Coalescence Theory, Structured Populations with Fast Migration

Ola Hössjer

April, 2011

Department of Mathematics, Stockholm University

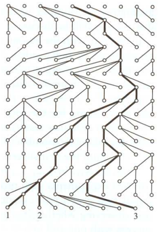
Wright-Fisher model

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

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

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

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



$N = 10$, $n = 3$

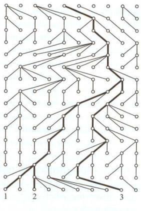
Discrete time coalescence process

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

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

$$\begin{aligned} P(X_N(\tau+1) < a | X_N(\tau) = a) &= 1 - \prod_{b=1}^{a-1}(1 - b/N) \\ &= \binom{a}{2}/N + o(N^{-1}),\ a = 2, \ldots, n \end{aligned}$$

.



$$X_N(0) = 3,\ X_N(2) = 2,\ X_N(9) = 1$$

.

Continuous time coalescence process

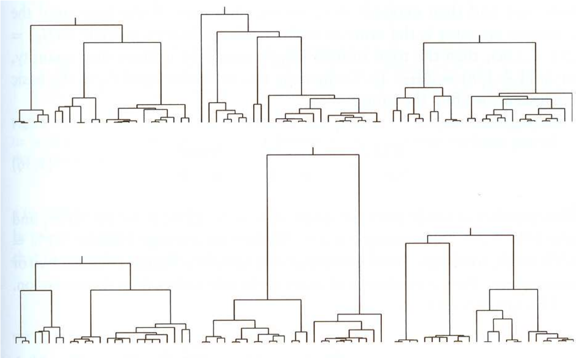
Let $N \to \infty$, keep n fixed.
Rescale time by factor N:

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

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

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

Six simulated Kingman coalescents with $n = 25$



$$T_a = |\{t;\, A(t) = a\}| \sim \mathrm{Exp}\left(\binom{a}{2}\right).$$

See Hein, Schierup and Wiuf (2005).

Generalizations and *some* references

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

$$\begin{aligned} N_e &= N/c \\ &= \text{coalescence effective population size} \\ &= \text{size of WF model with same ancestry asymptotically.} \end{aligned}$$

Nordborg and Krone (2002), Sjödin et al. (2005).

Coalescence rate and individual variability in reproductivity

One-sex population, constant size N
Nonoverlapping generations

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

$$(\nu^l)_{l=1}^N \text{ exchangeable random variables.}$$

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

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

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

$$
\begin{aligned}
N &= \text{population size} \\
L &= \text{number of subpopulations} \\
Na_i &= \text{size of subpopulation } i \;\; (\textstyle\sum_{i=1}^{L} a_i = 1), \\
m_{ki} &= \text{``migration rate'' from subp. } k \text{ to } i, \\
\nu_{ki}^{l} &= \text{nr of ``offspring'' of } l\text{th individual of subpop. } k \text{ that} \\
& \quad\; \text{end up in subpop. } i \text{ (possibly including parent itself)}
\end{aligned}
$$

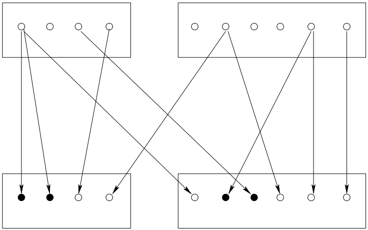
Constant subpopulation sizes is formulated as

$$
\begin{aligned}
\textstyle\sum_{l=1}^{Na_k} \nu_{ki}^{l} &= Na_k m_{ki}, \\
\textstyle\sum_{k=1}^{L} a_k m_{ki} &= a_i,
\end{aligned}
$$

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

$$
\begin{aligned}
X_{Ni}(\tau) &= \text{nr of ancestors of sample in subpop. } i, \tau \text{ generations back} \\
X_N(\tau) &= \text{total nr of ancestors of sample } \tau \text{ generations back} \\
&= \sum_{i=1}^{L} X_{Ni}(\tau), \\
X_N(0) &= \text{sample size} \\
&= n.
\end{aligned}
$$

Under certain conditions

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

in Skorohood 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

$$\begin{aligned} b_{ik} &= P(\text{parent of subpop. } i \text{ individual from subpop. } k) \\ &= a_k m_{ki}/a_i \end{aligned}$$

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

ν_{ki}^{l} = nr of children of lth individual of subpop k born in subpop i.

If WF type reproduction

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

gives local coalescence rate

$$c_{kij} = 1/a_k$$

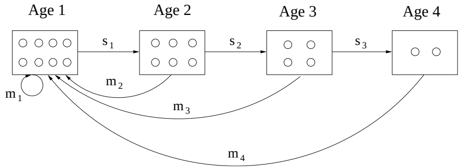
and

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

see Nordborg and Krone (2002).

Example 2: Age structured models

Subpopulation = age class



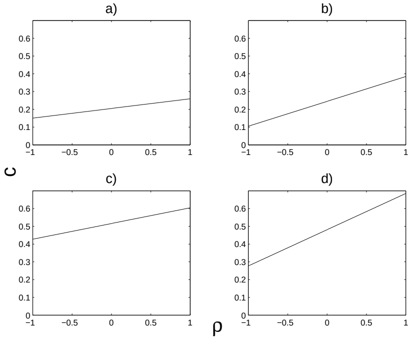
$$\begin{aligned} L &= \text{nr of age classes} = 4, \\ m_i &= m_{i1} = \text{exp. nr of children of parents of age } i, \\ s_i &= m_{i,i+1} = \text{survival prob from age class } i \text{ to } i+1, \end{aligned}$$

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).

Coalescence rate, age structured models

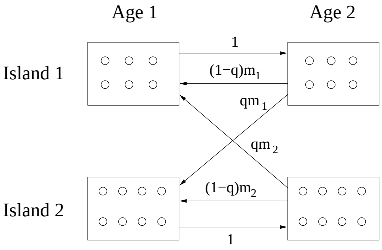


$$\rho = \text{Corr}(\nu^{l}_{k1}, \nu^{l}_{k,k+1}) = \text{correl. between nr of children and survival}$$

Figure	(s_1, s_2, s_3)	(m_1, m_2, m_3, m_4),	$c_{k11}a_k$
a)	(1, 1, 0.5)	(2,2,2,2)/7	-3/2
b)	(1, 1, 0.5)	(0,0,2,2)/3	1/2
c)	(0.5, 0.5, 0.5)	(8,8,8,8)/15	1/8
d)	(0.5, 0.5, 0.5)	(0,0,8,8)/3	13/8

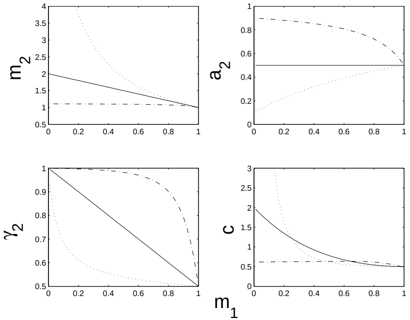
Example 3: Combined age and geographical structure

Subpopulation = (island nr,age nr)



$L = 4$
$N = 28$
$a_{(1,1)} = a_{(1,2)} = 3/14, a_{(2,1)} = a_{(2,2)} = 2/7$
q = prob that *all* children are born in island *different* from parent
m_i = fertility of adults of island i
$m_{(1,1),(1,2)} = 1, m_{(1,2),(2,1)} = qm_1$ 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

Summary

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 $\Longrightarrow$ effective population size

Proof uses partially techniques and results from Möhle (1998a), Kaj et al (2001) and Nordborg and Krone (2002).

References

Fu, Y.X. (1997). Coalescent theory for a partially selfing population. *Genetics* **146**, 1489-1499.

Hein, J., Schierup, M.H. and Wiuf, C. (2005). *Gene genealogies, variation and evolution.* Oxford University Press, Oxford.

Jagers, P. and Sagitov, S. (2004). Convergence to the coalescent in populations of substantially varying size. *J. Appl. Prob.* **41**, 368-378.

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

Kingman, J.F.C. (1982a). The coalescent. *Stoch. Proc. Appl.* **13**, 235-248.

Kingman, J.F.C. (1982b). On the genealogy of large populations. In *Essays in Statistical Science: Papers in Honour of P.A.P. Moran*, eds. Gani, J. and Hannan, E.J..

Möhle, M. (1998a). A convergence theorem for Markov chains arising in population genetics and the coalescent with selfing. *Adv. Appl. Prob.* **30**, 493-512.

Möhle, M. (1998b). Robustness results for the coalescent. *J. Appl. Prob.* **35**, 438-447.

Möhle, M. (1998c). Coalescent results for two-sex population models. *Adv. Appl. Probab.* **30**, 513-520.

Nordborg, M. and Donnelly, P. (1997). The coalescent process with selfing. *Genetics* **146**, 1185-1195.

Nordborg, M. and Krone, S. (2002). Separation of time scales and convergence to the coalescent in structured populations. In *Modern Development in Theoretical Population Genetics*, Slatkin, M. and Veuille, M., eds., 194-232. Oxford Univ Press.

Sagitov, S. and Jagers. P. (2005). The coalescent effective size of age-structured populations. *Ann. Appl. Probab.* **15**(3), 1778-1797.

Sjödin, P., Kaj, I., Krone, S., Lascoux, M. and Nordborg, M. (2005). On the meaning and existence of an effective population size. *Genetics* **169**, 1061-1070.