Ensemble behaviour in population processes with applications to ecological systems*

P.K. Pollett†

Abstract

We study Markovian models for population processes in continuous time, addressing questions concerning the behaviour of ensembles of individuals (equilibrium, quasi-equilibrium and time-dependent behaviour), and in particular what can be deduced from models for individual behaviour.

Keywords: Markov chains; population processes; stochastic models.

1 Introduction

Scientists frequently model population characteristics using individual-level models, because they are usually very simple to analyse. For example, suppose we were interested in modelling the numbers of individuals that share particular characteristics, such as genotype. Then, a typical state would be $\boldsymbol{n} = (n_j,\, j \in S)$, where S is the set of types, and thus our model could be quite complicated, especially if we were to allow individuals to enter or leave the population (say through birth, death or migration), or to change their type over time. However, if the population had a fixed size n, then it might be reasonable to assume that the *proportion* of individuals, n_j/n, of type j were approximately equal to the *probability* p_j that any given *individual* is of that type. The latter could be gleaned from a simpler one-dimensional model, the set of states of the individual being simply S. But, is it reasonable to approximate the random process of proportions by a set of probabilities, and, then, when is this reasonable? We examine these questions in some detail.

We begin by describing three examples of populations, and indicate how our methods and results can be used to simplify the analysis of models that are designed to study the complex behaviour of individuals in the population and the interaction between them. We make a distinction between the two kinds of models (or processes) by referring to them simply as the "individual model" or the "ensemble model" (or process).

Population 1. Our first example is a population network, frequently called a *metapopulation* (see for example Gilpin and Hanski, 1991), where a fixed number n of individuals occupies

*Based on a paper presented at the 17th Biennial Congress on Modelling and Simulation, Christchurch, New Zealand, December 2007.

†Department of Mathematics, University of Queensland, Qld 4072, AUSTRALIA pkp@maths.uq.edu.au.

geographically separated regions or *patches*. Patches may become empty, but can be recolonized through migration from other patches. From the point of view of the individual, it spends a period of time in a given patch and might then emigrate to another patch, spend a period there, and so forth. The state of the individual model is therefore very simple, being the patch currently occupied. Indeed, assuming individuals do not affect each other's progress through the network, one could model the progress of the individual as a random walk on the patches, and thus evaluate quantities such as the probability $p_j(t)$ that the individual occupies patch j at time t. Our intuition tells us that, for the ensemble, the *proportion* of individuals in patch j at time t should be approximately equal to $p_j(t)$.

In Section 3 we give a careful examination of whether it is reasonable to approximate random proportions of individuals that share a particular characteristic using probabilities derived from individual-level models. We are able to make a very precise statement for a very general class of models. For example, Theorem 1 can be used to show that, for Population 1, the proportion of individuals in patch j at time t can be approximated by the probability $p_j(t)$ that an individual occupies patch j at time t, the individuals being permitted to move amongst the patches according to a very general mechanism. Furthermore, it allows us to deduce that the quality of the approximation improves as the population size n grows.

Population 2. This is a variant of Population 1 where we allow death or external emigration from any patch. We will study two cases: (i) the *open* network, where there is external immigration to one or more patches, and (ii) the *closed* network, where there are n individuals to begin with, each eventually disappearing from the network through death or external emigration. As before, individuals are assumed not to affect one another's progress, but now individuals (perhaps arriving from outside the network) perform a random walk on the patches but then eventually leave. In contrast to Population 1, the total number of individuals is random. Yet, we would expect to be able to draw similar conclusions concerning ensemble proportions. Furthermore, as the population would be expected to settle down to a stable equilibrium, we might ask whether it is also reasonable to approximate the *equilibrium* proportion of individuals occupying patch j using the equilibrium probability that an individual is in patch j. We might expect this to be reasonable in the open case, but even closed metapopulations can exhibit "quasi" equilibrium behaviour over reasonable time scales before extinction occurs; see Pollett (1999).

Theorem 1 can be employed again to show that, for the closed population, the proportion of individuals in patch j at time t can be approximated by the probability that an individual occupies patch j at time t, and that the approximation is best for a large population. Theorem 2 allows us to draw a similar conclusion for the open population, but now the quality of the approximation improves as the arrival rate of individuals from outside the network grows. We examine *equilibrium* proportions in Sections 4 and 5, evaluating stationary and quasi-stationary distributions for the open and closed ensembles, respectively, and describing their relationship with the corresponding distributions for the individual model. For example, Theorem 3 gives an explicit expression for the quasi-stationary distribution of any closed ensemble model in terms of the quasi-stationary distribution of the corresponding

individual model. So, for Population 2, we are able to study equilibrium behaviour (quasi-equilibrium in the closed case) of the entire population network by examining a model that merely tracks the movement of the individual.

Population 3. Our final example is a population of organisms, each having a life time that consists of several distinct stages (for example, the butterfly life cycle comprises egg, larva, pupa and adult). Again our intuition tells us that the proportion of the population in stage s should be close to the proportion of time p_s that an individual spends in stage s of its life cycle.

Theorem 1 confirms our intuition about the process of proportions, and again the approximation is best when the number of organisms is large. The quasi-equilibrium behaviour of this population is examined in Section 5; Theorem 4 gives an explicit expression for the quasi-stationary distribution of the numbers of organisms at the various stages of life in terms of the expected times spent in those stages.

Finally, Section 6 summarises some standard results that are used to evaluate the chance of survival by time t, and the expected survival time, for general population models in terms of their quasi-stationary distribution. Theorem 5 is an intriguing result which shows, under mild conditions, that the numbers of individuals in a population that share particular characteristics are *independent* random variables, each having a Poisson distribution whose mean is *precisely* the survival probability for that type.

We begin by describing a general individual model and then construct the corresponding ensemble model. The individual model is Markovian with a specified set of transition rates Q. Since Q is arbitrary, the model is very flexible. In the ensemble model individuals are assigned to the various states, each then moving independently according to Q.

2 Individual and ensemble models

Let $X(t)$ be the state of an individual at time t (≥ 0); for Populations 1 and 2 it is the patch occupied by that individual, while for Population 3 it is the current stage in the individual's life cycle. We will suppose that $(X(t), t \geq 0)$ is a continuous-time Markov chain taking values in a discrete set S with a (conservative) set of transition rates $Q = (q_{ij}, i, j \in S)$. Thus, q_{ij} represents the rate of transition from state i to state j, for $j \neq i$, and $q_i := -q_{ii} = \sum_{j \neq i} q_{ij}$ represents the total rate out of state i. The sample behaviour is as follows: the process starts in a given state i, spends an exponentially distributed amount of time there with mean $1/q_i$ before jumping to another state j with probability $p_{ij} = q_{ij}/q_i$, where it spends another exponentially distributed period, but with mean $1/q_j$, and so on. If the process reaches a state j for which $q_j = 0$ (an *absorbing state*) it remains there. For further explanations and terminology see for example Anderson (1991) or Norris (1997). It will be convenient to restrict our attention to the case where S is a *finite* set, but I note that many of the arguments presented hold more generally. The state probabilities $\boldsymbol{p}(t) = (p_j(t), j \in S)$,

where $p_j(t) = \Pr(X(t) = j)$, can be obtained as the (unique) solution to the forward equation $\boldsymbol{p}'(t) = \boldsymbol{p}(t)\,Q$ satisfying $\boldsymbol{p}(0) = \boldsymbol{a}$, where $\boldsymbol{a} = (a_j, j \in S)$ is a given initial distribution (here and henceforth all vectors shall be interpreted as row vectors).

Population 1. Suppose that there are M patches in the population network. The states $S = \{1, \ldots, M\}$ are the patch labels and q_{ij} represents the intrinsic tendency for movement of individuals from patch i to patch j: when the individual leaves patch i it moves to patch j with probability $p_{ij} = q_{ij}/q_i$, where $q_i = \sum_{j \neq i} q_{ij}$ is the total (per-capita) rate of departure from patch i. For example, q_{ij} might be a function of the area of patch i and the distance between patches i and j, as in Gyllenberg and Silvestrov (1994), but our results do not require any particular prescription. We will suppose only that S is *irreducible*: any individual may pass between any two patches either directly or indirectly via a chain of other patches.

Population 2. Now $S = \{0, \ldots, M\}$ and we have the extra transition from i to 0 at (per-capita) rate q_{i0} for each $i \in C := \{1, \ldots, M\}$, corresponding to death or external emigration at patch i. We will suppose that $q_{i0} > 0$ for at least one i. In the open case there is an additional set of parameters $\boldsymbol{\nu} = (\nu_i, i \in C)$ with ν_i representing the rate at which new individuals arrive at patch i: when the individual starts its journey through the network it begins in patch i with probability $a_i := \nu_i / \sum_{k \in C} \nu_k$. In contrast to the model for Population 1, S here is *not* irreducible because 0 is an absorbing state. However, it will be convenient to suppose that C is irreducible: the individual can leave the network from any patch or, in the open case, reach any patch from outside the network (in both cases either directly or via a chain of other patches).

Population 3. In the life cycle example we let M be the number of stages of life, with M being the first stage and 1 being the last, and with state 0 representing death, so that $S = \{0, \ldots, M\}$. The rate at which an individual moves from stage s to the next stage $s-1$ ($s \geq 1$) is $q_s = q_{s,s-1} > 0$ (equivalently, the individual spends an exponentially distributed amount of time in stage s with mean $1/q_s$). Again S is not irreducible. Each state s, including the absorbing state 0, is in a class by itself.

Now we construct a model for the ensemble. Suppose that at time $t = 0$ the individuals are assigned to the states according to some rule and then each moves independently in S as a Markov chain governed by Q. The key assumption here is *independence*: individuals do not affect one another. We record only the *number* of individuals in the various states, rather than their positions. Let $N_j(t)$ be the number of individuals in state j at time t, and let $\boldsymbol{N} = (N_j, j \in S)$. The process $(\boldsymbol{N}(t), t \geq 0)$ is also a continuous-time Markov chain. In applied probability parlance, $\boldsymbol{N}(t)$ is an example of a *migration process* (Whittle, 1967), or, in queueing theory parlance, a *network of infinite-server queues*. Suffice it to say that systems like these are much studied and well understood; the books by Kelly (1979) and Serfozo (1999) together give an excellent summary of stochastic network theory. Since we are assuming that individuals move independently, the ensemble model can also be viewed as a (non-interacting) particle system, and thus dates back to at least Doob (1953).

The closed ensemble. We suppose that there is a fixed number n of individuals, each moving according to Q. Population models 1 and 3 and the closed version of model 2 fit this context. The ensemble process takes values in $E = \{\boldsymbol{n} \in \{0, \ldots, n\}^S : \sum_{j \in S} n_j = n\}$, and its transition rates $Q_{_E} = (q(\boldsymbol{n}, \boldsymbol{m}), \boldsymbol{n}, \boldsymbol{m} \in E)$ are given by $q(\boldsymbol{n}, \boldsymbol{n} + \boldsymbol{e}_j - \boldsymbol{e}_i) = n_i q_{ij}$, for all states $j \neq i$ in S, where $\boldsymbol{e}_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ is the unit vector with a 1 as its j-th entry. This follows because a transition from $\boldsymbol{n}$ to $\boldsymbol{n} + \boldsymbol{e}_j - \boldsymbol{e}_i$ is effected by the movement of a single individual from state i to state j and there are n_i individuals in state i when the state is $\boldsymbol{n}$, each moving to state j at rate q_{ij}. Notice that the total rate out of state $\boldsymbol{n}$ is

$$q(\boldsymbol{n}) := \sum_{\boldsymbol{m} \in E:\, \boldsymbol{m} \neq \boldsymbol{n}} q(\boldsymbol{n}, \boldsymbol{m}) = \sum_{i \in S} n_i \sum_{j \neq i} q_{ij} = \sum_{i \in S} n_i q_i \qquad (\boldsymbol{n} \in E).$$

The open ensemble. Now individuals may enter the system and, once there, move according to Q. New individuals appear in the various states according to independent Poisson processes, the rate for state j being ν_j. With Population 2 (open case) in mind, we will adopt the setup $S = \{0\} \cup C$, where C is a finite *irreducible* class and 0 is an absorbing state which is accessible from C ($q_{i0} > 0$ for at least one $i \in C$), so that an individual's departure from the system is effected by its entry into state 0. Thus, the ensemble process takes values in $E = \{0, 1, \ldots\}^C$, which is irreducible because C is irreducible, and we have additional sets of transition rates $q(\boldsymbol{n}, \boldsymbol{n} - \boldsymbol{e}_i) = n_i q_{i0}$ and $q(\boldsymbol{n}, \boldsymbol{n} + \boldsymbol{e}_i) = \nu_i$ ($i \in C$). Now the total rate out of state $\boldsymbol{n}$ is

$$q(\boldsymbol{n}) = \sum_{i \in C} \nu_i + \sum_{i \in C} n_i \sum_{j \in C,\, j \neq i} q_{ij} + \sum_{i \in C} n_i q_{i0} = \nu + \sum_{i \in C} n_i q_i \qquad (\boldsymbol{n} \in E),$$

where $q_i = \sum_{j \in S,\, j \neq i} q_{ij}$ ($i \in C$) and $\nu = \sum_{i \in C} \nu_i$.

3 Ensemble proportions

The closed ensemble. Our aim is to demonstrate that the proportion of individuals occupying state j can indeed be approximated by the probability that a single individual is in state j. To this end let $X_j(t) = N_j(t)/n$ be the proportion of individuals in state j at time t, where recall that n is the total number of individuals in the ensemble. Note that $(\boldsymbol{X}(t), t \geq 0)$, where $\boldsymbol{X} = (X_j, j \in S)$, is itself a continuous-time Markov chain. As we are going to vary n, let us make the dependence on n explicit in our notation by writing $X_j^{(n)}$ for X_j and $\boldsymbol{X}^{(n)}$ for $\boldsymbol{X}$, et cetera. Our first theorem establishes that, in the limit as n becomes large, the process of proportions converges (*uniformly in probability* over *finite* time intervals) to a deterministic (that is, non-random) trajectory, being the unique solution of the forward equations (referred to above) of the individual model. The only proviso is that the initial proportions (at $t = 0$) converge.

The proof of Theorem 1, along with the proofs of our other major results, are given in the appendix. Theorem 1, and Theorem 2 below, exploit important results of Kurtz (1970) on approximating limits of pure-jump Markov processes by solutions of ordinary differential equations.

Theorem 1. If $\boldsymbol{X}^{(n)}(0) \to \boldsymbol{a}$ as $n \to \infty$, then, for all $u > 0$, and for every $\epsilon > 0$,

$$\Pr\left(\sup_{0\le t\le u}\left|\boldsymbol{X}^{(n)}(t) - \boldsymbol{p}(t)\right| > \epsilon\right) \to 0 \quad \text{as } n \to \infty,$$

where $\boldsymbol{p}(t) = (p_j(t), j \in S)$ is the unique solution to $\boldsymbol{p}'(t) = \boldsymbol{p}(t)\,Q$ satisfying $\boldsymbol{p}(0) = \boldsymbol{a}$, namely $\boldsymbol{p}(t) = \boldsymbol{a}\exp(tQ)$, where $\exp(\,\cdot\,)$ is the matrix exponential.

The implications for our population models are obvious. For Model 1 and the closed version of Model 2, the proportion of individuals in patch j at time t can be approximated by the probability $p_j(t)$ that an individual occupies patch j at time t, the individual moving amongst the patches according to any given irreducible (finite) Markov chain. For Population 3, the same is true but for the proportions of individuals in the various stages of life. But, as individuals follow a *death process*, we can exhibit the limiting proportions explicitly. For example, if $q_1, q_2, \ldots$ are distinct, then (tweak Exercise 6.8.31 of Grimmett and Stirzaker (2001)) the limiting proportions are given by

$$p_j(t) = \frac{1}{q_j}\sum_{k=j}^{M} q_k e^{-q_k t} \prod_{l=j,\, l\ne k}^{M} \frac{q_l}{q_l - q_k},$$

for $j = 1, \ldots, M$, and $p_0(t) = 1 - \sum_{j=1}^{M} p_j(t)$.

The open ensemble. Now the total number of individuals is not fixed but random, yet there is an analogue of Theorem 1, which provides an approximation that is valid for large external arrival rates (in the specific context of Population 2, heavy immigration from outside the population network). As we are going to let the total exogenous arrival rate ν become large, we make the dependence on ν explicit in our notation. Let $\boldsymbol{v}^{(\nu)} = (v_j^{(\nu)}, j \in C)$, where $v_j^{(\nu)} = \nu_j/\nu$ is the probability that, when a new individual appears, it appears in state j. Our premise is that $\boldsymbol{v}^{(\nu)} \to \boldsymbol{v} = (v_j, j \in C)$ as $\nu \to \infty$. Similarly we let $\boldsymbol{X}^{(\nu)} = (X_j^{(\nu)}, j \in C)$, where $X_j^{(\nu)}(t) = N_j(t)/\nu$.

Theorem 2. Suppose that $\boldsymbol{v}^{(\nu)} \to \boldsymbol{v}$ as $\nu \to \infty$. Then, if $\boldsymbol{X}^{(\nu)}(0) \to \boldsymbol{a}$, we have, for all $u > 0$ and, for every $\epsilon > 0$,

$$\Pr\left(\sup_{0\le t\le u}\left|\boldsymbol{X}^{(\nu)}(t) - \boldsymbol{r}(t)\right| > \epsilon\right) \to 0 \quad \text{as } \nu \to \infty,$$

where $\boldsymbol{r}(t) = (r_j(t), j \in C)$ is the unique solution to $\boldsymbol{r}'(t) = \boldsymbol{v} + \boldsymbol{r}(t)\,Q_C$ satisfying $\boldsymbol{r}(0) = \boldsymbol{a}$, with Q_C being the restriction of transition rate matrix Q to C.

Notice that $\sum_{k\in C} X_k^{(\nu)}(t) = N(t)/\nu$, where $N(t) = \sum_{j\in C} N_j(t)$ is the total number of individuals in the system at time t. Therefore (formally), the proportion $N_j(t)/N(t)$ of individuals in state j can be approximated by $r_j(t)/\sum_{k\in C} r_k(t)$. The quantity $r_j(t)$ is clearly *not* a probability, but does it admit a probabilistic interpretation for the individual

model? It certainly does in the important special case where $\boldsymbol{a} = \mathbf{0}$, which happens if the ensemble system is initially empty. We can solve the differential equation in Theorem 2: since

$$\int_0^t \exp(uQ_C)du\, Q_C = \exp(tQ_C) - I,$$

we have

$$\boldsymbol{r}(t) = \boldsymbol{a}\exp(tQ_C) + \int_0^t \boldsymbol{v}\exp(uQ_C)du.$$

Thus $r_j(t)$ is the sum of two terms, the first being the probability that the individual is in state j at time t after his arrival under the initial distribution $\boldsymbol{a}$, and the second being the expected amount of time the individual spends in state j up to time t under the initial distribution $\boldsymbol{v}$ (the latter being the interpretation of $r_j(t)$ when $\boldsymbol{a} = \mathbf{0}$). This observation also permits us to make a connection with an important result of Kingman (1969). If the system is *initially empty*, then, remarkably, the numbers of individuals $N_j(t)$, $j \in C$, in the various states are *independent* with $N_j(t)$ having a Poisson distribution with mean $\nu r_j(t)$; no approximation is required. It follows that $\mathbb{E}(X_j^{(\nu)}(t)) = r_j(t)$ for all ν. Kingman's result holds in greater (and spectacular) generality; the essential feature is the *independence in the movement of individuals*.

Before leaving this section we remark that Kurtz (1971) provides a *Central Limit Theorem* that allows one to approximate fluctuations of the process of proportions about its macroscopic limit. These fluctuations follow a *Gaussian diffusion* $\boldsymbol{Z} = (\boldsymbol{Z}(t), t \geq 0)$ whose parameters can be determined. For example, in the context of Theorem 1, the process of *scaled proportions* $(\boldsymbol{Z}^{(n)}(t), t \geq 0)$, where

$$\boldsymbol{Z}^{(n)}(t) = \sqrt{n}(\boldsymbol{X}^{(n)}(t) - \boldsymbol{p}(t)),$$

can be shown to converge (over finite time intervals) to $\boldsymbol{Z}$. This implies, for example, that $(\boldsymbol{X}^{(n)}(t_1), \ldots, \boldsymbol{X}^{(n)}(t_k))$ has, for large n, an approximate Gaussian (normal) distribution with mean $\boldsymbol{p}(t)$ and a covariance structure that can be can be exhibited explicitly. Some of the implications of this for parameter estimation in population models are detailed in Ross et al. (2006). For brevity we will not pursue this further here.

4 Equilibrium behaviour

The closed ensemble. When S is irreducible (as it is in the model for Population 1), both the individual model and the ensemble model have a unique equilibrium distribution (because S is finite). For the individual model the equilibrium distribution is the unique probability distribution $\boldsymbol{p} = (p_j, j \in S)$ satisfying $\boldsymbol{p}Q = 0$; p_j is the equilibrium probability that the individual is in state j. For the ensemble model the equilibrium distribution $\boldsymbol{\pi}^{(n)} = (\boldsymbol{\pi}^{(n)}(\boldsymbol{n}), \boldsymbol{n} \in E)$ is given by

$$\boldsymbol{\pi}^{(n)}(\boldsymbol{n}) = n! \prod_{j \in S} \frac{p_j^{n_j}}{n_j!} \qquad (\boldsymbol{n} \in E). \tag{1}$$

To see this, apply Theorem 2.3 of Kelly (1979). So, in equilibrium, the numbers of individuals N_j, $j \in S$, in the various states have a multinomial distribution with parameters $(p_j, j \in S)$; it is as if the n individuals were allocated to the various states independently, with state j being allocated with probability p_j.

Since the two equilibrium distributions are also limiting (they are the limit of the state probabilities as $t \to \infty$), we might expect the equilibrium proportions in the ensemble model to converge to $\boldsymbol{p}$ as $n \to \infty$. Whilst this cannot be deduced from Theorem 1, because uniform convergence in probability is established over *finite* time intervals, it is obviously true from elementary considerations. Indeed a stronger result obtains. Suppose that $(N_j, j \in S)$ has the multinomial distribution (1). Then, N_j has a binomial distribution with parameters n and p_j. Hence, by the Strong Law of Large Numbers, $N_j/n \to p_j$ with probability 1. Indeed, $\boldsymbol{N}/n \to \boldsymbol{p}$ with probability 1. So for large ensembles the equilibrium proportions of individuals in the various states can be approximated by proportions that accord with the equilibrium distribution of the individual model.

The open ensemble. In equilibrium, the numbers of individuals N_j, $j \in C$, in the various states are *independent*, with N_j having a Poisson distribution with mean $\alpha_j := \nu r_j$, where $\boldsymbol{r} = (r_j, j \in C)$ is the unique solution to $\boldsymbol{v} + \boldsymbol{r}\, Q_C = \boldsymbol{0}$. This is a result due to Bartlett (1956) (and consistent with the result of Kingman mentioned earlier). Thus, the total number of individuals $N = \sum_{j \in C} N_j$ has a Poisson distribution with mean $\alpha := \nu \sum_j r_j$ and, given $N = n$, $(N_j, j \in C)$ has the multinomial distribution (1), but with $p_j = r_j / \sum_{k \in C} r_k$ $(= \alpha_j/\alpha)$. It follows, again from elementary considerations, that the equilibrium proportions converge, as ν becomes large, to $\boldsymbol{p}$ (with probability 1), where $\boldsymbol{p} = (p_j, j \in C)$.

5 Quasi-equilibrium behaviour

We restrict our attention to the following case: $S = \{0\} \cup C$, where C is an irreducible finite set, and 0 is an absorbing state which is accessible from C (we have in mind here the closed version of Population model 2). Each of n individuals reaches 0 with probability 1 in finite mean time, and so too the ensemble process reaches its absorbing state $(n, 0, \ldots, 0)$ in finite mean time. However, the ensemble process is *not* irreducible. It has irreducible classes $E_k = \{\boldsymbol{n} \in \{0, 1, \ldots\}^S : \sum_{j \in C} n_j = k\}$ $(k = 0, 1, \ldots, n)$ corresponding to there being k individuals in C, with E_0 having the single member $(n, 0, \ldots, 0)$, the process moving from E_k to E_{k-1} when one of the k individuals that remain in C reaches 0. The classes are therefore arranged as follows: $E_n \succ E_{n-1} \succ \cdots \succ E_1 \succ E_0$ ($D \succ E$ means $q_{ij} > 0$ for at least one i in D and at least one j in E).

Theorem 1 tells us that at any time t the ensemble proportions can be approximated by $\boldsymbol{p}(t) = \boldsymbol{a} \exp(tQ)$, the entries of $\boldsymbol{a}$ being the (large-n) initial proportions of individuals in the various states. But, what happens when t is large? Although both processes have degenerate equilibrium distributions assigning all mass to the absorbing state, they can settle down to a quasi equilibrium before extinction occurs. For this reason we usually model the long-term behaviour using a *limiting conditional* (or *quasi-stationary*) distribution. The idea is as follows. First suppose we have complete information: we know the probability $p_x(t)$ that

the process is in state x at time t, for *all* states x and for *every* time t. If at a fixed time t we observe that the process has not reached the absorbing state z, the appropriate probability to evaluate is $p_x(t)/(1-p_z(t))$, the probability that the process is in state x *conditional on* it not having been absorbed by time t; to be emphatic, this is the best possible information we could have about the chance of being in state x having observed that the process has not been absorbed. This being the case *for every* time t, it is natural to evaluate this probability in the limit as $t \to \infty$ to obtain information about the long-term behaviour of the process conditional on non-absorption (hence the term limiting conditional distribution). Under mild conditions this limiting conditional distribution (LCD) can be determined, directly from the transition matrix restricted to the transient states, as an appropriately normalized left eigenvector corresponding to the eigenvalue with maximum real part; it is not necessary for us to know $p_x(t)$ explicitly.

If the transient states form an irreducible class C, as is the case in the individual model, we can apply the classical result of Darroch and Seneta (1967). The restricted transition matrix $Q_{_C}$ has an eigenvalue $-\alpha$ with maximum real part, which is real, strictly negative and simple (multiplicity 1), and corresponding to it are strictly positive left and right eigenvectors, $\boldsymbol{u} = (u_j, j \in C)$ and $\boldsymbol{x}^\top = (x_j, j \in C)$. Furthermore, $p_j(t)/(1-p_0(t)) \to \pi_j$, where $\pi_j = u_j / \sum_{k \in C} u_k$ $(j \in C)$. The quantity α is called the *decay parameter*, because $p_j(t) = O(e^{-\alpha t})$ (see Kingman, 1963). For the (reducible) ensemble model, we must appeal to recent results of Van Doorn and Pollett (2008) to establish the existence of a LCD over the set of non-absorbing states $C_{_E} := E \backslash E_0 = \cup_{k=1}^n E_k$. Indeed, the ensemble process is an example of a *quasi-death process*, and Theorem 7 of Van Doorn and Pollett (2008) can be applied. The main result of this section shows that the LCD of the ensemble process assigns positive probability only to those states in E_1, being precisely the "unit vectors" $(n-1, \boldsymbol{e}_j) = (n-1, 0, \ldots, 0, 1, 0, \ldots, 0)$ $(j \in C)$ corresponding to the single remaining individual being in state j (and hence $n-1$ in state 0).

Theorem 3. The ensemble process admits a LCD $\boldsymbol{u}_{_E} = (u(\boldsymbol{m}), \boldsymbol{m} \in C_{_E})$. It assigns all its mass to E_1, with $u((n-1, \boldsymbol{e}_j)) = \pi_j$ $(j \in C)$, where $(\pi_j, j \in C)$ is the LCD of the individual process.

Whilst this result is not surprising (because each of the n individuals is lost at the same rate), we see that again ensemble proportions can be approximated by probabilities obtained from the individual model. It accords with Theorem 1; since the proportion $X_j^{(n)}(t)$ of individuals in state j at time t is approximately $p_j(t)$, we should expect $N_j(t)/(n - N_0(t)) = X_j^{(n)}(t)/(1 - X_0^{(n)}(t))$, the proportion in state j of those "alive" at time t, to be close to $p_j(t)/(1-p_0(t))$ $(j \in C)$.

Van Doorn and Pollett (2008) contains enough technology to extend Theorem 3 to the case where C is *not* irreducible, but where the (sole) absorbing state 0 is accessible from all states in C *for which the initial distribution assigns positive probability* (however, note that now the ensemble process is *not* a quasi-death process). We will content ourselves with a result for Population 3 (which of course has C reducible). Recall that there are M stages of life, with M being the first stage and 1 being the last, and with state 0 representing

death. Individuals move from stage s to stage $s-1$ at rate $q_s > 0$. Theorem 6 of Van Doorn and Pollett (2008) gives the LCD of the individual process. Let $C = \{1, \ldots, M\}$ and $q = \min_{s \in C} q_s$. One might have expected that those stages s for which $q_s = q$ (s could in fact be unique) would act as a "bottleneck" in that most individuals would be expected to be in those stages. But, it turns out that the limiting probability of being in stage s, conditional on being alive, is 0 *unless* $s \leq \kappa$, where $\kappa = \min\{k : q_k = q\}$.

Theorem 4. For the death process describing Population 3, the ensemble process admits a LCD $\boldsymbol{u}_{E} = (u(\boldsymbol{m}), \boldsymbol{m} \in C_{E})$, which does not depend on the initial distribution over states. It assigns all its mass to E_1, with $u((n-1, \boldsymbol{e}_s)) = \pi_s$ $(s \in C)$, where $(\pi_s, s \in C)$, the LCD of the individual process, is given by

$$\pi_s = \begin{cases} \dfrac{q}{q_s} \displaystyle\prod_{i=1}^{s-1} \left(1 - \frac{q}{q_i}\right) & \text{if } s \leq \kappa \\ 0 & \text{if } s > \kappa. \end{cases}$$

Remark. This result is at variance with Proposition 1 of Steinsaltz and Evans (2004), which, while formulated quite generally, is correct only if one assumes irreducibility.

6 Survival

We conclude proceedings with a result that brings together many of the ideas of the paper.

One of the many characterizations of the LCD is that if a Markov process starts in state $x \in \mathcal{X}$ (some irreducible set) with probability π_x, where $\boldsymbol{\pi} = (\pi_x, x \in \mathcal{X})$ is the LCD (conditional on not having left $\mathcal{X}$), then the chance that the process is in state $y \in \mathcal{X}$ at time t is $\pi_y e^{-\alpha t}$, where α is the decay parameter of $\mathcal{X}$. This result holds in great generality, and, depending on the setting, usually results from a simple calculation; for a recent discussion, see Steinsaltz and Evans (2004). It follows immediately that the time to exit from $\mathcal{X}$ is exponentially distributed with parameter α (because the probability of still being in $\mathcal{X}$ at time t is $\sum_{y \in \mathcal{X}} \pi_y e^{-\alpha t} = e^{-\alpha t}$). The implications for our models are obvious, the result for the ensemble being not particularly surprising because the ensemble survives until the last individual reaches 0. However, using an elaboration of the arguments presented in Section 3 of Derman (1955) (see also the proof of Theorem 2 of Brown (1970) for a more transparent discrete-time parallel), we can prove the following. The setup is $S = \{0\} \cup C$, where C is an irreducible finite set with decay parameter α and 0 is an absorbing state which is accessible from C.

Theorem 5. Let $\boldsymbol{\pi} = (\pi_j, j \in C)$ be the LCD of the individual process. If the initial numbers $N_j(0)$, $j \in C$, are chosen independently with $N_j(0)$ having a Poisson distribution with mean π_j, then, for all $t > 0$, $N_j(t)$, $j \in C$, are independent with $N_j(t)$ having a Poisson distribution with mean $\pi_j e^{-\alpha t}$.

For the aficionados. This result holds in much greater generality; C need not be finite, Q could be explosive, $\boldsymbol{\pi} = (\pi_j, j \in C)$ could be any α-subinvariant measure and, more remarkably still, $\boldsymbol{\pi}$ need not be finite (we could have $\sum_{j\in C} \pi_j = \infty$); for definitions, see Anderson (1991).

7 Conclusion

We have presented results on Markovian population processes that allow ensemble quantities, such as the proportions of individuals of different types, to be approximated by quantities derived from simpler individual-level models. Two theorems were presented that delineate conditions under which the random process of ensemble proportions can be approximated by probabilities obtained from the underlying models of individual behaviour. Results were given for time-dependent behaviour, as well as equilibrium and quasi-equilibrium behaviour. Particular attention was paid to the case where an individual moves through various stages of life, and the ensemble model keeps track of the numbers of individuals in each stage; the quasi-stationary distribution of the ensemble process was exhibited explicitly. The latter results apply more widely than the present ecological context: where there is stratification of risk, such as in epidemiological and clinical contexts.

8 Appendix

This section is necessarily technical. It contains proofs of our major results.

Proof of Theorem 1. The ensemble model is *density dependent* in the sense of Kurtz (1970): there is an open subset of $\mathbb{R}^S$, namely $E = (0,1)^S$, and functions $f : E \times \mathbb{Z}^S \to \mathbb{R}$ with the property that $q(\boldsymbol{n}, \boldsymbol{n}+\boldsymbol{l}) = nf(\boldsymbol{n}/n, \boldsymbol{l})$, $\boldsymbol{n}, \boldsymbol{n}+\boldsymbol{l} \in \mathbb{Z}^S$. Clearly $f(\boldsymbol{x}, \boldsymbol{e}_j - \boldsymbol{e}_i) = x_i q_{ij}$, $j \neq i$. Our theorem is then proved by applying Theorem 3.1 of Kurtz's paper. We first define $F : E \to \mathbb{R}^S$ by $F(\boldsymbol{x}) = \sum_{\boldsymbol{l}} \boldsymbol{l} f(\boldsymbol{x}, \boldsymbol{l})$. This function is required to be Lipschitz continuous on E (clearly true because F has bounded first partial derivatives on E), thus guaranteeing a solution $\boldsymbol{p}(t)$ to $\boldsymbol{p}'(t) = F(\boldsymbol{p}(t))$, $\boldsymbol{p}(t) \in E$, $0 \leq t \leq u$, satisfying $\boldsymbol{p}(0) = \boldsymbol{a}$. Next, two technical conditions must be fulfilled: (i) $\sup_{\boldsymbol{x}\in E} \sum_{\boldsymbol{l}\neq\boldsymbol{0}} |\boldsymbol{l}| f(\boldsymbol{x}, \boldsymbol{l}) < \infty$ and (ii) $\lim_{d\to\infty} \sup_{\boldsymbol{x}\in E} \sum_{|\boldsymbol{l}|>d} |\boldsymbol{l}| f(\boldsymbol{x}, \boldsymbol{l}) = 0$. They are trivially satisfied because S is a finite set.

Kurtz's theorem states that $\boldsymbol{X}^{(n)}(t)$ converges uniformly in probability over finite time intervals to $\boldsymbol{p}(t)$ as $n \to \infty$, so it remains for us to show that $F(\boldsymbol{x}) = \boldsymbol{x}Q$. This is straightforward:

$$
\begin{aligned}
F(\boldsymbol{x}) &= \sum_{i\in S} \sum_{j\neq i} (\boldsymbol{e}_j - \boldsymbol{e}_i) f(\boldsymbol{x}, \boldsymbol{e}_j - \boldsymbol{e}_i) = \sum_{i\in S} x_i \sum_{j\neq i} (\boldsymbol{e}_j - \boldsymbol{e}_i) q_{ij} \\
&= \sum_{i\in S} x_i \left(\sum_{j\neq i} q_{ij}\boldsymbol{e}_j - q_i \boldsymbol{e}_i \right) = \sum_{k\in S} x_k \sum_{i\in S} q_{ki} \boldsymbol{e}_i .
\end{aligned}
$$

Hence, $F(\boldsymbol{x}) = \boldsymbol{x}Q$, because (elementwise) $F_i(\boldsymbol{x}) = \sum_{k\in S} x_k q_{ki}$ $(i \in S)$.

Proof of Theorem 2. In addition to $f(\boldsymbol{x}, \boldsymbol{e}_j - \boldsymbol{e}_i) = x_i q_{ij}$, $j \neq i$, we have $f(\boldsymbol{x}, -\boldsymbol{e}_i) = x_i q_{i0}$ and $f(\boldsymbol{x}, \boldsymbol{e}_j) = v_j$ (asymptotically). The model is not strictly density dependent, but *asymptotically density dependent* in the sense of Pollett (1990), and the present result follows from Theorem 3.1 of Pollett's paper, which shows that $\boldsymbol{X}^{(\nu)}(t)$ converges uniformly in probability over finite time intervals to $\boldsymbol{r}(t)$ as $\nu \to \infty$, the unique solution to $\boldsymbol{r}'(t) = F(\boldsymbol{r}(t))$, $\boldsymbol{r}(t) \in E$, $0 \leq t \leq u$. We may evaluate $F(\boldsymbol{x})$ as follows:

$$\begin{aligned} F(\boldsymbol{x}) &= \sum_{i \in C} \sum_{j \in C,\, j \neq i} (\boldsymbol{e}_j - \boldsymbol{e}_i) f(\boldsymbol{x}, \boldsymbol{e}_j - \boldsymbol{e}_i) - \sum_{i \in C} \boldsymbol{e}_i f(\boldsymbol{x}, -\boldsymbol{e}_i) + \sum_{j \in C} \boldsymbol{e}_j f(\boldsymbol{x}, \boldsymbol{e}_j) \\ &= \sum_{i \in C} x_i \sum_{j \in C,\, j \neq i} (\boldsymbol{e}_j - \boldsymbol{e}_i) q_{ij} - \sum_{i \in C} x_i \boldsymbol{e}_i q_{i0} + \sum_{j \in C} \boldsymbol{e}_j v_j \\ &= \sum_{i \in C} x_i \left(\sum_{j \in C,\, j \neq i} q_{ij} \boldsymbol{e}_j - q_i \boldsymbol{e}_i \right) + \sum_{j \in C} \boldsymbol{e}_j v_j = \sum_{k \in C} x_k \sum_{i \in C} q_{ki} \boldsymbol{e}_i + \sum_{i \in C} \boldsymbol{e}_i v_i, \end{aligned}$$

where recall that $q_i = -q_{ii} = \sum_{j \in S,\, j \neq i} q_{ij}$. Hence, $F(\boldsymbol{x}) = \boldsymbol{v} + \boldsymbol{x} Q_{C}$, because (elementwise) $F_i(\boldsymbol{x}) = v_i + \sum_{k \in C} x_k q_{ki}$ $(i \in C)$.

Proof of Theorem 3. Let Q_k be the restriction to E_k of transition rate matrix Q_{E} of the ensemble process and let $-\alpha_k$ be the eigenvalue of Q_k with maximum real part $(k = 1, \ldots, n)$. Then, $\alpha_k = k\alpha$. To see this, observe that $\alpha_k = \lim_{t \to \infty} -(1/t) \log \Pr(T > t)$, where T is the time to first exit of the process from E_k (see Kingman, 1963); the limit does not depend on the initial distribution over states. However, $T = \min\{T_1, \ldots, T_k\}$, where T_i is the time it takes individual i to reach 0, and, since the individuals move independently, $\Pr(T > t) = \prod_{i=1}^{k} \Pr(T_i > t)$. Since the individuals move according to Q, $-(1/t) \log \Pr(T_i > t) \to \alpha$ as $t \to \infty$. Hence, $\alpha_k = k\alpha$. It follows immediately that $-\alpha$ is the eigenvalue of Q_{E} with maximum real part, and, moreover, that its algebraic, and hence geometric, multiplicity is equal to 1. We may therefore appeal directly to Theorem 5 of Van Doorn and Pollett (2008), which implies that the LCD of the ensemble process exists provided the initial distribution assigns mass to at least one of $E_1, \ldots, E_n$ (we assume that all n individuals are present initially, and so all this mass is assigned to E_n). Furthermore, the LCD is the unique *non-negative* solution $\boldsymbol{u}_{E} = (u(\boldsymbol{m}), \boldsymbol{m} \in C_{E})$ to $\boldsymbol{u}_{E} Q_{E} = -\alpha \boldsymbol{u}_{E}$ with $\boldsymbol{u}_{E} \boldsymbol{1}^{\top} = 1$. So, it remains to show that the given $\boldsymbol{u}_{E}$ satisfies these equations: $u(\boldsymbol{m}) = 0$ unless $\boldsymbol{m} \in E_1$, in which case $u((n-1, \boldsymbol{e}_j)) = \pi_j$ $(j \in C)$, where $\boldsymbol{\pi} = (\pi_j, j \in C)$ is the LCD of the individual process, that is, the unique (strictly positive) solution to $\boldsymbol{\pi} Q = -\alpha \boldsymbol{\pi}$ with $\boldsymbol{\pi} \boldsymbol{1}^{\top} = 1$. Writing out the latter we get

$$\sum_{i \in C,\, i \neq j} \pi_i q_{ij} = (q_j - \alpha) \pi_j \qquad (j \in C),$$

where $q_j = \sum_{k \in S,\, k \neq j} q_{jk}$, while the former eigenvector equation may be written as

$$\sum_{\boldsymbol{m} \in E_n,\, \boldsymbol{m} \neq \boldsymbol{n}} u(\boldsymbol{m}) q(\boldsymbol{m}, \boldsymbol{n}) = (q(\boldsymbol{n}) - \alpha) u(\boldsymbol{n}), \tag{2}$$

$\boldsymbol{n} \in E_n$, and, for $k = 1, \ldots, n-1$,

$$\sum_{\boldsymbol{m} \in E_{k+1}} u(\boldsymbol{m}) q(\boldsymbol{m}, \boldsymbol{n}) + \sum_{\boldsymbol{m} \in E_k,\, \boldsymbol{m} \neq \boldsymbol{n}} u(\boldsymbol{m}) q(\boldsymbol{m}, \boldsymbol{n}) = (q(\boldsymbol{n}) - \alpha) u(\boldsymbol{n}) \qquad (\boldsymbol{n} \in E_k). \tag{3}$$

Clearly (2), and (3) for $k = 2, \ldots, n$, are satisfied because $u(\boldsymbol{m}) = 0$ when $\boldsymbol{m} \in \cup_{k=2}^{n} E_k$, and the remaining equation ($k = 1$) will hold if and only if

$$\sum_{j \in C} \sum_{i \in C,\, i \neq j} u(\boldsymbol{n} + \boldsymbol{e}_j - \boldsymbol{e}_i)(n_j + 1) q_{ji} = \left(\sum_{k \in C} n_k q_k - \alpha \right) u(\boldsymbol{n}) \qquad (\boldsymbol{n} \in E_1), \tag{4}$$

remembering that the total rate out of state $\boldsymbol{n} \in E_k$ is $q(\boldsymbol{n}) = \sum_{i \in C} n_i q_i$. But, $\boldsymbol{m} \in E_1$ if and only if $\boldsymbol{m} = (n-1, \boldsymbol{e}_i)$ for some $i \in C$, and so we require

$$\sum_{i \in C,\, i \neq j} u((n-1, \boldsymbol{e}_i)) q_{ij} = (q_j - \alpha)\, u((n-1, \boldsymbol{e}_j)),$$

for $j \in C$. On substituting $u((n-1, \boldsymbol{e}_i)) = \pi_i$, we find that

$$\sum_{i \in C,\, i \neq j} \pi_j q_{ij} = (q_j - \alpha)\, \pi_j \qquad (j \in C),$$

which of course is the individual process eigenvector equation.

Proof of Theorem 4. This is similar to the proof of Theorem 3, but now we appeal to Theorem 3 of Van Doorn and Pollett (2008) for important information about accessibility; the initial distribution must assign all mass to $\cup_{k \geq \kappa_{_E}} E_k$, where $\kappa_{_E}$ is the smallest index k of those α_k which share the smallest value among $\alpha_1, \ldots, \alpha_n$. However, $\alpha_k = k\alpha$, implying that $\kappa_{_E} = 1$, and so the accessibility condition is satisfied automatically (we have assumed that the ensemble process starts in E_n). Evaluating the LCD as directed by Theorem 3 of Van Doorn and Pollett (2008) amounts to calculating $\boldsymbol{u}_{_E}$ as in the proof of our Theorem 3, but now the individual process is a simple death-process. Therefore, $\alpha = q$ and the unique (strictly positive) solution $\boldsymbol{\pi} = (\pi_j, j \in C)$ to $\boldsymbol{\pi} Q = -\alpha \boldsymbol{\pi}$ with $\boldsymbol{\pi} \mathbf{1}^\top = 1$ is given in Theorem 6 of Van Doorn and Pollett (2008), and written out in the statement of our theorem.

Acknowledgement. This work is supported by the Australian Research Council Centre of Excellence for Mathematics and Statistics of Complex Systems.

References

[1] Anderson, W.J. (1991) *Continuous-Time Markov Chains: An Applications-Oriented Approach*, Springer-Verlag, New York.

[2] Bartlett, M.S. (1956) Deterministic and stochastic models for recurrent epidemics. In *Proceedings of the 3rd Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, Vol. IV, Berkeley and Los Angeles, University of California Press, pp. 81–109.

[3] Brown, M. (1970) A property of Poisson processes and its application to macroscopic equilibrium of particle systems. *The Annals of Mathematical Statistics*, 41, 1935–1941.

[4] Darroch J.N. and Seneta, E. (1967) On quasi-stationary distributions in absorbing continuous-time finite Markov chains. *Journal of Applied Probability*, 4, 192–196.

[5] Derman, C. (1955) Some contributions to the theory of denumerable Markov chains. *Transactions of the American Mathematical Society*, 79, 541–555.

[6] Doob, J.L. (1953) *Stochastic Processes*, John Wiley & Sons Inc., New York.

[7] Gilpin, M.E. and Hanski, I. (1991) *Metapopulation Dynamics*, Academic Press, New York.

[8] Grimmett, G.R. and Stirzaker, D.R. (2001) *Probability and Random Processes*, 3rd edition, Oxford University Press, New York.

[9] Gyllenberg, M. and Silvestrov, D.S. (1994) Quasi-stationary distributions of a stochastic metapopulation model. *Journal of Mathematical Biology*, 33, 35–70.

[10] Kelly, F.P. (1979) *Reversibility and Stochastic Networks*, Wiley, Chichester.

[11] Kingman, J.F.C. (1963) The exponential decay of Markov transition probabilities. *Proceedings of the London Mathematical Society*, 13, 337–358.

[12] Kingman, J.F.C. (1969) Markov population processes. *Journal of Applied Probability*, 6, 1–18.

[13] Kurtz, T.G. (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of Applied Probability*, 7, 49–58.

[14] Kurtz, T.G. (1971) Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8, 344–356.

[15] Norris, J.R. (1997) *Markov Chains*, Cambridge University Press, Cambridge.

[16] Pollett, P.K. (1990) On a model for interference between searching insect parasites. *Journal of the Australian Mathematic Society, Series B*, 31, 133–150.

[17] Pollett, P.K. (1999) Modelling quasi-stationary behaviour in metapopulations. *Mathematics and Computers in Simulation*, 48, 393–405.

[18] Ross, J.V., Taimre, T. and Pollett, P.K. (2006) On parameter estimation in population models. *Theoretical Population Biology*, 70, 498–510.

[19] Serfozo, R. (1999) *Introduction to Stochastic Networks*, Springer-Verlag, New York.

[20] Steinsaltz, D. and Evans, S.N. (2004) Markov mortality models: implications of quasistationarity and varying initial distributions. *Theoretical Population Biology*, 65, 319–337.

[21] Van Doorn, E.A. and Pollett, P.K. (2008) Survival in a quasi-death process. *Linear Algebra and its Applications*, 429, 776–791.

[22] Whittle, P. (1967) Nonlinear migration processes. *Bulletin of the International Institute of Statistics*, 42, 642–647.