Adv. Appl. Prob. **47,** 1190–1211 (2015)
Printed in Northern Ireland

ESCAPE FROM THE BOUNDARY IN MARKOV POPULATION PROCESSES

A. D. BARBOUR,* *Universität Zürich*
K. HAMZA,** *Monash University*
HAYA KASPI,*** *Technion, Haifa*
F. C. KLEBANER,** *Monash University*

Abstract

Density dependent Markov population processes in large populations of size N were shown by Kurtz (1970), (1971) to be well approximated over finite time intervals by the solution of the differential equations that describe their average drift, and to exhibit stochastic fluctuations about this deterministic solution on the scale $\sqrt{N}$ that can be approximated by a diffusion process. Here, motivated by an example from evolutionary biology, we are concerned with describing how such a process leaves an absorbing boundary. Initially, one or more of the populations is of size much smaller than N, and the length of time taken until all populations have sizes comparable to N then becomes infinite as $N \to \infty$. Under suitable assumptions, we show that in the early stages of development, up to the time when all populations have sizes at least $N^{1-\alpha}$ for $\frac{1}{3} < \alpha < 1$, the process can be accurately approximated in total variation by a Markov branching process. Thereafter, it is well approximated by the deterministic solution starting from the original initial point, but with a random time delay. Analogous behaviour is also established for a Markov process approaching an equilibrium on a boundary, where one or more of the populations become extinct.

Keywords: Markov population process; boundary behaviour; branching process

2010 Mathematics Subject Classification: Primary 92D30
Secondary 60J27; 60B12

1. Introduction

A continuous-time version of the two morphs stage in the bare bones evolution model of Klebaner *et al.* (2011, Section 3) can be represented as a pure jump Markov process X_N on $\mathbb{Z}_+^2$, with the first component the count of wild-type individuals, initially around their carrying capacity, and the second the count of mutant individuals. The transition rates are

$$X \longrightarrow X + (1, 0) \quad \text{at rate } a_1 X_1,$$
$$X \longrightarrow X + (-1, 0) \quad \text{at rate } X_1\left\{\left(\frac{X_1}{N}\right) + \gamma\left(\frac{X_2}{N}\right)\right\},$$

Received 28 August 2014; revision received 22 September 2014.
* Postal address: Institut für Mathematik, Universität Zürich, Winterthurertrasse 190, CH-8057 Zürich, Switzerland. Email address: a.d.barbour@math.uzh.ch
** Postal address: School of Mathematical Sciences, Monash University, Clayton, VIC 3800, Australia.
*** Postal address: Faculty of Industrial Engineering and Management, Technion, Haifa, 32000, Israel.

$$X \longrightarrow X + (0, 1) \quad \text{at rate } a_2 X_2,$$
$$X \longrightarrow X + (0, -1) \quad \text{at rate } X_2 \left\{ \gamma \left(\frac{X_1}{N} \right) + \left(\frac{X_2}{N} \right) \right\}.$$

Initially, X_1 has a value near its carrying capacity Na_1, and $X_2 = 0$. At some time, which we call 0, Z_0 mutant individuals are introduced into the population; Z_0 is thought of as fixed, irrespective of the (large) value of N. The mutants and wild-type individuals differ only through their birth rates a_1 and a_2. Each species has *per capita* death rate given by the density of its own population, together with an additional component of γ multiplied by the density of individuals of the other species. If $\gamma > 1$, members of the other species cause a higher mortality rate than those of the same species; if $\gamma < 1$, they cause a lower mortality rate than those of the same species, favouring the possibility of coexistence. If $a_2 < \gamma a_1$, the mutants have negligible chance of survival, but if $a_2 > \gamma a_1$, there is a nonzero probability $p_N(Z_0) \approx 1 - (\gamma a_1/a_2)^{Z_0}$ that the mutant strain will become established. In this case, if also $a_1 > \gamma a_2$, the two populations will eventually come to coexist; if, instead, $a_1 < \gamma a_2$, the wild-type population will be driven to extinction. Note that, as expected, coexistence is impossible if $\gamma > 1$. In this paper we are primarily interested in describing how the process develops up to the time at which the mutants represent a positive fraction of the population when N is large. We also examine the detail of how the wild-type becomes extinct when $a_1 < \gamma a_2$.

This process is a particular example of a more general family of processes, that we now investigate. We suppose that X_N is a Markov population process on $\mathbb{Z}_+^d$ having transition rates

$$X \to X + J \quad \text{at rate } N g^J \left(\frac{X}{N} \right), \; X \in \mathbb{Z}_+^d, \; J \in \mathcal{J},$$

where $\mathcal{J}$ is a finite subset of $\mathbb{Z}^d$. We assume that the functions g^J are continuously differentiable for $x \in \mathbb{R}_+^d$ and we write

$$F(x) := \sum_{J \in \mathcal{J}} J g^J(x), \qquad x \in \mathbb{R}_+^d$$

to denote the infinitesimal drift of the process $x_N := N^{-1} X_N$. Letting $\{P^J : J \in \mathcal{J}\}$ be independent rate 1 Poisson processes, the evolution of x_N can be described (Kurtz (1978)) by

$$x_N(t) = x_N(0) + \frac{1}{N} \sum_{J \in \mathcal{J}} J P^J(N G_N^J(t)) = x_N(0) + \int_0^t F(x_N(u)) \, \mathrm{d}u + m_N(t), \tag{1.1}$$

where

$$m_N(t) := \sum_{J \in \mathcal{J}} J \left\{ \frac{P^J(N G_N^J(t)) - G_N^J(t)}{N} \right\}, \tag{1.2}$$

and $G_N^J(t) := \int_0^t g^J(x_N(u)) \, \mathrm{d}u$. The process m_N is a well-behaved vector-valued martingale. In differential form, (1.1) can be expressed as

$$\mathrm{d}x_N(t) = F(x_N(t))\mathrm{d}t + \mathrm{d}m_N(t),$$

and the corresponding 'deterministic equations', given by leaving out the martingale innovations, are

$$\frac{\mathrm{d}\xi}{\mathrm{d}t} = F(\xi). \tag{1.3}$$

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

Our interest here is in deriving an approximation to the process x_N in circumstances in which the initial state is close to $\bar{x}$, an unstable equilibrium point of (1.3), as in the bare bones example given above. In the seminal papers of Kendall (1956) and Whittle (1955), written in the context of Bartlett's (1949) Markovian SIR epidemic process, a basic description was proposed. Such processes should behave much like branching processes near $\bar{x}$, as far as those components in which numbers are small are concerned, and should then look more and more like solutions to the deterministic equations as the numbers grow. The deterministic part of the approximation was established for general Markov population processes in Kurtz (1970, Theorem (3.1)), who showed that, if $\lim_{N\to\infty} x_N(0) = x_0$ then $\sup_{0\le t\le T} |x_N(t) - \xi(t)| \to 0$ in distribution for any finite $T > 0$, where ξ satisfies (1.3) with $\xi(0) = x_0$. In particular, if $x_0 = \bar{x}$, Kurtz's (1970) theorem implies that $x_N(t)$ stays asymptotically close to $\bar{x}$ over any fixed finite time interval. However, the deterministic solution ξ_N starting with $x_N(0)$ close to $\bar{x}$ may still eventually escape from $\bar{x}$, but the time that it takes to do so is asymptotically infinite as $N \to \infty$, so that Kurtz's (1970) theorem is not suitable for describing what eventually happens. Such outcomes may nonetheless be of considerable practical importance in applications. The aim of this paper is to show that the Kendall–Whittle description can indeed be established in considerable generality, and to give some measure of the accuracy of the resulting approximation.

Under appropriate conditions, we prove that the process x_N, if it indeed escapes from x_0, then closely follows the path of the solution to the deterministic equations, but with a random time shift, and that the time required to escape from x_0 is of order $O(\log N)$. This behaviour is exactly what one might expect on the basis of the Kendall–Whittle description, with the random time shift reflecting the essential randomness that occurs in the early stages of the branching phase. However, proving that it is actually the case is not so easy. One main difficulty is presented by the asymptotically infinite length of time that elapses, while the process is escaping from the boundary, since this necessitates good control over the behaviour of the process over long time intervals. A related difficulty is to keep control of the branching approximation for a long enough time to ensure that the subsequent development is indeed almost deterministic. Our approach is to establish an extremely accurate approximation, in terms of the total variation distance between the probability distributions of the two processes, over a very long initial time interval. Once this has been achieved, the subsequent development can be described well enough by the deterministic solution. We then go on to prove complementary results, describing the behaviour of a process that approaches a stable equilibrium point of the deterministic equations at which some coordinates of the process take the value 0.

1.1. Assumptions

Our general setting is as follows. The specialization to the bare bones example is given in Section 1.4. Denote by $x^{(1)}$ the first d_1 components of x and by $x^{(2)}$ the remaining $d_2 = d - d_1$ components, and split $J = (j_1, \dots, j_d) = (J^{(1)}, J^{(2)})$ in the same way. For transitions with $J^{(2)} \neq 0$, suppose that the rates are always of the form $g^J(x) = \bar{g}^J(x) x_{s(J)}$ for some $s(J)$ such that $d_1 < s(J) \le d$, and that $\bar{g}^J(x_0) > 0$; we also assume that $J_i \ge 0$ for all $i \ne s(J)$ such that $d_1 < i \le d$, and that $J_{s(J)} \ge -1$. We denote the set of all such transitions by $\mathcal{J}_2$. These assumptions are natural in a population context; in particular, if the constraints on the elements of such J are violated, some of the components could become negative. The function F can now be written in the form

$$F(x) = \begin{pmatrix} A(x) \\ B(x) \end{pmatrix} x^{(2)} + \begin{pmatrix} c(x^{(1)}) \\ 0 \end{pmatrix},$$

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

where, for each x, $A(x)$ and $B(x)$ are $d_1 \times d_2$ and $d_2 \times d_2$ matrices, respectively, and $c(x^{(1)})$ is a d_1-vector. Suppose that $x_0^{(1)}$ is a strongly stable equilibrium of $\mathrm{d}\xi^{(1)}/\mathrm{d}t = c(\xi^{(1)})$ and that $x_0^{(2)} = 0$. Then the solution ξ of the deterministic equations starting at x_0 is the constant x_0, and the stochastic system x_N, if started near x_0 with $x_N^{(2)}(0) = 0$, typically spends an amount of time that is at least exponential in N before leaving the vicinity of x_0 (Barbour and Pollett (2012, Theorem 4.1)). However, if the initial value $x_N^{(2)}(0)$ is not 0, but takes the value $x_{N,0}^{(2)} = N^{-1}Z_0$ for some $0 \neq Z_0 \in \mathbb{Z}_+^{d_2}$, and if $B_0 := B(x_0)$ is such that ξ_N, the solution of (1.3) starting from this initial condition, leaves the neighbourhood of the boundary, then x_N has positive probability of doing so as well.

Henceforth, we shall suppose that $x_N(0) = x_{N,0}$ satisfies $|x_{N,0}^{(1)} - x_0^{(1)}| \leq N^{-5/12}$. Under the equilibrium distribution for $x_N^{(1)}$ when $x_N^{(2)} = 0$, typical values of $|x_{N,0}^{(1)} - x_0^{(1)}|$ are of order $O(N^{-1/2})$, so that such a starting condition is reasonable. Suppose also that $x_{N,0}^{(2)} = N^{-1}Z_0$. Our assumptions imply that B has nonnegative off-diagonal entries near x_0; we also assume that it is irreducible, and that the largest eigenvalue β_0 of B_0 is positive. In addition, the elements of the matrices A and B are assumed to be continuously differentiable functions of x. The stability of $x_0^{(1)}$ is expressed by assuming that the function c is of the form

$$c(w) = C(w - x_0^{(1)}) + \tilde{c}(w), \qquad w \in \mathbb{R}_+^{d_1},$$

where C is a fixed $d_1 \times d_1$ matrix such that, for some $\gamma_1 < \infty$,

$$|\mathrm{e}^{Ct}x| \leq \gamma_1|x|, \qquad x \in \mathbb{R}^{d_1},\ t \geq 0,$$

as is the case if all the eigenvalues of C have negative real part, and where for some $K_c, \rho_1 > 0$, and for $w_1, w_2 \in \mathbb{R}_+^{d_1}$ such that $\max_{i=1,2}|w_i - x_0^{(1)}| \leq \rho_1$,

$$|\tilde{c}(w_1) - \tilde{c}(w_2)| \leq K_c|w_1 - w_2|\Big\{|w_1 - w_2| + \min_{i=1,2}|w_i - x_0^{(1)}|\Big\}.$$

From the Perron–Frobenius theorem, there also exist $0 < \gamma_3 < \gamma_2 < \infty$ such that

$$|\mathrm{e}^{B_0t}x| \leq \gamma_2\mathrm{e}^{\beta_0t}|x|, \qquad x \in \mathbb{R}^{d_2},\ t \geq 0$$

and

$$|\mathrm{e}^{B_0t}x| \geq \gamma_3\mathrm{e}^{\beta_0t}|x|, \qquad x \in \mathbb{R}_+^{d_2},\ t \geq 0. \tag{1.4}$$

We also choose $0 < \rho_2 \leq \rho_1$ small enough so that

$$b_*^J := \inf_{|x-x_0|\leq\rho_2}|\bar{g}^J(x)| > 0 \quad \text{for all } J \in \mathcal{J}_2.$$

We denote by $\|G\|$ the matrix norm $\|G\| := \sup_{y:\,|y|=1}\{|Gy|\}$. For matrix functions $G(x)$, we write $\|G\|_\rho := \sup_{|x-x_0|\leq\rho}\|G(x)\|$ and

$$\|DG\|_\rho := \sup_{|x-x_0|\leq\rho,|x'-x_0|\leq\rho}\left\{\frac{\|B(x) - B(x')\|}{|x - x'|}\right\}.$$

In all the arguments that follow, constants involving the symbol k are defined solely in terms of the functions A, B, and c, and associated constants such as ρ_2, and do not vary, either with N, or with the choices made for the quantities $\varepsilon^{(i)}$, $1 \leq i \leq 4$, appearing in Lemmas 2.1 and 2.2. Constants involving the symbol δ are typically to be chosen suitably small, but again only with reference to the functions A, B, and c, and to associated constants such as ρ_2.

1.2. Main results

Under these assumptions, we carry out a programme indicated in Barbour (1980), but now in more general circumstances. We first show that the initial behaviour of $Nx_N^{(2)}$ is well approximated by that of a supercritical d_2-type Markov branching process Z, defined at the beginning of Section 3, whose mean growth rate matrix is $B_0^\top$. Let $\boldsymbol{u}^\top$ be the left eigenvector of $B_0^\top$ corresponding to β_0, normalized so that $\boldsymbol{u}^\top \mathbf{1} = 1$, and let the corresponding right eigenvector be $\boldsymbol{v}$, normalized so that $\boldsymbol{u}^\top \boldsymbol{v} = 1$. Then branching process theory (Athreya and Ney (1972, Chapter V.7, Theorem 2)) implies that $Z(t)\mathrm{e}^{-\beta_0 t} \to W\boldsymbol{u}$ almost surely (a.s.) as $t \to \infty$, where the random variable W has mean $Z_0^\top \boldsymbol{v}$ and satisfies $W > 0$ on the set of nonextinction, and in consequence, for as long as this approximation holds

$$x_N^{(2)}(t) \approx \frac{\mathrm{e}^{\beta_0\{t+\beta_0^{-1}\log W\}}\boldsymbol{u}}{N}; \tag{1.5}$$

the results that we use are proved in an online appendix; see Barbour *et al.* (2014). The development of $\xi_N^{(2)}$, the second group of components of the solution of the deterministic equation, also initially parallels that of $x_N^{(2)}$, in that the linear approximation to (1.3) near x_0 yields

$$\xi_N^{(2)}(t) \approx \frac{\mathrm{e}^{B_0 t} Z_0}{N} \sim \frac{\mathrm{e}^{\beta_0 t}(\boldsymbol{v}^\top Z_0)\,\boldsymbol{u}}{N} = \frac{\mathrm{e}^{\beta_0\{t+\beta_0^{-1}\log(\boldsymbol{v}^\top Z_0)\}}\boldsymbol{u}}{N}, \tag{1.6}$$

by virtue of the Perron–Frobenius theorem (Seneta (2006, Theorem 2.7)). The quantity W in (1.5) is replaced in (1.6) by its expectation, so that, apart from the random time shift $\beta_0^{-1}(\log W - \log \mathbb{E}W)$, the two paths are much the same. This simple description of the development of x_N turns out to be true also if *all* components, and not just those of the second group, are considered; the formal statement of this, together with some estimate of the accuracy of the approximation, is the main message of Theorem 1.1. Note that the approximations (1.5) and (1.6) need t to be large, so that in the first case the branching asymptotics and in the second the Perron–Frobenius asymptotics give good approximations. On the other hand, t should not be so large as to invalidate the linearizations around x_0, implicit in both approximations. It is the need to satisfy both requirements simultaneously, with sufficient accuracy, and for large enough values of t, that provides a major source of complication in the proofs.

In Section 3 we show that the branching approximation in fact holds *in total variation* up to a time $\tau_{N,\alpha}^x$, chosen so that $N\boldsymbol{v}^\top x_N^{(2)}(\tau_{N,\alpha}^x)$ is approximately $N^{1-\alpha}$ for any $\alpha > \frac{1}{3}$. As is shown by example in Section E of the online appendix (see Barbour *et al.* (2014)), approximation in total variation is typically not accurate for $\alpha \le \frac{1}{3}$, but it is essential to the subsequent argument that we can take $\alpha < \frac{1}{2}$; we take $\alpha = \frac{5}{12}$ for the remaining development. If the branching process is absorbed in 0, then so too, with high probability, is $x_N^{(2)}$. If not, then we show that $x_N(\tau_{N,\alpha}^x)$ is close to $\xi(t_{N,\alpha}^\xi)$, where $t_{N,\alpha}^\xi = \beta_0^{-1}(1-\alpha)\log N + O(1)$ is the approximate time t at which the deterministic solution ξ_N starting in $x_N(0)$ satisfies $\boldsymbol{v}^\top \xi_N(t) = N^{1-\alpha}$. The details are to be found in Proposition 3.1.

In Section 4 we show that the deterministic and stochastic paths $\tilde{\xi}_N$ and $\tilde{x}_N$, both starting at $x_N(\tau_{N,5/12}^x)$, and with time argument restarting at 0, stay asymptotically close for large N until an elapsed time $t_N(\delta)$, at which $\mathbf{1}^\top \tilde{\xi}_N$ first attains the value δ, for a small but fixed $\delta > 0$; note that $t_N(\delta) = \beta_0^{-1}\alpha \log N + O(1)$. The details are given in Proposition 4.1; the fact that $\alpha < \frac{1}{2}$ is needed to maintain the accuracy of approximation up to times at which the second components of the paths have attained asymptotically nonnegligible size. From this point onwards Kurtz's (1970) theorem, together with the Lipschitz continuity of the solutions

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

of the deterministic equations with respect to their initial conditions, can be used to justify the further deterministic approximation to x_N as long as the deterministic curve remains within some fixed, compact subset of $\mathbb{R}_+^d$. Thus, x_N closely follows the deterministic path, but at a random rate, with the randomness quickly settling down to a fixed time shift of order $O(1)$. The combined theorem is as follows; the parts not justified by theorems in Kurtz (1970), (1978) are proved in the following sections. For the statement of the theorem, we make the following general definitions:

$$\tau^Z(0) := \inf\{t > 0\colon Z(t) = 0\}, \qquad \tau_N^x(0) := \inf\{t > 0\colon x_N^{(2)}(t) = 0\}, \tag{1.7}$$

$$\tau_{N,\alpha}^Z := \inf\{t\colon \boldsymbol{v}^\top Z(t) \ge N^{1-\alpha} + \boldsymbol{v}^\top Z_0\}, \tag{1.8}$$

$$\tau_{N,\alpha}^x := \inf\{t\colon \boldsymbol{v}^\top N x_N^{(2)}(t) \ge N^{1-\alpha} + \boldsymbol{v}^\top Z_0\}, \tag{1.9}$$

$$t_{N,\alpha}^\xi := \beta_0^{-1}\{(1-\alpha)\log N - \log(\boldsymbol{v}^\top Z_0)\}$$

with the infimum of the empty set taken equal to ∞, and for the particular choice $\alpha = \frac{5}{12}$, we define

$$\tau_{N*}^Z := \tau_{N,5/12}^Z, \qquad \tau_{N*}^x := \tau_{N,5/12}^x, \qquad t_{N*}^\xi := t_{N,5/12}^\xi. \tag{1.10}$$

For the Markov branching process Z, defined at the beginning of Section 3, we set $W := \lim_{t\to\infty} \boldsymbol{v}^\top Z(t)\mathrm{e}^{-\beta_0 t}$.

Theorem 1.1. *With the assumptions and definitions of Section 1.1, suppose that* $x_N(0)$ *is such that* $|x_N^{(1)}(0) - x_0^{(1)}| \le N^{-5/12}$ *and that* $x_N^{(2)}(0) = N^{-1}Z_0$ *for fixed* $0 \ne Z_0 \in \mathbb{Z}_+^{d_2}$*. Then, except on an event* E_{N1}^c *of asymptotically negligible probability, the paths of* $Nx_N^{(2)}$ *and of* Z *can be coupled so as to be identical until the time* $\min\{\tau^Z(0), \tau_{N*}^Z\}$*, in which case* $\tau_{N*}^Z = \tau_{N*}^x = \beta_0^{-1}\{\frac{7}{12}\log N - \log W\} + O(N^{-7/48})$.

Let $\mathcal{K}$ *be any fixed compact subset of* $\mathbb{R}_+^d$*. Suppose that* T *is such that* $\xi_N(t_{N*}^\xi + t) \in \mathcal{K}$ *for all* $0 \le t \le T$*, where* ξ_N *denotes the solution to the deterministic equation starting with* $\xi_N(0) = x_N(0)$*. Then there exist constants* $\gamma > 0$*,* $k_T < \infty$*, and an event* E_{N2}^T *such that, on* $\{\tau_{N*}^x < \infty\} \cap E_{N1} \cap E_{N2}^T$,

$$\sup_{0\le t\le 5/12\beta_0^{-1}\log N + T} |x_N(\tau_{N*}^x + t) - \xi_N(t_{N*}^\xi + t)| \le k_T N^{-\gamma} \tag{1.11}$$

and $\lim_{N\to\infty} \mathbb{P}[E_{N2}^T \mid \{\tau_{N1}^Z < \infty\} \cap E_{N1}] = 1$.

The proof of the branching approximation is given in Section 3 and its content summarized in Proposition 3.1. The proof of the subsequent deterministic approximation, up to a time at which x_N is away from the boundary, is given in Section 4 and its content summarized in Proposition 4.1. The extension to further choices of T follows from Kurtz (1970, Theorem (3.1)) and approximation is then by a nondegenerate path. There is no universal choice possible for the exponent γ appearing in (1.11), which is a reflection of the greater delicacy required for the approximations derived here than in the setting of Kurtz (1970) when any $\gamma < \frac{1}{2}$ would satisfy; we give an example to illustrate this in Section E of the online appendix; see Barbour *et al.* (2014).

Theorem 1.1 can be interpreted in the sense that, to a first approximation, the random process x_N follows the deterministic curve starting at the same point, but with a random delay of $\tau_{N*}^x - t_{N*}^\xi \sim \beta_0^{-1}\{\log(\boldsymbol{v}^\top Z_0) - \log W\}$. The initial condition for Z_0 could be allowed to depend on N, in which case the distribution of W would depend on N, too: if $|Z_0^{(N)}| \to \infty$ then $\log(\boldsymbol{v}^\top Z_0^{(N)}) - \log W^{(N)} \to_d 0$, so that, to this level of approximation, the initial randomness would disappear.

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

1.3. Absorption

Our motivating example actually contains two periods in which the process is close to a boundary, the second being when the wild-type becomes extinct. The setting is then almost exactly as in Section 1.1, except for the fact that the deterministic solution converges to 0 in some of its coordinates, instead of moving away from 0. In the notation of Section 1.1, this corresponds to having the largest real part among the eigenvalues of B_0 being negative; we denote it by $-\beta_1$. In this setting, we also assume that the eigenvalues of C all have negative real parts.

Under these modified assumptions, we consider stochastic and deterministic processes $\tilde{\xi}_\delta$ and $x_{N,\delta}$ that are started close to one another, as is implied by the previous results, at a point where they are reasonably close to the stable equilibrium x_0. To be more precise, we first suppose that $|x_{N,\delta}(0) - x_0| \le \delta$, and that ξ_δ is the solution to the deterministic equations with $\xi_\delta(0) = x_{N,\delta}(0)$. We then show that for δ chosen small enough, the two processes remain close for a further time $t_N(\delta) := \beta_1^{-1}(\log \delta + \frac{5}{12} \log N)$, at which point the second group of coordinates, those that are converging to 0, are of magnitude approximately $N^{-5/12}$, and the first coordinates are at a similar distance from $x_0^{(1)}$. We also show that, if $|\xi_\delta(0) - \tilde{\xi}_\delta(0)| = O(N^{-\gamma})$ for some $\gamma > 0$, then, for δ chosen small enough, $|\xi_\delta^{(1)}(t_N(\delta)) - \tilde{\xi}_\delta^{(1)}(t_N(\delta))| = O(N^{-\gamma-\varepsilon})$ for some $\varepsilon > 0$, and $N^{5/12}|\xi_\delta^{(2)}(t_N(\delta)) - \tilde{\xi}_\delta^{(2)}(t_N(\delta))| = O(N^{-\gamma/2})$. After this time, the process $(Nx_N^{(2)}(t_N(\delta) + t),\ t \ge 0)$ is well approximated by a branching process Z in total variation, with rates as before. The following theorem summarizes these results; the proofs are given in Section A of the online appendix; see Barbour *et al.* (2014).

Theorem 1.2. *Suppose that the assumptions of Section 1.1 hold with the above modifications. Then there exist $\delta > 0$ and an event E_N, whose complement has asymptotically negligible probability, such that, on E_N, if $|x_{N,\delta}(0) - x_0| \le \delta$, and if $|x_{N,\delta}(0) - \tilde{\xi}_{N,\delta}(0)| = O(N^{-\gamma_1})$ for some $\gamma_1 > 0$, then*

$$\sup_{0\le t\le t_N(\delta)} |x_{N,\delta}^{(1)}(t) - \tilde{\xi}_{N,\delta}^{(1)}(t)| \le \tilde{k}^{(1)} N^{-\gamma}, \qquad \sup_{0\le t\le t_N(\delta)} \{e^{\beta_1 t}|x_{N,\delta}^{(2)}(t) - \tilde{\xi}_\delta^{(2)}(t)|\} \le \tilde{k}^{(2)} N^{-\gamma}$$

with $t_N(\delta) := \max\{\beta_1^{-1}(\log \delta + \frac{5}{12} \log N), 0\}$ and for suitable $\tilde{k}^{(1)}, \tilde{k}^{(2)}$, and $\gamma > 0$. After $t_N(\delta)$, the process $Nx_{N,\delta}^{(2)}(t_N(\delta) + \cdot)$ can be coupled to be identical until extinction to the (now subcritical) Markov branching process Z, except on an event of asymptotically negligible probability. In particular, for a suitable constant h^*, the time $t_N(\delta) + T_N$ at which $x_N^{(2)}$ is absorbed in 0 is such that $\mathcal{L}(\beta_1 T_N - \log N - \log(\boldsymbol{v}^\top \tilde{\xi}_{N,\delta}(t_N(\delta))) - \log(h^*))$ converges in total variation as $N \to \infty$ to a Gumbel distribution.

The approximation given by Theorem 1.2 shows that, to a first approximation, the random process x_N follows the deterministic curve starting at the same point until the time $t_N(\delta) = \beta_1^{-1}(\log \delta + \frac{5}{12} \log N)$. The law of large numbers for Z starting at $Nx_{N,\delta}^{(2)}(t_N(\delta))$ then shows that the same is true afterwards; however, for such times, $x_{N,\delta}^{(2)}$ is uniformly small and $x_N^{(1)}(t)$ is close to x_0, and so the conclusion is of little interest. By contrast, the branching approximation delivers more detailed information. In particular, the time taken by the deterministic solution $\tilde{\xi}_{N,\delta}$ from $t_N(\delta)$ until $\hat{t}_N := \inf\{t > 0\colon \boldsymbol{v}^\top \tilde{\xi}_{N,\delta}(t) = N^{-1}\}$ is such that

$$\hat{t}_N \sim {\beta_1}^{-1}\{\log N + \log(\boldsymbol{v}^\top \tilde{\xi}_{N,\delta}(t_N(\delta)))\}.$$

The deterministic solution itself never reaches $\tilde{\xi}_{N,\delta}^{(2)} = 0$ in finite time, but $\hat{t}_N$ is the sort of approximation that might be made for the time to extinction, based on deterministic considerations.

From Theorem 1.2 we see that this is reasonable, but that the duration in the stochastic model has an additional random component $\beta_1^{-1}\{G + \log(h^*)\}$.

1.4. The bare bones example

These results can all be applied to the bare bones example discussed earlier, which is of the form proposed in Section 1.1, with $d_1 = d_2 = 1$. In the initial stages, the matrices $A(x)$ and $B(x)$ are the scalars $-\gamma x_1$ and $(a_2 - \gamma x_1 - x_2)$, and the function $c(x_1) = -a_1(x_1 - a_1) - (x_1 - a_1)^2$, so that we have $C = -a_1 < 0$ and $\tilde{c}(x_1) = -(x_1 - a_1)^2$. Assuming that $a_2 > \gamma a_1$, the unstable equilibrium of the deterministic equations is $x_0 = (a_1, 0)^\top$, and $\beta_0 = B(x_0) = a_2 - \gamma a_1 > 0$. The set $\mathcal{J}_2$ consists of the transitions $\{(0, 1), (0, -1)\}$, and $s(J) = 2$ for both of them; the corresponding functions $\bar{g}^J$ are a_2 and $(\gamma x_1 + x_2)$, respectively. The process Z is a linear birth and death process with *per capita* birth and death rates a_2 and γa_1, respectively, and its behaviour is well understood. In particular, the limiting random variable W, conditional on the event of nonextinction, has a gamma distribution $\mathrm{Ga}(Z_0, 1)$. Hence, if $Z_0 = 1$, the delay in following the deterministic curve, given in general by

$$\tau^x_{N*} - t^\xi_{N*} \sim \beta_0{}^{-1}\{\log(\boldsymbol{v}^\top Z_0) - \log W\},$$

has the distribution of $\{a_2 - \gamma a_1\}^{-1} G_1$, where G_1 has a Gumbel distribution.

For the latter stages of the example, in the case when $a_1 < \gamma a_2$, the wild-type individuals eventually die out. To match the formulation in Section 1.3 it is necessary to swap the meaning of the coordinates so that the second coordinate now represents the remaining numbers of wild-type individuals. The matrices $A(x)$ and $B(x)$ become the scalars $-\gamma x_2$ and $(a_1 - \gamma x_2 - x_1)$, and the function $c(x_1)$ is given by $-a_2(x_1 - a_2) - (x_1 - a_2)^2$, so that we obtain $C = -a_2 = -\kappa$ and $\tilde{c}(x_1) = -(x_1 - a_2)^2$. The strongly stable equilibrium of the deterministic equations with the mutants established is given by $x_0 = (a_2, 0)^\top$, and $-\beta_1 = B(x_0) = a_1 - \gamma a_2 < 0$. The set $\mathcal{J}_2$ consists as before of the transitions $\{(0, 1), (0, -1)\}$, and $s(J) = 2$ for both of them; the corresponding functions $\bar{g}^J$ are a_1 and $(\gamma x_1 + x_2)$, respectively. The branching process Z is again a linear birth and death process, with *per capita* birth and death rates a_1 and γa_2, respectively. For this process, the constant h^* appearing in the final approximation can be evaluated using the definition in Heinzmann (2009, p. 299) as $1 - a_1/(\gamma a_2)$. Combining this with the above, we can deduce that the asymptotics of the entire time from the introduction of a single mutant until the extinction of the wild-type individuals is given by

$$\frac{G_1}{\{a_2 - \gamma a_1\}} + \frac{1}{\{\gamma a_2 - a_1\}}\left\{\log\left(\frac{1 - a_1}{(\gamma a_2)}\right) + G_2\right\} + T^{(N)},$$

where $T^{(N)} = (\{a_2 - \gamma a_1\}^{-1} + \{\gamma a_2 - a_1\}^{-1}) \log N + O(1)$ is the time taken for the deterministic curve to get from the initial state, where the proportion of mutants is N^{-1}, to the state in which the proportion of wild-type individuals is N^{-1}; and G_1 and G_2 are independent Gumbel random variables. The duration of the closed stochastic epidemic, studied in Barbour (1975), could also be approached in a similar way. In that example, however, the function c is identically 0, so that the final stages have to be treated differently.

2. The deterministic solutions

For use in our arguments, we collect some properties of the solutions to the deterministic equations in the neighbourhood of the initial point, deferring the proofs of the lemmas to the

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

online appendix; see Barbour *et al.* (2014). We first use variation of constants to rewrite the equations in the form

$$\begin{aligned}\xi^{(1)}(t) &= \xi^{(1)}(0) + \int_0^t \{A(\xi(u))\xi^{(2)}(u) + c(\xi^{(1)}(u))\}\,\mathrm{d}u \\ &= x_0^{(1)} + \mathrm{e}^{Ct}(\xi^{(1)}(0) - x_0^{(1)}) \\ &\quad + \int_0^t \mathrm{e}^{C(t-u)}\{A(\xi(u))\xi^{(2)}(u) + \tilde{c}(\xi^{(1)}(u))\}\,\mathrm{d}u, \qquad (2.1)\\ \xi^{(2)}(t) &= \xi^{(2)}(0) + \int_0^t B(\xi(u))\xi^{(2)}(u)\,\mathrm{d}u \\ &= \mathrm{e}^{B_0 t}\xi^{(2)}(0) + \int_0^t \mathrm{e}^{B_0(t-u)}\{B(\xi(u)) - B_0\}\xi^{(2)}(u)\,\mathrm{d}u. \qquad (2.2)\end{aligned}$$

We recall that, in the arguments that follow, constants involving the symbol k do not vary with the choices made for the quantities $\varepsilon^{(i)}$, $1 \le i \le 4$. In our applications, these quantities become small, as N increases, as negative powers of N, and the assumptions made about them in the lemmas are automatically satisfied for all sufficiently large N. For use in what follows, define

$$t_0(\delta,\varepsilon) := {\beta_0}^{-1}\log\left(\frac{\delta}{\varepsilon}\right), \qquad t_1(\delta,\varepsilon) := {\beta_1}^{-1}\log\left(\frac{\delta}{\varepsilon}\right) \quad \text{for } \delta \ge \varepsilon > 0, \qquad (2.3)$$

where β_0 is as in Section 1.1 and β_1 is as in Section 1.3.

Lemma 2.1. *Under the assumptions of Section 1.1, there exists a* δ_0 *with* $0 < \delta_0 \le 1$*, depending only on the functions* A, B, *and* c *and associated constants such as* ρ_2*, with the following properties. If* ξ *satisfies (2.1) and (2.2), with initial condition such that* $|\xi^{(1)}(0) - x_0^{(1)}| \le \varepsilon^{(1)}$ *and* $|\xi^{(2)}(0)| = \varepsilon^{(2)}$*, and if*

$$4\gamma_1\varepsilon^{(1)} \le \min\left\{1, \left(\frac{\rho_2}{4}\right)\right\}, \qquad \varepsilon^{(2)} \le \delta_0,$$

and if also

$$\varepsilon^{(1)}\log\left(\frac{1}{\varepsilon^{(2)}}\right) \le \min\left\{1, \frac{\beta_0}{(24\gamma_2\gamma_1\|DB\|_{\rho_2})}, \frac{\beta_0}{(32K_c\gamma_1^2)}\right\},$$

then, for all $0 \le t \le t_0(\delta_0, \varepsilon^{(2)})$,

$$\sup_{0\le u\le t}|\xi^{(1)}(u) - x_0^{(1)}| \le k^{(1)}\{\varepsilon^{(1)} + \varepsilon^{(2)}\mathrm{e}^{\beta_0 t}\}, \qquad \sup_{0\le u\le t}\mathrm{e}^{-\beta_0 u}|\xi^{(2)}(u)| \le k^{(2)}\varepsilon^{(2)},$$

$$\sup_{0\le u\le t}\mathrm{e}^{-\beta_0 u}|\xi^{(2)}(u) - \mathrm{e}^{B_0 u}\xi^{(2)}(0)| \le k^{(3)}\varepsilon^{(2)}\left\{\varepsilon^{(1)}\log\left(\frac{1}{\varepsilon^{(2)}}\right) + \varepsilon^{(2)}\mathrm{e}^{\beta_0 t}\right\}$$

for suitable $k^{(1)}, k^{(2)}$*, and* $k^{(3)}$*. Furthermore, if* $\tilde{\xi}$ *satisfies (2.1) and (2.2) with initial condition* $\tilde{\xi}(0)$ *satisfying* $|\tilde{\xi}^{(2)}(0) - \xi^{(2)}(0)| \le \varepsilon^{(3)} \le k^{(4)}(\varepsilon^{(2)})^{1+\gamma}$ *and* $|\tilde{\xi}^{(1)}(0) - \xi^{(1)}(0)| \le k^{(5)}\varepsilon^{(2)}\log(1/\varepsilon^{(2)})$ *for some* $k^{(4)}, k^{(5)} > 0$*, and* $0 < \gamma < 1$*, then there exist* $k^{(6)}, k^{(7)}, k^{(8)}$*, and* $0 < \delta_1 \le \delta_0$ *such that, for all* $\varepsilon^{(2)} \le \min\{k^{(8)}, \delta_1\}$,

$$\sup_{0\le u\le t_0(\delta_1,\varepsilon^{(2)})}|\xi^{(1)}(u) - \tilde{\xi}^{(1)}(u)| \le k^{(6)}(\varepsilon^{(2)})^{\gamma/2},$$

$$\sup_{0\le u\le t_0(\delta_1,\varepsilon^{(2)})}\{\mathrm{e}^{-\beta_0 u}|\xi^{(2)}(u) - \tilde{\xi}^{(2)}(u)|\} \le k^{(7)}(\varepsilon^{(2)})^{1+\gamma/2}.$$

Here, δ_1 *may depend on the choice of* γ*, as well as on the functions* A, B*, and* c.

We also consider the final stages of such a process, before absorption in a strongly stable equilibrium with the 2-components equal to 0. Under such circumstances, we can still work under assumptions similar to those made in Section 1.1. The main difference is to require that the eigenvalue of B_0 with largest real part is negative; we denote it by $-\beta_1$. We also assume that the equilibrium x_0 is strongly attractive, in the sense that

$$|e^{Ct}x| \le \gamma_1 e^{-\kappa t}|x|, \qquad x \in \mathbb{R}^{d_1},\ t \ge 0$$

for some $\kappa > 0$ and $\gamma_1 < \infty$; the previous assumptions of Section 1.1 only required $\kappa \ge 0$. The analogue of Lemma 2.1 is then as follows.

Lemma 2.2. *With the assumptions of Section 1.1, modified as in Section 1.3, let ξ_δ satisfy (2.1) and (2.2) with $\xi_\delta(0) =: x_{\delta 0}$ such that $|x_{\delta 0} - x_0| \le \delta$. Then, for any $0 < \kappa' < \min\{\kappa, \beta_1\}$, there exists a $\delta_0 > 0$ and constants $\hat{k}^{(i)}$ such that, for all $0 < \delta \le \delta_0$,*

$$\sup_{u\ge 0} e^{\kappa' u}|\xi_\delta^{(1)}(u) - x_0^{(1)}| \le \hat{k}^{(1)}\delta, \qquad \sup_{u\ge 0} e^{\beta_1 u}|\xi_\delta^{(2)}(u)| \le \hat{k}^{(2)}\delta,$$
$$\sup_{u\ge 0} e^{\beta_1 u}|\xi_\delta^{(2)}(u) - e^{B_0 u}x_\delta^{(2)}(0)| \le \hat{k}^{(3)}\delta^2.$$

Furthermore, for any $\theta > 0$, there exists a $\delta(\theta) > 0$ such that, for any $0 < \delta \le \delta(\theta)$, if $\tilde{\xi}_\delta$ satisfies (2.1) and (2.2) with $\tilde{\xi}_\delta(0)$ satisfying $|\tilde{\xi}_\delta(0) - x_{\delta 0}| \le \varepsilon^{(4)}$, and if $0 < \eta < \delta$ and $\varepsilon^{(4)}\eta^{-\theta} \le K$ for K defined implicitly in Equation (D.33) of the online appendix (see Barbour et al. (2014)), then

$$\sup_{0\le u\le t_1(\delta,\eta)} |\xi_\delta^{(1)}(u) - \tilde{\xi}_\delta^{(1)}(u)| \le \hat{k}^{(5)}\varepsilon^{(4)}\eta^{-\theta},$$
$$\sup_{0\le u\le t_1(\delta,\eta)} \{e^{\beta_1 u}|\xi_\delta^{(2)}(u) - \tilde{\xi}_\delta^{(2)}(u)|\} \le \hat{k}^{(6)}\varepsilon^{(4)}\eta^{-\theta}$$

for suitable $\hat{k}^{(5)}$ and $\hat{k}^{(6)}$.

Note that the estimates made in the discussion preceding Theorem 1.2 can be justified by the final statements of Lemma 2.2. Taking $\varepsilon^{(4)} = O(N^{-\gamma})$ for some $\gamma > 0$ and $\eta = N^{-5/12}$, choose θ such that $\theta < \max\{\kappa'/\beta_1, 6\gamma/5\}$.

3. The branching approximation

In this section we establish the approximation to $Nx_N^{(2)}$ by a Markov branching process Z in the early stages, starting with $x_N(0) = x_{N,0}$ such that $x_{N,0}^{(2)} = N^{-1}Z_0$ and $|x_{N,0}^{(1)} - x_0^{(1)}| \le \varepsilon_N^{(1)} := N^{-\alpha}$ for $\alpha > \frac{1}{3}$. The process Z is obtained by replacing $\bar{g}^J(x_N(t))$ by $\bar{g}^J(x_0)$ in the transition rates which have $J^{(2)} \ne 0$, and by taking its corresponding jumps to be $J^{(2)}$. It is a Markov branching process; for each J such that $J^{(2)} \ne 0$, an individual of type $s(J)$ gives birth to J_i individuals of type i, $d_1 < i \le d$, (if $J_{s(J)} = -1$, this represents the death of an individual of type $s(J)$) with *per capita* rate $\bar{g}^J(x_0)$. It is thus natural to index the components of Z by $\{d_1 + 1, \dots, d\}$ to match the indexing in X_N; we denote the resulting state space of Z by $\mathcal{Z}$. For $z \in \mathcal{Z}$, let

$$q^J(z) := \bar{g}^J(x_0)z_{s(J)}, \qquad q(z) := \sum_{J\in\mathcal{J}_2} q^J(z), \tag{3.1}$$

then, if Z is in state z, the time until its next jump is distributed as Exp $(q(z))$, and the probability

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

that it is a J transition, causing a corresponding change of $J^{(2)}$ in Z, is $q^J(z)/q(z)$, $J \in \mathcal{J}_2$. Since there are only finitely many $J \in \mathcal{J}$, the means and covariances of the offspring distributions of individuals of the different types are all finite. In particular, as noted in Section 1.2, the mean growth rate matrix is given by $B_0^\top$, whose positive left and right eigenvectors $\boldsymbol{u}^\top$ and $\boldsymbol{v}$ are normalized so that $\boldsymbol{u}^\top \mathbf{1} = \boldsymbol{u}^\top \boldsymbol{v} = 1$. Our approximation shows that, except on an event of negligible probability, the process $Nx_N^{(2)}$ can be constructed so as to have paths identical to those of Z, up to the time τ_{N1}^Z at which, if ever, $\boldsymbol{v}^\top Z$ has grown by at least the amount $N^{1-\alpha}$ from its initial value of $\boldsymbol{v}^\top Z_0$. The full details are given below in Proposition 3.1.

We begin by considering the first components $x_N^{(1)}(\cdot)$ of x_N. Under our assumptions on F, they satisfy the equation

$$\mathrm{d}x_N^{(1)}(t) = A(x_N(t))x_N^{(2)}(t) + C(x_N^{(1)}(t) - x_0^{(1)}) + \tilde{c}(x_N^{(1)}(t)) + \mathrm{d}m_N^{(1)}(t),$$

where m_N is as defined in (1.2), and this can be integrated by variation of constants to give

$$\begin{aligned} x_N^{(1)}(t) = x_0^{(1)} + \mathrm{e}^{Ct}(x_{N,0}^{(1)} - x_0^{(1)}) + \int_0^t \mathrm{e}^{C(t-u)}\{A(x_N(u))x_N^{(2)}(u) + \tilde{c}(x_N^{(1)}(u))\}\,\mathrm{d}u \\ + m_N^{(1)}(t) + C\int_0^t \mathrm{e}^{C(t-u)} m_N^{(1)}(u)\,\mathrm{d}u; \end{aligned} \tag{3.2}$$

note that

$$m_N^{(1)}(t) + C\int_0^t \mathrm{e}^{C(t-u)} m_N^{(1)}(u)\,\mathrm{d}u = \int_0^t \mathrm{e}^{C(t-u)}\,\mathrm{d}m_N^{(1)}(u),$$

explaining the stochastic term in (3.2). For $x_N^{(2)}$, up to the time at which it has made $n(N)$ jumps, it is enough for now to know that it is bounded by $N^{-1}\{|Z_0| + J^* n(N)\}$, where $J^* := \max_{J\in\mathcal{J}_2} |J|$.

We first use (3.2) to show that $x_N^{(1)}(t)$ moves away from $x_0^{(1)}$ rather slowly. For this, it is necessary to show that $|m_N|$ remains uniformly small with high probability for a long enough time interval. This is the substance of the following lemma. To state it, we define

$$\tau_N := \inf\{t > 0 \colon |x_N(t) - x_0| > \rho_2\}$$

and use $\mathbb{P}^0$ to denote probabilities given $x_N(0) = x_{N,0}$.

Lemma 3.1. *Let $T_N := k \log N$ for some $k > 0$, and define*

$$E_N(k) := \Big\{ \sup_{0\le t\le T_N\wedge\tau_N} |m_N(t)| \le \eta_{N1}(k) \Big\},$$

where $\eta_{N1}(k) := (2\sqrt{k}\sum_{J\in\mathcal{J}} |J|)N^{-1/2}(\log N)^{3/2}$. Then $\mathbb{P}^0[E_N(k)^c] = O(N^{-r})$ for any $r > 0$.

Proof. Let $E'_N(k)$ denote the event

$$E'_N(k) := \Big\{ \max_{J\in\mathcal{J}} \sup_{0\le t\le T_N\wedge\tau_N} \Big| \frac{P^J(NG_N^J(t))}{N} - G_N^J(t) \Big| \le 2\frac{\sqrt{T_N}\log N}{\sqrt{N}} \Big\}.$$

Note that the quantities $t^{-1}G_N^J(t)$ are uniformly bounded in $t \le \tau_N$, because the functions g^J are continuous and $x_N(t)$ is restricted to a compact set for such t. Denoting this bound by g^*, it follows from the Chernoff inequalities that, for N such that $T_N \geq 1$,

$$\mathbb{P}[E'_N(k)^c] \leq 2|\mathcal{J}|\sqrt{N} g^* T_N N \exp\left\{-\frac{(\log N)^2}{\{2(g^*+1)\}}\right\} = O(N^{-r}) \quad \text{for any } r > 0. \tag{3.3}$$

However, on the event $E'_N(k)$,

$$\sup_{0 \leq t \leq T_N \wedge \tau_N} |m_N(t)| \leq \left(2\sqrt{k}\sum_{J\in\mathcal{J}}|J|\right)\frac{(\log N)^{3/2}}{\sqrt{N}}$$

so that $E_N(k) \supset E'_N(k)$, which, with (3.3), proves the lemma.

Now define $\tau_1(m) := \inf\{t > 0\colon |x_N^{(2)}(t)| > m/N\}$ and write

$$d_N^{(1)}(t,m) := \sup_{0\leq u\leq t\wedge\tau_1(m)} |x_N^{(1)}(u) - x_0^{(1)}|.$$

Lemma 3.2. *With the assumptions and notation of Section 1.1, fix any $k > 0$, and assume that N is large enough so that*

$$k\log N \max\{\gamma_1|x_{N,0}^{(1)} - x_0^{(1)}|, \eta'_{N1}(k)\} \leq \frac{1}{(40\gamma_1 K_c)},$$

where $\eta'_{N1}(k) := \eta_{N1}(k)(1+\gamma_1\|C\|k\log N)$. Suppose that $E_N(k)$ occurs. Then, for all $0 \leq t \leq k\log N$ and $m < N/\{20k^2\gamma_1^2 K_c\|A\|_{\rho_2}(\log N)^2\}$,

$$d_N^{(1)}(t,m) \leq \frac{8}{7}\left\{\gamma_1\left(|x_{N,0}^{(1)} - x_0^{(1)}| + t\|A\|_{\rho_2}\left(\frac{m}{N}\right)\right) + \eta'_{N1}(k)\right\}.$$

Proof. From (3.2) and the assumptions on C and ρ_2, and from the definition of $E_N(k)$, it follows immediately that for $t \leq (\tau_1(m)\wedge k\log N)$ such that

$$\gamma_1 K_c \int_0^t d_N^{(1)}(u,m)\,\mathrm{d}u \leq \frac{1}{8}, \tag{3.4}$$

we have

$$\begin{aligned} |x_N^{(1)}(t) - x_0^{(1)}| &\leq \gamma_1|x_{N,0}^{(1)} - x_0^{(1)}| + \eta_{N1}(k) \\ &\quad + \gamma_1\int_0^t\left\{\|A\|_{\rho_2}\frac{m}{N} + K_c\{d_N^{(1)}(u,m)\}^2 + \|C\|\eta_{N1}(k)\right\}\mathrm{d}u \\ &\leq \gamma_1\left\{|x_{N,0}^{(1)} - x_0^{(1)}| + t\|A\|_{\rho_2}\frac{m}{N}\right\} + \frac{1}{8}d_N^{(1)}(t,m) + \eta'_{N1}(k). \end{aligned} \tag{3.5}$$

Now for $t \leq (\tau_1(m)\wedge k\log N)$, (3.5) implies that

$$\begin{aligned} \gamma_1 K_c\int_0^t d_N^{(1)}(u,m)\,\mathrm{d}u &\leq \gamma_1 K_c\frac{8}{7}\left\{\gamma_1\left(t|x_{N,0}^{(1)} - x_0^{(1)}| + \frac{1}{2}t^2\|A\|_{\rho_2}\left(\frac{m}{N}\right)\right) + t\eta'_{N1}(k)\right\} \\ &\leq \tfrac{8}{7}\left\{\tfrac{1}{40} + \tfrac{1}{40} + \tfrac{1}{40}\right\} \\ &= \tfrac{3}{35} \\ &< \tfrac{1}{8}, \end{aligned}$$

the bound assumed in (3.4).

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

Hence, since $\int_0^t d_N^{(1)}(u,m)\,\mathrm{d}u$ is continuous in t, we can apply Lemma D.1 of the online appendix (see Barbour *et al.* (2014)) with $\varphi = 0$ to show that, for all sufficiently large N, the inequality (3.5) holds for all $t \le (\tau_1(m) \wedge k \log N)$, and the lemma is proved.

Since, in the early phase, $x_N^{(2)}(t) \approx 0$ and Lemma 3.2 shows that $x_N^{(1)}(t) \approx x_0^{(1)}$, the process $Nx_N^{(2)}$ can plausibly be well approximated by replacing $\bar g^J(x_N(t))$ by $\bar g^J(x_0)$ in its transition rates, obtaining the Markov branching process Z. To show that this is indeed the case, we consider a path starting in Z_0, having $J_1,\dots,J_n$ as its first n transitions and $t_1,\dots,t_n$ their times. Then the probability density of this path segment is given by

$$\prod_{l=0}^{n-1}(\exp\{-(t_{l+1}-t_l)q(z_l)\}q^{J_{l+1}}(z_l)),$$

where $z_l := Z_0 + \sum_{i=1}^{l} J_i^{(2)}$, $t_0 = 0$, and the functions q and q^J are as in (3.1).

The corresponding expression for $Nx_N^{(2)}$ is more complicated, since the process is only Markovian if the state space is extended to include all the original coordinates. Define

$$q_N^J(x^{(1)},z) := \bar g^J\left(\left[x^{(1)}, \frac{z}{N}\right]\right) z_{s(J)}, \qquad q_N(x^{(1)},z) := \sum_{J\in\mathcal{J}_2} q_N^J(x^{(1)},z)$$

for $x \in \mathbb{R}_+^d$, $z \in \mathcal{Z}$, and $J \in \mathcal{J}_2$, with $[y_1, y_2]$ denoting $(y_1^\top, y_2^\top)^\top$. Writing

$$H^J(x^{(1)},z,t) := \mathbb{E}^{(x^{(1)},z)}\left(\exp\left\{-\int_0^t q_N(x_N^{(1)}(u),z)\,\mathrm{d}u\right\} q_N^J(x_N^{(1)}(t),z)\right),$$

the probability density at $\{(J_1,t_1),\dots,(J_n,t_n)\}$ is given by

$$\mathbb{E}^0\left(\prod_{l=0}^{n-1} H^{J_{l+1}}(x_N^{(1)}(t_l), z_l, t_{l+1}-t_l)\right);$$

here, $\mathbb{E}^{(x^{(1)},z)}$ denotes expectation conditional on $x_N(0) = \binom{x^{(1)}}{N^{-1}z}$, and $\mathbb{E}^0$ as before denotes expectation conditional on $x_N(0) = x_{N,0}$. Hence, the likelihood ratio, with respect to the branching process measure, of a path successively entering the states $z_{\{1,n\}} := z_1, z_2, \dots, z_n$ at times $t_{\{1,n\}} := t_1,\dots,t_n$ is given by

$$R_n(z_{\{1,n\}}, t_{\{1,n\}}) := \mathbb{E}^0\left(\prod_{l=0}^{n-1} \widetilde H^{J_{l+1}}(x_N^{(1)}(t_l), z_l, t_{l+1}-t_l)\right), \tag{3.6}$$

where

$$\begin{aligned}\widetilde H^J(x^{(1)},z,t) &:= \frac{H^J(x^{(1)},z,t)\mathrm{e}^{tq(z)}}{q^J(z)} \\ &= \mathbb{E}^{(x^{(1)},z)}\left(\exp\left\{-\int_0^t \{q_N(x_N^{(1)}(u),z) - q(z)\}\,\mathrm{d}u\right\}\frac{q_N^J(x_N^{(1)}(t),z)}{q^J(z)}\right). \end{aligned}\tag{3.7}$$

Let $\tau_2(n)$ denote the time at which $x_N^{(2)}$ makes its nth jump. Then $|x_N^{(1)}(u) - x_0^{(1)}|$ remains uniformly of order

$$O\left\{\frac{(\log N)^{5/2}}{\sqrt{N}} + N^{-\alpha} + \frac{n(N)\log N}{N}\right\} \quad \text{for } 0 \le u \le (\tau_2(n(N)) \wedge k\log N),$$

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

except on an event of probability $O(N^{-r})$, for any $r > 0$, because of Lemma 3.2. This implies that the rates $q_N^J(x_N^{(1)}(u), z)$ are close to $q^J(z)$, $J \in \mathcal{J}_2$, throughout the same u-interval. We now use this to show that the joint distributions of the times and values of the first $n(N)$ jumps of the processes Z and Nx_N are close to one another.

Lemma 3.3. *Let $x_N(0)$ be such that $|x_N^{(1)}(0) - x_0^{(1)}| = O(N^{-\alpha})$ and $x_N^{(2)}(0) = N^{-1}Z_0$. Then, for any fixed $k > 0$ and $\frac{1}{3} < \alpha < 1$, the total variation distance $d_{\mathrm{TV}}^{(N)}$ between the distributions of the paths of $Nx_N^{(2)}$ and those of Z, restricted to the first $kN^{1-\alpha}$ jumps, is such that* $\lim_{N\to\infty} d_{\mathrm{TV}}^{(N)} = 0$.

Proof. Letting $\tau_{\{1,n\}}$ denote the times of the first n jumps of Z, the main aim is to show that the likelihood ratio $R_n(Z_{\{1,n\}}, \tau_{\{1,n\}})$ defined in (3.6) is close to 1 with high probability. First, defining

$$\widehat{H}^J(x^{(1)}, z, t, y) := \mathbb{E}^{(x^{(1)},z)}\left(\exp\left\{-\int_0^t \{q_N(x_N^{(1)}(u), z) - q(z)\}\,\mathrm{d}u\right\} \frac{q_N^J(y, z)}{q^J(z)} \mathbf{1}_{\{x_N^{(1)}(t)=y\}}\right)$$

for $y \in N^{-1}\mathbb{Z}_+^{d_1}$, we can express the ratio

$$V_{n+1}(z_{\{1,n+1\}}, t_{\{1,n+1\}}) := \frac{R_{n+1}(z_{\{1,n+1\}}, t_{\{1,n+1\}})}{R_n(z_{\{1,n\}}, t_{\{1,n\}})}$$

as

$$V_{n+1}(z_{\{1,n+1\}}, t_{\{1,n+1\}}) = \mathbb{E}_n\{\widetilde{H}^{J_{n+1}}(Y, z_{n+1}, t_{n+1} - t_n)\}, \tag{3.8}$$

where $\mathbb{E}_n$ denotes expectation with respect to the measure with probabilities $p_n(y)$ given by

$$\frac{\mathbb{E}^0\left(\widehat{H}(x_N^{(1)}(t_{n-1}), z_{n-1}, t_n - t_{n-1}, y) \prod_{l=0}^{n-2} \widetilde{H}^{J_{l+1}}(x_N^{(1)}(t_l), z_l, t_{l+1} - t_l)\right)}{R_n(z_{\{1,n\}}, t_{\{1,n\}})} \quad \text{for } y \in \frac{\mathbb{Z}_+^{d_1}}{N}.$$

Now, defining the σ-fields $\Sigma_n := \sigma(Z_{\{1,n\}}, \tau_{\{1,n\}})$, the process $(R_n(Z_{\{1,n\}}, \tau_{\{1,n\}}), \Sigma_n,\ n \geq 0)$, being a likelihood ratio, is a martingale with expectation 1. We wish to show that it stays close to its expectation with high probability.

First, we consider the process x_N obtained by replacing $\bar{g}(x)$ with $\bar{g}(x_0)$ in the transition rates for jumps $J \in \mathcal{J}_2$, whenever $|x - x_0| > \theta_N$, yielding a new process $x_{N,\theta}$; the quantity $\theta_N \leq \rho_2$ is yet to be determined. We then conduct the whole analysis for $x_{N,\theta}$. Observe that, in (3.8), the quantity $\widetilde{H}^{J_{n+1}}(Y, z_{n+1}, t_{n+1} - t_n)$ is, from its definition (3.7), itself an expectation, and that, for the process $x_{N,\theta}$, the quantity within the expectation is itself close to 1. To see this, let $Q^* := \max_{J\in\mathcal{J}_2}\{\|D\bar{g}^J\|_{\rho_2}/b_*^J\}$; then

$$\begin{aligned}
&\left|\exp\left\{-\int_0^t \{q_N(x_{N,\theta}(u), z) - q(z)\}\,\mathrm{d}u\right\} \frac{q_N^J(x_{N,\theta}(t), z)}{q^J(z)} - 1\right| \\
&\quad \leq \exp\{q(z)tQ^*\theta_N\}(1 + Q^*\theta_N) - 1 \\
&\quad \leq Q^*\theta_N\left\{\left(\frac{5}{4}\right)q(z)t\exp\left(\frac{q(z)t}{4}\right) + 1\right\},
\end{aligned}$$

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

if also $\theta_N \le 1/\{4Q^*\}$. Hence, for $x_{N,\theta}$,

$$\begin{aligned}\mathbb{E}\{(V_{n+1}(Z_{\{1,n+1\}}, \tau_{\{1,n+1\}}) - 1)^2 \mid \Sigma_n\} \\ \le \{Q^*\theta_N\}^2 \int_0^\infty q(z_n)\mathrm{e}^{-q(z_n)t}\left\{\left(\frac{5}{4}\right)q(z_n)t\mathrm{e}^{q(z_n)t/4} + 1\right\}^2 \mathrm{d}t \\ \le 27\{Q^*\theta_N\}^2.\end{aligned}$$

Writing $\psi_N := 27\{Q^*\theta_N\}^2$ and $R_r := R_r(Z_{\{1,r\}}, \tau_{\{1,r\}})$, we obtain

$$\begin{aligned}\mathbb{E}^0\{(R_n - 1)^2\} &= \mathbb{E}^0\{(R_n - R_{n-1})^2\} + \mathbb{E}^0\{(R_{n-1} - 1)^2\} \\ &\le \psi_N \mathbb{E}^0 R_{n-1}^2 + \mathbb{E}^0\{(R_{n-1} - 1)^2\} \\ &= \mathbb{E}^0\{(R_{n-1} - 1)^2\}(1 + \psi_N) + \psi_N \\ &\le (1 + \psi_N)^n - 1.\end{aligned}$$

In consequence, for the process $x_{N,\theta}$, if $n\psi_N \le 1$,

$$\mathbb{E}^0\{(R_n(Z_{\{1,n\}}, \tau_{\{1,n\}}) - 1)^2\} \le ne\psi_N = 27ne\{Q^*\theta_N\}^2. \tag{3.9}$$

Now the total variation distance d_{TV} between probability measures $\mathbb{P}_1$ and $\mathbb{P}_2$ on a measurable space $(S, \mathcal{F})$ can be expressed as

$$d_{\mathrm{TV}}(\mathbb{P}_1, \mathbb{P}_2) := \sup_{A \in \mathcal{F}} |\mathbb{P}_1(A) - \mathbb{P}_2(A)| = \frac{1}{2}\int_S |R_{12} - 1|\, \mathrm{d}\mathbb{P}_2 = -\int_S \min\{0, R_{12} - 1\}\, \mathrm{d}\mathbb{P}_2,$$

where $R_{12} := \mathrm{d}\mathbb{P}_1/\mathrm{d}\mathbb{P}_2$. In view of (3.9), it thus follows that for $x_{N,\theta}$, $d_{\mathrm{TV}}^{(N,\theta)} = O(n(N)^{1/2}\theta_N)$. By Thorisson (2000, Chapter 3, Theorem 7.3, and Equation (8.19)), this also implies that the process $Nx_{N,\theta}$ and the branching process Z can be realized on the same probability space in such a way that their paths coincide up to the first $n(N)$ jumps, except on an event of probability of order $O(n(N)^{1/2}\theta_N)$.

Now fix $k > 0$, to be specified later, and for $m(N) := |Z_0| + J^*n(N)$, define

$$\theta_N^{(1)} := \frac{8}{7}\left\{\gamma_1\left\{|x_N^{(1)}(0) - x_0^{(1)}| + k\log N\|A\|_{\rho_2}\left(\frac{m(N)}{N}\right)\right\} + \eta'_{N1}(k)\right\}$$

and $\theta_N^{(2)} := N^{-1}m(N)$; set $\theta_N := \theta_N^{(1)} + \theta_N^{(2)}$. Note that with $n(N) = O(N^{1-\alpha})$ for $\frac{1}{3} < \alpha < 1$, this choice of θ_N satisfies $\theta_N \le 1/\{4Q^*\}$ for all large enough N, and that the total variation distance $d_{\mathrm{TV}}^{(N,\theta)}$ is of small order $O\{n(N)^{1/2}(N^{-1}n(N)\log N + N^{-\alpha})\}$. Now x_N and $x_{N,\theta}$ can be coupled by running their paths identically until $\tau_N(\theta_N) := \inf\{t > 0\colon |x_{N,\theta}(t) - x_0| > \theta_N\}$. So, for this choice of θ_N, let

$$\sigma_N^{(1)} := \inf\{t > 0\colon |x_{N,\theta}^{(1)}(t) - x_0^{(1)}| > \theta_N^{(1)}\}, \qquad \sigma_N^{(2)} := \inf\{t > 0\colon |x_{N,\theta}^{(2)}(t)| > \theta_N^{(2)}\}.$$

If, for $n(N) := k_0N^{1-\alpha}$, we can show that $\mathbb{P}[\sigma_N^{(1)} \wedge \sigma_N^{(2)} < \tau_{n(N)} \wedge \tau_N^x(0)]$ is asymptotically small as $N \to \infty$, where τ_n denotes the time of the nth jump of $x_{N,\theta}^{(2)}$ and $\tau_N^x(0)$, as in (1.7), its time of first hitting 0, the lemma will be proved.

It is immediate from the definition of $\theta_N^{(2)}$ that $\sigma_N^{(2)} \ge \tau_{n(N)}$ a.s. Then, by Lemma 3.2, $\mathbb{P}[\{\sigma_N^{(1)} < \tau_{n(N)} \wedge \tau_N^x(0)\} \cap \{\sigma_N^{(1)} \le k\log N\}] = O(N^{-r})$ for any $k, r > 0$. Finally,

$$\begin{aligned}\mathbb{P}[\{\sigma_N^{(1)} < \tau_{n(N)} \wedge \tau_N^x(0)\} \cap \{\sigma_N^{(1)} > k\log N\}] \\ \le \mathbb{P}[\{\tau_{n(N)} > k\log N\} \cap \{|x_{N,\theta}^{(2)}(k\log N)| > 0\}] \\ \le d_{\mathrm{TV}}^{(N,\theta)} + \mathbb{P}[0 < W^{\boldsymbol{v}}(k\log N) \le m(N)\exp\{-\beta_0 k\log N\}],\end{aligned}$$

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

where $W^{\mathbf{v}}(t) := \mathbf{v}^\top Z(t)\mathrm{e}^{-\beta_0 t}$. However, choosing any $k > \beta_0^{-1}(1-\alpha)$, the latter probability is asymptotically small, because $m(N) = O(N^{1-\alpha})$ and, writing $\tau^Z(0) := \inf\{t > 0\colon Z(t) = 0\}$, as in (1.7),

$$\Big\{\lim_{t\to\infty} W^{\mathbf{v}}(t) = 0\Big\} = \{\tau^Z(0) < \infty\} \quad \text{a.s.}$$

(Athreya and Ney (1972, Chapter V.7, Theorem 2, Equation (27))). This proves the lemma.

As a result of Lemma 3.3, for any fixed k, probabilities for the paths of $Nx_N^{(2)}$ up to the first $kN^{1-\alpha}$ jumps can, with only small error, be computed using the branching process Z instead. We complete our treatment of this phase of development by proving two further lemmas. The first shows that the branching approximation remains accurate until $t = \tau_{N,\alpha}^x$, defined in (1.9). The second shows that $x_N(\tau_{N,\alpha}^x)$ is close to a point on the solution ξ_N of (1.3) starting from $x_{N,0}$, except on an event $\widehat{E}_N$ whose complement has asymptotically negligible probability. The proofs are given in Section B of the online appendix; see Barbour *et al.* (2014).

Lemma 3.4. *For $\tau_{N,\alpha}^Z$ defined in (1.8), let $\nu_{N,\alpha}^Z$ denote the number of jumps made by Z until time $\tau_{N,\alpha}^Z$, infinite if $\tau_{N,\alpha}^Z = \infty$. Then, under the assumptions of Section 1.1, there are constants k_0 and θ_0 such that*

$$\mathbb{P}^0[k_0 N^{1-\alpha} < \nu_{N,\alpha}^Z < \infty] \le \mathrm{e}^{-\theta_0 N^{1-\alpha}}.$$

Lemma 3.5. *Suppose that $\frac{1}{3} < \alpha < \frac{1}{2}$. Then there is a $\gamma > 0$ and an event $\widehat{E}_N$ satisfying $\lim_{N\to\infty} \mathbb{P}^0[\widehat{E}_N^c \cap \{\tau_{N,\alpha}^x < \infty\}] = 0$ such that, on the event $\widehat{E}_N \cap \{\tau_{N,\alpha}^x < \infty\}$, we have*

$$|x_N^{(2)}(\tau_{N,\alpha}^x) - \xi_N^{(2)}(t_{N,\alpha}^\xi)| = O(N^{-\alpha-\gamma}), \qquad |x_N^{(1)}(\tau_{N,\alpha}^x) - \xi_N^{(1)}(t_{N,\alpha}^\xi)| = O(N^{-\alpha}\log N).$$

We summarize the results of this section in the following proposition. For use in the sections to come, we specialize to $\alpha = \frac{5}{12}$.

Proposition 3.1. *Suppose that $|x_N^{(1)}(0) - x_0^{(1)}| \le N^{-5/12}$ and that $Nx_N^{(2)}(0) = Z_0$ for some fixed Z_0. Define $\tau^Z(0)$ as in (1.7), and τ_{N*}^Z, τ_{N*}^x, and t_{N*}^ξ as in (1.10). Then, under the assumptions of Section 1.1, it is possible to couple the paths of $Nx_N^{(2)}$ and of the branching process Z in such a way that, except on an event of asymptotically negligible probability, they are identical until time $\min\{\tau^Z(0), \tau_{N*}^Z\}$, when, in particular, $\tau_{N*}^Z = \tau_{N*}^x$. Furthermore, there is a $\gamma > 0$ and constants $\bar{k}^{(1)}$ and $\bar{k}^{(2)}$ such that, if $\tau_{N*}^x < \infty$,*

$$|x_N^{(1)}(\tau_{N*}^x) - \xi_N^{(1)}(t_{N*}^\xi)| \le \bar{k}^{(1)} N^{-5/12}\log N, \qquad |x_N^{(2)}(\tau_{N*}^x) - \xi_N^{(2)}(t_{N*}^\xi)| \le \bar{k}^{(2)} N^{-5/12-\gamma},$$
$$\tau_{N*}^x = \beta_0^{-1}\{(1-\alpha)\log N - \log W\} + O(N^{-7/48}),$$

except on an event $\widehat{E}_N^c$ of negligible probability, where ξ_N is the solution to the deterministic equation starting with $\xi_N(0) = x_N(0)$.

4. Intermediate growth

In the previous section, it has been shown that, on $\{\tau_{N*}^x < \infty\} \cap \widehat{E}_N$, the point $x_N(\tau_{N*}^x)$ is close to $\xi_N(t_{N*}^\xi)$, where ξ_N is the solution to (1.3) with initial condition $\xi_N(0) = x_{N,0}$, and t_{N*}^ξ is a nonrandom time defined in (1.10). We now show that $x_N(\tau_{N*}^x + t)$ stays uniformly close to $\xi_N(t_{N*}^\xi + t)$ for all $0 \le t \le t_0(\delta', \varepsilon_N)$, for a suitably chosen $\delta' > 0$, not depending on N; here, and throughout the section, we define

$$\varepsilon_N := |x_N^{(2)}(\tau_{N*}^x)| \asymp N^{-5/12}, \tag{4.1}$$

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

with the last relation and the inequality $\varepsilon_N \geq N^{-5/12}/|\boldsymbol{v}|$ justified in view of the definition (1.10) of τ^x_{N*}. We also show that δ' can be chosen so that all the components of $\xi_N(t^{\xi}_{N*} + t_0(\delta', \varepsilon_N))$ are bounded away from 0. Hence, after this time Kurtz (1970, Theorem (3.1)) can be used to continue the approximation of x_N by ξ_N along a nondegenerate path, as stated in Theorem 1.1.

We start by using the Markov property to continue from τ^x_{N*}. Let $x_1 := x_N(\tau^x_{N*})$, i.e. $x^{(1)}_1 = x^{(1)}_N(\tau^x_{N*})$, $x^{(2)}_1 = x^{(2)}_N(\tau^x_{N*})$, and define $\tilde{x}_N(t) := x_N(\tau^x_{N*} + t)$. Note that, from Lemma 2.1,

$$|\xi^{(1)}_N(t^{\xi}_{N*}) - x^{(1)}_0| \leq k^{(1)}\left\{|x^{(1)}_{N,0} - x^{(1)}_0| + \left(\frac{|Z_0|}{\boldsymbol{v}^{\top} Z_0}\right)N^{-5/12}\right\} \leq \bar{k}^{(3)} N^{-5/12}$$

with $\bar{k}^{(3)} := k^{(1)}(1 + \max_{1\leq i\leq d}\{1/\boldsymbol{v}_i\})$. Then we can write

$$\tilde{x}^{(1)}_N(t) = x^{(1)}_1 + \int_0^t \{A(\tilde{x}_N(u))\tilde{x}^{(2)}_N(u) + c(\tilde{x}^{(1)}_N(u))\}\,\mathrm{d}u + \tilde{m}^{(1)}_N(t),$$

$$\tilde{x}^{(2)}_N(t) = x^{(2)}_1 + \int_0^t B(\tilde{x}_N(u))\tilde{x}^{(2)}_N(u)\,\mathrm{d}u + \tilde{m}^{(2)}_N(t),$$

or, using variation of constants

$$\tilde{x}^{(1)}_N(t) = x^{(1)}_0 + \mathrm{e}^{Ct}(x^{(1)}_1 - x^{(1)}_0) + \int_0^t \mathrm{e}^{C(t-u)}\{A(\tilde{x}_N(u))\tilde{x}^{(2)}_N(u) + \tilde{c}(\tilde{x}^{(1)}_N(u))\}\,\mathrm{d}u + \tilde{m}^{(1)}_N(t) + C\int_0^t \mathrm{e}^{C(t-u)}\tilde{m}^{(1)}_N(u)\,\mathrm{d}u, \tag{4.2}$$

$$\begin{aligned}\tilde{x}^{(2)}_N(t) &= x^{(2)}_1 + \int_0^t B(\tilde{x}_N(u))\tilde{x}^{(2)}_N(u)\,\mathrm{d}u + \tilde{m}^{(2)}_N(t)\\ &= \mathrm{e}^{B_0 t}x^{(2)}_1 + \int_0^t \mathrm{e}^{B_0(t-u)}\{B(\tilde{x}_N(u)) - B_0\}\tilde{x}^{(2)}_N(u)\,\mathrm{d}u\\ &\quad + \tilde{m}^{(2)}_N(t) + B_0\int_0^t \mathrm{e}^{B_0(t-u)}\tilde{m}^{(1)}_N(u)\,\mathrm{d}u,\end{aligned} \tag{4.3}$$

where $\tilde{m}_N(t) := m_N(\tau^x_{N*}+t) - m_N(\tau^x_{N*})$ and m_N is as in (1.2). The deterministic counterparts of (4.2) and (4.3) have been given previously in (2.1) and (2.2). We first use the comparison between these pairs of equations to show that $\tilde{x}_N$ stays close to $\tilde{\xi}_N$, where $\tilde{\xi}_N$ solves (1.3) with $\tilde{\xi}_N(0) = \tilde{x}_N(0) = x_1$. Afterwards, we can use Lemma 2.1 to show that $\tilde{\xi}_N(\cdot)$ stays uniformly close to $\xi_N(t^{\xi}_{N*} + \cdot)$ in the appropriate time interval, and that, at the end of this interval, ξ_N is away from the boundary. In preparation for the next result, taking δ_0 as in Lemma 2.1, note that $0 < t_0(\delta_0, \varepsilon_N) = \beta_0^{-1}\log(\delta_0/\varepsilon_N) \leq k_1 \log N$ for a suitable choice of k_1 and for all sufficiently large N.

Lemma 4.1. *There exist $\delta_1 > 0$ and an event E_N with $\mathbb{P}[E^c_N] = O(N^{-r})$ for any $r > 0$ such that, on E_N, for all $\delta \leq \delta_1$,*

$$\sup_{0\leq t\leq t_0(\delta_0,\varepsilon_N)} |\tilde{x}^{(1)}_N(t) - \tilde{\xi}^{(1)}_N(t)| \leq k^{(1)}(\delta_1)N^{-1/12+\chi(\delta)},$$

$$\sup_{0\leq t\leq t_0(\delta_0,\varepsilon_N)} \left\{\frac{|\tilde{x}^{(2)}_N(t) - \tilde{\xi}^{(2)}_N(t)|}{|\tilde{\xi}^{(2)}_N(t)|}\right\} \leq k^{(2)}(\delta_1)N^{-1/12+\chi(\delta)}$$

for some $k^{(1)}(\delta_1)$, $k^{(2)}(\delta_1)$, and $\chi(\delta) > 0$, where $\lim_{\delta\to 0}\chi(\delta) = 0$.

Proof. Define the (random) time

$$\tau_N := \inf\{t > 0\colon |\tilde{x}_N(t) - x_0| > \rho_2\},$$

and let E_N denote the event

$$\left\{\max_{J\in\mathcal{J}} \sup_{0\le t\le t_0(\delta_0,\varepsilon_N)\wedge\tau_N} \left|\frac{P^J(NG_N^J(t))}{N} - G_N^J(t)\right| \le 2\frac{\sqrt{k_1}(\log N)^{3/2}}{\sqrt{N}}\right\}$$

for k_1 as defined above. Then, by Lemma 3.1, $\mathbb{P}[E_N^c] = O(N^{-r})$ for any $r > 0$ and, on the event E_N,

$$\sup_{0\le t\le t_0(\delta_0,\varepsilon_N)\wedge\tau_N} |\tilde{m}_N(t)| \le \eta_{N1} := \eta_{N1}(k_1) = O\biggl(\frac{(\log N)^{3/2}}{\sqrt{N}}\biggr).$$

The remaining argument involves careful use of the Gronwall inequality on the event E_N, to translate the smallness of $\sup_{0\le t\le t_0(\delta_0,\varepsilon_N)\wedge\tau_N} |\tilde{m}_N(t)|$ into a corresponding closeness of $\tilde{x}_N$ and $\tilde{\xi}_N$ over a large part of this time interval. The main difficulty is that the length of the interval tends to ∞ with N.

Taking the difference of (4.2) and (2.1), we find that, on E_N,

$$\begin{aligned}
&|\tilde{x}_N^{(1)}(t) - \tilde{\xi}_N^{(1)}(t)| \\
&\quad\le \int_0^t |\mathrm{e}^{C(t-u)}\{A(\tilde{x}_N(u))\tilde{x}_N^{(2)}(u) - A(\tilde{\xi}_N(u))\tilde{\xi}_N^{(2)}(u)\}|\,\mathrm{d}u \\
&\qquad + \int_0^t |\mathrm{e}^{C(t-u)}\{\tilde{c}(\tilde{x}_N^{(1)}(u)) - \tilde{c}(\tilde{\xi}_N^{(1)}(u))\}|\,\mathrm{d}u + \eta'_{N1} \\
&\quad\le \|A\|_{\rho_2}\gamma_1 \int_0^t |\tilde{x}_N^{(2)}(u) - \tilde{\xi}_N^{(2)}(u)|\,\mathrm{d}u \\
&\qquad + \gamma_1 \int_0^t |\tilde{\xi}_N^{(2)}(u)|\|DA\|_{\rho_2}|\tilde{x}_N(u) - \tilde{\xi}_N(u)|\,\mathrm{d}u \\
&\qquad + \gamma_1 K_c \int_0^t |\tilde{x}_N^{(1)}(u) - \tilde{\xi}_N^{(1)}(u)|\{|\tilde{x}_N^{(1)}(u) - \tilde{\xi}_N^{(1)}(u)| + |\tilde{\xi}_N^{(1)}(u) - x_0^{(1)}|\}\,\mathrm{d}u + \eta'_{N1}
\end{aligned}$$

for $0 \le t \le t_0(\delta_0, \varepsilon_N) \wedge \tau_N$, where $\eta'_{N1} := (1 + \gamma_1\|C\|k_1 \log N)\eta_{N1}$. Writing

$$d_N^{(1)}(t) := \sup_{0\le u\le t\wedge\tau_N} |\tilde{x}_N^{(1)}(u) - \tilde{\xi}_N^{(1)}(u)|, \qquad d_N^{(2)}(t) := \sup_{0\le u\le t\wedge\tau_N} \mathrm{e}^{-\beta_0 u}|\tilde{x}_N^{(2)}(u) - \tilde{\xi}_N^{(2)}(u)|$$

it thus follows, for $t \le t_0(\delta_0, \varepsilon_N) \wedge \tau_N$ and on E_N, that

$$\begin{aligned}
d_N^{(1)}(t) \le \gamma_1 &\int_0^t \mathrm{e}^{\beta_0 u} d_N^{(2)}(u)\|A\|_{\rho_2}\,\mathrm{d}u \\
&+ \gamma_1\|DA\|_{\rho_2} \int_0^t \mathrm{e}^{\beta_0 u}\Delta_N^{(2)}(u)(d_N^{(1)}(u) + \mathrm{e}^{\beta_0 u} d_N^{(2)}(u))\,\mathrm{d}u \\
&+ \gamma_1 K_c d_N^{(1)}(t) \int_0^t \{d_N^{(1)}(u) + |\tilde{\xi}_N^{(1)}(u) - x_0^{(1)}|\}\,\mathrm{d}u + \eta'_{N1}.
\end{aligned}$$

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

Use Lemma 2.1 to bound $\Delta_N^{(2)}(t) := \sup_{0\le u\le t} \mathrm{e}^{-\beta_0 u}|\tilde{\xi}_N^{(2)}(u)|$. Here, we can take $\varepsilon^{(2)} = \varepsilon_N$ and $\varepsilon^{(1)} = (\bar{k}^{(1)} + \bar{k}^{(3)})N^{-5/12}\log N$ for $\tilde{\xi}_N(0) = x_N(\tau_{N*}^x)$, in view of Proposition 3.1, (4.1), and (4.2). This gives

$$d_N^{(1)}(t) \le \gamma_1\beta_0^{-1}\mathrm{e}^{\beta_0 t}\{d_N^{(2)}(t)\|A\|_{\rho_2} + k^{(2)}\varepsilon_N\|DA\|_{\rho_2}(d_N^{(1)}(t) + \mathrm{e}^{\beta_0 t}d_N^{(2)}(t))\} + \tfrac{1}{4}d_N^{(1)}(t) + \eta'_{N1}, \tag{4.4}$$

for all t such that

$$\gamma_1 K_c\int_0^t d_N^{(1)}(u)\,\mathrm{d}u \le \frac{1}{8}, \qquad \gamma_1 K_c\int_0^t |\tilde{\xi}_N^{(1)}(u) - x_0^{(1)}|\,\mathrm{d}u \le \frac{1}{8}. \tag{4.5}$$

Observe also that, from Lemma 2.1,

$$\gamma_1 K_c\int_0^t |\tilde{\xi}_N^{(1)}(u) - x_0^{(1)}|\,\mathrm{d}u \le \gamma_1 K_c k^{(1)}\beta_0^{-1}\left\{\varepsilon^{(1)}\log\left(\frac{1}{\varepsilon_N}\right) + \delta\right\}$$

for $t \le t_0(\delta, \varepsilon_N)$ and for any $\delta \le \delta_0$, where δ_0 is as in Lemma 2.1. With the above choice of $\varepsilon^{(1)}$ and for any $\delta = \delta'$ chosen small enough, smaller than δ_0 if necessary, the right-hand side is smaller than $\frac{1}{8}$ for all large enough N.

Now choose $0 < \delta_1 \le \min\{\delta_0, \delta'\}$ such that $\gamma_1 k^{(2)}\delta_1\beta_0^{-1}\|DA\|_{\rho_2} \le \frac{1}{4}$ and $\delta_1 \le \rho_2/2$, and consider $t \le t_0(\delta_1, \varepsilon_N)$ such that (4.5) is satisfied, and also such that

$$\max\{d_N^{(1)}(t), \mathrm{e}^{\beta_0 t}d_N^{(2)}(t)\} \le \delta_1 \tag{4.6}$$

for which, immediately, $t \le \tau_N$ and $\mathrm{e}^{\beta_0 t}\varepsilon_N \le \delta_1$. Then, from (4.4), it follows that, for such t and on E_N,

$$d_N^{(1)}(t) \le 2\gamma_1\beta_0^{-1}\mathrm{e}^{\beta_0 t}d_N^{(2)}(t)\{\|A\|_{\rho_2} + k^{(2)}\delta_1\|DA\|_{\rho_2}\} + 2\eta'_{N1}. \tag{4.7}$$

We now take the difference of (4.3) and (2.2), from which it follows that, for t as above and on E_N,

$$\begin{aligned}|\tilde{x}_N^{(2)}(t) - \tilde{\xi}_N^{(2)}(t)| \le &\int_0^t |\mathrm{e}^{B_0(t-u)}(B(\tilde{x}_N(u)) - B(\tilde{\xi}_N(u)))\tilde{x}_N^{(2)}(u)|\,\mathrm{d}u \\ &+ \int_0^t |\mathrm{e}^{B_0(t-u)}(B(\tilde{\xi}_N(u)) - B_0)(\tilde{x}_N^{(2)}(u) - \tilde{\xi}_N^{(2)}(u))|\,\mathrm{d}u \\ &+ \left|\tilde{m}_N^{(2)}(t) + B_0\int_0^t \mathrm{e}^{B_0(t-u)}\tilde{m}_N^{(1)}(u)\,\mathrm{d}u\right|,\end{aligned}$$

giving, with $\eta^*_{N1} := (1 + \gamma_2\|B_0\|/\beta_0)\eta_{N1}$, and from Lemma 2.1 and (4.6),

$$\begin{aligned}d_N^{(2)}(t) \le\, &\gamma_2\int_0^t d_N^{(2)}(u)\|B(\tilde{x}_N(u)) - B(\tilde{\xi}_N(u))\|\,\mathrm{d}u \\ &+ \gamma_2\int_0^t k^{(2)}\varepsilon_N\|B(\tilde{x}_N(u)) - B(\tilde{\xi}_N(u))\|\,\mathrm{d}u \\ &+ \gamma_2\int_0^t d_N^{(2)}(u)\|B(\tilde{\xi}_N(u)) - B_0\|\,\mathrm{d}u + \eta^*_{N1} \\ \le\, &k_2\left\{(\delta_1 + |x_{1N}^{(1)} - x_0^{(1)}|)\int_0^t d_N^{(2)}(u)\,\mathrm{d}u + \varepsilon_N\int_0^t d_N^{(1)}(u)\,\mathrm{d}u\right\} + \eta^*_{N1}\end{aligned} \tag{4.8}$$

for a suitable constant k_2. From (4.7), we have

$$\int_0^t d_N^{(1)}(u)\,\mathrm{d}u \le k_3 \mathrm{e}^{\beta_0 t} \int_0^t d_N^{(2)}(u)\,\mathrm{d}u + 2t\eta'_{N1}$$

and, substituting this into (4.8), we obtain

$$d_N^{(2)}(t) \le k_4(\delta_1 + |x_{1N}^{(1)} - x_0^{(1)}|) \int_0^t d_N^{(2)}(u)\,\mathrm{d}u + \eta^*_{N1} + k_5 \varepsilon_N \log\biggl(\frac{1}{\varepsilon_N}\biggr)\eta'_{N1}$$

for constants k_4, k_5, and for $t \le t_0(\delta_1, \varepsilon_N)$. Gronwall's inequality now yields

$$d_N^{(2)}(t) \le k_6 \eta_{N1} \exp\{k_4 t(\delta_1 + |x_{1N}^{(1)} - x_0^{(1)}|)\}$$

for suitable k_6. For $t = t_0(\delta_1, \varepsilon_N)$, the right-hand side can be made to be of order $O(N^{-1/2+\chi})$ for any $\chi > 0$ by choosing $\delta_1 = \delta_1(\chi)$ small enough. In particular, choosing $t = t_0(\delta_1(\chi), \varepsilon_N)$ and recalling (4.1), we have

$$\sup_{0\le u\le t} |\tilde{x}_N^{(2)}(u) - \tilde{\xi}_N^{(2)}(u)| \le \mathrm{e}^{\beta_0 t} d_N^{(2)}(t) = O(\varepsilon_N^{-1} N^{-1/2+\chi}) = O(N^{-1/12+\chi}) \tag{4.9}$$

on the event E_N, and also, in view of (1.4) and the third inequality in Lemma 2.1,

$$\sup_{0\le u\le t} \biggl\{ \frac{|\tilde{x}_N^{(2)}(u) - \tilde{\xi}_N^{(2)}(u)|}{|\tilde{\xi}_N^{(2)}(u)|} \biggr\} \le \varepsilon_N^{-1} d_N^{(2)}(t) = O(N^{-1/12+\chi}).$$

In addition, from (4.7) and (4.9), it follows that on the event E_N,

$$\sup_{0\le u\le t} |\tilde{x}_N^{(1)}(u) - \tilde{\xi}_N^{(1)}(u)| =: d_N^{(1)}(t) \le k_7\{\mathrm{e}^{\beta_0 t} d_N^{(2)}(t) + \eta'_{N1}\} = O(N^{-1/12+\chi}). \tag{4.10}$$

We now compare the assumed conditions (4.5) and (4.6), involving bounds on increasing processes with jumps bounded by $\varphi = N^{-1} \max_{J\in\mathcal{J}} |J|$ with the resulting estimates (4.9) and (4.10). It then follows immediately from Lemma D.1 of the online appendix (see Barbour *et al.* (2014)) that both (4.9) and (4.10) hold on E_N for all $t \le t_0(\delta_1(\chi), \varepsilon_N)$ and for all sufficiently large N, provided that $\chi < \frac{1}{12}$.

It remains to observe that the solution $\tilde{\xi}_N$ of the deterministic equations starting from $\tilde{\xi}_0 := x_1 = x_N(\tau^x_{N*})$ is close to the solution $\hat{\xi}_N$ starting from $\hat{\xi}_N(0) = \xi_N(t^\xi_{N*})$, up to the time $t_0(\delta_1, \varepsilon_N)$, because their starting points are close enough on $\widehat{E}_N$, as was shown in Proposition 3.1. From the final statements of Lemma 2.1, taking $\varepsilon^{(1)} = \bar{k}^{(1)} N^{-5/12}\log N$ and $\varepsilon^{(2)} = \bar{k}^{(2)} N^{-5/12}$, it follows that, for some $\gamma > 0$,

$$\sup_{0\le u\le t_0(\delta_1,\varepsilon^{(2)})} |\hat{\xi}_N^{(1)}(u) - \tilde{\xi}_N^{(1)}(u)| = O(N^{-\gamma}), \qquad \sup_{0\le u\le t_0(\delta_1,\varepsilon^{(2)})} |\hat{\xi}_N^{(2)}(u) - \tilde{\xi}_N^{(2)}(u)| = O(N^{-\gamma}).$$

One final result is needed, to show that continuation using Kurtz (1970, Theorem (3.1)) represents following the deterministic path along an asymptotically nondegenerate path. The proof is given in Section C of the online appendix; see Barbour *et al.* (2014).

https://doi.org/10.1239/aap/1449859806 Published online by Cambridge University Press

Lemma 4.2. *Define*

$$\hat{t}_N := t^{\xi}_{N*} + t_0(\delta', \varepsilon_N) = \beta_0^{-1}\left\{\log N + \log\left(\frac{\delta'}{\boldsymbol{v}^\top Z_0}\right)\right\}.$$

Then, for suitably chosen $\delta' \le \delta_1$, all the components of $\tilde{\xi}_N(t_0(\delta', \varepsilon^{(2)})) = \xi_N(\hat{t}_N)$ are uniformly bounded away from 0 for all large enough N.

We summarize the results of this section in the following proposition, which, with Proposition 3.1, completes the proof of Theorem 1.1. Theorem 1.2 is proved in Section A of the online appendix; see Barbour *et al.* (2014).

Proposition 4.1. *Let ξ_N denote the solution to the deterministic equation starting with $\xi_N(0) = x_{N,0}$ satisfying $|x^{(1)}_{N,0} - x^{(1)}_0| \le N^{-5/12}$ and $x^{(2)}_{N,0} = N^{-1}Z_0$. Let $\varepsilon_N \asymp N^{-5/12}$ be as defined in (4.1), $t_0(\delta, \varepsilon)$ as in (2.3), and τ^x_{N*} and t^{ξ}_{N*} as in (1.10). Then there exist $\delta' > 0$ and an event E_N, whose complement has asymptotically negligible probability, such that, on $E_N \cap \{\tau^x_{N*} < \infty\}$ and for all large enough N,*

$$\sup_{0 \le t \le t_0(\delta', \varepsilon_N)} |x_N(\tau^x_{N*} + t) - \xi_N(t^{\xi}_{N*} + t)| \le k(\delta')N^{-\gamma}$$

for some $\gamma > 0$ and $0 < k(\delta') < \infty$, and that all components of $\xi_N(t^{\xi}_{N*} + t_0(\delta', \varepsilon_N))$ are bounded uniformly away from 0. Note also that

$$t^{\xi}_{N*} + t_0(\delta', \varepsilon_N) = \beta_0^{-1}\{\tfrac{7}{12}\log N - \log(\boldsymbol{v}^\top Z_0) + \log\delta' - \log\varepsilon_N\} = \beta_0^{-1}\log N + O(1).$$

Acknowledgements

ADB was supported in part by the Australian Research Council (grant nos. DP120102728 and DP120102398), KH and FK by the Australian Research Council (grant no. DP120102728), and HK by the Israel Science Foundation (grant no. 764/13) and by the Milford Bohm Memorial Grant. The authors wish to warmly thank the anonymous referees of an earlier version, whose contributions have substantially improved this paper.

References

Athreya, K. B. and Ney, P. E. (1972). *Branching Processes.* Springer, New York.

Barbour, A. D. (1975). The duration of the closed stochastic epidemic. *Biometrika* **62,** 477–482.

Barbour, A. D. (1980). Density dependent Markov population processes. In *Biological Growth and Spread* (Lecture Notes Biomath. **38**). Springer, Berlin, pp. 36–49.

Barbour, A. D. and Pollett, P. K. (2012). Total variation approximation for quasi-equilibrium distributions, II. *Stoch. Process Appl.* **122,** 3740–3756.

Barbour, A. D., Hamza, K., Kaspi, H. and Klebaner, F. (2014). Escape from the boundary in Markov population processes. Preprint. Available at http://arxiv.org/abs/1312.5788v3.

Bartlett, M. S. (1949). Some evolutionary stochastic processes. *J. R. Statist. Soc.* **11,** 211–229.

Heinzmann, D. (2009). Extinction times in multitype Markov branching processes. *J. Appl. Prob.* **46,** 296–307.

Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. IV. University of California Press, Berkeley, pp. 149–165.

Klebaner, F. C., Sagitov, S., Vatutin, V. A., Haccou, P. and Jagers, P. (2011). Stochasticity in the adaptive dynamics of evolution: the bare bones. *J. Biol. Dyn.* **5,** 147–162.

Kurtz, T. G. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Prob.* **7,** 49–58.

KURTZ, T. G. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *J. Appl. Prob.* **8,** 344–356.

KURTZ, T. G. (1978). Strong approximation theorems for density dependent Markov chains. *Stoch. Process. Appl.* **6,** 223–240.

SENETA, E. (2006). *Non-negative Matrices and Markov Chains*. Springer, New York.

THORISSON, H. (2000). *Coupling, Stationarity, and Regeneration.* Springer, New York.

WHITTLE, P. (1955). The outcome of a stochastic epidemic—a note on Bailey's paper. *Biometrika* **42,** 116–122.