The Annals of Probability
1999, Vol. 27, No. 1, 166–205

PARTICLE REPRESENTATIONS FOR MEASURE-VALUED POPULATION MODELS[1]

BY PETER DONNELLY[2] AND THOMAS G. KURTZ[3]

University of Oxford and University of Wisconsin

Models of populations in which a *type* or *location*, represented by a point in a metric space E, is associated with each individual in the population are considered. A population process is *neutral* if the chances of an individual replicating or dying do not depend on its type. Measure-valued processes are obtained as infinite population limits for a large class of neutral population models, and it is shown that these measure-valued processes can be represented in terms of the total mass of the population and the de Finetti measures associated with an E^{∞}-valued *particle model* $X = (X_1, X_2, \dots)$ such that, for each $t \geq 0$, $(X_1(t), X_2(t), \dots)$ is exchangeable. The construction gives an explicit connection between genealogical and diffusion models in population genetics. The class of measure-valued models covered includes both neutral Fleming–Viot and Dawson–Watanabe processes. The particle model gives a simple representation of the Dawson–Perkins historical process and Perkins's historical stochastic integral can be obtained in terms of classical semimartingale integration. A number of applications to new and known results on conditioning, uniqueness and limiting behavior are described.

1. Introduction. We begin by considering two models for the evolution of a finite population. Although we concentrate mainly on continuous-time processes, we indicate the analogous results for discrete-time processes later.

1.1. *Model I.* Let $N(t)$ denote the total size of a population at time t, let $N_b(t)$ denote the number of births up to and including time t and let $N_d(t)$ denote the number of deaths, so

$$N(t) = N(0) + N_b(t) - N_d(t).$$

(Note that we are assuming that N, N_b and N_d are right continuous.) We allow simultaneous and/or multiple births and deaths, but we assume that all the births that happen simultaneously come from the same *parent*. At a birth event, the parent is selected at random (by which here and throughout

---

Received August 1996; revised June 1998.

[1]Part of the work presented here was done while the authors were in residence at the Institute for Mathematics and its Applications.

[2]Supported in part by NSF Grant DMS-95-05129, University of Chicago Block Fund and UK EPSRC Grant B/AF 1255.

[3]Supported in part by NSF Grants DMS-92-04866 and DMS-95-04323.

AMS 1991 *subject classifications*. 60J25, 60K35, 60J70, 60J80, 92D10.

Key words and phrases. Fleming–Viot process, Dawson–Watanabe process, superprocess, measure-valued diffusion, exchangeability, genealogical processes, coalescent, historical process, conditioning.

we mean uniformly at random) from the population; at a death event, the individuals that are eliminated from the population are selected at random, that is, if there are k deaths, the $\binom{N(t-)}{k}$ possible subsets of the population immediately prior to the death event are equally likely to be eliminated. For definiteness, assume that if birth and death events happen simultaneously, then the individuals who die are removed from the population before the parent of the new individuals is selected. We assume that at each time t, each individual in the population has a *type* or *location* in a space E, which we take to be a complete separable metric space. (Typically, in a genetic model, type is the appropriate interpretation, while in a model of a dispersing population, location is appropriate.) We assume that at a birth event, the *offspring* are given the same type as the parent and in between birth and death events, the types evolve as independent, E-valued Markov processes corresponding to a specified generator B. Therefore, the population at time t can be described by a vector $(Y_1(t), \dots, Y_{N(t)})$ in $E^{N(t)}$ in which we order the population by decreasing age or, since age and hence the above order do not play a role in the birth and death events, by the empirical measure

$$Z^I(t) = \sum_{i=1}^{N(t)} \delta_{Y_i(t)}.$$

Note that if N is Markov, then Z^I will be also. This model is *neutral* in the sense that the type of an individual does not affect its chances of dying or giving birth.

1.2. *Model II*. The population size is defined as in Model I, and in between birth and death events, the types or locations of the individuals evolve as independent Markov processes with generator B; however, the order of $(X_1(t), \dots, X_{N(t)}(t))$ plays a significant role in the birth and death events. The description of a death event is simple: the individuals removed are the individuals with the highest indices. Birth events, however, are more complex. Suppose there is a birth event at time t at which there are k offspring. The type of the offspring will again be the type of the parent. We must specify how to select the parent and how to specify the indices of the population after the birth event. Select $k+1$ indices, $i_1 < \cdots < i_{k+1}$, at random from $\{1, \dots, N(t)\}$. Note that the smallest of these indices, i_1, will be the index of some individual in the population immediately before the birth event. That individual will be the parent. After the birth event, the parent and the k offspring will be indexed by $i_1, \dots, i_{k+1}$. The remaining $N(t) - (k+1)$ individuals are reindexed by $\{1, \dots, N(t)\} - \{i_1, \dots, i_{k+1}\}$, maintaining their previous order. For example, if $k = 1$, then $X_i(t) = X_i(t-)$ for $i < i_2$, $X_{i_2}(t) = X_{i_1}(t-)$, and $X_i(t) = X_{i-1}(t-)$ for $i > i_2$.

Model II may seem strange; however, the following theorem explains its interest.

THEOREM 1.1. *Suppose that the initial population vectors* $(Y_1(0), \dots, Y_{N(0)}(0))$ *in Model I and* $(X_1(0), \dots, X_{N(0)}(0))$ *in Model II have the same* *exchangeable distribution and define*

$$Z^{II}(t) = \sum_{i=1}^{N(t)} \delta_{X_i(t)}.$$

Then Z^{II} *has the same distribution as* Z^I *and, for each* $t \geq 0$, $(X_1(t), \dots, X_{N(t)}(t))$ *is exchangeable.*

Theorem 1.1 is proved in Section 2 using a coupling argument. We also give the corresponding result for models with discrete generations. Intuitively, Model II can be obtained from Model I by looking into the future and ordering the individuals in terms of the time of survival of their line of descent. Neutrality assures that, conditioned on all information up to time t, each particle alive at time t has the same chance of having the longest line of descent, the second longest line of descent, etc. Consequently, this ordering is a random permutation of $(Y_1(t), \dots, Y_{N(t)})$. For example, this interpretation explains why in Model II we require the individuals with highest index to die first. The randomness of the permutation explains the exchangeability property for $(X_1(t), \dots, X_{N(t)})$.

Our primary interest in Theorem 1.1 is its implications for large population approximations. Special cases of Model I include neutral Moran models from population genetics [let $N_b(t) \equiv N_d(t)$] and branching Markov processes in which the offspring distribution does not depend on the location of the parent. Consequently, large population approximations of the measure-valued process $Z^I = Z^{II}$ include neutral Fleming–Viot processes and a large class of Dawson–Watanabe (super) processes. [See Dawson (1993) for a general discussion of these processes.]

In Section 3, for a sequence of these models, we assume that the normalized population size $P^n = n^{-1}N^n$ converges in distribution to a process P and show that, under additional technical assumptions, Model II $(X_1^n, \dots, X_{N_n}^n)$ converges to a process with values in E^∞. The limiting process has the property that, for each $t \geq 0$, $(X_1(t), X_2(t), \dots)$ is exchangeable and the sequence of normalized empirical measures

$$\frac{1}{n} \sum_{k=1}^{N^n(t)} \delta_{X_k^n(t)}$$

converges in distribution to PZ, where Z is the de Finetti measure

$$Z(t) = \lim_{m \to \infty} \frac{1}{m} \sum_{k=1}^{m} \delta_{X_k(t)}.$$

Section 4 discusses the martingale properties of the infinite population models and, in particular, gives conditions under which the measure valued process PZ is the unique solution of a martingale problem. Section 5 describes how the population genealogy is embedded in the model. In particular, the Dawson–Perkins historical process is constructed.

Section 6 includes a number of applications of the E^∞-representation of the measure-valued processes. In particular, generalizations of a variety of results on Dawson–Watanabe and Fleming–Viot processes can be obtained.

One of the advantages of the E^∞-valued limit process over the simpler measure-valued limit is that the E^∞-valued process retains information about the ancestral relationships of the individual particles. In the Fleming–Viot (genetic) setting, the model incorporates the full genealogical (coalescent) tree for the population at each time t. This fact is explored in more detail for a related but somewhat different construction in Donnelly and Kurtz (1996). In the Dawson–Watanabe setting, the model incorporates the "historical process' as studied by Dawson and Perkins (1991) and Perkins (1992, 1995) (cf. Section 5.2). In particular, we are able to represent the stochastic equation given by historical Brownian motion studied by Perkins in terms of an infinite system of ordinary Itô equations (cf. Section 6.5).

1.3. *Conditions on the type/location process.* Throughout we will assume that $P(t, x, \Gamma)$ is the transition function for a Markov process with sample paths in $D_E[0,\infty)$, where (E, r) is a complete, separable metric space. The corresponding semigroup on $B(E)$ is defined by

$$T(t)f(x) = \int f(y)P(t, x, dy),$$

and the weak infinitesimal operator [in the sense of Dynkin (1965)] is defined by

$$Bf = bp\text{-}\lim_{t\to 0} \frac{T(t)f - f}{t}$$

when the limit exists. Let P_x denote the distribution on $D_E[0,\infty)$ corresponding to the Markov process with initial position x. Under these assumptions, we have the following lemmas.

LEMMA 1.2. *There exists a countable subset $D \subset \mathscr{D}(B)$ that is separating in $\mathscr{P}(E)$ in the sense that, for $\mu, \nu \in \mathscr{P}(E)$, $\int f\,d\mu = \int f\,d\nu$ for all $f \in D$, implies that $\mu = \nu$.*

PROOF. See Donnelly and Kurtz (1996), Lemma 1.1.

Let $D = \{f_k,\ k \ge 1\} \subset \mathscr{D}(B)$ be separating and assume that $\|f_k\| \le 1$. Define the metric

$$\rho_B(\mu, \nu) = \sum_{k=1}^{\infty} \frac{1}{2^k}\left|\int f_k\,d\mu - \int f_k\,d\nu\right| \tag{1.1}$$

on $\mathscr{P}(E)$. The notation ρ_B is not really appropriate, since the metric depends on D rather than B. But B is part of the primary "data" for the process and the main restriction on D is that $D \subset \mathscr{D}(B)$. Consequently, it seems more important to emphasize the connection to B. Typically, D can be taken to be convergence determining, and the topology generated by ρ_B will be the weak topology. In particular, if E is locally compact and $\mathscr{D}(B)$ is dense in $\hat{C}(E)$, D can be selected to be convergence determining. In general, it is desirable to select D so as to make the topology generated by the metric as strong as possible. [See Donnelly and Kurtz (1996), Remark 2.5, for an example in which the topology is not the weak topology.]

LEMMA 1.3. *There exists a probability space* $(\Omega_0, \mathscr{F}_0, P_0)$ *and a measurable mapping* $M\colon E \times [0,\infty) \times \Omega_0 \to E$ *such that, for each* $x_0 \in E$, $\chi(t) = M(x_0, t, \cdot)$ *is a Markov process with transition function* $P(t, x, \Gamma)$ *and* $\chi(0) = x_0$. *If* $x \to P_x$ *is weakly continuous, then the mapping from E into* $D_E[0,\infty)$ *given by* $x \to M(x, \cdot, \omega)$ *can be taken to be almost surely continuous at each* $x \in E$. *If* $x \to P_x$ *is weakly continuous and* $P_x\{\chi(t) = \chi(t-)\} = 1$ *for all* $x \in E$ *and* $t \geq 0$ (*that is,* χ *has no fixed points of discontinuity*), *then for each* $(t_0, x_0) \in [0,\infty) \times E$, *with probability* 1, *the mapping* $(t, x) \to M(t, x, \cdot)$ *is continuous at* (t_0, x_0).

PROOF. The lemma follows by the construction of Blackwell and Dubins (1983) and the continuous mapping theorem. □

2. A coupling of finite population models.

2.1. *A coupling lemma.* The proof of Theorem 1.1 relies on a coupling of the two models

$$\big(Y_1(t), \dots, Y_{N(t)}\big) = \big(X_{\theta_1(t)}(t), \dots, X_{\theta_{N(t)}(t)}(t)\big),$$

in which $\theta(t)$ is uniformly distributed over all permutations of $(1, \dots, N(t))$ and is independent of $\mathscr{F}_t^Y = \sigma(Y(s)\colon s \leq t)$. θ will change only at birth/death event times, and we next describe an inductive procedure for its construction in a somewhat more general context.

For $n > 0$, let S_n denote the collection of permutations of $\{1, \dots, n\}$, let P_n denote the collection of all subsets of $\{1, \dots, n\}$ and let $P_{n,k} \subset P_n$ be the subcollection of subsets with cardinality k. We think of a permutation as a mapping from $\{1, \dots, n\}$ onto $\{1, \dots, n\}$.

Let n_0 be a positive integer and let $\{k_m\} \subset \mathbb{Z}$, $k_m \neq 0$. Define $n_m = n_{m-1} + k_m$ and $m^* = \min\{m\colon n_m \leq 0\}$. We construct a sequence of random permutations $\{\theta_m\}$, that is, a sequence of random variables with θ_m taking values in S_{n_m}, and a sequence of random subsets $\{\xi_m\}$, ξ_m taking values in $P_{n_{m-1}}$, in the following way. Let θ_0 be uniformly distributed over S_{n_0}. Let $\{\eta_m\colon 1 \leq m < m^*, k_m > 0\}$ be independent random sets, independent of θ_0, such that η_m is uniformly distributed over P_{n_m, k_m+1}, and let $\{\sigma_m\colon 1 \leq m < m^*, k_m > 0\}$ be independent random permutations, independent of θ_0 and $\{\eta_m\}$ such that σ_m is uniformly distributed over S_{k_m+1}. $\eta_m(i)$ will denote the ith largest element in η_m. Proceeding inductively, assume that θ_{m-1} is defined, $m < m^*$, and that θ_{m-1} is uniformly distributed over $S_{n_{m-1}}$.

If $k_m < 0$ (corresponding to a death event), let $\xi_m = \theta_{m-1}^{-1}(n_m + 1, \dots, n_{m-1})$ and let θ_m be the permutation in S_{n_m} with the same order as θ_{m-1} restricted to $\{1, \dots, n_{m-1}\} - \xi_m$. Note that θ_m is uniformly distributed over S_{n_m} and is independent of ξ_m. (The indices of the individuals removed from the population in Model II are $n_m + 1, \dots, n_{m-1}$. ξ_m determines the individuals to be removed from the population in Model I.)

If $k_m > 0$ (corresponding to a birth event), let $\kappa_m = \min \eta_m$. Define $\xi_m = \theta_{m-1}^{-1}(\kappa_m)$. (In this case, ξ_m is a singleton subset. We use ξ_m to denote both the subset and the value of the index in the subset.) Let θ_m restricted to $\{\xi_m, n_{m-1} + 1, \dots, n_m\}$ satisfy $\theta_m(\xi_m) = \eta_m(\sigma_m(1))$ and $\theta_m(n_{m-1} + i) = \eta_m(\sigma_m(i+1))$. Let θ_m restricted to $\{1, \dots, n_{m-1}\} - \xi_m$ be the mapping onto $\{1, \dots, n_m\} - \eta_m$ having the same order as θ_{m-1} restricted to $\{1, \dots, n_{m-1}\} - \xi_m$. ($\eta_m$ gives the set of indices determining the parent and the indices of the offspring in Model II. ξ_m determines the parent in Model I.)

Let $\mathscr{F}_m = \sigma\{\theta_k, \xi_k : k \le m\}$. The independence properties of the η_m and σ_m imply that

$$E\left[f(\theta_m, \xi_m) \mid \mathscr{F}_{m-1}\right] = E\left[f(\theta_m, \xi_m) \mid \theta_{m-1}\right]. \tag{2.1}$$

LEMMA 2.1. *For each m, $\xi_1, \dots, \xi_m, \theta_m$ are independent. If $k_m < 0$, ξ_m is uniformly distributed over $P_{n_{m-1}, |k_m|}$; if $k_m > 0$, ξ_m is uniformly distributed over $\{1, \dots, n_{m-1}\}$; and θ_m is uniformly distributed over S_{n_m}.*

PROOF. Proceeding by induction, assume that the result holds for m replaced by $m-1$. Then by (2.1) and the induction hypothesis, we have, for any choice of f and h_k,

$$\begin{aligned}
E\left[f(\theta_m) \prod_{k=1}^{m} h_k(\xi_k)\right] &= E\left[E\left[f(\theta_m) h_m(\xi_m) \mid \mathscr{F}_{m-1}\right] \prod_{k=1}^{m-1} h_k(\xi_k)\right] \\
&= E\left[E\left[f(\theta_m) h_m(\xi_m) \mid \theta_{m-1}\right] \prod_{k=1}^{m-1} h_k(\xi_k)\right] \\
&= E\left[f(\theta_m) h_m(\xi_m)\right] \prod_{k=1}^{m-1} E\left[h_k(\xi_k)\right].
\end{aligned}$$

It remains only to show that θ_m is independent of ξ_m and that they have the correct distributions. If $k_m < 0$, these observation follow immediately from the fact that θ_{m-1} is uniformly distributed. If $k_m > 0$, conditioning on ξ_m and η_m, it is clear that θ_m is uniformly distributed over all permutations that map $\{\xi_m, n_{m-1} + 1, \dots, n_m\}$ onto η_m and that conditioning on ξ_m, η_m is uniformly distributed on $P_{n_m, k_m + 1}$. It follows that the conditional distribution of θ_m given ξ_m is uniform on S_{n_m}, giving the desired independence and distribution. The uniformity of θ_{m-1} implies that ξ_m is uniformly distributed over $\{1, \dots, n_{m-1}\}$, completing the proof of the lemma. $\square$

2.2. *Proof of Theorem* 1.1. Suppose a realization of Model II is given. Let $\{t_m\}$ denote the sequence of times at which birth or death events occur, $0 \le t_1 \le t_2 \le \cdots$. If there are simultaneous birth and death events at time t, then, for the appropriate m, we have $t_m = t_{m+1} = t$. Under our convention of doing removals first, k_m is the negative of the number of deaths occurring at time t and k_{m+1} is the number of births. If $k_m > 0$, then η_m is the subset in Model II determining the indices of the parent and the offspring. Finally, let θ_0 be independent of X (and hence of $\{\eta_m\}$) and uniformly distributed over $S_{N(0)}$; for $k_m > 0$, let σ_m be independent (of everything) and uniformly distributed over S_{k_m+1}; and define θ_m as above. Set $\theta(t) = \theta_m$ for $t_m \le t < t_{m+1}$. Then by the properties of $\{\xi_m\}$ given in Lemma 2.1,

$$\big(Y_1(t), \dots, Y_{N(t)}\big) = \big(X_{\theta(t,1)}(t), \dots, X_{\theta(t, N(t))}(t)\big)$$

is the desired version of Model I. Since $Y(t)$ depends only on $Y(0)$, $\{\xi_m: t_m \le t\}$ and the evolution of the type processes between birth and death events, $\theta(t)$ must be conditionally independent of $\mathscr{F}_t^Y = \sigma(Y(s){:}\ s \le t)$ given $N(t)$. More generally, let $\mathscr{H} = \sigma(N(0), N_b(s), N_d(s){:}\ s \ge 0)$ and $\mathscr{G}_t = \mathscr{F}_t^Y \vee \mathscr{H}$. Then $\theta(t)$ is conditionally independent of $\mathscr{G}_t$ given $N(t)$. Consequently, the inverse permutation $\theta^{-1}(t)$ will also be conditionally independent of $\mathscr{G}_t$ and uniformly distributed over $S_{N(t)}$. Since

$$\big(X_1(t), \dots, X_{N(t)}\big) = \big(Y_{\theta^{-1}(t,1)}(t), \dots, Y_{\theta^{-1}(t, N(t))}(t)\big),$$

it follows that $(X_1(t), \dots, X_{N(t)}(t))$ is exchangeable. $\square$

2.3. *Exchangeability at stopping times*. As in the proof of Theorem 1.1, let $\mathscr{F}_t^Y = \sigma(Y(s){:}\ s \le t)$, $\mathscr{H} = \sigma(N(0), N_b(s), N_d(s){:}\ s \ge 0)$ and $\mathscr{G}_t = \mathscr{F}_t^Y \vee \mathscr{H}$.

PROPOSITION 2.2. *Let γ be a $\{\mathscr{G}_t\}$-stopping time. Then $(X_1(\gamma), \dots, X_{N(\gamma)}(\gamma))$ is exchangeable. (In particular, γ can be any nonnegative $\mathscr{H}$-measurable random variable.) If in addition, γ is a $\{\mathscr{G}_t\}$-predictable stopping time, then $(X_1(\gamma-), \dots, X_{N(\gamma-)}(\gamma-))$ is exchangeable.*

PROOF. As in the proof of Theorem 1.1, it is enough to show that $\theta(\gamma)$ is conditionally independent of $\mathscr{G}_\gamma$ given $N(\gamma)$. Assume first that γ is discrete. Let Π_n denote the uniform distribution over S_n. Then, for $A \in \mathscr{G}_\gamma$,

$$\begin{aligned} E\big[h(\theta(\gamma))I_A\big] &= \sum_{k=1}^{\infty} E\big[h(\theta(t_k))I_{A \cap \{\gamma = t_k\}}\big] \\ &= \sum_{k=1}^{\infty} E\bigg[\int h(\theta)\Pi_{N(t_k)}(d\theta)I_{A \cap \{\gamma = t_k\}}\bigg] \\ &= E\bigg[\int h(\theta)\Pi_{N(\gamma)}(d\theta)I_A\bigg], \end{aligned}$$

where the second equality follows from the fact that $\theta(t_k)$ is conditionally independent of $\mathscr{G}_{t_k}$ and $A \cap \{\gamma = t_k\} \in \mathscr{G}_{t_k}$. This identity gives the desired conditional independence. The result for general γ follows by approximating γ by a decreasing sequence of discrete stopping times. If γ is predictable, then there exists an increasing sequence $\{\gamma_n\}$ of $\{\mathscr{G}_t\}$-stopping times such hat $\gamma_n < \gamma$, a.s. and $\lim_{n\to\infty} \gamma_n = \gamma$. Consequently, the exchangeability of $(X_1(\gamma -), \dots, X_{N(\gamma-)}(\gamma -))$ follows from the exchangeability of $(X_1(\gamma_n), \dots, X_{N(\gamma_n)}(\gamma_n))$. $\square$

2.4. *Models with discrete generations.* We now consider the analogue of Theorem 1.1 for models with discrete generations. Let $N_0, N_1, \dots$ be positive integer-valued random variables giving the population size for each generation, and, for each $m \ge 1$, let K_m, $1 \le K_m \le N_{m-1}$, and $L_1^m, \dots, L_{K_m}^m$ be positive integer-valued random variables satisfying $\sum_{i=1}^{K_m} L_i^m = N_m$. The L_i^m are the litter sizes for the K_m members of generation $m-1$ that have descendants in the mth generation.

Model III. Let $Y_1^m, \dots, Y_{N_m}^m$ denote the types of the individuals in generation m. The parent of each litter in generation m is selected randomly, without replacement, from the members of generation $m-1$. For definiteness, the L_1^m members of the first litter are numbered $1, \dots, L_1^m$, the L_2^m members of the second litter are numbered $L_1^m + 1, \dots, L_1^m + L_2^m$, etc. If x is the type of the parent of litter i, then the type of each member of litter i has distribution $\eta(x, \cdot)$, where η is a transition function from E to E, and types of different individuals are conditionally independent given the types of their parents.

Model IV. Let $X_1^{m-1}, \dots, X_{N_{m-1}}^{m-1}$ denote the types of the individuals in generation $m-1$. Let the integers $\{1, \dots, N_m\}$ be partitioned randomly into set $A_1^m, \dots, A_{K_m}^m$ satisfying $|A_i^m| = L_i^m$, that is, the

$$\binom{N_m}{L_1^m \dots L_{K_m}^m}$$

distinct partitions are equally likely. Let σ^m be the permutation of $(1, \dots, K_m)$ defined so that the indices $\alpha_i^m = \min A_i^m$ are ordered

$$\alpha_{\sigma_1^m}^m < \alpha_{\sigma_2^m}^m < \cdots < \alpha_{\sigma_{K_m}^m}^m .$$

Then X_1^{m-1} becomes the parent for litter $A_{\sigma_1^m}^m$, X_2^{m-1} the parent for litter $A_{\sigma_2^m}^m$, etc. Conditioned on X^{m-1}, the types of the individuals in generation m with indices in $A_{\sigma_i^m}^m$ are iid with distribution $\eta(X_i^{m-1}, \cdot)$.

The proof of the following theorem is similar to that of Theorem 1.1.

Theorem 2.3. *Suppose that the initial population vector in Model IV* $(X_1^0, \dots, X_{N_0}^0)$ *is exchangeable and that the initial population vector* $(Y_1^0, \dots, Y_{N_0}^0)$ *in Model III satisfies* $Z^0 \equiv \sum_{i=1}^{N_0} \delta_{Y_i^0} = \tilde{Z}^0 \equiv \sum_{i=1}^{N_0} \delta_{X_i^0}$. *Define*

$$Z^m = \sum_{i=1}^{N_m} \delta_{Y_i^m}, \qquad \tilde{Z}^m = \sum_{i=1}^{N_m} \delta_{X_i^m}.$$

Then Z and $\tilde{Z}$ have the same distribution, and, for each $m \geq 0$, $(X_1^m, \dots, X_{N_m}^m)$ is exchangeable.

Proof. Let $Y^0 = (Y_1^0, \dots, Y_{N_0}^0)$ be given. Let $\{L_i^m\}$ be as above, and define $\mathscr{L} = \sigma(L_i^m\colon 1 \leq i \leq K_m,\ m = 1, 2, \dots)$. Let $H_0, H_1, \dots$ be a sequence of random permutations independent of Y^0 such that, conditioned on $\mathscr{L}$, $H_0, H_1, \dots$ are independent and H_m is uniformly distributed over S_{N_m}. Let $J_0^m = 0$, and for $i = 1, \dots, K_m$, define $J_i^m = \sum_{k=1}^i L_k^m$ and

$$A_i^m = \{H_m(j)\colon J_{i-1}^m < j \leq J_i^m\}.$$

Note that $A_1^m, \dots, A_{K_m}^m$ gives a random partition as in the description of Model IV. Starting with exchangeable $X^0 = (Y_{H_0(1)}^0, \dots, Y_{H_0(N_0)}^0)$, construct $X^1, X^2, \dots$ as prescribed in Model IV using the partitions $\{A_i^m, 1 \leq i \leq K_m\}$. Define $\theta_m(k) = j$ if $H_m(j) = k$ and set $Y_k^m = X_{\theta_m(k)}^m$. The parent of the ith litter in the mth generation is $Y_{H_{m-1}(i)}^{m-1} = X_i^{m-1}$, and if H_{m-1} is independent of $\{Y^0, \dots Y^{m-1}\}$, then the parents of Y^m are selected randomly from Y^{m-1} and $Y^0, Y^1, \dots$ will be a version of Model III. To see that this independence holds, first observe that Y^0 is independent of H_0 by assumption. Proceeding by induction, suppose H_{m-1} is independent of $\{Y^0, \dots, Y^{m-1}\}$. Let $\mathscr{G}_m = \sigma(Y^0, \dots, Y^{m-1}, H^0, \dots, H^{m-1}) \vee \mathscr{L}$ and $\mathscr{H}_m = \sigma(Y^0, \dots, Y^{m-1}, H^0, \dots, H^m) \vee \mathscr{L}$. Then

$$\begin{aligned} &E\big[f(H_m)h(Y^m)\big|\mathscr{H}_m\big] \\ &\quad = f(H_m)\int \cdots \int h(y_1, \dots, y_{N_m}) \prod_{i=1}^{K_m} \prod_{j=J_{i-1}^m+1}^{J_i^m} \eta\big(Y_{H_{m-1}(i)}^{m-1}, dy_j\big), \end{aligned}$$

and since H_m is independent of $\mathscr{G}_m$,

$$\begin{aligned} &E\big[f(H_m)h(Y^m)\big|\mathscr{G}_m\big] \\ &\quad = E[f(H_m)]\int \cdots \int h(y_1, \dots, y_{N_m}) \prod_{i=1}^{K_m} \prod_{j=J_{i-1}^m+1}^{J_i^m} \eta\big(Y_{H_{m-1}(i)}^{m-1}, dy_j\big), \end{aligned}$$

so the desired independence follows. □

2.5. *Models with simultaneous births to multiple parents.* The discrete generation model described in the previous section is a special case of a class of models in which simultaneous births may occur to multiple parents (in contrast to Models I and II in which only one parent is involved in each birth event). The analogous coupling for the more general models can be handled using essentially the same construction as in the previous section. For example, a birth event in which one parent has L_1 offspring and another L_2 offspring, increasing the population size from N to $N + L_1 + L_2$, can be treated as creating a "new generation" with one litter of size $L_1 + 1$, one of size $L_2 + 1$ and $N - 2$ litters each of size 1. Note that mutation/movement does not affect the coupling as long as it is defined the same way for both models and depends only on parental type.

3. Infinite population limit. In this section, we concentrate on continuous-time models in which each birth event involves only a single parent.

3.1. *Limit of total population size.*

3.1.1. *Birth and death processes.* In order to motivate the scaling that will be used in our general limit theorem, first consider a sequence of simple linear birth and death processes. These can be obtained as solutions of the equation

$$N^n(t) = N^n(0) + V_b\left(\int_0^t \lambda_n N^n(s)\,ds\right) - V_d\left(\int_0 \mu_n N^n(s)\,ds\right), \tag{3.1}$$

where V_b and V_d are independent, unit Poisson processes. If we rescale N^n, defining $P^n(t) = n^{-1}N^n(nt)$, (3.1) becomes

$$\begin{aligned} P^n(t) &= P^n(0) + \frac{1}{n}V_b\left(n^2\int_0^t \lambda_n P^n(s)\,ds\right) - \frac{1}{n}V_d\left(n^2\int_0^t \mu_n P^n(s)\,ds\right) \\ &= P^n(0) + \frac{1}{n}\tilde{V}_b\left(n^2\int_0^t \lambda_n P^n(s)\,ds\right) - \frac{1}{n}\tilde{V}_d\left(n^2\int_0^t \mu_n P^n(s)\,ds\right) \\ &\quad + n(\lambda_n - \mu_n)\int_0^t P^n(s)\,ds, \end{aligned} \tag{3.2}$$

here $\tilde{V}(u) = V(u) - u$. Note that

$$\left(\frac{1}{n}\tilde{V}_b(n^2\,\cdot), \frac{1}{n}\tilde{V}_d(n^2\,\cdot)\right)$$

is normalized so that it converges in distribution to (W_b, W_d), a pair of independent, standard Brownian motions. Consequently, if we assume that $\lambda_n \to \lambda$, $n(\lambda_n - \mu_n) \to c$ and $P^n(0) \Rightarrow P(0)$, P^n converges in distribution to a solution of

$$P(t) = P(0) + W_b\left(\int_0^t \lambda P(s)\,ds\right) - W_d\left(\int_0^t \lambda P(s)\,ds\right) + c\int_0^t P(s)\,ds. \tag{3.3}$$

Note, in addition, that the normalized total number of births satisfies

$$\frac{N_b^n(n\,\cdot)}{n^2} \Rightarrow \int_0^{\cdot} \lambda P(s)\,ds.$$

More generally, we can consider birth and death processes satisfying

$$\begin{aligned} N_b^n(t) &= V_1\left(n^2\int_0^t \lambda_n(P^n(s))\,ds\right) + V_3\left(n^2\int_0^t \tilde{\lambda}_n(P^n(s))\,ds\right), \\ N_d^n(t) &= V_2\left(n^2\int_0^t \mu_n(P^n(s))\,ds\right) + V_3\left(n^2\int_0^t \tilde{\lambda}_n(P^n(s))\,ds\right), \\ P^n(t) &= P^n(0) + \frac{1}{n}N_b^n(t) - \frac{1}{n}N_d^n(t). \end{aligned}$$

If $P^n(0) \Rightarrow P(0)$ and $\lambda_n(\cdot) \to \lambda(\cdot)$, $\tilde{\lambda}_n(\cdot) \to \tilde{\lambda}(\cdot)$ and $n(\lambda_n(\cdot) - \mu_n(\cdot)) \to b(\cdot)$ uniformly on compact sets, then P^n converges to a solution of

$$P(t) = P(0) + W_1\left(\int_0^t \lambda(P(s))\,ds\right) - W_2\left(\int_0^t \lambda(P(s))\,ds\right) + \int_0^t b(P(s))\,ds \tag{3.4}$$

and

$$\frac{N_b^n(\cdot)}{n^2} \Rightarrow \int_0^{\cdot} \left(\lambda(P(s)) + \tilde{\lambda}(P(s))\right) ds, \tag{3.5}$$

provided the solution of (3.4) does not blow up in finite time. In this case, P is a diffusion with generator

$$Gf(z) = \lambda(z) f''(z) + b(z) f'(z)$$

[see Ethier and Kurtz (1986), Theorem 6.5.4].

3.1.2. *Branching processes.* Another example of interest is for N^n to be a branching process. For each n, let $\{\xi_k^n,\ k = 1, 2, \dots\}$ be independent, integer-valued random variables with $\xi_k^n \ge -1$. Suppose that there exist $\alpha_n \to \infty$ such that

$$\sup_n E\left[\left|\frac{1}{n}\sum_{k=1}^{\alpha_n} \xi_k^n\right|\right] < \infty \tag{3.6}$$

and $(1/n)\sum_{k=1}^{\alpha_n} \xi_k^n \Rightarrow Y_1$. Let V be a unit Poisson process and define a compound Poisson process

$$\hat{V}^n(t) = \sum_{k=1}^{V(\alpha_n t)} \xi_k^n.$$

Then $Y^n \equiv n^{-1}\hat{V}^n \Rightarrow Y$ (in the Skorohod topology), where $Y(1)$ has the same distribution as Y_1. Let $N^n(0) = n$. Then the solution of

$$N^n(t) = N^n(0) + \hat{V}^n\left(\int_0^t n^{-1} N^n(s)\,ds\right)$$

is a continuous-time Markov branching process. Normalizing N^n, we have

$$P^n(t) = 1 + Y^n\left(\int_0^t P^n(s)\,ds\right)$$

and $P^n \Rightarrow P$ satisfying

$$P(t) = 1 + Y\left(\int_0^t P(s)\,ds\right).$$

[See Ethier and Kurtz (1986), Theorem 9.1.4.]

The limiting process Y can be any Lévy process with generator of the form

$$Hf(z) = \tfrac{1}{2} a f''(z) + b f'(z) + \int_{(0,\infty)} \left(f(z+y) - f(z) - y f'(z)\right) \nu(dy),$$

where ν satisfies $\int_{(0,\infty)} y(y \wedge 1)\nu(dy) < \infty$. In particular, Y has no negative jumps. The condition on ν is stronger than necessary for a general Lévy measure ($\int |y|^2 \wedge 1 \nu(dy) < \infty$). The stronger condition assures that $E[|Y(t)|] < \infty$ and that P does not blow up in finite time. The generator for P is

$$Gf(v) = vHf(v).$$

In addition, the convergence of $(1/n)\sum_{k=1}^{\alpha_n} \xi_k^n$ implies the convergence of $(1/n^2)\sum_{k=1}^{\alpha_n} (\xi_k^n)^2$. [This assertion follows from the central convergence criterion in Loeve (1963), Section 22.4.] This convergence implies the convergence of the quadratic variation $[Y_n]. \Rightarrow [Y]$. which in turn implies $[P^n]. \Rightarrow [P]$.. Note that $[Y]$ is a process with independent increments with generator

$$H_2 f(y) = af'(y) + \int_0^\infty \big(f(y+u^2) - f(y)\big)\nu(du),$$

and setting $\gamma(t) = \int_0^t P(s)\,ds$, $[P]_t = [Y]_{\gamma(t)}$.

3.1.3. *Population models with multiple simultaneous births and deaths.* Suppose that V_1 and V_2 are independent, unit Poisson processes and that $\{\zeta_k^b\}$ and $\{\zeta_k^d\}$ are iid sequences of nonnegative integer-valued random variables with finite mean and variance. [$E[\zeta_k^b] = m_b$, $\mathrm{Var}(\zeta_k^b) = \sigma_b^2$, $E[\zeta_k^d] = m_d$, $\mathrm{Var}(\zeta_k^d) = \sigma_d^2$.] Let

$$\begin{aligned}
\hat{N}_b^n(t) &= V_1\left(n^2 \int_0^t \lambda_n(P^n(s))\,ds\right),\\
N_b^n(t) &= \sum_{k=1}^{\hat{N}_b^n(t)} \zeta_k^b,\\
\hat{N}_d^n(t) &= V_2\left(n^2 \int_0^t \mu_n(P^n(s))\,ds\right),\\
N_d^n(t) &= \sum_{k=1}^{\hat{N}_d^n(t)} \zeta_k^d,\\
P^n(t) &= P^n(0) + \frac{1}{n}N_b^n(t) - \frac{1}{n}N_d^n(t).
\end{aligned}$$

If $P^n(0) \Rightarrow P(0)$ and $\lambda_n(\cdot) \to \lambda(\cdot)$ and $n(\lambda_n(\cdot)m_b - \mu_n(\cdot)m_d) \to b(\cdot)$ uniformly on compact sets, then P^n converges in distribution to a solution of

$$\begin{aligned}
P(t) = P(0) &+ \sigma_b W_1\left(\int_0^t \lambda(P(s))\,ds\right) + m_b W_2\left(\int_0^t \lambda(P(s))\,ds\right)\\
&- \sigma_d W_3\left(\frac{m_b}{m_d}\int_0^t \lambda(P(s))\,ds\right) - m_d W_4\left(\frac{m_b}{m_d}\int_0^t \lambda(P(s))\,ds\right)\\
&+ \int_0^t b(P(s))\,ds,
\end{aligned}$$

where the W_i are standard Brownian motions, provided the solution does not blow up in finite time. The quantity

$$U_n(t) = \frac{[N_b^n]_t + N_b^n(t)}{n^2},$$

where $[N_b^n]_t$ denotes the quadratic variation of N_b^n, will play a critical role in our discussion. Note that under the above assumptions, $U_n \Rightarrow U$ given by

$$U(t) = (m_b + \sigma_b^2 + m_b^2)\int_0^t \lambda(P(s))\,ds. \tag{3.7}$$

3.1.4. *Models with constant population size*. Assume that $N^n(0) = n$ and that $N_b^n = N_d^n$, so that $N^n(t) = n$ for all $t \geq 0$. Under our convention of "killing first," we must have $N_b^n(t) - N_b^n(t-) < n$. Again, consider

$$U^n(t) = \frac{[N_b^n]_t + N_b^n(t)}{n^2}.$$

If N_b^n has stationary, independent increments, then so does U^n. Under this hypothesis, the possible limits $U^n \Rightarrow U$ are the nondecreasing processes with stationary, independent increments and jumps bounded by 1, that is, processes with generators of the form

$$Df(u) = af'(u) + \int_0^1 (f(u+v) - f(u))\nu(dv),$$

where ν satisfies $\int_0^1 v\nu(dv) < \infty$. In Section 5, we will see that this model is related to coalescent models of Pitman (1997).

3.2. *Conditions on total population size*. With the above examples in mind, define

$$\begin{aligned} P^n(t) &= \frac{1}{n}N^n(0) + \frac{1}{n}N_b^n(t) - \frac{1}{n}N_d^n(t), \\ \tau^n &= \inf\{t: P^n(t) = 0\}, \\ U^n(t) &= \frac{[N_b^n]_t + N_b^n(t)}{n^2}, \\ H^n(t) &= \int_0^t \frac{1}{P^n(s)^2}\,dU^n(s), \end{aligned} \tag{3.8}$$

where $[N_b^n]_t$ denotes the quadratic variation of N_b^n. In the birth and death examples, Section 3.1.1, N_b^n is a counting process and $[N_b^n]_t = N_b^n(t)$, so by (3.5),

$$U_n \Rightarrow \int_0^t 2\big(\lambda(P(s)) + \tilde{\lambda}(P(s))\big)\,ds.$$

For the branching process examples, Section 3.1.2,

$$N_b^n(t) = \sum_{k=1}^{V(\alpha_n \int_0^t P^n(s)\,ds)} \xi_k^n \vee 0,$$

and observing that $(\xi_k^n \vee 0)^2 + \xi_k^n \vee 0 = \xi_k^n(\xi_k^n + 1)$, we see that

$$U_n(t) = [P^n]_t + \frac{1}{n} P^n(t) \Rightarrow [P]_t.$$

For the models in Section 3.1.3, the limit of U_n is given in (3.7). For the constant population size models of Section 3.1.4, the limit of U_n has stationary, independent increments.

We assume that there are no further births after τ^n. Our basic convergence assumption is that

$$(3.9) \qquad (P^n, U^n) \Rightarrow (P, U).$$

For $\varepsilon > 0$, let $\tau_\varepsilon^n = \inf\{t\colon\ P^n(t) \le \varepsilon\}$ and $\tau_\varepsilon = \inf\{t\colon\ P(t) \le \varepsilon\}$. In general, (3.9) does not imply $\tau_\varepsilon^n \Rightarrow \tau_\varepsilon$; however, this convergence will hold for all but countably many $\varepsilon > 0$. Define

$$(3.10) \qquad \tau \equiv \lim_{\varepsilon \to 0} \tau_\varepsilon = \inf\{t\colon P(t) \wedge P(t-) = 0\}$$

and

$$H(t) = \int_0^{t \wedge \tau} \frac{1}{P(s)^2}\, dU(s).$$

Then (3.9) implies the existence of a sequence $\varepsilon_n \to 0$ such that

$$(3.11) \qquad \left(P^n, U^n, H^n\left(\cdot \wedge \tau_{\varepsilon_n}^n\right), \tau_{\varepsilon_n}^n\right) \Rightarrow (P, U, H, \tau),$$

where the convergence in distribution is in $D_{[0,\infty)\times[0,\infty)\times[0,\infty]}[0,\infty) \times [0,\infty]$ with the Skorohod topology on $D_{[0,\infty)\times[0,\infty)\times[0,\infty]}[0,\infty)$. Note that we allow H_n and H to assume the value ∞ if the integrals diverge in finite time. For simplicity, we will usually assume

$$(3.12) \qquad (P^n, U^n, H^n, \tau^n) \Rightarrow (P, U, H, \tau).$$

In particular, if $\tau = \infty$ a.s., then (3.12) holds. If $(P^n, U^n, \tau^n) \Rightarrow (P, U, \tau)$ and $H(\tau) = \infty$ on $\{\tau < \infty\}$, then (3.12) also holds.

3.3. *Limit in E^∞.* Let X^n be a version of Model II of the previous section determined by N_b^n, N_d^n and a fixed Markov evolution with generator B satisfying the conditions in Section 1.3. In particular, a process χ corresponding to B has cadlag sample paths. We also assume that, for every initial condition, the process has no fixed points of discontinuity [i.e., $P\{\chi(t) = \chi(t-)\} = 1$ for all $t > 0$]. (This last condition is unnecessary most of the time, in particular, if the population size processes are Markov birth and death processes as described above.)

Let $X_1(0), X_2(0), \dots$ be an infinite exchangeable sequence in E, and assume that, for each n,

$$\left(X_1^n(0), \dots, X_{N^n(0)}^n(0)\right) = \left(X_1(0), \dots, X_{N^n(0)}(0)\right).$$

We will refer to X_k^n as the kth level process. It will be convenient to define $X_k^n(t)$ for $k > N^n(t)$ to be $X_k(0)$ if $\max_{s \le t} N^n(s) < k$ and to be $X_k^n(\beta_k^n(t)-)$, where $\beta_k^n(t) = \sup\{s < t: N^n(s) \ge k\}$, otherwise.

Note that the first-level process X_1^n is just an E-valued Markov process with generator B and fixed initial distribution stopped at $\tau^n = \inf\{t: N^n(t) = 0\}$, so X_1^n converges in distribution provided τ^n does. (The assumption of no fixed points of distinuity is needed here unless the limit of the τ^n "misses" all such points with probability 1.)

Next, recall that X_2^n evolves as a Markov process with generator B except at those tims when the first two levels are involved in a birth event. At each such time, the second-level process "copies" the value of the first-level process. Let $N_{12}^n(t)$ denote the number of birth events up to time t that involve the levels 1 and 2. Then (X_1^n, X_2^n) converges in distribution provided the counting process N_{12}^n converges in distribution. (Again, we need the assumption of no fixed points of discontinuity unless the jump times of the limit of N_{12}^n miss these.) Note that if there is a birth event at time t with k offspring, then, conditioning on N^n and N_b^n for all time (not just up to time t), the probability that levels 1 and 2 are involved is just

$$\frac{\binom{N^n(t)-2}{k-1}}{\binom{N^n(t)}{k+1}} = \frac{k(k+1)}{N^n(t)(N^n(t)-1)}.$$

Consequently,

$$\begin{aligned} N_{12}^n(t) &- \sum_{\{m:\, t_m \le t,\, k_m > 0\}} \frac{k_m(k_m+1)}{N^n(t_m)(N^n(t_m)-1)} \\ &= N_{12}^n(t) - \int_0^t \frac{1}{P^n(s)(P^n(s)-1/n)}\, dU^n(s) \end{aligned} \tag{3.13}$$

is a martingale with respect to the filtration

$$\mathscr{G}_t^n = \sigma\left(X^n(s), N_{ij}(s), s \le t, 1 \le i < j;\, N^n(u), N_b^n(u), u \ge 0\right),$$

at least if the process is stopped the first time the sum exceeds an arbitrary constant K. By (3.12), the sum converges in distribution to

$$H(\cdot) = \int_0^{\cdot \wedge \tau} \frac{1}{P(s)^2}\, dU(s). \tag{3.14}$$

[Note that the ratio $P^n(t)/(P^n(t) - 1/n)$ is bounded by 2 for $t < \tau^n$ and converges to 1 uniformly on $[0, \tau^n - \delta]$ for each $\delta > 0$.] By Lemma A.1, it follows that N_{12}^n converges in distribution to a counting process with distribution determined by H and hence that (X_1^n, X_2^n) converges in distribution.

In general, fix a level l, and let $K \subset \{1, \dots, l\}$. $|K|$ will denote the cardinality of the set. Define

$$N_K^n(t) = \left|\{m: t_m \le t, \eta_m \cap \{1, \dots, l\} = K\}\right|.$$

Then

$$(3.15) \qquad N_K^n(t) - \sum_{\{m:\, t_m \le t,\, k_m + 1 \ge |K|\}} \frac{\binom{N^n(t_m) - l}{k_m + 1 - |K|}}{\binom{N^n(t_m)}{k_m + 1}}$$

is a martingale with respect to $\{\mathscr{G}_t^n\}$. Let $H_K^n(t)$ denote the sum in (3.15) and U_c denote the continuous part of U. The summands can be rewritten as

$$\frac{\binom{N - l}{k + 1 - |K|}}{\binom{N}{k+1}} = \frac{(k+1)!}{(k + 1 - |K|)!\, N \dots (N - |K| + 1)} \times \frac{(N - k - 1) \dots (N - k - l + |K|)}{(N - |K|) \dots (N - l + 1)}.$$

If $|K| = 2$, it follows from (3.12) that H_K^n converges in distribution to

$$(3.16) \qquad \int_0^{\cdot \wedge \tau} \frac{1}{P(s)^2}\, dU_c(s) + \sum_{s \le \cdot \wedge \tau} \frac{\Delta U(s)}{P(s)^2} \left(1 - \frac{\sqrt{\Delta U(s)}}{P(s)}\right)^{l-2},$$

where $\Delta U(s) = U(s) - U(s-)$. Note that if $l = 2$, then (3.16) is just (3.14). If $|K| > 2$, then the sum converges in distribution to

$$(3.17) \qquad \sum_{s \le \cdot \wedge \tau} \left(\frac{\sqrt{\Delta U(s)}}{P(s)}\right)^{|K|} \left(1 - \frac{\sqrt{\Delta U(s)}}{P(s)}\right)^{l - |K|}.$$

In particular, if U is continuous and $|K| > 2$, then $N_K^n \Rightarrow 0$, that is, in the limit, only two levels are involved in any birth event. Note that typically U is continuous even when the original model has multiple simultaneous births. (See Section 3.1.3.)

In general, if $\Delta U(s) > 0$, then conditioned on U and P, levels are included in the birth event independently with probability $\sqrt{\Delta U(s)}\,/P(s)$. In any case, by Lemma A.1, the family of counting processes $\{N_K^n\colon K \subset \{1, \dots, l\}\}$ converges in distribution in the Skorohod topology on $D_{\mathbb{N}^{2^l}}[0, \infty)$.

Given U and P, we can construct the limiting process in the following manner. Let $\{V_{ij},\ i<j\}$ be independent unit Poisson processes, independent of U and P. Define

$$L_{ij}(t) = V_{ij}\left(\int_0^{t\wedge\tau} \frac{1}{P(s)^2}\,dU_c(s)\right). \tag{3.18}$$

L_{ij} determines the times of the birth events that involve only i and j. Let $\{\gamma_m\}$ be some ordering of the times of discontinuity of U, let $\alpha_m = \sqrt{\Delta U(\gamma_m)}\,/P(\gamma_m)$ and let $\{v_{jm}\}$ be independent, uniform $[0,1]$ random variables that are independent of U, P and the V_{ij}. Define

$$L_j(t) = \sum_{\gamma_m \le t} I_{\{v_{jm}\le \alpha_m\}}, \tag{3.19}$$

and, for $K \subset \{1,\dots,l\}$,

$$L_K^l(t) = \sum_{\gamma_m\le t}\prod_{j\in K} I_{\{v_{jm}\le\alpha_m\}}\prod_{j\in\{1,\dots,l\}-K} I_{\{v_{jm}>\alpha_m\}}. \tag{3.20}$$

L_j determines whether or not level j is involved in the birth event at each discontinuity of U and the L_K^l track the subsets of $\{1,\dots,l\}$ that are involved in birth events at each discontinuity of U.

We can construct the limit process $X=(X_1,X_2,\dots)$ inductively. X_1 is just a Markov process with generator B stopped at τ. Suppose that $(X_1,\dots,X_{l-1})$ has been constructed. Then between the jump times of L_j, $j\le l$, and L_{ij}, $i<j\le l$, and before τ, X_l evolves as a Markov process with generator B, dependent on the other levels only through its value at the most recent jump time. For $t>\tau$, $X_l(t)=X_l(\tau-)$. At a jump time t of L_{ij}, the level processes satisfy

$$\begin{aligned} X_k(t) &= X_k(t-), \qquad k<j,\\ X_j(t) &= X_i(t),\\ X_k(t) &= X_{k-1}(t-), \qquad k>j, \end{aligned}$$

and at a discontinuity time t of U, defining $i=\min\{j\colon \Delta L_j(t)>0\}$ and $K_k(t)=\sum_{j\le k}\Delta L_j(t)-1$, the level processes satisfy

$$\begin{aligned} X_k(t) &= X_k(t-), \qquad k\le i,\\ X_k(t) &= X_i(t), \qquad k>i,\ \Delta L_k(t)>0,\\ X_k(t) &= X_{k-K_k(t)}(t-), \qquad \text{otherwise.} \end{aligned}$$

Note that X can be explicitly constructed using the mappings $M(x,t,\cdot)$ described in Lemma 1.3, or X can be characterized by the requirement that there exists a filtration $\{\mathscr{G}_t\}$ such that L_{ij} is $\mathscr{G}_0$-measurable for all i, j and

$$\begin{aligned} f(X_k(t)) &- \int_0^t Bf(X_k(s))\,ds \\ &- \sum_{1\le i<j<k} \int_0^t \big(f(X_{k-1}(s-)) - f(X_k(s-))\big)\,dL_{ij}(s) \\ &- \sum_{1\le i<k} \int_0^t \big(f(X_i(s-)) - f(X_k(s-))\big)\,dL_{ik}(s) \\ &- \sum_{K\subset\{1,\dots,k\},\, k\in K} \int_0^t \Big(f\big(X_{\min(K)}(s-)\big) - f\big(X_k(s-)\big)\Big)\,dL_K^k(s) \\ &- \sum_{K\subset\{1,\dots,k\},\, k\notin K} \int_0^t \Big(f\big(X_{k-|K|+1}(s-)\big) - f\big(X_k(s-)\big)\Big)\,dL_K^k(s) \end{aligned}$$

is a $\{\mathscr{G}_t\}$-martingale for all $f \in \mathscr{D}(B)$.

For $t < \tau^n$, define

$$Z^n(t) = \frac{1}{N^n(t)} \sum_{k=1}^{N^n(t)} \delta_{X_k^n(t)},$$

and, for all $t \ge 0$, define

$$Z(t) = \lim_{l\to\infty} \frac{1}{l} \sum_{k=1}^{l} \delta_{X_k(t)}.$$

PROPOSITION 3.1. *For each $t \ge 0$, $(X_1(t), X_2(t), \dots)$ is exchangeable. More generally, let $\mathscr{H}_t = \sigma(Z(s)\colon s \le t) \vee \sigma(U(s), P(s)\colon s \ge 0)$ and let γ be an $\{\mathscr{H}_t\}$-stopping time. Then $(X_1(\gamma), X_2(\gamma), \dots)$ is exchangeable. If γ is $\{\mathscr{H}_t\}$-predictable, then $(X_1(\gamma-), X_2(\gamma-), \dots)$ is also exchangeable.*

PROOF. Let $\tau_l^n = \inf\{t\colon N^n(t) \le l\}$. The fact that $(P^n, \tau^n) \Rightarrow (P, \tau)$ implies $(P^n, \tau_l^n) \Rightarrow (P, \tau)$, since $P^n(\tau_l^n) \Rightarrow 0$. The exchangeability of $(X_1^n(t \wedge \tau_l^n), \dots, X_l^n(t \wedge \tau_l^n))$ then implies the exchangeability of $(X_1(t), \dots, X_l(t))$ and, since l is arbitrary, of $(X_1(t), X_2(t), \dots)$. The remainder of the proof is similar to the proof of Proposition 2.2. □

THEOREM 3.2. *Let X^n be a version of Model II of the previous section determined by N_b^n, N_d^n and a fixed Markov evolution with generator B satisfying the conditions in Section* 1.3. *Assume that, for each initial condition, the cadlag process corresponding to B has no fixed points of discontinuity. Let $X_1(0), X_2(0), \dots$ be an infinite exchangeable sequence in E and assume that, for each n,*

$$\big(X_1^n(0), \dots, X_{N^n(0)}^n(0)\big) = \big(X_1(0), \dots, X_{N^n(0)}(0)\big).$$

Let P^n, τ^n, U^n and H^n be defined as in (3.8), *and assume that* (3.12) *holds. Then*

$$(P^n, U^n, P^n Z^n, X^n) \Rightarrow (P, U, PZ, X) \tag{3.21}$$

in $D_{\mathbb{R}^2\times\mathscr{M}(E)\times E^\infty}[0,\infty)$. *If, in addition, U is continuous, then, for each* $f\in\mathscr{D}(B)$,

$$\int f(x)Z(\cdot\wedge\tau,dx)$$

is continuous a.s., and hence $Z(\cdot\wedge\tau)$ *is continuous in the* ρ_B *metric.*

REMARK 3.3. (a) If (3.9) holds but not (3.12), then there exists a sequence $\varepsilon_n\to 0$ such that

$$\big(P^n,U^n,P^nZ^n,X^n\big(\cdot\wedge\tau^n_{\varepsilon_n}\big)\big)\Rightarrow(P,U,PZ,X). \tag{3.22}$$

(b) As noted in Section 1.3, continuity in ρ_B is usually equivalent to continuity in the weak topology.

PROOF OF THEOREM 3.2. The assumptions and discussion above give $(P^n,U^n,X^n)\Rightarrow(P,U,X)$. To see that (3.21) holds, define

$$Z^n_l(t)=\frac{1}{l}\sum_{k=1}^{l}\delta_{X^n_k(t)},$$

and similarly for Z_l. Then $P^nZ^n_l(\cdot\wedge\tau^n_l)\Rightarrow PZ_l(\cdot\wedge\tau)$ by the convergence of (P^n,X^n).

Consequently, the theorem will follow if we show that Z^n_l approximates Z^n well enough. The following lemmas verify the necessary approximation.

Since the discontinuities of $\int f(x)Z_l(\cdot\wedge\tau)$ are bounded by $\|f\|/l$, the last statement follows by Lemma 3.5. □

LEMMA 3.4. *For each* $T>0$, $c>0$ *and* $\varepsilon>0$ *and each* $f\in\mathscr{D}(B)$, *there exists a sequence* δ_l *such that* $\sum_l\delta_l<\infty$ *and*

$$P\left\{\sup_{t\le T}\left|\int f(x)Z^n(t\wedge\tau^n_l,dx)-\int f(x)Z^n_l(t\wedge\tau^n_l,dx)\right|\ge 11\varepsilon,U^n(T)\le c\right\}$$
$$\le\delta_l.$$

PROOF. By Lemma A.2, for any $t\ge 0$ and $\varepsilon>0$,

$$P\left\{\left|\int f(x)Z^n(t\wedge\tau^n_l,dx)-\int f(x)Z^n_l(t\wedge\tau^n_l,dx)\right|\ge\varepsilon\right\}\le 2e^{-\eta l},$$

where η depends only on $\|f\|$ and ε. More generally, let

$$\mathscr{H}^n_t=\sigma(P^n(s),U^n(s):s\ge 0)\vee\sigma(Z^n(s):s\le t).$$

Then for any $\{\mathscr{H}^n_t\}$-stopping time α with $\alpha\le\tau^n_l$,

$$P\left\{\left|\int f(x)Z^n(\alpha,dx)-\int f(x)Z^n_l(\alpha,dx)\right|\ge\varepsilon\right\}\le 2e^{-\eta l}.$$

Fix l and ε. Define

$$\alpha^n_1=\inf\{t:U^n(t)>l^{-4}\}\wedge l^{-4}\wedge\tau^n_l$$

and

$$\alpha_{k+1}^n = \inf\{t: U^n(t) > U^n(\alpha_k^n) + l^{-4}\} \wedge (\alpha_k^n + l^{-4}) \wedge \tau_l^n.$$

In addition, define

$$\tilde{\alpha}_k^n = \inf\left\{t > \alpha_k^n: \left|\int f(x) Z^n(t, dx) - \int f(x) Z^n(\alpha_k^n, dx)\right| \ge 6\varepsilon\right\} \wedge \tau_l^n.$$

Note that, for $k_l = 2(c + T)l^4$, $P\{\alpha_{k_l}^n < T \wedge \tau_l^n, U^n(\alpha_{k_l}^n) < c\} = 0$. Defining

$$H_k^n = \left|\int f(x) Z^n(\alpha_k^n, dx) - \int f(x) Z_l^n(\alpha_k^n, dx)\right| \vee \left|\int f(x) Z^n(\tilde{\alpha}_k^n, dx) - \int f(x) Z_l^n(\tilde{\alpha}_k^n, dx)\right|,$$

we have that

$$P\left\{\sup_{k \le k_l} H_k^n \ge \varepsilon\right\} \le 8(c + T) l^4 e^{-\eta l}.$$

It remains to estimate the variation in the intervals $(\alpha_k^n, \alpha_{k+1}^n)$.

To simplify the notation, we suppress the index n. For each k and j, let $\gamma_{jk} = \inf\{s > \alpha_k: X_j(\alpha_k)$ has no descendants at time $s\}$. For $\alpha_k \le s < \gamma_{jk}$, let $\beta_{jk}(s)$ be the smallest index of a descendant of $X_j(\alpha_k)$ and define $\tilde{X}_j(s) = X_{\beta_j(s)}(s \wedge \gamma_{jk})$. Then, for $\alpha_k \le u < \alpha_{k+1}$,

$$(3.23) \quad \begin{aligned} &\int f(x) Z_l^n(u, dx) - \int f(x) Z_l^n(\alpha_k, dx) \\ &\quad = \int f(x) Z_l^n(u, dx) - \frac{1}{l}\sum_{j=1}^{l} f\big(X_{\beta_j(u)}(u)\big) \\ &\qquad + \frac{1}{l}\sum_{j=1}^{l}\Big(f\big(X_j(u)\big) - f\big(\tilde{X}_j(\alpha_k)\big)\Big). \end{aligned}$$

Let

$$K_1^n = \max_{k \le k_l} \sup_{\alpha_k \le u < \alpha_{k+1}} \left|\int f(x) Z_l^n(u, dx) - \frac{1}{l}\sum_{j=1}^{l} f\big(X_{\beta_j(u)}(u)\big)\right|$$

and

$$K_2^n = \max_{k \le k_l} \sup_{\alpha_k \le u < \alpha_{k+1}} \left|\frac{1}{l}\sum_{j=1}^{l}\Big(f\big(\tilde{X}_j(u)\big) - f\big(\tilde{X}_j(\alpha_k)\big)\Big)\right|.$$

Note that the first term on the right of (3.23) is bounded by $2\|f\| \cdot N_l^n(\alpha_k, \alpha_{k+1})/l$, where $N_l^n(\alpha_k, \alpha_{k+1})$ is the number of new individuals added to the population in the time interval (α_k, α_{k+1}) with initial index less than or equal to l. Then, with k_m and t_m as in the construction in the previous section, using the fact that

$$U^n(\alpha_{k+1}-) - U^n(\alpha_k) = \sum_{\alpha_k < t_m < \alpha_{k+1}} \frac{(k_m^+ + 1)k_m^+}{n_m(n_m - 1)} \le l^{-4}, \tag{3.24}$$

we have

$$E\big[N_l^n(\alpha_k, \alpha_{k+1})\big|U^n, P^n\big] \le \sum_{\alpha_k < t_m < \alpha_{k+1}} \binom{l}{2} \frac{(k_m^+ + 1)k_m^+}{n_m(n_m - 1)} \le \frac{1}{2l^2}.$$

Consequently, for $\|f\|l^{-3} < \varepsilon$,

$$P\{K_1^n \ge 2\varepsilon\}$$
$$\le \sum_{k=0}^{k_l - 1} P\left\{\big|N_l^n(\alpha_k, \alpha_{k+1}) - E\big[N_l^n(\alpha_k, \alpha_{k+1})\big|U^n, P^n\big]\big| \ge l\frac{\varepsilon}{2\|f\|}\right\}.$$

We can write

$$N_l^n(\alpha_k, \alpha_{k+1}) = \sum_{\alpha_k < t_m < \alpha_{k+1}} (\zeta_m - 1)^+,$$

where, conditional on U^n and P^n, the ζ_m are independent hypergeometric random variables. (Let $\zeta_m = 0$ if $k_m < 0$.) Using the fact that, for $k_m > 0$,

$$\begin{aligned} &E[\zeta_m(\zeta_m - 1)\cdots(\zeta_m - q)] \\ &\quad = l(l-1)\dots(l-q)\frac{(k_m + 1)\dots(k_m + 1 - q)}{n_m(n_m - 1)\dots(n_m - q)}, \end{aligned} \tag{3.25}$$

the inequalities

$$\big[(z-1)^+\big]^2 \le z(z-1),$$
$$\big[(z-1)^+\big]^4 \le 3z(z-1)(z-2)(z-3) + 3z(z-1),$$

and (3.24), we can estimate fourth moments to obtain

$$P\left\{\big|N_l^n(\alpha_k, \alpha_{k+1}) - E\big[N_l^n(\alpha_k, \alpha_{k+1})\big|U^n, P^n\big]\big| \ge l\frac{\varepsilon}{2\|f\|}\right\}$$
$$\le \frac{16\|f\|^4}{\varepsilon^4 l^4}[5l^{-4} + 3l^{-2}]$$

and hence

$$P\{K_1^n \ge 2\varepsilon\} \le \frac{k_l 16\|f\|^4}{\varepsilon^4 l^4}[5l^{-4} + 3l^{-2}].$$

The second term on the right-hand side of (3.23) can be rewritten as

$$\begin{aligned}\frac{1}{l}\sum_{j=1}^{l}\Big(f\big(\tilde X_j(u)\big)-f\big(\tilde X_j(\alpha_k)\big)\Big)\\ =\frac{1}{l}\sum_{j=1}^{l}\left(f\big(\tilde X_j(u)\big)-f\big(\tilde X_j(\alpha_k)\big)-\int_{\alpha_k}^{u\wedge\gamma_{jk}}Bf\big(\tilde X_j(s)\big)\,ds\right)\\ +\frac{1}{l}\sum_{j=1}^{l}\int_{\alpha_k}^{u\wedge\gamma_{jk}}Bf\big(\tilde X_j(s)\big)\,ds.\end{aligned}\tag{3.26}$$

Then

$$M_{lk}(u)=\frac{1}{l}\sum_{j=1}^{l}\left(f\big(\tilde X_j(u)\big)-f\big(\tilde X_j(\alpha_k)\big)-\int_{\alpha_k}^{u\wedge\gamma_{jk}}Bf\big(\tilde X_j(s)\big)\,ds\right)$$

is a martingale, and if $l^{-4}\|Bf\|\le\varepsilon$, we have

$$P\{K_2^n\ge 2\varepsilon\}\le\sum_{k=0}^{k_l-1}P\Big\{\sup_{\alpha_k\le u<\alpha_{k+1}}|M_{lk}(u)|\ge\varepsilon\Big\}\le Ce^{-\eta_2 l},$$

where C and η_2 depend only on ε and $2\|f\|+l^{-4}\|Bf\|$.

Finally, note that if $\max_{k\le k_l}H_k^n\le\varepsilon$, $K_1^n\le 2\varepsilon$ and $K_2^n\le 2\varepsilon$, then $\tilde\alpha_k^n\ge\alpha_{k+1}^n$ and $\sup_{\alpha_k^n\le t<\alpha_{k+1}^n}|\int f(x)Z^n(t,dx)-\int f(x)Z^n(\alpha_k^n,dx)|\le 6\varepsilon$. Hence, under these conditions,

$$\sup_{t\le\alpha_{k_l}^n}\left|\int f(x)Z^n(t\wedge\tau_l^n,dx)-\int f(x)Z_l^n(t\wedge\tau_l^n,dx)\right|\le 11\varepsilon,$$

and the conclusion of the lemma follows with

$$\delta_l=8(c+T)l^4e^{-\eta l}+\frac{k_l16\|f\|^4}{\varepsilon^4 l^4}\big[5l^{-4}+3l^{-2}\big]+Ce^{-\eta_2 l}.\qquad\square$$

LEMMA 3.5. *For each $T>0$, $c>0$ and $\varepsilon>0$ and each $f\in\mathscr{D}(B)$, there exists a sequence δ_l such that $\sum_l\delta_l<\infty$ and*

$$P\Big\{\sup_{t\le T}\Big|\int f(x)Z(t\wedge\tau,dx)-\int f(x)Z_l(t\wedge\tau,dx)\Big|\ge 11\varepsilon,\,U(T)\le c\Big\}\le\delta_l.$$

PROOF. For δ_l as in Lemma 3.4, by the same argument as above, we have

$$P\Big\{\sup_{t\le T}\Big|\int f(x)Z_m(t\wedge\tau,dx)-\int f(x)Z_l(t\wedge\tau,dx)\Big|\ge 11\varepsilon,\,U(T)\le c\Big\}\le\delta_l,$$

for all $m > l$. This inequality and the fact that $\sum_l \delta_l < \infty$ implies by the Borel–Cantelli lemma that, with probability 1,

$$\left\{ \int f(x) Z_l(\cdot \wedge \tau, dx) \right\} \tag{3.27}$$

is a Cauchy sequence in the complete metric on $C_{\mathbb{R}}[0, \infty)$,

$$d_u(x, y) = \int_0^\infty e^{-t} \sup_{s \le t} 1 \wedge |x(s) - y(s)| \, dt,$$

giving the topology of uniform convergence on bounded time intervals. Since for each fixed t, $\{\int f(x) Z_l(t \wedge \tau, dx)\}$ converges a.s. to $\int f(x) Z(t \wedge \tau, dx)$, the lemma follows. □

4. Martingale properties. In this section, we examine more carefully the martingale properties of the processes constructed in Section 3. In particular, we consider the martingale problem satisfied by the particle model, and more importantly, the martingale problem satisfied by the measure-valued process assuming that the order of the particles is unknown. We restrict our attention to models in which the population size process is given as a function of a Markov process Q, that is, $P(t) = p(Q(t))$, where Q has state space E_0 and generator G. For simplicity, we assume that E_0 is a locally compact, separable, metric space with metric r_0 and that the strong closure of G is the generator of a Feller semigroup $\{S(t)\}$ on $\hat{C}(E_0)$ extended so that $1 \in \mathscr{D}(G)$ and $G1 = 0$. In addition, we assume

$$U(t) = \int_0^t q_1(Q(s)) \, ds + \sum_{s \le t} q_2(Q(s-), Q(s)),$$

where $q_2(v, v) = 0$, that is, $U^c(t) = \int_0^t q_1(Q(s)) \, ds$ and $\Delta U(s) = q_2(Q(s-), Q(s)) > 0$ only if Q has a discontinuity at time s. Define $\alpha(v) = q_1(v)/p(v)^2$ and $\beta(v, v') = \sqrt{q_2(v, v')}/p(v')$, that is, $\alpha(Q(t))$ is the intensity for the L_{ij} and $\beta(Q(t-), Q(t))$ is the probability that a level is involved in a birth event at time t if there is a discontinuity in Q at time t. We assume that there exists a kernel η such that, for each $\epsilon > 0$, the closure of G_ε defined by

$$G_\varepsilon f(v) = Gf(v) - \int_{\{v' : r_0(v, v') > \varepsilon\}} (f(v') - f(v)) \eta(v, dv'), \qquad f \in \mathscr{D}(G),$$

generates a Feller semigroup corresponding to a Markov process Q_ε satisfying

$$\sup_s r_0(Q_\varepsilon(s-), Q_\varepsilon(s)) \le \varepsilon \quad \text{a.s.}$$

Of course, η is just the jump intensity measure for the process. For the branching process example (Section 3.1.2), $\beta(v, v') = (v' - v)/v'$ and $\eta(v, dv') = v\nu(dv')$.

Let E denote the type space which we continue to assume is a complete, separable, metric space. We do not require B to be all of the weak infintesimal operator. We assume that $\mathscr{D}(B) \subset \overline{C}(E)$ and is separating, $1 \in \mathscr{D}(B)$ with $B1 = 0$ and $\mathscr{R}(B) \subset B(E)$. In addition, we assume that the martingale problem for B is well posed, that is, for each $\mu \in \mathscr{P}(E)$, there exists a unique solution of the martingale problem for (B, μ), and that any solution of the martingale problem for B has a modification with sample paths in $D_E[0,\infty)$. It follows immediately that the martingale problem is well posed for the operator

$$C_m f(v, x_1, \dots, x_m) = Gf(v, x_1, \dots, x_m) + \sum_{i=1}^{m} B_i f(v, x_1, \dots, x_m),$$

where G operates on f as a function of v only and B_i operates on f as a function of x_i only. [See Ethier and Kurtz (1986), Theorem 4.10.1.] Note that C_m is the generator of the process with state space $E_0 \times E^m$ consisting of Q and m independent copies of the mutation process. We can take the domain for C_m to be

$$\mathscr{D}(C_m) = \left\{ f_0(v) \prod_{i=1}^{m} f_i(x_i) \colon f_0 \in \mathscr{D}(G),\, f_i \in \mathscr{D}(B),\, i = 1, \dots, m \right\}, \tag{4.1}$$

although for some purposes we may want to extend C_m to the closure of its linear extension.

The martingale problem for the first m levels in the particle model has generator

$$\begin{aligned} A_m f(v, x^{|m}) &= C_m f(v, x^{|m}) + \sum_{1 \le i < j \le m} \alpha(v)\Big(f\big(v, \theta_{ij}(x^{|m})\big) - f(v, x^{|m})\Big) \\ &\quad + \sum_{K \subset \{1, \dots, m\}} \int_{E_0} \beta(v, v')^{|K|}(1 - \beta(v, v'))^{m - |K|} \\ &\qquad \times \big(f\big(v', \theta_K(x^{|m})\big) - f(v', x)\big)\eta(v, dv'), \end{aligned} \tag{4.2}$$

where, for $x \in E^\infty$, $x^{|m} = (x_1, \dots, x_m)$,

$$\theta_{ij}(x_1, \dots, x_m) = x_1, \dots, x_{j-1}, x_i, x_j, \dots, x_{m-1},$$

and $\theta_K(x^{|m})$ is the element in E^m obtained from $x^{|m}$ by inserting copies of $x_{\min(K)}$ at the levels in $K - \{\min(K)\}$ and dropping the $|K| - 1$ components with highest indices. If $K = \{i, j\}$, then $\theta_{ij} = \theta_K$.

If α and $\beta_\eta \equiv \int_{E_0} \beta(\cdot, v')^2 \eta(\cdot, dv')$ are bounded, there is essentially no difficulty in verifying existence and uniqueness for the martingale problem for A_m. Existence follows by a direct construction. If B satisfies the Hille–Yosida range condition, then so will C_m. The range condition for A_m then follows since A_m is a bounded perturbation of C_m. Uniqueness for the martingale problems will typically follow from Theorem 4.4.1 or Corollary 4.4.4 of Ethier and Kurtz (1986). If $\beta = 0$ (i.e., U is continuous), uniqueness follows from Theorem 4.10.3 of Ethier and Kurtz (1986), at least if the state description is expanded to include the counting processes, V_{ij}. If the mutation process can be obtained as the unique solution of a stochastic differential equation, then a system of sde's can be written for X, and uniqueness of the solution of the system used to prove uniqueness for the martingale problem. See, for example, Section 6.4.

We will simply assume that the martingale problem is well posed for A_m for any α and β such that α and β_η are bounded. Taking $\mathscr{D}(A) = \bigcup_{m=1}^{\infty} \mathscr{D}(A_m)$ and defining $Af(v, x) = A_m f(v, x^{|m})$ for $f \in \mathscr{D}(A_m)$, we see that the martingale problem for A is well posed.

Models with unbounded α and/or β_η can be treated by a localization argument. Assume that there exist open subsets $U_k \subset E_0$ such that, for each k, α and β_η are bounded on U_k and $\bigcup_{k=1}^{\infty} U_k = U \equiv \{v\colon \alpha(v) + \beta_\eta(v) < \infty\}$. (In the diffusion models discussed in Section 3, we could take $U_k = \{v\colon p(v) > k^{-1}\}$.) Let $\tau_k = \inf\{t\colon Q(t) \notin U_k \text{ or } Q(t-) \notin U_k\}$, and define $\tau = \lim_{k\to\infty} \tau_k$ [which, for the diffusion models, is the extinction time defined in (3.10)]. Then the stopped martingale problem for $(A, \nu_0, U_k \times E^\infty)$ is well posed for each k [see Ethier and Kurtz (1986), Theorem 4.6.1] and hence the sequence of stopped martingale problems uniquely determines the process up to time τ. If we assume that $\tau_k < \tau$ a.s., then τ is predictable, and since by Ethier and Kurtz [(1986), Theorem 4.3.12], (Q, X_1) is quasi-left continuous, we have $Q(\tau) = Q(\tau-)$ and $X_1(\tau) = X_1(\tau-)$ on $\{\tau < \infty\}$. (Note that A_1 is independent of α and β, so (Q, X_1) is uniquely determined for all time.) If $\int_0^\tau (\alpha(Q(s)) + \beta_\eta(Q(s)))\,ds = \infty$ (i.e., there are infinitely many lookdowns prior to time τ), then as in Theorem 6.1, $X_i(\tau-) = X_1(\tau-) = X_1(\tau)$ a.s. for each i, and we will simply define $X_i(\tau) = X_i(\tau-) = X_1(\tau)$. If $\int_0^\tau (\alpha(Q(s)) + \beta_\eta(Q(s))\,ds < \infty$, then, for each i, there is a last lookdown at or below level i at a time strictly less than τ, and the exchangeability of the historical paths discussed in Section 5.2 shows that X_i and X_1 have the same distribution on the interval between the last lookdown and τ. Consequently, we will again set $X_i(\tau) = X_i(\tau-)$.

With the understanding that we can treat unbounded α and β_η by the above localization argument, we will focus our attention on the bounded case.

Since any process of this form arises as a limit of the type discussed in Section 3, we have that, if $X_1(0), X_2(0), \dots$ is exchangeable and independent of $Q(0)$, then $X_1(t), X_2(t), \dots$ is exchangeable for each $t > 0$. More generally, we have the following.

THEOREM 4.1. *Let A_m be as above, and assume that the martingale problem for A_m is well posed. Let $\nu_0 \in \mathscr{P}(E_0 \times E^\infty)$, and let $(Q, X_1, X_2, \dots)$ be a solution of the martingale problem for (A, ν_0). If there exists a transition function η_0 from E_0 to $\mathscr{P}(E)$ such that, for all $\Gamma \in \mathscr{B}(E_0)$ and $H_i \in \mathscr{B}(E)$,*

$$\nu_0(\Gamma \times H_1 \times \cdots \times H_m \times E^\infty) = \int_\Gamma \int_{\mathscr{P}(E)} \left(\prod_{i=1}^m \mu(H_i) \right) \eta_0(v, d\mu)\, \nu_0^0(d\nu), \tag{4.3}$$

where ν_0^0 is the E_0-marginal of ν_0, then, for each $t \ge 0$, $X_1(t), X_2(t), \dots$ is an exchangeable sequence. Denoting the corresponding de Finetti measure by

$Z(t)$, we have

$$E\left[h(X_1(t),\dots,X_m(t))\middle|\mathscr{F}_t^{Q,Z}\right]$$
$$= \int\cdots\int h(x_1,\dots,x_m)Z(t,dx_1)\dots Z(t,dx_m).$$

REMARK 4.2. Note that (4.3) is essentially just the exchangeability of the initial distribution.

PROOF. The result is an immediate consequence of Proposition 3.1. □

One consequence of Theorem 4.1 is that, for $h \in \mathscr{D}(A_m)$,

$$\left\langle h(Q(t),\cdot), Z(t)^m\right\rangle - \int_0^t \left\langle A_m h(Q(s),\cdot), Z(s)^m\right\rangle ds$$

is a martingale with respect to the filtration $\{\mathscr{F}_t^{Q,Z}\}$. Consequently, if we define an opertor

$$\mathbb{A}: \mathscr{D}(\mathbb{A}) \subset C(E_0 \times \mathscr{P}(E)) \to B(E_0 \times \mathscr{P}(E))$$

by taking

$$\mathscr{D}(\mathbb{A}) = \left\{F: F(v,\mu) = \langle h(v,\cdot), \mu^m\rangle,\ h \in \mathscr{D}(A_m),\ m = 1,2,\dots\right\}$$

and defining

$$\mathbb{A}F(v,\mu) = \langle A_m h(v,\cdot), \mu^m\rangle, \tag{4.4}$$

we have existence of solutions of the martingale problem for $\mathbb{A}$. Note that for α a constant, $\mathbb{A}$ gives the standard martingale problem for the Fleming–Viot process. [See, e.g., Ethier and Kurtz (1993).] We now consider the conditions for uniqueness of the martingale problem.

THEOREM 4.3. *Let α and β_η be bounded and suppose that there exists a $\lambda > 0$ such that $\mathscr{R}(\lambda - B)$ is bp-dense in $B(E)$.* [*Recall that we are assuming that the closure $\overline{G}$ of G generates a Feller semigroup on $\hat{C}(E_0)$, so $\mathscr{R}(\lambda - G) = \hat{C}(E_0)$ for every $\lambda > 0$.*] *Then, for every $\lambda > 0$, $\mathscr{R}(\lambda - \mathbb{A})$ is bp-dense in $B(E_0 \times \mathscr{P}(E))$ and the martingale problem for $\mathbb{A}$ is well posed.*

PROOF. Note that the conclusion of the theorem is valid if and only if it is valid with G replaced by $\overline{G}$, so without loss of generality, we assume $G = \overline{G}$. If X is a solution of the martingale problem for B, then it is a solution of the martingale problem for $\hat{B}$, the *bp*-closure of B. [See Ethier and Kurtz (1986), Proposition 4.3.1. In general, $\hat{B}$ will be multivalued and should be considered to be a set of ordered pairs; however, we will continue to use the more intuitive notation $\hat{B}f$.] Since the martingale problem for B is well posed, for $h \in \mathscr{R}(\lambda - B)$ we have

$$(\lambda - B)^{-1}h(x) = E\left[\int_0^\infty e^{-\lambda t} h(X_x(t))\,dt\right], \tag{4.5}$$

where X_x is a solution of the martingale problem for (B, δ_x), which, by assumption, we can take to be right continuous. Consequently, if $\{h_n\} \subset \mathscr{R}(\lambda - B)$ and $h = bp\text{-}\lim_{n\to\infty} h_n$, then the bp-limits of $f_n = (\lambda - B)^{-1}h_n$ and $Bf_n = \lambda f_n - h_n$ will exist. It follows from the assumption that $\mathscr{R}(\lambda - B)$ is bp-dense in $B(E)$ that $\hat{B}$ satisfies $\mathscr{R}(\lambda - \hat{B}) = B(E)$. Consequently, we may as well assume that B is bp-closed and hence that $\mathscr{R}(\lambda - B) = B(E)$. But if this condition holds for one $\lambda > 0$, it holds for all $\lambda > 0$ [See Ethier and Kurtz (1986), Lemma 4.2.3. Note that $(\lambda - \hat{B})^{-1}$ will be single-valued even if $\hat{B}$ is multivalued.]

Let $B_0 = \{(f, g) \in \hat{B}\colon g \in \overline{\mathscr{D}(\hat{B})}\}$. Then B_0 generates a strongly continuous (in the sup norm) contraction semigroup on $L = \overline{\mathscr{D}(\hat{B})}$. [Ethier and Kurtz (1986), Theorem 1.4.3.] The fact that X_x is right continuous implies $\mathscr{D}(\hat{B})$ is bp-dense in $B(E)$, so linear combinations of functions of the form

$$f(v, x_1, \dots, x_m) = f_0(v) f_1(x_1) \dots f_m(x_m),$$
$$f_0 \in \mathscr{D}(G),\ f_i \in \mathscr{D}(B_0),\ 1 \le i \le m,$$

will be bp-dense in $B(E_0 \times E^m)$. Call this collection of functions D_m, and note that the semigroup $\{S_m(t)\}$ corresponding to the process $(Q, X_1, \dots, X_m)$, where $X_1, \dots, X_m$ are independent solutions of the martingale problem for B, maps D_m into D_m. It follows by Ethier and Kurtz [(1986), Proposition 1.3.4], that the closure of C_m restricted to the linear span of D_m generates a strongly continuous contraction semigroup on L_m, the closure of the linear span of D_m. Since the strong closure of the linear span of $\mathscr{R}(\lambda - C_m)$ contains L_m, the bp-closure must equal $B(E_0 \times E^m)$. Finally, since A_m is a bounded perturbation of C_m, it follows that the linear span of $\mathscr{R}(\lambda - A_m)$ is bp-dense in $B(E_0 \times E^m)$. Since, for $f \in \mathscr{D}(A_m)$ and $F(v, \mu) = \langle f(v, \cdot), \mu^m \rangle$,

$$\lambda F(v, \mu) - \mathbb{A} F(v, \mu) = \left\langle \lambda f(v, \cdot) - A_m f(v, \cdot), \mu^m \right\rangle,$$

it follows that $\mathscr{R}(\lambda - \mathbb{A})$ is bp-dense in $B(E_0 \times \mathscr{P}(E))$. Consequently, by Ethier and Kurtz [(1986), Theorem 4.4.1], the martingale problem for $\mathbb{A}$ is well posed. □

THEOREM 4.4. *Let D_0 be a countable subset of $\mathscr{D}(G)$ that separates points in E_0, is closed under multiplication and vanishes nowhere, and let D_1 be a countable subset of $\mathscr{D}(B)$ that separates points in E, is closed under multiplication and vanishes nowhere. Suppose that, for $f \in D_1$, $Bf \in \overline{C}(E)$ and that the martingale problems for G restricted to D_0 and for B restricted to D_1 are well posed. If α and β are bounded and continuous and $\eta(v, \cdot)$ is weakly continuous in v, then the martingale problem for $\mathbb{A}$ is well posed.*

PROOF. Recall that we are assuming that $\mathscr{D}(B) \subset \overline{C}(E)$ and $\mathscr{D}(G) \subset \overline{C}(E_0)$. Note that the martingale problem for A restricted to the domain generated by D_0 and D_1 is still well posed and satisfies the conditions of Theorem 2.6 of Kurtz (1998) [i.e., Bhatt and Karandikar (1993), Theorem 4.1]. Define $\gamma(v, x_1, x_2, \dots) = (v, \mu) \in E_0 \times \mathscr{P}(E)$ if the limit

$$\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$$

exists, and define $\gamma(v, x_1, x_2, \dots) = (v, \delta_a)$ for some fixed $a \in E$ otherwise. Then uniqueness for the martingale problem for $\mathbb{A}$ follows by Kurtz (1998), Corollary 3.7, where the mapping γ is defined above and the transition function α (not to be confused with α in the present paper) is determined by

$$\alpha(v, \mu, \Gamma_0 \times \Gamma_1 \times \cdots \times \Gamma_m \times E^\infty) = \delta_v(\Gamma_0) \prod_{i=1}^{m} \mu(\Gamma_i). \qquad \square$$

If Q is continuous, the formulation of the martingale problem for $\mathbb{A}$ can be simplified [cf. El Karoui and Roelly (1991)].

THEOREM 4.5. *Let (Q, Z) be a process with sample paths in $C_{E_0\times\mathscr{P}(E)}[0,\infty)$, and assume that $\mathscr{D}(G)$ is an algebra. Suppose that, for $f_0 \in \mathscr{D}(G)$ and $f_1 \in \mathscr{D}(B)$,*

$$f_0(Q(t))\langle f_1, Z(t)\rangle - \int_0^t \langle f_1(\cdot)Gf_0(Q(s)) + f_0(Q(s))Bf_1(\cdot), Z(s)\rangle\, ds$$

is a continuous $\{\mathscr{F}_t^{Q,Z}\}$-martingale with quadratic variation

$$\int_0^t \Big(\langle f_1, Z(s)\rangle^2 \big(Gf_0^2(Q(s)) - 2f_0(Q(s))Gf_0(Q(s))\big)$$
$$+\alpha(Q(s))f_0^2(Q(s))\big(\langle f_1^2, Z(s)\rangle - \langle f_1, Z(s)\rangle^2\big)\Big)\, ds.$$

Then (Q, Z) is a solution of the martingale problem for $\mathbb{A}$.

PROOF. Apply Itô's formula to $\langle h(Q(t), \cdot), Z(t)^m\rangle$ for $h \in \mathscr{D}(A_m) = \mathscr{D}(C_m)$ defined by (4.1). $\square$

EXAMPLE 4.6 (Dawson–Watanabe process). Let $E_0 = [0, \infty)$, $Gf_0(v) = avf_0''(v) + bvf_0'(v)$, and let $p(v) = v$, that is, the population size is given by Q. Then $C_c^\infty[0,\infty)$, the space of continuously differentiable functions with compact support in $[0,\infty)$, is a core for G. Note also that 0 is absorbing, that is, if $\tau = \inf\{t\colon Q(t) = 0\}$, then $Q(t) = 0$ for all $t > \tau$. Let K be an $\mathscr{M}(E)$-valued process such that, for $f \in \mathscr{D}(B)$,

$$\langle f, K(t)\rangle - \int_0^t \langle bf + Bf, K(s)\rangle\, ds$$

is a continuuous $\{\mathscr{F}_t^K\}$-martingale with quadratic variation

$$\int_0^t 2a\langle f^2, K(s)\rangle\, ds.$$

Taking $f = 1$, and setting $Q(t) = |K(t)|$ [the total mass of $K(t)$] and $Z(t) = |K(t)|^{-1}K(t)$, we see that

$$Q(t) - \int_0^t bQ(s)\,ds$$

is a continuous martingale with quadratic variation

$$\int_0^t 2aQ(s)\,ds.$$

Consequently, Q is a solution of the martingale problem for G. Let $U_k = (k^{-1}, \infty)$, and set $\alpha(v) = 2a/v$, $v \neq 0$. Applying Itô's formula, we can see that (Q, Z) satisfies the martingale conditions of Theorem 4.5, provided we stop the process at $\tau_k = \inf\{t: Q(t) \notin U_k\}$. Consequently, $(Q(\cdot \wedge \tau_k), Z(\cdot \wedge \tau_k))$ is a solution of the stopped martingale problem for $(\mathbb{A}, U_k \times \mathscr{P}(E))$. Since Q absorbs at zero, if B satisfies the conditions of Theorem 4.3 or 4.4, the martingale conditions on K uniquely determine its distribution for any initial distribution on $\mathscr{M}(E)$. This result is a special case of the characterization of the Dawson–Watanabe process in El Karoui and Roelly (1991). □

EXAMPLE 4.7 (General diffusion population size). More generally, let $E_0 = [0, \infty)$, $p(v) = v$ and $Gf_0(v) = a(v)f_0''(v) + b(v)f_0'(v)$, and assume that $C_c^\infty[0, \infty)$ is a core for G. [See Ethier and Kurtz (1986), Theorem 8.2.1 for sufficient conditions on a and b.] Setting $\hat{b}(v) = b(v)/v$ and $\hat{a}(v) = a(v)/v$, let K be an $\mathscr{M}(E)$-valued process such that, for $f \in \mathscr{D}(B)$,

$$\langle f, K(t)\rangle - \int_0^t \left(\hat{b}(|K(s)|)\langle f, K(s)\rangle + \langle Bf, K(s)\rangle\right) ds$$

is a continuous $\{\mathscr{F}_t^K\}$-martingale with quadratic variation

$$\int_0^t 2\hat{a}(|K(s)|)\langle f^2, K(s)\rangle\,ds.$$

As in Example 4.6, let $Q(t) = |K(t)|$ and $Z(t) = |K(t)|^{-1}K(t)$. Taking $f = 1$, we see that

$$Q(t) - \int_0^t b(Q(s))\,ds$$

is a continuous martingale with quadratic variation

$$\int_0^t 2a(Q(s))\,ds,$$

and hence Q is a solution of the martingale problem for G. Setting $\alpha(v) = 2v^{-2}a(v)$, $U_k = \{v: \alpha(v) \le k\}$ and $\tau_k = \inf\{t: Q(t) \notin U_k\}$, $(Q(\cdot \wedge \tau_k), Z(\cdot \wedge \tau_k))$ is a solution of the stopped martingale problem for $(\mathbb{A}, U_k \times \mathscr{P}(E))$.

5. Genealogy.

5.1. *The Coalescent.* Let L_{ij} and L_K^l be defined as in (3.18) and (3.20). For each $t \ge 0$ and $k = 1, 2, \ldots,$ let $N_k^t(s)$, $0 \le s \le t$, be the level at time s of the ancestor of the particle at level k at time t. In terms of the L_{ij}, for $0 \le s \le t$,

$$\begin{aligned} N_k^t(s) = k &- \sum_{1 \le i < j < k} \int_s^t I_{\{N_k^t(u) > j\}} \, dL_{ij}(u) \\ &- \sum_{1 \le i < j \le k} \int_s^t (j - i) I_{\{N_k^t(u) = j\}} \, dL_{ij}(u) \\ &- \sum_{K \subset \{1, \dots, k\}} \int_s^t \big(N_k^t(u) - \min(K)\big) I_{\{N_k^t(u) \in K\}} \, dL_K^k(u) \\ &- \sum_{K \subset \{1, \dots, k\}} \int_s^t \big(\big|K \cap \{1, \dots, N_k^t(u)\}\big| - 1\big) \\ &\qquad \times I_{\{N_k^t(u) > \min(K),\, N_k^t(u) \notin K\}} \, dL_K^k(u). \end{aligned}$$

Fix $0 < t \le \tau$, and, for $s \le t$, define an equivalence relation, $\tilde{R}^t(s)$, by

$$\tilde{R}^t(s) = \big\{(k, l) \colon k, l = 1, 2, \dots, N_k^t(s) = N_l^t(s)\big\}. \tag{5.1}$$

Informally, $(k, l) \in \tilde{R}^t(s)$ iff the two levels k and l have the same ancestor at time s.

THEOREM 5.1. *Assume that U is continuous and that $t < \tau$. Let $\nu^t(u)$ be the time change determined for $u \le H(t) \equiv \int_0^t (1/P(s)^2) \, dU(s)$ by*

$$\int_{\nu^t(u)}^t \frac{1}{P(s)^2} \, dU(s) = u.$$

Up to time $H(t)$, the process R^t defined by $R^t(u) = \tilde{R}^t(\nu^t(u))$ is Kingman's (1982) *coalescent.*

PROOF. Observe that $R(0) = \{(i, i),\ i = 1, 2, \dots\}$. Define

$$V_{ij}^t(u) = V_{ij}\left(\int_0^t \frac{1}{P(s)^2} \, dU(s)\right) - V_{ij}\left(\int_0^{\nu^t(u)} \frac{1}{P(s)^2} \, dU(s)\right).$$

Since $V_{ij}^t(u)$ is the increment of a unit Poisson process over a (random) time interval of length u for which the location of the interval is independent of the process V_{ij}, it follows that (the right continuous modification of) V_{ij}^t is a unit Poisson process stopped at $H(t)$. Further, these processes are independent for distinct pairs (i, j).

The result then follows as in Theorems 3.1 and 3.2 of Donnelly and Kurtz (1996). □

Pitman (1997) considers coalescent models with multiple collisions. His models are given by (5.1), if the underlying population model is that described in Section 3.1.4. In particular, the finite measure Λ in the definition of Pitman's coalescent is related to ν by the identity

$$\int_0^1 g(x) \Lambda(dx) = a g(0) + \int_0^1 g(\sqrt{v}) v \nu(dv).$$

[Recall that $\int_0^1 v\nu(dv) < \infty$.] Compare (3.17) with the definition of $\lambda_{b,K}$ in Pitman's paper.

Suppose U is continuous and $\tau = \infty$. Then, on the event $\{\lim_{t\to\infty} H(t) = \infty\}$, R^t converges in distribution, as $t \to \infty$, to Kingman's coalescent. In particular, under these conditions, for large enough t, all the levels at time t share a common ancestor at time 0. If $\tau < \infty$ and $H(\tau) = \infty$, then with

$$\tau_\varepsilon = \inf\{t: P(t) < \varepsilon\},$$

R^{τ_ε} converges in distribution to Kingman's coalescent as $\varepsilon \to 0$. In particular, under these conditions, for some time t sufficiently close to τ, all the levels at time t share a common ancestor at time 0.

Dropping the continuity assumption on U, the existence of a common ancestor at time 0 will still hold on the event

$$\lim_{t\to\infty} \int_0^t \frac{1}{P(s)^2}\, dU^c(s) = \infty, \tag{5.2}$$

although $\tilde{R}^t$ can no longer be transformed by a time change to Kingman's coalescent. If (5.2) fails, then the question of the existence of a time at which all particles have a common ancestor at time 0 becomes more delicate. See Pitman [(1997), Section 3.6] for a discussion of this question for the models of Section 3.1.4, that is, models in which $P \equiv 1$ and U has stationary, independent increments.

The property that all the levels can be traced back to a single common ancestor in finite time is closely related to the ergodicity of the particle process, since it facilitates the coupling of versions of the process for different initial conditions. For details of the argument, see Donnelly and Kurtz (1996), Section 4. For example, under the assumptions on P and U in the first paragraph of Section 4, suppose that the process Q is strongly ergodic (i.e., for each initial distribution, the one-dimensional distributions converge in total variation to the unique stationary distribution) and, in addition, that the type/location process is strongly ergodic. Then the proof of Theorem 4.1 of Donnelly and Kurtz (1996) is easily extended to show that, if all the levels can be traced back to a single common ancestor in finite time, the particle process, and the associated measure-valued process, are also strongly ergodic.

5.2. *The Dawson–Perkins historical process.* Let N_k^t be as above. For each $t \ge 0$ and $k = 1, 2, \dots,$ define

$$\tilde{X}_k^t(s) = X_{N_k^t(s)}(s), \qquad 0 \le s \le t.$$

Then $\tilde{X}_k^t$, as a process on the interval $[0, t]$, is Markov with generator B, and the sequence $\{\tilde{X}_k^t\}$ is exchangeable as a sequence of $D_E[0, t]$-valued random variables. (Alternatively, we can define $\tilde{X}_k^t(s) = \tilde{X}_k^t(t)$ for $s \ge t$ and consider $\tilde{X}_k^t$ as a $D_E[0,\infty)$-valud random variable.) Let $\tilde{K}(t)$ denote the de Finetti measure corresponding to the sequence, and define $K(t) = P(t)\tilde{K}(t)$. In the branching case, K, viewed as an $\mathscr{M}(D_E[0,\infty))$-valued process, is the historical process of Dawson and Perkins (1991).

6. Applications and examples.

6.1. *Type distribution at the extinction time.* In the case of super Brownian motion, Tribe (1992) has shown that

$$\lim_{t\to\tau-} Z(t) = \delta_{\zeta_0} \quad \text{a.s.}$$

for some $\mathbb{R}^d$-valued random variable ζ_0. From the above construction, it is easy to see that $\zeta_0 = X_1(\tau)$. More generally, we have the following theorem.

THEOREM 6.1. *The limit*

$$\lim_{t\to\tau-} Z(t) = \delta_{X_1(\tau-)}$$

holds almost surely on $\{\tau < \infty\}$ *if and only if*

$$\int_0^\tau \frac{1}{P(s)^2}\, dU(s) = \int_0^\tau \frac{1}{P(s)^2}\, dU^c(s) + \sum_{s<\tau} \frac{\Delta U(s)}{P(s)^2} = \infty \tag{6.1}$$

holds almost surely on $\{\tau < \infty\}$.

PROOF. Let $N_{12}(t) = L_{12}(t) + L^2_{\{1,2\}}(t)$. By (6.1), for each $\varepsilon > 0$, $N_{12}(\tau) - N_{12}(\tau-\varepsilon) = \infty$ on $\{\tau < \infty\}$. It follows that $X_2(\tau-) = X_1(\tau-)$ a.s. But since τ is $\mathscr{G}_0$-measurable, it is $\{\mathscr{G}_t\}$-predictable and $(X_1(\tau-), X_2(\tau-), \dots)$ is exchangeable by Proposition 3.1. Consequently, we must have $X_k(\tau-) = X_1(\tau-)$ for all k and hence $Z(\tau-) = \delta_{X_1(\tau-)}$. □

6.2. *Conditioning*. In general, the effect of conditioning on U and P is clear. The only impact on the process is through the time change in the definition of L_{ij} at (3.18) and through the definition of L^l_K at (3.20). For example, if the original process is the Dawson–Watanabe process so that

$$\int_0^t \frac{1}{P(s)^2}\, dU(s) = \int_0^t \frac{c}{P(s)}\, ds,$$

then conditioning on $P \equiv 1$ is equivalent to setting $L_{ij}(t) \equiv V_{ij}(ct)$, which makes Z the Fleming–Viot (genetic) process, a result due to Etheridge and March (1991). See Perkins (1991) for related results.

Again, in the Dawson–Watanabe setting, conditioning P on nonextinction [cf. Evans and Perkins (1990) and Section 6.3] is equivalent to replacing P with generator $Gf(v) = avf''(v) - bvf'(v)$ $(b \ge 0)$ by a process with generator

$$\hat{G}f(v) = avf''(v) + (2a - bv)f'(v). \tag{6.2}$$

If $P(0) > 0$, then P never hits zero, but P grows slowly enough that

$$\int_0^\infty \frac{c}{P(s)}\, ds = \infty.$$

It follows that eventually all particles trace their ancestry back to the bottom-level particle. In particular, the bottom-level particle in our construction is the "immortal particle" of Evans (1993), and if $b > 0$, the ergodicity argument outlined in Section 5.1 applies whenever X_1 is ergodic.

6.3. *Asymptotic independence.* The following theorem extends a result of Evans and Perkins (1990) for critical superprocesses conditioned on nonextinction. As noted in Section 6.2, conditioning a Dawson–Watanabe process on nonextinction is equivalent to letting P be the diffusion with generator (6.2). It follows that in the critical case (i.e., $b = 0$), for $\alpha > 0$, $P^{\alpha}(t) = P(\alpha t)/\alpha$ is a diffusion with generator $\hat{G}$, and hence, as $\alpha \to \infty$, $P^{\alpha} \Rightarrow P_0$, where P_0 is the diffusion with generator $\hat{G}$ and $P(0) = 0$. Note that $P_0(t) > 0$ for all $t > 0$, and, consequently, that (6.3) below is satisfied.

THEOREM 6.2. *Suppose that $\tau = \infty$ a.s. and that the type process has stationary distribution π and is ergodic in the sense that $\lim_{t\to\infty} T(t)f(x) = \int f\,d\pi$ for every $f \in \overline{C}(E)$. Assume either that the convergence is uniform on compact subsets of E or that $X_1(0)$ has distribution π (that is, X_1 is stationary with marginal distribution π) and the convergence holds almost everywhere π.*

If

$$\lim_{t\to\infty} \int_t^{t+r} \frac{1}{P(s)^2}\, dU(s) = 0 \tag{6.3}$$

in probability for each $r > 0$, then $\lim_{t\to\infty} Z(t) = \pi$ in probability.

PROOF. Again, let $N_{12}(t) = L_{12}(t) + L^2_{\{1,2\}}(t)$. Let $\gamma(t) = \sup\{s \le t: N_{12}(s) \ne N_{12}(s-)\}$. Then by (6.3),

$$\lim_{t\to\infty} P\{t - \gamma(t) > r\} = 1$$

for each $r > 0$. By the conditional independence of the type processes and the assumption on $\{T(t)\}$, for each $f \in \overline{C}(E)$,

$$\begin{aligned}
\lim_{t\to\infty} E\left[\langle f, Z(t)\rangle^2\right] &= \lim_{t\to\infty} E\left[f(X_1(t))f(X_2(t))\right]\\
&= \lim_{t\to\infty} E\left[\left(T(t-\gamma(t))f(X_1(\gamma(t)))\right)^2\right]\\
&= \langle f, \pi\rangle^2.
\end{aligned}$$

[Note that since $\gamma(t)$ is independent of X_1, if X_1 is stationary with marginal distribution π, then $X_1(\gamma(t))$ has districution π.] But since $E[\langle f, Z(t)\rangle] \to \langle f, \pi\rangle$, it follows by the Chebychev inequality that $Z(t) \to \pi$ in probability. □

6.4. *Sochastic equations for diffusion type processes.* Let L_{ij} and L^k_K be defined as in (3.18) and (3.20). Suppose the type process is a diffusion in $\mathbb{R}^d$

given as the unique solution of an Itô equation,

$$X_0(t) = X_0(0) + \int_0^t \sigma(X_0(s))\,dW(s) + \int_0^t b(X_0(s))\,ds,$$

where σ is $d \times d$-matrix-valued function, W is a standard Brownian motion in $\mathbb{R}^d$, and b is an $\mathbb{R}^d$-valued function. Then the particle process satisfies the system of equations

$$\begin{aligned}
X_k(t) = {} & X_k(0) + \int_0^t \sigma(X_k(s))\,dW_k(s) + \int_0^t b(X_k(s))\,ds \\
& + \sum_{1 \le i < k} \int_0^t (X_i(s-) - X_k(s-))\,dL_{ik}(s) \\
& + \sum_{1 \le i < j < k} \int_0^t (X_{k-1}(s-) - X_k(s-))\,dL_{ij}(s) \\
& + \sum_{K \subset \{1,\dots,k\},\, k \in K} \int_0^t \big(X_{\min(K)}(s-) - X_k(s-)\big)\,dL_K^k(s) \\
& + \sum_{K \subset \{1,\dots,k\},\, k \notin K} \int_0^t \big(X_{k-|K|+1}(s-) - X_k(s-)\big)\,dL_K^k(s),
\end{aligned} \tag{6.4}$$

where the W_k are independent, $\mathbb{R}^d$-valued, standard Brownian motions.

6.5. *Measure-valued diffusion with spatial interaction.* In the Dawson–Watanabe setting, Perkins (1992) introduced stochastic equations driven by historical Brownian motion, that is, the historical process with Brownian location process (see Section 5.2). In our context, we can modify (6.4) to obtain a version of Perkins's models corresponding to our more general population models. We assume, for simplicity, that U is continuous and write

$$\begin{aligned}
X_k(t) = {} & X_k(0) + \int_0^t \sigma(P(s), Z(s), X_k(s))\,dW_k(s) \\
& + \int_0^t b(P(s), Z(s), X_k(s))\,ds \\
& + \sum_{1 \le i < k} \int_0^t (X_i(s-) - X_k(s-))\,dL_{ik}(s) \\
& + \sum_{1 \le i < j < k} \int_0^t (X_{k-1}(s-) - X_k(s))\,dL_{ij}(s).
\end{aligned} \tag{6.5}$$

For each $t \ge 0$, we require the solution $\{X_k(t)\}$ to be exchangeable with de Finetti measure $Z(t)$. The connection of this system to the equation of Perkins is more obvious if we first define

$$\tilde{W}_k^t(s) = \sum_{i=1}^k \int_0^s I_{\{N_k^t(u) = i\}}\,dW_i(u), \qquad 0 \le s \le t,$$

and note that the de Finetti measure for $\{\tilde{W}_k^t\}$ multiplied by $P(t)$ gives historical Brownian motion. Then, for each $t \geq 0$, $X_k(t) = \tilde{X}_k^t(t)$, where

$$\tilde{X}_k^t(s) = \tilde{X}_k^t(0) + \int_0^s \sigma\big(P(u), Z(u), \tilde{X}_k^t(u)\big)\, d\tilde{W}_k^t(u) + \int_0^s b\big(P(u), Z(u), \tilde{X}_k^t(u)\big)\, du. \tag{6.6}$$

Note that in (6.6), $Z(u)$ is still the de Finetti measure for $\{X_k(u)\}$, not that of $\{\tilde{X}_k^t(u)\}$. In the branching setting, (6.6) is essentially equation (SE) of Perkins (1992). Perkins aso considered more general equations in which the coefficients depend on the past of the processes.

Following Perkins, let ρ_w denote the Wasserstein metric on $\mathscr{P}(\mathbb{R}^d)$ and assume

$$|\sigma(p, z_1, x_1) - \sigma(p, z_2, x_2)| + |b(p, z_1, x_1) - b(p, z_2, x_2)| \leq K(\rho_w(z_1, z_2) + |x_1 - x_2|), \tag{6.7}$$

for $z_1, z_2 \in \mathscr{P}(\mathbb{R}^d)$ and $x_1, x_2 \in \mathbb{R}^d$. Consider the n-dimensional system, $1 \leq k \leq n$,

$$\begin{aligned} X_k^n(t) = X_k(0) &+ \int_0^t \sigma(P(s), Z^n(s), X_k^n(s))\, dW_k(s) \\ &+ \int_0^t b(P(s), Z^n(s), X_k^n(s))\, ds \\ &+ \sum_{1 \leq i < k} \int_0^t (X_i^n(s-) - X_k^n(s-))\, dL_{ik}(s) \\ &+ \sum_{1 \leq i < j < k} \int_0^t (X_{k-1}^n(s-) - X_k^n(s))\, dL_{ij}(s), \end{aligned}$$

where $Z^n(s) = (1/n)\sum_{k=1}^n \delta_{X_k^n(s)}$. The Lipschitz assumption (6.7) implies existence and uniqueness for (6.8) below.

Suppose that there exists a solution of (6.5), and note that, as in (6.6), $X_k^n(t) = \tilde{X}_k^{n,t}(t)$, where

$$\tilde{X}_k^{n,t}(s) = \tilde{X}_k^t(0) + \int_0^s \sigma\big(P(u), Z^n(u), \tilde{X}_k^{n,t}(u)\big)\, d\tilde{W}_k^t(u) + \int_0^s b\big(P(u), Z^n(u), \tilde{X}_k^{n,t}(u)\big)\, du. \tag{6.8}$$

By (6.7) and the usual Lipschitz estimates for Itô equations, for each $T > 0$, there exists a constant K_T such that, for all $0 \leq s \leq t \leq T$,

$$E\Big[\big|\tilde{X}_k^t(s) - \tilde{X}_k^{n,t}(s)\big|^2\Big] \leq K_T \int_0^s E\Big[\rho_w^2(Z(u), Z^n(u)) + \big|\tilde{X}_k^t(u) - \tilde{X}_k^{n,t}(u)\big|^2\Big]\, du,$$

and hence, by Gronwall's inequality, for $0 \le t \le T$,

$$\begin{aligned} E\Big[\big|X_k(t) - X_k^n(t)\big|^2\Big] &= E\Big[\big|\tilde{X}_k^t(t) - \tilde{X}_k^{n,t}(t)\big|^2\Big] \\ &\le \exp(TK_T)\int_0^t E\big[\rho_w^2(Z(u), Z^n(u))\big]\,du. \end{aligned} \tag{6.9}$$

Now let $\hat{Z}^n(t) = (1/n)\sum_{k=1}^n \delta_{X_k(t)}$, and note that

$$\rho_w^2\big(\hat{Z}^n(t), Z^n(t)\big) \le \frac{1}{n}\sum_{k=1}^n \big(X_k(t) - X_k^n(t)\big)^2.$$

By (6.9),

$$\begin{aligned} E\Big[\rho_w^2\big(\hat{Z}^n(t), Z^n(t)\big)\Big] \le 2\exp(TK_T)\int_0^t \Big(&E\Big[\rho_w^2\big(Z(u), \hat{Z}^n(u)\big)\Big] \\ &+E\Big[\rho_w^2\big(\hat{Z}^n(u), Z^n(u)\big)\Big]\Big)\,du, \end{aligned}$$

and again by Gronwall's inequality, we have

$$E\Big[\rho_w^2\big(\hat{Z}^n(t), Z^n(t)\big)\Big] \le \exp(t2e^{TK_T})\int_0^t E\Big[\rho_w^2\big(Z(u), \hat{Z}^n(u)\big)\Big]\,du. \tag{6.10}$$

By the requirement that $\{X_k(u)\}$ be exchangeable with de Finetti measure $Z(u)$, the right-hand side of (6.10) goes to zero as $n \to \infty$. It follows that the right-hand side of (6.9) goes to zero also, which, in particular, implies uniqueness for (6.5). Existence for (6.5) follows by using much the same argment to show that $\{Z^n(t)\}$ is a Cauchy sequence for each t.

Assume that P and U satisfy the conditions of Section 4. To simplify notation, assume that σ and b depend explicitly on Q rather than P, and set $a(v, z, x) = \sigma(v, z, x)\sigma(v, z, x)^T$. For $f \in \mathscr{D}(B) \equiv C_c^2(\mathbb{R}^d)$, define

$$Bf(v, z, x) = \tfrac{1}{2}\sum_{i,j=1}^d a_{ij}(v, z, x)\,\partial_i\partial_j f(x) + \sum_{i=1}^d b_i(v, z, x)\,\partial_i f(x).$$

The generator A for (Q, Z, X) is given by the obvious modification of (4.2). In formulating the corresponding martingale problem, we require that a solution have the exchangeability property, so that $Z(t)$ is defined to be the de Finetti measure for $X(t)$. Under the conditions above on σ and b, uniqueness for the system (6.5) implies uniqueness for the corresponding martingale problem. [Every solution of the martingale problem is a weak solution of (6.5).] If we define $\mathbb{A}$ as in (4.4), uniqueness for the martingale problem for $\mathbb{A}$ follows by the same argument as used in the proof of Theorem 4.4.

6.6. *Models with immigration.* A particle representation for models with immigration can be constructed in much the same way as for models without. We simply insert new "immigrants" at each level at a rate that is independent of the level or the current type at the level. In the case $\beta = 0$, the generator (4.2) becomes

$$A_m f(v, x^{|m}) = C_m f(v, x^{|m}) + \sum_{1 \le i < j \le m} \alpha(v)\Big(f\big(v, \theta_{ij}(x^{|m})\big) - f(v, x^{|m})\Big)$$
$$+ \sum_{i=1}^{m} \gamma(v) \int_E \Big(f\big(v, \theta_i(x^{|m}|y)\big) - f(v, x^{|m})\Big) q(v, dy),$$

where $\theta_i(x_1, \dots, x_m | y) = (x_1, \dots, x_{i-1}, y, x_i, \dots, x_{m-1})$ and q is a transition function from E_0 to E which gives the distribution of the type of the immigrant conditioned on the value of the Markov driving process Q. If α and γ are bounded, then uniqueness of the martingale problem for A_m will typically follow under the same conditions as in the case $\gamma = 0$ and the exchangeability results follow also. In particular, if we define $\mathbb{A}$ as before, that is, $\mathbb{A}F(v, \mu) = \langle A_m h(v, \cdot), \mu^m \rangle$, then Theorem 4.3 extends to the model with immigration under the assumption that α and γ are bounded. If γ is bounded and continuous and the mapping $v \to q(v, \cdot)$ from E_0 into $\mathscr{P}(E)$ is continuous, then Theorem 4.4 extends as well.

If the original finite population model is a branching Markov process with constant immigration rate and iid immigrant types with distribution q_0, then $Gf_0(v) = avf_0''(v) + (bv + c)f_0'(v)$, $p(v) = v$, $\alpha(v) = 2a/v$, $q(v, dy) = q_0(dy)$ and $\gamma(v) = c/v$ for some constant c. Defining $K = QZ$, for $f \in \mathscr{D}(B)$,

$$\langle f, K(t) \rangle - \int_0^t \big(\langle bf + Bf, K(s) \rangle + c\langle f, q_0 \rangle\big)\, ds \tag{6.11}$$

is a continuous $\{\mathscr{F}_t^K\}$-martingale with quadratic variation

$$\int_0^t 2a\langle f^2, K(s) \rangle\, ds. \tag{6.12}$$

As in Example 4.6, if B satisfies the conditions of Theorem 4.3 or 4.4, (6.11) and (6.12) determine a well-posed martingale problem.

Models with migration between colonies will be treated elsewhere.

APPENDIX

LEMMA A.1. *For each n, let $N_1^n, \dots, N_m^n$ be counting processes satisfying $[N_i^n, N_j^n]_t = 0$ for $i \neq j$ (i.e., there are no simultaneous jumps). Suppose that $\{H_i^n\}$ are nondecreasing processes with $H_i^n(t) - H_i^n(t-) \le 1$ for all i and $t \ge 0$, that*

$$N_i^n - H_i^n, \qquad i = 1, \dots, m,$$

are $\{\mathscr{G}_t^n\}$-martingales and that $H_i^n(t)$ is $\mathscr{G}_0^n$-measurable for each i and $t \ge 0$. If

$$(H_1^n, \dots, H_m^n) \Rightarrow H = (H_1, \dots, H_m)$$

in the Skorohod topology on $D_{\mathbb{R}^m}[0, \infty)$, then

$$(N_1^n, \dots, N_m^n) \Rightarrow (N_1, \dots, N_m),$$

where $(N_1, \dots, N_m)$ *are counting processes with joint distribution determined by*

$$\varphi_f(t) = E\left[\exp\left(-\sum_{i=1}^m \int_0^t f_i(s)\, dN_i(s)\right)\middle| H\right]$$
$$= 1 + \sum_{i=1}^m \int_0^t \varphi_f(u-)\big(\exp(-f_i(u)) - 1\big)\, dH_i(u)$$

for all nonnegative, continuous, $\mathbb{R}^m$*-valued functions* $f = (f_1, \dots, f_m)$.

Proof. Using the fact that there are no simultaneous jumps among the N_i^n,

$$\begin{aligned}
&\exp\left(-\sum_{i=1}^m \int_0^t f_i(s)\, dN_i^n(s)\right)\\
&\quad = 1 + \sum_{j=1}^m \int_0^t \big(\exp(-f_j(u)) - 1\big)\exp\left(-\sum_{i=1}^m \int_0^{u-} f_i(s)\, dN_i^n(s)\right)\\
&\qquad\qquad \times dN_j^n(u)\\
&\quad = 1 + \sum_{j=1}^m \int_0^t \big(\exp(-f_j(u)) - 1\big)\exp\left(-\sum_{i=1}^m \int_0^{u-} f_i(s) dN_i^n(s)\right)\\
&\qquad\qquad \times d\big(N_j^n(u) - H_j^n(u)\big)\\
&\qquad + \sum_{j=1}^m \int_0^t \big(\exp(-f_j(u)) - 1\big)\exp\left(-\sum_{i=1}^m \int_0^{u-} f_i(s) dN_i^n(s)\right) dH_j^n(u).
\end{aligned} \tag{A.1}$$

Using the martingale assumption and the measurability assumption, conditioning both sides of (A.1) on H^n, we have

$$\varphi_f^n(t) = E\left[\exp\left(-\sum_{i=1}^m \int_0^t f_i(s)\, dN_i^n(s)\right)\middle| H^n\right]$$
$$= 1 + \sum_{i=1}^m \int_0^t \varphi_f^n(u-)\big(\exp(-f_i(u)) - 1\big)\, dH_i^n(u),$$

and the convergence of H^n to H implies the convergence of H^n to H implies the convergence of φ_f^n to φ_f. [The convergence can be obtained by applying Theorem 5.4 of Kurtz and Protter (1991) or more directly from Avram (1988).]

Lemma A.2. *Let* $\xi_1, \dots, \xi_n$ *be exchangeable and suppose there exists a constant K such that* $|\xi_k| \le K$ *a.s. Define*

$$M_k = \frac{1}{k}\sum_{i=1}^k \xi_i.$$

Let $\varepsilon > 0$. *Then there exist* C *and* η *depending only on* K *and* ε, *such that, for* $l < n$,

$$P\{|M_n - M_l| \geq \varepsilon\} \leq C(\varepsilon, K)e^{-\eta(\varepsilon, K)l}.$$

(*In particular, the right-hand side does not depend on* n.)

PROOF. Note that $\{M_k\}$ is a reverse martingale and that

$$|M_{k+1} - M_k| \leq \frac{2K}{k+1}.$$

It follows that

$$\begin{aligned} &E[\exp(\lambda(M_n - M_l))] \\ &\quad = 1 + \sum_{k=l}^{n-1} E[\exp(\lambda(M_n - M_k)) - \exp(\lambda(M_n - M_{k+1})) \\ &\qquad\qquad - \lambda(M_{k+1} - M_k)\exp(\lambda(M_n - M_{k+1}))] \\ &\quad \leq 1 + \sum_{k=l}^{n-1} \left(\exp\left(\lambda\frac{2K}{k+1}\right) - 1 - \lambda\frac{2K}{k+1}\right) E[\exp(\lambda(M_n - M_{k+1}))] \end{aligned}$$

and, by Gronwall's inequality, that

$$\begin{aligned} E[\exp(\lambda(M_n - M_l))] &\leq \exp\left\{\sum_{k=l}^{n-1}\left(\exp\left(\lambda\frac{2K}{k+1}\right) - 1 - \lambda\frac{2K}{k+1}\right)\right\} \\ &\leq \exp\left\{\exp\left(\lambda\frac{2K}{l+1}\right)\frac{(\lambda 2K)^2}{l}\right\}. \end{aligned}$$

Hence,

$$\begin{aligned} P\{(M_n - M_l) \geq \varepsilon\} &\leq \exp\left\{\exp\left(\lambda\frac{2K}{l+1}\right)\frac{(\lambda 2K)^2}{l} - \lambda\varepsilon\right\} \\ &\leq \exp\left\{\left(e^{\delta}\delta^2 - \frac{\delta}{2K}\varepsilon\right)l\right\}, \end{aligned}$$

where we take $\lambda = \delta l/2K$. The same inequality holds with M_k replaced by $-M_k$. Consequently, we can take $\eta = -\inf_\delta(e^\delta\delta^2 - (\delta/2K)\varepsilon)$ and $C = 2$. □

LEMMA A.3. *For* $x \in D_E[0,\infty)$ *and* $\varepsilon > 0$, *define* $\tau_0^\varepsilon(x) = 0$ *and* $\tau_{k+1}^\varepsilon(x) = \inf\{t > \tau_k^\varepsilon(x): r(x(t), x(\tau_k^\varepsilon(x))) > \varepsilon\}$. *Let* $J(x, t, \varepsilon) = \min\{k: \tau_k^\varepsilon(x) > t\}$. *Then* $J(x, t, \varepsilon)$ *is bounded on compact subsets of* $D_E[0,\infty)$.

PROOF. The lemma follows easily from the characterization of the compact subsets of $D_E[0,\infty)$ in terms of a modulus of continuity. See, for example, Ethier and Kurtz (1986), Theorem 3.6.3. □

REFERENCES

AVRAM, F. (1988). Weak convergence of the variations, iterated integrals, and Doléans–Dade exponentials of sequences of semimartingales. *Ann. Probab.* **16** 246–250.

BHATT, A. G. and KARANDIKAR, R. L. (1993). Invariant measures and evolution equations for Markov processes characterized via martingale problems. *Ann. Probab.* **21** 2246–2268.

BLACKWELL, D. and DUBINS, L. E. (1983). An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* **89** 691–692.

DAWSON, D. A. (1993). Measure-valued Markov processes. *Ecole d'Eté de Probabilités de Saint-Flour XXI. Lecture Notes in Math.* **1541**. Springer, Berlin.

DAWSON, D. A. and PERKINS, E. A. (1991). Historical processes. *Mem. Amer. Math. Soc.* **93** (454).

DONNELLY, P. and KURTZ, T. G. (1996). A countable representation of the Fleming–Viot measure-valued diffusion. *Ann. Probab.* **24** 698–742.

DYNKIN, E. B. (1965). *Markov Processes I*. Springer, Berlin.

EL KAROUI, N. and ROELLY, S. (1991). Proprietes de martingales, explosion et representation de Lévy–Khinchine d'une classe du processus de branchement à valeurs mesures. *Stochastic Process. Appl.* **38** 239–266.

ETHERIDGE, A. and MARCH, P. (1991). A note on superprocesses. *Probab. Theory Related Fields* **89** 141–147.

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes: Characterization and Convergence*. Wiley, New York.

ETHIER, S. N. and KURTZ, T. G. (1993). Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* **31** 345–386.

EVANS, S. N. (1993). Two representations of a conditioned superprocess. *Proc. Roy. Soc. Edinburgh Sect. A* **123** 959–971.

EVANS, S. N. and PERKINS, E. (1990). Measure-valued Markov branching processes conditioned on non-extinction. *Israel J. Math.* **71** 329–337.

KINGMAN, J. F. C. (1982). The coalescent. *Stochastic Process. Appl.* **13** 235–248.

KURTZ, T. G. (1998). Martingale problems for conditional distributions of Markov processes. *Elec. J. Probab.* **3**.

KURTZ, T. G. and PROTTER, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Ann. Probab.* **19** 1035–1070.

LOEVE, M. (1963). *Probability Theory*, 3rd ed. Van Nostrand, Princeton.

PERKINS, E. A. (1991). Conditional Dawson–Watanabe processes and Fleming–Viot processes. In *Seminar on Stochastic Processes* 142–155. Birkhäuser, Boston.

PERKINS, E. A. (1992). Measure-valued branching diffusion with spatial interactions. *Probab. Theory Related Fields* **94** 189–245.

PERKINS, E. A. (1995). On the martingale problem for interactive measure-valued branching diffusions. *Mem. Amer. Math. Soc.* **115** 1–89.

PITMAN, J. (1997). Coalescents with multiple collisions. Preprint.

TRIBE, R. (1992). The behavior of superprocesses near extinction. *Ann. Probab.* **20** 286–311.

DEPARTMENT OF STATISTICS
UNIVERSITY OF OXFORD
1 SOUTH PARKS ROAD
OXFORD OX1 3TG
UNITED KINGDOM
E-MAIL: donnelly@stats.ox.ac.uk

DEPARTMENTS OF MATHEMATICS
AND STATISTICS
UNIVERSITY OF WISCONSIN—MADISON
480 LINCOLN DRIVE
MADISON, WISCONSIN 53706-1388
E-MAIL: kurtz@math.wisc.edu