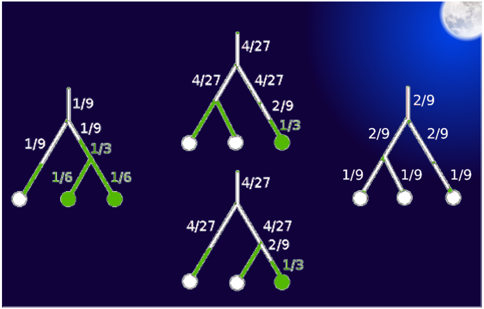
Time reversal of a Markov Chain on trees in a population-genetical context

PMSMA Montpellier 2023

Johannes Wirtz

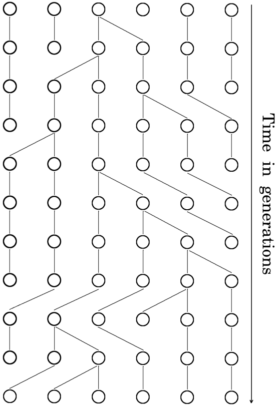


1. The Evolving Moran Genealogy

2. *EMG* time reversal

3. Transient Markov Chains

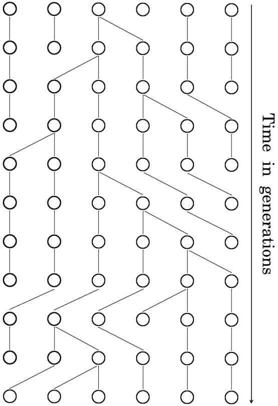
The basic population model



Moran Model (P. Moran, 1958) in discrete time:

- Constant population size n, one split and one death per generation

The basic population model



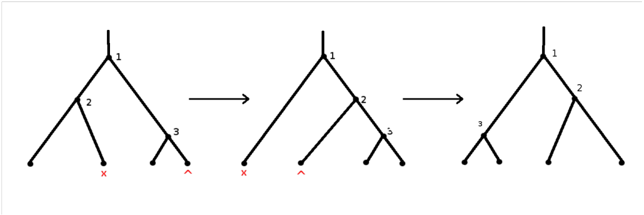
Moran Model (P. Moran, 1958) in discrete time:

- Constant population size n, one split and one death per generation
- The genealogy at any point $i > i^*$ can be interpreted as a Yule Tree T_i.

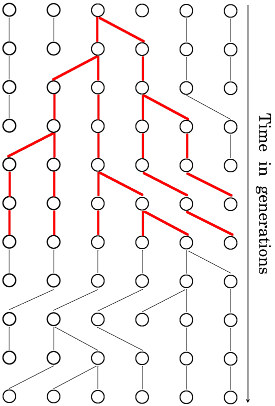
EMG definition and appearance

Definition

We call the process $(T_i)_{i\in\mathbb{N}}$, with $\Pr(T_{i+1} = T \in \mathcal{T}_n | T_i)$ dictated by the underlying Moran model, the *Evolving Moran Genealogy* (*EMG*).

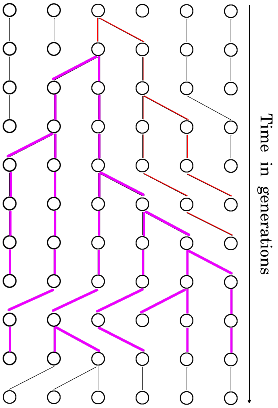


Motivating example: *MRCA* jumps



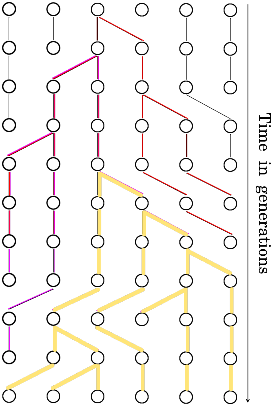
Occasionally, the root of the tree (*MRCA*) changes

Motivating example: *MRCA* jumps



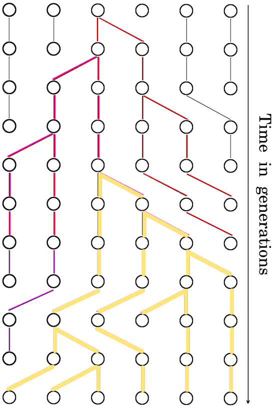
Occasionally, the root of the tree (*MRCA*) changes

Motivating example: ***MRCA*** jumps



Occasionally, the root of the tree (*MRCA*) changes

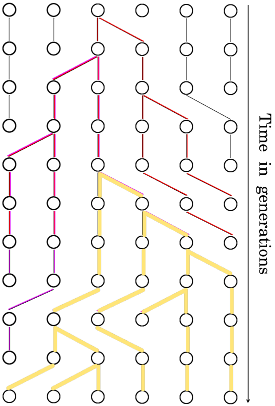
Motivating example: *MRCA* jumps



Occasionally, the root of the tree (*MRCA*) changes

Intervals in between changes are called *MRCA episodes*; changes are called *MRCA jumps* or *root jumps*.

Motivating example: *MRCA* jumps

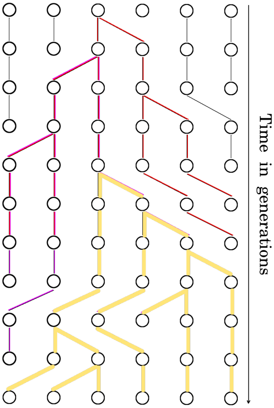


Occasionally, the root of the tree (*MRCA*) changes

Intervals in between changes are called *MRCA episodes*; changes are called *MRCA jumps* or *root jumps*.

Practical interest: Rate of fixations, loss of information (...)

Motivating example: *MRCA* jumps



Occasionally, the root of the tree (*MRCA*) changes

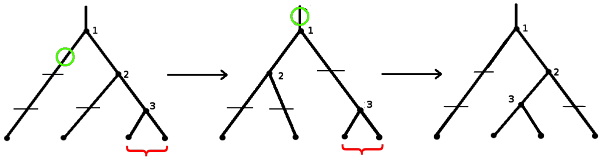
Intervals in between changes are called *MRCA episodes*; changes are called *MRCA jumps* or *root jumps*.

Practical interest: Rate of fixations, loss of information (...)

The derived process $(\mathcal{X}_i)_{i\in\mathbb{N}} \in \{0,1\}^{\mathbb{N}}$ with $\mathcal{X}_i = 1$ iff a root jump takes place from T_{i-1} to T_i is called *root jump process*.

Reversing the mechanics

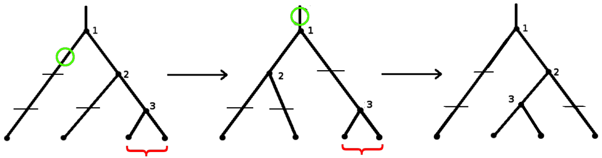
Backward in time: "Merge" the most recent split and revive an individual by "regrafting" it into the tree at some branch segment *b*.



Reversing the mechanics

Backward in time: "Merge" the most recent split and revive an individual by "regrafting" it into the tree at some branch segment *b*.

With what probability to choose the regrafting segment?



Kelly's Lemma

For a positive recurrent Markov Chain M on a state space S with transition matrix P, the matrix P' is the transition matrix of the reversed process iff

$$\forall s, t \in S : P^{\flat}_{st} = P_{ts} \cdot \frac{\pi_t}{\pi_s} \tag{1}$$

where π is the stationary distribution of M.

(Kolmogorov 1936, Kelly 1979, Lovasz & Winkler 1983)

Kelly's Lemma

For a positive recurrent Markov Chain M on a state space S with transition matrix P, the matrix P' is the transition matrix of the reversed process iff

$$\forall s, t \in S : P^{\flat}_{st} = P_{ts} \cdot \frac{\pi_t}{\pi_s} \tag{1}$$

where π is the stationary distribution of M.

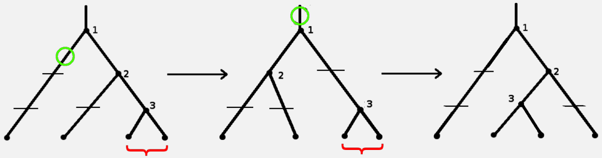
(Kolmogorov 1936, Kelly 1979, Lovasz & Winkler 1983)

Corollary (Loop Property)

Closed loops have the same probability in forward and reversed process.

Lemma (Wiehe, W. 19)

The conditions of Kelly's Lemma are satisfied for the reversed process of the EMG iff the regrafting segment is chosen uniformly.



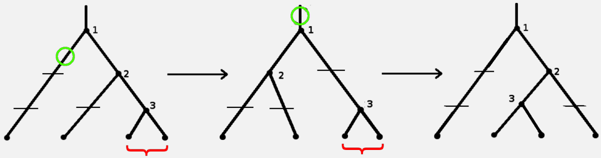
The $EMG^\flat$

Lemma (Wiehe, W. 19)

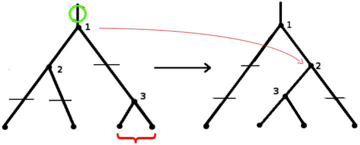
The conditions of Kelly's Lemma are satisfied for the reversed process of the EMG iff the regrafting segment is chosen uniformly.

Definition

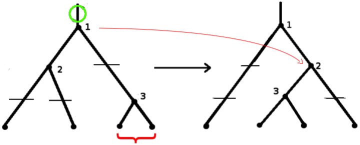
We call the reversed process *EMG backward in time*, for short $EMG^\flat$.



Analyzing jump times via $EMG^\flat$

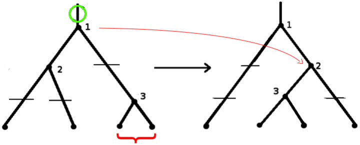


Analyzing jump times via $EMG^\flat$



Only one possibility of making the root jump in $EMG^\flat$...

Analyzing jump times via $EMG^\flat$

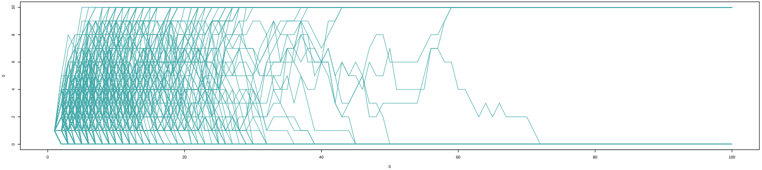


Only one possibility of making the root jump in $EMG^\flat$...

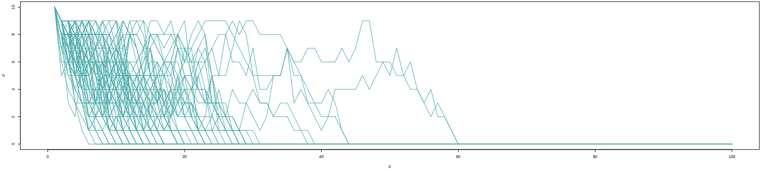
Under the $EMG^\flat$, the process of root jumps is geometric of intensity $2/n^2$.
Becomes a Poisson process of intensity 1 as $n \to \infty$ (Pfaffelhuber & Wakolbinger 2008).

Reversing a 2-allele Moran model

Assume one individual at time 0 has genotype *a*, all others have *b*.



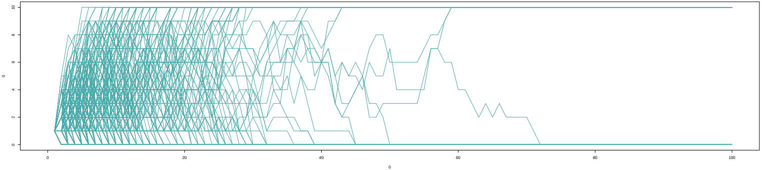
Neutral 2-allele Moran, initial frequency $f(a) = 1/n$.



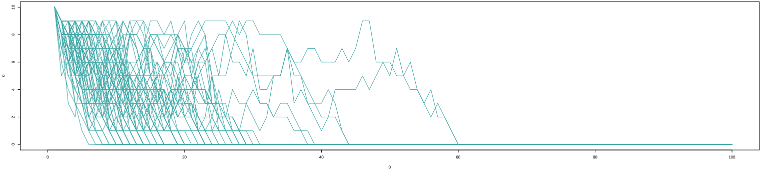
Fixation curves backward in time

Reversing a 2-allele Moran model

Assume one individual at time 0 has genotype *a*, all others have *b*.



Neutral 2-allele Moran, initial frequency $f(a) = 1/n$.



Fixation curves backward in time

The 2-allele Moran model is transient.

Let M be a Markov chain on a discrete state space $S \cup \{a, b\}$ such that for each path ω:

$$\forall i \leq \alpha(\omega) : \omega_i = a$$
$$\forall i \geq \beta(\omega) : \omega_i = b$$

where α, β are finite **a.s.**. Then

$$\forall s, t \in S : P^\flat_{st} = P_{st} \cdot \frac{\eta(t)}{\eta(s)} \qquad (2)$$

with $\eta(s)$ "average time" spent in state s

$$\eta(k) = \int_\omega \sum_{\alpha(\omega)<i<\beta(\omega)} \mathcal{X}_s(\omega_i) \mathrm{d}p(\omega)$$

(Hunt 1960)

Let M be a Markov chain on a discrete state space $S \cup \{a, b\}$ such that for each path ω:

$$\forall i \leq \alpha(\omega) : \omega_i = a$$
$$\forall i \geq \beta(\omega) : \omega_i = b$$

where α, β are finite **a.s.**. Then

$$\forall s, t \in S : P^{\flat}_{st} = P_{st} \cdot \frac{\eta(t)}{\eta(s)} \qquad (2)$$

with $\eta(s)$ "average time" spent in state s

$$\eta(k) = \int_{\omega} \sum_{\alpha(\omega)<i<\beta(\omega)} \mathcal{X}_s(\omega_i) \mathrm{d}p(\omega)$$

(Hunt 1960)
"Loop Property" still true!

Let M be a Markov chain on a discrete state space $S \cup \{a, b\}$ such that for each path ω:

$$\forall i \leq \alpha(\omega) : \omega_i = a$$
$$\forall i \geq \beta(\omega) : \omega_i = b$$

where α, β are finite **a.s.**. Then

$$\forall s, t \in S : P^\flat_{st} = P_{st} \cdot \frac{\eta(t)}{\eta(s)} \qquad (2)$$

with $\eta(s)$ "average time" spent in state s

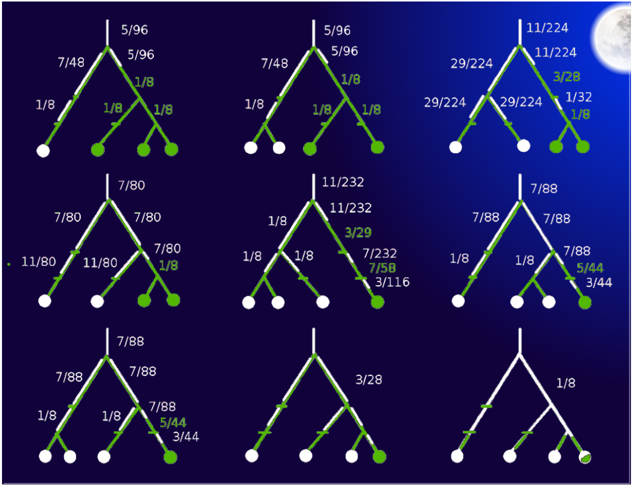
$$\eta(k) = \int_\omega \sum_{\alpha(\omega) < i < \beta(\omega)} \mathcal{X}_s(\omega_i) \mathrm{d}p(\omega)$$

(Hunt 1960)
"Loop Property" still true!

$$P_{f(a)} := \begin{array}{c} \begin{matrix} 0 & \qquad k/n \qquad & 1 \end{matrix} \\ \begin{pmatrix} 1 & 0 & & & \\ 0 & & \frac{k(n-k)}{n^2} & & \\ & & \frac{k^2+(n-k)^2}{n^2} & & \\ & & \frac{k(n-k)}{n^2} & & 0 \\ & & & 0 & 1 \end{pmatrix} \end{array} \begin{matrix} 0 \\ \\ k/n \\ \\ 1 \end{matrix}$$

$$P^\flat_{f(a)} := \begin{array}{c} \begin{matrix} * & \qquad k/n \qquad & 1 \end{matrix} \\ \begin{pmatrix} 1 & & & & \\ 0 & & \frac{k(n-k+1)}{n^2} & & \\ & & \frac{k^2+(n-k)^2}{n^2} & & \\ & & \frac{k(n-k-1)}{n^2} & & 1 \\ & & & 0 & 0 \end{pmatrix} \end{array} \begin{matrix} * \\ \\ k/n \\ \\ 1 \end{matrix}$$

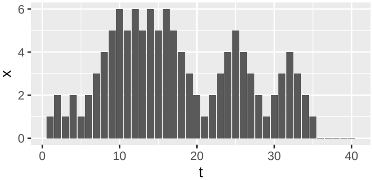
Associated tree-valued processes



Backward process on Yule Trees for $n = 4$.
Requires solving of a large nonlinear system (33 vars)

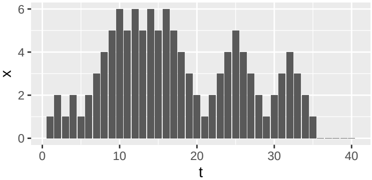
Birth-Death process with constant rate (MM1 queue)
Assume

$$
\begin{aligned}
&\Pr(X_{i+1} = X_i + 1)\\
=&\Pr(X_{i+1} = X_i - 1)\\
=&1/2
\end{aligned}
$$



Birth-Death process with constant rate (MM1 queue)
Assume

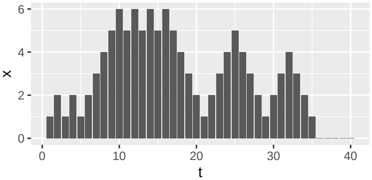
$$
\begin{aligned}
&\Pr(X_{i+1} = X_i + 1) \\
=&\Pr(X_{i+1} = X_i - 1) \\
=&1/2
\end{aligned}
$$



Choose an upper boundary n and condition the process on hitting n before 0.

Birth-Death process with constant rate (MM1 queue)
Assume

$$\begin{aligned} &\Pr(X_{i+1} = X_i + 1) \\ =&\Pr(X_{i+1} = X_i - 1) \\ =&1/2 \end{aligned}$$



Choose an upper boundary n and condition the process on hitting n before 0.

$$P_{f(a)} := \begin{array}{c} \begin{array}{ccccc} 0 & 1 & \dots & n-1 & n \end{array} \\ \begin{pmatrix} 1 & 1/2 & & & \\ & 0 & \ddots & & \\ & 1/2 & \ddots & 1/2 & \\ & & \ddots & 0 & \\ & & & 1/2 & 1 \end{pmatrix} \end{array} \begin{array}{c} 0 \\ \\ \\ n-1 \\ n \end{array}$$

$$P^{\flat}_{f(a)} := \begin{array}{c} \begin{array}{ccccc} * & & & n-2 & n-1 \end{array} \\ \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & \ddots & 2/3 & & \\ & \ddots & 0 & 3/4 & \\ & & 1/3 & 0 & 1 \\ & & & 1/4 & 0 \end{pmatrix} \end{array} \begin{array}{c} * \\ \\ \\ \\ n-2 \\ n-1 \end{array}$$

Time reversal of the genealogical process resulting from the standard population model

Time reversal of the genealogical process resulting from the standard population model

- Simplified process

Time reversal of the genealogical process resulting from the standard population model

- Simplified process
- Access to otherwise hidden statistical features

Time reversal of the genealogical process resulting from the standard population model

- Simplified process
- Access to otherwise hidden statistical features

Time reversal of Birth-Death processes

Time reversal of the genealogical process resulting from the standard population model

- Simplified process
- Access to otherwise hidden statistical features

Time reversal of Birth-Death processes

- Extendable to general birth and death probabilities and to continuous time

Time reversal of the genealogical process resulting from the standard population model

- Simplified process
- Access to otherwise hidden statistical features

Time reversal of Birth-Death processes

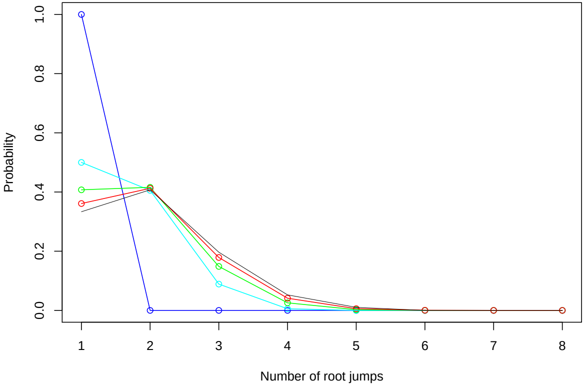
- Extendable to general birth and death probabilities and to continuous time
- Potential for application in phylogenetic/biostatistic simulation

Merci

Cher - *If I could turn back time*
written by Diane Warren
produced by Diane Warren and Guy Roche
©1989 Geffen Records

Lemma

In the time span between first occurrence and fixation of a mutant, the expected number of root jumps is $2 - \frac{2}{n}$.



Root jump distribution for $n = 2$ (blue), 5 (turquoise), 10 (green), 25 (red) and the