Coalescence Theory, Structured Populations with Fast Migration

Ola Hössjer

April, 2011

Department of Mathematics, Stockholm University
Wright-Fisher model

One-sex population, constant size $N$
Nonoverlapping generations
Each gener. children choose parents randomly and independently

$$\nu^l = \text{nr of children of parent } l$$

$$\nu = (\nu^l)_l^{N} \sim \text{Mult}(N; 1/N, \ldots, 1/N)$$

Follow ancestry of sample of $n \ll N$ individuals backwards

$N = 10, \ n = 3$
Discrete time coalescence process

\[ X_N(\tau) = \text{nr of ancestors of sample } \tau \text{ generations back in time} \]

Discrete time Markov chain with \( X_N(0) = n \) and

\[
P(X_N(\tau + 1) < a | X_N(\tau) = a) = 1 - \prod_{b=1}^{a-1} (1 - b/N)
\]

\[
= \left( \frac{a}{2} \right) / N + o(N^{-1}), \ a = 2, \ldots, n
\]

\[ X_N(0) = 3, \ X_N(2) = 2, \ X_N(9) = 1 \]
Continuous time coalescence process

Let $N \to \infty$, keep $n$ fixed.
Rescale time by factor $N$:

$$\{X_N(\lfloor Nt \rfloor); t \geq 0 \} \xrightarrow{\mathcal{L}} \{A(t); t \geq 0 \}$$

A is Kingman’s coalescent (Kingman, 1982a-b).
Continuous time Markov process
Infinitesimal generator ($q_{ab}$), with

$$q_{ab} = \begin{cases} \binom{a}{2}, & b = a - 1, \\ -\binom{a}{2}, & b = a, \\ 0, & \text{otherwise} \end{cases}$$

Ola Hössjer  Coalescence Theory, Structured Populations with Fast Migration
Six simulated Kingman coalescents with $n = 25$

\[ T_a = |\{t; A(t) = a\}| \sim \text{Exp} \left( \binom{a}{2} \right). \]

Kingman’s result has been generalized to populations with

- **Two sexes** (Möhle, 1998c)
- **Non-constant size** (Jagers and Sagitov, 2004)
- **Geographic structure** (Nordborg and Krone, 2002)
- **Age structure** (Kaj et al., 2001, Sagitov and Jagers, 2005)
- **Self-fertilization** (Fu, 1997, Nordborg and Donnelly, 1997)
- **Variable reproductivity** (Möhle, 1998b)

and many other models ...
Effective population size

$N$ is (current) population size

Kingman’s coalescent is robust, since for a large class of population genetic models

$$\{X_N(\lceil Nt \rceil); \ t \geq 0 \} \xrightarrow{\mathcal{L}} \{A(ct); \ t \geq 0 \}$$

where

$$c = \text{coalescence rate}$$

and

$$N_e = N/c$$

= coalescence effective population size

= size of WF model with same ancestry asymptotically.

Coalescence rate and individual variability in productivity

One-sex population, constant size $N$
Nonoverlapping generations

$$\nu^l = \text{number of children of parent } l.$$  

$$(\nu^l)_i^N \text{ exchangeable random variables.}$$

Hence, since $\sum_{i=1}^{N} \nu^l = N$,  

$$E(\nu^l) = 1$$

and it can be shown that  

$$c = \lim_{N \to \infty} E \left( \nu^l (\nu^l - 1) \right)$$

provided  

$$E \left( (\nu^l)^3 \right) = o(N),$$

otherwise multiple mergers in limit.
Structured Population Model

\[ N = \text{population size} \]
\[ L = \text{number of subpopulations} \]
\[ Na_i = \text{size of subpopulation } i \quad (\sum_{i=1}^{L} a_i = 1), \]
\[ m_{ki} = \text{“migration rate” from subp. } k \text{ to } i, \]
\[ \nu_{ki}^l = \text{nr of “offspring” of } l\text{th individual of subpop. } k \text{ that end up in subpop. } i \text{ (possibly including parent itself)} \]

Constant subpopulation sizes is formulated as

\[
\sum_{i=1}^{Na_k} \nu_{ki}^l = Na_k m_{ki}, \quad \sum_{k=1}^{L} a_k m_{ki} = a_i,
\]

Exchangeability of parental reproduction from subpop. \( k \) to \( i \):

\[ E(\nu_{ki}^l) = m_{ki} \text{ (independently of } N, \text{ i.e. fast migration)} \]
Example with $L = 2$ subpopulations

$N = 10,$
$a_1 = 0.4, a_2 = 0.6,$
$m_{11} = 3/4, m_{12} = 1/2, m_{21} = 1/6, m_{22} = 4/6,$
$\nu_{11}^1 = 2, \nu_{12}^1 = 1,$
$\nu_{21}^6 = 0, \nu_{22}^6 = 1,$
Ancestral process, main convergence result

\( X_{Ni}(\tau) \) = nr of ancestors of sample in subpop. \( i \), \( \tau \) generations back

\( X_N(\tau) \) = total nr of ancestors of sample \( \tau \) generations back

\[ = \sum_{i=1}^{L} X_{Ni}(\tau), \]

\( X_N(0) \) = sample size

\[ = n. \]

Under certain conditions

\[ \{ X_N([Nt]); \ t \geq 0 \} \overset{\mathcal{L}}{\longrightarrow} \{ A(ct); \ t \geq 0 \} \]

in Skorohod topology on \( D\{1,\ldots,n\}[0,\infty) \).

What is \( c \)?
Formula for coalescence rate

Any individual’s ancestral subpopulation history is a Markov chain with state space \( \{1, \ldots, L\} \) and trans. matrix \((b_{ik})\), where

\[
b_{ik} = P(\text{parent of subpop. } i \text{ individual from subpop. } k) = a_k m_{ki}/a_i
\]

and unique equilibrium distribution

\[(\gamma_1, \ldots, \gamma_L)\].

Then, under mild regularity conditions, conv. to Kingman’s coalescent with

\[c = \sum_{i,j,k=1}^{L} \gamma_i \gamma_j b_{ik} b_{jk} c_{kij}\]

where

\[
c_{kij} = \begin{cases} 
\lim_{N \to \infty} E \left( \nu_{ki}^l (\nu_{ki}^l - 1) \right) / (m_{ki}^2 a_k), & i = j, \\
\lim_{N \to \infty} E \left( \nu_{ki}^l \nu_{kj}^l \right) / (m_{ki} m_{kj} a_k), & i \neq j,
\end{cases}
\]

is the **local coalescence rate** for two lines that merge from subpop. \(i\) and \(j\) to subpop. \(k\).
Example 1: Geographical structure, nonoverlapping generations

Subpopulation = island

\( \nu_{ki} \) = nr of children of \( l \)'th individual of subpop \( k \) born in subpop \( i \).

If WF type reproduction

\[
(\nu_{ki}^l)_{l=1}^{Na_k} \sim \text{Mult}(Na_k m_{ki}; 1/(Na_k), \ldots, 1/(Na_k))
\]

gives local coalescence rate

\[
c_{kij} = 1/a_k
\]

and

\[
c = \sum_{i,j,k=1}^{L} \gamma_i \gamma_j b_{ik} b_{jk} a_k^{-1} = \sum_{k=1}^{L} \frac{\gamma_k^2}{a_k} \gamma_k \equiv a_k \ 1,
\]

see Nordborg and Krone (2002).
Example 2: Age structured models

Subpopulation = age class

\[ L = \text{nr of age classes} = 4, \]
\[ m_i = m_{i1} = \text{exp. nr of children of parents of age } i, \]
\[ s_i = m_{i,i+1} = \text{survival prob from age class } i \text{ to } i+1, \]

with

\[ a_1 = 0.4, a_2 = 0.3, a_3 = 0.2, a_1 = 0.1, \]
\[ s_1 = 3/4, s_2 = 2/3, s_3 = 1/2. \]

See also Jagers and Sagitov (2005).
\[ \rho = \text{Corr}(\nu_{k1}^l, \nu_{k,k+1}^l) = \text{correl. between nr of children and survival} \]

<table>
<thead>
<tr>
<th>Figure</th>
<th>((s_1, s_2, s_3))</th>
<th>((m_1, m_2, m_3, m_4))</th>
<th>(c_{k11} a_k)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a)</td>
<td>((1, 1, 0.5))</td>
<td>((2,2,2,2)/7)</td>
<td>-3/2</td>
</tr>
<tr>
<td>b)</td>
<td>((1, 1, 0.5))</td>
<td>((0,0,2,2)/3)</td>
<td>1/2</td>
</tr>
<tr>
<td>c)</td>
<td>((0.5, 0.5, 0.5))</td>
<td>((8,8,8,8)/15)</td>
<td>1/8</td>
</tr>
<tr>
<td>d)</td>
<td>((0.5, 0.5, 0.5))</td>
<td>((0,0,8,8)/3)</td>
<td>13/8</td>
</tr>
</tbody>
</table>
Example 3: Combined age and geographical structure

Subpopulation = (island nr, age nr)

\[ L = 4 \]
\[ N = 28 \]
\[ a_{(1,1)} = a_{(1,2)} = \frac{3}{14}, a_{(2,1)} = a_{(2,2)} = \frac{2}{7} \]
\[ q = \text{prob that all children are born in island different from parent} \]
\[ m_i = \text{fertility of adults of island } i \]
\[ m_{(1,1),(1,2)} = 1, m_{(1,2),(2,1)} = qm_1 \text{ etc} \]
Coalescence rate, combined geographical and age structure

\( q = 0.1 \) (dash-dotted), \( q = 0.5 \) (solid), \( q = 0.9 \) (dotted),
\( a_2 = a_{(2,1)} + a_{(2,2)} \) = relative size of island 2,
\( \gamma_2 = \gamma_{(2,1)} + \gamma_{(2,2)} \) = equilibrium prob of ancestor in island 2
\( m_i \) = fertility of adults of island \( i \),
WF type reproduction
Convergence to Kingman’s coalescent for
- General class of structured models
- Fast migration
- Includes geographical and/or age structure
- General dependency structure of offspring distribution

Coalescence rate $\Rightarrow$ effective population size

References


