Exact Transition Probabilities of the Number of Ancestral Sample Lineages in the Discrete Ancestral Recombination Graph

Raazesh Sainudiin[†] and Bhalchandra Thatte[*]

[†]Biomathematics Research Centre
University of Canterbury, Christchurch, New Zealand
`http://www.math.canterbury.ac.nz/~r.sainudiin/`

[*]Department of Statistics,
University of Oxford, Oxford, United Kingdom

ICM Satellite Meeting in Prob. & Stoch. Proc., Bangalore, India
August 14, 2010

Introduction with Disclaimer

Introduction with Disclaimer

- Somewhat basic maths...

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

The Big Picture:

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

The Big Picture:

- We have an asymptotically-neat Continuous-time model.

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

The Big Picture:

- We have an asymptotically-neat Continuous-time model.
- The model can get the sample genealogy — "space-time interrelations" — of size n from a population of size N

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

The Big Picture:

- We have an asymptotically-neat Continuous-time model.
- The model can get the sample genealogy — "space-time interrelations" — of size n from a population of size N
- The model requires large $N > 10^3$ for good approximation of small samples $2 < n < 100$.

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

The Big Picture:

- We have an asymptotically-neat Continuous-time model.
- The model can get the sample genealogy — "space-time interrelations" — of size n from a population of size N
- The model requires large $N > 10^3$ for good approximation of small samples $2 < n < 100$.
- The model isn't so fantastic for:
  - small N — currently threatened species.
  - for large n – current human genome samples in 10^3's

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

The Big Picture:

- We have an asymptotically-neat Continuous-time model.
- The model can get the sample genealogy — "space-time interrelations" — of size n from a population of size N
- The model requires large $N > 10^3$ for good approximation of small samples $2 < n < 100$.
- The model isn't so fantastic for:
  - small N — currently threatened species.
  - for large n – current human genome samples in 10^3's
- Discrete time transition probabilities do not exist in literature.

Introduction with Disclaimer

- Somewhat basic maths...
- ...but somewhat interesting.

The Big Picture:

- We have an asymptotically-neat Continuous-time model.
- The model can get the sample genealogy — "space-time interrelations" — of size n from a population of size N
- The model requires large $N > 10^3$ for good approximation of small samples $2 < n < 100$.
- The model isn't so fantastic for:
  - small N — currently threatened species.
  - for large n – current human genome samples in 10^3's
- Discrete time transition probabilities do not exist in literature.
- We set out to compute them.

Outline

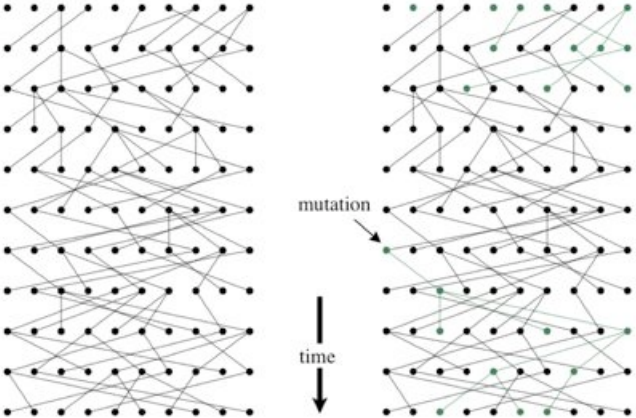
- Wright-Fisher Model – vanilla version
- The *n*-Coalescent Approximation
- Wright-Fisher Model with Recombination
- Ancestral Recombination Graph
- Exact Transition Probabilities
- Black Robins — a living example
- Detailed Derivation
- Blabber on Further Work

The Wright-Fisher Model – 1

Random Mating, Constant Size, No Recombination/Selection

A Population of $N = 10$ homologous DNA seqns. of length m and the Population History of site i

```
  : 1 2 3 4 5 6 7 8 9 10
1 : A A A A A A A A A C
2 : G G G G G G G G G G
...
i : T T A A A A A A A A
...
k : ...
```

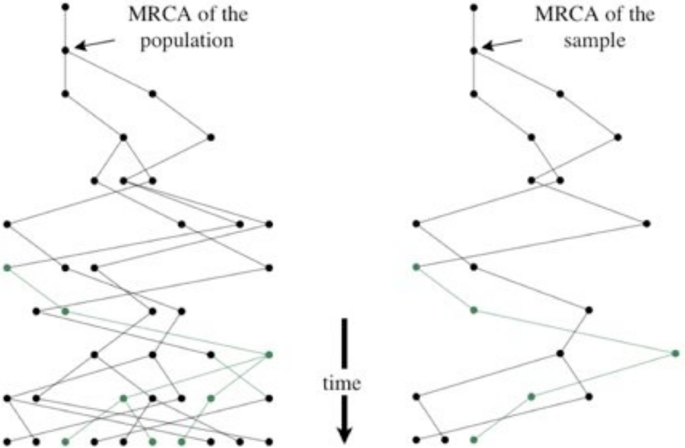


The Wright-Fisher Model – 2

Random Mating, Constant Size, No Recombination/Selection

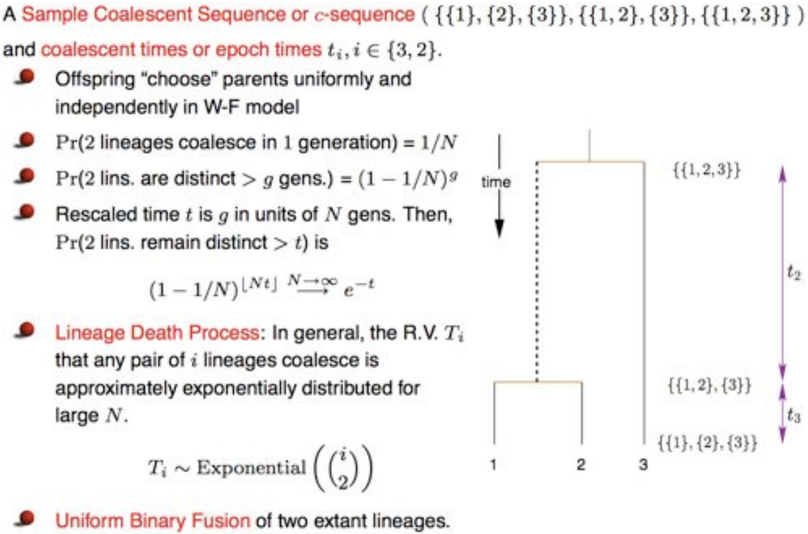
Ex: Data of 3 homologous DNA sequences at site i, its Population History and the Sample History of sampled individuals 1,2, and 3.

```
  : 1 2 3
i : T T A
```



The Wright-Fisher Model & the *n*-Coalescent – 1

Random Mating, Constant Size, No Recombination/Selection



A Sample Coalescent Sequence or c-sequence ($\{\{1\},\{2\},\{3\}\},\{\{1,2\},\{3\}\},\{\{1,2,3\}\}$) and coalescent times or epoch times $t_i, i \in \{3,2\}$.

- Offspring “choose” parents uniformly and independently in W-F model
- $\Pr$(2 lineages coalesce in 1 generation) = $1/N$
- $\Pr$(2 lins. are distinct $> g$ gens.) = $(1-1/N)^g$
- Rescaled time t is g in units of N gens. Then, $\Pr$(2 lins. remain distinct $> t$) is

$$(1-1/N)^{\lfloor Nt \rfloor} \overset{N\to\infty}{\longrightarrow} e^{-t}$$

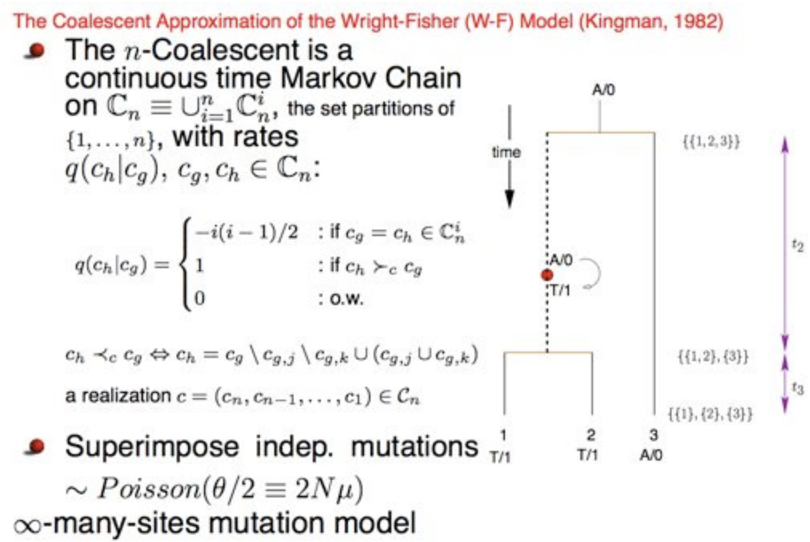
- Lineage Death Process: In general, the R.V. T_i that any pair of i lineages coalesce is approximately exponentially distributed for large N.

$$T_i \sim \text{Exponential}\left(\binom{i}{2}\right)$$

- Uniform Binary Fusion of two extant lineages.

The Wright-Fisher Model & the n-Coalescent – 2

Random Mating, Constant Size, No Recombination/Selection



The Coalescent Approximation of the Wright-Fisher (W-F) Model (Kingman, 1982)

- The n-Coalescent is a continuous time Markov Chain on $\mathbb{C}_n \equiv \cup_{i=1}^{n} \mathbb{C}_n^i$, the set partitions of $\{1,\ldots,n\}$, with rates $q(c_h|c_g)$, $c_g, c_h \in \mathbb{C}_n$:

$$q(c_h|c_g) = \begin{cases} -i(i-1)/2 & : \text{if } c_g = c_h \in \mathbb{C}_n^i \\ 1 & : \text{if } c_h \succ_c c_g \\ 0 & : \text{o.w.} \end{cases}$$

$c_h \prec_c c_g \Leftrightarrow c_h = c_g \setminus c_{g,j} \setminus c_{g,k} \cup (c_{g,j} \cup c_{g,k})$

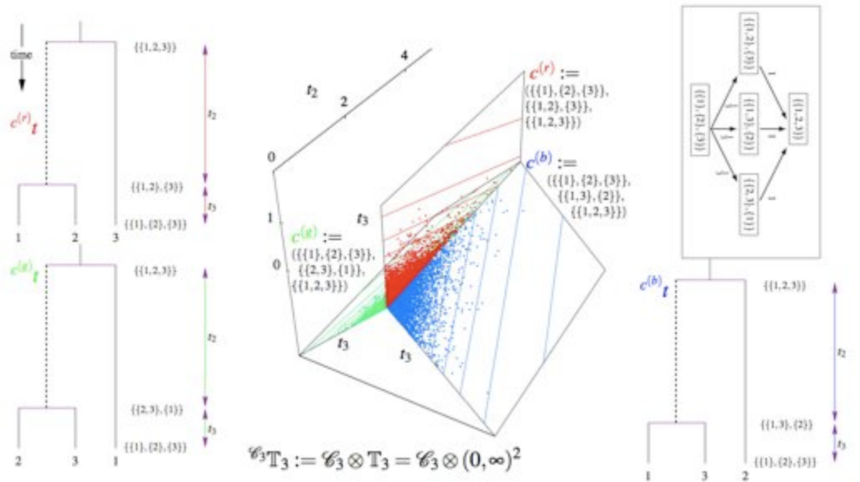
a realization $c = (c_n, c_{n-1}, \ldots, c_1) \in \mathcal{C}_n$

- Superimpose indep. mutations $\sim Poisson(\theta/2 \equiv 2N\mu)$

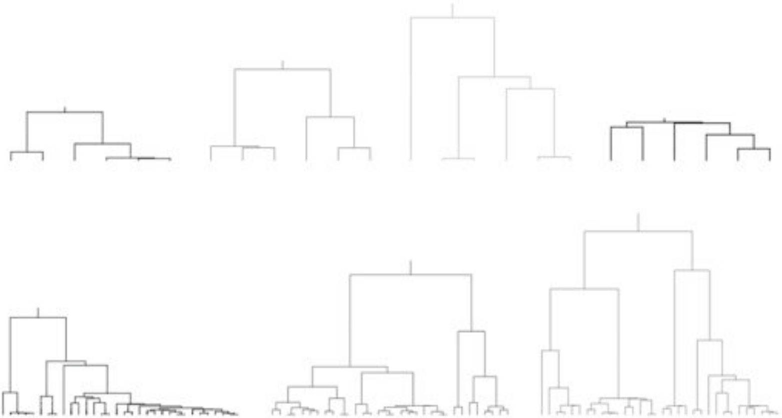
∞-many-sites mutation model

The n-Coalescent for $n = 3$

Random Mating, Constant Size, No Recombination/Selection – The Coalescent Tree Space

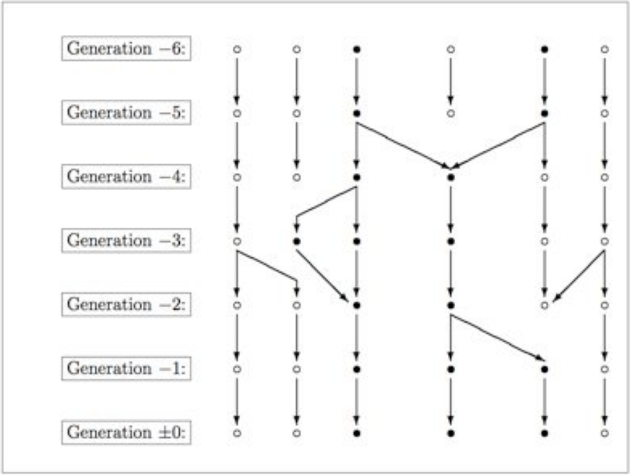


Realisations from the n-Coalescent for $n = 6$ and $n = 32$



The Wright-Fisher Model with Recombination – 1

Random Mating, Constant Size, Recombination, No Selection



The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing
- Label non-overlapping generations (forward in time) by $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, \pm 0, +1, +2, \ldots$

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing
- Label non-overlapping generations (forward in time) by $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, \pm 0, +1, +2, \ldots$
- Let r be the prob. of recombination per locus per generation

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing
- Label non-overlapping generations (forward in time) by $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, \pm 0, +1, +2, \ldots$
- Let r be the prob. of recombination per locus per generation
- Label the $2N$ individuals at $-k$ using $[2N]_+ := \{1, 2, \ldots, 2N\}$

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing
- Label non-overlapping generations (forward in time) by $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, \pm 0, +1, +2, \ldots$
- Let r be the prob. of recombination per locus per generation
- Label the $2N$ individuals at $-k$ using $[2N]_+ := \{1, 2, \ldots, 2N\}$
- $V_i =$ the number of non-recombinant offspring of individual i

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing
- Label non-overlapping generations (forward in time) by $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, \pm 0, +1, +2, \ldots$
- Let r be the prob. of recombination per locus per generation
- Label the $2N$ individuals at $-k$ using $[2N]_+ := \{1, 2, \ldots, 2N\}$
- $V_i =$ the number of non-recombinant offspring of individual i
- $U_{i,j} =$ the number of offspring that are recombinants of the haploid pair labeled by $\{i, j\}$ with $i < j$

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing
- Label non-overlapping generations (forward in time) by $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, \pm 0, +1, +2, \ldots$
- Let r be the prob. of recombination per locus per generation
- Label the $2N$ individuals at $-k$ using $[2N]_+ := \{1, 2, \ldots, 2N\}$
- $V_i =$ the number of non-recombinant offspring of individual i
- $U_{i,j} =$ the number of offspring that are recombinants of the haploid pair labeled by $\{i, j\}$ with $i < j$
- Let the non-recombinant and recombinant offspring numbers be $\Sigma V_\bullet := \sum_{i=1}^{2N} V_i$ and $\Sigma U_\bullet := \sum_{\{i,j \in [2N]_+ : i<j\}} U_{i,j}$, resp.

The Wright-Fisher Model with Recombination – 2

Random Mating, Constant Size, Recombination, No Selection

- $2N$ panmictic haploid individuals are meiotically reproducing
- Label non-overlapping generations (forward in time) by $\ldots, -k, -k+1, -k+2, \ldots, -2, -1, \pm 0, +1, +2, \ldots$
- Let r be the prob. of recombination per locus per generation
- Label the $2N$ individuals at $-k$ using $[2N]_+ := \{1, 2, \ldots, 2N\}$
- $V_i =$ the number of non-recombinant offspring of individual i
- $U_{i,j} =$ the number of offspring that are recombinants of the haploid pair labeled by $\{i, j\}$ with $i < j$
- Let the non-recombinant and recombinant offspring numbers be $\Sigma V_\bullet := \sum_{i=1}^{2N} V_i$ and $\Sigma U_\bullet := \sum_{\{i,j \in [2N]_+ : i<j\}} U_{i,j}$, resp.
- Then the lines of descent of the $2N$ homologous haploid loci into the next generation, follow from the multinomial random vector $(V, U) := (V_1, V_2, \ldots, V_{2N}, U_{1,2}, U_{1,3}, \ldots, U_{2N-1,2N})$ of length $2N + \binom{2N}{2}$, such that, $\Sigma V_\bullet + \Sigma U_\bullet = 2N$

The Wright-Fisher Model with Recombination – 3

Random Mating, Constant Size, Recombination, No Selection

Thus, we have the following multinomial scheme:

The Wright-Fisher Model with Recombination – 3

Random Mating, Constant Size, Recombination, No Selection

Thus, we have the following multinomial scheme:

$$\begin{aligned} P(V,U) &:= P(V_1 = v_1, \ldots, V_{2N} = v_{2N}, U_{1,2} = u_{1,2}, \\ &\qquad \ldots, U_{2N-1,2N} = u_{2N-1,2N}) \\ &= \frac{(2N)!}{v_1! \cdots v_{2N}! u_{1,2}! \cdots u_{2N-1,2N}!} r^{\Sigma u_\bullet} \binom{2N}{2}^{-\Sigma u_\bullet} (1-r)^{\Sigma v_\bullet} \left(\frac{1}{2N}\right)^{\Sigma v_\bullet} \quad . \end{aligned} \tag{1}$$

The Wright-Fisher Model with Recombination – 3

Random Mating, Constant Size, Recombination, No Selection

Thus, we have the following multinomial scheme:

$$\begin{aligned} P(V,U) &:= P(V_1 = v_1, \ldots, V_{2N} = v_{2N}, U_{1,2} = u_{1,2}, \\ &\qquad \ldots, U_{2N-1,2N} = u_{2N-1,2N}) \\ &= \frac{(2N)!}{v_1! \cdots v_{2N}! u_{1,2}! \cdots u_{2N-1,2N}!} r^{\Sigma u_\bullet} \binom{2N}{2}^{-\Sigma u_\bullet} (1-r)^{\Sigma v_\bullet} \left(\frac{1}{2N}\right)^{\Sigma v_\bullet} \quad . \end{aligned} \tag{1}$$

This reproduction scheme is independently and identically enforced in each generation to obtain the Wright-Fisher **population genealogy** with recombination as we go **forward in time**.

The Wright-Fisher Model with Recombination – 3

Random Mating, Constant Size, Recombination, No Selection

Thus, we have the following multinomial scheme:

$$\begin{aligned} P(V,U) &:= P(V_1 = v_1, \ldots, V_{2N} = v_{2N}, U_{1,2} = u_{1,2}, \\ &\qquad \ldots, U_{2N-1,2N} = u_{2N-1,2N}) \\ &= \frac{(2N)!}{v_1! \cdots v_{2N}! u_{1,2}! \cdots u_{2N-1,2N}!} r^{\Sigma u_\bullet} \binom{2N}{2}^{-\Sigma u_\bullet} (1-r)^{\Sigma v_\bullet} \left(\frac{1}{2N}\right)^{\Sigma v_\bullet} \quad . \end{aligned} \tag{1}$$

This reproduction scheme is independently and identically enforced in each generation to obtain the Wright-Fisher **population genealogy** with recombination as we go **forward in time**.
Now we want to track the **sample genealogy** of size n from current generation 0, within this population genealogy of size $2N$ as we go **backward in time** at the coarsest resolution of ancestral sample size exactly.

The Wright-Fisher Sample with Recombination – 1

The Wright-Fisher Sample with Recombination – 1

- Let I be a fixed set of haploid lineages in the current generation, $|I| = i$.

The Wright-Fisher Sample with Recombination – 1

- Let I be a fixed set of haploid lineages in the current generation, $|I| = i$.
- We are interested in the probability $P(j|I)$ that these i lineages descended from exactly j ancestral lineages in the previous generation.

The Wright-Fisher Sample with Recombination – 1

- Let I be a fixed set of haploid lineages in the current generation, $|I| = i$.
- We are interested in the probability $P(j|I)$ that these i lineages descended from exactly j ancestral lineages in the previous generation.
- $$P(j|I) = \sum_{J:|J|=j} P(J|I) = \binom{2N}{j} P(J|I) \ , \qquad (2)$$

The Wright-Fisher Sample with Recombination – 1

- Let I be a fixed set of haploid lineages in the current generation, $|I| = i$.
- We are interested in the probability $P(j|I)$ that these i lineages descended from exactly j ancestral lineages in the previous generation.
- $$P(j|I) = \sum_{J:|J|=j} P(J|I) = \binom{2N}{j} P(J|I) \ , \tag{2}$$
- where, $P(J|I)$ denotes the probability that the set of lineages ancestral to I in the previous generation is J.

The Wright-Fisher Sample with Recombination – 2

Let $P(J|I,K)$ be the probability that the set of lineages ancestral to I in the previous generation is J given that lineages in a fixed subset K of I are recombinants and the lineages in $I \setminus K$ are non-recombinants. Therefore,

$$P(J|I) = \sum_{K \subseteq I} r^{|K|}(1-r)^{|I|-|K|} P(J|I,K) \tag{3}$$

The Wright-Fisher Sample with Recombination – 2

Let $P(J|I, K)$ be the probability that the set of lineages ancestral to I in the previous generation is J given that lineages in a fixed subset K of I are recombinants and the lineages in $I \setminus K$ are non-recombinants. Therefore,

$$P(J|I) = \sum_{K \subseteq I} r^{|K|}(1-r)^{|I|-|K|} P(J|I, K) \qquad (3)$$

Next we describe how to calculate $P(J|I, K)$ for fixed sets of lineages I, J, K.

The Wright-Fisher Sample with Recombination – 3

Let $B(J|I, K)$ be the set of bipartite graphs with vertex set $I \cup J$, with bipartition $J|I$, such that the vertices in K are of degree 2, the vertices in $I \setminus K$ are of degree 1, and no vertices in J isolated.

The Wright-Fisher Sample with Recombination – 3

Let $B(J|I,K)$ be the set of bipartite graphs with vertex set $I \cup J$, with bipartition $J|I$, such that the vertices in K are of degree 2, the vertices in $I \setminus K$ are of degree 1, and no vertices in J isolated. Therefore,

$$P(J|I,K) = \frac{|B(J|I,K)|}{(2N)^{|I|-|K|}\binom{2N}{2}^{|K|}} \ , \tag{4}$$

since there are exactly $(2N)^{|I|-|K|}\binom{2N}{2}^{|K|}$ ways in which the lineages in I choose their ancestral lineages from the previous generation so that lineages in K are recombinants and the lineages in $I \setminus K$ are non-recombinants.

The Wright-Fisher Sample with Recombination – 4

Therefore, combining Eqs. 2 through 4 (and since only the cardinalities of $|I|, |J|$ and $|K|$ matter) we have the expression we wanted to find:

$$\boxed{P(j|i) = \binom{2N}{j} \sum_{k=0}^{i} \binom{i}{k} r^k (1-r)^{i-k} \frac{|B(j|i,k)|}{(2N)^{i-k} \binom{2N}{2}^k}} \qquad (5)$$

The Wright-Fisher Sample with Recombination – 5

From Eq. 5, the N-specific probability of i extant sample lineages in the current generation becoming j ancestral lineages in the previous generation is:

$$
{}^{N}P_{i,j} = \begin{cases}
\frac{(1-r)^i}{2}\left(1-\binom{i}{2}(2N)^{-1}\right) & \\
\qquad + \frac{i(i-1)(i+2)r(1-r)^{i-1}}{2N-1} + O\left((2N)^{-2}\right) & \text{if } j = i \\[2ex]
(1-r)^i\binom{i}{2}(2N)^{-1} + O\left((2N)^{-2}\right) & \text{if } j = i-1 \\[2ex]
\binom{i}{s}\frac{r^s(1-r)^{i-s}}{2}\left(1-\binom{i+s}{2}(2N)^{-1}\right) & \\
\qquad + \binom{i}{s+1} r^{s+1}(1-r)^{i-s-1}\frac{(i-s-1)(i+3s+2)+2s(s+1)}{2N-1} & \\
\qquad + O\left((2N)^{-2}\right) & \text{if } j = i+s, 0 \leq s \leq i \\[2ex]
O\left((2N)^{-2}\right) & \text{if } j = i-\ell, 1 < \ell \leq i-1 \\[2ex]
0 & \text{otherwise.}
\end{cases}
\qquad (6)
$$

The Wright-Fisher Sample with Recombination – 6

If $j = i + s, 0 \leq s \leq i$, we can further simplify ${}^{N}P_{i,\geq i}$ to:

$$O\left((2N)^{-2}\right) + \frac{1}{2} - \sum_{s=0}^{i} \binom{i+s}{2} \binom{i}{s} \frac{r^s(1-r)^{i-s}}{4N}$$
$$+ \binom{i}{s+1} r^{s+1}(1-r)^{i-s-1} \frac{(i-s-1)(i+3s+2)+2s(s+1)}{2N-1} \quad . \tag{7}$$

The Wright-Fisher Sample with Recombination – 7

We now derive the Hudson/Griffiths-Marjoram scaling condition for linear recombination rate.

The Wright-Fisher Sample with Recombination – 7

We now derive the Hudson/Griffiths-Marjoram scaling condition for linear recombination rate. When $\rho := 2rN$ is held constant, we obtain

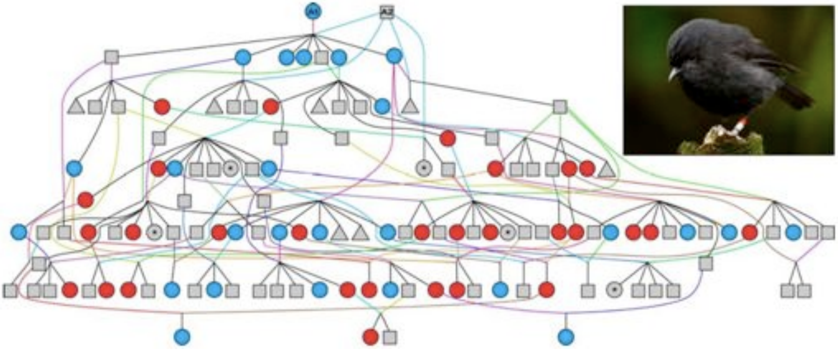
$$\begin{aligned} {}^{N}P_{i,>i} &= {}^{N}P_{i,\geq i} - {}^{N}P_{i,i} = \frac{1}{2} - \frac{(1-r)^{i}}{2} + O\left((2N)^{-2}\right) \\ &= \frac{i\rho}{4N} + O\left((2N)^{-2}\right) \end{aligned} \tag{8}$$

Therefore, in the discrete Wright-Fisher model with recombination, unlike the continuous-time ARG, the number of ancestral lineages does not necessarily reach 1 and may even oscillate about a critical ancestral size determined by r and N.

New Zealand Black Robin – a living Adam–Eve Story!

non-Random Mating, non-Constant Size, Recombination & managed-Selection in Applied Conservation Genetics

Joint work with Melanie Massaro, Antony Poole and Marie Hale, Biologists at Univ. of Canterbury, NZ.



Approximation of (6) in detail – 1

Here we derive the approximation of (6) in detail.
To evaluate $P(j|i)$, we count the number of bipartite graphs in the class $B(J|I,K)$, where sets I, J and K are fixed, and have cardinalities i, j and k, respectively. Only the cardinalities of I, J, K are relevant to counting $|B(J|I,K)|$, therefore, we write $|B(j|i,k)|$. First, we consider the case where $j = i + s$ and $s \geq 0$. In this case $k \geq s$. Let $k = s + m$ with $m \geq 0$.

$$\begin{aligned} & \binom{2N}{j} \frac{1}{(2N)^{i-k} \binom{2N}{2}^k} \\ = \; & \binom{2N}{i+s} \frac{2^k}{(2N)^i (2N-1)^k} \\ = \; & \frac{2^{s+m}}{(i+s)!} \frac{(2N)^s}{(2N-1)^{s+m}} \left(1 - \binom{i+s}{2}(2N)^{-1} + O\left((2N)^{-2}\right)\right) \\ = \; & O\left((2N)^{-m}\right) \qquad (9) \end{aligned}$$

Approximation of (6) in detail – 2

Therefore, we compute $|B(j|i,k)|$ only when $j = i + s$ and $k = s$ or $s + 1$, since the remaining contribution to $P(j|i)$ is $O\left((2N)^{-2}\right)$. By a direct counting we obtain

$$B(i+s|i,s) = \frac{(i+s)!}{2^s} \qquad \text{and}$$

$$B(i+s|i,s+1) = \frac{(i+s)!\,((i-s-1)(i+3s+2)+2s(s+1))}{2^{s+1}}$$

Now the first and the third expressions in (6) are obtained by substituting $B(i+s|i,s)$ and $B(i+s|i,s+1)$ in (5), and approximating with (10):

$$\frac{N_{[j]}}{N^j} := \frac{N(N-1)\cdots(N-(j-i))}{N^j} = \prod_{k=1}^{j-1}\left(1-kN^{-1}\right) 1 - \binom{j}{2} N^{-1} + O\left(N^{-2}\right) \quad . \tag{10}$$

Approximation of (6) in detail – 3

Next, we consider the case when $j = i - s$, where $s > 0$. In this case,

$$\begin{aligned}
& \binom{2N}{j} \frac{1}{(2N)^{i-k} \binom{2N}{2}^k} \\
= \; & \binom{2N}{i-s} \frac{2^k}{(2N)^i (2N-1)^k} \\
= \; & \frac{2^k}{(i-s)!} \frac{1}{(2N)^s (2N-1)^k} \left(1 - \binom{i-s}{2} (2N)^{-1} + O\left((2N)^{-2}\right) \right) \\
= \; & O\left((2N)^{-s-k}\right) \qquad (11)
\end{aligned}$$

Approximation of (6) in detail – 4

Therefore, we need to compute $|B(j|i,k)|$ only in the case when $j = i - 1$ and $k = 0$ since the contribution from other terms will be $O\left((2N)^{-2}\right)$. By direct counting, we have

$$B(i-1|i,0) = (i-1)!\binom{i}{2},$$

which, along with (5) and (10), implies the second and the forth expressions in (6).

Finally, the last expression in (6) follows from $0 < j \leq 2i$.

For arbitrary values of i, j, k, we can count $|B(j|i,k)|$ by the following inclusion-exclusion formula, which is useful for exact calculations.

$$|B(j|i,k)| = \sum_{m=0}^{j} (-1)^m \binom{j}{m} \frac{(j-m)^i (j-m-1)^k}{2^k}.$$

The End

- Many thanks to:
    - Robert C. Grifiths for looking at the Eqs.
    - Mike Steel for a discussion of problem satement.
    - Alison Ethridge for pointing out a “thought-o” in the slides!
    - Critically endangered birds of New Zealand for inspiration.

Bibliography:

R. Fisher, *The Genetical Theory of Natural Selection*, Clarenson, Oxford, UK, 1930.

S. Wright, *Evolution in mendelian populations*, Genetics **16** (1931), 97–159.

R.R. Hudson, *Properties of a Neutral Allele Model with Intragenic Recombination*, Theoretical Population Biology **23** (1981), 183–201.

R.C. Griffiths, *The Two-Locus Ancestral Graph*, Ishwar V. Basawa, and Robert L. Taylor, eds, Selected Proceedings of the Sheffield Symposium on Applied Probability: Held at the University of Sheffield, Sheffield, August 16–19, 1989 (Hayward, CA: Institute of Mathematical Statistics, 1991), 100-117.