Serik Sagitov
28 May 2009, Luminy

**Quenched effective population size**
(work in progress jointly with P.Jagers and V.Vatutin)

*Keywords*

- Wright-Fisher model and Kingman’s coalescent
- Coalescent effective population size $N_e$
- Geographically structured WFM with fast migration
- Randomly varying migration rates
- Markov chains with random transition matrices
- Random environment, $N_e^{\text{quenched}}$ and $N_e^{\text{annealed}}$

This work is supported by The Bank of Sweden Tercentenary Foundation
1 Motivating example

Constant migration rates

Variable migration rates

Time

Prob=1/2
2 Coalescent effective population size

Wright-Fisher model: \( \text{Mn}(N; N^{-1}, \ldots, N^{-1}) \) reproduction law.

Given \( X(0) = n \), the ancestral process \( X(t) \) is a Markov chain with a \( n \times n \) transition matrix \( \Pi = \Pi_N \).

Key decomposition: \( \Pi = I + N^{-1}Q + O(N^{-2}) \) with identity matrix \( I \) and

\[
Q = \begin{pmatrix}
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
1 & -1 & 0 & \ldots & 0 & 0 & 0 \\
& & & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & \binom{n-1}{2} & -\binom{n-1}{2} & 0 \\
0 & 0 & 0 & \ldots & 0 & \binom{n}{2} & -\binom{n}{2}
\end{pmatrix}
\]

Convergence to the Kingman coalescent: \( \Pi^{Nt} \to e^{tQ} \) as \( N \to \infty \).
Kingman’s coalescent is a robust approximation for $X(tN/c)$ in various population models. Coalescent $N_e = N/c$.


Usually $c \geq 1$. Example of $c \leq 1$: offspring numbers 0, 1, 2 with probabilities $(\alpha, 1 - 2\alpha, \alpha)$ imply $c = \sigma^2 = 2\alpha$. 
3 Geographically structured WFM

$L \geq 2$ connected islands: migration and WF reproduction.

Subpopulations of constant sizes $N_1, \ldots, N_L$ with

$$N_1 + \ldots + N_L = N \text{ and } N_i/N \to a_i, \ N \to \infty$$

Ancestral process: lineages migrate independently over the islands until they merge according to the WFM rules of the hosting islands. Configuration process of $n$ lineages:

$$\mathbf{X}(t) = (X_1(t), \ldots, X_L(t))$$

$X_i(t)$ is the number of lineages located on the $i$-th island at $t$-th generation backward in time.

The total number of lineages $X(t) = X_1(t) + \ldots + X_L(t)$ is not a Markov process except for the ”dummy islands” case.
\( \mathbf{X}(t) \) is a Markov chain with a finite state space \( S_1 \cup \ldots \cup S_n \), where \( S_r \) is the set of states \( \mathbf{x} \) satisfying \( x_1 + \ldots + x_L = r \).

The number of elements in \( S_r \) is \( d_r = (r^L - 1)/r \). The transition matrix \( \Pi \) of \( \mathbf{X}(t) \) is of size \((d_1 + \ldots + d_n) \times (d_1 + \ldots + d_n)\).

Key decomposition

\[
\Pi = \mathbf{B}(\mathbf{I} + \mathbf{N}^{-1}\mathbf{C}) + o(\mathbf{N}^{-1}).
\]

Backward migration probabilities \( \mathbf{B} = \text{diag}(\mathbf{B}_1, \ldots, \mathbf{B}_n) \), where \( \mathbf{B}_r \) is the \((d_r \times d_r)\) transition matrix for non-coalescing \( r \) lineages.

Coalescent rates

\[
\mathbf{C} = \begin{pmatrix}
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
\mathbf{C}_{21} & -\mathbf{C}_2 & 0 & \ldots & 0 & 0 & 0 \\
\mathbf{C}_{31} & \mathbf{C}_{32} & -\mathbf{C}_3 & \ldots & 0 & 0 & 0 \\
\mathbf{C}_{41} & \mathbf{C}_{42} & \mathbf{C}_{43} & \ldots & 0 & 0 & 0 \\
\mathbf{C}_{51} & \mathbf{C}_{52} & \mathbf{C}_{53} & \ldots & 0 & 0 & 0 \\
\mathbf{C}_{61} & \mathbf{C}_{62} & \mathbf{C}_{63} & \ldots & 0 & 0 & 0 \\
\end{pmatrix}
\]
\[ \mathbf{C}_r = \text{diag}(C(\mathbf{x}), \mathbf{x} \in S_r), \text{ where } C(\mathbf{x}) = \sum_{k=1}^{L} \frac{1}{a_k} (\frac{x_k}{2}) \]

\[ \mathbf{C}_{r,r-1} \text{ has } \frac{1}{a_k} (\frac{x_k}{2}) \text{ at positions } (\mathbf{x}, \mathbf{x} - \mathbf{e}_k) \text{ and zeros elsewhere.} \]

In particular, if \( L = 2 \), then \( d_r = r + 1 \) and

\[ \mathbf{C}_r = \begin{pmatrix}
\binom{r}{2} \frac{1}{a_1} & 0 & 0 & \ldots & 0 & \ldots & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & (r-k) \frac{1}{a_1} + (k) \frac{1}{a_2} & \ldots & 0 & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & 0 & \ldots & 0 & (\frac{r}{2}) \frac{1}{a_2}
\end{pmatrix} \]

\[ \mathbf{C}_{r,r-1} = \begin{pmatrix}
\binom{r}{2} \frac{1}{a_1} & 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & \binom{r-1}{2} \frac{1}{a_1} & 0 & \ldots & 0 & 0 & \ldots & 0 \\
0 & \frac{1}{a_2} & \binom{r-2}{2} \frac{1}{a_1} & \ldots & 0 & 0 & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & (k) \frac{1}{a_2} & (r-k) \frac{1}{a_1} & \ldots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \ldots & 0 & 0 & \ldots & (\frac{r}{2}) \frac{1}{a_2}
\end{pmatrix} \]
4 Convergence to coalescent

If \((\gamma_1, \ldots, \gamma_L)\) is stationary distr. for the backward migration, then

\[
B_r^u \to P_r, \ u \to \infty
\]

where \(P_r\) consists of \(d_r\) equal rows \((\pi_r(x), x \in S_r)\) with

\[
\pi_r(x) = \left( \begin{array}{c} r \\ x_1, \ldots, x_n \end{array} \right) \gamma_1^{x_1} \cdots \gamma_L^{x_L}.
\]

It follows

\[
B^u \to P = \text{diag}(P_1, \ldots, P_n), \ u \to \infty
\]

and according to M"ohle’s lemma, with \(G = \text{PCP}\)

\[
(B(I + N^{-1}C))^N t \to P - I + e^{tG}, \ N \to \infty
\]

For any \( x \in S_i \)

\[
\sum_{y \in S_j} G(x, y) = \sum_{y \in S_j} (\text{PCP})(x, y) = cQ_{ij}, \quad c = \sum_{k=1}^{L} \frac{1}{a_k} \gamma_k^2.
\]

Writing this as \( G^\downarrow = cQ \) we conclude

\[
(\Pi^{Nt})^\downarrow \to e^{ctQ}, \quad N \to \infty
\]

that the total number of lineages with scaled time \( X(Nt/c) \) is approximated by the number of branches in the Kingman coalescent.

Coalescent \( N_e = N/c \). By Jensen’s inequality \( c \geq 1 \)

\[
\sum_{k=1}^{L} \frac{1}{a_k} \gamma_k^2 = \sum_{k=1}^{L} a_k \left( \frac{\gamma_k}{a_k} \right)^2 \geq \left( \sum_{k=1}^{L} a_k \frac{\gamma_k}{a_k} \right)^2 = 1.
\]

Test example: WFM with dummy islands. In this case \( \gamma_i = a_i \) and \( c = 1 \).
5 Migration in random environment

Transition matrices of backward migration $B_1^{[1]}, B_1^{[2]}, \ldots$ are iid.


Irreducible case: for each pair $1 \leq i, j \leq L$ there is a $u$ such that

$$P(B_1^{[1]} \ldots B_1^{[u]}(i, j) > 0) > 0.$$  \hspace{1cm} (1)

If furthermore, for some $j$ and $u$

$$P(B_1^{[1]} \ldots B_1^{[u]}(i, j) > 0 \text{ for all } i) > 0,$$  \hspace{1cm} (2)

then there exist random stationary probabilities $(\gamma_1, \ldots, \gamma_L)$

$$B_1^{[1]} \ldots B_1^{[u]} \xrightarrow{d} \begin{pmatrix} \gamma_1 & \cdots & \gamma_L \\ \cdot & \cdots & \cdot \\ \gamma_1 & \cdots & \gamma_L \end{pmatrix}, \ u \to \infty.$$
6 Two examples

Example 1: $\gamma_1 = \gamma_2 = 0.5$

$\gamma_1 \sim U(0, 1)$
Exact distribution of $B_1^{[1]} \ldots B_1^{[u]}$ is uniform over $2^u$ matrices

$$
\begin{pmatrix}
    j2^{-u} & 1 - j2^{-u} \\
    (j-1)2^{-u} & 1 - (j-1)2^{-u}
\end{pmatrix}, \ j = 1, \ldots 2^u
$$

which is verified by induction

$$
\left( \frac{j}{2^u}, 1 - \frac{j}{2^u} \right) \begin{pmatrix}
    1 & 0 \\
    1/2 & 1/2
\end{pmatrix} = \left( \frac{j+2^u}{2^{u+1}}, 1 - \frac{j+2^u}{2^{u+1}} \right)
$$

$$
\left( \frac{j}{2^u}, 1 - \frac{j}{2^u} \right) \begin{pmatrix}
    1/2 & 1/2 \\
    0 & 1
\end{pmatrix} = \left( \frac{j}{2^{u+1}}, 1 - \frac{j}{2^{u+1}} \right)
$$

Weak convergence against almost sure convergence

$$
B_1^{[1]} \ldots B_1^{[u]} = \begin{pmatrix}
    Z_u & 1 - Z_u \\
    Z_u - 2^{-u} & 1 - Z_u + 2^{-u}
\end{pmatrix}, \ Z_{u+1} = Z_u/2 + 1/4 \pm 1/4
$$

$$
B_1^{[u]} \ldots B_1^{[1]} = \begin{pmatrix}
    Z_u^* & 1 - Z_u^* \\
    Z_u^* - 2^{-u} & 1 - Z_u^* + 2^{-u}
\end{pmatrix}, \ Z_{u+1}^* = Z_u^* + 2^{-u} (1/4 \pm 1/4)
$$

12
Example 2: \((\gamma_1, \ldots, \gamma_L) \sim \text{Mn}(1, 1/L, \ldots, 1/L)\)

Conditions (1) and (2) follow from

\[ P(B_1(i, j) > 0 \text{ for all } i) > 0, \quad \text{for all } j. \]
7 New formula for $N_e$

Our main assertion: if (1) and (2) hold, then

$$(\Pi^{[1]} \cdots \Pi^{[Nt]}) \downarrow a.s. \rightarrow e^{ctQ}, \quad N \rightarrow \infty$$

so that $N_e = N/c$ with

$$c = c^{[\text{quenched}]} = \sum_{k=1}^{L} \frac{1}{a_k} E\left(\gamma_k^2\right).$$

Notice that

$$c^{[\text{quenched}]} - c^{[\text{annealed}]} = \sum_{k=1}^{L} \frac{1}{a_k} \text{Var}(\gamma_k)$$

and therefore

$$N^{\text{[quenched]}}_e \leq N^{\text{[annealed]}}_e \leq N.$$
Example 1: $N_e^{\text{quenched}} = \frac{3}{4} N_e^{\text{annealed}}$ since

$$c^{\text{annealed}} = \frac{1}{4} \left( \frac{1}{a_1} + \frac{1}{a_2} \right),$$

$$c^{\text{quenched}} = \frac{1}{3} \left( \frac{1}{a_1} + \frac{1}{a_2} \right).$$

Example 2: $N_e^{\text{quenched}} = \frac{1}{2} N_e^{\text{annealed}}$ for $L = 2$

$$c^{\text{quenched}} = \frac{1}{2} \left( \frac{1}{a_1} + \frac{1}{a_2} \right)$$
Example 2 with general $L$ gives the harmonic mean formula

$$\frac{1}{N_e^{[\text{quenched}]}} = \frac{1}{L} \left( \frac{1}{Na_1} + \ldots + \frac{1}{Na_L} \right).$$

Viewed backward in time the population undergoes iid fluctuations of generation sizes.
8 About the proof

Key decomposition

$$\Pi^{[j]} = B^{[j]} (I + N^{-1}C) + o(N^{-1})$$

where again $B^{[j]} = \text{diag}(B_1^{[j]}, \ldots, B_n^{[j]})$.

We have weak convergence of random matrices

$$B^{[1]} \ldots B^{[u]} \xrightarrow{d} P, \ u \to \infty.$$ 

Switching the product order

$$\Pi^{[1]} \ldots \Pi^{[Nt]} \xrightarrow{d} \Pi^{[Nt]} \ldots \Pi^{[1]}$$

allows using a.s. convergence

$$B^{[u]} \ldots B^{[j]} \overset{a.s.}{\to} P^{[j]}, \ u \to \infty, \ j \geq 1.$$
Here $P^{[j]} \overset{d}{=} P$ are defined by $(\gamma_1^{[j]}, \ldots, \gamma_L^{[j]}) \overset{d}{=} (\gamma_1, \ldots, \gamma_L)$ satisfying for $i < j$

$$(\gamma_1^{[j]}, \ldots, \gamma_L^{[j]})B_1^{[j-1]} \cdots B_1^{[i]} = (\gamma_1^{[i]}, \ldots, \gamma_L^{[i]}). \tag{3}$$

An extension of Möhle’s lemma implies

$$(\Pi_1^{[Nt]} \cdots \Pi_1^{[1]}) \downarrow = e^{\frac{Q}{N} \sum_{j=1}^{[Nt]} c^{[j]}} + o_p(1)$$

where

$$c^{[j]} = \sum_{k=1}^{L} \frac{1}{a_k} (\gamma_k^{[j]})^2$$

form a strongly stationary sequence since the defining matrices $B_1^{[1]}, B_1^{[2]}, \ldots$ are iid.

The sequence $c^{[1]}, c^{[2]}, \ldots$ is mixing, because in view of (3), the vectors $(\gamma_1^{[j]}, \ldots, \gamma_L^{[j]})$ and $(\gamma_1^{[i]}, \ldots, \gamma_L^{[i]})$ are asymptotically independent as $j \to \infty.$
By the ergodic theorem

\[
\frac{1}{Nt} \sum_{j=1}^{[Nt]} c[j] \xrightarrow{a.s.} E \left( \sum_{k=1}^{L} \frac{1}{a_k} \gamma_k^2 \right) =: c
\]


we obtain convergence in probability

\[
(\Pi^{[1]} \cdots \Pi^{[Nt]}) \downarrow e^{ctQ}, \ N \to \infty.
\]

Finally, to show that convergence in (4) is a.s. we use a monotonocity property:

for the products of transition matrices \( P_k \) the discrepancy among rows is monotone \( \Delta_{u+1} \leq \Delta_u \), where

\[
\Delta_u = \sum_j \left( \max_i P_1 \cdots P_u(i, j) - \min_i P_1 \cdots P_u(i, j) \right).
\]
THANK YOU!

WOW, what an audience...