CONVERGENCE TO THE COALESCENT IN POPULATIONS OF SUBSTANTIALLY VARYING SIZE

PETER JAGERS ∗∗ AND
SERIK SAGITOV,∗ Chalmers University of Technology and Göteborg University

Abstract

Kingman’s classical theory of the coalescent uncovered the basic pattern of genealogical trees of random samples of individuals in large but time-constant populations. Time is viewed as being discrete and is identified with non-overlapping generations. Reproduction can be very generally taken as exchangeable (meaning that the labelling of individuals in each generation carries no significance). Recent generalisations have dealt with population sizes exhibiting given deterministic or (minor) random fluctuations. We consider population sizes which constitute a stationary Markov chain, explicitly allowing large fluctuations in short times. Convergence of the genealogical tree, as population size tends to infinity, towards the (time-scaled) coalescent is proved under minimal conditions. As a result, we obtain a formula for effective population size, generalising the well-known harmonic mean expression for effective size.

Keywords: Coalescent; exchangeability; population genetics

2000 Mathematics Subject Classification: Primary 92F25; 60J70
Secondary 92D15; 60F17

1. Introduction

The purpose of an evolutionary tree, based on a set of homologous biological sequences, is to reveal genealogical ties between the sequences (the tree topology) and to estimate divergence times (the lengths of the tree branches). There are a number of methods of tree reconstruction which yield estimators of both topology and branch lengths [19]. One approach to inference about the evolutionary processes that shaped gene genealogies relies upon the so-called coalescent, introduced by Kingman [12].

The coalescent is a simple stochastic process which describes the genealogical tree under assumptions of neutrality, no recombination, random mating, and a population size which is large (tends to infinity) but constant from generation to generation. Taken together, these four assumptions can be said to form a sort of null hypothesis for the analysis of the evolutionary features of a biological population.

The probability distribution of the coalescent tree is simple: any pair of branches followed backward in time may be the first to join together. Branch lengths are predicted to have huge variability, the main rule being that the fewer branches there remain in the tree, the longer the branch lengths. The branch-joining rate of the Kingman coalescent is equal to the number of pairs of branches.
When compared to the coalescent, the shape of a given tree might bear witness to various violations of the null hypothesis. For example, exponentially growing populations are characterised by star-like trees for all loci \cite{4}, \cite{7}. Positive selection and constant population size, on the other hand, would result in a star-like tree only for the selected locus and those linked to it. A third typical shape should be expected with population subdivision: for all loci, mergers occur faster within subpopulations, and this renders the branch lengths near the root (which correspond to mergers between subpopulations) longer than those predicted by the Kingman coalescent. A similar picture is to be expected for one locus (and those linked to it) in the case of balanced selection.

The coalescent was first obtained in \cite{12} as a large size approximation for the ancestral tree of the Wright–Fisher (WF) population model. The latter is a population caricature with non-overlapping generations and symmetric multinomial reproduction law (see Section 2). Reproduction is assumed to be asexual (haploid). The WF model of size $N$ (i.e. for populations of size $N$) requires the time-scale parameter $T_N = N$ for the basic coalescent approximation. The validity of the Kingman coalescent was considerably widened in \cite{11} and \cite{13}, where convergence towards it was established for a broad class of exchangeable reproduction distributions. The technical complication arising in the diploid case were described and handled in \cite{15}. In general, the proper time scale for the genealogy of a large population turns out to be $T_N = c_N^{-1}$, where $c_N$ is the probability that a random pair of individuals belonging to the same generation are siblings, and $N$ continues to denote the population size, still assumed constant.

Another important restriction, actually needed for the Kingman coalescent approximation, is a bound on the variation of the offspring numbers $(\nu_1, \ldots, \nu_N)$. It can be expressed either in terms of moments of the offspring numbers or else in terms of the tail distribution of the largest offspring number $\nu(1)$ observed in a single generation. According to \cite{18}, a necessary and sufficient condition for the Kingman coalescent approximation to be valid for an exchangeable reproduction model of size $N$ is that $P(\nu(1) > N\epsilon) = o(c_N)$ as $N \rightarrow \infty$ for $\epsilon > 0$. If instead the tail probability $P(\nu(1) > N\epsilon)$ is proportional to $c_N$, the asymptotic tree topology admits multiple mergers of branches. Furthermore, if the second largest offspring number also has such a heavy tail, then several (multiple) mergers might occur simultaneously.

The impact of variable population size was investigated in \cite{7}, \cite{13}, and \cite{14} under the assumption that the historical sizes are known (see also the review \cite{4} and references there). It was found that deterministic size variability results in a deterministic (nonlinear) time change $t \mapsto \lambda(t)$ of the Kingman coalescent reflecting the fluctuations of effective population size (see Section 3). Unknown past sizes $\{M_r\}_{r\geq 0}$ are naturally modelled as a Markov chain. In \cite{10} it was further assumed that population size is of order $N$ and changes accrue slowly, at a rate of order $1/N$, so that for large $N$ the random function $M_{\lfloor Nt \rfloor}/N$, $t \geq 0$, can be approximated in the Skorokhod topology by a càdlàg continuous-time stochastic process. Then the scaled genealogical trees converge weakly to a Kingman coalescent with a random time change (see \cite{10}, and also \cite{3} for a very general approach).

The topic of this paper is randomly and rapidly varying environments with an ergodic property. The population remains large, $M_r = O(N)$ thought of as tending to infinity, but we assume that a change of the same magnitude, a gain or loss of $N/2$ for example, can occur in one season. In the long run, the environment as reflected in the population size is, however, assumed ergodic. We study the genealogical process in such populations and establish weak convergence of the latter, as population size tends to infinity, to the Kingman coalescent with a linear time change $t \mapsto \lambda t$. The parameter $\lambda$ measures the variability of the population size (or the brutality of changes, the ‘roughness’ of the surrounding environment).
We start with a WF model with randomly rapidly varying population size, and claim that its effective population size is a stationary harmonic mean of the actual sizes. This result is extended in Section 4 to a class of exchangeable reproduction models. For the sake of simplicity we assume that the second moment of the offspring number has a finite limit as \( N \to \infty \). This leads to the simple time scale \( T_N = c_N^{-1} \) ensuring convergence. The case of a general arbitrary time scale \( T_N = c_N^{-1} \) is treated similarly but requires the use of central moment conditions.

Section 5 presents four examples illustrating the convergence theorem of Section 4, whose proof is given in Sections 6 and 7.

2. The Wright–Fisher model with random stationary size

The celebrated WF model describes an asexual population counted in successive seasons. These are identified with non-overlapping generations, all generations having the same size \( N \). If \( (\nu_1, \ldots, \nu_N) \) are the offspring numbers of individuals in one generation, then also \( \nu_1 + \cdots + \nu_N = N \) since this is the size of the next generation. The WF reproduction law specifies the joint distribution of \( (\nu_1, \ldots, \nu_N) \) as the symmetric multinomial distribution \( M_N(N, 1/N, \ldots, 1/N) \). The past genealogical tree of \( n \) individuals sampled from the WF population with time measured in units of \( N \) generations has a simple asymptotic distribution called the \( n \)-coalescent. In [12] it was defined as an absorbing Markov chain, whose states are the equivalence relations on the set of sampled individuals \( \{1, \ldots, n\} \). The Kingman coalescent topology is extremely simple: with (say) \( k \) branches in the tree, any pair of branches can join next with probability \( 1/(k^2) \). Due to the symmetry in the topology distribution, it is sufficient to consider only the number \( R_t \) of branches at the (continuously recorded) backward time \( t \). The Markov chain \( (R_t)_{t \geq 0} \) is a pure death process starting at \( R_0 = n \) with interdeath times exponentially distributed with mean values \( 1/(\binom{n}{2}), 1/(\binom{n-1}{2}), \ldots, 1 \).

Consider the WF population \( r \) generations ago and let \( Z_r \) be the number of common ancestors of \( n \) individuals sampled in the current generation. The time-scaled ancestral process \( (Z_{\lfloor Nt \rfloor})_{t \geq 0} \) converges weakly in the Skorokhod topology to \( (R_{\lambda h} t)_{t \geq 0} \) as \( N \to \infty \). By Proposition 1 below, this classical result is extended to the case of WF reproduction in a Markovian stationary environment, which we introduce next, by means of a retrospective view of reproduction.

In the Wright–Fisher model the retrospective view of reproduction is very simple: each individual in the current generation chooses a parent uniformly at random from the previous generation. This picture is easily adapted to variable population sizes. Let the past population sizes \( \{M_r\}_{r \geq 0} \) be a Markov chain with \( K \) possible values \( \{x_1 N, \ldots, x_K N\} \), where \( N \) as before is an integer intended to be sent to infinity, and the parameters \( (x_1, \ldots, x_K) \) are fixed numbers. The Markov chain is assumed regular and time-homogeneous with the stationary distribution \( P(M_r = x_j N) = v_j \), where \( 1 \leq j \leq K \) and \( r = 0, 1, 2, \ldots \). The next assertion is a corollary of the main result of this paper, Theorem 1 stated in Section 4.

Proposition 1. Consider a WF population with stationary population size, as described above, and put \( \lambda_h = v_1/x_1 + \cdots + v_K/x_K \). As \( N \to \infty \), the time-scaled ancestral process \( (Z_{\lfloor Nt \rfloor})_{t \geq 0} \) converges weakly to \( (R_{\lambda h} t)_{t \geq 0} \).

For example, if \( (M_r/N)_{r \geq 0} \) follows the Ehrenfest gas model with \( x_j \equiv j \) and \( v_j = K^{-1} \binom{K-1}{j-1} 2^{1-K} \), then

\[
\lambda_h = 2^{1-K} \sum_{i=1}^{K} \binom{K-1}{i-1} i^{-1} = \frac{2}{K} (1 - 2^{-K}).
\]
For a biological illustration, assume that there are good years with a large population, $L N$, average years when population size is $N$, and bad years with a population $SN$ with stationary probabilities $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{1}{4}$ respectively. Then $\lambda_h = (1/2L + 1 + 1/2S)/2$.

### 3. Effective population size

In population genetics, the WF population with constant population size $N$ serves as an idealised reference model. For other population models it is customary to calculate an inbreeding effective population size $N_e$, which reflects the population breeding structure and demographic history. Even if the actual population size, $N$, was constant in the past, the effective population size $N_e$ will turn out to be smaller than $N$, provided that a randomly sampled pair of individuals can be expected to be more closely related than a similar pair in the WF population of size $N$.

The definition of $N_e$ is given in terms of the tail distribution of the time $T_2$ to the most recent common ancestor for a pair of individuals sampled at random from the current generation. For the WF model, the geometric distribution, $P(T_{WF} > u) = (1 - 1/N)^u$, implies that, for $u = 1, 2, \ldots$,

$$\frac{1}{N_e} := 1 - [P(T_2 > u)]^{1/u},$$

In analogy with this, the effective population size, $N_e$, of any population observed during a certain number $u$ of generations is defined by

$$\frac{1}{N_e} := 1 - [P(T_2 > u)]^{1/u},$$

where the tail probability of $T_2$ depends on the given reproduction law and historical population sizes $\{M_r\}_{r=1}^u$.

Consider, for example, a WF population with known $\{M_r\}_{r=1}^u$. For large population sizes,

$$P(T_2 > u) = \prod_{k=1}^{u} (1 - \frac{1}{M_k}) \approx 1 - \sum_{k=1}^{u} \frac{1}{M_k} = 1 - \frac{u}{N_h},$$

where $N_h$ is the harmonic mean of actual population sizes. From the definition (1), straightforward calculations yield that $N_e \approx N_h$ for a given time interval $[0, u]$.

The above example can be considered as a calculation of the short-term effective population size. In contrast, Proposition 1 can be viewed as a result on the long-term effective population size with $u$ being of order $N$. It shows that the ancestral process $(Z_{[N\lambda^{-1}_h t]}^{WF})_{t \geq 0}$ is asymptotically similar to the ancestral process $(Z_{[Nt]}^{WF})_{t \geq 0}$ of the WF population with the constant size $N$. It follows that $P(T_2 > N\lambda^{-1}_h t) \approx P(T_{WF} > Nt)$. Thus, in the case of the WF model with stationary size,

$$\frac{1}{N_e} \approx \sum_j \frac{v_j}{x_j N},$$

and the effective population size $N_e$ is again approximated by the harmonic mean $N_h$ of actual sizes. A detailed comparison of $N_e$ and $N_h$ in the case $K = 2$ is given in [8]. Analysis of the exact formula for $N_e$ presented there confirms that the difference between $N_e$ and $N_h$ is asymptotically negligible for large values of $N$. 

4. Exchangeable reproduction

Our main result, Theorem 1 below, deals with a variable size version of a broad family of haploid population models with non-overlapping generations introduced in [1] and [2]. Given the sizes of two consecutive generations, \( M_{r-1} = l \) and \( M_r = m \), the corresponding offspring numbers \((\nu_1, \ldots, \nu_m)\) satisfy \( \nu_1 + \cdots + \nu_m = l \). We assume that the numbers \((\nu_1, \ldots, \nu_m)\) have an exchangeable joint distribution which depends on the involved generation sizes \((l, m)\) but is independent of the generation number \( r \). Exchangeability means that, for any permutation \((i_1, \ldots, i_m)\) of \((1, \ldots, m)\), the reordered offspring numbers \((\nu_{i_1}, \ldots, \nu_{i_m})\) have the same joint distribution as \((\nu_1, \ldots, \nu_m)\). This implies that all the offspring numbers \(\nu_j\) have the same marginal distribution.

As in Section 2, the past population sizes \(\{M_r\}_{r \geq 0}\) form a Markov chain with transition probabilities

\[
\pi_{ij} := P(M_r = x_jN \mid M_{r-1} = x_iN), \quad 1 \leq i, j \leq K.
\]

In the following \(E_{ij}\) and \(P_{ij}\) will stand for expectation and probability conditional upon the event \(\{M_{r-1} = x_iN, M_r = x_jN\}\), whereas \(E_{lm}^{ij}\) and \(P_{lm}^{ij}\) will denote conditioning upon the event \(\{M_{r-1} = l, M_r = m\}\).

In this notation, \(E_{ij}(\nu_1) = x_i/x_j\). Higher moments of offspring numbers will be supposed to satisfy the conditions

\[
E_{ij}(\nu_1^2) \rightarrow b_{ij}, \quad N \rightarrow \infty, \quad 1 \leq i, j \leq K, \quad (2)
\]

and

\[
E_{ij}(\nu_1^3) = o(N), \quad N \rightarrow \infty, \quad 1 \leq i, j \leq K. \quad (3)
\]

While (2) ensures that \(T_N = N\) is an efficient time scale leading to a coalescent approximation, the condition (3) is aimed at prohibiting multiple mergers of ancestral lines. Observe that, under the condition (2), the parameters \(\sigma_{ij}^2 := b_{ij} - x_i/x_j\) are always nonnegative since \(\nu_1(\nu_1 - 1) \geq 0\) and \(E_{ij}(\nu_1(\nu_1 - 1)) \rightarrow \sigma_{ij}^2\). Furthermore, if all \(\sigma_{ij}^2 = 0\), then \(P_{ij}(\nu_1 \leq 1) = 1\), and it follows that

\[x_iN = \nu_1 + \nu_2 + \cdots + \nu_{x_jN} \leq x_jN\]

for all \(i, j\). This cannot be true unless the \(x_i\) are all equal. Hence, in the case of truly varying population sizes,

\[
\sigma_{ij}^2 > 0 \quad \text{for at least one pair } (i, j), \quad (4)
\]

which we assume henceforth.

Our main result is the following theorem. It extends Kingman’s fundamental result in [11], where (5) with \(\lambda = \sigma^2\) was proved under the assumption of constant population size \(N\) and more restrictive moment conditions.

**Theorem 1.** Under the conditions (2), (3), and (4), the weak convergence

\[
(Z_{i|Nt})_{t \geq 0} \rightarrow (R_{\lambda t})_{t \geq 0}, \quad N \rightarrow \infty \quad (5)
\]

holds with

\[
\lambda = \sum_{i=1}^{K} \sum_{j=1}^{K} \nu_i \pi_{ij} \sigma_{ij}^2 x_j x_i^{-2} \quad (6)
\]

so that \(N_e \approx N\lambda^{-1}\).
The parameters $\sigma^2_{ij}$ describe how variation in fertility within generations depends on population-size fluctuations. They play a crucial role in determining the long-term effective size of the introduced population model. The larger the variation in offspring numbers, the faster goes the genetic drift, and therefore the smaller is the effective population size.

5. Examples

In this section we illustrate Theorem 1 with four examples simplifying the population model from Section 4 in different directions.

Example 1. A simple form of stationary fluctuations of population size would be to assume that $\{M_r\}_{r \geq 0}$ is an i.i.d. sequence with $P(M_r = x_j N) = v_j$. Then the general formula (6) reduces to

$$\lambda = \sum_{i=1}^{K} \sum_{j=1}^{K} v_i v_j \sigma^2_{ij} x_j \frac{x_i^2}{x_j^2}.$$ 

Note that the formulae for the parameter $\lambda$ in the following three examples remains the same for general stationary Markov $\{M_r\}_{r \geq 0}$ as for the case of i.i.d. population sizes.

Example 2. To verify that Theorem 1 entails Proposition 1, consider the WF reproduction law, so that the joint distribution of offspring numbers $(\nu_1, \ldots, \nu_m)$ is symmetric multinomial $M_n(l, 1/m, \ldots, 1/m)$ given two consecutive generation sizes $(l, m)$. Therefore, if $l = x_i N$ and $m = x_j N$, the conditional marginal distribution of $\nu_1$ is binomial $\text{Bin}(x_i N, (x_j N)^{-1})$, which tends to the Poisson distribution with parameter $x_i/x_j$ as $N \to \infty$. Also, moments converge so that

$$\sigma^2_{ij} = \frac{x_i^2}{x_j^2} \quad \text{and} \quad \lambda = \prod_{i=1}^{K} \pi_i x_i \sum_{j=1}^{K} v_j \frac{x_i}{x_j} = \lambda_h,$$

as asserted in Proposition 1.

Example 3. Now suppose that only the probability $P_{ij}(\nu_1 > 0)$ of getting any children at all is influenced by the population size, whereas the conditional distributions of how many children to get, $P_{ij}(\nu_1 = k \mid \nu_1 > 0)$, are independent of $(N, i, j)$. Then, so are the conditional moments

$$E(\nu_1 \mid \nu_1 > 0) = a, \quad E(\nu_1^2 \mid \nu_1 > 0) = b.$$

Since $E_{ij}(\nu_1) = x_i/x_j$, $P_{ij}(\nu_1 > 0) = x_i/x_j a$ and $E_{ij}(\nu_1^2) = x_i/x_j a$. Therefore, $\sigma^2_{ij} = (b/a - 1) x_i/x_j$ and

$$\lambda = \left(\frac{b}{a} - 1\right) \prod_{i=1}^{K} \frac{v_i \pi_i x_i^{-1}}{\sum_{j=1}^{K} v_i x_j^{-1}} = \left(\frac{b}{a} - 1\right) \sum_{i=1}^{K} v_i x_i^{-1} = \left(\frac{b}{a} - 1\right) \lambda_h.$$

We conclude that, for this two-parameter model, $N_e \approx N_h a/(b - a)$. 


Example 4. Here we introduce a flexible extension of the WF reproduction law. Consider a parent generation of size \( m \) with offspring numbers \( (\nu_1, \ldots, \nu_m) \) such that \( \nu_1 + \cdots + \nu_m = l \). Let the joint distribution of \( (\nu_1, \ldots, \nu_m) \) be the compound multinomial \( \text{Mn}(l, p_1, \ldots, p_m) \) with the vector \( (p_1, \ldots, p_m) \) having a Dirichlet distribution with parameters \( (\alpha_1, \ldots, \alpha_m) \). The resulting reproduction law is the so-called Dirichlet-compound multinomial distribution (see [9]) with

\[
P^{lm}(\nu_1 = k_1, \ldots, \nu_m = k_m) = \binom{l}{k_1, \ldots, k_m} \frac{\alpha_1^{[k_1]} \cdots \alpha_m^{[k_m]}}{\alpha_0^{[l]}},
\]

where \( \alpha_0 := \alpha_1 + \cdots + \alpha_m \) and \( \alpha^{[k]} := (\alpha + 1) \cdots (\alpha + k - 1) \) denotes the ascending factorial power. This model is particularly convenient for coalescent calculations due to the following formula available for joint moments:

\[
E^{lm}(v^{(1)}_{1_1} \cdots v^{(1)}_{u_1}) = \frac{l(u) \alpha^{[u]}_{1_1} \cdots \alpha^{[u]}_{u_1}}{\alpha_0^{[l]}},
\]

where \( u := u_1 + \cdots + u_k \) and \( \alpha^{(u)} := (\alpha - 1) \cdots (\alpha - u + 1) \) denotes the descending factorial power. For marginal factorial moments, it implies that \( E^{lm}(\nu^{(1)}_u) = \frac{l(l - 1)(\alpha + 1)}{m(m\alpha + 1)(\alpha + 2)} \).

In the exchangeable reproduction case with \( \alpha_1 = \cdots = \alpha_m = \alpha \) and \( \alpha_0 = m\alpha \), this relation yields that

\[
E^{lm}(v^{(2)}_1) = \frac{l(l - 1)(\alpha + 1)}{m(m\alpha + 1)}, \quad E^{lm}(v^{(3)}_1) = \frac{l(l - 1)(l - 2)(\alpha + 1)(\alpha + 2)}{m(m\alpha + 1)(m\alpha + 2)}.
\]

It is easy to verify that, if the parameter \( \alpha \) does not depend on \( N \), the conditions (2) and (3) hold with

\[
\sigma_{ij}^2 = \frac{(1 + 1/\alpha)x_i^2}{x_j^2} \quad \text{and} \quad \lambda = \frac{(1 + 1/\alpha)\lambda_0}{N}.
\]

We see that, the larger is \( \alpha \), the closer is the current model to the classical WF setup; on the other hand, the smaller is \( \alpha \), the more variable are offspring numbers, implying a higher coalescent rate.

6. Conditional transition probabilities of \( Z_r \)

The purpose of this section is to describe the asymptotics of the conditional transition probabilities

\[
p^{ij}_{uv} := P_{ij}(Z_r = v \mid Z_{r-1} = u)
\]

as \( N \to \infty \). We shall show that pairwise mergers occur at a rate proportional to \( 1/N \),

\[
p^{ij}_{uv} \sim N^{-1} \binom{u}{2} \phi(i, j) \quad \text{if} \quad v = u - 1,
\]

and that multiple mergers of ancestral lines are negligibly rare,

\[
p^{ij}_{uv} = o(N^{-1}) \quad \text{if} \quad v < u - 1.
\]

Combining (7) and (8) we obtain that

\[
p^{ij}_{iu} = 1 - N^{-1} \binom{u}{2} \phi(i, j) + o(N^{-1}), \quad \phi(i, j) := \sigma_{ij}^2 x_j x_i^{-2}.
\]
Coalescence in populations of varying size

Previous generation

Current generation

Figure 1: Two consecutive generations.

First we explain the formula for the conditional transition probability

$$P^{lm}(Z_r = v \mid Z_{r-1} = u) = \binom{m}{l} \sum_{u_1 + \cdots + u_v = u} E^{lm} \left( \frac{v_1}{u_1} \cdots \frac{v_v}{u_v} \right),$$

(10)

where the summation is done over all distinct vectors \((u_1, \ldots, u_v)\) with natural components satisfying \(u_1 + \cdots + u_v = u\). We use a combinatorial ‘putting balls into boxes’ argument illustrated by Figure 1, which depicts two consecutive generations with sizes \(l = 11\) and \(m = 10\) and the offspring numbers \(v_1 = 3, v_2 = 2, v_3 = v_5 = v_6 = v_9 = v_{10} = 0, v_4 = v_8 = 1, v_7 = 4\). If we follow the arrows going from the five leftmost individuals in the current generation, we end up with three individuals in the previous generation. This example corresponds to the values \(u = 5, v = 3, u_1 = 2, u_2 = 1, u_3 = 2, i_1 = 1, i_2 = 2, i_3 = 7\) of the parameters introduced next.

Think of \(u\) balls corresponding to \(u\) children sampled in a certain generation and \(m\) boxes representing potential parents from the previous generation. These boxes have \(l\) slots with \(v_1\) slots in box 1, \(v_2\) slots in box 2, \ldots, \(v_m\) slots in box \(m\). Then the random event

\[ A(i_1, \ldots, i_v) \]

\[ = \{ u_1 \text{ balls fall into box } i_1, u_2 \text{ balls fall into box } i_2, \ldots, u_v \text{ balls fall into box } i_v \} \]

has the conditional probability

\[ P^{lm}(A(i_1, \ldots, i_v) \mid v_1, \ldots, v_m) = \frac{1}{\binom{m}{l}} \left( \frac{v_1}{u_1} \right) \cdots \left( \frac{v_v}{u_v} \right). \]

Since there are \(\binom{m}{v}\) different ways to choose \(v\) hosting boxes \(i_1 < \cdots < i_v\), this equality entails (10).

With \(l = x_i N\) and \(m = x_j N\), (10) implies the asymptotic relation

\[ p^{ij}_{iu,v} \sim N^{v-u}(u!)^{-1} x_j x_i^{-u} \sum_{u_1 + \cdots + u_v = u} \binom{u}{u_1 \cdots u_v} E_{ij}(v_1^{(u_1)} \cdots v_v^{(u_v)}), \]

which has two important consequences. First, if \(v = u - 1\),

\[ p^{ij}_{iu,u-1} \sim N^{u-1} \binom{u}{2} x_j x_i^{-u} E_{ij}(v_1^{(2)} v_2 \cdots v_{u-1}). \]
Second, in order to prove (7) we must show that, under the conditions (2) and (3),

\[ \lim_{N \to \infty} E_{ij}(v_1^{(2)}v_2 \cdots v_{u-1}) = \left( \frac{x_i}{x_j} \right)^{u-2} \sigma_{ij}^2. \]  

(11)

On the other hand, according to the asymptotic formula for \( p_{uv}^{ij} \), (8) holds, provided that

\[ E_{ij}(v_1^{(u_1)} \cdots v_v^{(u_v)}) = o(N^{u-v-1}) \quad \text{if} \quad v \leq u - 2, u_1 + \cdots + u_v = u. \]  

(12)

In the rest of the section, we check (11) and (12) by induction over \((u_1, \ldots, u_v)\) with \(u_1 \geq \cdots \geq u_v\) using the following order:

- \{2\}, \{3\}, \{2, 1\},
- \{4\}, \{3, 1\}, \{2, 2\}, \{2, 1, 1\},
- \{5\}, \{4, 1\}, \{3, 2\}, \{3, 1, 1\}, \{2, 2, 1\}, \{2, 1, 1\},

and so on. Here, the last set in each row (except the first) concerns (11). In lieu of a formal induction proof, we verify (11) and (12) for the first three rows of this triangular array, and then outline the pattern for continuation.

In the cases \{2\}, \{3\}, \{4\}, \{3, 1\}, \{2, 2\}, \{2, 1, 1\}, the trick is to see that

\[ v_2 + \cdots + v_m = l - v_1 \implies (m - 1) E^{lm}(v_1^{(2)}v_2) = (l - 2) E^{lm}(v_1^{(2)}) - E^{lm}(v_1^{(3)}). \]

It follows that

\[ (m - 1) E^{lm}(v_1^{(2)}v_2) = (l - 2) E^{lm}(v_1^{(2)}) - E^{lm}(v_1^{(3)}), \]

confirming that (11) holds for \(u = 3\) as well.

For \{4\}, (12) follows easily from the upper bound

\[ E_{ij}(v_1^{(4)}) < N E_{ij}(v_1^{(3)}). \]

In the case \{3, 1\}, the trick to see that \( v_2 + \cdots + v_m = l - v_1 \) implies (12) via

\[ (m - 1) E^{lm}(v_1^{(3)}v_2) = (l - 4) E^{lm}(v_1^{(3)}) - E^{lm}(v_1^{(4)}). \]

The case \{2, 2\} follows from the upper bound

\[ E_{ij}(v_1^{(2)}v_2^{(2)}) < N E_{ij}(v_1^{(2)}v_2) \]

and the already established formula for \{2, 1\}. In the case \{2, 1, 1\}, the trick works in a slightly different way:

\[ l E^{lm}(v_1^{(2)}v_2) = \sum_{i=1}^{m} E^{lm}(v_1^{(2)}v_2 v_i) \]

\[ = (m - 2) E^{lm}(v_1^{(2)}v_2 v_3) + E^{lm}(v_1^{(2)}v_1 v_2) + E^{lm}(v_1^{(2)}v_2^{(2)}). \]
Coalescence in populations of varying size

implying that

\[(m - 2) E^{(m)}(v_1^{(2)}v_2v_3) = (l - 4) E^{(m)}(v_1^{(2)}v_2) - E^{(m)}(v_1^{(3)}v_2) - E^{(m)}(v_1^{(2)}v_2^{(2)}).\]

Thus, (11) for \([2, 1, 1]\) follows from (11) for \([2, 1]\) and from (12) for \([3, 1]\) and \([2, 2]\).

The continued argument then follows four paths (recall that \(v\) is the number of coordinates and \(u\) their sum, write \(e_i\) for a vector with \(i\)th component 1 and all others zero, and interpret \((u_1, \ldots, u_0, u_1 - 1)\) as \(u_1 - 1\)):

1. If \(uv \geq 3\), or \(v \geq 3\) and \(u_2 = 2\), or \(v = 1\) and \(u_1 \geq 3\), then prove (12) using (12) for \((u_1 - 1, u_v - 1, u_1 - 1)\).

2. If \(v = 2\), \(u_1 \geq 4\), and \(u_2 = 1\), then prove (12) using (12) for \((u_1)\) and \((u_1 + 1)\).

3. If \(v \geq 3\), \(u_1 \geq 3\), \(u_2 = 1\), and \(u_1 \geq 2\), then prove (12) using (12) for \((u_1, \ldots, u_{v-1})\) and \((u_1, \ldots, u_{v-1} + e_j)\), \(j = 1, 2, 3\).

4. If \(v \geq 4\), \(u_1 = 2\), and \(u_2 = 1\), then prove (11) using (11) for \((u_1, \ldots, u_{v-1})\) and \((u_1, \ldots, u_{v-1} + e_j)\), \(j = 1, 2\).

7. Proof of Theorem 1

To prove the weak convergence stated in Theorem 1, we apply Theorem 2.12 of [6, p. 173] which was called the projection theorem in [16]. It serves to show that it suffices to establish the convergence of one-dimensional distributions

\[P(Z_{[Nt]} = v \mid Z_0 = u, M_0 = Nx_1) \to P(R_{[t]} = v \mid R_0 = u), \quad N \to \infty. \quad (13)\]

That follows from the result of Section 6, which can be summarised as

\[P(i, j) = I + N^{-1} \phi(i, j)Q + o(N^{-1}), \quad N \to \infty, \quad (14)\]

where \(P(i, j)\) is the conditional one-step transition matrix of the process \((Z_r, r = 0, 1, 2, \ldots)\), \(I\) is the identity matrix, and

\[Q = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 1 & -1 & 0 & \cdots & 0 \\ 0 & 3 & -3 & \cdots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & (3) & (2) \end{pmatrix}\]

is the matrix of transition rates for the Markov chain \((R_t)_{t \geq 0}\).

Let \((S_r, r = 0, 1, 2, \ldots)\) be the index process defined by \(M_r = xS_rN\). According to (14), the \([Nt]\)-step transition matrix conditioned upon the population-size history \(S_r, r \leq [Nt]\) satisfies

\[\prod_{r=1}^{[Nt]} P(S_r, S_{r+1}) = \prod_{r=1}^{[Nt]} (I + N^{-1} \phi(S_r, S_{r+1})Q + o(N^{-1})) \]

\[= \exp \left\{ N^{-1} \sum_{r=1}^{[Nt]} \phi(S_r, S_{r+1}) \right\} (1 + o(1)).\]
The sum \( \sum_{r=1}^{[N_t]} \phi(S_{r-1}, S_r) \) is an additive functional of the regular two-dimensional stationary Markov chain \((S_{r-1}, S_r)\). The strong law of large numbers applied to this additive functional (see [5, pp. 323–324]) implies that

\[
\prod_{r=1}^{[N_t]} P(S_{r-1}, S_r) \rightarrow \exp \left\{ Qt \sum_{i=1}^{K} \sum_{j=1}^{K} v_{ij} \pi_{ij} \phi(i, j) \right\} = e^{\lambda t Q}
\]

almost surely as \( N \to \infty \). Taking the expectation, averaging over possible population-size histories and invoking bounded convergence show that the distribution of \( Z_{[N_t]} \) conditioned on \( \{Z_0 = u, M_0 = N_0\} \) converges to that of \( R_{\lambda t} \) conditioned on \( \{R_0 = u\} \). By this we have proved (13) and therefore Theorem 1.

Acknowledgements

We are grateful to the referee for close reading and useful suggestions. This work was supported by the Knut and Alice Wallenberg Foundation and The Bank of Sweden Tercentenary Foundation.

References