Laws of large numbers for epidemic models with countably many types

A. D. Barbour* and M. J. Luczak†
Universität Zürich and London School of Economics

CDAM Research report LSE-CDAM-2006-14

Abstract

In modelling parasitic diseases, it is natural to distinguish hosts according to the number of parasites that they carry, leading to a countably infinite type space. Proving the analogue of the deterministic equations, used in models with finitely many types as a 'law of large numbers' approximation to the underlying stochastic model, has previously either been done case by case, using some special structure, or else not attempted. In this paper, we prove a general theorem of this sort, and complement it with a rate of convergence in the ℓ_1-norm.

Keywords: Epidemic models, infinitely many types, quantitative law of large numbers
AMS subject classification: 92D30, 60J27, 60B12
Running head: Epidemics with countably many types

1 Introduction

This paper is concerned with generalizations of the stochastic models introduced in Barbour & Kafetzaki (1993) and developed in Luchsinger (2001a,b), which describe the spread of a parasitic disease. With such diseases, it is natural to distinguish hosts according to the number of parasites that they carry. Since it is not usually possible to

*Angewandte Mathematik, Universität Zürich, Winterthurertrasse 190, CH-8057 ZÜRICH; ADB was supported in part by Schweizerischer Nationalfonds Projekt Nr. 20–107935/1.

†London School of Economics; MJL was supported in part by the Nuffield Foundation.

prescribe a fixed upper limit for the parasite load, this leads to models with countably infinitely many types, one for each possible number of parasites. The model considered by Kretzschmar (1993) is also of this kind, though framed in a deterministic, rather than a stochastic form. Then there are models arising in cell biology, in which, for instance, hosts may be replaced by cells which are distinguished according to the number copies of a particular gene that they carry, a number which is again in principle unlimited; see Kimmel & Axelrod (2002, Chapter 7) for a selection of branching process examples. The metapopulation model of Arrigoni (2003) also allows for infinitely many types of patches, here distinguished by the size of the population in the patch.

The fact that there are infinitely many types can cause difficulty in problems which, for finitely many types, would be quite standard. To take a well known example, in a super-critical Galton–Watson branching process with finitely many types, whose mean matrix is irreducible and aperiodic and whose offspring distributions have finite variance, the proportions of individuals of the different types converge to fixed values if the process grows to infinity: a rather stronger result is to be found in Kesten & Stigum (1966). If there are infinitely many types, little is generally known about the asymptotics of the proportions, except when the mean matrix is r-positive, a condition which is automatically satisfied in the finite case; here, Moy (1967) was able to prove convergence under a finite variance condition. For epidemic models analogous to those above, but with only finitely many types, there are typically 'law of large numbers' approximations, which hold in the limit of large populations, and are expressed in the form of systems of differential equations: see, for example, Bailey (1968) or Kurtz (1980). Proving such limits for models with infinite numbers of types is much more delicate. Kretzschmar (1993) begins with the system of differential equations as the model, and so does not consider the question; in Barbour & Kafetzaki (1993) and Luchsinger (2001a,b), the arguments are involved, and make use of special assumptions about the detailed form of the transition rates.

In this paper, we show that a law of large numbers can be established in substantial generality. The models that we allow are constructed by superimposing state-dependent transitions upon a process with otherwise independent and well-behaved dynamics within the individuals; the state-dependent components are required to satisfy certain Lipschitz and growth conditions, to ensure that the perturbation of the underlying semi-group governing the independent dynamics is not too severe. The main approximation is stated in Theorem 4.1, and bounds the difference between the normalized process $N^{-1}X_N$ and a deterministic trajectory x with respect to the ℓ_1-norm, uniformly on finite time intervals. The theorem is sufficiently general to cover all the epidemic models mentioned above, except for that of Kretzschmar (1993), where only a version with a truncated infection rate can be treated. The proof is by way of an intermediate approximation, based on a system $\widetilde{X}_N$ consisting of independent particles, which has dynamics reflecting the average behaviour of X_N. The deterministic trajectory x is discussed in Section 3, the approximation of $N^{-1}\widetilde{X}_N$ by x in Section 4, culminating in Theorem 4.6, and the final approximation of $N^{-1}X_N$ by x in Section 5.

2 Specifying the model

Our model is expressed in terms of a sequence of processes X_N having state space $\mathcal{X} := \{X \in \mathbb{Z}_+^\infty\colon \sum_{i\geq 0} X^i < \infty\}$, where $\mathbb{Z}_+$ denotes the non-negative integers and X^i the i-th component of X. The process $X_N(t) := (X_N^j(t) : j \in \mathbb{Z}_+)$, $t \geq 0$, has $\sum_{j\geq 0} X_N^j(0) = N$, and evolves as a pure jump Markov process with transition rates given by

$$\begin{aligned}
X &\rightarrow X + e(j) - e(i) \text{ at rate } X^i\{\mu(i,j) + \alpha_{ij}(N^{-1}X)\}, \quad i \geq 0,\ j \geq 0,\ j \neq i;\\
X &\rightarrow X + e(i) \text{ at rate } N\beta_i(N^{-1}X), \quad i \geq 0;\\
X &\rightarrow X - e(i) \text{ at rate } X^i\{\bar{\delta}_i + \delta_i(N^{-1}X)\}, \quad i \geq 0,
\end{aligned}$$

where the non-negative quantities $\mu(i,j), \alpha_{ij}, \beta_i, \bar{\delta}_i$ and δ_i are used to model different aspects of the underlying parasite life cycle, and $e(i)$ denotes the ith coordinate vector.

We interpret $X_N^i(t)$ as the number of hosts who carry i parasites at time t. The first terms in the first and third transitions represent parasite communities developing independently within different hosts, according to a pure jump Markov process, which includes the possibility of host death at rate $\bar{\delta}_i$ if its parasite load is i. Note that the elements $\mu(0,\cdot)$ are all zero if only parasite mortality and reproduction are being modelled by the $\mu(i,j)$. However, we may also allow them to include a part of the force of infection, such as infection arising from external sources, that does not change with varying levels $x := N^{-1}X_N$ of infection among the host population. Hence it may be the case that elements $\mu(0,\cdot)$ could be positive.

The second term $\alpha_{ij}(x)$ in the first transition is usually the main infection term, allowing hosts to acquire further parasites at rates which can vary, depending on the levels of infection in the entire host population. However, the $\alpha_{ij}(x)$ can also be used to model state dependent loss of infection, as, for example, through treatment being offered when high levels of infection are observed, so that it is not necessarily the case that $\alpha_{ij}(x) = 0$ when $j < i$. For convenience, we define $\alpha_{ii}(x) = 0$ for all i. The remaining transitions allow one to model births, deaths and immigration of hosts, in the latter case possibly themselves infective, and they are all formulated so that dependence on the detailed levels of infection in the host population is allowed.

The components of the transitions rates will be required to satisfy a number of conditions. First, we address the $\mu(i,j)$ and $\bar{\delta}_i$. Letting Δ denote an absorbing 'cemetery' state, reached if a host dies, set

$$\mu(i,\Delta) \;:=\; \bar{\delta}_i, \quad \mu(i,i) \;:=\; -\mu(i) - \bar{\delta}_i, \quad i \geq 0, \tag{2.1}$$

where $\mu(i) := \sum_{j\geq 0, j\neq i} \mu(i,j)$. Then $\boldsymbol{\mu}$ is the infinitesimal matrix of a time homogeneous pure jump Markov process W on $\mathbb{Z}_+ \cup \Delta$. Writing

$$p_{ij}(t) \;:=\; \mathbf{P}[W(t) = j \,|\, W(0) = i],$$

for $i \geq 0$ and $j \in \mathbb{Z} \cup \Delta$, we shall assume that $\boldsymbol{\mu}$ is such that W is non-explosive and that

$$\mathbf{E}_i^0\{(W(t)+1)\} \;=\; \sum_{j\geq 0} (j+1) p_{ij}(t) \;\leq\; (i+1)e^{wt}, \quad i \geq 0, \tag{2.2}$$

for some $w \geq 0$, where we use the notation

$$\mathbf{E}_i^0(f(W(t))) \;:=\; \mathbf{E}\{f(W(t)) I[W(t) \notin \Delta] \,|\, W(0) = i\}.$$

We shall further require that, for some $1 \leq m_1, m_2 < \infty$,

$$\mu(i) + \bar{\delta}_i \;\leq\; m_1 (i+1)^{m_2} \quad \text{for all} \quad i \geq 0, \tag{2.3}$$

and also that, for each $j \geq 0$,

$$\limsup_{l\to\infty} \mu(l,j) \;<\; \infty. \tag{2.4}$$

The remaining elements depend on the state of the system through the argument $x := N^{-1}X$. In the random model, $x \in N^{-1}\mathcal{X}$ has only finitely many non-zero elements, but when passing to a law of large numbers approximation, this need not be appropriate in the limit. We shall instead work within the larger space $\ell_{11} := \{x \in \mathbb{R}^\infty \colon \sum_{i\geq 0}(i+1)|x^i| < \infty\}$

endowed with the norm $\|x\|_{11} := \sum_{i\ge 0}(i+1)|x^i|$. We then assume that α_{il}, β_i and δ_i are all locally ℓ_1-Lipschitz, and satisfy some extra conditions.

First, for $i \ge 0$ and $x, y \in \ell_{11}$, we assume that

$$\sum_{l\ge 0} \alpha_{il}(0) \;\le\; a_{00}, \tag{2.5}$$

$$\sum_{l\ge 0} |\alpha_{il}(x) - \alpha_{il}(y)| \;\le\; a_{01}(x,y)\|x-y\|_1, \tag{2.6}$$

$$\sum_{l\ge 0}(l+1)\alpha_{il}(0) \;\le\; (i+1)a_{10}, \tag{2.7}$$

$$\sum_{l\ge 0}(l+1)|\alpha_{il}(x) - \alpha_{il}(y)| \;\le\; (i+1)a_{11}(x,y)\|x-y\|_{11}, \tag{2.8}$$

where the a_{r0} are finite,

$$a_{r1}(x,y) \;\le\; \tilde{a}_{r1}(\|x\|_{11} \wedge \|y\|_{11}), \qquad r = 0,1,$$

and the $\tilde{a}_{r1}$ are bounded on bounded intervals; as usual, $\|x\|_1 := \sum_{i\ge 0}|x^i|$.

Then we assume that, for all $x, y \in \ell_{11}$,

$$\sum_{i\ge 0} \beta_i(0) \;\le\; b_{00}, \tag{2.9}$$

$$\sum_{i\ge 0} |\beta_i(x) - \beta_i(y)| \;\le\; b_{01}(x,y)\|x-y\|_1, \tag{2.10}$$

$$\sum_{i\ge 0}(i+1)\beta_i(0) \;\le\; b_{10}, \tag{2.11}$$

$$\sum_{i\ge 0}(i+1)|\beta_i(x) - \beta_i(y)| \;\le\; b_{11}(x,y)\|x-y\|_{11}, \tag{2.12}$$

where the b_{r0} are finite,

$$b_{r1}(x,y) \;\le\; \tilde{b}_{r1}(\|x\|_{11} \wedge \|y\|_{11}), \qquad r = 0,1,$$

and the $\tilde{b}_{r1}$ are bounded on bounded intervals; and finally that

$$\sup_{i\ge 0} \delta_i(0) \;\le\; d_0, \tag{2.13}$$

$$\sup_{i\ge 0} |\delta_i(x) - \delta_i(y)| \;\le\; d_1(x,y)\|x-y\|_1, \tag{2.14}$$

where d_0 is finite and

$$d_1(x,y) \;\le\; \tilde{d}_1(\|x\|_{11} \wedge \|y\|_{11}),$$

with $\tilde{d}_1$ bounded on finite intervals.

The various assumptions can be understood in the biological context. First, the two norms $\|\cdot\|_1$ and $\|\cdot\|_{11}$ have natural interpretations. The quantity $\|X-Y\|_1$ is the sum of the differences $|X^i - Y^i|$ between the numbers of hosts in states $i = 0, 1, 2, \ldots$ in two host populations X and Y; this can be thought of as the natural measure of difference as seen from the hosts' point of view. The corresponding 'parasite norm' is then $\|X-Y\|_{11}$, which weights each difference $|X^i - Y^i|$ by the factor $(i+1)$, the number of parasites plus one; in a similar way, writing $x = N^{-1}X$, one can interpret $\|x\|_{11}$ as a measure of 'parasite density'.

The simplest conditions are (2.5), (2.9) and (2.13), which, together with conditions (2.6), (2.10) and (2.14) with $y = 0$, ensure that the per capita infection, birth, immigration and death rates are all finite, and bounded by constant multiples of $\|x\|_1 + 1$. This for instance immediately excludes any model in which the per capita infection rate is a constant K times the parasite density $\|x\|_{11}$, and Kretzschmar's (1993) model is excluded for the same reason. Analogously, conditions (2.6), (2.10) and (2.14) for general y imply that cumulative differences in the above rates between population infection states x and y are limited by multiples of the host norm $\|x-y\|_1$ of the difference between x and y, and also that these multiples remain bounded provided that the *smaller* of $\|x\|_{11}$ and $\|y\|_{11}$ remains bounded.

The remaining conditions concern parasite weighted analogues of the preceding conditions. Conditions (2.11) and (2.12) constrain the overall rate of flow of parasites into the system through immigration to be finite, and bounded if the parasite density remains bounded; condition (2.12) also limits the way in which this influx may depend on the infection state. Conditions (2.7) and (2.8) impose analogous restrictions on the rates of influx of parasites into hosts through infection. Here, the limitations are imposed on the *multiplicative* rate of increase of parasites in a host, and may be useful for modelling systems in which parasites can directly reproduce in their hosts, and where this rate of reproduction can be influenced by an immune response to external parasite challenge.

In all, these assumptions are relatively mild. We now show that they cover the *stochastic non-linear model* introduced in Barbour & Kafetzaki (1993) and the *stochastic linear model* from Barbour (1994), both of which were generalized and studied in depth in Luchsinger (1999, 2001a,b), but not that of Kretzschmar (1993).

In Luchsinger's non-linear model, the total population size is fixed at N at all times, with $\beta_i(x) = \delta_i(x) = \bar{\delta}_i = 0$ for all $i \geq 0$ and $x \in \ell_{11}$. The matrix $\boldsymbol{\mu}$ is the superposition of the infinitesimal matrix of a pure death process with rate $\mu > 0$ and the catastrophe process which jumps from any state to 0 at constant rate $\kappa \geq 0$. The first of the above expresses the assumption that parasites die independently at rate μ. The second corresponds to the fact that hosts die independently at rate κ, and their parasites with them; whenever a host dies, it is instantly replaced by a healthy individual. Thus the positive elements of $\boldsymbol{\mu}$ are given by

$$\mu(i,i-1) = i\mu,\ \mu(i,0) = \kappa, \quad i \geq 2; \qquad \mu(1,0) = \mu + \kappa,$$

and $\mu(i,i)$, $i \geq 1$, is determined by (2.1). The elements $\mu(0,j)$ are all zero. As regards infection, hosts make potentially infectious contacts at rate $\lambda > 0$, and infection can only occur in a currently uninfected host. If a host carrying i parasites contacts a healthy one, infection with l parasites is developed by the healthy host with probability p_{il}, where $\sum_{l \geq 0} p_{il} = 1$ for all i and $p_{00} = 1$. Here, the distribution $F_i = (p_{il}, l \geq 0)$ is the i-fold convolution of F_1, modelling the assumption that, at such a contact, the parasites act independently in transmitting offspring to the previously healthy host. These rules are incorporated by taking

$$\alpha_{0l}(x) \;=\; \lambda \sum_{i \geq 1} x^i p_{il}, \quad l \geq 1,\ x \in \ell_{11},$$

and the remaining $\alpha_{il}(x)$ are all zero. Thus, for $z \geq 0$, we can take

$$a_{00} = a_{10} = 0,$$

$$\tilde{a}_{01}(z) \;=\; \lambda, \qquad \tilde{a}_{11}(z) \;=\; \lambda \max\{\theta, 1\},$$

where θ is the mean of F_1, the mean number of offspring transmitted by a parasite during an infectious contact: thus $\sum_{l \geq 0} p_{il}(l+1) = i\theta + 1$. In Luchsinger's linear model, there is tacitly assumed to be an *infinite* pool of potential infectives, so that the 0-coordinate is not required, and its value may if desired be set to 0; the population of interest consists of the infected hosts, whose number may vary. The matrix $\boldsymbol{\mu}$ is the infinitesimal matrix of a simple death process with rate $\mu > 0$, but now restricted to the reduced state space, giving the positive elements

$$\mu(i,i-1) \;=\; i\mu, \quad i \geq 2;$$

hosts losing infection are now incorporated by using the $\bar{\delta}_i$, with

$$\bar{\delta}_i \;=\; \kappa, \quad i \geq 2; \qquad \bar{\delta}_1 \;=\; \kappa + \mu,$$

again with $\mu(i,i)$, $i \geq 2$, determined by (2.1). Only a member of the pool of uninfected individuals can be infected, and infections with i parasites occur at a rate $\lambda \sum_{l\geq 1} X^l p_{li}$, so that we have

$$\beta_i(x) \;=\; \lambda \sum_{l\geq 1} x^l p_{li}, \quad i \geq 1,$$

with all the $\alpha_{il}(x)$ and $\delta_i(x)$ equal to zero. Here, for $z > 0$, we can take $b_{00} = b_{10} = 0$ and

$$\tilde{b}_{01}(z) \;=\; \lambda, \qquad \tilde{b}_{11}(z) \;=\; \lambda \max\{\theta, 1\}.$$

Both models exhibit an epidemic threshold behaviour as the parameter θ increases, but its form is rather unexpected. For instance, if $\kappa = 0$ and $\mu/\lambda \leq e$, the critical value of θ is μ/λ. This is in no way surprising, since the mean number of offspring of a parasite over its whole lifetime is just $R_0 := \lambda\theta/\mu$, and the usual branching process heuristic would suggest that the critical value of R_0 should be 1. However, if $\mu/\lambda > e$, the critical value of θ is instead $e^{\mu/e\lambda}$; see Luchsinger (2001a,b) for further details of the still more complicated behaviour when $\kappa > 0$.

In Kretzschmar's model, infection only takes place a single parasite at a time, but at a complicated state dependent rate. Mortality of parasites is modelled as before, and hosts can both die and give birth; here, death rates increase with parasite load, and birth rates decrease. In our formulation, the positive elements of $\boldsymbol{\mu}$ are given by

$$\mu(i, i-1) \;=\; i\mu, \quad i \geq 1,$$

and we take

$$\bar{\delta}_i \;=\; \kappa + i\alpha, \qquad \delta_i(x) = 0,$$

for non-negative constants κ and α; again, the $\mu(i,i)$ are determined by (2.1). The $\alpha_{il}(x)$ are all zero except for $l = i+1$, when

$$\alpha_{i,i+1}(x) \;=\; \lambda \sum_{j\geq 1} j x^j \Big/ \Big\{ c + \sum_{j\geq 0} x^j \Big\}, \quad i \geq 0;$$

hosts are born free of parasites, so that $\beta_i(x) = 0$ for $i \geq 1$, and

$$\beta_0(x) \;=\; \beta \sum_{i\geq 0} x^i \xi^i,$$

for some $\beta > 0$ and $0 \leq \xi \leq 1$. Here, we can take $b_{00} = b_{10} = 0$, $b_{01}(z) = b_{11}(z) = \beta$; but, for $c > 0$, we cannot improve substantially on the choices

$$a_{00} \;=\; a_{10} \;=\; 0$$
$$a_{01}(x,y) \;=\; a_{11}(x,y) \;=\; \frac{\lambda}{c}\Big\{|\|x\|_{11} - \|y\|_{11}| + c^{-1}(\|x\|_{11} \wedge \|y\|_{11})|\|x\|_1 - \|y\|_1|\Big\},$$

and these do not yield a suitable bound for $a_{01}(x,y)$. If, however, we take instead

$$\alpha_{i,i+1}(x) \;=\; \lambda \sum_{j\geq 1}(j \wedge M)x^j \Big/ \Big\{c + \sum_{j\geq 0} x^j\Big\}, \quad i \geq 0, \tag{2.15}$$

for any $M < \infty$, there is no problem in satisfying our conditions.

Remark. The assumptions made about the $\alpha_{ij}(x)$ and $\beta_i(x)$ have certain general consequences. One is that the total number of hosts has to be finite almost surely for all t. This can be seen by comparison with a pure birth process, since the number of hosts $\|X\|_1$ only increases through immigration, and the total rate of immigration $N\sum_{i\geq 0}\beta_i(N^{-1}X)$ does not exceed $Nb_{00} + \tilde{b}_{01}(0)\|X\|_1$. Hence, for any $T > 0$,

$$\lim_{M\to\infty} \mathbf{P}[\sup_{0\leq t\leq T} \|X_N(t)\|_1 > NM] \;=\; 0. \tag{2.16}$$

Now, if $N^{-1}\|X\|_1 \leq M$, it follows from (2.5) and (2.6) that

$$\sum_{l\geq 0} \alpha_{il}(N^{-1}X) \;\leq\; a_{00} + \tilde{a}_{01}(0)M,$$

for all $i \geq 0$. Hence, and because W is non-explosive, it follows that, on the event $\{\sup_{0\leq t\leq T}\|X_N(t)\|_1 \leq NM\}$, the X-chain makes a.s. only finitely many transitions in $[0,T]$. Letting $M \to \infty$, it follows from (2.16) that a.s. only finitely many transitions can occur in the X-chain in any finite time interval.

3 The differential equations

We assume deterministic initial conditions $X_N(0)$ for each N. Our aim is to approximate the evolution of the process $N^{-1}X_N(t)$ when N is large. A natural candidate approximation is given by the solution to the 'average drift' infinite dimensional differential equation

$$\frac{dx^i(t)}{dt} = \sum_{l\geq 0} x^l(t)\mu(l,i) + \sum_{l\neq i} x^l(t)\alpha_{li}(x(t)) - x^i(t)\sum_{l\neq i}\alpha_{il}(x(t)) + \beta_i(x(t)) - x^i(t)\delta_i(x(t)), \quad i \geq 0, \tag{3.1}$$

with initial condition $x(0) = N^{-1}X_N(0)$; for the time being, we suppress the N-dependence in x. We shall see shortly that this differential equation has a unique 'mild' solution $x = (x^i(t),\, i \geq 0)$ in some bounded t-interval (see (3.6) below). We are then able to show that x is in fact a 'classical' solution to (3.1), in the usual sense.

Note that, in contrast to the stochastic model, it is not perhaps immediate that a solution $x(t)$ to these equations has to belong to the nonnegative cone. To accommodate this possibility, we extend the definitions of α_{il}, β_i and δ_i, setting

$$\alpha_{il}(x) = \alpha_{il}(x_+), \quad \beta_i(x) = \beta_i(x_+), \quad \delta_i(x) = \delta_i(x_+),$$

where $x^i_+ := \max(x^i, 0)$, $i \geq 0$, and observing that conditions (2.5) – (2.14) are still satisfied, with x and y replaced by x_+ and y_+, respectively, as arguments of $a_{01}, a_{11}, b_{01}, b_{11}$ and d_1.

Equation (3.1), as in Arrigoni (2003), can be compactly expressed in the form

$$\frac{dx}{dt} = Ax + F(x), \qquad x(0) = N^{-1}X_N(0), \tag{3.2}$$

where A is a linear operator given by

$$(Ax)^i = \sum_{l\geq 0} x^l\mu(l,i), \quad i \geq 0, \tag{3.3}$$

and F is an operator given by

$$F(x)^i = \sum_{l\neq i} x^l\alpha_{li}(x) - x^i\sum_{l\neq i}\alpha_{il}(x) + \beta_i(x) - x^i\delta_i(x), \quad i \geq 0. \tag{3.4}$$

This representation is crucial to establishing the existence and uniqueness of the solution x.

The matrix of the adjoint A^* of A is easily seen to be the infinitesimal matrix $\boldsymbol{\mu}$, so that A^* is the generator of a non-explosive pure-jump Markov chain on the countable set $\mathbb{Z}_+ \cup \Delta$. Standard theory implies that A^* corresponds to a Feller, and hence strongly continuous (Kallenberg 2002, Theorem 19.6) semigroup on the Banach space $C_0(\mathbb{Z}_+)$ of all sequences $x = (x^i : i \in \mathbb{Z}_+)$ such that $x^i \to 0$ as $i \to \infty$, endowed with supremum norm, that is $\| x \|_\infty = \sup_i |x^i|$. See, for example, Kallenberg (2002, Theorem 12.22 and Proposition 19.2, together with the remarks following the theorem, and the definition of a Feller semigroup on p. 369). See also Ethier and Kurtz (1986, Chapter 4, Theorem 2.2) or Norris (1997, Section 2.8).

We may now apply the following theorem.

Theorem 3.1 (Pazy 1983, Theorem 10.4, Chapter 1) *Let $T(t)$ be a C_0 (i.e. strongly continuous) semigroup on S with infinitesimal generator A and let $T(t)^*$ be its adjoint semigroup. If A^* is the adjoint of A and $\overline{D(A^*)}$ is the closure of $D(A^*)$ in the dual space S^*, then the restriction $T(t)^+$ of $T(t)^*$ to $\overline{D(A^*)}$ is a C_0 semigroup on $\overline{D(A^*)}$. The infinitesimal generator A^+ of $T(t)^+$ is the part of A^* in $\overline{D(A^*)}$.*

In the above theorem, given a linear operator in L and a subspace V of S, the *part of L in V* is the operator $\tilde{L}$ defined by

$$D(\tilde{L}) = \{x \in D(L) \cap V : Lx \in V\}.$$

In our case, $S = C_0(\mathbb{Z}_+)$ and its dual S^* can be identified with the space ℓ_1. For the adjoint $\boldsymbol{\mu}^*$ of the infinitesimal matrix $\boldsymbol{\mu}$, the closure of $\overline{D(\boldsymbol{\mu}^*)}$ is precisely equal to S^*. This is because every sequence x such that $\|x\|_1 < \infty$ can be approximated by sequences with bounded support, which are all in the domain of $\boldsymbol{\mu}^*$ since $\mu(i) < \infty$ for each i, so $D(A)$ is dense in S^*. Further, it is easily checked that the domain of the infinitesimal generator A^+ contains the set of all sequences $x \in S^*$ such that

$$|x^0| + \sum_{i\geq 1}(\mu(i) + \bar{\delta}_i)|x^i| < \infty.$$

It follows that our operator A generates a strongly continuous semigroup on the Banach space ℓ_1 of real-valued sequences.

We shall further show that A in fact also generates a strongly continuous semigroup $T(t)$ on the Banach subspace ℓ_{11} of ℓ_1. For this note that again every sequence x such that $\|x\|_{11} < \infty$ can be approximated by sequences with bounded support, which are all in the domain of $\boldsymbol{\mu}^*$, so $D(A)$ is dense in ℓ_{11}. We need to check that $T(t)\ell_{11} \subseteq \ell_{11}$, which is equivalent to checking that for every sequence x with $\|x\|_{11} < \infty$, we have $\|T(t)x\|_{11} < \infty$. Recalling that A corresponds to the adjoint of the semigroup $P = (p_{ij}(t))$ associated with W, we must show that for every x with $\|x\|_{11} < \infty$, we have $\|xP(t)\|_{11} < \infty$ for all times t; but this follows from (2.2), since

$$\|xP(t)\|_{11} \;=\; \sum_{j\ge 0}(j{+}1)\Bigl|\sum_{i\ge 0} x^i p_{ij}(t)\Bigr| \;\le\; \sum_{i\ge 0}|x^i|(i{+}1)e^{wt} \;=\; e^{wt}\|x\|_{11} \;<\; \infty. \tag{3.5}$$

Strong continuity also follows, since, for $x \in \ell_{11}$,

$$\begin{aligned}
&\lim_{t\downarrow 0}\|xP(t)-x\|_{11}\\
&\quad\le\; \lim_{t\downarrow 0}\sum_{j\ge 0}(j+1)\left\{\sum_{i\ne j}|x^i|p_{ij}(t)+|x^j|(1-p_{jj}(t))\right\}\\
&\quad= \lim_{t\downarrow 0}\left\{\sum_{i\ge 0}|x^i|\{\mathbf{E}_i^0(W(t)+1)-(i+1)\mathbf{P}_i[W(t)=i]\}\right.\\
&\qquad\qquad\left. +\sum_{j\ge 0}(j+1)|x^j|(1-p_{jj}(t))\right\}\\
&\quad\to\; 0,
\end{aligned}$$

by dominated convergence, since $\|x\|_{11} < \infty$ and, by (2.2),

$$0 \;\le\; \mathbf{E}_i^0(W(t)+1)-(i+1)\mathbf{P}_i[W(t)=i] \;\le\; (i+1)e^{wt},$$

and finally, again by (2.2),

$$\begin{aligned}
\limsup_{t\downarrow 0}\mathbf{E}_i^0(W(t)+1) \;&\le\; \lim_{t\downarrow 0}(i+1)e^{wt} \;=\; i+1;\\
\liminf_{t\downarrow 0}\mathbf{E}_i^0(W(t)+1) \;&\ge\; \liminf_{t\downarrow 0}\sum_{j=0}^{i} p_{ij}(t)(j+1) \;=\; i+1.
\end{aligned}$$

Now every solution x of (3.2) also satisfies the integral equation

$$x(t) = T(t)x(0) + \int_0^t T(t-s)F(x(s))\,ds, \tag{3.6}$$

where $T(t)$ is the C_0 semigroup generated by A. Conversely, a continuous solution x of the integral equation (3.6) is called a *mild* solution of the initial value problem (3.2). The following result guarantees the existence and uniqueness of a mild solution of the problem (3.6) if F is Lipschitz.

Theorem 3.2 (Pazy 1983, Theorem 1.4, Chapter 6) *Let $F : S \to S$ be locally Lipschitz continuous. If A is the infinitesimal generator of a $\mathcal{C}_0$ semigroup e^{tA} on S then for every $x_0 \in S$ there is a $t_{\max} \leq \infty$ such that the initial value problem (3.6) has a unique mild solution x on $[0, t_{\max})$. Moreover, if $t_{\max} < \infty$, then $\lim_{t \uparrow t_{\max}} \| x \| = \infty$.*

Note that, as in Pazy (1983), Theorem 3.2 in fact holds more generally, for a function $F = F(t, u)$: $[0, \infty) \times S \to S$, continuous in time t uniformly on bounded intervals, and locally Lipschitz continuous in u.

We shall apply this theorem to our equation (3.2), with S the space ℓ_{11}. In order to do so, we require F to be locally ℓ_{11}-Lipschitz continuous. This is established as follows.

Lemma 3.3 *The function F defined in* (3.4) *is locally Lipschitz continuous in the ℓ_{11}-norm.*

Proof. For $x, y \geq 0$ such that $\|x\|_{11}, \|y\|_{11} \leq M$, using assumptions (2.5) – (2.14), we have

$$
\begin{aligned}
&\|F(x) - F(y)\|_{11} \\
&\quad \leq \sum_{i\geq 0}(i+1)\sum_{l\neq i}|x^l \alpha_{li}(x) - y^l \alpha_{li}(y)| + \sum_{i\geq 0}(i+1)\Big|x^i \sum_{l\neq i}\alpha_{il}(x) - y^i \sum_{l\neq i}\alpha_{il}(y)\Big| \\
&\qquad + \sum_{i\geq 0}(i+1)|\beta_i(x) - \beta_i(y)| + \sum_{i\geq 0}(i+1)|x^i\delta_i(x) - y^i\delta_i(y)| \\
&\quad \leq \sum_{l\geq 0}|x^l - y^l|\sum_{i\neq l}(i+1)\alpha_{li}(x) + \sum_{l\geq 0}y^l\sum_{i\neq l}(i+1)|\alpha_{li}(x) - \alpha_{li}(y)| \\
&\qquad + \sum_{i\geq 0}(i+1)|x^i - y^i|\sum_{l\neq i}\alpha_{il}(x) + \sum_{i\geq 0}(i+1)y^i\sum_{l\neq i}|\alpha_{il}(x) - \alpha_{il}(y)| \\
&\qquad + b_{11}(x,y)\|x - y\|_{11} + \sum_{i\geq 0}(i+1)|x^i - y^i|\delta_i(x) + \sum_{i\geq 0}(i+1)y^i|\delta_i(x) - \delta_i(y)| \\
&\quad \leq \{a_{10} + \tilde{a}_{11}(0)\|x\|_{11}\}\|x - y\|_{11} + \tilde{a}_{11}(M)\|x - y\|_{11}\|y\|_1 \\
&\qquad + \{a_{00} + \tilde{a}_{01}(0)\|x\|_1\}\|x - y\|_{11} + \tilde{a}_{01}(M)\|y\|_{11}\|x - y\|_1 + \tilde{b}_{11}(M)\|x - y\|_{11} \\
&\qquad + \{d_0 + \tilde{d}_1(0)\|x\|_1\}\|x - y\|_{11} + \tilde{d}_1(M)\|x - y\|_1\|y\|_{11} \\
&\quad \leq F_M\|x - y\|_{11},
\end{aligned}
$$

where

$$
\begin{aligned}
F_M \;:=\; & a_{10} + \tilde{a}_{11}(0)M + M\tilde{a}_{11}(M) + a_{00} + \tilde{a}_{01}(0)M + M\tilde{a}_{01}(M) \\
& + \tilde{b}_{11}(M) + d_0 + \tilde{d}_1(0)M + M\tilde{d}_1(M). \qquad (3.7)
\end{aligned}
$$

∎

Remark. It would naturally be good to have $t_{\max} = \infty$. However, our assumptions may not be enough to guarantee that this is true. On the other hand, $t_{\max} = \infty$ if, for some $C < \infty$,

$$\|F(x)\|_{11} \;\le\; C\|x\|_{11}, \tag{3.8}$$

with $F(x)$ as given in (3.4). For then, from (3.5), we can bound $\|T(t)x\|_{11} \le \|x\|_{11} e^{wt}$, where $w > 0$ is the constant in (2.2). And then, from (3.6) and (3.5), it follows that

$$\|x(t)\|_{11} \;\le\; \|x(0)\|_{11} e^{wt} + \int_0^t C\|x(s)\|_{11} e^{w(t-s)}\,ds, \tag{3.9}$$

and Gronwall's inequality then implies that $\|x(t)\|_{11}$ is bounded on finite intervals, whereas, from Theorem 3.2, it converges to infinity as $t \to t_{\max}$ if the latter is finite. However, under the conditions of this paper, the inequality (3.8) is not automatic.

The specific examples that we consider satisfy slightly stronger conditions, which we can use to bound

$$\begin{aligned} \|F(x)\|_{11} \;&=\; \sum_{i\ge 0}(i+1)|F(x)^i| \\ &\le\; \sum_{i\ge 0}(i+1)\Bigl\{\sum_{l\ge 0} x^l \alpha_{li}(x) + x^i \sum_{l\ge 0}\alpha_{il}(x) + \beta_i(x) + x^i\delta_i(x)\Bigr\}. \end{aligned} \tag{3.10}$$

In all of them, we have $\sup_{i\ge 0}\delta_i(x) < \infty$, uniformly in x, so that the last term in (3.10) causes no problems. In Luchsinger's non-linear model, the $\beta_i(x)$ are all zero, as are the $\alpha_{ij}(x)$ for $i \ge 1$, and it therefore remains to check

$$x^0 \sum_{i\ge 0}(i+1)\alpha_{0i}(x) + x^0 \sum_{l\ge 0}\alpha_{0l}(x),$$

bounded by $\lambda x^0(\max\{\theta, 1\}\|x\|_{11} + \|x\|_1)$; furthermore, x^0 remains bounded by $\|x(0)\|_1$, so that (3.8) is satisfied and $t_{\max} = \infty$. In Luchsinger's linear model, all the $\alpha_{ij}(x)$ are zero, and

$$\sum_{i\ge 0}(i+1)\beta_i(x) \;\le\; \lambda \max\{\theta, 1\}\|x\|_{11},$$

again satisfying (3.8). In Kretzschmar's model, however, we only have the bounds

$$\sum_{i\ge0}(i+1)\sum_{l\ge0}x^l\alpha_{li}(x) \;=\; \sum_{i\ge1}(i+1)x^{i-1}\alpha_{i-1,i}(x) \;\le\; 2\lambda\|x\|_{11}^2/(c+\|x\|_1);$$

$$\sum_{i\ge0}(i+1)x^i\sum_{l\ge0}\alpha_{il}(x) \;=\; \sum_{i\ge0}(i+1)x^i\alpha_{i,i+1}(x) \;\le\; \lambda\|x\|_{11}^2/(c+\|x\|_1),$$

and these are not enough for (3.8). However, if the infection rate is truncated as in (2.15), there is no difficulty, and then $t_{\max}=\infty$ once again.

The next lemma shows that the solution of (3.6) depends smoothly on the initial conditions within the interval of existence. This is useful for approximating our sequence of processes, when it will rarely be possible to have the initial condition fixed for all N. Instead, we would typically have $N^{-1}X_N(0)\to x(0)$ as $N\to\infty$, and approximation throughout by the single solution x of (3.6) starting in $x(0)$ might seem appropriate.

Lemma 3.4 *Fix a solution x to the integral equation* (3.6)*, and suppose that $T<t_{\max}$. Then there is an $\varepsilon>0$ such that, if y is a solution with initial condition $y(0)$ satisfying $\|y-x\|_{11}\le\varepsilon$, then*

$$\sup_{0\le t\le T}\|x(t)-y(t)\|_{11} \;\le\; \|x(0)-y(0)\|_{11}C_T,$$

for a constant $C_T<\infty$.

Proof. From the integral equation (3.6) together with (3.5), it follows that, if $\|x(0)-y(0)\|_{11}\le\varepsilon$, then, as in (3.9),

$$\|x(t)-y(t)\|_{11} \;\le\; \varepsilon e^{wt}+\int_0^t F_{2M_T}\|x(s)-y(s)\|_{11}e^{w(t-s)}\,ds,$$

where F_M is defined in (3.7) and

$$M_T \;:=\; \sup_{0\le t\le T}\|x(t)\|_{11},$$

provided also that $\sup_{0\le t\le T}\|y(t)\|_{11}\le 2M_T$. Again by Gronwall's inequality, it then follows that

$$\sup_{0\le t\le T}\|x(t)-y(t)\|_{11} \;\le\; \|x(0)-y(0)\|_{11}C_T \;\le\; \varepsilon C_T,$$

for a constant $C_T < \infty$. This implies that $\sup_{0\le t\le T} \|y(t)\|_{11} \le 2M_T$ is indeed satisfied if $\varepsilon < M_T/C_T$, and the lemma follows. ∎

Let x_N denote the solution starting with $x_N(0) = N^{-1}X_N(0)$ and x that starting with $x(0)$. It follows that, if $\|x_N(0) - x(0)\|_{11} \to 0$, then for all N large enough

$$\left|\sup_{0\le t\le T} \|x_N(t)\|_{11} - M_T\right| \le \|x_N(0) - x(0)\|_{11} C_T, \tag{3.11}$$

provided that $T < t_{\max}$. In particular, if $t^N_{\max}$ denotes the maximum time such that x_N is uniquely defined on $[0, t^N_{\max})$, then $\liminf_{N\to\infty} t^N_{\max} \ge t_{\max}$.

4 The independent sum approximation

Our aim is to prove the following quantitative law of large numbers for $N^{-1}X_N$. As before, we write x_N for the solution to (3.6) with initial condition $N^{-1}X_N(0)$, and x for the solution starting at $x(0)$.

Theorem 4.1 *Suppose that* (2.2)–(2.14) *hold, and that* $x_N(0) := N^{-1}X_N(0)$ *satisfies* $\|x_N(0) - x(0)\|_{11} \to 0$ *as* $N \to \infty$. *Then, for any* $1/2 < \gamma \le 1$, *and for any* $T < t_{\max}$, *there exists a constant* $K_\gamma(T)$ *such that, as* $N \to \infty$,

$$\mathbf{P}[N^{-1} \sup_{0\le t\le T} \|X_N(t) - Nx_N(t)\|_1 > K_\gamma(T) N^{\gamma-1} \log N] \;=\; O(N^{-\gamma}).$$

If also $\|x_N(0) - x(0)\|_{11} = O(N^{\gamma-1})$, *then* x_N *can be replaced by* x *in the statement, without altering the order of the approximation.*

Instead of embarking on a direct proof, we first consider an approximating model $\tilde{X}_N(\cdot)$, starting with $\tilde{X}_N(0) = X_N(0)$, and consisting of independent individuals. Each individual's parasite load evolves according to a time inhomogeneous Markov process $\widetilde{W}$ on $\mathbb{Z}_+ \cup \Delta$ with infinitesimal matrix defined by

$$\begin{aligned} q_{lj}(t) &= \mu(l,j) + \tilde{a}_{lj}(t), & j \ne l, \Delta,\ l \ge 0, \\ q_{ll}(t) &= -\sum_{j\ne l} q_{lj}(t) - \bar{\delta}_l - \tilde{d}_l(t), & l \ge 0, \\ q_{l\Delta}(t) &= \bar{\delta}_l + \tilde{d}_l(t), & l \ge 0, \end{aligned} \tag{4.1}$$

where

$$\tilde{a}_{il}(t) := \alpha_{il}(x_N(t)); \quad \tilde{d}_i(t) := \delta_i(x_N(t)); \tag{4.2}$$

and, for $i, j \in \mathbb{Z}_+ \cup \Delta$, we shall write

$$\tilde{p}_{ij}(s,t) \;:=\; \mathbf{P}[\widetilde{W}(t) = j \,|\, \widetilde{W}(s) = i], \quad s < t. \tag{4.3}$$

In addition, individuals may immigrate, with rates

$$N\tilde{b}_i(t) := N\beta_i(x_N(t)). \tag{4.4}$$

The process $\widetilde{X}_N$ differs from X_N in having the non-linear elements of the transition rates made linear, by replacing the Lipschitz state–dependent elements $\alpha_{ij}(x)$, $\beta_i(x)$ and $\delta_i(x)$ at any time t by their 'typical' values $\tilde{a}_{ij}(t)$, $\tilde{b}_i(t)$ and $\tilde{d}_i(t)$. Our strategy will be first to show that the process $\widetilde{X}_N$ stays close to the deterministic process $Nx_N(t)$ with high probability, and then to show that, if this is the case, then X_N also stays close to $\widetilde{X}_N$, again with high probability. However, we shall first use the process $\widetilde{X}_N$ to improve our knowledge about the weak solution x to (3.2).

Let us start by introducing some further notation. Fixing $T < t_{\max}$, define

$$M_T \;:=\; M_T(x) \;:=\; \sup_{0\le t\le T} \sum_{i\ge 1} (i+1)|x^i(t)|; \tag{4.5}$$

$$G_T(\gamma) \;:=\; G_T(\gamma, x) \;:=\; \sup_{0\le t\le T} \sum_{i\ge 0} |x^i(t)|^\gamma, \;\; 1/2 \le \gamma \le 1, \tag{4.6}$$

and write $M_T^N := M_T(x_N)$, $G_T^N(\gamma) := G_T(\gamma, x_N)$. Note that M_T is finite if $x(0) \in \ell_{11}$, because the mild solution x is ℓ_{11}-continuous, and that $M_T \ge 1$ and $G_T(\gamma) \ge 1$ for all $1/2 < \gamma \le 1$ whenever $\|x(0)\|_1 = 1$, as is always the case here.

It is immediate from Lemma 3.4 that if $\|x_N(0) - x(0)\|_{11} \to 0$, then $M_T^N \le M_T + 1$ for all N large enough. Furthermore,

$$\sum_{i \ge |x^i|^{\gamma-1}} |x^i|^\gamma \;\le\; \sum_{i \ge |x^i|^{\gamma-1}} i|x^i|,$$

whereas

$$\sum_{i < |x^i|^{\gamma-1}} |x^i|^\gamma \;\le\; \sum_{i < |x^i|^{\gamma-1}} i^{-\gamma/(1-\gamma)} \;=:\; c_\gamma \;<\; \infty,$$

this last provided that $\gamma > 1/2$. Replacing x by $x(t)$ in the above thus shows that

$$G_T(\gamma) \le M_T + c_\gamma \le (c_\gamma + 1)M_T < \infty \tag{4.7}$$

for all $1/2 < \gamma \le 1$. Hence, using also dominated convergence, we deduce that $G_T^N(\gamma) \le G_T(\gamma) + 1$ for all N sufficiently large.

Our first result of the section controls the mean of the process $N^{-1}\widetilde{X}_N$ in the ℓ_{11}-norm.

Lemma 4.2 *Under Conditions* (2.2)–(2.14)*, for any* $X_N(0)$ *in* ℓ_{11} *and for any* $T < t^N_{max}$*, we have*

$$\sup_{0\le t\le T} N^{-1}\sum_{l\ge 0}(l+1)\mathbf{E}\widetilde{X}^l_N(t)$$
$$\le \{N^{-1}\|X_N(0)\|_{11} + T(b_{00} + \tilde{b}_{01}(0)G^N_T(1)\}e^{(w+a_0^N+a_1^N M_T^N)T} < \infty,$$

where w *is as in Assumption* (2.2)*,* $a_0^N := a_{10} - a_{00}$ *and* $a_1^N := a_{11} - a_{01}$.

Proof. Neglecting the individuals in the cemetery state Δ, the process $\widetilde{X}_N$ can be represented by setting

$$(4.8) \qquad \widetilde{X}_N(t) = \sum_{i\ge 0}\sum_{j=1}^{X^i_N(0)} e(\widetilde{W}_{ij}(t)) + \sum_{i\ge 0}\sum_{j=1}^{R_i(t)} e(\widetilde{W}'_{ij}(t-\tau_{ij})),$$

where $\widetilde{W}_{ij}$ and $\widetilde{W}'_{ij}$, $i \ge 0, j \ge 1$, are independent copies of $\widetilde{W}$, with $\widetilde{W}_{ij}$ and $\widetilde{W}'_{ij}$ starting at i, and the τ_{ij}, $j \ge 1$, are the successive event times of independent time inhomogeneous Poisson (counting) processes R_i with rates $N\tilde{b}_i(t)$, which are also independent of all the $\widetilde{W}_{ij}$ and $\widetilde{W}'_{ij}$; as usual, $e(l)$ denotes the l-th co-ordinate vector. Hence it follows that, given $X_N(0)$,

$$\sum_{l\ge 0}(l+1)\mathbf{E}\widetilde{X}^l_N(t)$$
$$= \sum_{i\ge 0}\Big\{X^i_N(0)\mathbf{E}^0_i\{\widetilde{W}(t)+1\} + N\int_0^t \tilde{b}_i(u)\mathbf{E}^0_i\{\widetilde{W}(t-u)+1\}\,du\Big\}.$$

where $\mathbf{E}^0_i$ is as defined for (2.2).

Now $\widetilde{W}$ has paths which are piecewise paths of W, but with extra killing because of the $\tilde{d}_i(u)$ components of the rates, and with sporadic jumps because of the $\tilde{a}_{ij}(u)$ components. The killing we can neglect, since it serves only to reduce $\mathbf{E}^0_i\widetilde{W}(t)$. For the remainder, by Assumptions (2.5) and (2.6), the rate of occurrence is at most $\chi := \{a_{00} + \tilde{a}_{01}(0)G^N_T(1)\}$, irrespective of state and time. So, defining

$$c_{ij}(u) := \tilde{a}_{ij}(u)/\chi, \quad i,j \ge 0,\ j \ne i;$$
$$c_{ii}(u) := 1 - \sum_{j\ne i} c_{ij}(u), \quad i \ge 0,$$

we can construct the α-jumps by taking them to occur at the event times of a Poisson process R of rate χ, with jump distribution for a jump at time u given by $c_{i\cdot}(u)$ if $\widetilde{W}(u-) = i$. Note that, in this case, no jump is realized with probability $c_{ii}(u)$. Conditional on R having events at times $0 < t_1 < \cdots < t_r < t$ between 0 and t, we thus have

$$
\begin{aligned}
&\mathbf{E}_i^0\{\widetilde{W}(t) + 1 \,|\, t_1, \ldots, t_r\} \\
&\quad \leq \sum_{j_1 \geq 0} p_{ij_1}(t_1) \sum_{l_1 \geq 0} c_{j_1 l_1}(t_1) \sum_{j_2 \geq 0} p_{l_1 j_2}(t_2 - t_1) \sum_{l_2 \geq 0} c_{j_2 l_2}(t_2) \ldots \\
&\qquad \ldots \sum_{j_r \geq 0} p_{l_{r-1} j_r}(t_r - t_{r-1}) \sum_{l_r \geq 0} c_{j_r l_r}(t_r) \sum_{j \geq 0} (j+1) p_{l_r j}(t - t_r),
\end{aligned}
$$

where, as before, $p_{ij}(t) = \mathbf{P}[W(t) = j \,|\, W(0) = i]$. Applying Assumptions (2.2), (2.7) and (2.8) to the last two sums, we have

$$
\begin{aligned}
\sum_{l_r \geq 0} c_{j_r l_r}(t_r) \sum_{j \geq 0} (j+1) p_{l_r j}(t - t_r) &\leq \sum_{l_r \geq 0} c_{j_r l_r}(t_r)(l_r + 1) e^{w(t - t_r)} \\
&\leq a_3^N (j_r + 1) e^{w(t - t_r)},
\end{aligned}
$$

where $a_3^N := \{a_{10} + \tilde{a}_{11}(0) M_T^N\}/\chi$. It thus follows that

$$
\mathbf{E}_i^0\{\widetilde{W}(t) + 1 \,|\, t_1, \ldots, t_r\} \;\leq\; a_3^N \mathbf{E}_i^0\{\widetilde{W}(t_r) + 1 \,|\, t_1, \ldots, t_{r-1}\} e^{w(t - t_r)}.
$$

Arguing inductively, this implies that

$$
\mathbf{E}_i^0\{\widetilde{W}(t) + 1 \,|\, t_1, \ldots, t_r\} \;\leq\; (i+1)\{a_3^N\}^r e^{wt},
$$

and hence, unconditionally, that

$$
\begin{aligned}
\mathbf{E}_i^0\{\widetilde{W}(t) + 1\} &\leq (i+1) e^{wt} \mathbf{E}\{(a_3^N)^{R(t)}\} \\
&\leq (i+1) \exp\{(w + (a_3^N - 1)\chi) t\}. \qquad (4.9)
\end{aligned}
$$

The remainder of the proof is immediate. ∎

Armed with this estimate, we can now proceed to identify $N^{-1}\mathbf{E}\widetilde{X}_N(t)$ with $x_N(t)$, at the same time proving that the mild solution x_N is in fact a classical solution to the infinite differential equation (3.1) with initial condition $N^{-1}X_N(0)$.

First, define the 'linearized' version of (3.1):

$$
\begin{aligned}
\frac{dy^i(t)}{dt} &= \sum_{l \geq 1} y^l(t) \mu(l, i) + \sum_{l \geq 0} y^l(t) \tilde{a}_{li}(t) - y^i(t) \sum_{l \geq 0} \tilde{a}_{il}(t) \\
&\quad + \tilde{b}_i(t) - y^i(t) \tilde{d}_i(t), \quad i \geq 0, \qquad (4.10)
\end{aligned}
$$

where $\tilde{a}$, $\tilde{b}$ and $\tilde{d}$ are as in (4.2) and (4.4), to be solved in $t \in [0,T]$ for an unknown function y. Clearly these equations have x_N itself as a mild solution, and, by the remark following Lemma 3.2, the mild solution is unique under our assumptions, since now

$$\widetilde{F}(t,u)^i \ := \ \sum_{l\geq 0} u^l \tilde{a}_{li}(t) - u^i \sum_{l \geq 0} \tilde{a}_{il}(t) + \tilde{b}_i(t) - u^i \tilde{d}_i(t)$$

is ℓ_{11} locally Lipschitz in $u \in S$ with constant $a_{00} + a_{10} + d_0 + M_T^N\{\tilde{a}_{01}(0) + \tilde{a}_{11}(0) + \tilde{d}_1(0)\}$, and

$$\begin{aligned} &\|\widetilde{F}(s,u) - \widetilde{F}(t,u)\|_{11} \\ &\quad \leq \ (M_T^N\{\tilde{a}_{11}(M_T^N) + \tilde{a}_{01}(M_T^N) + \tilde{d}_1(M_T^N)\} + \tilde{b}_{11}(M_T^N))\|x_N(s) - x_N(t)\|_{11}, \end{aligned}$$

whenever $\|u\|_{11} \leq M_T^N$, so that the t-uniform continuity of $\widetilde{F}$ on compact time intervals (contained in $[0, t^N_{\max})$) follows from that of x_N. We now show that $y(t) = N^{-1}\mathbf{E}\widetilde{X}_N(t)$ solves (4.10), and indeed as a *classical* solution. Since it also therefore solves (3.6), and since this equation has a unique solution, it follows that y is the same as x_N, and that it is the classical solution to equation (3.1) with initial condition $N^{-1}X_N(0)$.

Theorem 4.3 *Under Conditions* (2.2)–(2.14)*, for any fixed* $X_N(0) \in \ell_{11}$*, the function* $y(t) := N^{-1}\mathbf{E}\widetilde{X}_N(t)$ *satisfies the system* (4.10) *with initial condition* $N^{-1}X_N(0)$ *on any interval* $[0,T]$ *with* $T < t^N_{max}$*. It is hence the unique classical solution* x_N *to* (3.1) *for this initial condition.*

Proof. Let $\widetilde{X}^j_{N1}(t)$ denote the number of particles present at time 0 that are still present and in state j at time t; let $\widetilde{X}^j_{N2}(t)$ denote the number of particles that immigrated after time 0 and are present and in state j at time t. Then

$$\begin{aligned} \mathbf{E}\,\widetilde{X}^j_N(t) &= \mathbf{E}\,\widetilde{X}^j_{N1}(t) + \mathbf{E}\,\widetilde{X}^j_{N2}(t) \\ &= \sum_{i\geq 0} X^i_N(0)\tilde{p}_{ij}(0,t) + N\int_0^t \sum_{i\geq 0} \tilde{b}_i(u)\tilde{p}_{ij}(u,t)\,du, \end{aligned} \qquad (4.11)$$

where $\tilde{p}_{ij}(u,v)$ is as defined in (4.3). Note that the expectations are finite, since, from Conditions (2.9) and (2.10), uniformly in $u \in [0,T]$,

$$\sum_{i\geq 0} \tilde{b}_i(u) \ \leq \ b_{00} + \tilde{b}_{01}(0)G^N_T(1) \ < \ \infty; \qquad (4.12)$$

here, we have used the fact that, by the subadditivity of the function x^γ in $x \geq 0$, $G_T(1) \leq G_T(\gamma)^{1/\gamma} < \infty$. Then, defining $q_{jl}(t)$ as in (4.1), it follows for each $j \geq 0$ that the quantities $q_j(t) := -q_{jj}(t)$ are bounded, uniformly in t, since by Conditions (2.13), (2.14) (2.5) and (2.6),

$$\tilde{d}_j(t) \;\leq\; d_0 + \tilde{d}_1(0)G_T^N(1); \quad \sum_{l\geq 0} \tilde{a}_{jl}(t) \;\leq\; a_{00} + \tilde{a}_{01}(0)M_T^N, \tag{4.13}$$

and since $\mu(j) = \sum_{l\neq j} \mu(l,j)$ is finite. Note further that, from the forward equations,

$$\frac{\partial}{\partial t}\tilde{p}_{ij}(u,t) \;=\; \sum_{l\geq 0} \tilde{p}_{il}(u,t)q_{lj}(t);$$

see Iosifescu & Tautu (1973, Corollary to Theorem 2.3.8, p. 214).

Now we have

$$\begin{aligned}
\mathbf{E}\,\widetilde{X}^j_{N1}(t) &= \sum_{i\geq 0} X^i_N(0)\tilde{p}_{ij}(0,t) \\
&= X^j_N(0) + \sum_{i\geq 0} X^i_N(0) \int_0^t \sum_{l\geq 0} \tilde{p}_{il}(0,u)q_{lj}(u)\,du \\
&= X^j_N(0) + \int_0^t \sum_{l\geq 0} \mathbf{E}\,\widetilde{X}^l_{N1}(u)q_{lj}(u)\,du, \qquad (4.14)
\end{aligned}$$

with no problems about reordering, because of the uniform boundedness discussed above, and since only one of the q_{lj} is negative. Then also, with rearrangements similarly justified, we define

$$Q_t \;:=\; N\int_0^t \Big\{\int_0^v \sum_{i\geq 0} \tilde{b}_i(u) \sum_{l\geq 0} \tilde{p}_{il}(u,v)q_{lj}(v)\,du\Big\}\,du;$$

taking the i-sum first, and then the u-integral, we obtain

$$Q_t \;=\; \int_0^t \sum_{l\geq 0} \mathbf{E}\,\widetilde{X}^l_{N2}(v)q_{lj}(v)\,dv;$$

taking the l-sum first, we have

$$\begin{aligned}
Q_t &= N\int_0^t \Big\{\int_0^t \sum_{i\geq 0} \tilde{b}_i(u)\frac{\partial}{\partial v}\tilde{p}_{ij}(u,v)\mathbf{1}_{[0,v]}(u)\,du\Big\}\,dv \\
&= N\int_0^t \sum_{i\geq 0} \tilde{b}_i(u)\{\tilde{p}_{ij}(u,t) - \tilde{p}_{ij}(u,u)\}\,du \\
&= \mathbf{E}\,\widetilde{X}^j_{N2}(t) - N\int_0^t \tilde{b}_j(u)\,du.
\end{aligned}$$

From these two representations of Q_t, it follows that

$$\mathbf{E}\,\widetilde{X}^j_{N2}(t) \;=\; \int_0^t \Big\{ N\tilde{b}_j(u) + \sum_{l\geq 0} \mathbf{E}\,\widetilde{X}^l_{N2}(u) q_{lj}(u) \Big\}\, du; \tag{4.15}$$

combining (4.14) and (4.15), we thus have

$$N^{-1}\mathbf{E}\,\widetilde{X}^j_N(t) \;=\; N^{-1}X^j_N(0) + \int_0^t \Big\{ \tilde{b}_j(u) + \sum_{l\geq 0} N^{-1}\mathbf{E}\,\widetilde{X}^l_N(u) q_{lj}(u) \Big\}\, du. \tag{4.16}$$

Since the right hand side is an indefinite integral up to t, it follows that $N^{-1}\mathbf{E}\,\widetilde{X}^j_N(t)$ is continuous in t, for each j. The quantities $q_{jl}(t)$ are all continuous, because x_N is ℓ_{11}-continuous in t and the $\alpha_{il}(x)$ and $\delta_l(x)$ are ℓ_{11}-Lipschitz, and also, for $q_{ll}(t)$, from assumption (2.6). Then, from Lemma 4.2, we also have

$$\sum_{j\geq J} \mathbf{E}\,\widetilde{X}^j_N(t) \leq (J+1)^{-1}\{\|X_N(0)\|_{11} + NTb_{00}G^N_T(1)\} e^{(w+a^N_0+a^N_1 M^N_T)T},$$

so that, in view of assumption (2.4) and of (4.13), the sum on the right hand side of (4.16) is uniformly convergent, and its sum continuous. Hence (4.16) can be differentiated with respect to t to recover the system (4.10), proving the theorem. ∎

Our next result shows that, under appropriate conditions, $N^{-1}\widetilde{X}_N(t)$ and $x(t)$ are close in ℓ_1-norm at any *fixed* t, with very high probability.

Lemma 4.4 *Suppose that Conditions* (2.2)–(2.14) *are satisfied, and that* $X_N(0) \in \ell_{11}$. *Then, for any* $t \in [0,T]$ *with* $T < t^N_{max}$ *and any* $\gamma \in (1/2, 1]$,

$$\mathbf{E}\|\widetilde{X}_N(t) - Nx_N(t)\|_1 \;\leq\; 3N^\gamma G^N_T(\gamma).$$

Furthermore, for any $r > 0$, *there exist constants* $K^{(1)}_r > 1$ *and* $K^{(2)}_r$ *such that*

$$\mathbf{P}[\|\widetilde{X}_N(t) - Nx_N(t)\|_1 > K^{(1)}_r G^N_T(\gamma) N^\gamma \log N] \;\leq\; K^{(2)}_r G^N_T(1) N^{-r}.$$

Proof. For a sum W of independent indicator random variables with mean M, and for any $\delta > 0$, the Chernoff inequalities give

$$\max\{\mathbf{P}[W > M(1+\delta)], \mathbf{P}[W < M(1-\delta)]\} \;\leq\; \exp\{-M\delta^2/(2+\delta)\}. \tag{4.17}$$

Now the quantity $\widetilde{X}^j(t)$ can be expressed as a sum of independent indicator random variables $Y_1^j, \ldots, Y_N^j$, where Y_k^j is the indicator of the event that the k-th initial individual is in state j at time t, and an independent Poisson random variable Y' with mean $N\int_0^t \sum_{i\ge 0} \beta_i(s)\tilde{p}_{ij}(s,t)\,ds$. Hence

$$\mathbf{E}|\widetilde{X}_N^i(t) - Nx_N^i(t)| \;\le\; \sqrt{Nx_N^i(t)} \wedge \{2Nx_N^i(t)\}, \tag{4.18}$$

and, by (4.17), for any $a \ge 2$ and $N \ge 3$,

$$\mathbf{P}[|\widetilde{X}_N^i(t) - Nx_N^i(t)| > a\sqrt{Nx_N^i(t)}\log N] \;\le\; 2N^{-a/2},$$

so long as $Nx_N^i(t) \ge 1$.

Let $I_N(t) := \{i\colon x_N^i(t) \ge 1/N\}$. Then, recalling that we need $\gamma > 1/2$ to be sure that $G_T^N(\gamma) < \infty$, we have

$$\begin{aligned} \sum_{i\in I_N(t)} \mathbf{E}|\widetilde{X}_N^i(t) - Nx_N^i(t)| &\;\le\; \sum_{i\in I(t)} \sqrt{Nx_N^i(t)} \\ &\;\le\; \sum_{i\ge 0} \{Nx_N^i(t)\}^\gamma \;=\; N^\gamma G_T^N(\gamma). \end{aligned} \tag{4.19}$$

Furthermore, $|I_N(t)| \le NG_T^N(1)$, so that, if

$$B_N(t) \;:=\; \bigcap_{i\in I_N(t)} \{|\widetilde{X}_N^i(t) - Nx_N^i(t)| \le a\sqrt{Nx_N^i(t)}\log N\},$$

then

$$\mathbf{P}[B_N^C(t)] \;\le\; 2G_T^N(1)N^{1-a/2}, \tag{4.20}$$

whereas, on the event $B_N(t)$, arguing as above, we have

$$\begin{aligned} \sum_{i\in I_N(t)} |\widetilde{X}_N^i(t) - Nx_N^i(t)| &\;\le\; a\,\log N \sum_{i\in I(t)} \sqrt{Nx_N^i(t)} \\ &\;\le\; a\,\log N\, N^\gamma G_T^N(\gamma). \end{aligned} \tag{4.21}$$

For the remaining indices, we first have

$$\sum_{i\notin I_N(t)} Nx_N^i(t) \;\le\; \sum_{i\notin I_N(t)} N\{x_N^i(t)\}^\gamma\, N^{\gamma-1} \;\le\; N^\gamma G_T^N(\gamma), \tag{4.22}$$

from which, with (4.18) and (4.19), it follows that

$$\mathbf{E}\|\widetilde{X}_N(t) - Nx_N(t)\|_1 \;=\; \sum_{i\ge 0}\mathbf{E}|\widetilde{X}^i_N(t) - Nx^i_N(t)| \;\le\; 3N^\gamma G^N_T(\gamma),$$

proving the first part of the lemma. Then $S_N(t) := \sum_{i\notin I_N(t)} \widetilde{X}^i_N(t)$ is also a sum of many independent indicator random variables plus an independent Poisson component. Using (4.17), we thus have

$$\mathbf{P}\Big[S_N(t) > \sum_{i\notin I_N(t)} Nx^i_N(t) + N^\gamma G^N_T(\gamma)\Big] \;\le\; \exp\{-N^\gamma G^N_T(\gamma)/3\} \;\le\; \exp\{-N^\gamma/3\}, \tag{4.23}$$

since $\delta := N^\gamma G^N_T(\gamma)/\sum_{i\notin I_N(t)} Nx^i_N(t) \ge 1$ from (4.22); otherwise, we have

$$\sum_{i\notin I_N(t)} \widetilde{X}^i_N(t) \;\le\; 2N^\gamma G^N_T(\gamma). \tag{4.24}$$

Now, fixing any $r > 0$ and taking $a = 2(r+1)$, the second part of the lemma follows from (4.20) – (4.24). ∎

The next lemma is used to control the fluctuations of $\widetilde{X}_N$ between close time points. We define the quantity

$$\begin{aligned} H^N_T &:= 2^{m_2-1}m_1 + \{b_{00} + a_{00} + \tilde{b}_{01}(0) + d_0 + G^N_T(1)(\tilde{a}_{01}(0) + \tilde{d}_1(0))\}/\lceil NM^N_T\rceil^{m_2-1} \\ &\ge 1, \end{aligned}$$

which will be used as part of an upper bound for the transition rates of the process $\widetilde{X}_N$ on $[0,T]$.

Lemma 4.5 *Suppose that Conditions* (2.2)–(2.14) *are satisfied. Then, if* $h \le 1/(2\lceil NM^N_T\rceil^{m_2} H^N_T)$, $t \le T-h$, *and if* $\|\widetilde{X}_N(t) - Nx_N(t)\|_1 \le KN^\gamma \log N$, *it follows that*

$$\mathbf{P}\Big[\sup_{0\le u\le h}\|\widetilde{X}_N(t+u) - \widetilde{X}_N(t)\|_1 > KN^\gamma\log N + a\log N\Big] \;\le\; N^{-a/6},$$

for any $a \ge 2$ *and* $N \ge 3$.

Proof. At time t, there are $\|\widetilde{X}_N(t)\|_1$ individuals in the system, each of which evolves independently of the others over the interval $s \in [t, t+h]$; in addition, new immigrants may arrive. During the interval $[t, t+h]$, an individual in state $i \geq 0$ at time t has probability

$$\exp\Big\{-\Big(h(\mu(i)+\bar\delta_i) + \int_0^h \delta_i(x(t+u))\,du + \int_0^h \sum_{l\neq i} \alpha_{il}(x(t+u))\,du\Big)\Big\}$$

of not changing state; and the expected number of immigrants is Poisson distributed with mean

$$N\sum_{i\geq 0}\int_0^h \beta_i(x(t+u))\,du.$$

Now consider the total number $R(t,h)$ of individuals that change state during the interval $[t,t+h]$. For each $i\geq 0$, the individuals in state i at time t can be split into two groups, the first containing $\widetilde X_N^i(t)\wedge Nx_N^i(t)$ randomly chosen individuals, and the second containing the remainder. Adding over i, the numbers in the second group add up to at most $KN^\gamma \log N$, by assumption, whereas, from the observations above, the mean number of individuals who change state in the first group is at most

$$\begin{aligned}
&\sum_{i=0}^{2\lceil NM_T^N\rceil-1} hNx_N^i(t)m_1(i+1)^{m_2} + \sum_{i\geq 2\lceil NM_T^N\rceil} Nx_N^i(t) + \sum_{i\geq 0} hNx_N^i(t)\{d_0+\tilde d_1(0)G_T^N(1)\}\\
&\qquad + \sum_{i\geq 0} hNx_N^i(t)\{a_{00}+\tilde a_{01}(0)G_T^N(1)\} + hN\{b_{00}+\tilde b_{01}(0)G_T^N(1)\}\\
&\quad\leq\ hNM_T^N m_1\{2\lceil NM_T^N\rceil\}^{m_2-1} + \frac12\\
&\qquad + hN\{b_{00}+G_T^N(1)(d_0+a_{00}+\tilde b_{01}(0)) + [G_T^N(1)]^2(\tilde d_1(0)+\tilde a_{01}(0))\} \qquad (4.25)\\
&\quad\leq\ \frac12 + h\lceil NM_T^N\rceil^{m_2}H_T^N\ \leq\ 1,
\end{aligned}$$

from (2.3) and (2.5) – (2.14), and with the last inequality by the assumption on h. Applying the Chernoff bounds (4.17), the probability that more than $a\log N\geq 2$ individuals in the first group change state is thus at most $N^{-a/6}$, implying in sum that

$$\mathbf{P}[R(t,h)\geq KN^\gamma\log N + a\log N]\ \leq\ N^{-a/6}.$$

Since $\sup_{0\leq u\leq h}\|\widetilde X_N(t+u)-\widetilde X_N(t)\|_1 \leq R(t,h)$, the lemma follows. ∎

We are now in a position to prove the main result of the section, showing that the independent sum process $N^{-1}\widetilde X_N$ is a good approximation to x_N, uniformly in $[0,T]$.

Theorem 4.6 *Under Conditions* (2.2)–(2.14), *and for* $T < t^N_{max}$, *there exist constants* $K^{(3)}_r, K^{(4)}_r < \infty$ *such that for* N *large enough*

$$\mathbf{P}[\sup_{0\le t\le T} \|N^{-1}\widetilde{X}_N(t) - x_N(t)\|_1 > K^{(4)}_r (M^N_T)^2 N^{\gamma-1}\log N] \;\le\; K^{(3)}_r (M^N_T)^{m_2+1} N^{-r}.$$

Proof. Divide the interval $[0,T]$ into $\lceil 2T\lceil NM^N_T\rceil^{m_2} H^N_T\rceil$ intervals $[t_l, t_{l+1}]$ of lengths $h_l = t_{l+1} - t_l \le 1/(2\lceil NM^N_T\rceil^{m_2} H^N_T)$. Apply Lemma 4.4 with $r + m_2$ for r and with $t = t_l$ for each l, and apply Lemma 4.5 with $a = 6(r + m_2)$ and with $t = t_l$ and $h = h_l$ for each l; except on a set of probability at most

$$\lceil 2T\lceil NM^N_T\rceil^{m_2} H^N_T\rceil (K^{(2)}_{r+m_2} G^N_T(1) N^{-r-m_2} + N^{-r-m_2}),$$

we have

$$\begin{aligned}
&\sup_{0\le t\le T} \|N^{-1}\widetilde{X}_N(t) - x_N(t)\|_1 \\
&\qquad \le\; 2K^{(1)}_{r+m_2} G^N_T(\gamma) N^{\gamma-1}\log N + 6(r+m_2)N^{-1}\log N \\
&\qquad\quad + \sup_{0\le s,t\le T;|s-t|\le 1/(2\lceil NM^N_T\rceil^{m_2} H^N_T)} \|x_N(s) - x_N(t)\|_1. \qquad (4.26)
\end{aligned}$$

Now, since x_N satisfies (3.6), it follows that, for $0 \le u \le h$,

$$\|x_N(t+u) - x_N(t)\|_1 \;\le\; \|x_N(t)P(u) - x_N(t)\|_1 + \int_0^u \|F(x_N(t+v))P(h-v)\|_1\,dv.$$

By (3.5), we have

$$\|F(x_N(t+v))P(h-v)\|_{11} \;\le\; e^{wu}\|F(x_N(t+v))\|_{11} \;\le\; e^{wh} M^N_T F_{M^N_T},$$

the last inequality following from Lemma 3.3, so that therefore

$$\int_0^u \|F(x_N(t+v))P(h-v)\|_1\,dv \;\le\; he^{wh} M^N_T F_{M^N_T}.$$

Then

$$\|x_N(t)P(u) - x_N(t)\|_1 \;\le\; 2\sum_{j\ge 0} |x^j_N(t)|(1 - p_{jj}(u)),$$

and, by part of the calculation in (4.25), we find that

$$\begin{aligned}
\sum_{j\ge 0} |x^j_N(t)|(1 - p_{jj}(u)) &\le u \sum_{j=0}^{2\lceil NM^N_T\rceil - 1} |x^j_N(t)| m_1 (j+1)^{m_2} + \sum_{j\ge 2\lceil NM^N_T\rceil} |x^j_N(t)| \\
&\le hm_1(2\lceil NM^N_T\rceil^{m_2-1}) M^N_T + \frac{1}{2N},
\end{aligned}$$

which is at most $1/N$ if $h \le 1/(2\lceil NM_T^N \rceil^{m_2} H_T^N)$.

Hence, using also (3.7) and the fact that $m_2 \ge 1$, the third term in (4.26) is of order $(M_T^N)^2 N^{-1}$ under the conditions of the theorem, and the result follows from (4.7). ∎

5 The main approximation

We now turn to estimating the deviations of the process $\widetilde{X}_N$ from the actual process X_N of interest. We do so by coupling the processes in such a way that the "distance" between them cannot increase too much over any finite time interval. In our coupling, we pair each individual in state $i \ge 1$ in $X_N(0)$ with a corresponding individual in state i in $\widetilde{X}_N(0)$ so that all their μ- and $\bar{\delta}$-transitions are identical. This process entails an implicit labelling, which we suppress from the notation. Now the remaining transitions have rates which are not quite the same in the two processes, and hence the two can gradually drift apart. Our strategy is to make their transitions identical as far as we can, but, once a transition in one process is not matched in the other, the individuals are decoupled thereafter. For our purposes, it is simply enough to show that the *number* of decoupled pairs is small enough; what pairs of states these individuals occupy is immaterial.

We realize the coupling between X_N and $\widetilde{X}_N$ in terms of a four component process $Z(\cdot)$ with

$$Z(t) = ((Z_l^i(t),\, i \ge 0,\, 1 \le l \le 3), Z_4(t)) \in \mathcal{X}^3 \times \mathbb{Z}_+,$$

constructed in such a way that we can define $X_N(\cdot) = Z_1(\cdot) + Z_2(\cdot)$ and $\widetilde{X}_N(\cdot) = Z_1(\cdot) + Z_3(\cdot)$, and starting with $Z_1(0) = X_N(0) = \widetilde{X}_N(0)$, $Z_2(0) = Z_3(0) = 0 \in \mathcal{X}$, and $Z_4(0) = 0$. The component Z_4 is used only to keep count of certain uncoupled individuals; either unmatched Z_2-immigrants, or Z_3 individuals that die, or Z_2 individuals created at the death of (one member of) a coupled pair. The transition rates of Z are given as follows, using the notation $e_l(i)$ for the ith coordinate vector in the lth copy of $\mathcal{X}$, and writing $X = Z_1 + Z_2$. For the $\boldsymbol{\mu}$- and α-transitions, at time t and for any $i \ne l$, we have

$$\begin{aligned}
Z &\to Z + (e_1(l) - e_1(i)) \quad \text{at rate} \quad Z_1^i\{\mu(i,l) + (\alpha_{il}(N^{-1}X) \wedge \alpha_{il}(x_N(t)))\};\\
Z &\to Z + (e_2(l) + e_3(i) - e_1(i)) \quad \text{at rate} \quad Z_1^i\{\alpha_{il}(N^{-1}X) - \alpha_{il}(x_N(t))\}^+;\\
Z &\to Z + (e_2(i) + e_3(l) - e_1(i)) \quad \text{at rate} \quad Z_1^i\{\alpha_{il}(N^{-1}X) - \alpha_{il}(x_N(t))\}^-;
\end{aligned}$$

$$
\begin{aligned}
Z &\to Z + (e_2(l) - e_2(i)) \quad \text{at rate} \quad Z_2^i\{\mu(i,l) + \alpha_{il}(N^{-1}X)\};\\
Z &\to Z + (e_3(l) - e_3(i)) \quad \text{at rate} \quad Z_3^i\{\mu(i,l) + \alpha_{il}(x_N(t))\},
\end{aligned}
$$

with possibilities for individuals in the two processes to become uncoupled, when $N^{-1}X \neq x(t)$. For the immigration transitions, we have

$$
\begin{aligned}
Z &\to Z + e_1(i) \quad \text{at rate} \quad N\{\beta_i(N^{-1}X) \wedge \beta_i(x_N(t))\}, \quad i \geq 0;\\
Z &\to Z + e_2(i) + e_4 \quad \text{at rate} \quad N\{\beta_i(N^{-1}X) - \beta_i(x_N(t))\}^+, \quad i \geq 0;\\
Z &\to Z + e_3(i) \quad \text{at rate} \quad N\{\beta_i(N^{-1}X) - \beta_i(x_N(t))\}^-, \quad i \geq 0,
\end{aligned}
$$

with some immigrations not being precisely matched; the second transition includes an e_4 to ensure that each individual in Z_2 has a counterpart in either Z_3 or Z_4. For the deaths, we have

$$
\begin{aligned}
Z &\to Z - e_1(i) \quad \text{at rate} \quad Z_1^i\{\bar{\delta}_i + (\delta_i(N^{-1}X) \wedge \delta_i(x_N(t)))\}, \quad i \geq 0;\\
Z &\to Z - e_1(i) + e_3(i) \quad \text{at rate} \quad Z_1^i\{\delta_i(N^{-1}X) - \delta_i(x_N(t))\}^+, \quad i \geq 0;\\
Z &\to Z - e_1(i) + e_2(i) + e_4 \quad \text{at rate} \quad Z_1^i\{\delta_i(N^{-1}X) - \delta_i(x_N(t))\}^-, \quad i \geq 0;\\
Z &\to Z - e_2(i) \quad \text{at rate} \quad Z_2^i\{\bar{\delta}_i + \delta_i(N^{-1}X)\}, \quad i \geq 0;\\
Z &\to Z - e_3(i) + e_4 \quad \text{at rate} \quad Z_3^i\{\bar{\delta}_i + \delta_i(x_N(t))\}, \quad i \geq 0,
\end{aligned}
$$

where $Z_4(\cdot)$ is also used to count the deaths of uncoupled Z_3-individuals, and uncoupled deaths in $\tilde{X}_N$ of coupled Z_1 individuals. With this joint construction, we have arranged that

$$
\sum_{i\geq 0} Z_2^i(t) \leq Z_4(t) + \sum_{i\geq 0} Z_3^i(t) \tag{5.1}
$$

for all t, and that

$$
V_N(t) \;:=\; Z_4(t) + \sum_{i\geq 0} Z_3^i(t) \tag{5.2}
$$

is a counting process. We allow unmatched deaths in the Z_2-process. We thus have the bound

$$
\begin{aligned}
\|X_N(t) - \widetilde{X}_N(t)\|_1 &= \|(Z_1(t) + Z_2(t)) - (Z_1(t) + Z_3(t))\|_1\\
&\leq \sum_{i\geq 0}\{Z_2^i(t) + Z_3^i(t)\} \;\leq\; 2\left\{Z_4(t) + \sum_{i\geq 0} Z_3(t)\right\} \;=\; 2V_N(t),
\end{aligned} \tag{5.3}
$$

for all t, by (5.1).

Now V_N has a compensator A_N with intensity a_N, satisfying

$$\begin{aligned} a_N(t) &= \sum_{i\ge 0} Z_1^i(t) \sum_{l\ge 0} |\alpha_{il}(N^{-1}X_N(t)) - \alpha_{il}(x_N(t))| \\ &\quad + N\sum_{i\ge 0} |\beta_i(N^{-1}X_N(t)) - \beta_i(x_N(t))| + \sum_{i\ge 0} Z_1^i(t)|\delta_i(N^{-1}X_N(t)) - \delta_i(x_N(t))| \\ &\le \sum_{i\ge 0} \widetilde{X}_N^i(t) \sum_{l\ge 0} |\alpha_{il}(N^{-1}X_N(t)) - \alpha_{il}(x_N(t))| \\ &\quad + N\sum_{i\ge 0} |\beta_i(N^{-1}X_N(t)) - \beta_i(x_N(t))| + \sum_{i\ge 0} \widetilde{X}_N^i(t)|\delta_i(N^{-1}X_N(t)) - \delta_i(x_N(t))|. \end{aligned}$$

Now, Condition (2.6) implies that, uniformly in i,

$$\sum_{l\ge 0} |\alpha_{il}(N^{-1}X_N(t)) - \alpha_{il}(x_N(t))| \le \tilde{a}_{01}(\|x_N(t)\|_{11})\|N^{-1}X_N(t) - x_N(t)\|_1.$$

Hence,

$$\begin{aligned} &N^{-1}a_N(t) \\ &\quad\le \Big(\sum_{i\ge 0} x_N^i(t)\tilde{a}_{01}(M_T^N) + \tilde{b}_{01}(M_T^N) + \sum_{i\ge 0} x_N^i(t)\tilde{d}_1(M_T^N)\Big)\|N^{-1}X_N(t) - x_N(t)\|_1 \\ &\qquad + \|N^{-1}\widetilde{X}_N(t) - x_N(t)\|_1\Big(\tilde{a}_{01}(M_T^N) + \tilde{d}_1(M_T^N)\Big)\|N^{-1}X_N(t) - x_N(t)\|_1 \\ &\quad\le \{H_T^{(1,N)} + H_T^{(2,N)}\|N^{-1}\widetilde{X}_N(t) - x_N(t)\|_1\}\|N^{-1}X_N(t) - x_N(t)\|_1, \end{aligned}$$

where

$$\begin{aligned} H_T^{(1,N)} &= G_T^N(1)\tilde{a}_{01}(M_T^N) + \tilde{b}_{01}(M_T^N) + G_T^N(1)\tilde{d}_1(M_T^N); \\ H_T^{(2,N)} &= \tilde{a}_{01}(M_T^N) + \tilde{d}_1(M_T^N). \end{aligned}$$

In particular, defining

$$\tau_N := \inf\{t\ge 0 : \|N^{-1}\widetilde{X}_N(t) - x_N(t)\|_1 \ge 1\},$$

it follows that

$$\begin{aligned} &N^{-1}a_N(t\wedge\tau_N) \\ &\quad\le \{H_T^{(1,N)} + H_T^{(2,N)}\}\|N^{-1}X_N(t\wedge\tau_N) - x_N(t\wedge\tau_N)\|_1. \end{aligned} \tag{5.4}$$

For the next two lemmas, we shall restrict the range of N in a way that is asymptotically unimportant. We shall suppose that N satisfies the inequalities

$$K_{r_0}^{(4)}(M_T^N)^2 N^{\gamma-1}\log N \le 1; \qquad N > \{K_{r_0}^{(3)}\}^{1/(m_2+1)} M_T^N, \tag{5.5}$$

where $r_0 = m_2 + 2$, and the quantities $K_r^{(3)}$ and $K_r^{(4)}$ are as for Theorem 4.6.

Lemma 5.1 *Under Conditions* (2.2)–(2.14)*, for any* $t \in [0,T]$ *and* $1/2 < \gamma \leq 1$*, and for all* N *satisfying* (5.5)*, we have*

$$N^{-1}\mathbf{E}\|X_N(t\wedge\tau_N) - \widetilde{X}_N(t\wedge\tau_N)\|_1 \\ \leq\ 8N^{\gamma-1}G_T^N(\gamma)\, t(H_T^{(1,N)} + H_T^{(2,N)})\exp\{t(H_T^{(1,N)} + H_T^{(2,N)})\}.$$

Proof. Write $M_N(\cdot) = V_N(\cdot) - A_N(\cdot)$ for the martingale part of V_N. Then, because also

$$\|N^{-1}X_N(t)-x_N(t)\|_1 \ \leq\ \|N^{-1}X_N(t)-N^{-1}\widetilde{X}_N(t)\|_1+\|N^{-1}\widetilde{X}_N(t)-x_N(t)\|_1, \tag{5.6}$$

and using (5.4), we have

$$\begin{aligned}(2N)^{-1}&\|X_N(t\wedge\tau_N) - \widetilde{X}_N(t\wedge\tau_N)\|_1 \\ &\leq\ N^{-1}V(t\wedge\tau_N) \\ &\leq\ N^{-1}M_N(t\wedge\tau_N) \\ &\quad + \int_0^{t\wedge\tau_N}\{H_T^{(1,N)} + H_T^{(2,N)}\}\{N^{-1}\|X_N(s\wedge\tau_N) - \widetilde{X}_N(s\wedge\tau_N)\|_1 \\ &\qquad\qquad + \|N^{-1}\widetilde{X}_N(s\wedge\tau_N) - x_N(s\wedge\tau_N)\|_1\}\, ds \\ &\leq\ N^{-1}M_N(t\wedge\tau_N) \\ &\quad + \int_0^{t}\{H_T^{(1,N)} + H_T^{(2,N)}\}\{N^{-1}\|X_N(s\wedge\tau_N) - \widetilde{X}_N(s\wedge\tau_N)\|_1 \\ &\qquad\qquad + \|N^{-1}\widetilde{X}_N(s\wedge\tau_N) - x_N(s\wedge\tau_N)\|_1\}\, ds.\end{aligned} \tag{5.7}$$

Taking expectations, it thus follows that

$$\begin{aligned}(2N)^{-1}&\mathbf{E}\|X_N(t\wedge\tau_N) - \widetilde{X}_N(t\wedge\tau_N)\|_1 \\ &\leq\ \int_0^{t}\{H_T^{(1,N)} + H_T^{(2,N)}\}\{N^{-1}\mathbf{E}\|X_N(s\wedge\tau_N) - \widetilde{X}_N(s\wedge\tau_N)\|_1 \\ &\qquad\qquad + \mathbf{E}\|N^{-1}\widetilde{X}_N(s\wedge\tau_N) - x_N(s\wedge\tau_N)\|_1\}\, ds.\end{aligned} \tag{5.8}$$

Now we have

$$\begin{aligned}\mathbf{E}&\|N^{-1}\widetilde{X}_N(s\wedge\tau_N) - x_N(s\wedge\tau_N)\|_1 \\ &=\ \mathbf{E}\{\|N^{-1}\widetilde{X}_N(s) - x_N(s)\|_1 I[\tau_N \geq s]\} \\ &\qquad + \mathbf{E}\{\|N^{-1}\widetilde{X}_N(\tau_N) - x_N(\tau_N)\|_1 I[\tau_N < s]\} \\ &\leq\ \mathbf{E}\|N^{-1}\widetilde{X}_N(s) - x_N(s)\|_1 + (1+1/N)\mathbf{P}[\tau_N < s] \\ &\leq\ \mathbf{E}\|N^{-1}\widetilde{X}_N(s) - x_N(s)\|_1 + (1+1/N)\mathbf{P}[\tau_N < T].\end{aligned} \tag{5.9}$$

The first term in (5.9) is bounded by $3N^{\gamma-1}G_T^N(\gamma)$ by Lemma 4.4, and, for N satisfying (5.5), the event $\{\tau_N < T\}$ lies in the exceptional set for Theorem 4.6 with $r = r_0$, implying that

$$\mathbf{P}[\tau_N < T] \;\le\; K_{r_0}^{(3)}(M_T^N)^{m_2+1}N^{-r_0}, \tag{5.10}$$

so that the second term is no larger than $N^{\gamma-1}G_T^N(\gamma)$ if

$$N > \{K_{r_0}^{(3)}(M_T^N)^{m_2+1}\}^{1/(r_0+\gamma-1)},$$

which is also true if (5.5) is satisfied. This implies that, for such N,

$$\mathbf{E}\|N^{-1}\widetilde{X}_N(s\wedge\tau_N) - x_N(s\wedge\tau_N)\|_1 \;\le\; 4N^{\gamma-1}G_T^N(\gamma). \tag{5.11}$$

Using (5.11) in (5.8) and applying Gronwall's inequality, the lemma follows. ∎

Lemma 5.2 *Under Conditions* (2.2)–(2.14)*, for any $t \in [0,T]$ and $1/2 < \gamma \le 1$, and for all N satisfying* (5.5)*, we have*

$$\mathbf{P}[\sup_{0\le s\le t}|N^{-1}M_N(s\wedge\tau_N)| \ge N^{\gamma-1}G_T^N(\gamma)] \;\le\; g(t(H_T^{(1,N)}+H_T^{(2,N)}))N^{-\gamma}/G_T^N(\gamma),$$

where $g(x) := 4x(1+xe^x)$.

Proof. Since $V_N(\cdot)$ is a counting process with continuous compensator A_N, we have from (5.4) and (5.6) that

$$\begin{aligned}\mathbf{E}M_N^2(t\wedge\tau_N) &= \mathbf{E}A_N(t\wedge\tau_N)\\ &\le \{H_T^{(1,N)}+H_T^{(2,N)}\}\int_0^t\{\mathbf{E}\|X_N(s\wedge\tau_N)-\widetilde{X}_N(s\wedge\tau_N)\|_1\\ &\qquad + \mathbf{E}\|\widetilde{X}_N(s\wedge\tau_N)-Nx_N(s\wedge\tau_N)\|_1\}\,ds.\end{aligned}$$

The first expectation is bounded using Lemma 5.1, the second from (5.11), from which it follows that

$$\mathbf{E}M_N^2(t\wedge\tau_N) \;\le\; g(t(H_T^{(1,N)}+H_T^{(2,N)}))\,N^{\gamma}G_T^N(\gamma).$$

The lemma now follows from the Lévy–Kolmogorov inequality. ∎

We are finally in a position to complete the proof of Theorem 4.1. Suppose that N satisfies (5.5). Returning to the almost sure inequality (5.7), we can now write

$$
\begin{aligned}
(2N)^{-1} \sup_{0\le s\le t} \|X_N(s\wedge\tau_N) - \widetilde{X}_N(s\wedge\tau_N)\|_1 \;&\le\; N^{-1}V(t\wedge\tau_N) \\
\le\; N^{-1}M_N(t\wedge\tau_N) & \\
+ \int_0^t \{H_T^{(1,N)} + H_T^{(2,N)}\}\{N^{-1}\|X_N(s\wedge\tau_N) - \widetilde{X}_N(s\wedge\tau_N)\|_1 & \\
+ \|N^{-1}\widetilde{X}_N(s\wedge\tau_N) - x_N(s\wedge\tau_N)\|_1\}\, ds. &
\end{aligned}
$$

From Lemma 5.2, we can bound the martingale contribution uniformly on $[0,T]$ by $N^{\gamma-1}G_T^N(\gamma)$, except on an event of probability at most

$$
g(T(H_T^{(1,N)} + H_T^{(2,N)}))N^{-\gamma}.
$$

By Theorem 4.6, for any $r>0$, we can find a constant K_r such that

$$
\sup_{0\le t\le T} \|N^{-1}\widetilde{X}_N(t) - x(t)\|_1 \le K_r(M_T^N)^2 N^{\gamma-1}\log N,
$$

except on an event of probability $O((M_T^N)^{m_2+1}N^{-r})$. Hence, once again by Gronwall's inequality, it follows that, except on these exceptional events,

$$
\begin{aligned}
N^{-1} \sup_{0\le s\le t} \|X_N(s\wedge\tau_N) - \widetilde{X}_N(s\wedge\tau_N)\|_1 & \\
\le\; 2N^{\gamma-1}\{G_T^N(\gamma) + T(H_T^{(1,N)} + H_T^{(2,N)})K_r(M_T^N)^2\log N\}e^{2t(H_T^{(1,N)}+H_T^{(2,N)})}. &
\end{aligned}
$$

Combining this with Theorem 4.6, and since also, by (5.10), $\mathbf{P}[\tau_N < T] = O((M_T^N)^{m_2+1}N^{-r})$ for any r, the first part of the theorem follows:

$$
\mathbf{P}[N^{-1} \sup_{0\le t\le T} \|X_N(t) - Nx_N(t)\|_1 > K_\gamma(T)N^{\gamma-1}\log N] \;=\; O(N^{-\gamma}).
$$

Note that the inequalities (5.5) are satisfied for all N sufficiently large, and that the constant $K_\gamma(T)$ and the implied constant in $O(N^{-\gamma})$ can be chosen uniformly in N, because, under the conditions of the theorem, $\|x_N(0) - x(0)\|_{11} \to 0$ as $N\to\infty$, with the result that, for all large enough N, $G_T^N(\gamma)$, $G_T^N(1)$ and M_T^N can be replaced by $G_T(\gamma)+1$, $G_T(1)+1$ and M_T+1 respectively, with the corresponding modifications in $H_T^{(1,N)}$ and $H_T^{(2,N)}$.

That x_N can be replaced by x in the theorem without changing the order of approximation, provided that $\|x_N(0) - x(0)\|_{11} = O(N^{\gamma-1})$, follows directly from (3.11).

Acknowledgement

ADB wishes to thank the Institute for Mathematical Sciences of the National University of Singapore for their warm hospitality while part of this work was accomplished.

References

[1] F. Arrigoni, Deterministic approximation of a stochastic metapopulation model. *Adv. Appl. Prob.* **35** (2003) 691–720.

[2] N. T. J. Bailey, A perturbation approximation to the simple stochastic epidemic in a large population. *Biometrika* **55** (1968) 199–209.

[3] A. D. Barbour, Threshold phenomena in epidemic theory. In: Probability, statistics and optimization: a tribute to Peter Whittle, Ed. F.P. Kelly, pp. 101–116: Wiley, New York, 1994.

[4] A. D. Barbour & M. Kafetzaki, A host–parasite model yielding heterogeneous parasite loads. *J. Math. Biology* **31** (1993) 157–176.

[5] S. N. Ethier & T. G. Kurtz, *Markov Processes: Characterization and Convergence*, Wiley, 1986.

[6] M. Iosifescu & P. Tautu, *Stochastic processes and applications in biology and medicine I*, Springer, 1973.

[7] O. Kallenberg, *Foundations of Modern Probability*, Springer, 2002.

[8] H. Kesten & B. I. Stigum, A limit theorem for multidimensional Galton–Watson processes. *Ann. Math. Statist.* **37** (1966) 1211–1223.

[9] M. Kimmel & D. E. Axelrod, *Branching processes in biology.* Springer, Berlin, 2002.

[10] M. Kretzschmar, Comparison of an infinite dimensional model for parasitic diseases with a related 2-dimensional system. *J. Math. Analysis Applics* **176** (1993) 235–260.

[11] T. G. Kurtz, Relationships between stochastic and deterministic population models. *Lecture Notes in Biomathematics* **38** (1980) 449–467: Springer, Berlin.

[12] C. J. Luchsinger, Mathematical Models of a Parasitic Disease, Ph.D. thesis, University of Zürich, 1999.

[13] C. J. Luchsinger, Stochastic models of a parasitic infection, exhibiting three basic reproduction ratios. *J. Math. Biol* **42**, (2001a) 532–554.

[14] C. J. Luchsinger, Approximating the long term behaviour of a model for parasitic infection. *J. Math. Biol* **42**, (2001b) 555–581.

[15] S.-T. C. Moy, Extensions of a limit theorem of Everett, Ulam and Harris on multitype branching processes with countably many types. *Ann. Math. Statist.* **38** (1967) 992–999.

[16] J. R. Norris, *Markov Chains*, CUP, 1997.

[17] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer, Berlin, 1983.