

Stochastic Demography, Coalescents, and Effective Population Size

Steve Krone

University of Idaho
Department of Mathematics & IBEST

I. Kaj, Uppsala Univ. (Mathematics)

M. Nordborg, USC (Molecular and Computational Biology)

M. Lascoux, Uppsala Univ. (Cons. Biology & Genetics)

P. Sjödin, Uppsala Univ. (Cons. Biology & Genetics)

NSF DMS-00-72198

NIH P20 RR16448

Wright–Fisher model

- discrete time (generations)
- constant population size N
- panmictic
- no selection, no recombination
- ancestry: each individual chooses (haploid) parent at random (prob $1/N$ each) from previous generation

Effective population size

Other population models (reproduction, variable pop size, structure, . . .) sometimes behave **in certain respects** like a W-F model with an “effective population size” N_e .

- inbreeding effective size (probability of identity by descent)
- variance effective size (variance in offspring allele frequency)
- eigenvalue effective size (leading non-unit eigenvalue for allele frequency transition matrix)
- “**coalescent effective size**” (if it exists) supersedes all of these.

The coalescent

- $P(2 \text{ indiv choose same parent}) = 1/N$
- Takes $O(N)$ generations to find common ancestor (per pair)
- Measure time in units of N generations $\dots [Nt]$
- $A_N(\tau) = \# \text{ ancestors } \tau \text{ generations in past}$
- $A_N([Nt]) \Rightarrow A(t) \dots$ Kingman coalescent

All genetic information about a sample (**polymorphism data**) is embedded in the coalescent.

Fu and Li's F statistic

$$F = F(\pi, \eta_s, S) = \frac{\pi - \left(\frac{n-1}{n}\right)\eta_s}{\sqrt{c_1 S + c_2 S^2}}$$

where n = sample size

π = ave. # pairwise differences (influenced by deep branches)

η_s = # singletons (influenced by external branches)

S = # segregating sites

Tajima's D statistic

$$D = D(\pi, \eta_s, S) = \frac{\pi - \frac{S}{a_n}}{\sqrt{c'_1 S + c'_2 S^2}}$$

where

$$a_n = \sum_{i=1}^{n-1} \frac{1}{i}$$

Both statistics have mean ≈ 0 , variance ≈ 1 .

Deviations from assumptions (neutrality, constant pop size, panmixia,...) produce changes in F and D .

Relative time scales

Coalescence events have prob $\sim O(1/N)$.

- Events that are “faster” have prob $\sim O(1/N^\alpha)$, where $0 \leq \alpha < 1$. Effects appear in coalescent only in average sense. (All demographic processes “fast” \Rightarrow coalescent effective size exists.)
- Events with prob $\sim O(1/N)$ are incorporated in the coalescent and affect pattern of variation in nonhomogeneous way. (No coalescent effective size)

Fluctuating population size

(backward) size process $M_N(1), M_N(2), M_N(3), \dots$

Markov chain with state space $\{N_1, N_2, \dots\}$

$$N_i = Nx_i$$

How does this affect the coalescent?

Depends on time it takes for “large” size changes (i.e., $O(N)$) to occur.

Harmonic mean size

Special case: $M_N(1), M_N(2), M_N(3), \dots$ (i.i.d.)

with $p_i = P(M_N(\tau) = N_i)$

P_2 (no coalescence in $[Nt]$ generations)

$$\begin{aligned} &= E\left[\prod_{\tau=1}^{[Nt]} \left(1 - \frac{1}{M_N(\tau)}\right)\right] \\ &= \left(1 - \sum_i p_i \cdot \frac{1}{Nx_i}\right)^{[Nt]} \rightarrow \exp\{-t \sum p_i/x_i\} \end{aligned}$$

⇒ limiting coalescent is a linear time change of standard coalescent:

$$A_N([Nt]) \Rightarrow A(ct)$$

where $c = \sum \frac{p_i}{x_i}$. . . pairwise coalescence rate

⇒ pairwise coalescence prob $\approx \frac{1}{N} \sum \frac{p_i}{x_i} \equiv \frac{1}{N_e}$

⇒ $N_e = \left(\sum \frac{p_i}{N_i} \right)^{-1}$... harmonic mean of sizes

This is the “**coalescent effective size**”: $N_e = N/c$

General fast size fluctuations

size process stationary distribution $(\gamma_1, \gamma_2, \dots)$

P_2 (no coalescence in $[Nt]$ generations)

$$\begin{aligned} &= E \left[\prod_{\tau=1}^{[Nt]} \left(1 - \frac{1}{M_N(\tau)} \right) \right] \\ &\sim \left(1 - \sum_i \gamma_i \cdot \frac{1}{Nx_i} \right)^{[Nt]} \rightarrow \exp \left\{ -t \sum \gamma_i / x_i \right\} \end{aligned}$$

Limiting coalescent . . . linear time change of standard coalescent:

$$A_N([Nt]) \Rightarrow A(ct)$$

where $c = \sum \frac{\gamma_i}{x_i}$. . . pairwise coalescence rate

$\Rightarrow N_e = \frac{N}{c} = \left(\sum \frac{\gamma_i}{N_i} \right)^{-1}$... harmonic mean of sizes

Intermediate fluctuations—stochastic time change

What if macroscopic changes in pop. size (i.e., $O(N)$) occur on coalescent time scale (i.e., $O(N)$ generations)?

Pop. size τ generations in past (Markov chain):

$$M_N(\tau) = NX_N(\tau),$$

where relative size proc. $X_N([Nt]) = \frac{M_N([Nt])}{N} \Rightarrow X(t)$

... cont-time Markov (e.g., diffusion proc. or cont-time jump chain)

“Large” size changes occur on same time scale as coalescence events; do not “average out.” Limiting coalescent is of form

$$A_N([Nt]) \Rightarrow A(Y(t)),$$

where the time change

$$Y(t) \equiv \int_0^t \frac{1}{X(s)} ds$$

is nonlinear and stochastic (coalescence intensity).

No (coalescent) effective size! Behavior different from any standard W-F model. Effects should show up in **polymorphism data**.

Idea

$$\begin{aligned} P_2(\text{no coalescence in } [Nt] \text{ generations} | \{M_N(\cdot)\}) \\ &= \prod_{\tau=1}^{[Nt]} \left(1 - \frac{1}{M_N(\tau)}\right) \\ &= \prod_{\tau=1}^{[Nt]} \left(1 - \frac{1}{NX_N(\tau)}\right) \\ &\sim \exp\left(-\frac{1}{N} \sum_{\tau=1}^{[Nt]} \frac{1}{X_N(\tau)}\right) \Rightarrow \exp\left(-\int_0^t \frac{1}{X(s)} ds\right) \end{aligned}$$

Time change

$$\int_0^t \frac{1}{X_s} ds = \int_E \frac{1}{x} \cdot L_t^x m(dx)$$

L_t^x . . . diffusion local time

$m(dx)$. . . speed measure

More general demography

Let

$$c_N(M_N(\tau - 1), M_N(\tau))$$

denote prob. that two lineages coalesce when going from gen. $\tau - 1$ to gen. τ (in past). Assume

$$c_N(k, m) = \frac{1}{N} H_N\left(\frac{k}{N}, \frac{m}{N}\right),$$

where $H_N\left(\frac{k}{N}, \frac{m}{N}\right) \rightarrow H(x, y)$ as $k/N \rightarrow x$ and $m/N \rightarrow y$.
Time change becomes

$$\int_0^t H(X_s, X_s) ds.$$

Cannings-type models

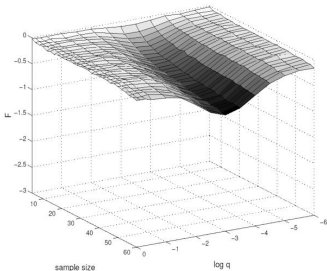
$$\begin{aligned}c_N(M_N(\tau - 1), M_N(\tau)) \\ = \frac{1}{(M_N(\tau - 1))_2} \sum_{i=1}^{M_N(\tau)} E[(\nu_i^{(\tau)})_2]\end{aligned}$$

$\nu_i^{(\tau)}$. . . number of offspring produced by i th indiv in gen τ .
With exchangeable reproduction, get

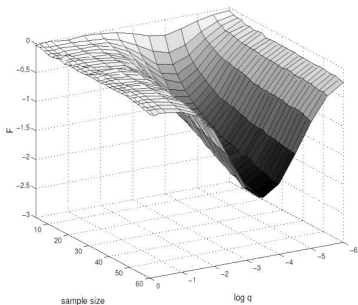
$$\begin{aligned}H_N\left(\frac{k}{N}, \frac{m}{N}\right) \\ = \left(\frac{k}{N}\left(\frac{k}{N} - \frac{1}{N}\right)\right)^{-1} \frac{md}{N} \rightarrow \frac{yd}{x^2} \equiv H(x, y)\end{aligned}$$

Simulations for fluctuating size

2 sizes N_1, N_2 ; equal prob of size change $q_1 = q_2 \equiv q$; mutation prob $u = .001$; 10,000 runs per data pt.; stationary starting size. Plot of F_u and Li's F



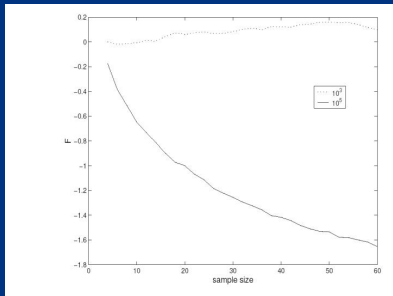
$$N_1 = 10^3, N_2 = 10^4$$



$$N_1 = 10^3, N_2 = 10^5$$

Rule of thumb: $q \in \left(\frac{10^{-1}}{N_2}, \frac{10^1}{N_1}\right) \Rightarrow$ no averaging;
too close to coalescent scale.

Dependence on initial size



$$N_1 = 10^3, N_2 = 10^5; q_1 = q_2 = 10^{-4}.$$

Top curve: initial size 10^3

Bottom curve: initial size 10^5

Structured Populations

Population of total size N , subdivided into L islands (demes), connected by migration. Pop. size in deme k is $N_k = Na_k$ ($a_1 + \dots + a_L = 1$).

- Migration on **same time scale** as coalescence events (i.e., migration prob. for lineage $b_{ij} = \beta_{ij}/N$)

⇒ limiting coalescent is “structured.” (no averaging, no coalescent effective size)

- **Fast** migration (i.e., $b_{ij} = \beta_{ij}/N^\alpha$, $0 \leq \alpha < 1$), and stationary distribution for locations $(\gamma_1, \gamma_2, \dots, \gamma_L)$

⇒ averaging occurs w/ coalescent time change

$$c = \sum_{k=1}^L \frac{\gamma_k^2}{a_k}.$$

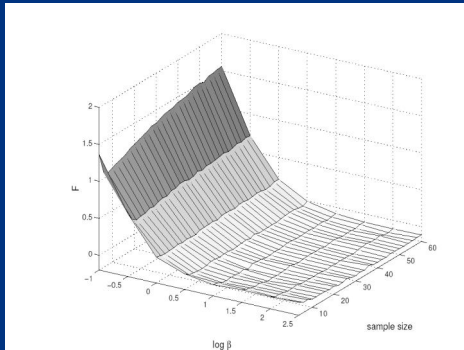
⇒ coalescent effective size is

$$N_e = \frac{N}{c} = \left(\sum \frac{\gamma_k^2}{N_k} \right)^{-1} \quad \text{“harmonic mean”}$$

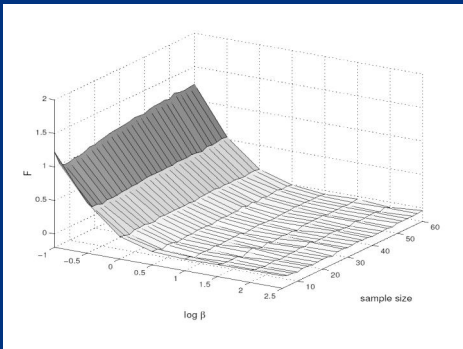
In case of fast migration, structured model can be thought of as panmictic W-F model with pop. size N_e .

Simulations for population subdivision

2 demes, equal size, equal migration rate $\beta = 2Nb$



$$N = 10^3$$



$$N = 10^4$$

- P. Sjödin, I. Kaj, S. Krone, M. Lascoux, M. Nordborg (2004) On the meaning and existence of an effective population size. Submitted to *Genetics*.
- I. Kaj and S. Krone (2003) The coalescent process in a population with stochastically varying size. *J. Appl. Probab.* 40: 33-48.
- M. Nordborg and S. Krone (2002) Separation of time scales and convergence to the coalescent in structured populations. In *Modern Developments in Theoretical Population Genetics*. M. Slatkin and M. Veuille (eds.).