J. Appl. Prob. **40,** 33–48 (2003)
Printed in Israel

THE COALESCENT PROCESS IN A POPULATION WITH STOCHASTICALLY VARYING SIZE

INGEMAR KAJ,* *Uppsala University*
STEPHEN M. KRONE,** *University of Idaho*

Abstract

We study the genealogical structure of a population with stochastically fluctuating size. If such fluctuations, after suitable rescaling, can be approximated by a nice continuous-time process, we prove weak convergence in the Skorokhod topology of the scaled ancestral process to a stochastic time change of Kingman's coalescent, the time change being given by an additive functional of the limiting backward size process.

Keywords: Coalescent; stochastically fluctuating population size; weak convergence; stochastic time change

AMS 2000 Subject Classification: Primary 92D10; 60J70; 60F05
Secondary 92D25

1. Introduction

We investigate the genealogical structure of a neutral, haploid population with stochastically fluctuating size. If such fluctuations, after suitable rescaling, can be approximated by a nice continuous-time process (for example, a one-dimensional diffusion), then we are led to genealogies given by Kingman's coalescent run on a nonlinear, stochastic time-scale. More explicitly, our main result provides weak convergence in the Skorokhod topology of the scaled ancestral process to a stochastic time change of Kingman's coalescent, the time change being given by an additive functional of the limiting backward size process.

To date, there have been a number of papers in the literature dealing with coalescent theory in the presence of deterministically varying population sizes. Perhaps the most popular of these is the case in which the population size is assumed to be large and growing (forward in time) exponentially fast. Mathematically, the deterministic size model is much easier to handle. If we have a stochastic size model but condition on knowing the sizes, the deterministic case can be applied provided that we have a neutral model. The case of stochastic sizes presented here is fundamentally different. We do not condition on knowing the size process and hence we keep the full stochasticity in the limiting coalescent. Indeed, assuming that the sequence of past population sizes is known removes an important source of randomness from the coalescent.

The results in this paper can be thought of as another manifestation of the robustness of the coalescent. The standard coalescent is the same for all exchangeable models which have the same diffusion limit as that of the Wright–Fisher model. Our results show that, if we introduce stochastic fluctuations in population size, then all models whose (scaled) backward size process converges to the same process will have the same coalescent.

Received 18 July 2002; revision received 22 November 2002.
* Postal address: Department of Mathematics, Uppsala University, S-751 06 Uppsala, Sweden.
** Postal address: Department of Mathematics, University of Idaho, Moscow, ID 83844-1103, USA.
Email address: krone@uidaho.edu

A related result on coalescents with stochastically varying size can be found in the paper by Donnelly and Kurtz (1999). Their approach is quite different from ours in several ways. First of all, their result is written in terms of the forward-size process and involves quantities and assumptions that are perhaps not the most natural to most people working in population genetics. Their paper is remarkable in its generality, the coalescent result being but one of many interesting theorems embedded in their 'look-down' construction. The cost is the heavy machinery that must be developed beforehand. We will take a more direct, and we hope more transparent, approach, our proof relying on fairly standard machinery in the weak convergence arena.

1.1. The population-size process

Assume that we have a single neutral, haploid population whose size r generations in the past is denoted by $M_N(r)$. Assume further that this backward population-size process is a Markov chain and write

$$M_N(r) = NX_N(r),$$

where N is a parameter that we will take to be large and $X_N(\cdot)$ is the 'relative size' process. Thus, as is typical in coalescent theory, population sizes in the unscaled process M_N are of order N for large N. Finally, we assume that the process $\{M_N(r) : r = 0, 1, \dots\}$, after a suitable rescaling, converges weakly to a continuous-time Markov process $\{X(t) : t \in [0, \infty)\}$ with state space S, an interval contained in $(0, \infty)$:

$$N^{-1}M_N(\lfloor N \times \cdot\rfloor) = X_N(\lfloor N \times \cdot\rfloor) \Rightarrow X(\cdot), \qquad N \to \infty.$$

Here, $\Rightarrow$ denotes weak convergence in the Skorokhod space $D_S[0, \infty)$ of right-continuous functions with left limits and values in S and $\lfloor\cdot\rfloor$ denotes the integer-part function. The state spaces for the above processes are thus given by

$$\begin{aligned}
&X(t) \in S,\\
&X_N(r) \in S_N \equiv S \cap \mathbb{Z}_N, \qquad \text{where } \mathbb{Z}_N := \frac{1}{N}\mathbb{Z},\\
&M_N(r) \in NS_N.
\end{aligned}$$

To simplify notation, let

$$\tilde{X}_N(t) := X_N(\lfloor Nt\rfloor) = N^{-1}M_N(\lfloor Nt\rfloor)$$

be the right-continuous piecewise-flat extension of the backward population-size process, properly renormalized. The functions $\tilde{X}_N(\cdot)$ are elements in the path space $D = D_S[0, \infty)$. We equip D with the Skorokhod topology of weak convergence. Denote by $\mathrm{C_b}(S)$ the space of bounded continuous functions $f : S \to \mathbb{R}$.

The transition semigroups of the relevant backward size processes are defined on $f \in \mathrm{C_b}(S)$ by

$$\begin{aligned}
T_t^N f(x) &= \mathrm{E}[f(X_N(\lfloor Nt\rfloor)) \mid X_N(0) = x]\\
&= \mathrm{E}[f(\tilde{X}_N(t)) \mid \tilde{X}_N(0) = x], \qquad x \in S_N,
\end{aligned}$$

and

$$T_t f(x) = \mathrm{E}[f(X(t)) \mid X(0) = x], \qquad x \in S.$$

We assume that the latter semigroup satisfies the Feller property.

Our main assumption regarding convergence of the backward-size processes is the following.

Assumption 1. *For all* $f \in \mathrm{C_b}(S)$,

$$\sup_{x \in S_N} |\,\mathrm{E}^x f(\tilde{X}_N(t)) - \mathrm{E}^x f(X(t))| \to 0 \quad \text{as } N \to \infty.$$

Because of the Feller property of the limit semigroup, Assumption 1 is equivalent to

$$T_t^N f(x_N) \to T_t f(x) \quad \text{as } N \to \infty, x_N \to x,$$

uniformly in x for all $f \in \mathrm{C_b}(S)$. Letting L denote the generator of X and $L_N := N(T_N - I)$ the corresponding discrete generator, the above is also equivalent to $L_N f(x_N) \to Lf(x)$ uniformly in x as $N \to \infty$ and $x_N \to x$. Here, $T_N f(x) \equiv \sum_{y \in S_N} p(x, y) f(y)$, with $p(x, y)$ denoting the one-step transition probability for the process X_N. The above equivalent conditions are slightly stronger than just weak convergence, but are typically satisfied by approximations arising in applications; cf. Ethier and Kurtz (1986, p. 415) for an example.

1.2. The ancestral process

In coalescent theory, the number of ancestors (i.e. ancestral lineages) is followed back in time starting from a sample in the current generation. Let $A_N(\cdot)$ be the ancestral process defined by $A_N(0) = n$ and, for $r \geq 1$,

$$A_N(r) = \text{number of distinct ancestors } r \text{ generations in the past},$$

where n is the original sample size and the index N indicates the dependence on the underlying population size M_N. We will be interested in the limiting behaviour of the continuous-time, scaled ancestral process $A_N(\lfloor Nt \rfloor),\ t \geq 0$.

The dynamics of A_N are determined by the coalescence probabilities as we move back in time one generation at a time. A key role will be played by a function H_N, essentially describing pairwise coalescence rates, according to the following assumption.

Assumption 2. *If* $M_N(r-1) = k$ *and* $M_N(r) = m$, *then the coalescence probability for a randomly chosen pair of ancestral lineages at time* r *is of the form*

$$c_N(k, m) = \frac{1}{N} H_N\left(\frac{k}{N}, \frac{m}{N}\right).$$

It is further assumed that there is a nonnegative, bounded, jointly continuous function $H(x, y)$ *on* $S \times S$ *such that*

$$H_N\left(\frac{k}{N}, \frac{m}{N}\right) \to H(x, y),$$

uniformly in $S \times S$, *as* $N \to \infty$, $k/N \to x$ *and* $m/N \to y$. *Finally, we assume that, for any finite number* i *of lineages, the event that more than two are involved in a coalescence event in a single generation has probability* $\mathcal{O}(1/N^2)$, *and the probability of exactly one pair coalescing is*

$$\binom{i}{2} c_N(k, m) + \mathcal{O}\left(\frac{1}{N^2}\right).$$

For models in which parents are 'chosen' independently (e.g. Wright–Fisher), the last condition follows from the first. We remark that the $\mathcal{O}(1/N^2)$ terms can be replaced throughout the paper by terms of order $o(1/N)$. In addition to making it more clear where these terms are coming from, the reason for our choice is that, in the most common examples, the $o(1/N)$ terms are of the form $\mathcal{O}(1/N^2)$.

To see that these conditions are natural, consider first the Wright–Fisher model with fluctuating population size. If we consider a sample of two randomly chosen lineages $r-1$ generations in the past, the probability that they coalesce (i.e. have the same parent) in generation r is given by

$$c_N(M_N(r-1), M_N(r)) := \frac{1}{M_N(r)} = \frac{1}{NX_N(r)}. \tag{1}$$

Thus, we can write

$$c_N(k, m) = \frac{1}{m} = \frac{1}{N} H_N\left(\frac{k}{N}, \frac{m}{N}\right),$$

where $H_N(x, y) = 1/y = H(x, y)$. Multiple coalescence events have probability $\mathcal{O}(1/N^2)$. Note that H will be bounded and jointly continuous, and the convergence of H_N to H will be uniform when, for example, $S \subseteq [a, \infty)$, with $a > 0$.

The above development can be generalized to a large class of reproduction models that includes the Wright–Fisher model as a special case. Let $\nu_i^{(r)}$ denote the random number of offspring produced by the ith individual in generation r (in the past). Thus,

$$\sum_{i=1}^{M_N(r)} \nu_i^{(r)} = M_N(r-1).$$

Assume that the offspring vectors $(\nu_1^{(r)}, \ldots, \nu_{M_N(r)}^{(r)})$ are independent from generation to generation and, conditioned on the offspring numbers, all assignments of offspring to parents which honour these numbers are equally likely. Such models have been studied, in the setting of deterministic population-size processes, by Möhle (2000), (2002), Donnelly (1986), Griffiths and Tavaré (1994) and others. It can be shown that the pairwise coalescence probability in generation r is given by

$$c_N(M_N(r-1), M_N(r)) = \frac{1}{(M_N(r-1))_2} \sum_{i=1}^{M_N(r)} \mathrm{E}[(\nu_i^{(r)})_2], \tag{2}$$

where $(k)_2 := k(k-1)$. (This expression reduces to (1) in the special case of Wright–Fisher dynamics.) If the distributions of offspring numbers do not change with time, we have the following expression, independent of r:

$$\begin{aligned} c_N(k, m) &= \frac{1}{k(k-1)} \sum_{i=1}^{m} \mathrm{E}[\nu_i(\nu_i - 1)] \\ &= \left(N^2 \frac{k}{N}\left(\frac{k}{N} - \frac{1}{N}\right)\right)^{-1} \cdot \sum_{i=1}^{N \cdot m/N} d(i), \end{aligned}$$

where $d(i) := \mathrm{E}[\nu_i(\nu_i - 1)]$. So we can write the coalescence probability in the form of Assumption 2 by setting

$$H_N\left(\frac{k}{N}, \frac{m}{N}\right) = \left(\frac{k}{N}\left(\frac{k}{N} - \frac{1}{N}\right)\right)^{-1} \cdot \frac{1}{N} \sum_{i=1}^{N \cdot m/N} d(i).$$

Under mild conditions, this will converge, as $k/N \to x$ and $m/N \to y$, to a function of the form

$$H(x, y) = \frac{\bar{d}y}{x^2},$$

where $\bar{d}$ is a constant. Note that the special case of exchangeable reproduction would yield $d(i) = d$ for all i, and hence

$$H_N\left(\frac{k}{N}, \frac{m}{N}\right) = \left(\frac{k}{N}\left(\frac{k}{N} - \frac{1}{N}\right)\right)^{-1} \cdot \frac{md}{N} \to \frac{yd}{x^2}.$$

Again, we can arrange for H to be bounded and jointly continuous, and the convergence of H_N to H to be uniform if, for example, the interval S is of the form $[a, b]$, with $0 < a < b < \infty$. This would give the first part of Assumption 2. To prevent multiple coalescences in the limiting coalescent, we need a further condition; i.e. this does not follow from (2) alone. Such a condition can be found in Möhle (2000).

Remark 1. We note that, in the general setting of (2), the pairwise coalescence probabilities can depend on the time r through more than just the dependence on $M_N(r-1)$ and $M_N(r)$. In such a situation, we have functions of the form $c_{N,r}(k, m)$ and $H_{N,r}$ and we would need convergence of the form $H_{N,\lfloor Nt \rfloor}(k/N, m/N) \to H_t(x, y)$ for some limiting coalescence rate that depends on time explicitly and not just the population sizes. We will not treat this case in the present paper.

One final comment about Assumption 2 is that, as we will see below, we really only need the conditions to hold near the diagonal $H(x, x)$. The key is that we get the convergence discussed next.

1.3. Coalescence intensity

To see what the limiting coalescent should look like, we argue heuristically as follows. Intuitively, the rate of coalescence for a pair of lineages is obtained by adding pairwise coalescence probabilities over $\lfloor Nt \rfloor$ generations and taking the limit as $N \to \infty$. Thus, if we define the *cumulative coalescence intensity* over k generations by

$$Y_N(k) := \sum_{r=1}^{k} c_N(M_N(r-1), M_N(r)),$$

the limiting cumulative coalescence intensity over the continuous-time interval $[0, t]$ grows like

$$Y_N(\lfloor Nt \rfloor) \Rightarrow \int_0^t H(X(s-), X(s))\,\mathrm{d}s = \int_0^t H(X(s), X(s))\,\mathrm{d}s, \tag{3}$$

where the last equality follows from the fact that a D-valued process has at most a countable set of discontinuities. To demonstrate the weak convergence in (3), write

$$\sum_{r=1}^{\lfloor Nt \rfloor} c_N(M_N(r-1), M_N(r)) = \int_0^{\lfloor Nt \rfloor/N} H_N(\tilde{X}_N(s - 1/N), \tilde{X}_N(s))\,\mathrm{d}s.$$

Now, $\tilde{X}_N \Rightarrow X$ in $D_S[0, \infty)$, and so the desired result follows from the continuity of H and the fact that $H_N \to H$ uniformly (cf. Billingsley (1968)). Hence

$$Y_N((\lfloor N \times \cdot \rfloor) \Rightarrow \int_0^{\cdot} H(X(s-), X(s))\,\mathrm{d}s.$$

Because of the limit in (3), it seems reasonable that, if $A(t)$ denotes the standard Kingman (1982) coalescent, then the coalescent for the above stochastically varying population-size model should be given by $A(Y_t)$, where

$$Y_t = \int_0^t H(X(s), X(s))\,\mathrm{d}s,$$

i.e. the Kingman coalescent run according to the clock $t \mapsto Y_t$.

We assume that the limit process X satisfies

$$\int_0^\infty H(X(s), X(s))\,\mathrm{d}s = \infty,$$

so that all ancestral lineages are guaranteed to coalesce in finite time. For simplicity, we also assume that

$$\int_0^t H(X(s), X(s))\,\mathrm{d}s < \infty$$

for each $t \in (0, \infty)$. If this were not the case, we would just run the process up to the stopping time $\tau := \inf\{t : Y_t = \infty\}$; all coalescences will have occurred by then. Note that the two conditions above are trivially satisfied in the Wright–Fisher example when the interval S is of the form $[a, b]$, with $0 < a < b < \infty$.

In the case where X is a regular one-dimensional diffusion (cf. Rogers and Williams (1987)), the last integral in (3) can be expressed as:

$$\int_0^t H(X(s), X(s))\,\mathrm{d}s = \int_S H(x, x) L_t^x m(\mathrm{d}x), \tag{4}$$

where m is the speed measure for X and $\{L_t^x : x \in S, t \geq 0\}$ is its local time. This comes from the fact that $L_t^x m(\mathrm{d}x)$ serves as an occupation measure for the diffusion. Thus, the expression $\int_S H(x, x) L_t^x m(\mathrm{d}x)$ is essentially a sum of coalescence intensities weighted by the amount of time in $[0, t]$ that this intensity was in force.

According to (4), the time change can be written in the form

$$\int_0^t H(X(s), X(s))\,\mathrm{d}s = \int_S \frac{1}{x} \cdot x H(x, x) L_t^x m(\mathrm{d}x),$$

with the usual Wright–Fisher term, $1/x$, weighted by the factor $xH(x, x)$. An interpretation of this expression is that, when the actual population size is given by xN, the population acts like a Wright–Fisher model with 'instantaneous effective population size' given by $N_{\mathrm{e}}(x) = N/xH(x, x)$. This kind of coalescent effective size is discussed in Nordborg and Krone (2002) for cases in which the time change is linear; in such cases, the usual $1/x$ term becomes c/x and the effective size is N/c. Of course, one of the main points of this paper is that the models considered here do not have population sizes that average out on the coalescent time-scale, and hence there is no linear scaling of time that would give an actual effective population size.

2. Convergence theorem

The main result of the paper follows. In addition to giving the advertised convergence of scaled ancestral processes to a time-changed version of Kingman's coalescent, we have joint convergence of the backward size process and ancestral process. Recall that $A(t)$ is the standard

Kingman coalescent and L is the generator of the limiting backward size process. Denote the usual decreasing and increasing factorials by

$$(i)_\ell = i(i-1)\cdots(i-\ell+1),$$
$$i_{(\ell)} = i(i+1)\cdots(i+\ell-1).$$

Theorem 1. *Suppose that Assumptions 1 and 2 hold and set*

$$Y_t = \int_0^t H(X(s), X(s))\,\mathrm{d}s.$$

Then, as $N \to \infty$,

$$(X_N(\lfloor Nt \rfloor), A_N(\lfloor Nt \rfloor)) \Rightarrow (X(t), A(Y_t))$$

holds in the sense of weak convergence in the space $D_{S\times\{1,\dots,n\}}[0,\infty)$ whenever $X_N(0) \Rightarrow X(0)$ in S. The transition semigroup $(\mathcal{T}_t)$ for the limit process, defined by $\mathcal{T}_t f(x,i) = \mathrm{E}^{(x,i)}[f(X(t), A(Y_t))]$, can be decomposed as

$$\mathcal{T}_t f(x,i) = \sum_{j=1}^{i}\sum_{\ell=j}^{i} C_\ell(i,j)\,\mathrm{E}^{(x,i)}\left[f(X(t),j)\exp\left(-\binom{\ell}{2}Y_t\right)\right],$$

where, for ℓ with $j \le \ell \le i$,

$$C_\ell(i,j) := \prod_{j+1\le s\le i}\binom{s}{2}\prod_{j\le r\le i, r\ne\ell}\frac{1}{\binom{r}{2}-\binom{\ell}{2}} \tag{5}$$
$$= \frac{(2\ell-1)(-1)^{\ell-j}\,j_{(\ell-1)}(i)_\ell}{j!\,(\ell-j)!\,i_{(\ell)}}. \tag{6}$$

In (5), empty products are defined to have value 1. *The generator of the limit process is given by*

$$\mathcal{L}f(x,i) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{T}_t f(x,i)\Big|_{t=0}$$
$$= Lf(x,i) + \binom{i}{2}H(x,x)(f(x,i-1) - f(x,i)).$$

To explain the strategy of the proof, for $1 \le i \le n$, $x \in \mathbb{Z}_N$ and $r \ge 0$, let

$$\mathcal{T}_r^N f(x,i) = \mathrm{E}^{(x,i)}[f(X_N(r), A_N(r))], \qquad f \in \mathrm{C_b}(\mathbb{Z}_N \times \{1,\dots,n\})$$

be the transition operator for the vector-valued process (X_N, A_N). (Note that our notation is set up so that the symbol $\mathcal{T}$ is used for operators corresponding to the joint process and the symbol T is for operators corresponding to the backward size process alone.) We are going to show that $\mathcal{T}_t$ is a Feller semigroup on $\mathrm{C_b}(\mathbb{Z}_N \times \{1,\dots,n\})$ such that, for each $f \in \mathrm{C_b}(\mathbb{Z}_N \times \{1,\dots,n\})$ and $t \ge 0$,

$$\sup_{x,i}|\mathcal{T}_{\lfloor Nt\rfloor}^N f(x,i) - \mathcal{T}_t f(x,i)| \to 0, \qquad N \to \infty, \tag{7}$$

where the supremum is over $x \in S_N$ and $i \ge j$. Then the weak convergence $(X_N(\lfloor N\times\cdot\rfloor), A_N(\lfloor N\times\cdot\rfloor)) \Rightarrow (X(\cdot), A(Y_\cdot))$ follows by applying Theorem 4.2.12 of Ethier and Kurtz (1986). The remaining statements in Theorem 1 will appear in the process of proving (7).

3. Proof of the theorem

The approach we take here in proving the convergence result for the ancestral process is in many ways similar to that of Kaj *et al.* (2001). In both cases semigroup convergence is used, as in (7), and explicit representations for the r-step transition probabilities of the underlying Markov chain are derived; cf. Lemma 1 below.

3.1. Properties of the discrete-time process

To prepare for the proof of (7) we first study the Markov chain (M_N, A_N) for fixed N. Let

$$P(k,m) = \mathrm{P}(M_N(r) = m \mid M_N(r-1) = k)$$

be the transition probabilities and let $\boldsymbol{P} = (P(k,m))$ be the transition probability matrix for the discrete-time Markov chain M_N. The one-step transition probabilities for the process $(M_N(r), A_N(r))_{r\geq 0}$ are given by

$$\mathrm{P}^{(k,i)}((M_N(1), A_N(1)) = (m, i-1)) = P(k,m)\binom{i}{2}c_N(k,m) + \mathcal{O}\left(\frac{1}{N^2}\right)$$

and

$$\mathrm{P}^{(k,i)}((M_N(1), A_N(1)) = (m, i)) = P(k,m)\left(1 - \binom{i}{2}c_N(k,m) + \mathcal{O}\left(\frac{1}{N^2}\right)\right),$$

with $c_N(k,m)$ as given in Assumption 2, and all other transitions involving multiple coalescences having probabilities of order $\mathcal{O}(1/N^2)$. In addition, we let $\hat{\boldsymbol{P}} = (\hat{P}(k,m))$ be the matrix with elements

$$\hat{P}(k,m) = P(k,m)c_N(k,m).$$

We have suppressed the dependence on N in the matrices $\boldsymbol{P}$ and $\hat{\boldsymbol{P}}$ to simplify the presentation of the next lemma and its proof. These matrices are of size $K \times K$ where $K = |S_N| < \infty$.

The transition probability matrix for $(M_N(\cdot), A_N(\cdot))$ is conveniently represented as a block matrix in the form $\boldsymbol{\Pi} + \mathcal{O}(1/N^2)$, where

$$\boldsymbol{\Pi} = \boldsymbol{\Pi}_N = \begin{pmatrix} \boldsymbol{P} & \boldsymbol{0} & \cdots & \cdots & \boldsymbol{0} \\ \binom{2}{2}\hat{\boldsymbol{P}} & \boldsymbol{P} - \binom{2}{2}\hat{\boldsymbol{P}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \binom{n-1}{2}\hat{\boldsymbol{P}} & \boldsymbol{P} - \binom{n-1}{2}\hat{\boldsymbol{P}} & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \binom{n}{2}\hat{\boldsymbol{P}} & \boldsymbol{P} - \binom{n}{2}\hat{\boldsymbol{P}} \end{pmatrix}.$$

Here, the blocks $\boldsymbol{\Pi}_{i,j}$ are zero except for the diagonal and subdiagonal blocks, and, when $i \geq 2$,

$$\boldsymbol{\Pi}_{i,i} = \boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}, \qquad \boldsymbol{\Pi}_{i,i-1} = \binom{i}{2}\hat{\boldsymbol{P}}.$$

Lemma 1. *For any $t \geq 0$ fixed and for i, j with $1 \leq j \leq i \leq n$, as $N \to \infty$,*

$$\begin{aligned}
&\mathrm{P}^{(k,i)}((M_N(\lfloor Nt \rfloor), A_N(\lfloor Nt \rfloor)) = (m, j)) \\
&\quad = \sum_{\ell=j}^{i} C_\ell(i, j)\left(\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}} + \mathcal{O}\left(\frac{1}{N^2}\right)\right)^{\lfloor Nt \rfloor}(k, m) \qquad (8) \\
&\quad = \sum_{\ell=j}^{i} C_\ell(i, j)\left(\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}}\right)^{\lfloor Nt \rfloor}(k, m) + \mathcal{O}\left(\frac{1}{N}\right), \qquad (9)
\end{aligned}$$

where $C_\ell(i, j)$ is given by (5) and (6) in Theorem 1.

Remark 2. Note that the matrices $\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}}$ appearing in the sum are the same for each i, j. In the Wright–Fisher model with constant size $M_N(r) \equiv N$, the above matrices are actually the scalars $\boldsymbol{P} = 1$ and $\hat{\boldsymbol{P}} = 1/N$. In this special case, summing over m and letting $N \to \infty$ in the lemma leads to the classical result that

$$\begin{aligned}
\mathrm{P}^{i}(A_N(\lfloor Nt \rfloor) = j) &= \sum_{\ell=j}^{i} C_\ell(i, j)\left(1 - \binom{\ell}{2}\frac{1}{N} + \mathcal{O}\left(\frac{1}{N^2}\right)\right)^{\lfloor Nt \rfloor} \\
&\to \sum_{\ell=j}^{i} C_\ell(i, j)\exp\left\{-\binom{\ell}{2}t\right\},
\end{aligned}$$

where the last term is the transition probability $\mathrm{P}^{i}(A(t) = j)$ for Kingman's coalescent (cf. Tavaré (1984)).

Proof. The asymptotic equivalence of (8) and (9) follows from a simple argument using the binomial theorem. Write $(\boldsymbol{\Pi}^r)_{i,j}$ for the (i, j)th block in the matrix $\boldsymbol{\Pi}^r$ and $(\boldsymbol{\Pi}_{i,j})^r$ for the matrix $\boldsymbol{\Pi}_{i,j}$ raised to the power r. These block matrices are all finite and of the same size as $\boldsymbol{P}$. The matrix powers $\boldsymbol{\Pi}^r$ are lower triangular block matrices given recursively by

$$(\boldsymbol{\Pi}^r)_{i,i} = (\boldsymbol{\Pi}_{i,i})^r = \left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right)^r, \qquad (10)$$

$$\begin{aligned}
(\boldsymbol{\Pi}^r)_{i,j} &= \boldsymbol{\Pi}_{i,i-1}(\boldsymbol{\Pi}^{r-1})_{i-1,j} + \boldsymbol{\Pi}_{i,i}(\boldsymbol{\Pi}^{r-1})_{i,j} \\
&= \binom{i}{2}\hat{\boldsymbol{P}}(\boldsymbol{\Pi}^{r-1})_{i-1,j} + \left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right)(\boldsymbol{\Pi}^{r-1})_{i,j}, \qquad 1 \leq j \leq i-1, \qquad (11)
\end{aligned}$$

and, of course, $(\boldsymbol{\Pi}^r)_{i,j} = \boldsymbol{0}$ when $j > i$. In view of the fact that

$$\begin{aligned}
\mathrm{P}^{(k,i)}((M_N(\lfloor Nt \rfloor), A_N(\lfloor Nt \rfloor)) = (m, j)) &= \left(\left(\boldsymbol{\Pi} + \mathcal{O}\left(\frac{1}{N^2}\right)\right)^{\lfloor Nt \rfloor}\right)_{i,j}(k, m) \\
&= (\boldsymbol{\Pi}^{\lfloor Nt \rfloor})_{i,j}(k, m) + \mathcal{O}\left(\frac{1}{N}\right),
\end{aligned}$$

the lemma will be proved if we can show that, for each $r \geq 1$ and $j \leq i$,

$$(\boldsymbol{\Pi}^r)_{i,j} = \sum_{\ell=j}^{i} C_\ell(i, j)\left(\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}}\right)^r. \qquad (12)$$

We begin with the case $j = i$, in which case the index sets in both products defining $C_\ell(i, j)$ in (5) are empty. Hence (12) takes the form

$$(\mathbf{\Pi}^r)_{i,i} = \left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right)^r,$$

in accordance with (10).

The proof of (12) for the case $j < i$ is by induction on r, starting with $r = 1$. For $r = 1$ and $j = i - 1$ the right-hand side of (12) becomes

$$\frac{\binom{i}{2}}{\binom{i}{2} - \binom{i-1}{2}}\left(\boldsymbol{P} - \binom{i-1}{2}\hat{\boldsymbol{P}}\right) + \frac{\binom{i}{2}}{\binom{i-1}{2} - \binom{i}{2}}\left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right) = \binom{i}{2}\hat{\boldsymbol{P}}$$
$$= (\mathbf{\Pi})_{i,i-1},$$

as desired. Moreover, $(\mathbf{\Pi})_{i,j} = \mathbf{0}$ for $1 \le j \le i - 2$. To see that the right-hand side of (12) is zero in this case, use the facts that, for $1 \le j \le i - 2$,

$$\sum_{\ell=j}^{i}\left(\prod_{j \le r \le i, r \ne \ell} \frac{1}{\binom{r}{2} - \binom{\ell}{2}}\right) = 0,$$
$$\sum_{\ell=j}^{i}\binom{\ell}{2}\left(\prod_{j \le r \le i, r \ne \ell} \frac{1}{\binom{r}{2} - \binom{\ell}{2}}\right) = 0.$$

These relations follow from general combinatorial properties of positive numbers. Indeed, for any distinct positive $x_r,\ r \in \{j, \dots, i\}$, by putting all terms in the sum over a common denominator,

$$\sum_{\ell=j}^{i}\left(\prod_{j \le r \le i; r \ne \ell} \frac{1}{x_r - x_\ell}\right) = \left(\prod_{j \le s < r \le i} \frac{1}{x_r - x_s}\right)\sum_{\ell=j}^{i}(-1)^{\ell-j} \prod_{j \le s < r \le i; r,s \ne \ell} (x_r - x_s).$$

The right-most sum above vanishes since each term that results from expanding the product appears twice, with opposite signs. Also, by a similar argument,

$$\sum_{\ell=j}^{i} x_\ell(-1)^{\ell-j} \prod_{j \le s < r \le i; r,s \ne \ell} (x_r - x_s) = 0.$$

Now fix $r \ge 2$ and suppose that (12) is true for all indices up to and including $r - 1$. Then, for any $j \le i - 1$,

$$\binom{i}{2}\hat{\boldsymbol{P}}(\mathbf{\Pi}^{r-1})_{i-1,j} = \sum_{\ell=j}^{i-1}\binom{i}{2}C_\ell(i-1, j)\hat{\boldsymbol{P}}\left(\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}}\right)^{r-1}$$

and

$$\left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right)(\mathbf{\Pi}^{r-1})_{i,j} = \sum_{\ell=j}^{i} C_\ell(i, j)\left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right)\left(\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}}\right)^{r-1}.$$

Therefore, by (11), we have

$$\begin{aligned}
(\mathbf{\Pi}^r)_{i,j} &= \binom{i}{2}\hat{\boldsymbol{P}}(\mathbf{\Pi}^{r-1})_{i-1,j} + \left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right)(\mathbf{\Pi}^{r-1})_{i,j} \\
&= \sum_{\ell=j}^{i-1} C_\ell(i,j)\left\{\left(\binom{i}{2} - \binom{\ell}{2}\right)\hat{\boldsymbol{P}} + \boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right\}\left(\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}}\right)^{r-1} \\
&\quad + C_i(i,j)\left(\boldsymbol{P} - \binom{i}{2}\hat{\boldsymbol{P}}\right)^r \\
&= \sum_{\ell=j}^{i} C_\ell(i,j)\left(\boldsymbol{P} - \binom{\ell}{2}\hat{\boldsymbol{P}}\right)^r.
\end{aligned}$$

This shows that (12) holds for index r, hence by induction for any index.

It remains to prove the alternative representation of the coefficients given in (6). To this end, we use

$$\binom{r}{2} - \binom{\ell}{2} = \frac{(r-\ell)(r+\ell-1)}{2}$$

to rewrite (5) as

$$C_\ell(i,j) = \frac{\prod_{j+1\le r\le i} r(r-1)}{\prod_{j\le r\le i;r\ne\ell}(r-\ell)(r+\ell-1)}.$$

Here we have

$$\begin{aligned}
\prod_{j+1\le r\le i} r(r-1) &= \frac{i!}{j!}j\cdots(i-1), \\
\prod_{j\le r\le i;r\ne\ell}(r-\ell) &= (j-\ell)\cdots 1\cdot(i-\ell)! = (-1)^{\ell-j}(\ell-j)!\,(i-\ell)!
\end{aligned}$$

and

$$\prod_{j\le r\le i;r\ne\ell}(r+\ell-1) = \frac{(j+\ell-1)\cdots(i+\ell-1)}{2\ell-1}.$$

Thus,

$$\begin{aligned}
C_\ell(i,j) &= \frac{(2\ell-1)(-1)^{\ell-j}}{j!\,(\ell-j)!}\frac{i!}{(i-\ell)!}\frac{j\cdots(i-1)}{(j+\ell-1)\cdots(i+\ell-1)} \\
&= \frac{(2\ell-1)(-1)^{\ell-j}(i)_\ell\, j_{(\ell-1)}}{j!\,(\ell-j)!\,i_{(\ell)}},
\end{aligned}$$

which is (6). This completes the proof of Lemma 1.

3.2. Limit distribution

It was pointed out earlier that to prove the weak convergence in Theorem 1 it suffices to establish (7), a criterion that we now reformulate slightly as follows. Let

$$(\mathcal{F}_r^{X_N})_{r\ge 0}, \qquad \mathcal{F}_r^{X_N} = \sigma\{X_N(k) : k \le r\}$$

be the discrete-time filtration generated by the Markov chain X_N. Then, for any $r \geq 0$,

$$\mathcal{T}_r^N f(x,i) = \sum_{j=1}^{i} \mathcal{T}_r^{N,j} f(x,i), \tag{13}$$

where

$$\mathcal{T}_r^{N,j} f(x,i) = \mathrm{E}^{(x,i)}[f(X_N(r), j)\, \mathrm{P}^{(x,i)}(A_N(r) = j \mid \mathcal{F}_r^{X_N})]. \tag{14}$$

Similarly, the transition semigroup for the limit process in Theorem 1 has the representation

$$\mathcal{T}_t f(x,i) = \sum_{j=1}^{i} \mathcal{T}_t^j f(x,i), \tag{15}$$

where

$$\mathcal{T}_t^j f(x,i) = \mathrm{E}^{(x,i)}[f(X(t), j)\, \mathrm{P}^i(A(Y_t) = j \mid \mathcal{F}_t^X)]$$

and

$$(\mathcal{F}_t^X)_{t\geq 0}, \qquad \mathcal{F}_t^X = \sigma\{X(s) : s \leq t\}$$

denotes the continuous-time filtration generated by X. It follows from (13) and (15) that to prove (7) it is enough to show that, for fixed t and each j,

$$\sup_{x,i} |\mathcal{T}_{\lfloor Nt \rfloor}^{N,j} f(x,i) - \mathcal{T}_t^j f(x,i)| \to 0, \qquad N \to \infty, \tag{16}$$

where the supremum is over $x \in S_N$ and $i \geq j$.

To compute the conditional probability appearing in (14), we begin by writing

$$\mathrm{P}^{(x,i)}(A_N(r) = i \mid \mathcal{F}_r^{X_N}) = \prod_{s=1}^{r} \left(1 - \binom{i}{2} c_N(M_N(s-1), M_N(s)) + \mathcal{O}\left(\frac{1}{N^2}\right)\right).$$

The next step is to apply the time-scale $r = \lfloor Nt \rfloor$ and rewrite the above in terms of $\tilde{X}_N$. Then, as $N \to \infty$, by obvious approximations that can be taken uniform in x and i,

$$\begin{aligned}\mathrm{P}^{(x,i)}(A_N(\lfloor Nt \rfloor) = i \mid \mathcal{F}_{\lfloor Nt \rfloor}^{X_N}) &\sim \exp\left\{-\binom{i}{2} \frac{1}{N} \sum_{r=1}^{\lfloor Nt \rfloor} H_N(X_N(r-1), X_N(r))\right\} \\ &= \exp\left\{-\binom{i}{2} \int_0^t H_N\left(\tilde{X}_N\left(s - \frac{1}{N}\right), \tilde{X}_N(s)\right) \mathrm{d}s\right\}.\end{aligned}$$

In the remaining proofs, it is convenient to have a separate notation for the multiplicative functionals of X defined by

$$M_\ell(t) = \exp\left\{-\binom{\ell}{2} Y_t\right\} = \exp\left\{-\binom{\ell}{2} \int_0^t H(X(s), X(s))\, \mathrm{d}s\right\}$$

and, similarly,

$$M_{\ell,N}(t) = \exp\left\{-\binom{\ell}{2} \int_0^t H_N\left(\tilde{X}_N\left(s - \frac{1}{N}\right), \tilde{X}_N(s)\right) \mathrm{d}s\right\}.$$

Since $Y_t \in \mathcal{F}_t^X$, we can use the classical expression for transition probabilities involving Kingman's coalescent (cf. Remark 2) to obtain that

$$\mathrm{P}^{(x,i)}(A(Y_t) = j \mid \mathcal{F}_t^X) = \sum_{\ell=j}^{i} C_\ell(i,j) M_\ell(t), \qquad j \le i.$$

In view of Lemma 1 and the above asymptotics, we can also write

$$\mathrm{P}^{(x,i)}(A_N(\lfloor Nt \rfloor) = j \mid \mathcal{F}_{\lfloor Nt \rfloor}^{X_N}) \sim \sum_{\ell=j}^{i} C_\ell(i,j) M_{\ell,N}(t), \qquad N \to \infty.$$

This gives

$$\begin{aligned} \mathcal{T}_{\lfloor Nt \rfloor}^{N,j} f(x,i) &= \mathrm{E}^{(x,i)}[f(\tilde{X}_N(t), j)\, \mathrm{P}^{(x,i)}(A_N(\lfloor Nt \rfloor) = j \mid \mathcal{F}_{\lfloor Nt \rfloor}^{X_N})] \\ &\sim \sum_{\ell=j}^{i} C_\ell(i,j)\, \mathrm{E}^{(x,i)}[f(\tilde{X}_N(t), j) M_{\ell,N}(t)] \end{aligned} \tag{17}$$

and

$$\begin{aligned} \mathcal{T}_t^j f(x,i) &= \mathrm{E}^{(x,i)}[f(X(t), j)\, \mathrm{P}^{(x,i)}(A(Y_t) = j \mid \mathcal{F}_t^X)] \\ &= \sum_{\ell=j}^{i} C_\ell(i,j)\, \mathrm{E}^{(x,i)}[f(X(t), j) M_\ell(t)]. \end{aligned} \tag{18}$$

We remark that, for each i, j, the semigroup $\mathrm{E}^{(x,i)}[f(X(t), j) M_\ell(t)]$ inherits the Feller property in x from that of the diffusion X.

By (17) and (18),

$$\begin{aligned} &\sup_{x,i} |\mathcal{T}_{\lfloor Nt \rfloor}^{N,j} f(x,i) - \mathcal{T}_t^j f(x,i)| \\ &\quad \le \sum_{\ell=j}^{i} C_\ell(i,j) \sup_{x,i} |\,\mathrm{E}^{(x,i)}[f(\tilde{X}_N(t), j) M_{\ell,N}(t)] - \mathrm{E}^{(x,i)}[f(X(t), j) M_\ell(t)]|. \end{aligned}$$

Since i is restricted to the finite set $\{2, \ldots, n\}$, we may fix an initial condition $A(0) = n$ and limit our attention in the sequel to a supremum over x. Furthermore, for fixed j, the function $f_j(x) = f(x, j)$ is just a particular choice of a bounded continuous function acting on x. To prove the criterion (16) it is therefore enough to prove the following uniform convergence of Feynman–Kac semigroups.

Lemma 2. *For t fixed and for all choices of $f \in \mathrm{C_b}(S)$,*

$$\sup_{x \in S_N} |\,\mathrm{E}^x[f(\tilde{X}_N(t)) M_{\ell,N}(t)] - \mathrm{E}^x[f(X(t)) M_\ell(t)]| \to 0 \quad \textit{as } N \to \infty.$$

Proof. For given $\delta > 0$, consider the equidistant sequence $t_i = \delta i,\ i \ge 0$, and let

$$\begin{aligned} \Delta M_\ell(t_i) &= M_\ell(t_i) - M_\ell(t_{i-1}) \\ &= M_\ell(t_{i-1}) \left(\exp\left\{ -\binom{\ell}{2} \int_{t_{i-1}}^{t_i} H(X(s), X(s))\, \mathrm{d}s \right\} - 1 \right). \end{aligned}$$

Since H is assumed to be bounded,

$$\Delta M_\ell(t_i) = -M_\ell(t_{i-1})\binom{\ell}{2}\int_{t_{i-1}}^{t_i} H(X(s), X(s))\,\mathrm{d}s + \mathcal{O}(\delta^2), \qquad \delta \to 0,$$

almost surely, where the remainder term is uniform in i and the initial state x.

Hence, for any $k \geq 1$,

$$\begin{aligned}
&\mathrm{E}^x[f(X(t_k))M_\ell(t_k)] - f(x)\\
&\quad = \sum_{i=1}^{k} \mathrm{E}^x[f(X(t_i))\Delta M_\ell(t_i) + (f(X(t_i)) - f(X(t_{i-1})))M_\ell(t_{i-1})]\\
&\quad = S_k^1 + S_k^2 + \mathcal{O}(t_k\delta),
\end{aligned}$$

where

$$S_k^1 = -\binom{\ell}{2}\sum_{i=1}^{k} \mathrm{E}^x\left[M_\ell(t_{i-1})f(X(t_i))\int_{t_{i-1}}^{t_i} H(X(s), X(s))\,\mathrm{d}s\right]$$

and

$$S_k^2 = \sum_{i=1}^{k} \mathrm{E}^x[M_\ell(t_{i-1})(f(X(t_i)) - f(X(t_{i-1})))],$$

and the remainder term $\mathcal{O}(t_k\delta) = \mathcal{O}(k\delta^2)$ is again uniform in $X(0) = x$. Next, using the Markov property, we observe that

$$\begin{aligned}
S_k^1 &= -\binom{\ell}{2}\sum_{i=1}^{k} \mathrm{E}^x\left[M_\ell(t_{i-1})\int_{t_{i-1}}^{t_i} H(X(s), X(s))T_{t_i-s}f(X(s))\,\mathrm{d}s\right]\\
&= -\binom{\ell}{2}\sum_{i=1}^{k} \mathrm{E}^x\left[M_\ell(t_{i-1})\int_{t_{i-1}}^{t_i} T_{s-t_{i-1}}(HT_{t_i-s}f)(X(t_{i-1}))\,\mathrm{d}s\right]\\
&= -\binom{\ell}{2}\sum_{i=1}^{k}\int_{t_{i-1}}^{t_i} \mathrm{E}^x[M_\ell(t_{i-1})R_sf(X(t_{i-1}))]\,\mathrm{d}s,
\end{aligned}$$

where, for each s, $R_sf(x) = R_s^{t_{i-1},t_i}f(x) := T_{s-t_{i-1}}(HT_{t_i-s}f)(x)$ is a bounded, continuous function. Similarly,

$$S_k^2 = \sum_{i=1}^{k} \mathrm{E}^x[M_\ell(t_{i-1})Q_{t_1}f(X(t_{i-1}))],$$

where $Q_tf(x) = T_tf(x) - f(x)$ is a bounded, continuous function.

In summary,

$$\begin{aligned}
\mathrm{E}^x[f(X(t_k))M_\ell(t_k)] - f(x) = -\binom{\ell}{2}\sum_{i=1}^{k}\int_{t_{i-1}}^{t_i} \mathrm{E}^x[M_\ell(t_{i-1})R_sf(X(t_{i-1}))]\,\mathrm{d}s\\
+ \sum_{i=1}^{k} \mathrm{E}^x[M_\ell(t_{i-1})Q_{t_1}f(X(t_{i-1}))] + \mathcal{O}(t_k\delta). \qquad (19)
\end{aligned}$$

The same derivation in modified form applies to the process $\tilde{X}_N(t)$. Namely, take $Q_t^N f(x) = T_t^N f(x) - f(x)$ and $R_s^N f(x) = T_{s-t_{i-1}}^N (H T_{t_i - s}^N f)(x)$ and observe that these functions are all uniformly bounded in N. This gives, for any N,

$$\mathrm{E}^x[f(\tilde{X}_N(t_k))M_{\ell,N}(t_k)] - f(x) = -\binom{\ell}{2}\sum_{i=1}^{k}\int_{t_{i-1}}^{t_i} \mathrm{E}^x[M_{\ell,N}(t_{i-1})R_s^N f(\tilde{X}_N(t_{i-1}))]\,\mathrm{d}s$$
$$+\sum_{i=1}^{k}\mathrm{E}^x[M_{\ell,N}(t_{i-1})Q_{t_1}^N f(\tilde{X}_N(t_{i-1}))] + \mathcal{O}(t_k\delta), \quad (20)$$

where the remainder term is independent of f and uniform in both x and N as $\delta \to 0$.

Now fix $\varepsilon > 0$. For the given time t, choose δ to be a number of the form $\delta = t/m$, where m is an integer which is so large that the remainder terms in both (19) and (20), which are both of the form $\mathcal{O}(t_k\delta) = \mathcal{O}(k\delta t/m)$, are bounded by ε for $k = 1, \dots, m$. With this δ we now apply (19) and (20) to obtain, for arbitrary N, for $1 \le k \le m$ and for $f \in \mathrm{C_b}(S)$,

$$\sup_{x\in S_N} |\,\mathrm{E}^x[f(X(t_k))M_\ell(t_k)] - \mathrm{E}^x[f(\tilde{X}_N(t_k))M_{\ell,N}(t_k)]|$$
$$\le \binom{\ell}{2}\sum_{i=0}^{k-1}\int_{t_i}^{t_{i+1}} \sup_{x\in S_N} |\,\mathrm{E}^x[M_\ell(t_i)R_s f(X(t_i))] - \mathrm{E}^x[M_{\ell,N}(t_i)R_s^N f(\tilde{X}_N(t_i))]|\,\mathrm{d}s$$
$$+\sum_{i=0}^{k-1}\sup_{x\in S_N} |\,\mathrm{E}^x[M_\ell(t_i)Q_{t_1} f(X(t_i))] - \mathrm{E}^x[M_{\ell,N}(t_i)Q_{t_1}^N f(\tilde{X}_N(t_i))]| + 2\varepsilon. \quad (21)$$

To prove the lemma it remains to show that the expression in (21) for $k = m$ (recall that $t_m = t$) goes to zero as $N \to \infty$. To this end, we prove the following discretized version of the lemma. Namely, for the above partition $t_1, \dots, t_m$, we claim that

$$\sup_{x\in S_N} |\,\mathrm{E}^x[g(X(t_k))M_\ell(t_k)] - \mathrm{E}^x[g(\tilde{X}_N(t_k))M_{\ell,N}(t_k)]| \to 0, \qquad N \to \infty, \quad (22)$$

for all choices of $g \in \mathrm{C_b}(S)$ and $k = 1, \dots, m$. We prove the claim by induction on k. First, take $k = 1$ in (21) to get

$$\sup_{x\in S_N} |\,\mathrm{E}^x[M_\ell(t_1)g(X(t_1))] - \mathrm{E}^x[M_{\ell,N}(t_1)g(\tilde{X}_N(t_1))]|$$
$$\le \binom{\ell}{2}\int_0^{t_1} \sup_{x\in S_N} |R_s g(x) - R_s^N g(x)|\,\mathrm{d}s + \sup_{x\in S_N} |Q_{t_1} g(x) - Q_{t_1}^N g(x)| + 2\varepsilon. \quad (23)$$

Based on the contraction property of the semigroups, the uniform convergence in Assumption 1 and the uniform convergence of H_N in Assumption 2, the suprema in each term on the right-hand side of (23) converge to zero as $N \to \infty$. By the dominated convergence theorem, the integral also vanishes as $N \to \infty$. Hence,

$$\sup_{x\in S_N} |\,\mathrm{E}^x[M_\ell(t_1)g(X(t_1))] - \mathrm{E}^x[M_{\ell,N}(t_1)g(\tilde{X}_N(t_1))]| \le 3\varepsilon, \qquad N \ge N_0(1),$$

providing the first induction step. Now suppose as induction hypothesis that the claim (22) is true for each index $1, \dots, k-1$ less than k. Then, by (21), using the same arguments that led to (23), we can find $N_0(k)$ so large and each term on the right-hand side of (21) so small that

$$\sup_{x\in S_N} |\mathrm{E}^x[M_\ell(t_k)g(X(t_k))] - \mathrm{E}^x[M_{\ell,N}(t_k)g(\tilde{X}_N(t_k))]| \le 3\varepsilon, \qquad N \ge N_0(k),$$

which verifies the induction step for the proof of the claim (22). This concludes the proof of the lemma and hence the proof of Theorem 1.

Acknowledgements

We thank Martin Lascoux, Magnus Nordborg and Tom Kurtz for discussions. Thanks also go to an anonymous referee for suggestions. SMK was supported in part by NSF grants DMS-00-72198 and EPS-00-80935.

References

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

Donnelly, P. (1986). A genealogical approach to variable-population-size models in population genetics. *J. Appl. Prob.* **23,** 283–296.

Donnelly, P. and Kurtz, T. G. (1999). Particle representations for measure-valued population models. *Ann. Prob.* **27,** 166–205.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*. John Wiley, New York.

Griffiths, R. C. and Tavaré, S. (1994). Sampling theory for neutral alleles in a varying environment. *Phil. Trans. R. Soc. London B* **344,** 403–410.

Kaj, I., Krone, S. M. and Lascoux, M. (2001). Coalescent theory for seed bank models. *J. Appl. Prob.* **38,** 285–300.

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

Möhle, M. (2000). Ancestral processes in population genetics—the coalescent. *J. Theoret. Biol.* **204,** 629–638.

Möhle, M. (2002). The coalescent in population models with time-inhomogeneous environment. *Stoch. Process. Appl.* **97,** 199–227.

Nordborg, M. and Krone, S. M. (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Developments in Theoretical Population Genetics*, eds M. Slatkin and M. Veuille, Cambridge University Press, pp. 194–232.

Rogers, L. C. G. and Williams, D. (1987). *Diffusions, Markov Processes, and Martingales*, Vol. 2. John Wiley, Chichester.

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