The Annals of Probability
1992, Vol. 20, No. 1, 322–341

WEAK CONVERGENCE OF POPULATION GENEALOGICAL PROCESSES TO THE COALESCENT WITH AGES

BY PETER DONNELLY[1] AND PAUL JOYCE

Queen Mary and Westfield College and University of Southern California, Los Angeles

The population genealogical processes associated with a wide range of exchangeable reproductive models (including the Wright–Fisher model) are shown to converge weakly, as the population size becomes large, to a particularly tractable limiting process, the age-ordered analog of Kingman's coalescent. This result extends the known convergence results for sample processes and effectively completes the robustness theory for neutral genealogies. Its consequences, which include a unification of the results for neutral models, have already been exploited elsewhere. The techniques used rely heavily on knowledge of sample behavior, together with consistency arguments. They may be of more general interest.

1. Introduction.

The n-coalescent was introduced by Kingman (1982a, c) as a robust method of describing the genealogy of a sample of fixed size, n, taken from a large haploid population evolving according to one of a large class of "exchangeable" reproductive models. Watterson (1984) extended the process to the infinitely many alleles setting, by explicitly allowing for mutation, and Donnelly and Tavaré (1986) observed that it is possible to keep track of the order in which the new classes arise and as a consequence to study the age structure of the alleles.

While many questions of genetic interest may be answered from a knowledge of the genealogy of samples of fixed but arbitrary size, some require information about the genealogy of the whole population. The natural processes of interest are the coalescent [Kingman (1982b)] and for models with mutation, its close relative, the coalescent with ages [Donnelly and Tavaré (1987)]. These processes possess a rich structure which naturally lends itself to calculations of interest (in fact, in the infinite alleles case, explicit attention to age ordering actually leads to a considerable simplification) and present themselves as natural candidates for the limits of population genealogical processes. In spite of this, these processes have retained a somewhat ambivalent status. The convergence results for sample genealogies rely heavily on the fact that the sample size remains fixed as the population size increases. Until recently, questions concerning convergence of the population processes associated with discrete models have remained open. Donnelly (1991) verified a conjecture in Kingman (1982c) in proving weak convergence of so-called population ances-

---

Received April 1989; revised April 1990.

[1] Supported in part by NSF Grant DMS-86-08857.

AMS 1980 *subject classifications*. Primary 60F99; secondary 60J75, 92A10.

Key words and phrases. Coupling, sampling functions, neutral models, exchangeability, genealogy.

tral processes, the processes which count numbers of ancestors, and provide the natural "time scale" for the more informative genealogical processes. Donnelly and Joyce (1991) prove weak convergence of the equilibrium distributions of the population processes. Our purpose here is to complete this part of the theory by proving that the processes themselves converge weakly (as random elements of the appropriate function space) to the coalescent with ages, thus verifying conjectures implicit in a number of earlier works [e.g., Kingman (1982b), Tavaré (1984) and Donnelly and Tavaré (1987)].

Our second aim is to exhibit an approach to certain weak convergence problems which may be more widely applicable. The raw materials available for proving weak convergence, and the nature of the limiting process, differ from those in many standard settings. The basic tool is knowledge that samples from these processes behave in the correct way and the challenge is to exploit this and the internal "consistency" of the processes to force the population genealogies to behave correctly. Heavy use is made of monotonicity properties and two different couplings between population and sample processes. One bonus of the approach is that while the results apply to a wide range of models, no further model-specific calculations are necessary. The techniques may be useful in studying certain additive interactive particle systems, where processes with different starting configurations may be naturally coupled. For example, they apply in the context studied in Cox and Griffeath (1990), to provide an alternative to the approach adopted there.

We conclude this section with some notation and point out that an outline of the general strategy of the paper is given at the end of the next section. Denote the infinite-dimensional unit simplex by Δ:

$$\Delta = \left\{ x = (x_1, x_2, \dots) : x_i \geq 0,\ \sum_{i=1}^{\infty} x_i = 1 \right\}$$

and let

$$\overline{\Delta} = \left\{ x = (x_1, x_2, \dots) : x_i \geq 0,\ \sum_{i=1}^{\infty} x_i \leq 1 \right\}.$$

We regard Δ and $\overline{\Delta}$ as subspaces of the (metrizable) space $[0,1]^{\mathbb{N}}$ endowed with the product topology. Note that with this topology $\overline{\Delta}$ is closed and hence compact. Throughout we give $\overline{\Delta}$ the Borel σ-algebra. We denote $\mathscr{P}(\overline{\Delta})$ as the set of probability measures on $\overline{\Delta}$. We use the symbol $\Rightarrow$ to denote weak convergence, and for $x \in \mathbb{R}$, $[x]$ will denote the largest integer less than or equal to x. For a set A, $|A|$ will denote the cardinality of A and $I\{A\}$ the indicator function of A.

2. Exchangeable reproductive models and the genealogical processes.

We shall consider a class of neutral models with nonoverlapping generations introduced by Cannings (1974). Randomly label the N individuals in a particular generation, and let ν_i be the number of offspring born to the ith individual, $i = 1, 2, \dots, N$. The random variables $\{\nu_i\}$ are exchangeable, and constant population size is maintained by requiring $\nu_1 + \nu_2 + \cdots + \nu_N = N$. We shall also assume that the distribution of $\{\nu_1, \dots, \nu_N\}$ is invariant over time, and the assumption of neutrality implies that the ν_i are independent of family sizes in preceding generations. Each offspring individual may be subject to mutation at birth. Mutations occur independently for each individual (and independently of the past history of the process) with probability u, and result in a novel allelic type.

Consider the population at some time labeled 0 and ask about its composition with respect to the ancestral population m generations into the past. Divide the population into two types of equivalence classes which we label old and new. We say individuals i and j are in the same old class at time m, if i and j have the same ancestor m generations into the past with no intervening mutation along their lines of descent. We say that i and j are in the same new class at time m if for some r satisfying $0 \le r < m$, individuals i and j have the same ancestor r generations into the past, this ancestor itself being a mutant, with no mutations between 0 and r.

Denote the number of old classes at time m by $A_N(m)$ and write $X_{N,1}(m), X_{N,2}(m), \dots, X_{N, A_N(m)}(m)$ for the proportion of the population in each of these classes. Similarly, let $F_N(m)$ denote the number of new classes at time m, and write $Y_{N,1}(m), Y_{N,2}(m), \dots, Y_{N, F_N(m)}(m)$ for the proportion of the population in the oldest, second oldest, ..., youngest new class. (The age of a class is the time from the mutation event which gives rise to the class until the present, so that older alleles arise further in the past. Break ties in ages at random.) In some settings it is convenient to describe genealogies (and genealogical processes) in terms of equivalence classes, while in others one keeps track only of the frequencies in each of the classes. Exchangeability arguments [see, e.g., Kingman (1982b)] mean that the two approaches are equivalent. Define the process $\{X_N(t),\ t \ge 0\}$ with state space $\overline{\Delta}$, by

$$X_N(0) = (1/N, 1/N, \dots, 1/N, 0, 0, \dots),$$

$$X_N(t) = \big(X_{N,1}(t), \dots, X_{N, A_N(t)}(t), Y_{N,1}(t), \dots, Y_{N, F_N(t)}(t), 0, 0, \dots\big), \quad t > 0.$$

We write the frequencies of the old classes in an arbitrary order. If at time t, old classes of sizes $x_{i_1}, x_{i_2}, \dots, x_{i_l}$, say, coalesce, $i_1 < i_2 < \cdots < i_l$, where $X_N(t-) = (x_1, \dots, x_L, y_1, \dots, y_F, 0, 0, \dots)$, then $X_N(t)$ will take one of the values

$$\big(x_1, \dots, x_{i_1 - 1}, x_{i_1 + 1}, \dots, x_{i_2 - 1}, x_{i_2 + 1}, \dots, x_{i_k - 1}, x_{i_1} + \cdots + x_{i_l},$$
$$x_{i_k + 1}, \dots, x_{i_l - 1}, x_{i_l + 1}, \dots, x_L, y_1, \dots, y_F, 0, 0, \dots\big),$$

$k = 1, 2, \dots, l$, each with probability l^{-1}, this choice being independent of the past history of the process.

The limiting population process $\{(D(t), X(t)),\ t \ge 0\}$ is described in Donnelly and Tavaré (1987). It is Markov with state space

$$E \equiv \{0, 1, 2, \dots, \infty\} \times \overline{\Delta},$$

where $D(t)$ is the Markov death process started from an entrance boundary at ∞ with death rates $k(k+\theta-1)/2$ from state k. Also, $X(t)$ may be represented as $\mathscr{M}_{D(t)}$, where the (discrete-time, Markov) jump chain $\{\mathscr{M}_k;\ k = 0,1,2,\dots\}$ is independent of $\{D(t),\ t \geq 0\}$ and has transition probabilities

$$P\big[\mathscr{M}_{k-1} = (x_1,\dots,x_{i-1},x_{i+1},\dots,x_k,x_i,y_1,y_2,\dots)\ \big|\ \mathscr{M}_k = (x_1,\dots,x_k,y_1,y_2,\dots)\big] = \frac{\theta}{k(k+\theta-1)},$$

and for $1 \leq i < j \leq k$,

$$P\big[\mathscr{M}_{k-1} = (x_1,\dots,x_{i-1},x_i+x_j,x_{i+1},\dots,x_{j-1},x_{j+1},\dots,x_k,y_1,y_2,\dots)\ \big|\ \mathscr{M}_k = (x_1,\dots,x_k,y_1,y_2,\dots)\big] = \frac{1}{k(k+\theta-1)},$$

$$P\big[\mathscr{M}_{k-1} = (x_1,\dots,x_{i-1},x_{i+1},\dots,x_{j-1},x_i+x_j,x_{j+1},\dots,x_k,y_1,y_2,\dots)\ \big|\ \mathscr{M}_k = (x_1,\dots,x_k,y_1,y_2,\dots)\big] = \frac{1}{k(k+\theta-1)}.$$

The distribution of $D(t)$ is known; see, for example, Tavaré (1984). It suffices here to observe that as $t \to 0$, $D(t) \to \infty$ a.s., but that for $t > 0$, $E(D(t)) < \infty$.

The marginal distribution of the jump chain is also known. In fact for $k = 1,2,\dots,$ $\mathscr{M}_k$ has the same distribution as the random element of Δ given by

$$(1)\qquad (U_1V_k, U_2V_k,\dots,U_kV_k,(1-V_k)Z_1,(1-V_k)(1-Z_1)Z_2,\ (1-V_k)(1-Z_1)(1-Z_2)Z_3,\dots),$$

where $Z_1, Z_2,\dots,$ are independent and identically distributed random variables with density $\theta(1-x)^{\theta-1}$, $0 \leq x \leq 1$, $(U_1,U_2,\dots,U_k)$ are jointly distributed uniformly on the simplex $\{(x_1,\dots,x_k)\colon x_i \geq 0,\ x_1 + \cdots + x_k = 1\}$, and V_k has density

$$(2)\qquad f_k(x) = \frac{\Gamma(k+\theta)x^{k-1}(1-x)^{\theta-1}}{\Gamma(\theta)(k-1)!},\qquad 0 \leq x \leq 1,$$

with the collections, $Z_1, Z_2,\dots,(U_1,\dots,U_k)$, and V_k mutually independent.

In the sequel we will often write $M(\cdot)$ for $(D(\cdot),X(\cdot))$ and $M_N(\cdot)$ for $(A_N([N\sigma^{-2}\cdot]), X_N([N\sigma^{-2}\cdot]))$. Further let $\tilde{X}(\cdot)$ and $\tilde{X}_N(\cdot)$ be the processes which just keep track of old class sizes: If $M(t)$, respectively $M_N(t)$, takes the value $(k,(x_1,\dots,x_k,y_1,\dots))$ then $\tilde{X}(t)$, respectively $\tilde{X}_N(t)$, takes the value $(x_1,\dots,x_k,0,0,\dots)$.

Sample genealogical processes are defined analogously. Take a sample of n individuals at time 0 and consider *its* genealogy with respect to the ancestral population m generations into the past. Denote by $A_N^n(m)$ and $F_N^n(m)$ the number of old and new classes respectively and by $X_N^n(\cdot)$ the process analogous to $X_N(\cdot)$ which describes the evolution of the proportion of the sample in each of these classes.

The limiting genealogical process for samples $\{(D^n(t), X^n(t)), t \geq 0\}$ is the "frequency" analog of the n-coalescent with ages of Donnelly and Tavaré (1986). It is Markov with state space $E_n \equiv \{0, 1, \dots, n\} \times \bar{\Delta}$. Again, $\{D^n(t), t \geq 0\}$ is a (Markov) death process with death rate $k(k + \theta - 1)/2$ from state k; this time $D^n(0) = n$. Also [Donnelly and Tavaré (1986)] $(D^n(\cdot), X^n(\cdot))$ may be represented as $(D^n(\cdot), \mathscr{M}^n_{D^n(\cdot)})$ with the jump chain $\{\mathscr{M}^n_k; k = n, n-1, \dots, 0\}$ Markov and independent of the death process $D^n(\cdot)$. Its marginal distributions are given by

$$
\begin{aligned}
(3)\qquad & P\left[\mathscr{M}^n_k = \left(\frac{\lambda_1}{n}, \dots, \frac{\lambda_k}{n}, \frac{\eta_1}{n}, \dots, \frac{\eta_l}{n}, 0, 0, \dots\right)\right] \\
&= \frac{(n-k)!\theta^l}{(k+\theta)_{(n-k)}} \frac{1}{\eta_l(\eta_l + \eta_{l-1}) \cdots (\eta_l + \cdots + \eta_1)},
\end{aligned}
$$

provided $l \leq n$ and the λ_i, $i = 1, 2, \dots, k$, and η_j, $j = 1, 2, \dots, l$, are positive integers with $\lambda_1 + \cdots + \lambda_k + \eta_1 + \cdots + \eta_l = n$. Here and throughout, $x_{(n)} \equiv x(x+1) \cdots (x+n-1)$ denotes the ascending factorial of x.

For the sake of completeness we now restate the known convergence results for these processes. Let ρ denote the ranking function on $\bar{\Delta}$. That is, if

$$
\bar{\nabla} = \left\{x = (x_1, x_2, \dots,): x_i \geq 0, x_1 \geq x_2 \geq \cdots, \sum_{i=1}^{\infty} x_i \leq 1\right\},
$$

then $\rho\colon \bar{\Delta} \to \bar{\nabla}$, and $\rho(x)$ is the rearrangement of the components of x into nonincreasing order.

Theorem 1 [Kingman (1982c) and Donnelly and Tavaré (1986)]. *Assume that as* $N \to \infty$,

1. $\mathrm{Var}(\nu_1) \to \sigma^2 > 0$,
2. $\sup_N E[\nu_1^k] < \infty$, $k = 1, 2, \dots,$
3. $u = \theta\sigma^2/(2N) + o(N^{-1})$ *for some* $\theta > 0$.

Then the process $\{(A^n_N([N\sigma^{-2}t]), X^n_N([N\sigma^{-2}t])), t \geq 0\}$ *converges weakly in* $D_{E_n}[0, \infty)$ *to* $\{(D^n(t), X^n(t)), t \geq 0\}$ *as* $N \to \infty$.

Theorem 2 [Donnelly (1991)]. *Under the assumptions of Theorem* 1, *as* $N \to \infty$,

$$
A_N([N\sigma^{-2} \cdot]) \Rightarrow D(\cdot)
$$

as elements of $D_{\mathbb{N} \cup \{\infty\}}[0, \infty)$.

Theorem 3 [Donnelly and Joyce (1991)]. *Under the assumptions of Theorem* 1, *as* $N \to \infty$,

$$
\rho(\tilde{X}_N([N\sigma^{-2}t])) \Rightarrow \rho(\tilde{X}(t))
$$

as random elements of $\bar{\nabla}$, *for each fixed* $t > 0$.

The main result of this paper is the following theorem.

THEOREM 4. *Under the assumptions of Theorem* 1, *as* $N \to \infty$,

$$M_N([N\sigma^{-2}\cdot]) \equiv (A_N([N\sigma^{-2}\cdot]), X_N([N\sigma^{-2}\cdot])) \Rightarrow (D(\cdot), X(\cdot)) \equiv M(\cdot),$$

as elements of $D_E[0, \infty)$.

Now some comments on strategy. We will first prove weak convergence of the one-dimensional marginal distributions of the process, then extend this to prove convergence of finite-dimensional distributions and finally establish tightness. Theorem 2 gives convergence of marginal distributions of the first component; it then turns out to be sufficient to establish convergence of the conditional distribution of $X_N([N\sigma^{-2}t])$ given $A_N([N\sigma^{-2}t]) = k$ to that of $X(t)$ given $D(t) = k$ (which is exactly the distribution of $\mathscr{M}_k$) for each k. The key here is to observe that (by definition) one way of obtaining the distribution of $X_N^n([N\sigma^{-2}t])$ is by "taking a sample" from $X_N([N\sigma^{-2}t])$. Theorem 1 (and the structure of sample and population limiting processes) ensures that these sampling distributions converge. That is, for certain f_N, f,

$$E(f_N(X_N([N\sigma^{-2}t]))) \to E(f(X(t))).$$

The problems with extending this to give weak convergence are effectively that none of these f's is continuous on $\overline{\Delta}$ and that our concern in any case is with conditional distributions. The first was overcome in an analogous situation in Donnelly and Joyce (1991); it is necessary to show that any weak limits of the sequence $\{X_N([N\sigma^{-2}t])\}$ live on especially nice subsets of Δ. These ideas can be pushed through in the conditional case. Section 3 contains the ground work (properties of limit points of sequences of conditional distributions) and Section 4 gives the required conditional, and hence marginal, convergence.

The rest is a little easier. For any $t > 0$, $X_N([N\sigma^{-2}t])$ and $X(t)$ will have a finite number, say k, of old classes. (That both processes have the same number of old classes, at least for large N, is a consequence of Theorem 2.) The behavior of the processes as t increases effectively only depends on the genealogy of these k ancestors, but that is described by a process with the same distribution as $(A_N^k(\cdot), X_N^k(\cdot))$. Now apply Theorem 1 again and check the behavior of the limiting process to observe that the conditional distribution of $X_N([N\sigma^{-2}t])$ from t onwards converges to that of $X(t)$. Judicious application of this observation (in fact it is a coupling) and the Markov property converts convergence of marginal distributions into convergence of finite-dimensional distributions.

For tightness, we wish to show that the prelimiting processes do not jump too far in small time intervals. Remember that for all large enough N, $(A_N([N\sigma^{-2}t]), X_N([N\sigma^{-2}t]))$ is close in distribution to $(D(t), X(t))$. First choose $t_1 > 0$ very small so that this latter process has not moved far from its starting position of $(\infty, (0, 0, \dots))$ and use monotonicity properties to bound the jumps of the prelimiting processes over $[0, t_1]$. Now from t_1 on the process makes only a finite number of jumps. So it is straightforward to show that the process on $[t_1, \infty)$ has sample paths whose jump sizes are at least ε with high probability.

Before continuing we note that convergence of age-ordered allele frequencies at equilibrium to the so-called GEM distribution follows easily from Theorem 4. This was proved directly in Donnelly and Joyce (1991).

COROLLARY 5. *Let* $\rho_N = \inf\{t \geq 0\colon A_N([N\sigma^{-2}t]) = 0\}$, $\rho = \inf\{t \geq 0\colon D(t) = 0\}$. *Then as* $N \to \infty$, $X_N([N\sigma^{-2}\rho_N]) \Rightarrow X(\rho) =_{\mathscr{D}} (Z_1, (1 - Z_1)Z_2, (1 - Z_1)(1 - Z_2)Z_3, \dots)$, *where, as before,* $Z_1, Z_2, \dots$ *are i.i.d. with common density* $\theta(1 - x)^{\theta - 1}$, $0 \leq x \leq 1$.

PROOF. Define $\xi\colon D_{\bar{\Delta}}[0, \infty) \times [0, \infty) \to \mathbb{R}$ by $\xi(x, t) = x(t)$. Since $X(\cdot)$ has no fixed jumps on $[0, \infty)$ and $P(\rho \in (0, \infty)) = 1$, the function ξ is a.s. continuous with respect to the distribution of $X(\cdot)$. The result will follow if we show that $(\rho_N, X_N([N\sigma^{-2} \cdot])) \Rightarrow (\rho, X(\cdot))$ as $N \to \infty$.

Let $E_1 = \{0, 1, 2, \dots, \infty\}$. Define $\hat{\pi}\colon D_{E_1 \times \bar{\Delta}}[0, \infty) \to D_{E_1}[0, \infty) \times D_{\bar{\Delta}}[0, \infty)$ in the canonical way. Noting that the projection maps from $E_1 \times \bar{\Delta} \to E_1$ and $E_1 \times \bar{\Delta} \to \bar{\Delta}$ are continuous, it is a straightforward exercise to show that $\hat{\pi}$ is continuous. Let $\pi_t\colon D_{E_1}[0, \infty) \to E_1$ be the evaluation map, and note that since $D(\cdot)$ has no fixed jumps on $(0, \infty)$, π_t is continuous $D(\cdot)$ a.s. for each fixed $t > 0$. Let B be a closed subset of $D_{\bar{\Delta}}[0, \infty)$. The Portmanteau and continuity theorem guarantee that

$$
\begin{aligned}
&\limsup_{N \to \infty} P(\rho_N \leq t,\ X_N \in B) \\
&\quad = \limsup_{N \to \infty} P\big(A_N([N\sigma^{-2}t]) = 0,\ X_N([N\sigma^{-2} \cdot]) \in B\big) \\
&\quad = \limsup_{N \to \infty} P\big(\hat{\pi}(M_N) \in \pi_t^{-1}(0) \times B\big) \\
&\quad \leq \limsup_{N \to \infty} P\big(\hat{\pi}(M) \in \pi_t^{-1}(0) \times B\big) \\
&\quad = P(\rho \leq t, X(\cdot) \in B). \qquad \square
\end{aligned}
$$

3. Properties of the limiting processes. In this section we collect together some results concerning the limiting processes.

THEOREM 6. $(D^n(t), X^n(t)) \Rightarrow (D(t), X(t))$ *for each* $t > 0$ *as* $n \to \infty$.

PROOF. It is straightforward to show that $D^n(t) \Rightarrow D(t)$ as $n \to \infty$. In view of the independence of $D^n(\cdot)$ and the jump chain $\mathscr{M}^n$ of the coalescent, it only remains to show that for each $k = 0, 1, \dots, n$, $\mathscr{M}_k^n \Rightarrow \mathscr{M}_k$ as $n \to \infty$.

Write the components of $\mathscr{M}_k$ as $(R_1, \dots, R_k, S_1, S_2, \dots)$ and those of $\mathscr{M}_k^n$ as $(R_1^n \dots, R_1^k, S_1^n, S_2^n, \dots)$. Denote by $F^n = \max\{l\colon S_l^n > 0\}$, the number of new classes in $\mathscr{M}_k^n$ and note [Donnelly and Tavaré (1986)] that $F^n \to \infty$ a.s. as

$n \to \infty$. It is sufficient to prove that for each fixed r, $r = 1, 2, \dots$, as $n \to \infty$,

$$(R_1^n, \dots, R_k^n, S_1^n, \dots, S_r^n) \Rightarrow (R_1, \dots, R_k, S_1, \dots, S_r),$$

which in turn will follow from Scheffé's theorem [see Serfling (1980), page 17] if

$$\begin{aligned}\lim_{n \to \infty} n^{r+k} P\big(nR_i^n = [ny_i], i = 1, 2, \dots, k, nS_j^n = [nx_j],& \\ j = 1, 2, \dots, r; F^n \geq r\big)& \\ = f(y_1, \dots, y_k, x_1, \dots, x_r),&\end{aligned}$$

where f is the joint density function of $(R_1, \dots, R_k, S_1, \dots, S_r)$. This last is just a calculation involving (1), (2) and (3) which is similar to that on page 15 of Donnelly and Tavaré (1986) and we omit the details. □

Now, and throughout the remainder of this section, fix $t > 0$. Denote by μ_N^k the distribution of $X_N([N\sigma^{-2}t])$ given that $A_N([N\sigma^{-2}t]) = k$. (It is possible that μ_N^k depends on t, but for convenience we will not make this dependence explicit.) Our aim is to show that if ν^k is a limit point of $\{\mu_N^k\}_{N=1}^\infty$ then ν^k is concentrated on the set $\{x \in \Delta: x_i > 0 \text{ for each } i\}$. We do this by exploiting two key properties of the discrete genealogical processes:

1. Monotonicity property: As the process progresses, the surviving old classes will either increase in size or become new.
2. Conditional independence property: Conditional on there being l old classes at time t, the ancestry from the time t onwards has the same structure as the process $(A_N^l(\cdot), X_N^l(\cdot))$ concerned with genealogy of a sample of size l, starting at time 0.

For convenience we will denote $A_N([N\sigma^{-2} \cdot])$ by $A_N(\cdot)$ and $X_N([N\sigma^{-2} \cdot])$ by $X_N(\cdot)$.

THEOREM 7. *If ν^k is a limit point of the sequence $\{\mu_N^k\}_{N=1}^\infty$, then $\nu^k(\Delta) = 1$.*

PROOF. First note that it is sufficient to show that given ε, $\delta > 0$, $\exists$ m with

$$\nu^k\left(\left\{\sum_{i=1}^m x_i \geq 1 - \delta\right\}\right) \geq 1 - \varepsilon.$$

This in turn will follow from the Portmanteau theorem provided

$$\limsup_{N \to \infty} P\left[\sum_{i=1}^k X_{N,i}(t) + \sum_{i=1}^{m-k} Y_{N,i}(t) \geq 1 - \delta \,\middle|\, A_N(t) = k\right] \geq 1 - \varepsilon.$$

Observe that conditional on $A_N(s) = j$, the monotonicity property guarantees that for $t \geq s$ the sum of the first j components of $X_N(t)$ is a nondecreasing function of t. The idea behind the proof is effectively to choose a small time s with the property that the sum of the old class frequencies at time s is large and then choose an m larger than the number of old classes at that time. The observation above then ensures that the sum of the first m components of $X_N(t)$ will also be large.

Now, for any $R < L \in \mathbb{N}$, provided $m \geq L$,

$$P\left[\sum_{i=1}^{k} X_{N,i}(t) + \sum_{i=1}^{m-k} Y_{N,i}(t) \geq 1 - \delta \,\middle|\, A_N(t) = k\right]$$

$$\begin{aligned} &= \sum_{j=k}^{\infty} P\left[\sum_{i=1}^{k} X_{N,i}(t) + \sum_{i=1}^{m-k} Y_{N,i}(t) \geq 1 - \delta \,\middle|\, A_N(s) = j, A_N(t) = k\right] \\ &\qquad \times P[A_N(s) = j | A_n(t) = k] \\ &\geq \sum_{j=R}^{L} P\left[\sum_{i=1}^{j} X_{N,i}(s) \geq 1 - \delta \,\middle|\, A_N(s) = j\right] P[A_N(s) = j | A_N(t) = k]. \end{aligned} \tag{4}$$

Rewrite the components $X_{N,1}(s), X_{N,2}(s), \dots, X_{N,j}(s)$, in nonincreasing order as $X_{N,(1)}(s), X_{N,(2)}(s), \dots, X_{N,(j)}(s)$, and note that on $\{A_N(s) = j\}$,

$$\sum_{i=1}^{j} X_{N,i}(s) = \sum_{i=1}^{j} X_{N,(i)}(s). \tag{5}$$

Invoke Theorem 3 and observe that for $l \geq j$, the boundary of event $\{\sum_{i=1}^{j} x_i \geq 1 - \delta, x_l > 0\}$ is contained in the set $\{\sum_{i=1}^{j} x_i = 1 - \delta\}$, which has probability 0 (recall $\delta > 0$) under the distribution of the limit process. Thus

$$\begin{aligned} &P\left[\sum_{i=1}^{j} X_{N,(i)}(s) \geq 1 - \delta, A_N(s) > l\right] \\ &\qquad = P\left[\sum_{i=1}^{j} X_{N,(i)}(s) \geq 1 - \delta, X_{N,l}(s) > 0\right] \\ &\qquad \to P\left[\sum_{i=1}^{j} X_{(i)}(s) \geq 1 - \delta, X_{(l)}(s) > 0\right] \\ &\qquad = P\left[\sum_{i=1}^{j} X_{(i)}(s) \geq 1 - \delta, D(s) > l\right], \end{aligned} \tag{6}$$

as $N \to \infty$.

It follows from (4), (5) and (6) and Theorem 2, that with V_l defined at (2),

$$\begin{aligned} &\limsup_{N \to \infty} P\left[\sum_{i=1}^{k} X_{N,i}(t) + \sum_{i=1}^{m-k} Y_{N,i}(t) \geq 1 - \delta \,\middle|\, A_N(t) = k\right] \\ &\qquad \geq \sum_{j=R}^{L} P[V_j \geq 1 - \delta] P[D(s) = j | D(t) = k] \\ &\qquad \geq P[V_R \geq 1 - \delta] P[R \leq D(s) \leq L | D(t) = k], \end{aligned} \tag{7}$$

which can be made arbitrary close to 1, as required, by first choosing L large, then choosing s small, and finally choosing R large. □

THEOREM 8. *If* ν^k *is a limit point of* $\{\mu_N^k\}$, *then* $\nu^k\{x \in \Delta\colon x_i > 0$ *for all* $i\} = 1$.

PROOF. Recall that $t > 0$ is fixed. Let $Q_N^{l,k}(t)$, respectively $Q^{l,k}(t)$, denote the number of new classes formed in the process $(A_N^l(\cdot), X_N^l(\cdot))$, respectively $(D^l(\cdot), X^l(\cdot))$, between time 0 and time t, conditional on $A_N^l(t) = k$, respectively $D^l(t) = k$. Further let $C_N(s,t)$ be the number of new classes formed in $(A_N^l(\cdot), X_N^l(\cdot))$ from time s to time t.

Now, for any $r, m \in \mathbb{N}$ with $r < m$, and $s < t$,

$$\begin{aligned}\mu_N^k(x_{k+i} > \delta) &\ge P[Y_{N,i}(t) \ge \delta, C_N(s,t) \ge i \mid A_N(t) = k]\\ &\ge \sum_{l=r}^{m} P\big[X_{N,(l)}(s) \ge \delta, C_N(s,t) \ge i \mid A_N(t) = k, A_N(s) = l\big]\\ &\qquad \times P[A_N(s) = l \mid A_N(t) = k]\\ &= \sum_{l=r}^{m} P\big[X_{N,(l)}(s) \ge \delta \mid A_N(s) = l\big] P\big[Q_N^l(t-s) \ge i\big]\\ &\qquad \times P[A_N(s) = l \mid A_N(t) = k],\end{aligned} \tag{8}$$

since by the conditional independence property, $X_N(s)$ is conditionally independent of $C_N(s,t)$ and $A_N(t)$, given $A_N(s)$, and given that $A_N(s) = l$, $A_N(t) = k$, $C_N(s,t)$ has the same distribution as $Q_N^{l,k}(t,s)$. Now as $N \to \infty$, Theorems 1, 2 and 3, and an argument similar to the one used in the proof of Theorem 6 guarantee that the right-hand side of (8) converges to

$$\sum_{l=r}^{m} P\big[X_{(l)}(s) \ge \delta \mid D(s) = l\big] P\big[Q^{l,k}(t-s) \ge i\big] P[D(s) = l \mid D(t) = k].$$

With $U_1, \dots, U_l, V_l$ defined at (1), the Portmanteau theorem gives

$$\begin{aligned}\nu^k(x_{k+i} > 0) &= \lim_{\delta \downarrow 0} \nu^k(x_{k+i} \ge \delta)\\ &\ge \lim_{\delta \downarrow 0} \sum_{l=r}^{m} P[U_1V_l \ge \delta, U_2V_l \ge \delta, \dots, U_lV_l \ge \delta]\\ &\qquad \times P\big[Q^{l,k}(t-s) \ge i\big] P[D(s) = l \mid D(t) = k]\\ &= \sum_{l=r}^{m} P\big[Q^{l,k}(t-s) \ge i\big] P[D(s) = l \mid D(t) = k]\\ &= P\big[Q^{r,k}(t-s) \ge i\big] P[r \le D(s) \le m \mid D(t) = k].\end{aligned}$$

Now let $m \to \infty$ and then $s \downarrow 0$ to obtain

$$\nu^k(x_{k+i} > 0) \ge P\big[Q^{r,k}(t) \ge i\big] \to 1$$

as $r \to \infty$ in view of the structure of the chain $\mathscr{M}$. From Theorems 2 and 3, it also follows that $\nu^k(x_i > 0) = 1$ for $i = 1, 2, \dots, k$, so that the desired result follows. □

4. Convergence of one-dimensional distributions. Let $B_n^k = \{(l; \eta_1, \dots, \eta_l)\colon l = k, k+1, \dots, \eta_i \in \{1, 2, \dots\}, \eta_1 + \cdots + \eta_l = n\}$. Think of a point $(l; \eta_1, \dots, \eta_l) \in B_n^k$ as describing a sample of size n in which the individuals form l classes, k of which are old and $(l - k)$ of which are new. The number of individuals in the old classes are $\eta_1, \dots, \eta_k$, and the numbers in the oldest, second oldest, ..., youngest new classes are $\eta_{k+1}, \dots, \eta_l$, respectively. If the sample is taken (with replacement) from an infinite population with description $(j, x) \equiv (j, (x_1, x_2, \dots))$ with the interpretation that the population has j old classes with frequencies $x_1, x_2, \dots, x_j$, and new classes with frequencies x_{j+1} for the oldest, x_{j+2} for the second oldest, ..., and so on, then we denote the probability that the sample will have configuration $\eta \in B_n^k$ by $\phi_\eta^{k,j}(x)$. Specifically, for $k \le j$, $\eta \in B_n^k$, define the sampling function $\phi_\eta^{k,j}\colon \overline{\Delta} \to R$ by

$$\phi_\eta^{k,j}(x) = \sum_{i_1 < i_2 < \cdots < i_k \le j < i_{k+1} < \cdots < i_l} \frac{n!}{\eta_1! \cdots \eta_l!} x_{i_1}^{\eta_1} x_{i_2}^{\eta_2} \cdots x_{i_l}^{\eta_l}.$$

We regard a point $(l; \eta_1, \dots, \eta_l) \in B_n^k$ as describing a sample with old class sizes $\eta_1, \eta_2, \dots, \eta_k$ written in a fixed but arbitrary order and distinguish between $(l; \eta_1, \dots, \eta_k, \eta_{k+1}, \dots, \eta_l)$ and $(l; \eta_{i_1}, \dots, \eta_{i_k}, \eta_{k+1}, \dots, \eta_l)$ in B_n^k if $i_1, \dots, i_k$ is a (nonidentical) permutation of $1, 2, \dots, k$. In the sampling context, and this is implicit in the definition of $\phi_\eta^{i,j}$, the old and new classes in the sample inherit their ordering from the population.

Note that for $x \in \Delta$,

$$\sum_{k=1}^{j} \sum_{\eta \in B_n^k} \phi_\eta^{k,j}(x) = 1. \tag{9}$$

It is also easy to check that as a function on $\overline{\Delta}$, $\phi_\eta^{k,j}$ is lower semicontinuous. It then follows from (9) that $\phi_\eta^{k,j}$ is actually continuous on Δ.

To cover sampling without replacement, we define for $\eta = (l, \eta_1, \dots, \eta_l) \in B_n^k$, $N, k, j \in \mathbb{N}$, $k \le j \le n$, $\phi_\eta^{N,k,j}\colon \overline{\Delta} \to \mathbb{R}$ by

$$\phi_\eta^{N,k,j}(x) = \sum_{i_1 < \cdots < i_i \le j < i_{k+1} < \cdots < i_l} \frac{\binom{Nx_{i_1}}{\eta_1} \cdots \binom{Nx_{i_l}}{\eta_l}}{\binom{N}{n}}.$$

[If x does describe a population of size N, then $Nx_i \in \mathbb{N}$ for each i; for convenience cover the general case by defining $\binom{a}{b}$ to be $\Gamma(a+1)(\Gamma(a-b+$

$1)\Gamma(b+1))^{-1}$.] Again $\phi_\eta^{N,k,j}$ is lower semicontinuous on $\overline{\Delta}$ and it is actually continuous on Δ.

Let $\mathscr{P}(\overline{\Delta})$ be the space of probability measures on $\overline{\Delta}$. For each $n, k \in \mathbb{N}$, $n \le k$, define $G^{n,k}\colon B_n^k \times \mathscr{P}(\overline{\Delta}) \to \mathbb{R}$ by

$$G^{n,k}(\eta, \mu) = \int_{\overline{\Delta}} \phi_\eta^{k,k}(x)\, d\mu. \tag{10}$$

We interpret $G^{n,k}(\eta,\mu)$ as the probability of getting a sample described by $\eta = (l, \eta_1, \eta_2, \dots, \eta_l)$ from a population described by μ. The sample is taken in such a way that the first k types in the population appear in the sample.

Define

$$S_n^k = \left\{ y \in \Delta \colon \exists l \ge k \text{ with } ny_i \in \mathbb{N}, y_i > 0, i = 1, 2, \dots, l, \sum_{i=1}^{l} y_i = 1 \right\}$$

and with $l_y = \max\{i\colon y_i > 0\}$, define $H\colon S_n^k \to B_n^k$ by

$$H(y) = \left(l_y; ny_1, ny_2, \dots, ny_{l_y}\right).$$

Now let $\hat{G}_\mu^{n,k}$ be the atomic measure on $\overline{\Delta}$ corresponding to $G^{n,k}(\eta,\mu)$:

$$\hat{G}_\mu^{n,k}(A) = \sum_{y \in A \cap S_n^k} G^{n,k}(H(y), \mu). \tag{11}$$

THEOREM 9. *Suppose* $\mu\{x \in \Delta\colon x_i > 0$ *for all* $i\} = 1$. *If* $G^{n,k}(\eta,\mu) = \int_\Delta \phi_\eta^{k,k}(x)\, d\mu(x)$, *then for any bounded continuous function* $f\colon \overline{\Delta} \to \mathbb{R}$, *as* $n \to \infty$,

$$\int_{\overline{\Delta}} f\, d\hat{G}_\mu^{n,k} \to \int_{\overline{\Delta}} f\, d\mu.$$

PROOF. Suppose $x \in \Delta$ has $x_i > 0$ for all i, and let $Z_1, Z_2, \dots$ be i.i.d. with $P[Z_1 = i] = x_i$, $i = 1, 2, \dots$. Put $Y_i^{(n)} = n^{-1}\sum_{j=1}^{n} I\{Z_j = i\}$ and define for $k = 2, 3, \dots$,

$$J_1^{(n)} = \min\left\{i\colon Y_i^{(n)} > 0\right\}, \qquad J_k^{(n)} = \min\left\{i > J_{k-1}^{(n)}\colon Y_i^{(n)} > 0\right\}.$$

If $J_k^{(n)} = \infty$, put $Y_{J_k^{(n)}}^{(n)} = 0$. It is straightforward to prove that for $k = 1, 2, \dots, m$, as $n \to \infty$,

$$I\left\{J_k^{(n)} = k\right\} \to 1 \quad \text{a.s.} \quad \text{and} \quad \left(Y_{J_1^{(n)}}^{(n)}, \dots, Y_{J_k^{(n)}}^{(n)}\right) \to (x_1, \dots, x_k) \quad \text{a.s.} \tag{12}$$

Let $f\colon \overline{\Delta} \to \mathbb{R}$ be a bounded continuous function depending only on a finite number, m say, of coordinates and write $f(y)$ for $f(y_1, y_2, \dots, y_m)$. Then

$$\begin{aligned}
\lim_{n\to\infty} \int_{\overline{\Delta}} f(y)\, d\hat{G}_\mu^{n,k}(y) &= \lim_{n\to\infty} \int_\Delta \sum_{y \in S_n^k} f(y)\, \phi_{H(y)}^{k,k}(x)\, d\mu(x) \\
&= \lim_{n\to\infty} \int_\Delta E\left[f\left(Y_{J_1^{(n)}}^{(n)}, \dots, Y_{J_k^{(n)}}^{(n)}\right) I\left\{J_k^{(n)} = k\right\}\right] d\mu(x) \qquad (13) \\
&= \int_\Delta f(x)\, d\mu(x),
\end{aligned}$$

using dominated convergence and the fact that [again by dominated convergence and (12)] the integrand in (13) converges to $f(x)$ as $n \to \infty$. An application of the Stone–Weierstrass theorem completes the proof. □

Once more fix $t > 0$. Recall that μ_N^k is the conditional distribution of $X_N([N\sigma^{-2}t])$ given that $A_N([N\sigma^{-2}t]) = k$. Denote by μ^k the distribution of $X(t)$ given that $D(t) = k$. (In fact it follows from the structure of the limiting process that μ^k does not depend on t.)

THEOREM 10. *For each* k, $k = 0, 1, 2, \dots, \mu_N^k \Rightarrow \mu^k$ *as* $N \to \infty$.

PROOF. Throughout, we denote $A_N([N\sigma^{-2}t])$ by A_N and $X_N([N\sigma^{-2}t])$ by

$$X_N \equiv \left(X_{N,1}, \dots, X_{N,A_N}, Y_{N,1}, Y_{N,2}, \dots\right).$$

It will be convenient to define distributions for samples corresponding to μ_N^k and μ_N in terms of class sizes rather than proportions. Specifically, for $k, l, n, N \in \mathbb{N}$ with $k \le l \le n \le N$, define $P_N^{n,k}\colon B_n^l \to \mathbb{R}$ by

$$\begin{aligned} P_N^{n,k}(\eta) = P\big[nX_{N,i}^n = \eta_i, 0 < i \le k, nY_{N,i}^n = \eta_i, k < i \le l,\\ nY_{N,i}^n = 0, i > l \mid A_N^n = k\big]. \end{aligned}$$

Further, for $k, l, n \in \mathbb{N}$, $k \le l \le n$, define $P^{n,k}\colon B_n^l \to \mathbb{R}$ by

$$\begin{aligned} P^{n,k}(\eta) = P\big[nX_i^n = \eta_i, 0 < i \le k, nY_i^n = \eta_i, k < i \le l,\\ nY_i^n = 0, i > l \mid A^n = k\big]. \end{aligned}$$

Finally, let $\hat{P}^{n,k}$ be the distribution of X^n given $D^n = k$.

It follows that

$$\begin{aligned} P_N^{n,k}(l; \eta_1, \dots, \eta_l) &= \sum_{j=k}^{\infty} \int_\Delta \varphi_\eta^{N,k,j}(x)\, d\mu_N^j(x) \frac{P[A_N = j]}{P[A_N^n = k]}\\ &= \int_\Delta \varphi_\eta^{N,k,k}(x)\, d\mu_N^k(x) \frac{P[A_N = k]}{P[A_N^n = k]} + h_n^N(\eta), \end{aligned}$$

where

$$h_n^N(\eta) = \sum_{j>k} \int_\Delta \varphi_\eta^{N,k,j}(x)\, d\mu_N^j(x) \frac{P[A_N = j]}{P[A_N^n = k]}.$$

Suppose ν^k is a limit point of μ_N^k and index the convergent subsequence by N; then by Theorem 7, ν^k is concentrated on $\{x \in \Delta\colon x_i > 0 \text{ for all } i\}$. Recall that the functions $\phi_\eta^{k,j}$ and $\phi_\eta^{N,k,j}$ are continuous on Δ. It is straightforward to check that for $x \in \Delta$,

$$\phi_\eta^{N,k,j}(x) \to \phi_\eta^{k,j}(x), \quad \text{uniformly in } x,$$

as $N \to \infty$. It then follows from the continuity theorem [Billingsley (1968), Theorem 5.5] that

$$\lim_{N\to\infty}\int_\Delta \phi_\eta^{N,k,j}(x)\,d\mu_N^j(x)=\int_\Delta \phi_\eta^{k,j}(x)\,d\nu^j. \tag{14}$$

Thus

$$\lim_{N\to\infty} P_N^{n,k}(\eta)=\int_\Delta \frac{\varphi_\eta^{k,k}(x)\,d\nu^k(x)P[D=k]}{P[D^n=k]}+\lim_{N\to\infty} h_n^N(\eta).$$

But by Theorem 1,

$$\lim_{N\to\infty} P_N^{n,k}(\eta)=P^{n,k}(\eta).$$

So for $G^{n,k}(\eta,\nu^k)$ defined by (10) and $\hat{G}_{\nu^k}^{n,k}$ defined by (11), it follows that if $f\colon \bar{\Delta}\to\mathbb{R}$ is bounded and continuous,

$$\begin{aligned}\int_\Delta f(x)\,d\hat{P}^{n,k}(x)&=\int_\Delta f(x)\,d\hat{G}_{\nu^k}^{n,k}(x)\frac{P[D=k]}{P[D^n=k]}\\&\quad+\lim_{N\to\infty}\sum_{\eta\in B_n^k} f(n^{-1}\eta_1,\dots)h_n^N(\eta).\end{aligned}\tag{15}$$

However,

$$\left|\sum_{\eta\in B_n^k} f(n^{-1}\eta_1,\dots)h_n^N(\eta)\right|\le \sup_{x\in\bar{\Delta}}|f(x)|\sum_{\eta\in B_n^k} h_n^N(\eta)$$

and

$$\sum_{\eta\in B_n^k} h_n^N(\eta)\le\sum_{j=k+1}^{M}\int_\Delta\sum_{\eta\in B_n^k}\varphi_\eta^{N,k,j}(x)\,d\mu_N^j(x)\frac{P[A_N=j]}{P[A_N^n=k]}+\frac{P[A_N>M]}{P[A_N^n=k]}.$$

Thus by (14) and Theorem 2,

$$\lim_{N\to\infty}\sum_{\eta\in B_n^k} h_n^N(\eta)\le\sum_{j=k+1}^{M}\int_\Delta\sum_{\eta\in B_n^k}\varphi_\eta^{k,j}(x)\,d\nu^j(x)\frac{P[D=j]}{P[D^n=k]}+\frac{P[D>M]}{P[D^n=k]}.$$

But for $j>k$, it follows from (12) that for $x\in\{x\in\Delta\colon x_i>0 \text{ for each } i\}$,

$$\sum_{\eta\in B_n^k}\varphi_\eta^{k,j}(x)\le P\left[J_j^{(n)}\ne j\right]\to 0\quad\text{as } n\to\infty.$$

Thus

$$\lim_{n\to\infty}\lim_{N\to\infty}\sum_{\eta\in B_n^k} h_n^N(\eta)\le\frac{P[D>M]}{P[D=k]}.$$

Now let $M\to\infty$ to get

$$\lim_{n\to\infty}\lim_{N\to\infty}\left|\sum_{\eta\in B_n^k} f(n^{-1}\eta_1,\dots)h_n^N(\eta)\right|=0. \tag{16}$$

Recall that $\nu^k\{x\in\Delta\colon x_i>0 \text{ for all } i\}=1$, and apply Theorem 8 together with

(17) and (18) to obtain

$$\lim_{n\to\infty}\int_{\Delta} f(x)\,d\hat{P}^{n,k}(x)=\int_{\Delta} f(x)\,d\nu^k(x),$$

implying $\hat{P}^{n,k}\Rightarrow\nu^k$ as $n\to\infty$. By Theorem 4 (recalling that $\hat{P}^{n,k}$ is the distribution of X^n given $A^n=k$), $\hat{P}^{n,k}\Rightarrow\mu^k$ as $n\to\infty$, implying $\nu^k=\mu^k$. Thus all the limit points of the sequence $\{\mu_N^k\}$ are equal to μ^k, and the result follows. (Tightness is immediate since $\bar{\Delta}$ is compact.) □

We are now in a position to prove the main result of the section.

THEOREM 11. *Fix $t>0$ and assume that as $N\to\infty$,*

1. $\mathrm{Var}(\nu_1)\to\sigma^2>0$,
2. $\sup_N E[\nu_1^k]<\infty$, $k=1,2,\dots$,
3. $u=\theta\sigma^2/(2N)+o(N^{-1})$ *for some* $\theta>0$.

Then $(A_N([N\sigma^{-2}t]), X_N([N\sigma^{-2}t]))\Rightarrow(D(t),X(t))$ *as* $N\to\infty$.

PROOF. Again write $A_N(t)$ and $X_N(t)$ for $A_N([N\sigma^{-2}t])$ and $X_N([N\sigma^{-2}t])$, respectively. Let $F\colon E\to\mathbb{R}$ be a bounded continuous function.

Define $f^F, f_N^F\colon\{0,1,2,\dots,\infty\}\to\mathbb{R}$ by

$$f^F(k)=\int_{\bar{\Delta}}F(k,x)\,d\mu^k(x),\qquad f_N^F(k)=\int_{\bar{\Delta}}F(k,x)\,d\mu_N^k(x).$$

Theorem 10 and the topology of $\mathbb{N}$ then guarantee that as $N\to\infty$, $f_N^F(k_N)\to f^F(k)$, for any sequence $k_N\to k$, so that Theorem 2, the continuity theorem and the fact that $D(t)\in\mathbb{N}$ a.s. ensure that

$$f_N^F(A_N(t))\Rightarrow f^F(D(t))\quad\text{as } N\to\infty.$$

For each k, $|f_N^F(k)|\le\|F\|$ so that by dominated convergence, as $N\to\infty$,

$$\begin{aligned}E[F(A_N(t),X_N(t))]&=E[f_N^F(A_N(t))]\to E[f^F(A_N(t))]\\&=E[F(D(t),X(t))]\end{aligned}$$

as required. □

5. Convergence of finite-dimensional distributions.

Conditional on the value of $M_N(t)$, its behavior from time t onwards depends only on the genealogy of the $A_N([N\sigma^{-2}t])$ ancestors of the population. [This is Kingman's (1982a) "temporal coupling".] For $t>0$, this number remains finite as $N\to\infty$ (Theorem 2) and so its limiting genealogy is described by Theorem 1.

Recall that $M_N(t)=(A_N([N\sigma^{-2}t]), X_N([N\sigma^{-2}t]))$ and $M(t)=(D(t),X(t))$.

THEOREM 12. *Suppose $0\le t_1<t_2<\cdots<t_k$. Under the hypotheses of Theorem 1, as $N\to\infty$,*

$$(M_N(t_1),M_N(t_2),\dots,M_N(t_k))\Rightarrow(M(t_1),\dots,M(t_k)).$$

PROOF. First suppose $t_1 > 0$. Theorem 11 gives the result for $k = 1$. We will proceed by induction and assume the result is true for $k - 1$.

It is again convenient to go back to describing genealogies of samples by equivalence relation valued processes. These processes are the obvious analogs of $X_N^n(\cdot)$ and $X^n(\cdot)$. We denote them by $R_N^n(\cdot)$ and $R^n(\cdot)$, respectively; they are described in detail in Donnelly and Tavaré (1986). Their state space is the set of two type equivalence relations on the set $\{1, 2, \dots, n\}$, which we denote by $\mathscr{R}_n$. If $\alpha = (\xi_1, \xi_2, \dots, \xi_k; \eta_1, \eta_2, \dots, \eta_l) \in \mathscr{R}_n$, then the sets $\xi_1, \dots, \xi_k, \eta_1, \dots, \eta_l$, form a partition of $\{1, 2, \dots, n\}$ and the interpretation is that α has k old classes and l new classes, with the new classes being listed in order of decreasing age, and say $\eta_i = \{j_1, \dots, j_m\}$ means that the individuals labeled $j_1, \dots, j_m$ comprise the ith oldest new class. [In Donnelly and Tavaré (1986), the new classes were listed in the opposite order, i.e., from youngest to oldest.] Thus, for example, if

$$R_N^n(t) = (\xi_1, \xi_2, \dots, \xi_k; \eta_1, \dots, \eta_l)$$

then

$$X_N^n(t) = \left(\frac{|\xi_1|}{n}, \dots, \frac{|\xi_k|}{n}, \frac{|\eta_1|}{n}, \dots, \frac{|\eta_l|}{n}, 0, 0, \dots \right),$$

where $|A|$ represents the number of elements in A.

Define f_n: $\mathscr{R}_n \times \overline{\Delta} \to E$ as follows: Suppose $\alpha \in \mathscr{R}_n$ has k old classes $\xi_1, \xi_2, \dots, \xi_k$ and l new classes $\eta_1, \eta_2, \dots, \eta_l$ where η_i is the ith oldest new class, then define

$$f_n(\alpha, x) = (k; z_1, z_2, \dots),$$

where $z_i = \sum_{j \in \xi_i} x_j$ for $i \le k$, $z_{k+i} = \sum_{j \in \eta_i} x_j$ for $i \le l$ and $z_{k+l+i} = x_{n+i}$ for $i > l$.

Note that $f_n(\alpha, \cdot)$ is continuous. Further, we endow $\mathscr{R}_n$ with the discrete topology so that $f_n(\cdot, x)$ is also continuous. The interpretation of $f_n(\alpha, x)$ is (via the temporal coupling) that if the population genealogy at some time t is given by (m, x) and the m-sample genealogical process $R_N^m(\cdot)$ takes the value $\alpha \in \mathscr{R}_m$ at time s, then the population genealogy at time $t + s$ has the same distribution as $f_m(\alpha, x)$.

Suppose $y_1, y_2, \dots, y_{k-1} = (n, x) \in E$ and denote $(y_1, \dots, y_{k-1}) \in E^{k-1}$ by y. Let F: $E^k \to \mathbb{R}$ $(k \ge 2)$ be bounded and continuous. Define h_N, h: $E^{k-1} \to \mathbb{R}$, by

$$\begin{aligned} h_N(y) &= E\left[F(y, f_n(R_N^n(t_k - t_{k-1}), x))\right], \\ h(y) &= E\left[F(y, f_n(R^n(t_k - t_{k-1}), x))\right]. \end{aligned}$$

The Markov property and temporal coupling give

$$E\left[F(M_N(t_1), \dots, M_N(t_k))\right] = E\left[h_N(M_N(t_1), \dots, M_N(t_{k-1}))\right]$$

and

$$E\left[F(M(t_1), \dots, M(t_k))\right] = E\left[h(M(t_1), \dots, M(t_{k-1}))\right].$$

Now suppose $y^{(N)} \equiv (y_1^{(N)}, y_2^{(N)}, \dots, y_{k-1}^{(N)}) \to y \equiv (y_1, \dots, y_{k-1})$ and throughout write $y_{k-1}^{(N)} = (n^{(N)}, x^{(N)})$, $y_{k-1} = (n, x)$. Further suppose $n \in \mathbb{N}$, $x \in \Delta$ and N large enough to ensure that $n^{(N)} = n$. Define $g_N, g\colon \mathscr{R}_n \to \mathbb{R}$ by

$$g_N(\alpha) = F\big(y^{(N)}, f_n(\alpha, x^{(N)})\big), \qquad g(\alpha) = F\big(y, f_n(\alpha, x)\big).$$

The continuity of f_n, F and the discrete topology on $\mathscr{R}_n$ ensure that $g_N(\alpha_N) \to g(\alpha)$ whenever $\alpha_N \to \alpha$. The continuity theorem, Theorem 1 and the boundness of g, imply that as $N \to \infty$,

$$h_N(y^{(N)}) = E\big[g_N(R_N^n(t_k - t_{k-1}))\big] \to E\big[g(R^n(t_k - t_{k-1}))\big] = h(y).$$

The distribution of $(M(t_1), \dots, M(t_k))$ is concentrated on $\mathbb{N} \times \Delta$, so that the induction hypothesis, the boundedness of h and another application of the continuity theorem give

$$E\big[h_N(M_N(t_1), \dots, M_N(t_{k-1}))\big] \to E\big[h(M(t_1), \dots, M(t_k))\big]$$

as $N \to \infty$, as required.

Finally it is straightforward to check that $M_N(0) \Rightarrow M(0)$. If $t_1 = 0$, then Theorem 11 effectively gives the desired result for $k = 2$ and the above induction argument still applies, to extend the result to arbitrary k. □

6. Tightness. Recall that $E = \{0, 1, 2, \dots, \infty\} \times \bar{\Delta}$ is a compact metric space. We use the metric

$$r((n,x),(m,y)) = d_1(n,m) + d_2(x,y),$$

where d_1 and d_2 are metrics on $\{0, 1, \dots, \infty\}$ and $\bar{\Delta}$, respectively. Without loss of generality assume that for $m < n$, $d_1(n,m)$ is increasing in n for fixed m and $d_1(n,\infty)$ is decreasing in n. For the sake of concreteness define d_2 by

$$d_2((x_1, x_2, \dots), (y_1, y_2, \dots)) = \sum_{k=1}^{\infty} \frac{|x_k - y_k|}{2^k}.$$

In order to show that $M_N(\cdot)$ is tight we will verify the conditions of Corollary 3.7.4 in Ethier and Kurtz (1986). First note that the compact containment condition (a) is immediate. It remains to verify:

(b) For every $\eta > 0$ and $T > 0$, there exists a $\delta > 0$ such that

$$\limsup_N P\big(w'_0(M_N, \delta, T) \geq \eta\big) \leq \eta, \tag{17}$$

where

$$w'_\alpha(M_N, \delta, T) = \inf_{\{t_i\}} \max_i \sup_{s,t \in [t_{i-1}, t_i)} r(M_N(s), M_N(t))$$

with $\{t_i\}$ ranging over all partitions of the form $\alpha = t_0 < t_1 < \cdots < t_{n-1} < T \leq t_n$ with

$$\min_{1 \leq i \leq n} (t_i - t_{i-1}) > \delta.$$

Henceforth treat $\eta > 0$ and $T > 0$ as fixed. Since for any $\alpha \geq \delta$,

$$w'_0(M_N, \delta, T) \leq \sup_{s, t \in [0, \alpha)} r(M_N(s), M_N(t)) + w'_\alpha(M_N, \delta, T)$$

the required result (17) will follow if we can first find an $\alpha > 0$ such that

$$\limsup_N P\Big(\sup_{s, t \in [0, \alpha)} r(M_N(s), M_N(t)) \geq \eta/2 \Big) \leq \eta/2 \tag{18}$$

and then find a $\delta(\eta, \alpha) > 0$ (with $\delta < \alpha$), where

$$\limsup_N P(w'_\alpha(M_N, \delta, T) \geq \eta/2) \leq \eta/2. \tag{19}$$

THEOREM 13. *The family of processes* $\{M_N(t); t \geq 0\}$, $N = 1, 2, \dots$, *is tight in* $D_E[0, \infty)$.

PROOF. Choose $z < \infty$ such that $d_1(\infty, z) \leq \eta/4$. Recall that $A_N(t) \Rightarrow D(t)$ for all $t > 0$ and that $A_N(t)$ is nondecreasing. Then

$$\begin{aligned} &\limsup_N P\Big(\sup_{s, t \in [0, \alpha)} d_1(A_N(s), A_N(t)) > \frac{\eta}{4} \Big) \\ &\quad \leq \limsup_N P\Big(d_1(N, A_N(\alpha)) > \frac{\eta}{4} \Big) \\ &\quad \leq \limsup_N P(A_N(\alpha) < z) \\ &\quad = P(D(\alpha) < z). \end{aligned} \tag{20}$$

Next, let $S: \bar{\Delta} \to \mathbb{R}$ be defined by $S((x_1, x_2, \dots)) = \max\{x_1, x_2, \dots\}$. Note that $X_N(\alpha) \Rightarrow X(\alpha)$, $P(X(\alpha) \in \Delta) = 1$ for $\alpha > 0$ and S is continuous on Δ [Donnelly and Joyce (1989)]. Note also that $S(X_N(\cdot))$ is nondecreasing. Thus

$$\begin{aligned} &\limsup_N P\Big(\sup_{s, t \in [0, \alpha)} d_2(X_N(s), X_N(t)) > \frac{\eta}{4} \Big) \\ &\quad \leq \limsup_N P\Big(S(X_N(\alpha)) > \frac{\eta}{4} \Big) \\ &\quad = P\Big(S(X(\alpha)) > \frac{\eta}{4} \Big). \end{aligned} \tag{21}$$

It follows from the structure of $X(\alpha)$ that

$$S(X(\alpha)) =_{\mathscr{D}} \max\big(V_{D(\alpha)} U_{(1)}^{D(\alpha)}, (1 - V_{D(\alpha)}) W \big),$$

where for each k, V_k has a Beta distribution given by (2), $U_{(1)}^k$ is the largest order statistic of $U_1, \dots, U_k$ defined at (1) and W is the first component of the Poisson–Dirichlet distribution with parameter θ [see, e.g., Kingman (1978)].

Note that $D(\alpha) \Rightarrow \infty$ as $\alpha \to 0$, and it is a straightforward exercise to show that $S(X(\alpha)) \Rightarrow 0$ as $\alpha \to 0$. Therefore for sufficiently small α, (18) follows from (20) and (21).

Let M_N^α be the restriction of the process M_N to $[\alpha, \infty)$. Define $S_\alpha^\varepsilon \subset D_E[\alpha, \infty)$ by

$$S_\alpha^\varepsilon = \{\text{step functions in } D_E[\alpha, \infty) \text{ whose interjump times are at least } \varepsilon\}.$$

We will show that M_N^α lives on S_α^ε with high probability for sufficiently small ε. To see that this guarantees (19), we note that S_α^ε is relatively compact in $D_E[\alpha, \infty)$ [see Ethier and Kurtz (1986), Lemma 3.6.1 and Theorem 3.6.3].

Let $\tau_1^{N,n}, \tau_2^{N,n}, \dots, \tau_n^{N,n}$ be the interjump times for the process $A_N^n(\cdot)$. Let $\tau_1^n, \tau_2^n, \dots, \tau_n^n$ be the interjump times for $D^n(\cdot)$. Note that the interjump times for $D^n(\cdot)$ are independent and that τ_1^n is exponential with parameter $n(n+\theta-1)/2$ and $\tau_i^n =_{\mathscr{D}} \tau_1^{n-i}$.

Thus

$$\begin{aligned} P(M_N^\alpha \in S_\alpha^\varepsilon) &\geq P(M_N^\alpha \in S_\alpha^\varepsilon, A_N(\alpha) \leq L) \\ &= \sum_{i=1}^{L} P(M_N^\alpha \in S_\alpha^\varepsilon | A_N(\alpha) = i) P(A_N(\alpha) = i) \\ &= \sum_{i=1}^{L} P\big(\tau_j^{N,i} > \varepsilon \text{ for all } j \leq i\big) P(A_N(\alpha) = i), \end{aligned}$$

in view of the conditional independence property and the fact that $M_N(\cdot)$ jumps only when $A_N(\cdot)$ does.

Finally

$$\begin{aligned} \liminf_N P(M_N^\alpha \in S_\alpha^\varepsilon) &\geq \sum_{i=1}^{L} P\big(\tau_j^i > \varepsilon \text{ for all } j \leq i\big) P(D(\alpha) = i) \\ &\geq \big(P(\tau_1^L > \varepsilon)\big)^L P(D(\alpha) \leq L). \end{aligned} \tag{22}$$

We can make the right-hand side of the above inequality arbitrarily close to 1 by first choosing L sufficiently large and then choosing ε small. Thus (19) follows from (22). □

REFERENCES

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

CANNINGS, C. (1974). The latent roots of certain Markov chains arising in genetics: A new approach. I. Haploid models. *Adv. in Appl. Probab.* **6** 260–290.

COX, J. T. and GRIFFEATH, D. (1990). Mean field asymptotics for the planar stepping stone model. *Proc. London Math. Soc.* **61** 189–208.

DONNELLY, P. (1991). Weak convergence to a Markov chain with an entrance boundary: Ancestral processes in population genetics. *Ann. Probab.* **19** 1102–1117.

DONNELLY, P. and JOYCE, P. (1989). Continuity and weak convergence of ranked and size-biased permutations on the infinite simplex. *Stochastic Process. Appl.* **31** 89–104.

DONNELLY, P. and JOYCE, P. (1991). Consistent ordered sampling distributions: Characterization and convergence. *Adv. in Appl. Probab.* **23**.

DONNELLY, P. and TAVARÉ, S. (1986). The ages of alleles and a coalescent. *Adv. in Appl. Probab.* **18** 1–19.

DONNELLY, P. and TAVARÉ, S. (1987). The population genealogy of the infinitely-many neutral alleles model. *J. Math. Biol.* **25** 381–391.

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

KINGMAN, J. F. C. (1978). Random partitions in population genetics. *Proc. Roy. Soc. London A* **361** 1–20.

KINGMAN, J. F. C. (1982a). On the genealogy of large populations. *J. Appl. Probab.* **19A** 27–43.

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

KINGMAN, J. F. C. (1982c). Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.) 97–112. North-Holland, Amsterdam.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

TAVARÉ, S. (1984). Line-of-descent and genealogical processes and their applications in population genetics models. *Theoret. Population Biol.* **26** 119–164.

WATTERSON, G. A. (1984). Lines of descent and the coalescent. *Theoret. Population Biol.* **25** 77–92.

SCHOOL OF MATHEMATICAL SCIENCES
QUEEN MARY AND WESTFIELD COLLEGE
UNIVERSITY OF LONDON
MILE END ROAD
LONDON E1 4NS
ENGLAND

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF IDAHO
MOSCOW, IDAHO 83843