WEAK CONVERGENCE TO A MARKOV CHAIN WITH AN ENTRANCE BOUNDARY: ANCESTRAL PROCESSES IN POPULATION GENETICS[1]

By Peter Donnelly

Queen Mary and Westfield College

We derive conditions under which a sequence of processes will converge to a (continuous-time) Markov chain with an entrance boundary. Our main application of this result is in proving weak convergence of the so-called population ancestral processes, associated with a wide class of exchangeable reproductive models, to a particular death process with an entrance boundary at infinity. This settles a conjecture of Kingman. We also prove weak convergence of the absorption times of many neutral genetics models to that of the Wright–Fisher diffusion, and convergence of population line-of-descent processes to another death process.

1. Introduction. In spite of recent advances and the continuing development of the theory of weak convergence, together with its extensive application [see for example Ethier and Kurtz (1986)], one setting which does not seem to have attracted special attention is that in which the limit process has an entrance boundary which is the accumulation point of a countable state space. The work described here is motivated by an outstanding problem in population genetics in which such processes arise naturally. The other novelty in that context is that the raw materials available for proving weak convergence differ from those usually available. Here one does not have knowledge of finite dimensional distributions but of the behavior of the prelimiting processes after they have reached "finite" values. Equivalently, in trying to infer the behavior of processes describing the whole population, in the limit as the population size tends to infinity, one is armed with knowledge of the limiting behavior of samples of fixed but arbitrary size.

In Section 2 we prove a general result. As well as highlighting the essential features of the problem, the result and the associated techniques may be of wider interest. Section 3 introduces the genetics models, defines the ancestral process and gives the finite sample result of Kingman (1982c). Next we verify that, in an appropriate sense, the prelimiting processes do not "get stuck at infinity." The convergence result may then be applied to prove that with an appropriate time scaling, population ancestral processes associated with a wide range of exchangeable models converge to a particular death process with an

---

entrance boundary at infinity. Convergence of (neutral) discrete genetic models forward in time to the familiar diffusion approximations is [Donnelly (1985)] a consequence of the finite sample result. In Section 4 we show that weak convergence of absorption times to those of the diffusion is a consequence of convergence of population ancestral processes, thus extending some results of Ethier (1979) to a wide class of neutral models. In fact weak convergence of absorption times is equivalent to weak convergence of the marginal distribution of the ancestral process to that of the limiting death process. Griffiths (1981) used known results about absorption times in the specific case of the Wright–Fisher reproductive scheme to prove convergence of ancestral processes for that model. The final section discusses the related problem of population lines of descent in genetic models incorporating mutation, and again uses the general result to prove weak convergence to a (different) death process.

Throughout, we will use the symbol $\Rightarrow$ to denote weak convergence, $\overset{\text{st}}{\leq}$ for stochastic ordering of elements of a partially ordered set [Kamae, Krengel and O'Brien (1977)], and write $[x]$ for the integer part of x.

2. The general convergence result. The form of the conditions for Theorem 1 is motivated by the genetics problem we discuss later. Note that the property of (stochastic) monotonicity in the initial value is natural in that context, as it is indeed for many death, or birth–death type processes. We remark, however, that while this property is used repeatedly throughout our proof, similar techniques are applicable in related situations in which other conditions may apply. In connection with condition (A3) of Theorem 1, one may well have hoped [Kingman (1982c)] that the "finite sample" behavior (A2) might play a role analogous to that of finite dimensional distributions in other invariance results. This does not seem to be the case, and some condition like (A3), whose effect is to ensure that the prelimiting processes do not get "stuck" at the entrance boundary, appears essential.

THEOREM 1. *Let $\{B^N(\cdot)\}_{N=1}^\infty$ be a sequence of Markov processes with respective state spaces $E_1 \subseteq E_2 \subseteq \cdots \subseteq \{0, 1, 2, \ldots\}$. Denote by E' the union $\bigcup_{N=1}^\infty E_N$ and for $n \in E_N$, by $B_n^N(\cdot)$ the process $B^N(\cdot)$ conditional on $B^N(0) = n$. Suppose that $\{D(t), t \geq 0\}$ is Markov with state space $E = E' \cup \{\infty\}$ [again we write $D_n(\cdot)$ for the process $D(\cdot)$ conditional on $D(0) = n$] and that:*

(i) *The state $\infty \in E$ is an entrance boundary for the process.*
(ii) *No states $x \in E'$ are instantaneous.*
(iii) *Whenever $k, n \in E'$ with $k \leq n, t > 0$, $P(D_n(t) = k) > 0$.*

In addition, we make the following assumptions:
(A1) *For $l, n \in E_N$ with $l < n$,*

$$B_l^N(\cdot) \overset{\text{st}}{\leq} B_n^N(\cdot)$$

as elements of $D_{E_N}[0, \infty]$.

(A2) *For each $n \in E'$, as $N \to \infty$ (with $N > \min\{N: n \in E_N\}$),*

$$B_n^N(\cdot) \Rightarrow D_n(\cdot)$$

as elements of $D_E[0, \infty)$.
  (A3) *For each $t > 0$,*

$$\lim_{M \to \infty} \liminf_{N \to \infty} P\big(B_{n_N}^N(t) \le M\big) = 1$$

for some sequence $\{n_N\}_{N=1}^{\infty}$.
  Then, provided $n_N \to \infty$ as $N \to \infty$,

$$B_{n_N}^N(\cdot) \Rightarrow D_\infty(\cdot)$$

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

PROOF. We will first show that the sequence $\{B_{n_N}^N\}_{N=1}^{\infty}$ is tight, then collect together properties inherited by any weak limit of a subsequence of the $B_{n_N}^N$ and finally show that these properties characterize any such limit as the process $D_\infty(\cdot)$.
  For tightness we verify Aldous' sufficient condition: For each T, each η and each $\varepsilon > 0$, there exists a $\delta > 0$ and N_0 such that

$$P\big(d\big(B_{n_N}^N(\rho_N + \delta'), B_{n_N}^N(\rho_N)\big) \ge \eta\big) < \varepsilon \quad \text{for } N > N_0,$$

whenever ρ_N is a stopping time for $B_{n_N}^N$ that takes values in $[0, T]$ and δ' is a real number with $0 \le \delta' \le \delta$. [Here and throughout we use $d(\cdot, \cdot)$ to denote a metric which generates the usual topology on E. Without loss of generality we will further assume that $d(x, \infty)$ is decreasing in x.] Pollard (1984) shows that an analogous condition guarantees tightness in $D_{\mathbb{R}}[0, \infty]$. The compactness of E and Theorem 3.9.1 of Ethier and Kurtz (1986), for example, gives the desired extension.
  Choose $z \in E$ with $d(z, \infty) < \eta/2$ and suppose that N is large enough to ensure $z \in E_N$. Then, conditional on $B_{n_N}^N(\rho_N) = K \ge z$,

$$\sup_{\delta' \le \delta} d\big(B_{n_N}^N(\rho_N + \delta'), B_{n_N}^N(\rho_N)\big) \le d(K, \infty) + d\Big(\inf_{\delta' \le \delta} B_K^N(\delta'), \infty\Big)$$

$$\overset{\text{st}}{\le} \eta/2 + d\Big(\inf_{\delta' \le \delta} B_z^N(\delta'), \infty\Big) \quad \text{by (A1).}$$

Since by (A2), $B_z^N \Rightarrow D_z$, it follows that

$$\inf_{\delta' \le \delta} B_z^N(\delta') \Rightarrow \inf_{\delta' \le \delta} D_z(\delta')$$

and, in particular, that

$$\lim_{N \to \infty} P\Big(\inf_{\delta' \le \delta} B_z^N(\delta') \le z - 1\Big) = P\Big(\inf_{\delta' \le \delta} D_z(\delta') \le z - 1\Big)$$

$$\le 1 - P\big(\{D_z(\delta') = z, 0 \le \delta' \le \delta\}\big).$$

Assumption (ii) guarantees that there is a value of δ, δ_z say, which ensures

that this last is bounded above by $\varepsilon/2$ and hence that for some N_z,

$$P\left(\inf_{\delta' \le \delta} B_z^N(\delta') \le z - 1 \right) < \varepsilon \quad \text{for all } N > N_z.$$

Thus for $0 \le \delta' \le \delta_z$, if $K \in E_N$, $K \ge z$,

$$P\left(d\left(B_{n_N}^N(\rho_N + \delta'), B_{n_N}^N(\rho_N) \right) \ge \eta \,\middle|\, B_{n_N}^N(\rho_N) = K \right) < \varepsilon$$

for all $N > N_z$.

Now for $y \in E_N$ with $y < z$,

$$P\left(d\left(B_{n_N}^N(\rho_N + \delta'), B_{n_N}^N(\rho_N) \right) < \eta \,\middle|\, B_{n_N}^N(\rho_N) = y \right)$$
$$\ge P\left(\left\{ B_y^N(\delta') = y, 0 \le \delta' \le \delta \right\} \right)$$
$$\to P\left(\left\{ D_y(\delta') = y, 0 \le \delta' \le \delta \right\} \right)$$

as $N \to \infty$, as a consequence of (A2). Further, assumption (ii) ensures that there is a value of δ, δ_y say, for which this last probability is at least $1 - \varepsilon/2$ and hence that for some N_y,

$$P\left(\left\{ B_y^N(\delta') = y, 0 \le \delta' \le \delta_y \right\} \right) \ge 1 - \varepsilon \quad \text{for all } N > N_y.$$

Thus whenever $0 \le \delta' \le \delta_y$,

$$P\left(d\left(B_{n_N}^N(\rho_N + \delta'), B_{n_N}^N(\rho_N) \right) \ge \eta \,\middle|\, B_{n_N}^N(\rho_N) = y \right) < \varepsilon \quad \text{for all } N > N_y.$$

Finally, put

$$\delta_0 = \min\{\delta_y : y \in E, y \le z\}$$

and

$$N_0 = \max\{N_y : y \in E, y \le z\}.$$

Thus for all δ' with $0 \le \delta' \le \delta_0$, and all $N > N_0$,

$$P\left(d\left(B_{n_N}^N(\rho_N + \delta'), B_{n_N}^N(\rho_N) \right) \ge \eta \right)$$
$$= \sum_{y \in E_N} P\left(d\left(B_{n_N}^N(\rho_N + \delta'), B_{n_N}^N(\rho_N) \right) \ge \eta, \, B_{n_N}^N(\rho_N) = y \right)$$
$$\le \varepsilon \sum_{y \in E_N} P\left(B_{n_N}^N(\rho_N) = y \right)$$
$$= \varepsilon,$$

so that the sequence $\{B_{n_N}^N(\cdot)\}_{N=1}^{\infty}$ is tight.

Now focus attention on a particular weakly convergent subsequence of $\{B_{n_N}^N(\cdot)\}_{N=1}^{\infty}$ and denote its limit by $X(\cdot)$. (For convenience, here and below, we will also index the subsequence by n_N.) Note that as Jacod, Memin and Metevier (1983) observe, the weak limit of any sequence of processes satisfying Aldous' tightness condition will be quasi-left-continuous. In particular

$$\{t : P(X(t-) = X(t)) = 1\} = (0, \infty)$$

so that, for example, if $0 \le t_0 < t_1 < \cdots < t_n < \infty$, then as $N \to \infty$,

$$\left(B_{n_N}^N(t_0), B_{n_N}^N(t_1), \ldots, B_{n_N}^N(t_n) \right) \Rightarrow \left(X(t_0), X(t_1), \ldots, X(t_n) \right).$$

Next we show that the limiting process $X(\cdot)$ enjoys the following properties:

(P1) $P(X(0) = \infty) = 1$.

(P2) For any $t > 0$, $P(X(t) < \infty) = 1$.

(P3) For any $t > 0$, $k \in E'$, $P(X(t) = k) > 0$.

(P4) For $t > s > 0$, $k, j \in E'$, $P(X(t) = k|X(s) = j) = P(D_j(t - s) = k)$.

(P5) For any $t > 0$, $k \in E'$, $P(X(t) = k) = P(D_\infty(t) = k)$.

Recall that (along the subsequence in question) $n_N \to \infty$. Thus for any M,

$$P(X(0) \ge M) \ge \limsup_{N \to \infty} P\left(B_{n_N}^N(0) \ge M \right) = 1,$$

from which (P1) follows.

For (P2), note that for $t > 0$,

$$\begin{aligned}
P(X(t) < \infty) &= \lim_{M \to \infty} P(X(t) \le M) \\
&= \lim_{M \to \infty} \lim_{N \to \infty} P\left(B_{n_N}^N(t) \le M \right) \\
&= 1
\end{aligned}$$

by (A3).

Next suppose $t > 0$, $k \in E'$. Then with $0 < s < t$, Fatou's lemma and (A2) guarantee that

$$\begin{aligned}
\text{(1)} \qquad P(X(t) = k) &= \lim_{N \to \infty} P\left(B_{n_N}^N(t) = k \right) \\
&\ge \sum_{l=k}^{\infty} P(X(s) = l) P(D_l(t - s) = k).
\end{aligned}$$

Now assumption (iii) ensures that for each $l \ge k$, $P(D_l(t - s) = k) > 0$ for any $0 < s < t$, so that (P3) will follow provided we can find an $l \ge k$, $0 < s < t$, with $P(X(s) = l) > 0$. The existence of such a pair follows from (P2) and the fact that $\lim_{s \downarrow 0} P(X(s) > k) = 1$ [recall (P1) and the right continuity of $X(\cdot)$].

Now for $t > s > 0$, $k, j \in E$,

$$\begin{aligned}
P(X(s) = j) P(X(t) = k|X(s) = j) &= P(X(s) = j, X(t) = k) \\
&= \lim_{N \to \infty} P\left(B_{n_N}^N(s) = j, B_{n_N}^N(t) = k \right) \\
&= \lim_{N \to \infty} P\left(B_{n_N}^N(s) = j \right) P\left(B_{n_N}^N(t) = k \big| B_{n_N}^N(s) = j \right) \\
&= P(X(s) = j) P\left(D_j(t - s) = k \right)
\end{aligned}$$

by (A2), so that (P4) now follows from (P3).

Before establishing (P5), we derive several properties of the process $D(\cdot)$. Note that in many applications these will be immediate; the point here is that they follow from the assumptions of the theorem. First observe that $D(\cdot)$

inherits monotonicity properties from the $B^N(\cdot)$. For our purposes it is enough that as a consequence of (A1) and (A2), for $L, L' \in E'$ with $L < L'$ and $k \in E'$, $t > 0$,

$$P\big(D_L(t) \geq k\big) \leq P\big(D_{L'}(t) \geq k\big).$$

Next we prove that for any $t > 0$, $k \in E'$,

$$P\big(D_L(t - h) \geq k\big) \to P\big(D_L(t) \geq k\big)$$

as $h \to 0$, uniformly in $L \in E$. For

$$\big|P\big(D_L(t - h) \geq k\big) - P\big(D_L(t) \geq k\big)\big|$$

$$= P\big(D_L(t - h) \geq k, D_L(t) < k\big) + P\big(D_L(t - h) < k, D_L(t) \geq k\big)$$

$$= \sum_{\substack{j \in E' \\ j \geq k}} P\big(D_L(t - h) = j\big)P\big(D_j(h) < k\big)$$

$$+ \sum_{\substack{j \in E' \\ j < k}} P\big(D_L(t - h) = j\big)P\big(D_j(h) \geq k\big)$$

$$\leq P\big(D_k(h) < k\big) + P\big(D_{\tilde{k}}(h) \geq k\big)$$

by monotonicity, where we have written $\tilde{k} = \max\{j \in E' : j < k\}$. By assumption (ii), each of these may be made small by taking h small enough, from which the desired result follows. Now we show that for any $t > 0$, $k \in E'$,

$$P\big(D_\infty(t) \geq k\big) = \lim_{L \to \infty} P\big(D_L(t) \geq k\big).$$

Write π_k for the limit of the bounded, monotone sequence $\{P(D_L(t) \geq k)\}_{L \in E'}$ and suppose $\varepsilon > 0$ is given. Then, with h chosen to ensure

$$\big|P\big(D_L(t - h) \geq k\big) - P\big(D_L(t) \geq k\big)\big| \leq \varepsilon$$

for all $L \in E'$,

$$P\big(D_\infty(t) \geq k\big) = \sum_{j \in E'} P\big(D_\infty(h) = j\big)P\big(D_j(t - h) \geq k\big)$$

$$\leq \sum_{j \in E'} P\big(D_\infty(h) = j\big)P\big(D_j(t) \geq k\big) + \varepsilon$$

$$\leq \pi_k + \varepsilon.$$

In the other direction, for the same h,

$$P\big(D_\infty(t) \geq k\big) = \sum_{j \in E'} P\big(D_\infty(h) = j\big)P\big(D_j(t - h) \geq k\big)$$

$$\geq \sum_{\substack{j \in E' \\ j \geq M}} P\big(D_\infty(h) = j\big)P\big(D_j(t - h) \geq k\big)$$

$$\geq P\big(D_M(t) \geq k\big) - \varepsilon$$

$$\geq \pi_k - 2\varepsilon$$

for suitably large M.

To prove (P5), write, for $0 < h < t$,

$$P(X(t) \geq k) = \sum_{j \in E'} P(X(h) = j) P(X(t) \geq k | X(h) = j)$$

$$= \sum_{j \in E'} P(X(h) = j) P(D_j(t - h) \geq k)$$

$$\geq P(D_L(t - h) \geq k) P(X(h) \geq L)$$

$$\rightarrow P(D_\infty(t) \geq k)$$

as first $h \rightarrow 0$ (recall that X is right continuous) and then $L \rightarrow \infty$. Also

$$P(X(t) \geq k) = \sum_{j \in E'} P(X(h) = j) P(D_j(t - h) \geq k)$$

$$\leq P(D_\infty(t - h) \geq k)$$

$$\rightarrow P(D_\infty(t) \geq k)$$

as $h \rightarrow 0$, from which (P5) follows.

Finally, we are in a position to establish the theorem by proving that X and D_∞ have the same finite dimensional distributions. Thus suppose $t_1 < t_2 < \cdots < t_n$ and $k_1, k_2, \ldots, k_n \in E$. Note that in view of (P1), we may assume $t_1 > 0$ and in view of (P2) and (i), we may take $k_1, k_2, \ldots, k_n \in E'$, all without loss of generality to what follows. Then

$$P(X(t_1) = k_1, X(t_2) = k_2, \ldots, X(t_n) = k_n)$$

$$= \lim_{N \rightarrow \infty} P(B_{n_N}^N(t_1) = k_1, B_{n_N}^N(t_2) = k_2, \ldots, B_{n_N}^N(t_n) = k_n)$$

$$= \lim_{N \rightarrow \infty} P(B_{n_N}^N(t_1) = k_1) P(B_{n_N}^N(t_2) = k_2 | B_{n_N}^N(t_1) = k_1) \cdots$$

$$P(B_{n_N}^N(t_n) = k_n | B_{n_N}^N(t_{n-1}) = k_{n-1})$$

$$= P(X(t_1) = k_1) P(D_{k_1}(t_2 - t_1) = k_2) \cdots P(D_{k_{n-1}}(t_n - t_{n-1}) = k_n)$$

$$= P(D_\infty(t_1) = k_1, D_\infty(t_2) = k_2, \ldots, D_\infty(t_n) = k_n)$$

in view of (P5) and the Markov structure of B^N and D_∞. This is enough [recall the topology of E' and, for example, Lemma 3.4.5(a) of Ethier and Kurtz (1986)] to show that X and D_∞ do have the same finite dimensional distributions. $\square$

3. Convergence of ancestral processes. Much recent work in mathematical population genetics has focussed on the genealogy induced by genetic models. The central result [Kingman (1982a, c)] states that with an appropriate time scaling the genealogy of a sample of *fixed size n*, say, arising from any of a wide class of reproductive models, converges as the population size tends to infinity to a particularly tractable process called an *n*-coalescent. A consequence of this is that (in rescaled time) the so-called ancestral process, which simply counts the number of ancestors of the sample, converges to a particular

death process. While many questions of genetic interest may be expressed in terms of the genealogy of samples of fixed size, some problems require knowledge of the genealogy of the whole population. Potential limiting processes must take particular forms and have been closely studied [Kingman (1982b), Donnelly and Tavaré (1987)], but no convergence results exist. In many senses the most fundamental of the open problems, conjectured but unsolved in Kingman (1982c), concerns the convergence of the ancestral process associated with the whole population. Our purpose here is to prove that conjecture.

Our starting point is a general model, initially introduced by Cannings (1974), for the reproduction of a haploid population. [See Ewens (1979) for background.] Assume that the population evolves in discrete, nonoverlapping generations and that each generation is of fixed size N. [In fact, as noted in Kingman (1982a) and elsewhere, the following results apply to a number of models with overlapping generations (notably the Moran model) and varying population sizes, but in these cases a different time scaling is appropriate.] Label the members of the rth generation in random order and let ν_i be the number of children born to the ith member. The random variables $\nu_1, \nu_2, \ldots, \nu_N$ are then exchangeable and $\nu_1 + \nu_2 + \cdots + \nu_N = N$. Further assume that the ν_j are independent of family sizes in other generations (this is an assumption of genetic neutrality) and that the joint distribution of the ν_j does not depend on r. The most celebrated model of population genetics, the Wright–Fisher model, corresponds to the case in which the joint distribution of these offspring numbers is symmetric multinomial.

Choose a sample of n ($\leq N$) members at random from the rth generation. For $t = 0, 1, \ldots, r$, let $A_n^N(t)$ denote the number of distinct ancestors of the sample in generation $r - t$. Then $A_n^N(0) = n$ and $A_n^N(t + 1) \leq A_n^N(t)$. Furthermore the conditional distribution of $A_n^N(t + 1)$ given $A_n^N(0), A_n^N(1), \ldots, A_n^N(t)$ depends only on $A_n^N(t)$. Thus $\{A_n^N(t): t = 0, 1, \ldots, r\}$ is a Markov chain on $E_n = \{1, 2, \ldots, n\}$ whose transition probabilities depend only on N, n and the joint distribution of the offspring numbers ν_j. Although only defined for $t \leq r$, its distributions do not depend on r and are therefore those of a Markov chain $\{A_n^N(t), t = 0, 1, 2, \ldots\}$. We will identify these two chains throughout the sequel. Thus for example, for the Wright–Fisher model,

$$P\big(A_n^N(t + 1) = j \mid A_n^N(t) = i\big)$$

$$= \sum_{1 \leq k_1 \leq \cdots \leq k_j \leq N} \sum_{\substack{(s_1, \ldots, s_j) \in N^j \\ s_1 + \cdots + s_j = i}} E\left(\frac{\binom{\nu_{k_1}}{s_1} \cdots \binom{\nu_{k_j}}{s_j}}{\binom{N}{i}}\right).$$

The appropriate limiting process is a death process. Throughout this and the next section we denote by $\{D(t); t \geq 0\}$ the simple Markov death process in continuous time with state space $E = \{1, 2, \ldots, \infty\}$ and death rate $l(l - 1)/2$ from l to $l - 1$. Recall that $D_n(\cdot)$ denotes the process $D(\cdot)$ conditional on $D(0) = n$.

THEOREM 2 [Kingman (1982c)]. *Suppose that, as $N \to \infty$, the variance of ν_1 converges to a nonzero limit σ^2 and that*

$$\sup_N E(\nu_1^s) < \infty$$

for all $s \geq 1$. Then the finite dimensional distributions of the process $A_n^N(t) = A_n^N([N\sigma^{-2}t])$ converge to those of $D_n(t)$.

The result is a consequence of Theorem 1 of Kingman (1982c), which concerns a more informative, equivalence–relation-valued, genealogical process. In fact not all the moment conditions are essential to Kingman's proof, although all are satisfied by the familiar genetic models.

Note that because of the quadratic nature of the death rates, the state ∞ is in fact an entrance boundary for the process $D_\infty(\cdot)$. For example $D_\infty(\cdot)$ could be constructed by defining independent exponential random variables $\tau_2, \tau_3, \ldots$ on a common probability space with τ_k having mean $2/k(k-1)$, defining $\tau_1 = \infty$ a.s. and putting $D_\infty(t) = s$ on the set

$$\left\{ \sum_{k=s+1}^{\infty} \tau_k \leq t < \sum_{k=s}^{\infty} \tau_k \right\}.$$

Note also that it is clear from this construction that $D_\infty(\cdot) \in D_E[0, \infty)$.

It is natural to ask about the process $\{A_N^N(t), \ t \geq 0\}$ which counts the number of ancestors of the whole population. In view of Theorem 2 and the behavior of $D_\infty(t)$, a sensible conjecture, implicit in the final section of Kingman (1982c), is that the sequence $A_N^N([N\sigma^{-2} \cdot])$ converges weakly to $D_\infty(\cdot)$.

THEOREM 3. *Suppose that as $N \to \infty$, the variance of ν_1 converges to a nonzero limit σ^2 and that*

$$\sup_N E(\nu_1^s) < \infty$$

for $s = 1, 2, \ldots$. Then for any sequence of positive integers $\{n_N\}_{N=1}^{\infty}$ with $n_N \leq N$ and $n_N \to \infty$ as $N \to \infty$, the sequence of processes $A_{n_N}^N([N\sigma^{-2} \cdot])$ converges weakly to $D_\infty(\cdot)$ as $N \to \infty$. In particular

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

as $N \to \infty$.

PROOF. We will apply Theorem 1 with $B_{n_N}^N(\cdot) = A_{n_N}^N([N\sigma^{-2} \cdot])$. First note that the limiting process $\{D(t); \ t \geq 0\}$ clearly satisfies Theorem 1(i), (ii) and (iii). For $l < n$, an easy coupling argument, founded on the observation that one way of taking a sample of size l is first to take a sample of size n and then to take a subsample of size l from this first sample, guarantees the stochastic monotonicity in (A1):

$$A_l^N([N\sigma^{-2} \cdot]) \overset{st}{\leq} A_n^N([N\sigma^{-2} \cdot]).$$

Condition (A2) is exactly Kingman's finite sample result, Theorem 2. The result will follow then provided (A3) holds. In the case $n_N = N$, this is a consequence of the next proposition and, say, Markov's inequality. The general result then follows from the fact that

$$P\big(A_{n_N}^N([N\sigma^{-2}t]) \le M\big) \ge P\big(A_N^N([N\sigma^{-2}t]) \le M\big),$$

which is itself a consequence of monotonicity. □

PROPOSITION 3.1. *For any $t \ge 0$, there is a finite constant C (depending on t) with $E(A_N^N([N\sigma^{-2}t])) \le C$ for all N.*

We first extend Theorem 2 and bound the expected number of ancestors of a sample whose size increases with the population size N (in fact $N^{1/3}$ will do) and then exploit the structure of the process, and monotonicity, to bound the expected number of ancestors of the whole population. Throughout we will denote by K_N ($N = 2, 3, \ldots$) the smallest integer larger than $N^{1/3}$: $K_N \equiv [N^{1/3}] + 1$, for $N = 2, 3, \ldots$, and put $K_1 = 1$.

LEMMA 1. *For any $t > 0$, there is a finite constant C_1 (depending on t) with $E(A_{K_N}^N([N\sigma^{-2}t])) \le C_1$ for all N.*

PROOF. It follows from Kingman [(1982c), page 103] that for any k and any $M = k, k + 1, \ldots N$,

$$P\big(A_{K_N}^N(t + 1) = k - 1 \big| A_{K_N}^N(t) = k\big)$$

$$\ge \binom{k}{2}\bigg[N^{-1}\operatorname{var}(\nu_1) - (k - 2)N^{-2}E(\nu_1^3) - \binom{k - 2}{2}N^{-2}E(\nu_1^2\nu_2^2)\bigg].$$

Throughout what follows we will have $k \le K_N$ and we will assume that N is large enough to ensure that

$$N^{-1}\operatorname{var}(\nu_1) - (K_N - 2)N^{-2}E(\nu_1^3) + \binom{K_N - 2}{2}N^{-2}E(\nu_1^2\nu_2^2) \ge \tfrac{1}{2}N^{-1}\sigma^2.$$

(Recall the moment conditions in the statement of the theorem.) Then, for $k = 2, 3, \ldots, K_N$,

$$P\big(A_{K_N}^N(t + 1) = k - 1 \big| A_{K_N}^N(t) = k\big) \ge \frac{1}{2}\binom{k}{2}N^{-1}\sigma^2 = \Pi_k^N, \quad \text{say}.$$

Now for $k = 2, 3, \ldots, K_N$, let W_k^N be independent geometric random variables with respective success probabilities Π_k^N. Define

$$T_{K_N, n}^N = \sum_{k=n+1}^{K_N} W_k^N.$$

Elementary calculation then gives

$$E\big(T_{K_N, n}^N\big) \le \frac{4N}{\sigma^2 n} \quad \text{and} \quad \operatorname{var}\big(T_{K_N, n}^N\big) \le \frac{64N^2}{3\sigma^4 n^3}.$$

An easy monotonicity argument shows that

$$P\big(A_{K_N}^N([N\sigma^{-2}t]) > n\big) \le P\big(T_{K_N, n}^N > N\sigma^{-2}t\big)$$

and by Markov's inequality this last is bounded above by

$$\frac{\sigma^4}{N^2 t^2} E\Big(\big(T_{K_N, n}^N\big)^2\Big) \le \frac{64}{3t^2 n^3} + \frac{16}{t^2 n^2}.$$

Thus

$$E\big(A_{K_N}^N([N\sigma^{-2}t])\big) = \sum_{n=1}^{\infty} P\big(A_{K_N}^N([N\sigma^{-2}t]) \ge n\big) \le 1 + t^{-2}C' < \infty,$$

where C' is a constant which does not depend on N, and the result follows. $\square$

LEMMA 2.   *For any n, $n = 1, 2, \ldots, N$, and any t,*

$$E\big(A_n^N(t)\big) \le (nN^{-1/3} + 1)E\big(A_{K_N}^N(t)\big).$$

PROOF.   For $n \le K_N$, monotonicity means that $A_n^N(t)$ is stochastically smaller than $A_{K_N}^N(t)$ and the result is immediate.

For $n > K_N$, partition the sample of size n into j ($\le nN^{-1/3} + 1$) disjoint subsamples of size at most K_N, and denote by $A_{(1)}^N, \ldots, A_{(J)}^N$ the ancestral process corresponding to each of the J subsamples, and for convenience suppose that $A_{(1)}^N$ corresponds to the largest of these subsamples. These processes are not independent, but they have identical transition mechanisms and would be identically distributed were it not for the fact that the subsamples may be of different sizes. A coupling argument, founded on the observation that an ancestor, t generations into the past, of any of the original sample of n must be an ancestor of at least one individual in at least one of the subgroups, gives

$$A_n^N(t) \overset{st}{\le} A_{(1)}^N(t) + A_{(2)}^N(t) + \cdots + A_{(J)}^N(t).$$

Thus

$$E\big(A_n^N(t)\big) \le JE\big(A_{(1)}^N(t)\big) \le (nN^{-1/3} + 1)E\big(A_{K_N}^N(t)\big)$$

as required, where we have again exploited the stochastic monotonicity of $A_k^N(t)$ in k. $\square$

Now for $0 \le u < s$, let $\mathscr{F}_t$ denote the σ-algebra generated by $\{A_N^N(u), u = 0, 1, \ldots, [t]\}$. Note that for all s and u,

$$[N\sigma^{-2}s] - [N\sigma^{-2}u] \ge [N\sigma^{-2}(s - u)],$$

so that as a consequence of the fact that $A_N^N(t)$ is stochastically nonincreasing

in t, we may write

$$E\big(A_N^N([N\sigma^{-2}s])\big) = E\big(E\big(A_N^N([N\sigma^{-2}s])|\mathscr{F}_{N\sigma^{-2}u}\big)\big)$$

$$= E\Big(E\big(A_{A_N^N([N\sigma^{-2}u])}^N([N\sigma^{-2}s] - [N\sigma^{-2}u])\big)\Big)$$

$$\le E\Big(E\big(A_{A_N^N([N\sigma^{-2}u])}^N([N\sigma^{-2}(s-u)])\big)\Big)$$

$$\le E\big((N^{-1/3}A_N^N([N\sigma^{-2}u]) + 1)$$

$$\times E\big(A_{K_N}^N([N\sigma^{-2}(s-u)])\big)\big)\quad \text{(by Lemma 2)}$$

$$\le C_1\big(N^{-1/3}E\big(A_N^N([N\sigma^{-2}u])\big) + 1\big)$$

for some constant $C_1 < \infty$, depending on $s - u$ but not on N, by Lemma 1. The proof of Proposition 3.1 is completed by applying this last inequality three times, first with $s = t$, $u = t_2$, then with $s = t_2$, $u = t_1$, and finally with $s = t_1$, $u = 0$, for some $t_1, t_2 \in (0, t)$ with $t_1 < t_2$. $\square$

4. Convergence of absorption times. A central technique in mathematical population genetics has been the approximation of suitably rescaled discrete models (evolving forward through time) by more tractable diffusion processes, with generator techniques predominating in the proof of appropriate limit theorems [see, e.g., Ethier and Kurtz (1986), Chapter 10]. It was shown in Donnelly (1985) that weak convergence to the so-called Wright-Fisher diffusion is a consequence of Kingman's finite sample result (Theorem 2), and in this way the class of models known to converge was extended to include all of those to which that result applies. A more delicate question, in the case of no mutation, concerns the weak convergence of absorption times.

Consider a population evolving according to one of the exchangeable models satisfying the conditions of Theorem 2, and suppose that there are two genetic types, A and a, say, and that offspring always adopt the type of their parent. (The restriction to only two types is one of convenience. The ensuing analysis will apply to any K-allele model without mutation.) Suppose also that in some initial generation, which we call 0, types are assigned to individuals independently, in such a way that each individual is of type A with probability p, $0 < p < 1$. Denote by $Y(t)$ the number of individuals of type A in generation t. It is immediate that $\{Y(t), t \ge 0\}$ is a Markov chain on $\{0, 1, \ldots, N\}$ for which 0 and N are absorbing states. To make the dependence of N explicit we will write $Y^N(t)$ for $Y(t)$.

The limiting process $\{Z(t), t \ge 0\}$ is the (so-called Wright–Fisher) diffusion on $[0, 1]$ with $Z(0) = p$, absorbing barriers at 0 and 1 and generator

$$\frac{1}{2}x(1-x)\frac{\partial^2}{\partial x^2}.$$

The convergence result of Donnelly (1985) is:

THEOREM 4. *If the Markov chain* $\{Y^N(t), \ t = 0, 1, 2, \ldots\}$ *arises from one of the exchangeable reproductive models satisfying the conditions of Theorem 2, then as* $N \to \infty$, $N^{-1}Y^N([N\sigma^{-2} \cdot]) \Rightarrow Z(\cdot)$.

Define $\xi_0, \xi_1 \colon D_{[0,1]}[0, \infty] \to [0, \infty]$ by
$$\xi_0(x) = \inf\{t \ge 0 \colon x(t) = 0 \text{ or } x(t-) = 0\},$$
$$\xi_1(x) = \inf\{t \ge 0 \colon x(t) = 1 \text{ or } x(t-) = 1\},$$
where $\inf \phi = \infty$. Then ξ_0 and ξ_1 are Borel measurable, so we can define the absorption times
$$S_i^N = \xi_i\big(N^{-1}Y^N([N\sigma^{-2} \cdot])\big) \quad \text{and} \quad S_i = \xi_i(Z) \quad \text{for } i = 0, 1.$$

We now prove the equivalence of the weak convergence of these absorption times and the weak convergence of the marginal distribution of the ancestral process. Note that for the next theorem we are not assuming the conditions of Theorem 2.

THEOREM 5. *As* $N \to \infty$, $S_0^N \Rightarrow S_0$ *and* $S_1^N \Rightarrow S_1$ *for all* $p \in (0, 1)$ *iff* $A_N^N([N\sigma^{-2}t]) \Rightarrow D_\infty(t)$ *for each* $t \ge 0$.

PROOF. The process $Y^N(t)$ will take the value N iff all the ancestors in generation 0 of the individuals in generation t are of type A. Condition on the number of ancestors and invoke the independence of initial types to write

$$(2) \qquad P\big(S_1^N \le t\big) = \sum_{k=1}^{N} P\big(A_N^N([N\sigma^{-2}t]) = k\big)p^k.$$

The distribution of S_1 is known [see, e.g., Ewens (1979), page 141] and in fact can be written as

$$(3) \qquad P(S_1 \le t) = \sum_{k=1}^{\infty} P(D_\infty(t) = k)p^k.$$

Now each $t \ge 0$ is a continuity point of the (improper) distribution of S_1, so writing $p = e^{-\lambda}$ for $\lambda > 0$ in (2) and (3), we have $S_1^N \Rightarrow S_1$ for all $p \in (0, 1)$ iff $\Phi_t^N(\lambda) \to \Phi_t(\lambda)$ for each $t \ge 0$ and each $\lambda > 0$, where

$$\Phi_t^N(\lambda) = E\big(\exp\big(-\lambda A_N^N([N\sigma^{-2}t])\big)\big) \quad \text{and} \quad \Phi_t(\lambda) = E\big(\exp(-\lambda D_\infty(t))\big)$$

are the Laplace transforms of the distributions of $A_N^N([N\sigma^{-2}t])$ and $D_\infty(t)$, respectively. The result follows from the continuity theorem of, for example, Feller [(1971), Theorem XIII.1.2)] and the observation that by symmetry $S_1^N \Rightarrow S_1$ for all $p \in (0, 1)$ iff $S_0^N \to S_0$ for all $p \in (0, 1)$. □

REMARKS.

1. It is perhaps worth noting that $S_1^N \Rightarrow S_1$ does not follow from Theorem 4. The function ξ_1 is discontinuous at every $x \in D_{[0,1]}[0, \infty]$ for which $\xi_1(x) < \infty$ and it is known that $P(S_1 < \infty) > 0$. [In fact $P(S_0 \wedge S_1 < \infty) = 1$.]

2. Theorems 3 and 5 now guarantee the convergence of absorption times for any of the models satisfying the moment conditions of Theorem 2. This result was known for the specific case of the discrete Wright–Fisher model [see Ethier (1979) and references therein]. Even in this case the direct proof is formidable (although the results apply to more general genetic models, in particular in allowing selection). Griffiths (1981) used this result and an argument similar to the one above to prove Theorem 3 (or at least weak convergence of one dimensional distributions) for the Wright–Fisher model. The novelty of the current approach is that diffusion approximations and approximations for absorption times follow readily from genealogical approximations, and these in turn require few model specific calculations.

3. The observation at the start of the above proof is *exactly* analogous to the duality relationship for voter models. See, for example, Durrett [(1981), equation (3.7)]. In fact the connection has recently been shown to be rather more intimate [Cox and Griffeath (1986)]. One interpretation of this is that with suitable spatial and temporal rescaling the two dimensional voter model shares the dual process $D_\infty(\cdot)$ with the genetic models.

4. The assumption of a binomial distribution for $Y^N(0)$ is standard in the genetics context. More generally, denote by $f_N(k)$, $k = 1, 2, \ldots, N$, the probability that a random sample of size k taken from the population in its initial generation will consist entirely of individuals of type A:

$$f_N(k) = \binom{N}{k}^{-1} E\left(\binom{Y^N(0)}{k}\right), \qquad k = 1, 2, \ldots, N,$$

and put $f_N(k) = 0$, $k \in E \setminus \{1, 2, \ldots N\}$. Fix $p \in (0, 1)$ and define $f: E \to R$ by

$$f(k) = p^k, \qquad k = 1, 2, \ldots$$

and $f(\infty) = 0$. If $A_N^N([N\sigma^{-2}t]) \Rightarrow D_\infty(t)$ for each $t \geq 0$ (and in particular if Theorem 3 obtains) then the continuity theorem [Billingsley (1968)], the topology of E and the fact that for $t > 0$, $P(D_\infty(t) = \infty) = 0$, guarantee that as $N \to \infty$,

$$P\left(S_1^N \leq t\right) = E\left(f_N\left(A_N^N[N\sigma^{-2}t]\right)\right) \to E\left(f(D_\infty(t))\right) = P(S_1 \leq t)$$

for each t, whenever $f_N(k) \to f(k)$ for each $k = 1, 2, 3, \ldots$.

5. Lines of descent. When genetic models incorporate the effects of mutation it is often more useful to count "lines of descent" than numbers of ancestors. In the interests of brevity we refer the interested reader to Griffiths (1980) or Tavaré (1984) for motivation and exact definitions. Loosely speaking the number of lines of descent of a sample is the number of ancestors without intervening mutation. We will denote by $R_n^N(\cdot)$ the Markov chain which counts the number of lines of descent of a sample of size n from a population of size N. Again, finite sample behavior is well understood. Let $\{\breve{D}(t); t \geq 0\}$ denote the death process with state space $\breve{E} = \{0, 1, 2, \ldots, \infty\}$ and death rates $\frac{1}{2}l(l + \theta - 1)$ from l to $l - 1$. (Here $\theta > 0$ is a mutation parameter.) It follows from the main result of Donnelly and Tavaré (1986) that as $N \to \infty$,

$B_n^N([N\sigma^{-2} \cdot])$ converges weakly to $\tilde{D}_n(\cdot)$, but the problem concerning population lines of descent has remained open.

THEOREM 6. *Suppose that the population is evolving according to one of the exchangeable models of Section 3 and that the finite sample result of Donnelly and Tavaré (1986) obtains. (This requires the moment conditions of Theorem 2 and a condition on the mutation rates.) Let $\{n_N\}_{N=1}^{\infty}$ be a sequence of positive integers with $n_N \leq N$ for each N and $\lim_{N \to \infty} n_N = \infty$. Then as $N \to \infty$ the sequence of processes $B_{n_N}^N([N\sigma^{-2} \cdot])$ converges weakly to $\tilde{D}_{\infty}(\cdot)$ in $D_{\bar{E}}[0, \infty]$.*

PROOF.   Again apply Theorem 1. Stochastic monotonicity of $B_n^N(\cdot)$ in n follows as before, and Proposition 3.1 and the (immediate) observation that for $n_N \leq N$ and any t,

$$B_{n_N}^N(t) \overset{st}{\leq} A_N^N(t)$$

ensure condition (A3).   □

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 1. Haploid models. *Adv. Appl. Probab.* **6** 260–290.

COX, J. T. and GRIFFEATH, D. (1986). Diffusive clustering in the two dimensional voter model. *Ann. Probab.* **14** 347–370.

DONNELLY, P. (1985). Dual processes and an invariance result for exchangeable models in population genetics. *J. Math. Biol.* **23** 103–118.

DONNELLY, P. and TAVARÉ, S. (1986). The ages of alleles and a coalescent. *Adv. 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** 281–391.

DURRETT, R. (1981). An introduction to infinite particle systems. *Stochastic Processes Appl.* **11** 109–150.

ETHIER, S. N. (1979). Limit theorems for absorption times of genetic models. *Ann. Probab.* **7** 622–638.

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

EWENS, W. (1979). *Mathematical Populations Genetics.* Springer, New York.

FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

GRIFFITHS, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright–Fisher models. *Theor. Pop. Biol.* **17** 35–50.

GRIFFITHS, R. C. (1981). Transient distribution of the number of segregating sites in a neutral infinite-sites model with no recombination. *J. Appl. Probab.* **18** 42–51.

JACOD, J., MEMIN, J. and METEVIER, M. (1983). On tightness and stopping times. *Stochastic Process. Appl.* **14** 109–146.

KAMAE, T., KRENGEL, U. and O'BRIEN, G. L. (1977). Stochastic inequalities on partially ordered spaces. *Ann. Probab.* **5** 899–912.

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.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

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

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