T^{W_N}\le t\}\cap\mathcal{E}_N(t)\bigr).$$

Hence, for any fixed w,

$$\mathbb{P}\bigl(\kappa_N\le t(N^{3/4},w)\bigr)\le\mathbb{P}\bigl(T^{W_N}\le t(N^{3/4},w)\bigr)+\mathbb{P}\bigl(\overline{\mathcal{E}_N(t(N^{3/4},w))}\bigr)\to e^{-e^{-w}}$$

and

$$\mathbb{P}\bigl(\kappa_N\le t(N^{3/4},w)\bigr)\ge\mathbb{P}\bigl(T^{Z_N}\le t(N^{3/4},w)\bigr)-\mathbb{P}\bigl(\overline{\mathcal{E}_N(t(N^{3/4},w))}\bigr)\to e^{-e^{-w}},$$

which completes the proof of Lemma 8. □

6. Proof of Theorem 2. By assumption, $x_N(0)=(\alpha_N,\beta_N)^T$, where $\alpha_N\to\alpha$ and $\beta_N\to\beta$ as $N\to\infty$. We let $x(0)=x_N(0)$ as the initial condition for (1). By Theorem 4 and the discussion following it, if $\lambda_1/\mu_1>\lambda_2/\mu_2$ and $\lambda_1/\mu_1>1$, then the fixed point $x^*=(\frac{\lambda_1-\mu_1}{\lambda_1},0)^T$ of (1) is globally attractive, so that there exists $t_0>0$ such that, with $L=\min\{\eta_1,\eta_2\}$, $L_1=(\lambda_1+|\lambda_1-\lambda_2|)(\eta_1+\eta_2)/\eta_1$, as defined in Section 2.2, $\max\{|\tilde{x}_1(t_0)|,|a|^{-1}\tilde{x}_2(t_0)\}\le L/8L_1$. It is also not hard to see that we can choose a finite t_0 that works for every value of N, for N-dependent initial conditions as above.

Let $t_N=\inf\{t\ge t_0:x_2(t)\le N^{-1/4}\}$. Lemma 2 implies that, as $N\to\infty$,

$$t_N=\frac{1}{4\eta_2}\log N+O(1).$$

It then also follows from Lemmas 1 and 2 that there exists $0 < \varepsilon < 1/4$ such that, if N is large enough, then

$$\left| x_1(t_N) - \frac{\lambda_1 - \mu_1}{\lambda_1} \right| \le \frac{1}{2} N^{-\varepsilon}.$$

By Lemma 5 with $\delta = N^{-5/12}$ and Lemma 6 with $\omega = N^{1/4}$, if N is large enough, then with probability at least $1 - e^{-N^{1/12}}$

$$\sup_{t \le t_N} \left| x_{N,1}(t) - x_1(t) \right| \le \frac{1}{2} N^{-1/3} \le \frac{1}{2} N^{-\varepsilon};$$

$$\sup_{t \le t_N} \left| x_{N,2}(t) - x_2(t) \right| \le \frac{1}{2} N^{-1/3}. \tag{16}$$

By (6), the length t_0 of the first phase can be written as

$$\frac{\lambda_2}{\mu_2\lambda_1 - \mu_1\lambda_2} \log\bigl(x_1(t_0)/x_1(0)\bigr) - \frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} \log\bigl(x_2(t_0)/x_2(0)\bigr),$$

and the length $t_N - t_0$ of the second phase can be written as

$$\frac{\lambda_2}{\mu_2\lambda_1 - \mu_1\lambda_2} \log\bigl(x_1(t_N)/x_1(t_0)\bigr) - \frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} \log\bigl(x_2(t_N)/x_2(t_0)\bigr).$$

When (16) holds, we take $x_{N,1}(t_N)$ and $x_{N,2}(t_N)$ as initial values in Lemma 8. On the event $\tilde{\mathcal{E}}$ that (16) holds, by Lemma 8, the length of the third phase is

$$\frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} \left(\log N^{3/4} + \log\left(1 - \frac{\mu_1\lambda_2}{\lambda_1\mu_2}\right)\right) + \frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} G_N,$$

where G_N converges in distribution to a standard Gumbel variable G. Hence, on the event $\tilde{\mathcal{E}}$, the total time κ_N until the extinction of $X_{N,2}$ is

$$\begin{aligned} &\frac{\lambda_2}{\mu_2\lambda_1 - \mu_1\lambda_2} \log\bigl(x_1(t_N)/\alpha_N\bigr) - \frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} \log\bigl(N^{-1/4}/\beta_N\bigr) \\ &\quad + \frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} \left(\log N^{3/4} + \log\left(1 - \frac{\mu_1\lambda_2}{\mu_2\lambda_1}\right)\right) + \frac{\lambda_1}{\mu_2\lambda_1 - \mu_1\lambda_2} G_N. \end{aligned}$$

Since $\mathbb{P}(\tilde{\mathcal{E}}) \to 1$, $\alpha_N \to \alpha$, $\beta_N \to \beta$ as $N \to \infty$, and, for large N, $|x_1(t_N) - (\lambda_1 - \mu_1)/\lambda_1| \le N^{-\varepsilon}$, we conclude

$$\frac{\mu_2\lambda_1 - \mu_1\lambda_2}{\lambda_1} \kappa_N - \left(\log N\beta\left(1 - \frac{\mu_1\lambda_2}{\mu_2\lambda_1}\right) + \frac{\lambda_2}{\lambda_1} \log\left(\frac{1 - \mu_1/\lambda_1}{\alpha}\right)\right) \to G, \tag{17}$$

in distribution, with G a standard Gumbel, so the first part of Theorem 2 follows. The second part of Theorem 2 then follows from Theorem 1, since, after the extinction of the weaker species, the stronger species evolves as a single supercritical logistic epidemic, and we have shown that κ_N is with high probability negligible in comparison with $\tilde{\kappa}_N$.

7. Near-critical phenomena.

In this section, we will prove Theorem 3, showing that the formulae for the extinction times κ_N and $\tilde{\kappa}_N$ of the weaker and stronger species in Theorem 2 extend to near-criticality.

Recall that $\mu_1 = \mu_2 = 1$, and $\lambda_1 = \lambda_1(N)$, $\lambda_2 = \lambda_2(N)$ are such that $\lambda_1 > \lambda_2$, λ_1 is bounded, and $(\lambda_1 - \lambda_2)(\lambda_1 - 1)^{-1} \to 0$. Recall from (2) that we further assume that

$$N(\lambda_1 - \lambda_2)^3(\lambda_1 - 1)^{-1} / \log\log\bigl(N(\lambda_1 - \lambda_2)^2\bigr) \to \infty.$$

We believe the last condition is not best possible, and we only need $N(\lambda_1-\lambda_2)^2 \to \infty$ for Theorem 3 to hold—see Section 7.1.

Note that under our assumptions $(\lambda_2-1)(\lambda_1-1)^{-1} \to 1$. Also, the quantity a defined in (8) satisfies $0<a<1$ for N large enough, and converges to 1 as $N\to\infty$. As before, we assume that $x_{N,1}(0)=N^{-1}X_{N,1}(0)=\alpha_N$, $x_{N,2}(0)=N^{-1}X_{N,2}(0)=\beta_N$, where $\alpha_N\to\alpha>0$ and $\beta_N\to\beta>0$.

The idea of the proof is as follows. The sum $X_{N,1}(t)+X_{N,2}(t)$ behaves approximately like a single stochastic SIS logistic epidemic process with recovery rate 1 and infection rate somewhere between λ_2 and λ_1. (The difference between λ_2 and λ_1 is smaller than the difference between either λ_1 or λ_2 and the recovery rate 1.) Initially, we track the sum $X_{N,1}(t)+X_{N,2}(t)$ by sandwiching it between two SIS logistic epidemics, until it reaches near $(\lambda_1-1)/\lambda_1$; it will then stay near there for a very long time with high probability. Simultaneously, we track the ratio $X_{N,1}(t)/X_{N,2}(t)$, which has a positive but slow drift and stays close to $X_{N,1}(0)/X_{N,2}(0)$ for a time almost as long as $(\lambda_1-\lambda_2)^{-1}$.

Subsequently, we show that the transformed variables $\tilde{x}_{N,1}(t)=x_{N,1}(t)-\frac{\lambda_1-1}{\lambda_1}+\frac{1}{a}x_{N,2}(t)$, $\tilde{x}_{N,2}(t)=x_{N,2}(t)$ follow closely the deterministic process for a long time, until $\tilde{x}_{N,2}(t)=N^{-1}X_{N,2}(t)$ is $o(\lambda_1-\lambda_2)$ and $N^{-1}X_{N,1}(t)=\tilde{x}_{N,1}(t)+\frac{\lambda_1-1}{\lambda_1}-\frac{1}{a}\tilde{x}_{N,2}(t)$ is within distance $o(\lambda_1-\lambda_2)$ of its carrying capacity $(\lambda_1-1)/\lambda_1$. From then on until extinction, we approximate $X_{N,2}(t)$ by a linear birth-and-death chain with birth rate λ_2/λ_1 and death rate 1.

The first lemma compares $X_{N,1}(t)+X_{N,2}(t)$ to single stochastic SIS logistic epidemic processes with suitable parameter values.

LEMMA 10. *Let $Y_N(t)$ denote the number of infectives in a stochastic logistic SIS epidemic with infection rate λ_2 and recovery rate 1. If $X_{N,1}(0)+X_{N,2}(0)\ge Y_N(0)$, then $X_{N,1}(t)+X_{N,2}(t)$ stochastically dominates $Y_N(t)$.*

Let $Z_N(t)$ denote the number of infectives in a stochastic SIS logistic epidemic with infection rate λ_1 and recovery rate 1. If $X_{N,1}(0)+X_{N,2}(0)\le Z_N(0)$, then $X_{N,1}(t)+X_{N,2}(t)$ is stochastically dominated by $Z_N(t)$.

PROOF. The process $X_{N,1}(t)+X_{N,2}(t)$ jumps by $+1$ at rate at least $\lambda_2(X_{N,1}+X_{N,2})\times(1-X_{N,1}-X_{N,2})$ and jumps by -1 at rate $X_{N,1}+X_{N,2}$. Thus we can couple $X_{N,1}+X_{N,2}$ and $Y_N(t)$ so they always jump down together as much as possible, and jump up together as much as possible, and otherwise each jumps on its own with any excess rate in either direction. With this coupling, $X_{N,1}(t)+X_{N,2}(t)\ge Y_N(t)$. The second part can be proved analogously. $\square$

Next, we establish concentration of measure for the supercritical stochastic SIS logistic epidemic, showing that it follows the deterministic epidemic for a long time. We will use this to show that $X_{N,1}(t)+X_{N,2}(t)$ rapidly arrives near the carrying capacity of the stronger species.

LEMMA 11. *Let $Y_N(t)$ be the number of infectives in a stochastic SIS logistic epidemic with infection rate $\lambda=\lambda(N)$ and recovery rate $\mu=\mu(N)$, where $\lambda>\mu>0$. Let $y(t)$ denote the proportion of infectives in the corresponding deterministic SIS logistic epidemic. Suppose that λ,μ are bounded, and $(\lambda-\mu)^2N\to\infty$ as $N\to\infty$. Let $\omega=\omega(N)>0$ be such that $N(\lambda-\mu)^2/\omega\to\infty$ as $N\to\infty$.*

We assume that $0<y(0)\le 2(\lambda-\mu)/\lambda$, and that $|N^{-1}Y_N(0)-y(0)|\le\frac{1}{2}\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}$. If $y(0)\ge(\lambda-\mu)/\lambda$, then, for N sufficiently large,

$$\mathbb{P}\left(\sup_{t\le(\lambda-\mu)^{-1}e^{\omega/8}}|N^{-1}Y_N(t)-y(t)|>4e^2\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}\right)\le 4e^{-\omega/8}.$$

If $y(0) < (\lambda-\mu)/\lambda$, then, for N sufficiently large,

$$\mathbb{P}\left(\sup_{t\le(\lambda-\mu)^{-1}e^{\omega/8}}\left|N^{-1}Y_N(t)-y(t)\right| > 4e^{\frac{2(\lambda-\mu)}{\lambda y(0)}}\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}\right)\le 4e^{-\omega/8}.$$

PROOF. Since $\lambda > \mu$, the solution $y(t)$ to

$$\frac{dy(t)}{dt} = \lambda y(t)\bigl(1-y(t)\bigr)-\mu y(t), \tag{18}$$

converges to $(\lambda-\mu)/\lambda$ as $t\to\infty$.

Let $\tilde{y}(t) = y(t)-(\lambda-\mu)/\lambda$ and let $\tilde{y}_N(t) = N^{-1}Y_N(t)-(\lambda-\mu)/\lambda$. Then

$$\tilde{y}(t) = \tilde{y}(0)-(\lambda-\mu)\int_0^t \tilde{y}(s)\,ds-\lambda\int_0^t\tilde{y}(s)^2\,ds$$

and

$$\tilde{y}_N(t) = \tilde{y}_N(0)-(\lambda-\mu)\int_0^t\tilde{y}_N(s)\,ds-\lambda\int_0^t\tilde{y}_N(s)^2\,ds+m_N(t),$$

where $(m_N(t))$ is a zero-mean martingale. Treating this equation as a perturbation of a linear equation, see Chapter 6 of Pazy [22], cf. Section 2.2 and equation (14) in the present paper, it follows that

$$\tilde{y}(t) = e^{-(\lambda-\mu)t}\tilde{y}(0)-\lambda\int_0^t e^{-(\lambda-\mu)(t-s)}\tilde{y}(s)^2\,ds$$

and

$$\tilde{y}_N(t) = e^{-(\lambda-\mu)t}\tilde{y}_N(0)-\lambda\int_0^t e^{-(\lambda-\mu)(t-s)}\tilde{y}_N(s)^2\,ds+\int_0^t e^{-(\lambda-\mu)(t-s)}\,dm_N(s).$$

Letting $f_N(t) = |\tilde{y}(t)-\tilde{y}_N(t)| = |y(t)-N^{-1}Y_N(t)|$, we thus have

$$\begin{aligned}f_N(t) &\le f_N(0)e^{-(\lambda-\mu)t}+\lambda\int_0^t e^{-(\lambda-\mu)(t-s)}\bigl(f_N(s)\bigr)^2\,ds\\&\quad+2\lambda\int_0^t e^{-(\lambda-\mu)(t-s)}f_N(s)\bigl|\tilde{y}(s)\bigr|\,ds+\left|\int_0^t e^{-(\lambda-\mu)(t-s)}\,dm_N(s)\right|.\end{aligned}$$

Let $y_N(t) = N^{-1}Y_N(t)$, and $T_1 = \inf\{t : y_N(t) > 2y(t)\}$. To estimate the deviations of the martingale transform $\int_0^t e^{-(\lambda-\mu)(t-s)}\,dm_N(s)$, let $v(y,u) = e^{-2(\lambda-\mu)u}N^{-1}(\lambda y(1-y)+\mu y)$, and note that, on the event $t < T_1$,

$$\begin{aligned}\int_0^t v\bigl(\tilde{y}_N(s),t-s\bigr)\,ds &= \frac{\lambda}{N}\int_0^t y_N(s)\bigl(1-y_N(s)\bigr)e^{-2(\lambda-\mu)(t-s)}\,ds\\&\quad+\frac{\mu}{N}\int_0^t y_N(s)e^{-2(\lambda-\mu)(t-s)}\,ds\\&\le\frac{2(\lambda+\mu)}{N}\int_0^t e^{-2(\lambda-\mu)(t-s)}y(s)\,ds.\end{aligned}$$

As is well known, the solution $y(t)$ to (18) above satisfies

$$y(t) = \frac{y(0)(\lambda-\mu)}{\lambda y(0)+\lambda e^{-(\lambda-\mu)t}(\frac{\lambda-\mu}{\lambda}-y(0))}, \tag{19}$$

and so $\tilde{y}(t) = y(t)-(\lambda-\mu)/\lambda$ satisfies

$$\tilde{y}(t) = \frac{\tilde{y}(0)e^{-(\lambda-\mu)t}}{y(0)\frac{\lambda}{\lambda-\mu}(1-e^{-(\lambda-\mu)t})+e^{-(\lambda-\mu)t}}. \tag{20}$$

Assuming first $y(0) \geq (\lambda-\mu)/\lambda$, we have $(\lambda-\mu)/\lambda \leq y(t) \leq y(0) \leq 2(\lambda-\mu)/\lambda$. It follows that, on the event $t < T_1$,

$$\int_0^t v(\tilde{y}_N(s), t-s)\,ds \leq \frac{2(\lambda+\mu)}{N\lambda}.$$

Given $\omega = \omega(N) > 0$, let T_2 be the infimum of times t such that

$$\left|\int_0^t e^{-(\lambda-\mu)(t-s)}\,dm_N(s)\right| > 3\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}.$$

By Lemma 7 applied to $(\tilde{y}_N(t))$, with $k=1$, $B=1/N$, $\tilde{A} = -(\lambda-\mu)$, $D=1$, $\sigma = \frac{1}{\lambda-\mu}$, $\eta = \lambda-\mu$, $K = \frac{2(\lambda+\mu)}{N\lambda}$, we see that, if $\omega \leq 8(\log 2)^2 N(\lambda+\mu)/\lambda$ (which holds for N large enough if $N(\lambda-\mu)^2/\omega \to \infty$) and $t_0(N) \leq \lceil e^{\omega/8}\rceil/(\lambda-\mu)$, then

$$\mathbb{P}(T_2 \leq T_1 \wedge t_0) \leq 4e^{-\omega/8}.$$

Also, by the above, and using the assumption that $f_N(0) \leq \frac{1}{2}\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}$, on the event $t < T_1 \wedge T_2$,

$$\begin{aligned} f_N(t) &\leq \lambda \int_0^t e^{-(\lambda-\mu)(t-s)}\big(f_N(s)\big)^2\,ds + 2\lambda\int_0^t e^{-(\lambda-\mu)(t-s)} f_N(s)\big|\tilde{y}(s)\big|\,ds \\ &\quad + \frac{7}{2}\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}. \end{aligned}$$

Let T_3 be the infimum of times t such that $f_N(t) > 5e^4\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}$. Then, if N is large enough, on the event $t < T_1 \wedge T_2 \wedge T_3$,

$$f_N(t) \leq \frac{\lambda}{\lambda-\mu}\frac{50e^8\omega(\lambda+\mu)}{N\lambda} + 2\lambda\int_0^t e^{-(\lambda-\mu)(t-s)} f_N(s)\big|\tilde{y}(s)\big|\,ds + \frac{7}{2}\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}.$$

Since $N(\lambda-\mu)^2/\omega \to \infty$, then, for N large enough,

$$\frac{\lambda}{\lambda-\mu}\frac{50e^8\omega(\lambda+\mu)}{N} \leq \frac{1}{2}\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}},$$

and so, for N large enough, on the event $t < T_1 \wedge T_2 \wedge T_3$,

$$f_N(t)e^{(\lambda-\mu)t} \leq 2\lambda\int_0^t e^{(\lambda-\mu)s} f_N(s)\big|\tilde{y}(s)\big|\,ds + 4e^{(\lambda-\mu)t}\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}.$$

From (20), $\tilde{y}(t) \leq e^{-(\lambda-\mu)t}\tilde{y}(0) \leq e^{-(\lambda-\mu)t}(\lambda-\mu)/\lambda$, so, by Grönwall's inequality, for $t < T_1 \wedge T_2 \wedge T_3$,

$$f_N(t) \leq 4\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}e^2.$$

Hence, if T_4 is the infimum of t such that $f_N(t) > 4\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}e^4$, then

$$\mathbb{P}(T_4 \leq t_0) \leq \mathbb{P}(T_1 \wedge T_2 \wedge T_3 \leq T_4 \wedge t_0),$$

and so, since $\mathbb{P}(T_3 \leq T_4) = 0$ and $\mathbb{P}(T_1 \leq T_4) = 0$ for N large enough,

$$\mathbb{P}(T_4 \leq t_0) \leq \mathbb{P}(T_1 \leq T_4) + \mathbb{P}(T_3 \leq T_4) + \mathbb{P}(T_2 \leq T_1 \wedge t_0) \leq 4e^{-\omega/8},$$

and so, as claimed, for N large enough,

$$\mathbb{P}\Big(\sup_{t\le(\lambda-\mu)^{-1}\lceil e^{\omega/8}\rceil}\big|N^{-1}Y_N(t)-y(t)\big|>4e^4\sqrt{\frac{2\omega(\lambda+\mu)}{N\lambda}}\Big)\le 4e^{-\omega/8}.$$

The case $0<y(0)\le\frac{\lambda-\mu}{\lambda}$ is similar. Here we use that, from (19), $y(0)\le y(t)\le\frac{\lambda-\mu}{\lambda}$, and, from (20),

$$\big|\tilde{y}(t)\big|\le\frac{|\tilde{y}(0)|(\lambda-\mu)}{y(0)\lambda}e^{-(\lambda-\mu)t}.\qquad\square$$

LEMMA 12. *Let $Y_N(t)$ denote the number of infective individuals in a stochastic SIS logistic epidemic with infection rate $\lambda=\lambda(N)$ and recovery rate $\mu=\mu(N)$. Let $y(t)$ denote the proportion of infectives in the corresponding deterministic SIS logistic epidemic. We assume that $\lambda=\lambda(N)$ and $\mu=\mu(N)$, where λ,μ are bounded, $\lambda>\mu>0$, and $(\lambda-\mu)^2N\to\infty$ as $N\to\infty$.*

Let $y_N(t)=N^{-1}Y_N(t)$, and assume that $y_N(0)>2(\lambda-\mu)/\lambda$ and $y(0)=y_N(0)$. Let $\tau=\tau(N)$ be such that $y(\tau)=2(\lambda-\mu)/\lambda$. Then, for large N,

$$\mathbb{P}\Big(\sup_{t\le\tau}\big|y_N(t)-y(t)\big|>4\big(N(\lambda-\mu)^2\big)^{1/16}\sqrt{\frac{y_N(0)(\lambda+\mu)}{N(\lambda-\mu)}}\Big)\le 2e^{-(N(\lambda-\mu)^2)^{1/8}/8}.$$

PROOF. We have

$$y_N(t)-y(t)=\big(y_N(0)-y(0)\big)-\lambda\int_0^t\big(y_N(s)-y(s)\big)\Big(y_N(s)+y(s)-\frac{\lambda-\mu}{\lambda}\Big)\,ds+m_N(s).$$

For all s, $y(s)\ge(\lambda-\mu)/\lambda$, so $y_N(s)+y(s)-(\lambda-\mu)/\lambda\ge 0$. By Lemma 3.2 in Brightwell, House and Luczak [6],

$$\sup_{t\le\tau}\big|y_N(t)-y(t)\big|\le 2\big|y_N(0)-y(0)\big|+2\sup_{t\le\tau}\big|m_N(s)\big|=2\sup_{t\le\tau}\big|m_N(s)\big|.$$

From (19),

$$e^{(\lambda-\mu)\tau}-1=1-\frac{2(\lambda-\mu)}{\lambda y(0)}.$$

Arguing as in the proof of Lemma 3.1 in Brightwell, House and Luczak [6], using standard martingale techniques (as in the proof of Lemma 5), and the fact that

$$\int_0^t y(s)\,ds=\frac{1}{\lambda}\log\big(\lambda y(0)\big(e^{(\lambda-\mu)t}-1\big)+(\lambda-\mu)\big)-\frac{1}{\lambda}\log(\lambda-\mu),$$

we see that, if $\phi=\phi(N)\le N^{1/2}$, then

$$\mathbb{P}\Big(\sup_{t\le\tau}\big|y_N(t)-y(t)\big|>4\sqrt{y_N(0)\phi(\lambda+\mu)/N(\lambda-\mu)}\Big)\le 2e^{-\phi/8},$$

and so, taking $\phi=(N(\lambda-\mu)^2)^{1/8}$,

$$\mathbb{P}\Big(\sup_{t\le\tau}\big|y_N(t)-y(t)\big|>4\big(N(\lambda-\mu)^2\big)^{1/16}\sqrt{\frac{y_N(0)(\lambda+\mu)}{N(\lambda-\mu)}}\Big)\le 2e^{-(N(\lambda-\mu)^2)^{1/8}/8}.\qquad\square$$

Next we consider the ratio $Q(X_N(t))=X_{N,1}(t)/X_{N,2}(t)$. We will show that the value of $Q(X_N(t))$ does not change very much over a time period of length "nearly" $(\lambda_1-\lambda_2)^{-1}$. Recall that $N^{-1}X_{N,1}(0)=\alpha_N$ and $N^{-1}X_{N,2}(0)=\beta_N$, where $\alpha_N\to\alpha$ and $\beta_N\to\beta$ as $N\to\infty$.

LEMMA 13. *Suppose that $X_{N,2}(0) \geq 2N(\lambda_1 - \lambda_2)$. Let $\psi = \psi(N) \to \infty$ in such a way that $(\lambda_1 - \lambda_2)\psi \to 0$ as $N \to \infty$. Let $t_0 = (\lambda_1 - \lambda_2)^{-1}\psi^{-1}$. Then, for N large enough,*

$$\mathbb{P}\left(\sup_{t \leq t_0}\left|\frac{X_{N,1}(t)}{X_{N,2}(t)} - \frac{X_{N,1}(0)}{X_{N,2}(0)}\right| > 2\psi^{-1/4}\right) \leq 2e^{-\frac{\psi^{1/2}\beta^2}{128\alpha(\alpha+\beta)}} + 4e^{-\sqrt{N}(\lambda_2 - 1)}.$$

PROOF. Let $T_0 = \inf\{t \geq 0 : X_{N,2}(t) \leq 1\}$. Given a vector $X = (X_1, X_2)^T$ with integer components, such that $X_1 \geq 0$, $X_2 > 1$ and $X_1 + X_2 \leq N$, the drift $g_N(X)$ in $Q(X)$ is

$$\begin{aligned}
&\frac{1}{X_2}\lambda_1 X_1\left(1 - \frac{X_1}{N} - \frac{X_2}{N}\right) - \frac{X_1}{X_2} \\
&\qquad + X_1\left(\frac{1}{X_2 + 1} - \frac{1}{X_2}\right)\lambda_2 X_2\left(1 - \frac{X_1}{N} - \frac{X_2}{N}\right) + X_1\left(\frac{1}{X_2 - 1} - \frac{1}{X_2}\right)X_2 \\
&\quad = \left(1 - \frac{X_1}{N} - \frac{X_2}{N}\right)\left(\frac{\lambda_1 X_1}{X_2} - \frac{\lambda_2 X_1}{X_2 + 1}\right) - \frac{X_1}{X_2} + \frac{X_1}{X_2 - 1}.
\end{aligned}$$

Clearly, we see that $g_N(X_N(t)) \geq 0$ for all $t < T_0$. Also,

$$\begin{aligned}
g_N(X) &= \frac{X_1}{X_2}(\lambda_1 - \lambda_2) + \lambda_2 X_1\left(\frac{1}{X_2} - \frac{1}{X_2 + 1}\right) - X_1\left(\frac{1}{X_2} - \frac{1}{X_2 - 1}\right) \\
&\qquad + \left(\frac{X_1}{N} + \frac{X_2}{N}\right)X_1\left(\frac{\lambda_2}{X_2 + 1} - \frac{\lambda_1}{X_2}\right) \\
&\leq \frac{X_1}{X_2}(\lambda_1 - \lambda_2) + \lambda_2 X_1\left(\frac{1}{X_2} - \frac{1}{X_2 + 1}\right) - X_1\left(\frac{1}{X_2} - \frac{1}{X_2 - 1}\right) \\
&= \frac{X_1}{X_2}\left(\lambda_1 - \lambda_2 + \frac{\lambda_2}{X_2 + 1} + \frac{1}{X_2 - 1}\right).
\end{aligned}$$

Let $T_1 = \inf\{t \geq 0 : X_{N,2}(t) - 1 < N(\lambda_1 - \lambda_2)\}$. Then, since $N(\lambda_1 - \lambda_2)^2 \to \infty$, we have $0 \leq g_N(X_N(t)) \leq 3Q(X_N(t))(\lambda_1 - \lambda_2)$ for $t < T_0 \wedge T_1$, if N is large enough.

We write $Q(X_N(t)) = Q(X_N(0)) + \int_0^t g_N(X_N(s))\,ds + M_N(t)$, where $M_N(t)$ is a martingale. Let $R_N(X)$ be given by

$$\begin{aligned}
&\frac{\lambda_1 X_1}{X_2^2}\left(1 - \frac{X_1}{N} - \frac{X_2}{N}\right) + \frac{X_1}{X_2^2} + X_1^2\left(\frac{1}{X_2 + 1} - \frac{1}{X_2}\right)^2 \lambda_2 X_2\left(1 - \frac{X_1}{N} - \frac{X_2}{N}\right) \\
&\qquad + X_1^2\left(\frac{1}{X_2 - 1} - \frac{1}{X_2}\right)^2 X_2.
\end{aligned}$$

We denote $q_0 := Q(X_N(0)) = \alpha_N/\beta_N$. Let T_2 be the infimum of times t such that $Q(X_N(t)) > 2q_0$. It is easily seen that, for $t < T_0 \wedge T_1 \wedge T_2$, if N is large enough,

$$R_N(X_N(t)) \leq 4q_0(1 + q_0)(\lambda_1 - \lambda_2).$$

Letting $T(\delta) = \inf\{t \geq 0 : |M_N(t)| > \delta\}$, a standard exponential martingale argument, as in the proof of Lemma 5, shows that, if $\delta \leq t_0$ and N is large enough, then

$$\mathbb{P}(T(\delta) \leq t_0 \wedge T_0 \wedge T_1 \wedge T_2) \leq 2e^{-\delta^2/16t_0q_0(1+q_0)(\lambda_1-\lambda_2)}.$$

By Grönwall's inequality, on the event $t_0 < T_0 \wedge T_1 \wedge T_2 \wedge T(\delta)$,

$$\sup_{t \leq t_0} Q(X_N(t)) \leq \left(q_0 + \sup_{t \leq t_0}|M_N(t)|\right)e^{3(\lambda_1 - \lambda_2)t_0} \leq (q_0 + \delta)e^{3(\lambda_1 - \lambda_2)t_0}.$$

Furthermore, on the event $t_0 < T_0 \wedge T_1 \wedge T_2 \wedge T(\delta)$, $\inf_{t \le t_0} Q(X_N(t)) \ge q_0 - \delta$. In other words, on the event $t_0 < T_0 \wedge T_1 \wedge T_2 \wedge T(\delta)$,

$$\sup_{t \le t_0} |Q(X_N(t)) - q_0| \le \delta e^{3(\lambda_1 - \lambda_2)t_0} + q_0(e^{3(\lambda_1 - \lambda_2)t_0} - 1).$$

Let $\delta = \psi(N)^{-1/4}$, so $\delta \le t_0$ for N sufficiently large. It follows that, for N sufficiently large, on the event $t_0 < T_0 \wedge T_1 \wedge T_2 \wedge T(\psi^{-1/4})$,

$$\sup_{t \le t_0} |Q(X_N(t)) - q_0| \le 2\psi^{-1/4}.$$

Let $T_3 = \inf\{t \ge 0 : |Q(X_N(t)) - q_0| > 2\psi(N)^{-1/4}\}$. Clearly, $\mathbb{P}(T_2 \le T_3) = 0$ and $\mathbb{P}(T_0 \le T_1) = 0$ for large N. Then, from the above,

$$\begin{aligned}
\mathbb{P}(T_3 \le t_0) &\le \mathbb{P}(T_0 \wedge T_1 \wedge T_2 \wedge T(\psi^{-1/4}) \le t_0 \wedge T_3) \\
&\le \mathbb{P}(T(\psi^{-1/4}) \le t_0 \wedge T_0 \wedge T_1 \wedge T_2) + \mathbb{P}(T_1 \le t_0 \wedge T_3) \\
&\le 2e^{-\psi(N)^{1/2}/(16q_0(1+q_0))} + \mathbb{P}(T_1 \le t_0 \wedge T_3).
\end{aligned}$$

Let $T_4 = \inf\{t \ge 0 : X_{N,1}(t) + X_{N,2}(t) < N(\lambda_1 - 1)/4\}$. Then, if N is sufficiently large, $\mathbb{P}(T_1 \le t_0 \wedge T_3) \le \mathbb{P}(T_4 \le t_0)$, since $\alpha/2\beta \le q_0 \le 2\alpha/\beta$. We will use Lemma 10, and Lemma 11, with $\lambda = \lambda_2$, $\mu = 1$, $\omega = 8\sqrt{N(\lambda_2 - 1)^2}$. Note $(\lambda_2 - 1)^{-1} e^{\sqrt{N(\lambda_2-1)^2}} \ge (\lambda_1 - \lambda_2)^{-1} \ge t_0$ for large N, since

$$e^{\sqrt{N(\lambda_2-1)^2}} \ge e^{(\frac{N(\lambda_2-1)^2}{N(\lambda_1-\lambda_2)^2})^{1/2}} \ge \left(\frac{N(\lambda_2 - 1)^2}{N(\lambda_1 - \lambda_2)^2}\right)^{1/2} = \frac{\lambda_2 - 1}{\lambda_1 - \lambda_2}.$$

Hence $\mathbb{P}(T_4 \le t_0) \le 4e^{-\sqrt{N}(\lambda_2 - 1)}$, and the result follows. $\square$

After time $(\lambda_1 - \lambda_2)^{-1}\psi^{-1}$, we approximate vector $\tilde{x}_N(t)$ by the solution $\tilde{x}(t)$ to (9). When $\mu_1 = \mu_2 = 1$, equation (9) takes the form:

$$\begin{aligned}
\frac{d\tilde{x}_1(t)}{dt} &= -(\lambda_1 - 1)\tilde{x}_1(t) - \lambda_1 \tilde{x}_1(t)^2 - \frac{(\lambda_1 - \lambda_2)^2}{\lambda_1(\lambda_1 - 1)} \left(\frac{\tilde{x}_2(t)}{a}\right)^2 \\
&\quad + \frac{(\lambda_1 - \lambda_2)\lambda_1}{\lambda_1 - 1} \tilde{x}_1(t) \frac{\tilde{x}_2(t)}{a},
\end{aligned}$$

$$\frac{d\tilde{x}_2(t)}{dt} = -\frac{\lambda_1 - \lambda_2}{\lambda_1} \tilde{x}_2(t) - \lambda_2 \tilde{x}_2(t)\tilde{x}_1(t) + \frac{\lambda_2}{\lambda_1 a} \frac{\lambda_1 - \lambda_2}{\lambda_1 - 1} \tilde{x}_2(t)^2. \tag{21}$$

The next lemma gives what seems to be the best possible error for $\tilde{x}_{N,1}(t)$, with magnitude close to the corresponding martingale transform.

LEMMA 14. *Assume that*

$$N(\lambda_1 - 1)/\lambda_1 - Nh_N \le X_{N,1}(0) + X_{N,2}(0) \le 2N(\lambda_1 - 1)/\lambda_1$$

for some function h_N such that $h_N/(\lambda_1 - 1) \to 0$, and that $X_{N,1}(0)/X_{N,2}(0) = \alpha/\beta + \varepsilon_N$, where $\varepsilon_N \to 0$ as $N \to \infty$. Let $x_i(0) = N^{-1}X_{N,i}(0)$ for $i = 1, 2$. Let $\omega(N) \to \infty$ be such that $N(\lambda_1 - \lambda_2)^2/\omega(N) \to \infty$ as $N \to \infty$. Let

$$f_N(t) = \max\left\{\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2} |\tilde{x}_{N,1}(t) - \tilde{x}_1(t)|, |\tilde{x}_{N,2}(t) - \tilde{x}_2(t)|\right\}, \quad t \ge 0.$$

Then, for N large enough,

$$\mathbb{P}\left(\sup_{t \le (\lambda_1 - 1)^{-1} e^{\omega/8}} f_N(t) > 9 \frac{\lambda_1 - 1}{\lambda_1 - \lambda_2} \sqrt{\frac{\omega}{N}} e^{32(\lambda_1 \frac{\beta}{\alpha} + 1)}\right) \le 12e^{-\omega/8}.$$

PROOF. From the integral form of (21) and its stochastic analogue, noting that $\eta_1 = \lambda_1 - 1$, $\eta_2 = (\lambda_1 - \lambda_2)/\lambda_1$ and that $\eta_1 > \eta_2$ for large N, we obtain

$$\begin{aligned}
\left|\tilde{x}_{N,1}(t) - \tilde{x}_1(t)\right| \le{}& \left|\tilde{x}_{N,1}(0) - \tilde{x}_1(0)\right| e^{-t\eta_1} + \left|\int_0^t e^{-\eta_1(t-s)}\, dM_{N,1}(s)\right| \\
&+ \lambda_1 \int_0^t e^{-\eta_1(t-s)} \left|\tilde{x}_{N,1}(s) - \tilde{x}_1(s)\right| \left|\tilde{x}_{N,1}(s) + \tilde{x}_1(s)\right| ds \\
&+ \frac{(\lambda_1 - \lambda_2)^2}{\lambda_1(\lambda_1 - 1)a^2} \int_0^t e^{-\eta_1(t-s)} \left|\tilde{x}_{N,2}(s) - \tilde{x}_2(s)\right| \bigl(\tilde{x}_{N,2}(s) + \tilde{x}_2(s)\bigr)\, ds \\
&+ \frac{(\lambda_1 - \lambda_2)\lambda_1}{a(\lambda_1 - 1)} \int_0^t e^{-\eta_1(t-s)} \left|\tilde{x}_{N,1}(s) - \tilde{x}_1(s)\right| \tilde{x}_{N,2}(s)\, ds \\
&+ \frac{(\lambda_1 - \lambda_2)\lambda_1}{a(\lambda_1 - 1)} \int_0^t e^{-\eta_1(t-s)} \left|\tilde{x}_{N,2}(s) - \tilde{x}_2(s)\right| \left|\tilde{x}_1(s)\right| ds
\end{aligned}$$

and

$$\begin{aligned}
&\left|\tilde{x}_{N,2}(t) - \tilde{x}_2(t)\right| \\
&\quad\le \left|\tilde{x}_{N,2}(0) - \tilde{x}_2(0)\right| e^{-t\eta_2} + \lambda_2 \int_0^t e^{-(t-s)\eta_2} \left|\tilde{x}_{N,2}(s) - \tilde{x}_2(s)\right| \left|\tilde{x}_{N,1}(s)\right| ds \\
&\qquad + \lambda_2 \int_0^t e^{-(t-s)\eta_2} \left|\tilde{x}_{N,1}(s) - \tilde{x}_1(s)\right| \tilde{x}_2(s)\, ds + \left|\int_0^t e^{-(t-s)\eta_2}\, dM_{N,2}(s)\right| \\
&\qquad + \frac{\lambda_2}{\lambda_1 a} \frac{\lambda_1 - \lambda_2}{\lambda_1 - 1} \int_0^t e^{-(t-s)\eta_2} \left|\tilde{x}_{N,2}(s) - \tilde{x}_2(s)\right| \bigl(\tilde{x}_{N,2}(s) + \tilde{x}_2(s)\bigr)\, ds.
\end{aligned}$$

Letting $g_N(t) = e^{t\eta_2} f_N(t)$, we now have

$$\begin{aligned}
&\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2} \left|\tilde{x}_{N,1}(t) - \tilde{x}_1(t)\right| e^{t\eta_2} \\
&\quad\le g_N(0) + \lambda_1 \int_0^t g_N(s) \left|\tilde{x}_{N,1}(s) + \tilde{x}_1(s)\right| ds \\
&\qquad + \frac{\lambda_1 - \lambda_2}{\lambda_1 a^2} \int_0^t g_N(s) \bigl(\tilde{x}_{N,2}(s) + \tilde{x}_2(s)\bigr)\, ds \\
&\qquad + \frac{(\lambda_1 - \lambda_2)\lambda_1}{a(\lambda_1 - 1)} \int_0^t g_N(s) \tilde{x}_{N,2}(s)\, ds + \frac{\lambda_1}{a} \int_0^t g_N(s) \left|\tilde{x}_1(s)\right| ds \\
&\qquad + \frac{\lambda_1 - 1}{\lambda_1 - \lambda_2} e^{t\eta_2} \left|\int_0^t e^{-\eta_1(t-s)}\, dM_{N,1}(s)\right|
\end{aligned}$$

and

$$\begin{aligned}
&\left|\tilde{x}_{N,2}(t) - \tilde{x}_2(t)\right| e^{t\eta_2} \\
&\quad\le g_N(0) + \lambda_2 \int_0^t g_N(s) \left|\tilde{x}_{N,1}(s)\right| ds + \lambda_2 \frac{\lambda_1 - \lambda_2}{\lambda_1 - 1} \int_0^t g_N(s) \tilde{x}_2(s)\, ds \\
&\qquad + \frac{\lambda_2}{\lambda_1 a} \frac{\lambda_1 - \lambda_2}{\lambda_1 - 1} \int_0^t g_N(s) \bigl(\tilde{x}_{N,2}(s) + \tilde{x}_2(s)\bigr)\, ds \\
&\qquad + e^{t\eta_2} \left|\int_0^t e^{-(t-s)\eta_2}\, dM_{N,2}(s)\right|.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\big|\tilde{x}_{N,1}(t) - \tilde{x}_1(t)\big|e^{t\eta_2} \\
&\quad\le g_N(0) + 2\lambda_1 \int_0^t g_N(s)\big|\tilde{x}_1(s)\big|\,ds + \frac{\lambda_1(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds \\
&\qquad + \frac{2(\lambda_1 - \lambda_2)}{\lambda_1 a^2}\int_0^t g_N(s)\tilde{x}_2(s)\,ds + \frac{\lambda_1 - \lambda_2}{\lambda_1 a^2}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds \\
&\qquad + \frac{(\lambda_1 - \lambda_2)\lambda_1}{a(\lambda_1 - 1)}\int_0^t g_N(s)\tilde{x}_2(s)\,ds + \frac{(\lambda_1 - \lambda_2)\lambda_1}{a(\lambda_1 - 1)}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds \\
&\qquad + \frac{\lambda_1}{a}\int_0^t g_N(s)\big|\tilde{x}_1(s)\big|\,ds + \frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}e^{t\eta_2}\left|\int_0^t e^{-\eta_1(t-s)}\,dM_{N,1}(s)\right|
\end{aligned}$$

and

$$\begin{aligned}
\big|\tilde{x}_{N,2}(t) - \tilde{x}_2(t)\big|e^{t\eta_2} &\le g_N(0) + \lambda_2\int_0^t g_N(s)\big|\tilde{x}_1(s)\big|\,ds + \lambda_2\frac{\lambda_1 - \lambda_2}{\lambda_1 - 1}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds \\
&\quad + \lambda_2\frac{\lambda_1 - \lambda_2}{\lambda_1 - 1}\int_0^t g_N(s)\tilde{x}_2(s)\,ds + \frac{2\lambda_2(\lambda_1 - \lambda_2)}{a\lambda_1(\lambda_1 - 1)}\int_0^t g_N(s)\tilde{x}_2(s)\,ds \\
&\quad + \frac{\lambda_2}{\lambda_1 a}\frac{\lambda_1 - \lambda_2}{\lambda_1 - 1}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds + e^{t\eta_2}\left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right|.
\end{aligned}$$

So, since $a \to 1$ as $N \to \infty$, for N large enough,

$$\begin{aligned}
&\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\big|\tilde{x}_{N,1}(t) - \tilde{x}_1(t)\big|e^{t\eta_2} \\
&\quad\le g_N(0) + 4\lambda_1\int_0^t g_N(s)\big|\tilde{x}_1(s)\big|\,ds + \frac{2\lambda_1(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)\tilde{x}_2(s)\,ds \\
&\qquad + \frac{4\lambda_1(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds + \frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}e^{t\eta_2}\left|\int_0^t e^{-\eta_1(t-s)}\,dM_{N,1}(s)\right|
\end{aligned}$$

and

$$\begin{aligned}
&\tilde{x}_{N,2}(t) - \tilde{x}_2(t)|e^{t\eta_2} \\
&\quad\le g_N(0) + \lambda_2\int_0^t g_N(s)\big|\tilde{x}_1(s)\big|\,ds + \frac{4\lambda_2(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)\tilde{x}_2(s)\,ds \\
&\qquad + \frac{4\lambda_2(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds + e^{t\eta_2}\left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right|.
\end{aligned}$$

Hence, if N is large enough,

$$\begin{aligned}
g_N(t) \le g_N(0) &+ 4\lambda_1\int_0^t g_N(s)\big|\tilde{x}_1(s)\big|\,ds + \frac{4\lambda_1(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)\tilde{x}_2(s)\,ds \\
&+ \frac{4\lambda_1(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)^2 e^{-s\eta_2}\,ds + e^{t\eta_2}M,
\end{aligned}$$

where

$$M = \max\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\left|\int_0^t e^{-\eta_1(t-s)}\,dM_{N,1}(s)\right|, \left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right|\right).$$

Let $T_1 = \inf\{t \geq 0 : x_{N,1}(t) + x_{N,2}(t) > 4(\lambda_1 - 1)/\lambda_1\}$. Analogously to the proof of Lemma 6, on the event $t < T_1$, for N large enough,

$$\int_0^t \sum_y q(\tilde{x}_N(s), \tilde{x}_N(s) + y)(e^{\tilde{A}(t-s)}y)_1^2\, ds \\ \leq \frac{8(\lambda_1+1)}{N} \int_0^t \frac{\lambda_1 - 1}{\lambda_1} e^{-2(t-s)\eta_1}\, ds \leq \frac{4(\lambda_1+1)}{\lambda_1 N} \leq \frac{8}{N}$$

and

$$\int_0^t \sum_y q(\tilde{x}_N(s), \tilde{x}_N(s) + y)(e^{\tilde{A}(t-s)}y)_2^2\, ds \\ \leq \frac{4(\lambda_1+1)(\lambda_1-1)}{N(\lambda_1-\lambda_2)} \int_0^t \frac{\lambda_1 - \lambda_2}{\lambda_1} e^{-2(t-s)\eta_2}\, ds \leq \frac{4\lambda_1(\lambda_1-1)}{N(\lambda_1-\lambda_2)}.$$

Let T_2 be the infimum of times t such that

$$\left|\int_0^t e^{-(\lambda_1-1)(t-s)}\, dM_{N,1}(s)\right| > 8\sqrt{\frac{\omega}{N}}$$

or

$$\left|\int_0^t e^{-(t-s)\frac{\lambda_1-\lambda_2}{\lambda_1}}\, dM_{N,2}(s)\right| > 6\sqrt{\frac{\omega\lambda_1(\lambda_1-1)}{N(\lambda_1-\lambda_2)}}.$$

Then on the event $t < T_1 \wedge T_2$, for N large enough,

$$g_N(t) \leq g_N(0) + 4\lambda_1 \int_0^t g_N(s)|\tilde{x}_1(s)|\, ds + \frac{4\lambda_1(\lambda_1-\lambda_2)}{\lambda_1-1} \int_0^t g_N(s)\tilde{x}_2(s)\, ds \\ + \frac{4\lambda_1(\lambda_1-\lambda_2)}{\lambda_1-1} \int_0^t g_N(s)^2 e^{-s\eta_2}\, ds + 8e^{t\eta_2} \frac{\lambda_1-1}{\lambda_1-\lambda_2}\sqrt{\frac{\omega}{N}}.$$

Let T_3 be the infimum of times t such that

$$g_N(t) > 10e^{32(\lambda_1\beta/\alpha+1)} e^{t\eta_2} \frac{\lambda_1-1}{\lambda_1-\lambda_2}\sqrt{\frac{\omega}{N}}.$$

On the event $t < T_1 \wedge T_2 \wedge T_3$, for N large enough,

$$\frac{4\lambda_1(\lambda_1-\lambda_2)}{\lambda_1-1} \int_0^t g_N(s)^2 e^{-s\eta_2}\, ds \leq 400\lambda_1^2 e^{64(\lambda_1\beta/\alpha+1)} \frac{(\lambda_1-1)}{(\lambda_1-\lambda_2)^2} \frac{\omega}{N} e^{t\eta_2} \\ \leq e^{t\eta_2} \frac{\lambda_1-1}{\lambda_1-\lambda_2}\sqrt{\frac{\omega}{N}},$$

since we have assumed that $N(\lambda_1-\lambda_2)^2/\omega \to \infty$. It follows that, for N large enough, on the event $t < T_1 \wedge T_2 \wedge T_3$,

$$g_N(t) \leq g_N(0) + 4\lambda_1 \int_0^t g_N(s)|\tilde{x}_1(s)|\, ds + \frac{4\lambda_1(\lambda_1-\lambda_2)}{\lambda_1-1} \int_0^t g_N(s)\tilde{x}_2(s)\, ds \\ + 9e^{t\eta_2} \frac{\lambda_1-1}{\lambda_1-\lambda_2}\sqrt{\frac{\omega}{N}} \\ \leq 9e^{t\eta_2} \frac{\lambda_1-1}{\lambda_1-\lambda_2}\sqrt{\frac{\omega}{N}} + 4\lambda_1 \int_0^t g_N(s)|\tilde{x}_1(s)|\, ds + \frac{4\lambda_1(\lambda_1-\lambda_2)}{\lambda_1-1} \int_0^t g_N(s)\tilde{x}_2(s)\, ds,$$

since, by assumption, $g_N(0)=0$. By Grönwall's inequality, on the event $t<T_1\wedge T_2\wedge T_3$, for N large enough,

$$g_N(t)\le 9e^{t\eta_2}\frac{\lambda_1-1}{\lambda_1-\lambda_2}\sqrt{\frac{\omega}{N}}e^{4\lambda_1\int_0^t|\tilde{x}_1(s)|\,ds+\frac{4\lambda_1(\lambda_1-\lambda_2)}{\lambda_1-1}\int_0^t\tilde{x}_2(s)\,ds}.$$

Now, by (5),

$$\begin{aligned}\tilde{x}_2(t)&=x_2(t)=\big(x_1(t)/x_1(0)\big)^{\lambda_2/\lambda_1}x_2(0)e^{-t\frac{\lambda_1-\lambda_2}{\lambda_1}}\\ &\le\big(x_2(0)/x_1(0)\big)^{\lambda_2/\lambda_1}x_1(t)^{\lambda_2/\lambda_1}e^{-t\frac{\lambda_1-\lambda_2}{\lambda_1}}\qquad(22)\\ &=\left(\frac{\beta}{\alpha}\big(1+o(1)\big)\right)^{\lambda_2/\lambda_1}x_1(t)^{\lambda_2/\lambda_1}e^{-t\frac{\lambda_1-\lambda_2}{\lambda_1}}\le 4(\beta/\alpha)\frac{\lambda_1-1}{\lambda_1}e^{-t\frac{\lambda_1-\lambda_2}{\lambda_1}},\end{aligned}$$

where we have also used the facts that $x_1(0)+x_2(0)\le 2(\lambda_1-1)/\lambda_1$ implies $x_1(t)+x_2(t)\le 2(\lambda_1-1)/\lambda_1$ for all t, and that $(\lambda_1-\lambda_2)\log(\lambda_1-1)\to 0$ (and so $x_1(t)^{\lambda_2/\lambda_1}\le 3(\lambda_1-1)/\lambda_1$ for N large enough). Thus

$$\frac{4\lambda_1(\lambda_1-\lambda_2)}{\lambda_1-1}\int_0^t\tilde{x}_2(s)\,ds\le 16\lambda_1(\beta/\alpha).$$

Also, while $\tilde{x}_1(t)\ge 0$, using (21), we have for N large enough,

$$\frac{d\tilde{x}_1(t)}{dt}\le-\tilde{x}_1(t)\left((\lambda_1-1)-\frac{(\lambda_1-\lambda_2)\lambda_1}{(\lambda_1-1)a}\tilde{x}_2(t)\right)\le-\frac{\lambda_1-1}{2}\tilde{x}_1(t),$$

since $\tilde{x}_2(t)=x_2(t)\le 2(\lambda_1-1)/\lambda_1$ for all t. On the other hand, if and when $\tilde{x}_1(t)$ becomes negative, and while $|\tilde{x}_1(t)|\le(\lambda_1-1)/4\lambda_1$, then for N large enough, using $x_2(t)\le 4(\beta/\alpha)\frac{\lambda_1-1}{\lambda_1}e^{-t(\lambda_1-\lambda_2)/\lambda_1}$,

$$\frac{d\tilde{x}_1(t)}{dt}\ge-\frac{\lambda_1-1}{2}\tilde{x}_1(t)-\frac{(\lambda_1-\lambda_2)^2}{\lambda_1(\lambda_1-1)}\left(\frac{\tilde{x}_2(t)}{a}\right)^2,$$

so, since $-\tilde{x}_1(0)=o(\lambda_1-1)$ (as $x_1(0)+x_2(0)\ge(\lambda_1-1)/\lambda_1-h_N$),

$$\tilde{x}_1(t)\ge-\frac{32\beta^2}{\alpha^2a^2\lambda_1^2}(\lambda_1-\lambda_2)^2e^{-2t(\lambda_1-\lambda_2)/\lambda_1}.$$

Since $\tilde{x}_1(0)\le 4(\lambda_1-1)/\lambda_1$ for N large enough, we see that, if N is large enough, then $4\lambda_1\int_0^t|\tilde{x}_1(s)|\,ds\le 32$.

It follows that if N is large enough, then on the event $t<T_1\wedge T_2\wedge T_3$,

$$g_N(t)\le 9e^{t(\lambda_1-\lambda_2)/\lambda_1}\frac{\lambda_1-1}{\lambda_2-\lambda_1}\sqrt{\frac{\omega}{N}}e^{32(\lambda_1\beta/\alpha+1)},$$

and so

$$f_N(t)\le 9\frac{\lambda_1-1}{\lambda_1-\lambda_2}\sqrt{\frac{\omega}{N}}e^{32(\lambda_1\beta/\alpha+1)}.$$

Let $T_4=\inf\{t:f_N(t)>9\frac{\lambda_1-1}{\lambda_2-\lambda_1}\sqrt{\frac{\omega}{N}}e^{32(\lambda_1\beta/\alpha+1)}\}$. Let $t_0=t_0(N)=(\lambda_1-1)^{-1}e^{\omega/8}$. By the above, and since $\mathbb{P}(T_3\le T_4)=0$ for N large enough,

$$\mathbb{P}(T_4\le t_0)\le\mathbb{P}(T_1\wedge T_2\wedge T_3\le T_4\wedge t_0)\le\mathbb{P}(T_1\le t_0)+\mathbb{P}(T_2\le T_1\wedge t_0).$$

By Lemma 10, and by Lemma 11 with $\lambda=\lambda_1$ and $\mu=1$, $\mathbb{P}(T_1\le t_0)\le 4e^{-\omega/8}$ if N is large enough. By Lemma 7 applied to $(\tilde{x}_N(t))$, this time taking $\sigma=(\lambda_1-1)^{-1}$, $K_1=8/N$, $K_2=$

$4\lambda_1(\lambda_1-1)/N(\lambda_1-\lambda_2)$, $\mathbb{P}(T_2 \le T_1 \wedge t_0) \le 8e^{-\omega/8}$. It follows that $\mathbb{P}(T_4 \le t_0) \le 12e^{-\omega/8}$, as required. $\square$

The next lemma improves the upper bound on the approximation error $|\tilde{x}_{N,2}(t) - \tilde{x}_2(t)|$ by a factor $\sqrt{(\lambda_1-\lambda_2)(\lambda_1-1)^{-1}}$, which allows us to get closer to criticality. (There is a second term in our bound below, but it is always of order smaller than $x_2(t)$.)

LEMMA 15. *Let $\omega_1 = \omega_1(N) \to \infty$ as $N \to \infty$. Let the assumptions of Lemma 14 on $X_N(0)$ and $x(0)$ be satisfied. Assume further that*

$$\frac{N(\lambda_1-\lambda_2)^2}{\log(\frac{\lambda_1-1}{\lambda_1-\lambda_2})e^{\omega_1/2}} \to \infty.$$

For $t \ge 0$, let $f_N(t) = |\tilde{x}_{N,2}(t) - \tilde{x}_2(t)|$. Set

$$\delta_N(t) = \left(2(\lambda_1-\lambda_2)^{-1}\log^{1/2}\left(\frac{\lambda_1-1}{\lambda_1-\lambda_2}\right)e^{\omega_1/8}x_2(t) + 8\sqrt{\frac{\lambda_1(\lambda_1-1)}{\lambda_1-\lambda_2}}\right)\sqrt{\frac{\omega_1}{N}}e^{16(1+\beta/\alpha)}.$$

Then, for N large enough,

$$\mathbb{P}\Big(\sup_{t \le \frac{\lambda_1}{\lambda_1-\lambda_2}e^{\omega_1/8}} \big(f_N(t)/\delta_N(t)\big) > 1\Big) \le 16e^{-\omega_1/8}.$$

PROOF. Let $\omega_2 = \log(\frac{\lambda_1(\lambda_1-1)}{\lambda_1-\lambda_2}) + \omega_1$, and note that $N(\lambda_1-\lambda_2)^2/\omega_2 \to \infty$. As in the proof of Lemma 14, with $\eta_2 = (\lambda_1-\lambda_2)/\lambda_1$,

$$\begin{aligned}
f_N(t) &= \big|\tilde{x}_{N,2}(t) - \tilde{x}_2(t)\big| \\
&\le \big|\tilde{x}_{N,2}(0) - \tilde{x}_2(0)\big|e^{-t\eta_2} + \lambda_2\int_0^t e^{-(t-s)\eta_2}\big|\tilde{x}_{N,2}(s) - \tilde{x}_2(s)\big|\big|\tilde{x}_{N,1}(s)\big|\,ds \\
&\quad + \lambda_2\int_0^t e^{-(t-s)\eta_2}\big|\tilde{x}_{N,1}(s) - \tilde{x}_1(s)\big|\tilde{x}_2(s)\,ds + \left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right| \\
&\quad + \frac{\lambda_2}{\lambda_1 a}\frac{\lambda_1-\lambda_2}{\lambda_1-1}\int_0^t e^{-(t-s)\eta_2}\big|\tilde{x}_{N,2}(s) - \tilde{x}_2(s)\big|\big(\tilde{x}_{N,2}(s) + \tilde{x}_2(s)\big)\,ds.
\end{aligned}$$

Let $g_N(t) = f_N(t)e^{t\eta_2}$. Let T_1 be the infimum of times t such that

$$\big|\tilde{x}_{N,1}(t) - \tilde{x}_1(t)\big| > 9\sqrt{\frac{\omega_2}{N}}e^{32(\lambda_1\beta/\alpha+1)}.$$

Let T_2 be the infimum of times t such that

$$\left|\int_0^t e^{-(t-s)\eta_2}\,dM_{N,2}(s)\right| > 6\sqrt{\frac{\omega_1\lambda_1(\lambda_1-1)}{N(\lambda_1-\lambda_2)}}.$$

Then for $t < T_1 \wedge T_2$, if N is large enough,

$$\begin{aligned}
g_N(t) &\le g_N(0) + \lambda_2\int_0^t g_N(s)\big|\tilde{x}_{N,1}(s)\big|\,ds \\
&\quad + 9\lambda_2\sqrt{\frac{\omega_2}{N}}e^{32(\lambda_1\beta/\alpha+1)}\int_0^t e^{s\eta_2}\tilde{x}_2(s)\,ds + 6e^{t\eta_2}\sqrt{\frac{\omega_1\lambda_1(\lambda_1-1)}{N(\lambda_1-\lambda_2)}} \\
&\quad + \frac{4(\lambda_1-\lambda_2)}{\lambda_1-1}\int_0^t g_N(s)\tilde{x}_2(s)\,ds + \frac{2(\lambda_1-\lambda_2)}{\lambda_1-1}\int_0^t g_N(s)^2e^{-s\eta_2}\,ds.
\end{aligned}$$

Now, by (22),

$$x_2(t) \le 4(\beta/\alpha)\frac{\lambda_1 - 1}{\lambda_1} e^{-t\eta_2},$$

and, by (5) and (22), the fact that, for all t, $x_1(t) + x_2(t) \ge \frac{1}{2}\frac{\lambda_1 - 1}{\lambda_1}$ if N is large enough, and the fact that $x_1(t)/x_2(t)$ is increasing, also

$$x_2(t) \ge \frac{1}{4}\frac{\beta}{\alpha+\beta}\frac{\lambda_1 - 1}{\lambda_1} e^{-t\eta_2}. \tag{23}$$

Let T_3 be the infimum of times t such that

$$\begin{aligned} g_N(t) > {} & 10\sqrt{\frac{\lambda_1(\lambda_1 - 1)}{\lambda_1 - \lambda_2}}\sqrt{\frac{\omega_1}{N}} e^{16(1+\beta/\alpha)} e^{t\eta_2} \\ & + 4\sqrt{\frac{\omega_1}{N(\lambda_1 - \lambda_2)^2}} \log^{1/2}\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) e^{16(1+\beta/\alpha)} e^{\omega_1/8} x_2(t) e^{t\eta_2}. \end{aligned}$$

Then for $t < T_3$, if N is large enough,

$$\begin{aligned} & \frac{2(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)^2 e^{-s\eta_2}\, ds \\ & \le \frac{200\omega_1\lambda_1^2}{(\lambda_1 - \lambda_2)N} e^{32(1+\beta/\alpha)} e^{t\eta_2} + \frac{1024(\beta/\alpha)^2(\lambda_1 - 1)\omega_1}{N(\lambda_1 - \lambda_2)^2} e^{32(1+\beta/\alpha)} \log\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) e^{\omega_1/4} \\ & \le e^{t\eta_2}\sqrt{\frac{\omega_1\lambda_1(\lambda_1 - 1)}{N(\lambda_1 - \lambda_2)}} + \sqrt{\frac{\omega_1}{N(\lambda_1 - \lambda_2)^2}} \log^{1/2}\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) e^{\omega_1/8} x_2(t) e^{t\eta_2}, \end{aligned}$$

since our assumptions imply that

$$N(\lambda_1 - 1)(\lambda_1 - \lambda_2)/\omega_1 \to \infty,$$
$$N(\lambda_1 - \lambda_2)^2 / \log\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) e^{\omega_1/4} \to \infty.$$

So, for $t < T_1 \wedge T_2 \wedge T_3$, if N is large enough and since $g_N(0) = 0$,

$$\begin{aligned} g_N(t) \le {} & \lambda_2 \int_0^t g_N(s)\left|\tilde{x}_1(s)\right| ds + \frac{4(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t g_N(s)\tilde{x}_2(s)\, ds \\ & + 36\sqrt{\frac{\omega_2}{N}} e^{32(\lambda_1\beta/\alpha+1)} \frac{\beta}{\alpha}(\lambda_1 - 1)t + 8e^{t\eta_2}\sqrt{\frac{\omega_1\lambda_1(\lambda_1 - 1)}{N(\lambda_1 - \lambda_2)}} \\ & + \sqrt{\frac{\omega_1}{N(\lambda_1 - \lambda_2)^2}} \log^{1/2}\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) e^{\omega_1/8} x_2(t) e^{t\eta_2}. \end{aligned}$$

By Grönwall's lemma, on the event $t < T_1 \wedge T_2 \wedge T_3$,

$$\begin{aligned} g_N(t) \le {} & \left(36\sqrt{\frac{\omega_2}{N}} e^{32(\lambda_1\beta/\alpha+1)} \frac{\beta}{\alpha}(\lambda_1 - 1)t + 8e^{t\eta_2}\sqrt{\frac{\omega_1\lambda_1(\lambda_1 - 1)}{N(\lambda_1 - \lambda_2)}} \right. \\ & \left. + \sqrt{\frac{\omega_1}{N(\lambda_1 - \lambda_2)^2}} \log^{1/2}\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) e^{\omega_1/8} x_2(t) e^{t\eta_2}\right) e^{H_N(t)}, \end{aligned}$$

where

$$H_N(t) = \lambda_2 \int_0^t \left|\tilde{x}_1(s)\right| ds + \frac{4(\lambda_1 - \lambda_2)}{\lambda_1 - 1}\int_0^t \tilde{x}_2(s)\, ds \le 16\left(1 + \frac{\beta}{\alpha}\right)$$

for N large enough, since, as in the proof of the previous lemma, $\lambda_2 \int_0^t |\tilde{x}_1(s)|\,ds \le 8$. It follows that for large N, on the event $t < T_1 \wedge T_2 \wedge T_3$,

$$f_N(t) \le \biggl(36\sqrt{\frac{\omega_2}{N}} e^{32(\lambda_1\beta/\alpha+1)} \frac{\beta}{\alpha}(\lambda_1 - 1) t e^{-t\eta_2} + 8\sqrt{\frac{\omega_1\lambda_1(\lambda_1-1)}{N(\lambda_1-\lambda_2)}} + \sqrt{\frac{\omega_1}{N(\lambda_1-\lambda_2)^2}} \log^{1/2}\Bigl(\frac{\lambda_1-1}{\lambda_1-\lambda_2}\Bigr) e^{\omega_1/8} x_2(t)\biggr) e^{16(1+\frac{\beta}{\alpha})}.$$

Hence, from (22) and (23), for N large enough, on the event $t < T_1 \wedge T_2 \wedge T_3$,

$$f_N(t) \le \biggl(144\frac{\alpha+\beta}{\alpha}\sqrt{\frac{\omega_2}{N}} e^{32(\lambda_1\beta/\alpha+1)} \lambda_1 x_2(t) t + 8\sqrt{\frac{\omega_1\lambda_1(\lambda_1-1)}{N(\lambda_1-\lambda_2)}} + \sqrt{\frac{\omega_1}{N(\lambda_1-\lambda_2)^2}} \log^{1/2}\Bigl(\frac{\lambda_1-1}{\lambda_1-\lambda_2}\Bigr) e^{\omega_1/8} x_2(t)\biggr) e^{16(1+\frac{\beta}{\alpha})}.$$

Hence, if also $t \le t_0(N) := \frac{\lambda_1}{\lambda_1-\lambda_2} e^{\omega_1/8}$, then $f_N(t) \le \delta_N(t)$, where

$$\delta_N(t) = \biggl(\frac{2}{\lambda_1-\lambda_2} \log^{1/2}\Bigl(\frac{\lambda_1-1}{\lambda_1-\lambda_2}\Bigr) e^{\omega_1/8} x_2(t) + 8\sqrt{\frac{\lambda_1(\lambda_1-1)}{\lambda_1-\lambda_2}}\biggr)\sqrt{\frac{\omega_1}{N}} e^{16(1+\frac{\beta}{\alpha})}.$$

Letting $T_4 = \inf\{t : f_N(t) > \delta_N(t)\}$, $\mathbb{P}(T_4 \le t_0)$ is at most $\mathbb{P}(T_1 \wedge T_2 \wedge T_3 \le T_4 \wedge t_0)$, which is bounded above by $\mathbb{P}(T_1 \le t_0) + \mathbb{P}(T_2 \le T_1 \wedge t_0) + \mathbb{P}(T_3 \le T_4)$. By Lemma 14, $\mathbb{P}(T_1 \le t_0) \le 12e^{-\omega_2/8}$. By Lemma 7 applied to $(\tilde{x}_N(t))$, this time taking $\sigma = (\lambda_1 - \lambda_2)^{-1}$, $K_2 = 4\lambda_1(\lambda_1-1)/N(\lambda_1-\lambda_2)$, and $\eta = \eta_2$, $\mathbb{P}(T_2 \le T_1 \wedge t_0) \le 4e^{-\omega_1/8}$. Also, clearly, for N large enough, $\mathbb{P}(T_3 \le T_4) = 0$. It follows that $\mathbb{P}(T_4 \le t_0) \le 16e^{-\omega_1/8}$, as required. $\square$

We are now ready to prove the formulae for the distribution of the extinction times of the two species.

PROOF OF THEOREM 3. Let $t_1 = t_1(N) = (\lambda_1 - 1)^{-1/2}(\lambda_1 - \lambda_2)^{-1/2}$. By Lemma 13 with $\psi = (\lambda_1 - 1)^{1/2}(\lambda_1 - \lambda_2)^{-1/2}$, for N large enough,

$$\mathbb{P}\biggl(\biggl|\frac{X_{N,1}(t_1)}{X_{N,2}(t_1)} - \frac{X_{N,1}(0)}{X_{N,2}(0)}\biggr| > 2\Bigl(\frac{\lambda_1-\lambda_2}{\lambda_1-1}\Bigr)^{1/8}\biggr) \le 4e^{-\sqrt{N}(\lambda_2-1)} + e^{-(\frac{\lambda_1-1}{\lambda_1-\lambda_2})^{1/8}}.$$

Let $x = \min\{x_{N,1}(0) + x_{N,2}(0), N^{-1}\lfloor N(\lambda_2-1)(N(\lambda_2-1)^2)^{1/8}\rfloor\}$, and let T be the infimum of times t such that $x_{N,1}(t) + x_{N,2}(t) = x$. Note that $T = 0$ if $x_{N,1}(0) + x_{N,2}(0) \le N^{-1}\lfloor N(\lambda_2-1)(N(\lambda_2-1)^2)^{1/8}\rfloor = N^{-1}\lfloor N^{9/8}(\lambda_2-1)^{5/4}\rfloor$, which is an initial state up to "just above" the approximate carrying capacity $\lambda_2 - 1$ of the single stochastic SIS logistic epidemic process approximating $x_{N,1}(t) + x_{N,2}(t)$.

Lemma 4.1 in Brightwell, House and Luczak [6] is still valid for a supercritical stochastic SIS logistic epidemic $(Y_N(t))$, stating that $N^{-1}\mathbb{E}Y_N(t) \le y(t)$, where $y(t)$ solves (18), provided that $N^{-1}Y_N(0) = y(0)$. So, by that lemma, combined with Lemma 10 in the present paper and Markov's inequality, for N large enough,

$$\mathbb{P}\bigl(x_{N,1}(t_1/4) + x_{N,2}(t_1/4) > x\bigr) \le 2\bigl(N(\lambda_2-1)^2\bigr)^{-1/8},$$

and so $T \le t_1/4$ with probability at least $1 - 2(N(\lambda_2-1)^2)^{-1/8}$.

Let $y(t)$ solve equation (18) with $\lambda = \lambda_1$ and $\mu = 1$, $y(T) = x$, and let $z(t)$ solve equation (18) with $\lambda = \lambda_2$ and $\mu = 1$, and $z(T) = x$. Let $Y_N(t)$ and $Z_N(t)$ be the corresponding stochastic SIS logistic epidemics satisfying $Y_N(T) = Nx$ and $Z_N(T) = Nx$ respectively. It is easily seen from (20) that, if N is sufficiently large and $T \le t_1/4$, then $y(t_1/2) \le 2(\lambda_1-1)/\lambda_1$

and $z(t_1/2) \le 2(\lambda_2 - 1)/\lambda_2$. Furthermore, using (20) over the time-interval $[t_1/2, t_1]$, we see that in that case

$$\left|y(t_1) - (\lambda_1 - 1)/\lambda_1\right| \le \frac{\lambda_1 - 1}{\lambda_1} e^{-\frac{1}{2}\sqrt{\frac{\lambda_1-1}{\lambda_1-\lambda_2}}}$$

and

$$\left|z(t_1) - (\lambda_2 - 1)/\lambda_2\right| \le \frac{\lambda_2 - 1}{\lambda_2} e^{-(\lambda_2-1)t_1/2} \le \frac{\lambda_2 - 1}{\lambda_2} e^{-\frac{1}{4}\sqrt{\frac{\lambda_1-1}{\lambda_1-\lambda_2}}}.$$

We will now apply Lemma 12 twice, both starting at time T, once with $\lambda = \lambda_1$, $\mu = 1$, ending at time $\tau_1 = \inf\{t \ge T : y(t) \le 2(\lambda_1 - 1)/\lambda_1\}$, and the second time with $\lambda = \lambda_2$, $\mu = 1$, ending at time $\tau_2 = \inf\{t \ge T : z(t) \le 2(\lambda_2 - 1)/\lambda_2\}$. We further apply Lemma 11 twice, once to $Y_N(t)$, starting at time τ_1, with $\lambda = \lambda_1$, $\mu = 1$ and $\omega = \sqrt{N(\lambda_1 - 1)^2}$, and once to $Z_N(t)$, starting at time τ_2, with $\lambda = \lambda_2$, $\mu = 1$ and $\omega = \sqrt{N(\lambda_2 - 1)^2}$. Additionally applying Lemma 10, we see that, for N sufficiently large,

$$\begin{aligned}
&\mathbb{P}\left(\left|x_{N,1}(t_1) + x_{N,2}(t_1) - \frac{\lambda_1 - 1}{\lambda_1}\right| > 6e^4 \frac{(N(\lambda_1 - 1)^2)^{1/4}\sqrt{\lambda_1 + 1}}{\sqrt{N}} + \lambda_1 - \lambda_2\right) \\
&\quad \le 12e^{-(N(\lambda_2-1)^2)^{1/8}/8} + 2\big(N(\lambda_2 - 1)^2\big)^{-1/8} \le 3\big(N(\lambda_2 - 1)^2\big)^{-1/8}.
\end{aligned}$$

In particular, we see that, with probability $1 - \delta_N$, event $\mathcal{E}_{N,1}$ holds that $(\lambda_1 - 1)/\lambda_2 - h_N \le x_{N,1}(t_1) + x_{N,2}(t_1) \le 2(\lambda_1 - 1)/\lambda_1$, for some $h_N = o(\lambda_1 - 1)$, and $x_{N,1}(t_1)/x_{N,2}(t_1) = \alpha/\beta + \varepsilon_N$, where $\delta_N, \varepsilon_N \to 0$ as $N \to \infty$.

Let $\omega_1 = \omega_1(N) \to \infty$ be such that $\omega_1 \le \log(\frac{\lambda_1-1}{\lambda_1-\lambda_2})$ and let $\omega_2 = 16\log(\frac{\lambda_1-1}{\lambda_1-\lambda_2})$. Let $t_2 = t_2(N) = t_1(N) + \frac{\lambda_1}{\lambda_1-\lambda_2}e^{\omega_1/8}$, and note that $t_2 - t_1 \le (\lambda_1 - 1)^{-1}e^{\omega_2/8}$.

Consider the solution $x(t) = (x_1(t), x_2(t))^T$ to (1) subject to the condition $x_1(t_1) = N^{-1}X_{N,1}(t_1)$, $x_2(t_1) = N^{-1}X_{N,2}(t_1)$. Let also $\tilde{x}(t) = (\tilde{x}_1(t), \tilde{x}_2(t))^T$ be the corresponding solution to (9). For N large enough, by (22),

$$x_2(t_2) \le 4(\beta/\alpha)\frac{\lambda_1 - 1}{\lambda_1} e^{-(t_2-t_1)(\lambda_1-\lambda_2)/\lambda_1},$$

and by (23)

$$x_2(t_2) \ge \frac{1}{4}\frac{\beta}{\alpha + \beta}\frac{\lambda_1 - 1}{\lambda_1} e^{-(t_2-t_1)(\lambda_1-\lambda_2)/\lambda_1}.$$

Note that ω_1 can be chosen in such a way that $(\lambda_1 - \lambda_2)^{-1}x_2(t_2) \to 0$: for instance, we choose ω_1 satisfying $e^{\omega_1/8} = \log(\frac{\lambda_1-1}{\lambda_1-\lambda_2}) + \phi$ for a suitable $\phi = \phi(N) \to \infty$ such that $\phi \le \log(\frac{\lambda_1-1}{\lambda_1-\lambda_2})$. Since then $x_2(t_2) \ge \frac{1}{4\lambda_1}\frac{\beta}{\alpha+\beta}(\lambda_1 - \lambda_2)^2(\lambda_1 - 1)^{-1}$, we further have $Nx_2(t_2)(\lambda_1 - \lambda_2) \to \infty$.

Note that, since $N(\lambda_1 - \lambda_2)^3(\lambda_1 - 1)^{-1} \to \infty$, conditions of Lemmas 14 and 15 are satisfied. By Lemma 14 with $\omega = \omega_2$ and by Lemma 15 with the value of ω_1 above, with $X_{N,1}(t_1), X_{N,2}(t_1)$ as initial values, with probability at least $1 - 16e^{-\omega_1/8} - 12e^{-\omega_2/8} \to 1$ as $N \to \infty$, the event $\mathcal{E}_{N,2}$ holds that

$$\left|\tilde{x}_{N,1}(t_2) - \tilde{x}_1(t_2)\right| \le 9\sqrt{\frac{\omega_2}{N}}e^{32(\lambda_1\beta/\alpha+1)}$$

and

$$\begin{aligned}
\left|\tilde{x}_{N,2}(t_2) - \tilde{x}_2(t_2)\right| &\le \left[2(\lambda_1 - \lambda_2)^{-1}\log^{1/2}\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right)e^{\omega_1/8}x_2(t_2) + 8\sqrt{\frac{\lambda_1(\lambda_1 - 1)}{\lambda_1 - \lambda_2}}\right] \\
&\quad \times \sqrt{\frac{\omega_1}{N}}e^{16(1+\beta/\alpha)}.
\end{aligned}$$

Hence, also using bounds on $|\tilde{x}_1(t)|$ in the proof of Lemma 14, there exists a constant c_1 such that, for N large enough, on $\mathcal{E}_{N,2}$,

$$\left|x_{N,1}(t_2) - \frac{\lambda_1 - 1}{\lambda_1}\right| \le \left[2(\lambda_1 - \lambda_2)^{-1} \log^{1/2}\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) e^{\omega_1/8} x_2(t_2) + 10\sqrt{\frac{\lambda_1(\lambda_1 - 1)}{\lambda_1 - \lambda_2}}\right] \times \sqrt{\frac{\omega_1}{N}} e^{32(1+\beta/\alpha)} + c_1(\lambda_1 - 1)e^{-(t_2-t_1)(\lambda_1-\lambda_2)/\lambda_1}.$$

Note that

$$N(\lambda_1 - \lambda_2)^3(\lambda_1 - 1)^{-1}/\log\log\bigl(N(\lambda_1 - \lambda_2)^2\bigr) \to \infty$$

implies that

$$N(\lambda_1 - \lambda_2)^3(\lambda_1 - 1)^{-1}/\log\log\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right) \to \infty,$$

and so we can choose ϕ satisfying $e^{\omega_1/8} = \log(\frac{\lambda_1-1}{\lambda_1-\lambda_2}) + \phi$ so that

$$N(\lambda_1 - \lambda_2)^3(\lambda_1 - 1)^{-1}/\omega_1 e^{2\phi} \to \infty.$$

With such a choice of ϕ, it follows that

$$(\lambda_1 - 1)e^{-(t_2-t_1)\frac{\lambda_1-\lambda_2}{\lambda_1}} \gg \sqrt{\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}}\sqrt{\frac{\omega_1}{N}},$$

and so, on the event $\mathcal{E}_{N,2}$, both $x_{N,1}(t_2)$ and $x_{N,2}(t_2)$ are concentrated around $(\lambda_1 - 1)/\lambda_1$ and $x_2(t_2)$ respectively, $x_{N,1}(t_2)$ with error of size $o(\lambda_1 - \lambda_2)$ and $x_{N,2}(t_2)$ with error of size $o(x_2(t_2)) = o(\lambda_1 - \lambda_2)$.

Let

$$t_3 = t_2 + \frac{10\lambda_1}{\lambda_1 - \lambda_2} \log\bigl(N(\lambda_1 - \lambda_2)^2\bigr) \le t_1 + (\lambda_1 - 1)^{-1} e^{\omega_3/8},$$

where $\omega_3 = 32\log((\lambda_1 - 1)/(\lambda_1 - \lambda_2)) + 32\log\log(N(\lambda_1 - \lambda_2)^2)$. Note that ω_3 satisfies the condition of Lemma 14, so we can apply this lemma on the interval $[t_1, t_3]$.

For $t \ge t_2$, let $\mathcal{E}_N(t)$ be the event that, for all $s \in [t_2, t]$,

$$\left|x_{N,1}(s) - \frac{\lambda_1 - 1}{\lambda_1}\right| \le \left[4(\lambda_1 - \lambda_2)^{-1} \log\left(\frac{\lambda_1 - 1}{\lambda_1 - \lambda_2}\right)^{1/2} e^{\omega_1/8} x_2(t_2) + 20\sqrt{\frac{\lambda_1(\lambda_1 - 1)}{\lambda_1 - \lambda_2}}\right] \times \sqrt{\frac{\omega_4}{N}} e^{32(1+\beta/\alpha)} + (c_1 + 16\beta/\alpha)(\lambda_1 - 1)e^{-(t_2-t_1)(\lambda_1-\lambda_2)/\lambda_1},$$

and $x_{N,2}(s) \le 2x_{N,2}(t_2)$. Note that on the event $\mathcal{E}_N(t)$, for all $t_2 \le s \le t$, $x_{N,1}(s)$ is concentrated around $(\lambda_1 - 1)/\lambda_1$ with an $o(\lambda_1 - \lambda_2)$ error.

On the event $\mathcal{E}_{N,2}$, and while $\mathcal{E}_N(t)$ holds, using a standard argument similar to the proof of Lemma 8 and the proof of Lemma 2.1 in Brightwell, House, and Luczak [6], we couple the subsequent evolution of $X_{N,2}(t)$ with two linear birth-and-death chains, each with birth rate $\frac{\lambda_2}{\lambda_1} + o(\lambda_1 - \lambda_2)$, and death rate 1, so as to sandwich it between two such chains. The next event after time $t \ge t_2$ in each of the three chains can be coupled together, as long as event $\mathcal{E}_N(t)$ holds. Since $N(\lambda_1 - \lambda_2)x_2(t_2) \to \infty$, we have $X_{N,2}(t_2)(\lambda_1 - \lambda_2) \to \infty$ and so Lemma 9 can be applied to show that the randomness in the duration of the final phase after t_2 is approximately Gumbel distributed.

If event $\mathcal{E}_N(\kappa_N)$ holds, then the length of the final phase is

$$\begin{aligned}&\frac{\lambda_1}{\lambda_1-\lambda_2}\left(\log N+\log x_2(t_2)+\log\frac{\lambda_1-\lambda_2}{\lambda_1}+o(1)+G_N\right)\\&\quad=\frac{\lambda_1}{\lambda_1-\lambda_2}\big(\log N(\lambda_1-\lambda_2)^2-\phi(N)+O(1)\big),\end{aligned}$$

with high probability, where G_N converges to a Gumbel random variable G as $N\to\infty$. Event $\mathcal{E}_N(\kappa_N)$ holds with high probability by Lemma 14 and by Lemma 2.2 in Brightwell, House and Luczak [6], as extinction occurs before t_3 with high probability. The details are as in the proof of Lemma 8.

The length of the "fluid-limit" phase can be expressed as

$$\frac{\lambda_2}{\lambda_1-\lambda_2}\log\big(x_1(t_2)/x_1(t_1)\big)-\frac{\lambda_1}{\lambda_1-\lambda_2}\log\big(x_2(t_2)/x_2(t_1)\big),$$

and the length of the first phase is $t_1=(\lambda_1-\lambda_2)^{-1}\sqrt{(\lambda_1-\lambda_2)/(\lambda_1-1)}=o((\lambda_1-\lambda_2)^{-1})$.

Hence, using the fact that $(\lambda_1-\lambda_2)\log(\lambda_1-1)\to 0$, the total time to extinction is, with high probability,

$$\frac{\lambda_1}{\lambda_1-\lambda_2}\left(\log\left(\frac{N(\lambda_1-1)(\lambda_1-\lambda_2)\beta}{\lambda_1^2\alpha}+o(1)+G_N\right)\right),$$

thus proving the first part of Theorem 3.

The second part follows from Theorem 1 and its extension to the barely supercritical regime due to Foxall [9]—see comments following the statement of that theorem. Also, our assumptions imply that κ_N, which is of order $(\lambda_1-\lambda_2)^{-1}\log(N(\lambda_1-1)(\lambda_1-\lambda_2))$, is negligible compared to $\exp(N(\lambda_1-1)^2)N^{-1/2}(\lambda_1-1)^{-2}$, and so also compared to $\tilde{\kappa}_N$; this follows from the fact that, as $N\to\infty$, $N^{-1}(\lambda_1-1)^{-2}\exp(N(\lambda_1-1)^2)\to\infty$, and hence is much larger than $N^{-1/2}(\lambda_1-\lambda_2)^{-1}$, which tends to 0. $\square$

7.1. *Relaxing the assumption on separation from criticality.* As stated above, we believe Theorem 3 is in fact valid under the weaker condition $N(\lambda_1-\lambda_2)^2\to\infty$ (still assuming $\mu_1=\mu_2=1$ and $(\lambda_1-\lambda_2)(\lambda_1-1)^{-1}\to\infty$). Here is a sketch of how one might go about proving such an extension. The differential equation approximation phase can be split into a number of subphases, each corresponding to a refined version of Lemma 15 with a smaller value of $x_{N,2}(0)$ and thus a smaller bound on the deviation of the martingale transform term. Roughly speaking the first subphase would have $x_{N,2}(0)$ of order λ_1-1 and the martingale transform term of order $N^{-1/2}(\lambda_1-\lambda_2)^{-1/2}(\lambda_1-1)^{1/2}$. The first subphase would last until $x_{N,2}(t)$ is of size about $N^{-1/2}(\lambda_1-\lambda_2)^{-1/2}(\lambda_1-1)^{1/2}\omega$, for a suitable $\omega(N)\to\infty$, and would thus take time just slightly less than $(\lambda_1-\lambda_2)^{-1}\log\sqrt{N(\lambda_1-1)(\lambda_1-\lambda_2)}$. The second subphase would have $x_{N,2}(0)$ of order $N^{-1/2}(\lambda_1-\lambda_2)^{-1/2}(\lambda_1-1)^{1/2}\omega$ and the martingale transform term of order $N^{-3/4}(\lambda_1-\lambda_2)^{-3/4}(\lambda_1-1)^{1/4}\omega^{1/2}$. It would last until $x_{N,2}(t)$ is of size about $N^{-3/4}(\lambda_1-\lambda_2)^{-3/4}(\lambda_1-1)^{1/4}\omega$, and would thus take time just slightly less than $(\lambda_1-\lambda_2)^{-1}\log(N(\lambda_1-1)(\lambda_1-\lambda_2))^{1/4}$. In the third subphase, $x_{N,2}(0)$ would be of order $N^{-3/4}(\lambda_1-\lambda_2)^{-3/4}(\lambda_1-1)^{1/4}\omega$ and the martingale transform term would be of order $N^{-7/8}(\lambda_1-\lambda_2)^{-7/8}(\lambda_1-1)^{1/8}\omega^{1/2}$. It would last until $x_{N,2}(t)$ is of size about $N^{-7/8}(\lambda_1-\lambda_2)^{-7/8}(\lambda_1-1)^{1/8}\omega$, and would thus take time just slightly less than $(\lambda_1-\lambda_2)^{-1}\log(N(\lambda_1-1)(\lambda_1-\lambda_2))^{1/8}$. And, in principle, one should be able to carry on this process. The phases would be joined together using the end value of $x_{N,2}(t)$ from the previous phase as initial condition for the differential equation in the next phase, and the various deterministic solutions with different random initial conditions can be shown to get closer and closer together over time, using techniques similar to those already used in this paper.

As many phases would be used as needed to "reach" $x_{N,2}(t)$ of order $o(\lambda_1 - \lambda_2)$, while keeping the deviation smaller than the mean. Since the "limiting" error size in the above process is $N^{-1}(\lambda_1 - \lambda_2)^{-1}$, this should in principle be possible as long as $N(\lambda_1 - \lambda_2)^2 \to \infty$, and as $N(\lambda_1 - \lambda_2)^2$ tends to infinity more and more slowly, the time spent in the differential equation phase becomes closer and closer to $(\lambda_1 - \lambda_2)^{-1} \log(N(\lambda_1 - 1)(\lambda_1 - \lambda_2))$.

After the condition $N(\lambda_1 - \lambda_2)^2 \to \infty$ fails, one can still carry out the differential equation phase from the time when $x_{N,2}$ is of order $\lambda_1 - 1$ through the various phases until it is of order $N^{-1}(\lambda_1 - \lambda_2)^{-1}$, which takes time of order about $(\lambda_1 - \lambda_2)^{-1} \log(N(\lambda_1 - 1)(\lambda_1 - \lambda_2))$. After that, the fluctuations dominate, and the remaining time is about $(\lambda_1 - \lambda_2)^{-2}$ steps, translating to a time of order $N^{-1}(\lambda_1 - \lambda_2)^{-2} = o((\lambda_1 - \lambda_2)^{-1})$, and the randomness is not Gumbel distributed.

A differential equation approximation phase is possible as long as the order of the initial martingale transform deviation, $N^{-1/2}(\lambda_1 - \lambda_2)^{-1/2}(\lambda_1 - 1)^{1/2}$, is $o(\lambda_1 - 1)$, which is as long as $N(\lambda_1 - 1)(\lambda_1 - \lambda_2) \to \infty$. Our formulae above would suggest a critical window with $|N(\lambda_1 - 1)(\lambda_1 - \lambda_2)| = O(1)$, or $|N(\lambda_1 - \lambda_2)| = O(1)$ if $\lambda_1 - 1$ is positive and bounded away from 0. This would join up our result nicely with that of Kogan et al. [12], showing that when the two basic reproductive ratios are equal (but not close to 1), then the time to extinction is of order N. We expect all the details above, while somewhat tedious, can be filled in to yield a complete proof of the behaviour away from criticality. However, we leave this till a subsequent paper, along with analysis of the behaviour inside the critical window.

Acknowledgements. The authors are grateful to Graham Brightwell for helpful comments that inspired Lemma 15, and as a result helped us achieve an improved understanding of near-critical phenomena in this model.

FL was first a postdoctoral research assistant of ML funded by EPSRC Leadership Fellowship EP/J004022/2, and was later supported partly by Conicyt/Fondecyt Proyecto Postdoctorado N.3160163.

ML was supported partly by EPSRC Leadership Fellowship EP/J004022/2 and partly by ARC Future Fellowship FT170100409.

REFERENCES

[1] ANDERSSON, H. and DJEHICHE, B. (1998). A threshold limit theorem for the stochastic logistic epidemic. *J. Appl. Probab.* **35** 662–670. MR1659540 https://doi.org/10.1239/jap/1032265214

[2] ARROWSMITH, D. K. and PLACE, C. M. (1992). *Dynamical Systems: Differential Equations, Maps and Chaotic Behaviour. Chapman and Hall Mathematics Series*. CRC Press, London. MR1195127 https://doi.org/10.1007/978-94-011-2388-4

[3] BAHL, J., VIJAYKRISHNA, D., HOLMES, E. C., SMITH, G. J. D. and GUAN, Y. (2009). Gene flow and competitive exclusion of avian influenza A virus in natural reservoir hosts. *Virology* **390** 289–297.

[4] BARBOUR, A. D., HAMZA, K., KASPI, H. and KLEBANER, F. C. (2015). Escape from the boundary in Markov population processes. *Adv. in Appl. Probab.* **47** 1190–1211. MR3433302 https://doi.org/10.1239/aap/1449859806

[5] BARBOUR, A. D. and LUCZAK, M. J. (2012). A law of large numbers approximation for Markov population processes with countably many types. *Probab. Theory Related Fields* **153** 727–757. MR2948691 https://doi.org/10.1007/s00440-011-0359-2

[6] BRIGHTWELL, G., HOUSE, T. and LUCZAK, M. (2018). Extinction times in the subcritical stochastic SIS logistic epidemic. *J. Math. Biol.* **77** 455–493. MR3830272 https://doi.org/10.1007/s00285-018-1210-5

[7] BRIGHTWELL, G. and LUCZAK, M. (2012). Vertices of high degree in the preferential attachment tree. *Electron. J. Probab.* **17** Art. ID 14. MR2892327 https://doi.org/10.1214/EJP.v17-1803

[8] DARLING, R. W. R. and NORRIS, J. R. (2008). Differential equation approximations for Markov chains. *Probab. Surv.* **5** 37–79. MR2395153 https://doi.org/10.1214/07-PS121

[9] FOXALL, E. (2018). Extinction time of the logistic process. Preprint. Available at arXiv:1805.08339.

[10] HUMPLIK, J., HILL, A. L. and NOWAK, M. A. (2014). Evolutionary dynamics of infectious diseases in finite populations. *J. Theoret. Biol.* **360** 149–162. MR3255502 https://doi.org/10.1016/j.jtbi.2014.06.039

[11] KIRUPAHARAN, N. and ALLEN, L. J. S. (2004). Coexistence of multiple pathogen strains in stochastic epidemic models with density-dependent mortality. *Bull. Math. Biol.* **66** 841–864. MR2255779 https://doi.org/10.1016/j.bulm.2003.11.007

[12] KOGAN, O., KHASIN, M., MEERSON, B., SCHNEIDER, D. and MYERS, C. R. (2014). Two-strain competition in quasineutral stochastic disease dynamics. *Phys. Rev. E* **90** Art. ID 042149.

[13] KURTZ, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Probab.* **7** 49–58. MR0254917 https://doi.org/10.2307/3212147

[14] LEVIN, S. A. (1970). Community equilibria and stability, and an extension of the competitive exclusion principle. *Amer. Nat.* **104** 413–423.

[15] LOTKA, A. J. (1925). *Elements of Physical Biology*. Williams and Wilkins, Baltimore, MD.

[16] MORAN, P. A. P. (1958). Random processes in genetics. *Proc. Camb. Philos. Soc.* **54** 60–71. MR0127989 https://doi.org/10.1017/s0305004100033193

[17] NÅSELL, I. (2011). *Extinction and Quasi-Stationarity in the Stochastic Logistic SIS Model*. *Lecture Notes in Math.* **2022**. Springer, Heidelberg. MR2808078 https://doi.org/10.1007/978-3-642-20530-9

[18] NEUHAUSER, C. (1992). Ergodic theorems for the multitype contact process. *Probab. Theory Related Fields* **91** 467–506. MR1151806 https://doi.org/10.1007/BF01192067

[19] PARSONS, T. L. and QUINCE, C. (2007). Fixation and haploid populations exhibiting density dependence I: The non-neutral case. *Theor. Popul. Biol.* **72** 121–135.

[20] PARSONS, T. L. and QUINCE, C. (2007). Fixation and haploid populations exhibiting density dependence II: The quasi-neutral case. *Theor. Popul. Biol.* **72** 468–479.

[21] PARSONS, T. L., QUINCE, C. and PLOTKIN, J. B. (2008). Absorption and fixation times for neutral and quasi-neutral populations with density dependence. *Theor. Popul. Biol.* **74** 302–310.

[22] PAZY, A. (1983). *Semigroups of Linear Operators and Applications to Partial Differential Equations*. *Applied Mathematical Sciences* **44**. Springer, New York. MR0710486 https://doi.org/10.1007/978-1-4612-5561-1

[23] RENSHAW, E. (2011). *Stochastic Population Processes*: *Analysis*, *Approximations*, *Simulations*. Oxford Univ. Press, Oxford. MR2865609 https://doi.org/10.1093/acprof:oso/9780199575312.001.0001

[24] THORISSON, H. (2000). *Coupling*, *Stationarity*, *and Regeneration*. *Probability and Its Applications* (*New York*). Springer, New York. MR1741181 https://doi.org/10.1007/978-1-4612-1236-2

[25] VOLTERRA, V. (1931). *Leçons sur la Théorie Mathématique de la Lutte Pour la Vie*. Gautier-Villars, Paris. MR1189803

[26] WALTER, W. (1998). *Ordinary Differential Equations*. *Graduate Texts in Mathematics* **182**. Springer, New York. MR1629775 https://doi.org/10.1007/978-1-4612-0601-9

[27] ZEEMAN, M. L. (1995). Extinction in competitive Lotka–Volterra systems. *Proc. Amer. Math. Soc.* **123** 87–96. MR1264833 https://doi.org/10.2307/2160613