Adv. Appl. Prob. **30,** 493–512 (1998)
Printed in Northern Ireland

A CONVERGENCE THEOREM FOR MARKOV CHAINS ARISING IN POPULATION GENETICS AND THE COALESCENT WITH SELFING

M. MÖHLE,* *University of Chicago and Johannes Gutenberg-Universität Mainz*

Abstract

A simple convergence theorem for sequences of Markov chains is presented in order to derive new 'convergence-to-the-coalescent' results for diploid neutral population models.

For the so-called diploid Wright–Fisher model with selfing probability s and mutation rate θ, it is shown that the ancestral structure of n sampled genes can be treated in the framework of an n-coalescent with mutation rate $\tilde{\theta} := \theta(1 - s/2)$, if the population size N is large and if the time is measured in units of $(2 - s)N$ generations.

Keywords: Coalescent; diploid population models; genealogical process; population genetics; robustness; selfing

AMS 1991 Subject Classification: Primary 60F05; 92D10
Secondary 60237; 92D25

1. Introduction

Continuous-time Markov chains are usually characterized by their transition matrix $\Pi(t) = \mathrm{e}^{tG}$, where G is the so-called infinitesimal generator. Note that $\Pi(0+) = I$, i.e. this characterization is only satisfied for Markov chains which have no instantaneous jumps at time $t = 0$. In Section 2 a simple time-scaling convergence theorem for sequences of discrete-time Markov chains on the same finite state space is presented. The transition matrix of the corresponding limit process has the more general form $\Pi(t) = P - I + \mathrm{e}^{tG} = P\mathrm{e}^{tG}$. The matrix P describes the instantaneous jumps at time $t = 0$.

The theorem is, for example, useful to derive convergence results for ancestral processes arising in population genetics. A special stochastic process, called the coalescent, is of fundamental interest in population genetics. For a large class of haploid population models this process is the appropriate tool to analyse the ancestral structure of a sample of n genes (or individuals), if the total number of genes in the population is sufficiently large. A corresponding convergence theorem for a large class of exchangeable population models was first proved by Kingman (see [6, 7, 8]). More recently the coalescent-theory has been extended to more general and more complicated models, for example for models with underlying mutation, selection or recombination, for models with variable population size or for non-exchangeable models. Only some of the publications are listed here: [14, 5, 3, 4, 10]. One speaks of the robustness of the coalescent, as this process appears in a lot of quite different models when the total population size tends to infinity.

Received 15 July 1996; revision received 10 October 1996.
* Postal address: (1) The University of Chicago, Department of Statistics, 5734 University Avenue, Chicago, IL 60637, USA. (2) Johannes Gutenberg-Universität Mainz, Fachbereich Mathematik, Saarstraße 21, 55099 Mainz, Germany. E-mail address: (1) moehle@galton.uchicago.edu, (2) moehle@mathematik.uni-mainz.de

The purpose is to illustrate how the convergence results given in Section 2 can be used to extend the coalescent-theory to more complex, for example, diploid population models (this paper) or two-sex population models (see [11]).

In Section 3 a typical diploid plant population system with a mixture of self-fertilization and random mating ([1, 9]) is studied. An application of the convergence results of Section 2 leads to the so-called coalescent with selfing. Previous applications of the coalescent to diploid models [9] ignored the correlations in offspring numbers between genes within individuals. The results presented here now provide a formal justification of this problem.

2. A simple but useful convergence result

The convergence results presented in this paper are all based on the following Lemma 1, which is a generalization of the well known matrix equation $\lim_{N\to\infty}(I + A/N)^N = \mathrm{e}^A$. A proof is given in the appendix. Throughout this paper, for the matrix $A = (a_{ij})$, the norm $\|A\| := \max_i \sum_j |a_{ij}|$ is used. Note that $\|AB\| \le \|A\|\,\|B\|$ and that $\|A\| = 1$ if A is a stochastic matrix.

Lemma 1. *Let $t, K \ge 0$ be fixed and let $(c_N)_{N\in\mathbb{N}}$ be a sequence of positive real numbers with $\lim_{N\to\infty} c_N = 0$. Further let A be a matrix with $\|A\| = 1$ such that $P := \lim_{m\to\infty} A^m$ exists. Then*

$$\lim_{N\to\infty} \sup_{\|B\|\le K} \|(A + c_N B)^{[t/c_N]} - (P + c_N B)^{[t/c_N]}\| = 0.$$

If $(B_N)_{N\in\mathbb{N}}$ is a matrix sequence such that $G := \lim_{N\to\infty} PB_NP$ exists, then

$$\lim_{N\to\infty} (A + c_N B_N)^{[t/c_N]} = P - I + \mathrm{e}^{tG} \qquad \forall\, t > 0.$$

Remarks.

1. Note that P is a projection, i.e. $P^2 = P$ and therefore $GP = PG = G$ and $P - I + \mathrm{e}^{tG} = P\mathrm{e}^{tG} = \mathrm{e}^{tG}P$.
2. The convergence of the sequence $(B_N)_{N\in\mathbb{N}}$ is not required. Of course, if
$$B := \lim_{N\to\infty} B_N$$
exists, then $G = PBP$.

Lemma 1 can be used to derive time-scaling convergence results for sequences of Markov chains with the same finite state space, which appear in many fields in applied probability, especially in population genetics.

Theorem 1. *Let $X_N = (X_N(r))_{r\in\mathbb{N}_0}$ be a sequence of time homogeneous Markov chains on a probability space $(\Omega, \mathcal{F}, P)$ with the same finite state space S and let Π_N denote the transition matrix of X_N. Assume that the following conditions are satisfied.*

1. *$A := \lim_{N\to\infty} \Pi_N$ exists and $\Pi_N \ne A$ for all sufficiently large N.*
2. *$P := \lim_{m\to\infty} A^m$ exists.*
3. *$G := \lim_{N\to\infty} PB_NP$ exists, where $B_N := (\Pi_N - A)/c_N$ and $c_N := \|\Pi_N - A\|$ for all $N \in \mathbb{N}$.*

If the sequence of initial probability measures $P_{X_N(0)}$ converge weakly to some probability measure μ, then the finite-dimensional distributions of the process $(X_N([t/c_N]))_{t\geq 0}$ converge to those of a time continuous Markov process $(X_t)_{t\geq 0}$ with initial distribution

$$X_0 \stackrel{\mathrm{d}}{=} \mu,$$

transition matrix $\Pi(t) := P - I + \mathrm{e}^{tG} = P\mathrm{e}^{tG}$, $t > 0$, and infinitesimal generator G.

Remarks.

1. The equation $\Pi(t) = P\mathrm{e}^{tG}$, i.e.

$$\pi_{ij}(t) = \sum_{k\in S} p_{ik}(\mathrm{e}^{tG})_{kj},$$

leads to a helpful interpretation. The limit process jumps instantaneously from a state $i \in S$ at time $t = 0$ to a state $k \in S$ at time $t = 0+$, with probability p_{ik}, i.e. the matrix P describes the instantaneous jumps. Then the evolution behaves like a Markov chain with initial value k and infinitesimal generator G. In most applications only a few entries of the projection P are not equal to zero. Hence the process moves instantaneously to a state k belonging to some 'small' subset $S' \subset S$.

2. Obviously $\|\Pi_N\| = 1$ for all $N \in \mathbb{N}$. Hence there exists at least a subsequence $(N_k)_{k\in\mathbb{N}}$ with $\lim_{k\to\infty} N_k = \infty$ such that $A := \lim_{k\to\infty} \Pi_{N_k}$ exists. Now $\|B_{N_k}\| = 1$ for all $k \in \mathbb{N}$ and hence (eventually, after changing to another subsequence) $B := \lim_{k\to\infty} B_{N_k}$ exists. Thus the assumptions 1 and 3 of the theorem are not as strong as they seem to be at the first glance.

Proof of Theorem 1. From Lemma 1 it follows that

$$\lim_{N\to\infty} \Pi_N^{[t/c_N]} = \lim_{N\to\infty} (A + c_N B_N)^{[t/c_N]} = P - I + \mathrm{e}^{tG} = \Pi(t)$$

for all $t > 0$. Hence the finite-dimensional distributions of $(X_N([t/c_N]))_{t\geq 0}$ converge to those of a Markov process, $(X_t)_{t\geq 0}$, with initial distribution μ and transition matrix $\Pi(t)$, $t > 0$. The infinitesimal generator of the process $(X_t)_{t\geq 0}$ is given by

$$\lim_{t\searrow 0} \frac{\Pi(t) - \Pi(0+)}{t} = \lim_{t\searrow 0} \frac{P - I + \mathrm{e}^{tG} - P}{t} = \lim_{t\searrow 0} \frac{\mathrm{e}^{tG} - I}{t} = G.$$

The next section is the applied part of this article. The ancestral structure of a special diploid plant population model is studied in detail. An application of the above convergence results leads to the so-called coalescent with selfing.

3. The coalescent with selfing

The mechanism of reproduction in plant population mating systems often involves a mixture of self fertilizations and out-cross fertilization. Several models have been suggested to characterize such population systems [1, 2]. Most of them are based on the fraction s of individuals

produced by self fertilization. The parameter s is called the selfing probability or the selfing rate.

Here the standard Wright–Fisher model with a fixed number of N diploid individuals is assumed. This means that genes belonging to different individuals choose independently their parent genes, and that the parent genes of two genes of the same individual belong with probability s to the same parent individual, and with probability $1 - s$ to different parent individuals.

Consider a single diploid locus. Fix $n \leq 2N$ and choose n genes (without replacement) from the current generation 0. Label these n genes randomly from 1 to n and consider the set of ancestral genes at time $r \in \mathbb{N}_0 := \{0, 1, 2, \dots\}$, i.e. r generations backwards in time.

3.1. The case of $n = 2$ sampled genes

For simplicity the case $n = 2$ is considered for the moment. The general case will be studied later. Define

$$\mathcal{D}_r := \begin{cases} 1, & \text{if the two ancestral genes are the same (identical by descent),} \\ 2, & \text{if the genes belong to different individuals,} \\ 3, & \text{if the genes are distinct, but belong to the same individual.} \end{cases}$$

Obviously the so-called backward process $(\mathcal{D}_r)_{r\in\mathbb{N}_0}$ is a Markov chain with state space $S_2 = \{1, 2, 3\}$, initial distribution

$$(P(\mathcal{D}_0 = 1), P(\mathcal{D}_0 = 2), P(\mathcal{D}_0 = 3)) = (0, (2N - 2)(2N - 1)^{-1}, (2N - 1)^{-1})$$

and transition matrix

$$\Pi_N = (\pi_{ij})_{i,j\in S_2} = \begin{pmatrix} 1 & 0 & 0 \\ 1/(2N) & 1 - 1/N & 1/(2N) \\ s/2 & 1 - s & s/2 \end{pmatrix}.$$

Here s is the probability that the ancestral genes of the two genes of an individual belong to the same ancestral individual. Obviously 1 is an absorbing state. The transition matrix has a decomposition of the form $\Pi_N = A + B/N$, where

$$A := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s/2 & 1 - s & s/2 \end{pmatrix} \quad \text{and} \quad B := \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

The matrix A contains the transition probabilities due to the underlying 'selfing mechanism', while the matrix B contains the transition probabilities coming from the 'random mating mechanism' due to the assumed Wright–Fisher model. The second mechanism is of order N 'slower' than the first one. Note that

$$P = \lim_{m\to\infty} A^m = \lim_{m\to\infty} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{s}{2}\frac{1 - (s/2)^m}{1 - s/2} & (1 - s)\frac{1 - (s/2)^m}{1 - s/2} & \left(\frac{s}{2}\right)^m \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p & q & 0 \end{pmatrix},$$

with

$$p := \frac{s}{2 - s} \quad \text{and} \quad q := 1 - p = \frac{2(1 - s)}{2 - s}. \tag{1}$$

Now apply Lemma 1 to show that

$$\lim_{N\to\infty} \Pi_N^{[Nt]} = \lim_{N\to\infty} (A + B/N)^{[Nt]} = P - I + \mathrm{e}^{tPBP} = P\mathrm{e}^{tPBP},$$

or equivalently,

$$\lim_{N\to\infty} \Pi_N^{[(2-s)Nt]} = P - I + \mathrm{e}^{(2-s)tPBP} = P\mathrm{e}^{(2-s)tPBP} =: \Pi(t)$$

for all $t > 0$. Hence the finite-dimensional distributions of the process $(\mathcal{D}_{[(2-s)Nt]})_{t\geq 0}$ converge to those of a time continuous Markov process $(D_t)_{t\geq 0}$ with initial distribution (0, 0, 1), transition matrix

$$\Pi(t) = P - I + \mathrm{e}^{(2-s)tPBP} = \begin{pmatrix} 1 & 0 & 0 \\ 1-\mathrm{e}^{-t} & \mathrm{e}^{-t} & 0 \\ 1-q\mathrm{e}^{-t} & q\mathrm{e}^{-t} & 0 \end{pmatrix}.$$

and infinitesimal generator

$$G = \lim_{t\searrow 0} \frac{\Pi(t) - \Pi(0+)}{t} = (2-s)PBP = \begin{pmatrix} 0 & 0 & 0 \\ 1 & -1 & 0 \\ q & -q & 0 \end{pmatrix}.$$

Call this process the 2-coalescent with selfing probability s. The state 3 is instantaneous. Eliminating this state leads to the usual 2-coalescent (see [6, 7, 8]).

Remarks.

1. The effective population size is (approximately) given by $N_\mathrm{e} = N(2-s)/2$ [13, 15]. Thus it is reasonable to measure the time in units of $2N_\mathrm{e} = (2-s)N$ generations.

2. In applications it is often assumed that the selfing probability is not a constant s but something of the form $s + O(N^{-1})$, for example $s + (1-s)/N$ (see [9, 12]). In this case an appropriate decomposition of Π_N is given by $\Pi_N = A + C/N$ with A defined as before and

$$C := \begin{pmatrix} 0 & 0 & 0 \\ \frac{1}{2} & -1 & \frac{1}{2} \\ (1-s)/2 & s-1 & (1-s)/2 \end{pmatrix}.$$

   C divers slightly from B but $PCP = PBP$ and hence the limit process is the same. Nordborg and Donnelly [12] present an informal proof of this result. They also discuss its application to the problem of estimating simultaneously the selfing probability and the mutation rate.

3.2. The general case of *n* sampled genes

In principle the arguments given above can be carried over to a sample of size n. Unfortunately the state space is more complicated and hence the formal proof becomes more difficult. Call an individual of type $k \in \{0, 1, 2\}$, if exactly k of the two genes of this individual belong to the set of all the ancestral genes. Now consider the Markov process $(\mathcal{D}_r)_{r\in\mathbb{N}_0} := (\mathcal{N}_r, \mathcal{X}_r)_{r\in\mathbb{N}_0}$, where $\mathcal{N}_r$ denotes the number of ancestral genes and $\mathcal{X}_r$ denotes the number of individuals of type 2 in generation r, i.e. r steps backwards in time. A typical state is of the

form $i = (b, x)$ with $b \in \{1, \dots, n\}$ and $x \in \{0, \dots, [b/2]\}$ and the state space S_n has size $|S_n| = \sum_{b=1}^{n}([b/2] + 1) = [n(n+4)/4]$. In order to analyse the initial distribution of the process $(\mathcal{D}_r)_{r\in\mathbb{N}_0}$, i.e. the distribution of $\mathcal{D}_0$, let $p(N, n, x) := P(\mathcal{D}_0 = (n, x)) = P(\mathcal{X}_0 = x)$ denote the probability that the n genes of the current generation 0 are sampled in such a way that there are exactly x individuals of type 2 involved. Obviously $p(N, 1, x) = \delta_{0x}$ and

$$p(N, n+1, x) = \frac{n - 2x + 2}{2N - n}\, p(N, n, x-1) \;+\; \frac{2(N - n + x)}{2N - n}\, p(N, n, x),$$

for $n \geq 1$. The solution of this recursion is given by

$$\begin{aligned} p(N, n, x) &= \frac{(n)_{2x}}{2^x\, x!}\, \frac{2^{n-x}\, (N)_{n-x}}{(2N)_n}, \qquad (2)\\ &= \begin{cases} 1 - q_n/(2N) + O(N^{-2}) & \text{if } x = 0,\\ q_n/(2N) + O(N^{-2}) & \text{if } x = 1,\\ O(N^{-2}) & \text{if } x \geq 2; \end{cases} \end{aligned}$$

where the notation $(n)_0 := 1$ and $(n)_k := n(n-1)\cdots(n-k+1)$ for all $k \in \mathbb{N}$ is used and $q_n := (n)_2/2 = n(n-1)/2$. For $i = (b, x)$, $j = (a, y) \in S_n$ let $\pi_{ij} := P(\mathcal{D}_r = j \,|\, \mathcal{D}_{r-1} = i)$ denote the transition probabilities of the process $(\mathcal{D}_r)_{r\in\mathbb{N}_0}$. Transitions from i to j occur with probability

$$\pi_{ij} = \binom{x}{b-a}\binom{x-(b-a)}{y}(1-s)^{x-y-(b-a)}\left(\frac{s}{2}\right)^{y+b-a} + O(N^{-1}). \qquad (3)$$

The exact form of π_{ij} is not very important except for the case $x = 0$, where it can be shown that

$$\begin{aligned} \pi_{ij} &= S(b, a)(2N)^{-b}(2N)_a\; p(N, a, y),\\ &= \begin{cases} 1 - q_b/N + O(N^{-2}) & \text{if } j = i,\\ q_b/(2N) + O(N^{-2}) & \text{if } j = (b-1, 0) \text{ or } j = (b, 1),\\ O(N^{-2}) & \text{otherwise,} \end{cases} \qquad (4) \end{aligned}$$

where $S(b, a)$ denotes the Stirling numbers of the second kind. From (3) it follows that the transition matrix $\Pi_N = (\pi_{ij})_{i,j\in S_n}$ of the process $(\mathcal{D}_r)_{r\in\mathbb{N}_0}$ has a decomposition of the form $\Pi_N = A + B_N/N$ where $A := \lim_{N\to\infty} \Pi_N$ and $B_N := N(\Pi_N - A)$. The matrix-sequence $(B_N)_{N\in\mathbb{N}}$ is bounded and the entries a_{ij}, $i = (b, x)$, $j = (a, y) \in S_n$, of A are given by

$$a_{ij} = \binom{x}{b-a}\binom{x-(b-a)}{y}(1-s)^{x-y-(b-a)}\left(\frac{s}{2}\right)^{y+b-a}.$$

These are exactly the probabilities of the trinomial distribution $Mn(x, 1-s, s/2, s/2)$ evaluated at the point $(x-y-(b-a), b-a, y)$. Note that for given $(b, x) \in S_n$ these probabilities are positive only if $0 \leq x \leq b-a$ and $0 \leq y \leq x-(b-a)$. Thus a_{ij} can be positive only if $1 \leq a \leq b$ and

$$y \leq x - b + a = x - b/2 + (a-b)/2 + a/2 \leq 0 + 0 + a/2 = a/2.$$

A is a lower triangular matrix if the states are ordered such that $i \leq j :\Longleftrightarrow a < b$ or $(a = b$ and $x \leq y)$. It is not easy to find closed forms for all the entries $b_{ij}^{(N)}$ of B_N. For the special case $i = (b, 0)$ it follows from (4) that

$$b_{ij} := \lim_{N\to\infty} b_{ij}^{(N)} = \begin{cases} -q_b & \text{if } j = i, \\ q_b/2 & \text{if } j = (b-1, 0) \text{ or } j = (b, 1), \\ 0 & \text{otherwise.} \end{cases}$$

As A is a stochastic matrix there exists a Markov process $(Z_m)_m$ with corresponding transition matrix A. Note that $a_{ii} = 1$ if and only if $i = (b, 0)$, i.e. the states $(b, 0)$, $b \in \{1, \ldots, n\}$, are the absorbing states of the process $(Z_m)_m$. For $i, j \in S_n$ define $p_{ij} := P(Z_m = j \text{ finally} \mid Z_0 = i)$. From $P = AP$, i.e. $p_{ij} = \sum_{k\in S_n} a_{ik} p_{kj} = \sum_{j\leq k\leq i} a_{ik} p_{kj}$, it follows that the p_{ij} can be recursively (recursion on i) calculated via

$$p_{ij} = \begin{cases} \delta_{ij} & \text{if } a_{ii} = 1, \\ \dfrac{1}{1 - a_{ii}} \displaystyle\sum_{j\leq k<i} a_{ik} p_{kj} & \text{if } a_{ii} < 1, \end{cases}$$

where δ_{ij} denotes the Kronecker symbol. It is shown below that the solution of this recursion is given by

$$p_{ij} = \begin{cases} \dbinom{x}{b-a} p^{b-a} q^{x-(b-a)} & \text{if } y = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

where $i = (b, x)$, $j = (a, y) \in S_n$ and p and q are given by (1). Thus if $y = 0$ then p_{ij} is equal to the binomial probability $B(x, p, b - a)$ with the parameters x and p evaluated at $b - a$. Now the proof of (5) is given. From the recursion it follows that there is only one solution of the equation $P = AP$. Thus one has only to verify that the p_{ij}, as given in (5), solve the equation $\sum_{k\in S_n} a_{ik} p_{kj} = p_{ij}$. This is obviously the case for $y > 0$. Assume now that $y = 0$. Then it follows that

$$\sum_{k\in S_n} a_{ik}\, p_{kj} = \sum_{c=a}^{b} \sum_{z=c-a}^{x-(b-c)} \binom{x}{b-c} \binom{x-(b-c)}{z} \times (1-s)^{x-(b-c)-z} \left(\frac{s}{2}\right)^{z+b-c} \binom{z}{c-a} p^{c-a}\, q^{z-(c-a)}.$$

From

$$\binom{x}{b-c}\binom{x-(b-c)}{z}\binom{z}{c-a} = \binom{x}{b-a}\binom{b-a}{c-a}\binom{x-(b-a)}{z-(c-a)}$$

and

$$\begin{aligned} &(1-s)^{x-(b-c)-z} \left(\tfrac{1}{2}s\right)^{z+b-c} p^{c-a}\, q^{z-(c-a)} \\ &\quad = ((1-s)(1+p))^{x-(b-c)-z} \left(\tfrac{1}{2}s(1+p)\right)^{z+b-c} (1+p)^{-x}\, p^{c-a}\, q^{z-(c-a)} \\ &\quad = q^{x-(b-c)-z}\, p^{z+b-c}\, (1+p)^{-x}\, p^{c-a}\, q^{z-(c-a)} \\ &\quad = p^{z+b-a}\, q^{x-(b-a)}\, (1+p)^{-x}, \end{aligned} \tag{6}$$

it follows that

$$\begin{aligned}\sum_{k\in S_n} a_{ik}\, p_{kj} &= \binom{x}{b-a} p^{b-a} q^{x-(b-a)} (1+p)^{-x} \sum_{c=a}^{b} \binom{b-a}{c-a} \sum_{z=c-a}^{x-(b-c)} \binom{x-(b-a)}{z-(c-a)} p^z \\ &= p_{ij}\,(1+p)^{-x} \sum_{c=a}^{b} \binom{b-a}{c-a} \sum_{l=0}^{x-(b-a)} \binom{x-(b-a)}{l} p^{l+(c-a)} \\ &= p_{ij}\,(1+p)^{-x} \sum_{c=a}^{b} \binom{b-a}{c-a} p^{c-a} (1+p)^{x-(b-a)} \\ &= p_{ij}\,(1+p)^{-(b-a)} \sum_{l=0}^{b-a} \binom{b-a}{l} p^l = p_{ij},\end{aligned}$$

and (5) is proven. Because the state space is finite and from each state i there is at least one absorbing state reachable with positive probability, it follows that $\lim_{m\to\infty} a_{ij}^{(m)}$ exists and is equal to p_{ij}. Thus the entries of $P = \lim_{m\to\infty} A^m$ are given by (5). For example for the case $n = 4$ it follows that

$$A = \begin{pmatrix} 1 & & & & & & & \\ 0 & 1 & & & & & & \\ s/2 & 1-s & s/2 & & & & & \\ 0 & 0 & 0 & 1 & & & & \\ 0 & s/2 & 0 & 1-s & s/2 & & & \\ 0 & 0 & 0 & 0 & 0 & 1 & & \\ 0 & 0 & 0 & s/2 & 0 & 1-s & s/2 & \\ 0 & s^2/4 & 0 & s(1-s) & s^2/2 & (1-s)^2 & s(1-s) & s^2/4 \end{pmatrix}$$

and

$$P = \begin{pmatrix} 1 & & & & & & & \\ 0 & 1 & & & & & & \\ p & q & 0 & & & & & \\ 0 & 0 & 0 & 1 & & & & \\ 0 & p & 0 & q & 0 & & & \\ 0 & 0 & 0 & 0 & 0 & 1 & & \\ 0 & 0 & 0 & p & 0 & q & 0 & \\ 0 & p^2 & 0 & 2pq & 0 & q^2 & 0 & 0 \end{pmatrix}.$$

The calculation of the entries h_{ij} of the matrix $H := \lim_{N\to\infty} P B_N P$ is now straightforward. Obviously $h_{ij} = 0$ for $y > 0$. Assume now $y = 0$. Then it follows that

$$\begin{aligned} h_{ij} &= \lim_{N\to\infty} \sum_{k,l\in S_n} p_{ik}\, b_{kl}^{(N)}\, p_{lj} \\ &= \sum_{k=(c,0)\in S_n} p_{ik} \sum_{l=(d,v)\in S_n} b_{kl}\, p_{lj} \\ &= \sum_{k=(c,0)\in S_n} p_{ik} \left(\tfrac{1}{2} q_c\, p_{(c-1,0),j} - q_c\, p_{kj} + \tfrac{1}{2} q_c\, p_{(c,1),j}\right) \end{aligned}$$

$$
\begin{aligned}
&= p_{ij}(\tfrac{1}{2}q_a p_{(a-1,0),j} - q_a p_{jj} + \tfrac{1}{2}q_a p_{(a,1),j}) \\
&+ p_{i,(a+1,0)}(\tfrac{1}{2}q_{a+1} p_{(a,0),j} - q_{a+1} p_{(a+1,0),j} + \tfrac{1}{2}q_{a+1} p_{(a+1,1),j}) \\
&= p_{ij}(0 - q_a + \tfrac{1}{2}q_a q) + p_{i,(a+1,0)}(\tfrac{1}{2}q_{a+1} - 0 + \tfrac{1}{2}q_{a+1} p) \\
&= p_{ij} q_a(\tfrac{1}{2}q - 1) + \tfrac{1}{2}(p+1)q_{a+1} p_{i,(a+1,0)} \\
&= \frac{1}{2-s}(q_{a+1} p_{i,(a+1,0)} - q_a p_{ij}) \\
&= \frac{1}{2-s}(q_{a+1} B(x, p, b-a-1) - q_a B(x, p, b-a)).
\end{aligned}
$$

As for the case $n = 2$, it follows from Lemma 1 that $\lim_{N\to\infty} \Pi_N^{[(2-s)Nt]} = P - I + \mathrm{e}^{tG} =: \Pi(t)$, where $G := (2-s)H$. Thus the following theorem holds.

Theorem 2. *The finite-dimensional distributions of $(\mathcal{D}_{[(2-s)Nt]})_{t\geq 0}$ converge to those of a continuous time Markov process $(D_t)_{t\geq 0}$, with transition matrix $\Pi(t) = P - I + \mathrm{e}^{tG} = P\mathrm{e}^{tG}$, $t > 0$, where the entries of P are given by (5), and the entries of the infinitesimal generator G are given by*

$$
\begin{gathered}
g_{ij} = q_{a+1} B(x, p, b-a-1) - q_a B(x, p, b-a), \\
i = (b, x),\ j = (a, y) \in S_n,
\end{gathered}
$$

where

$$
\begin{aligned}
q_a &:= a(a-1)/2, \\
p &:= s/(2-s)
\end{aligned}
$$

and

$$
B(x, p, k) := \binom{x}{k} p^k (1-p)^{x-k}.
$$

Remarks.

1. For the special case $x = 0$ the entries of the infinitesimal generator G are given by

$$
g_{ij} = \begin{cases} -q_b & \text{if } y = 0 \text{ and } a = b, \\ q_b & \text{if } y = 0 \text{ and } a = b-1, \\ 0 & \text{otherwise.} \end{cases} \tag{7}
$$

   These are exactly the entries of the infinitesimal generator of the death process of the usual n-coalescent. Thus it is reasonable to call the process $(D_t)_{t\geq 0}$ the death process of the n-coalescent with selfing probability s. (It is shown later that there exists a process, called the n-coalescent with selfing probability s.)

2. From $\Pi(0+) := \lim_{t\searrow 0} \Pi(t) = P$ it follows that the process moves instantaneously from a state i to a state j with probability p_{ij} given by (5). Then the process will be in a state of the form $j = (a, 0)$ with $a \in \{1, \ldots, n\}$. From (7) it follows that after reaching such a state $j = (a, 0)$ the process behaves like the death process of the usual a-coalescent.

3.3. The n-coalescent with selfing

So far only the number of ancestral genes has been studied which leads to the so-called death process of the n-coalescent with selfing. Now the n-coalescent with selfing will be introduced.

The n-coalescent [6, 7, 8] is a Markov chain with state space $\mathcal{E}_n$, the set of all equivalence relations on $\{1, \dots, n\}$. A typical state $\xi \in \mathcal{E}_n$ can be written in the form $\xi = \{C_1, \dots, C_b\}$, where $C_1, \dots, C_b$ are the equivalence classes of ξ and $b = |\xi|$ denotes the number of equivalence classes of ξ. By definition, i and j belong to the same equivalence class of ξ, if and only if the genes i and j have a common ancestor (identical by descent) in generation r, i.e. r generations backwards in time.

As the diploid population model considered here is more complex, some additional information is needed in order to analyse the backward structure. Recall that an individual is of type $k \in \{0, 1, 2\}$, if exactly k of the two genes of this individual belong to the set of ancestral genes. Looking now r generations backwards in time, one can distinguish between ancestral genes belonging to individuals of type 2 and ancestral genes belonging to individuals of type 1. Thus the typical state of a new backward process $(\mathcal{C}_r)_{r\in\mathbb{N}_0}$, with more detailed information, can be written in the form

$$\xi = \big\{\{C_1, C_2\}, \dots, \{C_{2x-1}, C_{2x}\}, C_{2x+1}, \dots, C_b\big\}, \tag{8}$$

where the classes $C_1, \dots, C_b$ are given as before and $x := \|\xi\| \in \{0, \dots, [b/2]\}$ denotes the number of individuals of type 2. Obviously $\mathcal{E}_n$ is a subset of S_n. Note that for given classes $C_1, \dots, C_b$ and for given $x \in \{0, \dots, [b/2]\}$ there exist exactly

$$\frac{1}{x!}\prod_{k=0}^{x-1}\frac{(b-2k)(b-2k-1)}{2} = \frac{(b)_{2x}}{2^x\,x!}$$

states ξ of the form (8). Hence the state space S_n has size

$$|S_n| = \sum_{\xi\in S_n} 1 = \sum_{b=1}^{n}\sum_{\substack{\xi\in\mathcal{E}_n\\|\xi|=b}}\sum_{x=0}^{[b/2]}\frac{(b)_{2x}}{2^x\,x!} = \sum_{b=1}^{n} S(n,b)\sum_{x=0}^{[b/2]}\frac{(b)_{2x}}{2^x\,x!},$$

where $S(n, b)$ denotes the Stirling numbers of the second kind. The values of $|S_n|$ are shown below.

n	$\lvert S_n\rvert$	n	$\lvert S_n\rvert$
1	1	6	1 539
2	3	7	10 299
3	11	8	75 905
4	49	9	609 441
5	257	10	5 284 451

The initial distribution of the process $(\mathcal{C}_r)_{r\in\mathbb{N}_0}$ can be calculated using (2). For $\xi \in S_n$ with $|\xi| = n$ it follows that,

$$\begin{aligned} P(\mathcal{C}_0 = \xi) &= p(N,n,x)\,\frac{2^x\,x!}{(n)_{2x}} = \frac{2^{n-x}\,(N)_{n-x}}{(2N)_n} \\ &= \prod_{k=0}^{n-x-1}\frac{2N-2k}{2N-k}\prod_{k=n-x}^{n-1}\frac{1}{2N-k}, \\ &= \begin{cases} 1-q_n/(2N)+O(N^{-2}) & \text{if } x=0, \\ 1/(2N)+O(N^{-2}) & \text{if } x=1, \\ O(N^{-2}) & \text{if } x\geq 2, \end{cases} \end{aligned}$$

where $q_n = n(n-1)/2$ and $x := \|\xi\|$ is the number of individuals of type 2 in the current generation 0. Note that for large N the process starts with high probability in the state $\xi = \Delta := \{\{1\},\dots,\{n\}\}$, i.e. if N is large, the n sampled genes belong to different individuals with high probability.

Let $\pi_{\xi\eta} := P(\mathcal{C}_r = \eta \,|\, \mathcal{C}_{r-1} = \xi)$ denote the transition probabilities of the process $(\mathcal{C}_r)_{r\in\mathbb{N}_0}$. Transitions from

$$\xi = \{\{C_1, C_2\},\dots,\{C_{2x-1}, C_{2x}\}, C_{2x+1},\dots, C_b\}$$

to

$$\eta = \{\{D_1, D_2\},\dots,\{D_{2y-1}, D_{2y}\}, D_{2y+1},\dots, D_a\}$$

occur with probability

$$\pi_{\xi\eta} = (1-s)^{x-y-(b-a)}(s/2)^{y+b-a} + O(N^{-1}), \tag{9}$$

if $\xi \sqsubseteq \eta$, and with probability $\pi_{\xi\eta} = O(N^{-1})$ otherwise; $\xi \sqsubseteq \eta$ denotes that there occur $x-y-(b-a)$ non-selfing events $(\{C_k, C_l\} \mapsto C_k, C_l)$, $b-a$ selfing events with coalescence $(\{C_k, C_l\} \mapsto C_k \cup C_l)$ and y selfing events with no coalescence $(\{C_k, C_l\} \mapsto \{C_k, C_l\})$, i.e. $D_i := C_i$ for all $i \in \{1,\dots,2y\}$, $D_i := C_{2i-2y+1} \cup C_{2i-2y}$ for all $i \in \{2y+1,\dots,2y + (b-a)\}$ and $D_i := C_{b-a+i}$ for all $i \in \{2y+(b-a)+1,\dots,a\}$.

The exact form of the transition probabilities $\pi_{\xi\eta}$ is not very important except for the case $\|\xi\| = 0$, where it can be shown that

$$\begin{aligned} \pi_{\xi\eta} &= \frac{(2N)_a}{(2N)^b}\,p(N,a,y)\,\frac{2^y\,y!}{(a)_{2y}} = \frac{2^{a-y}\,(N)_{a-y}}{(2N)^b}, \\ &= \begin{cases} 1-q_b/N+O(N^{-2}) & \text{if } \xi=\eta, \\ 1/(2N)+O(N^{-2}) & \text{if } \xi \prec \eta \text{ or } \xi \rightsquigarrow \eta, \\ O(N^{-2}) & \text{otherwise,} \end{cases} \end{aligned} \tag{10}$$

where

$$\begin{aligned} \xi \prec \eta &:\Longleftrightarrow \xi \subseteq \eta,\ |\xi| = |\eta|+1 \text{ and } \|\xi\| = \|\eta\|, \\ \xi \rightsquigarrow \eta &:\Longleftrightarrow \xi \subseteq \eta,\ |\xi| = |\eta| \text{ and } \|\xi\| = \|\eta\|-1, \end{aligned}$$

and the notation $\xi \subseteq \eta$ is used for the case when each class of ξ is a subset of a class of η.

From (9) it follows that the transition matrix $\Pi_N := (\pi_{\xi\eta})_{\xi,\eta\in S_n}$ has a decomposition of the form

$$\Pi_N = A + \frac{1}{N} B_N, \tag{11}$$

where $B_N := N(\Pi_N - A)$ defines a bounded matrix-sequence $(B_N)_{N\in\mathbb{N}}$ and the entries of $A := \lim_{N\to\infty} \Pi_N$ are given by $a_{\xi\eta} = (1-s)^{x-y-(b-a)}(s/2)^{y+b-a}$, if $\xi \sqsubseteq \eta$ and $a_{\xi\eta} = 0$ otherwise. Note that for given $\xi \in S_n$, given a and given y, there exist exactly

$$\binom{x}{b-a}\binom{x-(b-a)}{y}$$

states $\eta \in S_n$ with $|\eta| = a$ and $\|\eta\| = y$ such that $\xi \sqsubseteq \eta$. Hence

$$\begin{aligned}
\sum_{\eta\in S_n} a_{\xi\eta} &= \sum_{a=b-x}^{b} \binom{x}{b-a} \sum_{y=0}^{x-(b-a)} \binom{x-(b-a)}{y} (1-s)^{x-y-(b-a)} \left(\frac{s}{2}\right)^{y+b-a} \\
&= \sum_{a=b-x}^{b} \binom{x}{b-a} \left(\frac{s}{2}\right)^{b-a} \sum_{y=0}^{x-(b-a)} \binom{x-(b-a)}{y} (1-s)^{x-y-(b-a)} \left(\frac{s}{2}\right)^{y} \\
&= \sum_{a=b-x}^{b} \binom{x}{b-a} \left(\frac{s}{2}\right)^{b-a} \left(1-s+\frac{s}{2}\right)^{x-(b-a)} \\
&= \sum_{c=0}^{x} \binom{x}{c} \left(\frac{s}{2}\right)^{c} \left(1-\frac{s}{2}\right)^{x-c} = 1,
\end{aligned}$$

i.e. A is a stochastic matrix, which is also directly clear from (11) by taking $N \to \infty$. It is not easy to find closed forms for all the entries $b_{\xi\eta}^{(N)}$ of B_N. For the special case $\xi \in S_n$ with $\|\xi\| = 0$ it follows from (10) that

$$b_{\xi\eta} := \lim_{N\to\infty} b_{\xi\eta}^{(N)} = \begin{cases} -q_b & \text{if } \xi = \eta, \\ 1/2 & \text{if } \xi \prec \eta \text{ or } \xi \rightsquigarrow \eta, \\ 0 & \text{otherwise.} \end{cases}$$

As A is a stochastic matrix there exists a Markov process $(Z_m)_m$ with corresponding transition matrix A. Note that $a_{\xi\xi} = (s/2)^{\|\xi\|}$ for all $\xi \in S_n$ and that $a_{\xi\xi} = 1$ if and only if $\|\xi\| = 0$, i.e. exactly the states $\xi \in \mathcal{E}_n$ are the absorbing states of the process $(Z_m)_m$. For $\xi, \eta \in S_n$ define $p_{\xi\eta} := P(Z_m = \eta \text{ finally} \mid Z_0 = \xi)$ and $P := (p_{\xi\eta})_{\xi,\eta\in S_n}$. Note that P satisfies the equation $P = AP$. The solution of this equation is given by

$$p_{\xi\eta} = \begin{cases} p^{b-a} q^{x-(b-a)} & \text{if } \|\eta\| = 0 \text{ and } \xi \sqsubseteq \eta, \\ 0 & \text{otherwise,} \end{cases} \tag{12}$$

where $x := \|\xi\|, a := |\eta|, b := |\xi|$ and p and q are given by (1). In order to verify (12), one has only to prove that the $p_{\xi\eta}$ given in (12) solve the equation $P = AP$, i.e. $\sum_{\nu\in S_n} a_{\xi\nu} p_{\nu\eta} = p_{\xi\eta}$

for all $\xi, \eta \in S_n$. This is obviously true except for the case that $\|\eta\| = 0$ and $\xi \sqsubseteq \eta$. In this case it follows that

$$\begin{aligned}
&\sum_{\nu \in S_n} a_{\xi\nu}\, p_{\nu\eta} \\
&\quad = \sum_{c=a}^{b} \binom{b-a}{c-a} \sum_{z=c-a}^{x-(b-c)} \binom{x-(b-a)}{z-(c-a)} (1-s)^{x-(b-c)-z} \left(\frac{s}{2}\right)^{z+b-c} p^{c-a}\, q^{z-(c-a)} \\
&\quad \overset{(6)}{=} p^{b-a} q^{x-(b-a)} (1+p)^{-x} \sum_{c=a}^{b} \binom{b-a}{c-a} \sum_{z=c-a}^{x-(b-c)} \binom{x-(b-a)}{z-(c-a)} p^{z} \\
&\quad = p_{\xi\eta}\, (1+p)^{-x} \sum_{c=a}^{b} \binom{b-a}{c-a} \sum_{l=0}^{x-(b-a)} \binom{x-(b-a)}{l} p^{l+(c-a)} \\
&\quad = p_{\xi\eta}\, (1+p)^{-x} \sum_{c=a}^{b} \binom{b-a}{c-a} p^{c-a} (1+p)^{x-(b-a)} \\
&\quad = p_{\xi\eta}\, (1+p)^{-(b-a)} \sum_{l=0}^{b-a} \binom{b-a}{l} p^{l} = p_{\xi\eta},
\end{aligned}$$

and (12) is proven. Using the same argument as the previous section it follows that

$$\lim_{m \to \infty} A^m = P.$$

The calculation of the entries $h_{\xi\eta}$ of the matrix $H := \lim_{N\to\infty} P B_N P$ is somewhat technical. Obviously $h_{\xi\eta} = 0$ except for the case that $\|\eta\| = 0$ and $\xi \subseteq \eta$. In this case it follows that

$$\begin{aligned}
h_{\xi\eta} &= \lim_{N\to\infty} \sum_{\nu,\mu \in S_n} p_{\xi\nu}\, b^{(N)}_{\nu\mu}\, p_{\mu\eta} = \sum_{\substack{\nu,\mu \in S_n \\ \|\nu\|=0}} p_{\xi\nu}\, b_{\nu\mu}\, p_{\mu\eta}, \\
&= \underbrace{\frac{1}{2} \sum_{\substack{\nu,\mu \in S_n \\ \nu \prec \mu}} p_{\xi\nu}\, p_{\mu\eta}}_{=: I} + \underbrace{\frac{1}{2} \sum_{\substack{\nu,\mu \in S_n \\ \nu \rightsquigarrow \mu}} p_{\xi\nu}\, p_{\mu\eta}}_{=: II} - q_a\, p_{\xi\eta}.
\end{aligned}$$

In the first sum only the state $\mu = \eta$ provides a contribution, i.e.

$$I = \sum_{\substack{\nu \in S_n \\ \nu \prec \eta}} p_{\xi\nu}.$$

The second sum can be split up into two parts, where the first part includes the summation over all ν with $|\nu| = a$ $(:= |\eta|)$ and the second part includes the summation over all ν with $|\nu| = a + 1$. For the first part, only $\nu = \eta$ provides a contribution, i.e.

$$\sum_{\substack{\nu,\mu \in S_n \\ \nu \rightsquigarrow \mu \\ |\nu|=a}} p_{\xi\nu}\, p_{\mu\eta} = \sum_{\substack{\mu \in S_n \\ \eta \rightsquigarrow \mu}} p_{\xi\eta}\, p_{\mu\eta} = p_{\xi\eta}\, q\, q_a.$$

For the second part, there is only one μ with $\mu \prec \eta$ which provides a contribution, i.e.

$$\sum_{\substack{\nu,\mu\in S_n\\ \nu\rightsquigarrow\mu\\ |\nu|=a+1}} p_{\xi\nu}\, p_{\mu\eta} = p \sum_{\substack{\nu\in S_n\\ \nu\prec\eta}} p_{\xi\nu} = p\cdot I.$$

Thus $II = p\cdot I + p_{\xi\eta}\, q\, q_a$ and hence

$$h_{\xi\eta} = \tfrac{1}{2}\cdot I + \tfrac{1}{2}\cdot II - q_a\, p_{\xi\eta} = \frac{1+p}{2} I + (\tfrac{1}{2}q - 1) q_a\, p_{\xi\eta} = \frac{1}{2-s}\Bigg(\sum_{\substack{\nu\in S_n\\ \nu\prec\eta}} p_{\xi\nu} - q_a\, p_{\xi\eta}\Bigg).$$

Now apply Lemma 1 to verify the following theorem.

Theorem 3. *The finite-dimensional distributions of $(\mathcal{C}_{[(2-s)Nt]})_{t\geq 0}$ converge to those of a continuous-time Markov process $(C_t)_{t\geq 0}$, with transition matrix $\Pi(t) = P - I + \mathrm{e}^{tG} = P\mathrm{e}^{tG}$, $t > 0$, where the entries of P are given by (12) and the entries of the infinitesimal generator G are given by*

$$g_{\xi\eta} = \sum_{\substack{\nu\in S_n\\ \nu\prec\eta}} p_{\xi\nu} - \frac{|\eta|(|\eta|-1)}{2}\, p_{\xi\eta} \qquad \forall\, \xi, \eta \in S_n.$$

Remarks.

1. For $\xi \in S_n$ with $\|\xi\| = 0$, the entries of the generator G are given by

$$g_{\xi\eta} = \begin{cases} -|\xi|(|\xi|-1)/2 & \text{if } \xi = \eta,\\ 1 & \text{if } \xi \prec \eta,\\ 0 & \text{otherwise.}\end{cases} \tag{13}$$

   These are exactly the entries of the infinitesimal generator of the usual n-coalescent. Thus it makes sense to call $(C_t)_{t\geq 0}$ the n-coalescent with selfing probability s.

2. From $\Pi(0+) := \lim_{t\searrow 0}\Pi(t) = P$ it follows that the process moves instantaneously from the state ξ to the state η, with probability $p_{\xi\eta}$ given by (12). Then the process will be in a state η with $\|\eta\| = 0$. From (13) it follows that after reaching such a state η, the process behaves like the usual $|\eta|$-coalescent. Especially, given $C_0 \equiv \Delta$, the process behaves like the usual n-coalescent. In other words all the states $\xi \in S_n$ with $\|\xi\| > 0$ are instantaneous. Eliminating all these states leads to the usual n-coalescent.

3.4. Overlying mutations

Assume that mutations occur with probability μ per gene per generation. The state space of the ancestral process then becomes more complex, as non-mutant and mutant equivalence classes are distinguished. The non-mutant classes are sometimes also called 'old' classes and the mutant ones 'new' classes. Proceeding one generation backwards in time, the following two types of events may occur (perhaps several times).

1. Some of the non-mutant classes coalesce to form a 'larger' non-mutant class.

2. A non-mutant class becomes a mutant class.

More precisely, the transition probabilities are given by $\pi_{\xi\eta}(\mu) = \mu^m(1-\mu)^b\pi_{\xi\eta}$, where m is the number of mutations, $b+m$ is the number of old equivalence-classes of ξ and the $\pi_{\xi\eta} = a_{\xi\eta} + b_{\xi\eta}^{(N)}/N$ are the transition probabilities of the corresponding process with no mutations (see (9) or (11)). If $\theta := \lim_{N\to\infty} 4N\mu$ exists, then it follows that

$$\pi_{\xi\eta}(\mu) = \begin{cases} a_{\xi\eta} + (b_{\xi\eta}^{(N)} - \theta\, b\, a_{\xi\eta}/4)N^{-1} + o(N^{-1}) & \text{if } m = 0, \\ \theta\, a_{\xi\eta}/(4N) + o(N^{-1}) & \text{if } m = 1, \\ o(N^{-1}) & \text{if } m \geq 2. \end{cases}$$

Thus the transition probabilities are of the form

$$\pi_{\xi\eta}(\mu) = a_{\xi\eta}(\mu) + \frac{1}{N}\, b_{\xi\eta}^{(N)}(\mu),$$

where

$$a_{\xi\eta}(\mu) := \begin{cases} a_{\xi\eta} & \text{if } m = 0, \\ 0 & \text{if } m \geq 1, \end{cases}$$

and $b_{\xi\eta}^{(N)}(\mu) := N(\pi_{\xi\eta}(\mu) - a_{\xi\eta}(\mu))$. For the special case $\xi \in S_n$ with $\|\xi\| = 0$ it follows that

$$b_{\xi\eta}(\theta) := \lim_{N\to\infty} b_{\xi\eta}^{(N)}(\mu) = \begin{cases} b_{\xi\eta} - \theta\, b\, a_{\xi\eta}/4 & \text{if } m = 0, \\ \theta\, a_{\xi\eta}/4 & \text{if } m = 1, \\ 0 & \text{if } m \geq 2. \end{cases}$$

In order to calculate the entries $h_{\xi\eta}(\theta)$ of $H(\theta) := \lim_{N\to\infty} P(\mu)B_N(\mu)P(\mu)$ consider first the case that there occurs exactly one mutation while the process jumps from ξ to η. Then it follows that

$$\begin{aligned} h_{\xi\eta}(\theta) &= \lim_{N\to\infty} \sum_{\nu,\mu} p_{\xi\nu}(\mu)\, b_{\nu\mu}^{(N)}(\mu)\, p_{\mu\eta}(\mu) \\ &= \sum_{\nu,\mu} p_{\xi\nu}\, (\theta\, a_{\nu\mu}/4)\, p_{\mu\eta} \\ &= \frac{1}{4}\theta \sum_{\nu,\mu} p_{\xi\nu}\, a_{\nu\mu}\, p_{\mu\eta} = \frac{1}{4}\theta \sum_{\nu} p_{\xi\nu}\, p_{\nu\eta} = \frac{1}{4}\theta p_{\xi\eta}. \end{aligned}$$

Now assume that there occurs no mutation while proceeding from ξ to η. Then

$$\begin{aligned} h_{\xi\eta}(\theta) &= \lim_{N\to\infty} \sum_{\nu,\mu} p_{\xi\nu}(\mu)\, b_{\nu\mu}^{(N)}(\mu)\, p_{\mu\eta}(\mu) \\ &= \sum_{\nu,\mu} p_{\xi\nu}\, (b_{\nu\mu} - \theta c\, a_{\nu\mu}/4)\, p_{\mu\eta} \\ &= h_{\xi\eta} - \frac{1}{4}\theta \sum_{\nu} p_{\xi\nu}\, c\, p_{\nu\eta} = h_{\xi\eta} - \tfrac{1}{4}\theta a\, p_{\xi\eta}. \end{aligned}$$

Here c denotes the number of old classes of ν and a the number of old classes of η. Multiplication with $2/(p+1) = 2 - s$ leads to

$$g_{\xi\eta}(\theta) := (2-s)h_{\xi\eta}(\theta) = \begin{cases} g_{\xi\eta} - \frac{1}{2}\tilde{\theta} a\, p_{\xi\eta} & \text{if } m = 0, \\ \frac{1}{2}\tilde{\theta}\, p_{\xi\eta} & \text{if } m = 1, \\ 0 & \text{if } m \geq 2, \end{cases}$$

where $\tilde{\theta} := \theta(1 - s/2)$. Note that for the special case $\xi \in S_n$ with $\|\xi\| = 0$, the entries $g_{\xi\eta}(\theta)$ are given by

$$g_{\xi\eta}(\theta) = \begin{cases} -b(b+\tilde{\theta}-1)/2 & \text{if } \xi = \eta \text{ and } m = 0, \\ 1 & \text{if } \xi \prec \eta \text{ and } m = 0, \\ \tilde{\theta}/2 & \text{if } \xi = \eta \text{ and } m = 1, \\ 0 & \text{otherwise.} \end{cases}$$

These are exactly the entries of the infinitesimal generator of the usual n-coalescent with mutation rate $\tilde{\theta} = \theta(1 - s/2)$.

Remark. Estimating the selfing probability s and the mutation rate θ is of major interest in statistical population genetics. Several methods are known, most of them based on frequency data [1], some others based on DNA sequence data [9], but all these estimators have, in some sense, undesirable statistical properties. For a recent paper focusing on these problems see Nordborg and Donnelly [12].

Appendix

Proof of Lemma 1. Assume $K = 1$ without loss of generality. (Otherwise choose $B' := B/K, c'_N := Kc_N$ and $t' := Kt$.) Fix $t \geq 0$ and $\varepsilon > 0$. Choose $M \in \mathbb{N}$ such that $\|A^m - P\| < \varepsilon$ for all $m \geq M$. For convenience, define $n := [t/c_N]$. Now

$$\|(A + c_N B)^n - (P + c_N B)^n\| \leq \sum_{k=0}^{n} c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k}} \left\| A^{m_1} \prod_{j=2}^{k+1} B A^{m_j} - P^{m_1} \prod_{j=2}^{k+1} B P^{m_j} \right\|.$$

If $m_j \geq M$ for all $j \in \{1, \dots, k+1\}$, then

$$\|A^{m_1} B A^{m_2} B \cdots B A^{m_{k+1}} - P^{m_1} B P^{m_2} B \cdots B P^{m_{k+1}}\| \leq \sum_{j=1}^{k+1} \|A^{m_j} - P^{m_j}\| = \sum_{j=1}^{k+1} \|A^{m_j} - P\| < (k+1)\varepsilon.$$

Otherwise the weaker inequality

$$\begin{aligned} &\|A^{m_1} B A^{m_2} B \cdots B A^{m_{k+1}} - P^{m_1} B P^{m_2} B \cdots B P^{m_{k+1}}\| \\ &\quad \leq \|A^{m_1} B A^{m_2} B \cdots B A^{m_{k+1}}\| + \|P^{m_1} B P^{m_2} B \cdots B P^{m_{k+1}}\|, \\ &\quad \leq 1 + 1 = 2, \end{aligned}$$

is available. Thus it follows that $\|(A+c_N B)^n-(P+c_N B)^n\| \le \|A^n-P\|+S_1+S_2$, where,

$$\begin{aligned}
S_1 &:= \sum_{k=1}^{n} c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\ge M \\ m_1+\cdots+m_{k+1}=n-k}} (k+1)\varepsilon \\
&\le \varepsilon \sum_{k=0}^{n}(k+1)c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k}} 1 = \varepsilon \sum_{k=0}^{n}(k+1)\binom{n}{k}c_N^k \\
&= \varepsilon(nc_N(1+c_N)^{n-1}+(1+c_N)^n) \sim \varepsilon \mathrm{e}^t(t+1);
\end{aligned}$$

(note that $\sum_{k=0}^{n}\binom{n}{k}x^k=(1+x)^n$ and $\sum_{k=0}^{n}k\binom{n}{k}x^k=nx(1+x)^{n-1}$)
and

$$\begin{aligned}
S_2 &:= \sum_{k=1}^{n} c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k \\ \exists j \text{ with } m_j<M}} 2 \\
&\le \sum_{k=1}^{n} c_N^k (k+1) \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k \\ m_{k+1}<M}} 2 \\
&\le \sum_{k=1}^{n} c_N^k (k+1) \sum_{m_{k+1}=0}^{M-1} \sum_{\substack{m_1,\dots,m_k\in\mathbb{N}_0 \\ m_1+\cdots+m_k=n-m_{k+1}-k}} 2 \\
&= \sum_{k=1}^{n} c_N^k (k+1) \sum_{m_{k+1}=0}^{M-1} 2\binom{n-m_{k-1}-1}{k-1} \\
&\le 2M\sum_{k=1}^{n}(k+1)\binom{n-1}{k-1}c_N^k = 2Mc_N\sum_{k=0}^{n-1}(k+2)\binom{n-1}{k}c_N^k \\
&= 2Mc_N((n-1)c_N(1+c_N)^{n-2}+2(1+c_N)^{n-1}) \\
&\sim 2Mc_N\mathrm{e}^t(t+2) = O(c_N).
\end{aligned}$$

As ε can be chosen arbitrarily the first part of Lemma 1 is established. Now fix $t>0$ and assume that N is large such that $n=[t/c_N]>0$. If $G:=\lim_{N\to\infty} PB_NP$ exists, then

$$\begin{aligned}
&(P+c_N B_N)^n - P + I - \mathrm{e}^{tG} \\
&\quad = P^n + \sum_{k=1}^{n} c_N^k \Bigg(\sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k}} P^{m_1}\prod_{j=2}^{k+1} B_N P^{m_j} \Bigg) - P + I - \sum_{k=0}^{\infty}\frac{t^k}{k!}G^k
\end{aligned}$$

$$= \sum_{k=1}^{n} \Bigg(c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k}} P^{m_1} \prod_{j=2}^{k+1} B_N P^{m_j} - \frac{t^k}{k!} G^k \Bigg) - \sum_{k=n+1}^{\infty} \frac{t^k}{k!} G^k$$

$$= \sum_{k=1}^{n} c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k}} \Bigg(P^{m_1} \prod_{j=2}^{k+1} B_N P^{m_j} - (P B_N P)^k \Bigg)$$

$$+ \sum_{k=1}^{n} \Bigg(c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k}} (P B_N P)^k - \frac{t^k}{k!} G^k \Bigg) - \sum_{k=n+1}^{\infty} \frac{t^k}{k!} G^k.$$

From $P^2 = P$ it follows that $P^{m_1} B_N P^{m_2} \cdots B_N P^{m_{k+1}} = (P B_N P)^k$ if $m_j > 0$ for all $j \in \{1, \dots, k+1\}$. Hence

$$(P + c_N B_N)^n - P + I - \mathrm{e}^{tG}$$

$$= \sum_{k=1}^{n} c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k \\ \exists j \text{ with } m_j=0}} (P^{m_1} B_N P^{m_2} \cdots B_N P^{m_{k+1}} - (P B_N P)^k)$$

$$+ \sum_{k=1}^{n} \Bigg(c_N^k \binom{n}{k} (P B_N P)^k - \frac{t^k}{k!} G^k \Bigg) - \sum_{k=n+1}^{\infty} \frac{t^k}{k!} G^k.$$

The norm of the first sum is not larger than

$$\sum_{k=1}^{n} c_N^k \sum_{\substack{m_1,\dots,m_{k+1}\in\mathbb{N}_0 \\ m_1+\cdots+m_{k+1}=n-k \\ \exists j \text{ with } m_j=0}} 2 \leq \sum_{k=1}^{n} c_N^k (k+1) \sum_{\substack{m_1,\dots,m_k\in\mathbb{N}_0 \\ m_1+\cdots+m_k=n-k}} 2$$

$$= 2 \sum_{k=1}^{n} (k+1) \binom{n-1}{k-1} c_N^k = 2 c_N \sum_{k=0}^{n-1} (k+2) \binom{n-1}{k} c_N^k$$

$$= 2 c_N ((n-1) c_N (1+c_N)^{n-2} + 2(1+c_N)^{n-1})$$

$$\sim 2 c_N \mathrm{e}^t (t+2) = O(c_N).$$

The last sum, $\sum_{k=n+1}^{\infty} (tG)^k / k!$, is the tail of the exponential series e^{tG}. Hence this sum converges to zero as N tends to infinity. To finish the proof we show that

$$\lim_{N\to\infty} \sum_{k=1}^{n} \Bigg(c_N^k \binom{n}{k} (P B_N P)^k - \frac{t^k}{k!} G^k \Bigg) = 0.$$

Obviously $c_N^k (n)_k \leq (c_N n)^k \leq t^k$ and from $t - c_N = c_N (t/c_N - 1) < c_N n$, it follows that

$$c_N^k (n)_k = \prod_{i=0}^{k-1} c_N (n-i) > \prod_{i=0}^{k-1} t - c_N (i+1)$$

$$= \prod_{i=1}^{k}(t - c_N i) \geq t^k - t^{k-1}c_N(1 + \cdots + k)$$
$$= t^k - t^{k-1}c_N k(k+1)/2 = t^k - (c_N/2)k(k+1)t^{k-1}.$$

Thus

$$\left| c_N^k \binom{n}{k} - \frac{t^k}{k!} \right| \leq (c_N/2)(k+1)t^{k-1}/(k-1)!$$

and

$$\begin{aligned}
&\left\| \sum_{k=1}^{n} c_N^k \binom{n}{k} (PB_N P)^k - \frac{t^k}{k!} G^k \right\| \\
&\qquad \leq \sum_{k=1}^{n} \left| c_N^k \binom{n}{k} - \frac{t^k}{k!} \right| \|(PB_N P)^k\| + \sum_{k=1}^{n} \frac{t^k}{k!} \|(PB_N P)^k - G^k\| \\
&\qquad \leq \sum_{k=1}^{n} \left| c_N^k \binom{n}{k} - \frac{t^k}{k!} \right| + \sum_{k=1}^{n} \frac{t^k}{k!} k \|PB_N P - G\| \\
&\qquad \leq \sum_{k=1}^{n} \frac{c_N}{2}(k+1)\frac{t^{k-1}}{(k-1)!} + \|PB_N P - G\| \sum_{k=1}^{n} \frac{t^k}{(k-1)!} \\
&\qquad \leq \frac{c_N}{2} \sum_{k=0}^{n-1} (k+2)\frac{t^k}{k!} + \|PB_N P - G\| \sum_{k=0}^{n-1} \frac{t^{k+1}}{k!} \\
&\qquad \leq \frac{c_N}{2}(t+2)\mathrm{e}^t + \|PB_N P - G\| \, t\mathrm{e}^t,
\end{aligned}$$

which converges to zero as N tends to infinity.

Acknowledgements

The author wishes to thank Professor Peter Donnelly and Dr. Magnus Nordborg for many helpful discussions and comments and Professor Wolgang Bühler and Professor Hans-Jürgen Schuh for their general assistance. The author was supported by the 'Deutsche Forschungsgemeinschaft' (DFG) during the preparation of this article.

References

[1] BROWN, A. H. D. (1990). Genetic characterization of plant mating systems. In *Plant Population Genetics, Breeding, and Genetic Resources*, eds. A. H. D. Brown, M. T. Clegg, A. L. Kahler, B. S. Weir. Sinauer Associates Inc., Sunderland, Massachusetts, pp. 145–162.

[2] CLEGG, M. T. (1980). Measuring plant mating systems. *BioScience* **30**, 814–818.

[3] DONNELLY, P. AND TAVARÉ, S. (1995). Coalescents and genealogical structure under neutrality. *A. Rev. Genet.* **29**, 401–421.

[4] GRIFFITHS, R. C. AND MAJORAM, P. (1997). An ancestral recombination graph. In *Progress in Population Genetics and Human Evolution,* ed. P. Donnelly. Springer, pp. 257–270.

[5] HUDSON, R. R. AND KAPLAN, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics* **120**, 831–840.

[6] KINGMAN, J. F. C. (1982). On the genealogy of large populations. *J. Appl. Prob.* **19A**, 27–43.

[7] KINGMAN, J. F. C. (1982). Exchangeability and the evolution of large populations. In *Exchangeability in Probability and Statistics*, eds. G. Koch, F. Spizzichino. North-Holland Publishing Company, pp. 97–112.

[8] KINGMAN, J. F. C. (1982). The coalescent. *Stoch. Process. Appl.* **13**, 235–248.
[9] MILLIGAN, B. G. (1996). Estimating long-term mating systems using DNA sequences. *Genetics* **142**, 619–627.
[10] MÖHLE, M. Robustness results for the coalescent. *J. Appl. Prob.* **35**, 437–446.
[11] MÖHLE, M. Coalescent results for two-sex population models. *Adv. Appl. Prob.* **30**, 513–520.
[12] NORDBORG, M. AND DONNELLY, P. The coalescent process with selfing. *Genetics* **146**, 1185–1195.
[13] POLLAK, E. (1987). On the theory of partially inbreeding finite populations. I. Partial selfing. *Genetics* **117**, 353–360.
[14] TAVARÉ, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetics models. *Theor. Pop. Biol.* **26**, 119–164.
[15] WRIGHT, S. (1969). *Evolution and the Genetics of Populations*. Volume 2. University of Chicago Press, Chicago.