Ancestral processes with selection

Ellen Baake

Bielefeld University

Universität Bielefeld

SFB 1283

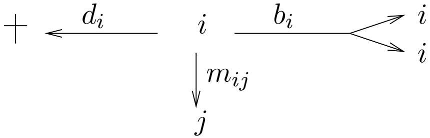
Agenda

1. mutation-selection differential equation
   & multitype branching process
   - forward and backward
   - ancestral distribution, variational principle
     (with H.-O. Georgii, 2007)

2. Moran model with selection and mutation
   - forward and backward
   - ancestral distribution, lookdown ancestral selection graph
     (with A. Wakolbinger, S. Kluth, U. Lenz, 2015, 2016, 2018)

MuSe Model

individual = (geno)type $i \in S$ (finite)



(Malthusian) fitness: $r_i := b_i - d_i$

LLN

$y_i(t)$ abundance of type i at time t $\quad (i \in S)$

$$\dot{y}_i(t) = r_i y_i(t) + \sum_{j:j \neq i} \big(y_j(t) m_{ji} - y_i(t) m_{ij}\big)$$

or

$$\dot{y}(t) = y(t)\big(\overbrace{\underset{\substack{\uparrow \\ m_{ij}}}{\mathcal{M}} + \underset{\substack{\uparrow \\ r_i}}{\mathcal{R}}}^{\mathcal{A}}\big)$$

with solution

$$y(t) = y(0)\,\mathrm{e}^{t\mathcal{A}}$$

MuSe differential equation

relative frequencies:

$$\dot{p}_i(t) = (r_i - \bar{r}(t))p_i(t) + \sum_{j:j \neq i} (p_j(t)m_{ji} - p_i(t)m_{ij})$$

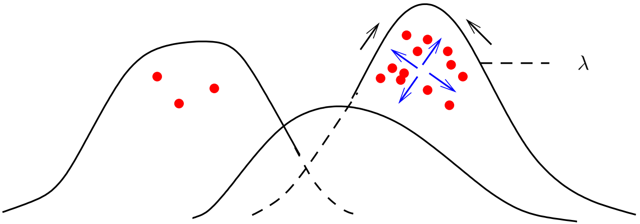
$$p_i(t) := \frac{y_i(t)}{\sum_j y_j(t)}, \qquad \bar{r}(t) := \sum_j r_j p_j(t)$$

MuSe Equilibrium

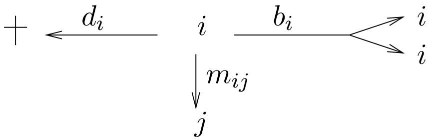
$\mathcal{M}$ irreducible $\rightsquigarrow$ Perron-Frobenius: $p\mathcal{A} = \lambda p \quad (\langle p, 1 \rangle = 1)$

$$p(t) \xrightarrow{t \to \infty} p \quad \text{(stationary type distribution)}$$

$$\lambda = \sum_i r_i p_i = \langle p, r \rangle = \bar{r} \qquad \text{(equilibrium mean fitness)}$$



Multitype Branching

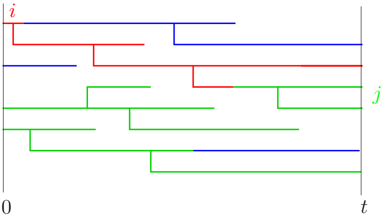


i-individual:

waiting time $\sim \mathcal{E}(a_i), \quad a_i = b_i + d_i + \sum_{j:\, j \neq i} m_{ij}$

then: birth, death, mutation to j with probability $\dfrac{b_i}{a_i}, \dfrac{d_i}{a_i}, \dfrac{m_{ij}}{a_i}$

Multitype branching



$(Z(t))$, $Z(t)$ counting measure on S
$Z_j(t)$ # ind. of type j at time t
first-moment generator: $\mathcal{A} = \mathcal{M} + \mathcal{R}$

$\mathbb{E}^i(Z_j(t)) = (\mathrm{e}^{t\mathcal{A}})_{ij}$
assumption: $\lambda > 0 \rightsquigarrow$ branching supercritical
(with asymptotic growth rate λ)

Connections "branching" $\leftrightarrow$ "MuSe":

1. $$\frac{Z(t)}{|Z(t)|} \overset{t\to\infty}{\longrightarrow} p \overset{\infty\leftarrow t}{\longleftarrow} p(t)$$
   $\uparrow$
   a.s., on {non-extinction} (Kesten–Stigum '66)

   p left PF-EV of $\mathcal{A}$ (stationary distribution of types)

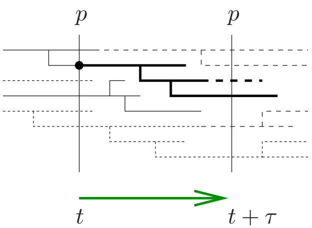
2. $$\mathbb{E}^i(|Z(t)|\,\mathrm{e}^{-\lambda t}) \overset{t\to\infty}{\longrightarrow} h_i \overset{\infty\leftarrow t}{\longleftarrow} \frac{\sum_j(\,\mathrm{e}^{t\mathcal{A}})_{ij}}{|p(0)\,\mathrm{e}^{t\mathcal{A}}|}$$

   $h = (h_i)_{i\in S}$ right PF-EV of $\mathcal{A}$ (asymptotic offspring expectation)

$$\langle p, 1\rangle = 1 = \langle p, h\rangle$$

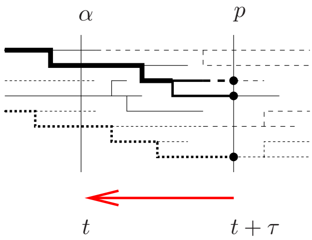
Forward and backward

forward



backward

(NO coalescent)



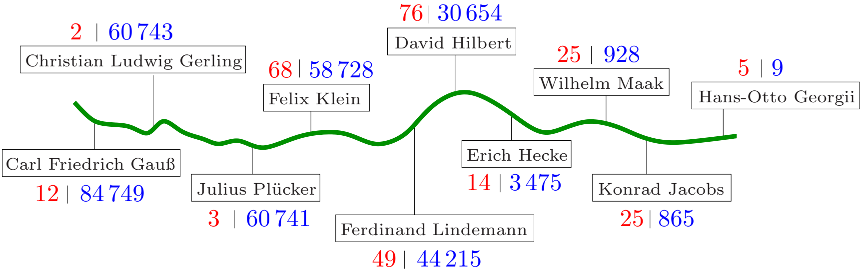
$\tau \to \infty \rightsquigarrow$

$\alpha_i := p_i h_i$

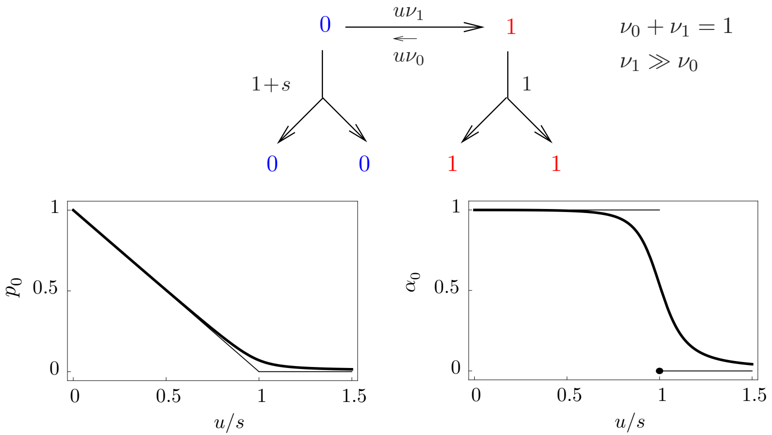
ancestral type distribution

Jagers, Nerman 1992 . . .

Mathematics genealogy project

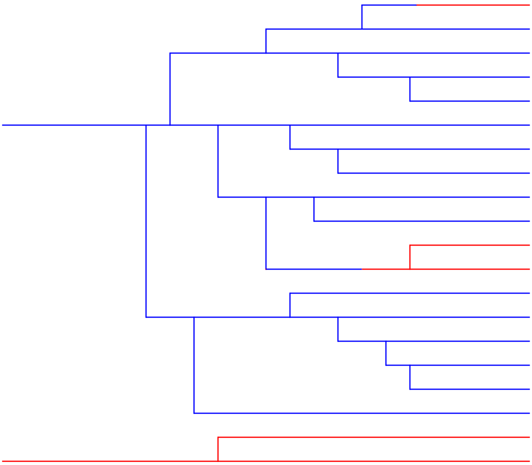


Forward and backward

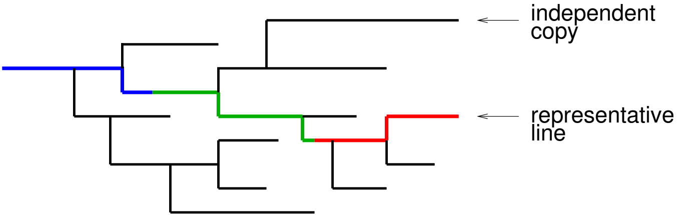


$s = 0.001$, $\nu_0 = 0.005$ and $\nu_0 \to 0$

Forward and backward

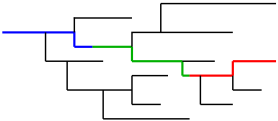


Large deviations



mutation process (M_t) on representative line (M_t type at time t)
(generator $\mathcal{M}$, stationary distribution π)

Large deviations



empirical measure L_t on S: $L_t(j) = \frac{1}{t}\int_0^t \mathbf{1}_{\{M_\tau = j\}}\,\mathrm{d}\tau$ (random!)

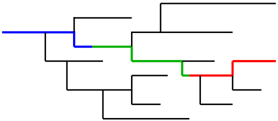
LDP: $\mathbb{P}(L_t \sim \nu) \approx \mathrm{e}^{-t\,I_{\mathcal{M}}(\nu)}$ (large t)

$$\left(\lim_{t\to\infty} \frac{1}{t}\log \mathbb{P}(L_t \in A) = -\inf_{\nu \in A} I_{\mathcal{M}}(\nu)\right)$$

rate function: $I_{\mathcal{M}}(\nu) = \sup_{v>0}\left(-\langle \nu, \frac{\mathcal{M}v}{v}\rangle\right)$

(M_t) reversible $\rightsquigarrow$ $I_{\mathcal{M}}(\nu) = -\left\langle \sqrt{\frac{\nu}{\pi}}, \mathcal{M}\sqrt{\frac{\nu}{\pi}}\right\rangle_\pi$

Variational principle



line with $L_t = \nu$

experiences

- mutation: changes ν

  $\mathbb{P}(L_t \sim \nu)$ decays with $I_{\mathcal{M}}(\nu)$ $\quad$ (> 0 for $\nu \neq \pi$ (stat. distr.), $= 0$ for $\nu = \pi$)

  $L_t \underset{t\to\infty}{\longrightarrow} \pi$ a.s.

- reproduction: duplicates ν at rate r_{M_t} at time t, mean rate $\langle \nu, r\rangle$

Theorem (EB & Georgii 2007)

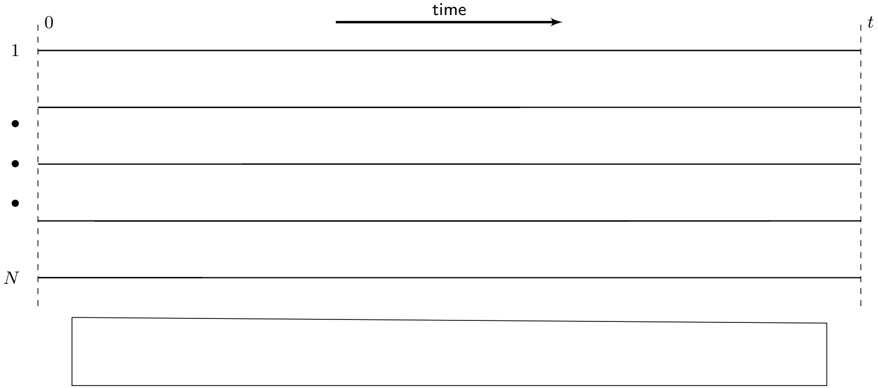
$$\langle p, r\rangle = \lambda = \sup_{\nu \in \mathcal{P}(S)} [\langle \nu, r\rangle - I_{\mathcal{M}}(\nu)] = \langle \alpha, r\rangle - I_{\mathcal{M}}(\alpha)$$

$\langle p, r\rangle$ ↑ "present"; $\langle \nu, r\rangle$ ↑ energy; $I_{\mathcal{M}}(\nu)$ ↑ entropy; $\langle \alpha, r\rangle - I_{\mathcal{M}}(\alpha)$ ↑ "past"

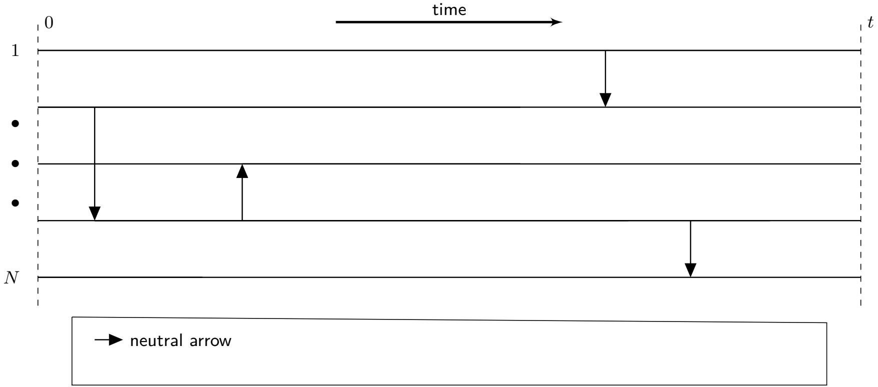
2-type Moran model with mutation and selection

- population of fixed size N
- types: 0 ('fit') and 1 ('unfit')
- individuals of type 1 reproduce at rate 1
- individuals of type 0 reproduce at rate $1 + s^N, \quad s^N \geqslant 0$
- single offspring inherits parent's type and replaces uniformly chosen individual
- mutation at rate $u^N > 0$
- resulting type: 0 with probability ν_0; 1 with probability ν_1 $(\nu_0 + \nu_1 = 1)$

Interacting particle system: untyped

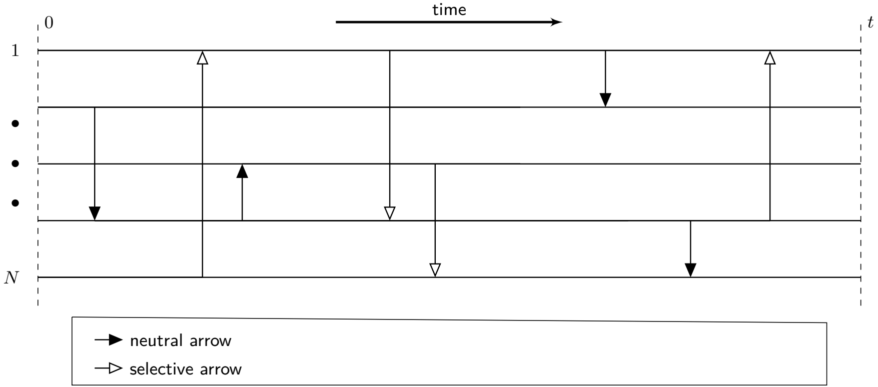


Interacting particle system: untyped



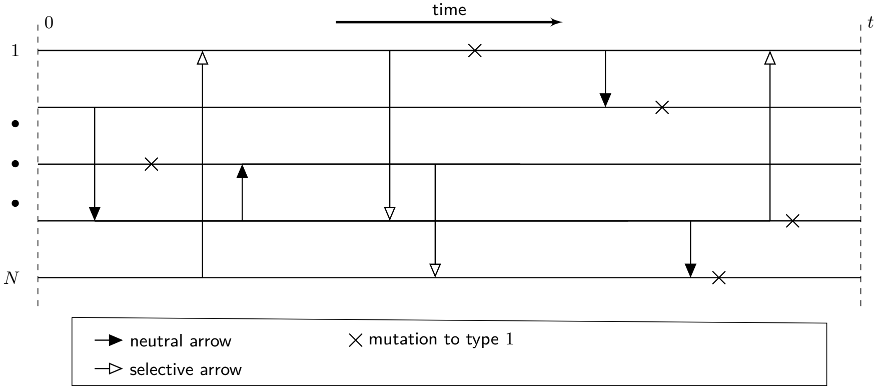
neutral arrows: rate 1,

Interacting particle system: untyped



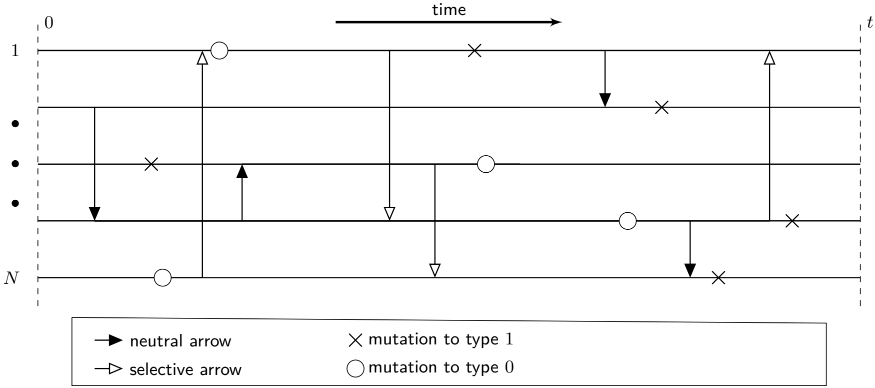
neutral arrows: rate 1, selective arrows: rate s^N,

Interacting particle system: untyped



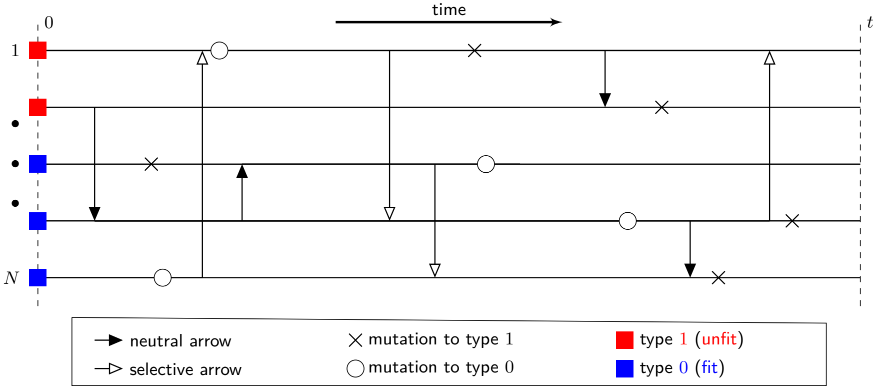
neutral arrows: rate 1, selective arrows: rate s^N,
mutation to type 1: rate $u^N \nu_1$,

Interacting particle system: untyped



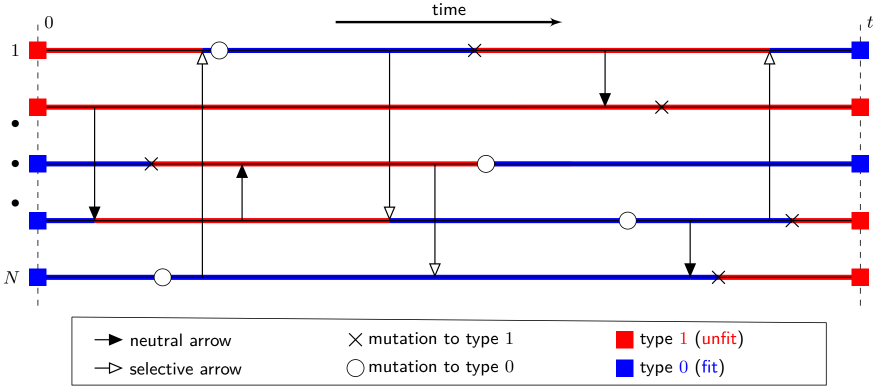
neutral arrows:	rate 1,	selective arrows:	rate s^N,
mutation to type 1:	rate $u^N \nu_1$,	mutation to type 0:	rate $u^N \nu_0$

Interacting particle system: typed



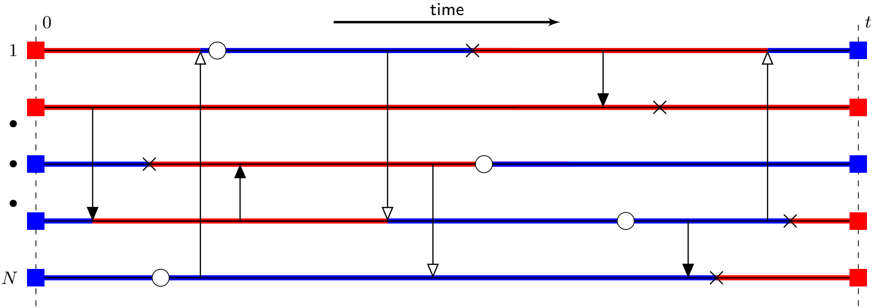
neutral arrows:	rate 1,	selective arrows:	rate s^N,
mutation to type 1:	rate $u^N \nu_1$,	mutation to type 0:	rate $u^N \nu_0$

Interacting particle system: typed



neutral arrows:	rate 1,	selective arrows:	rate s^N,
mutation to type 1:	rate $u^N \nu_1$,	mutation to type 0:	rate $u^N \nu_0$

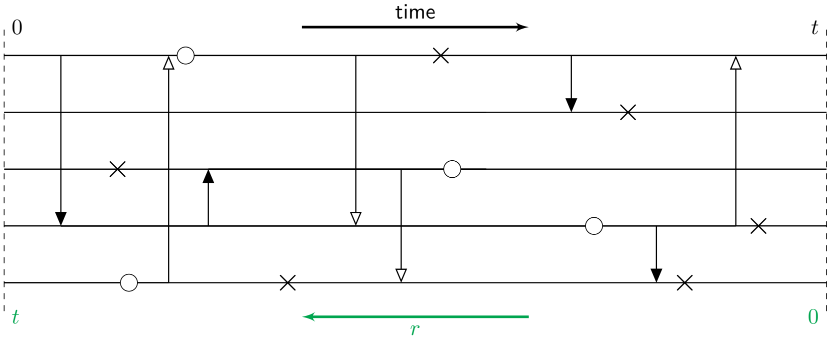
Interacting particle system: typed



$Y_t^N :=$ proportion of individuals of type 1 at time t in MoMo of size N

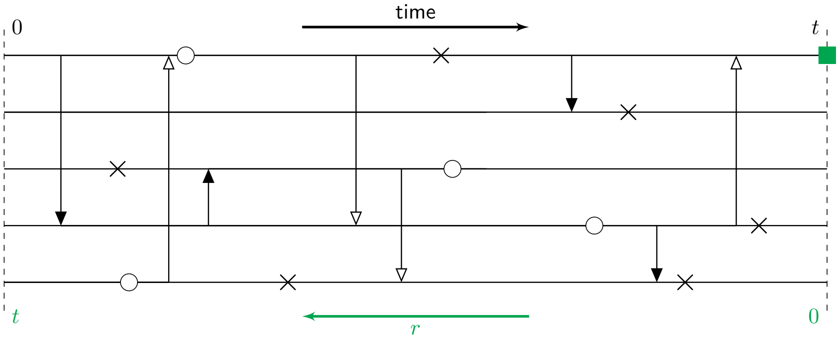
Ancestral selection graph (ASG)

Krone and Neuhauser 1997



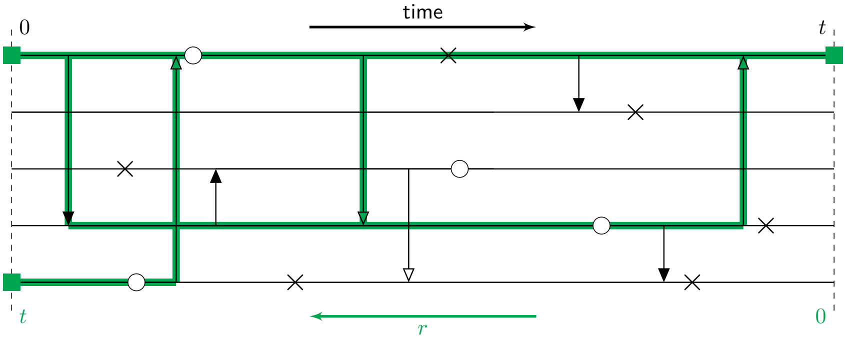
Ancestral selection graph (ASG)

Krone and Neuhauser 1997



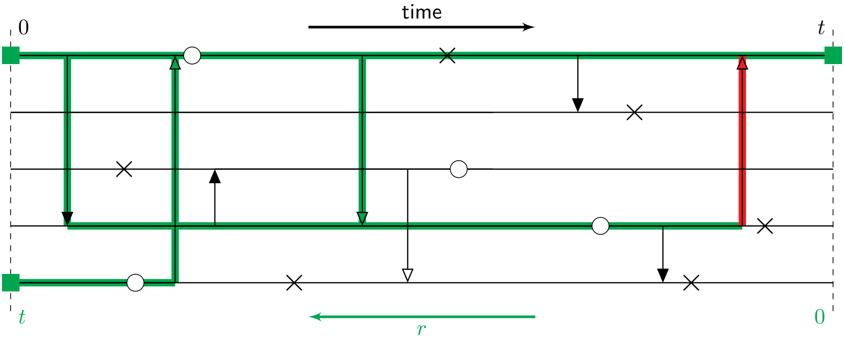
Ancestral selection graph (ASG)

Krone and Neuhauser 1997



Ancestral selection graph (ASG)

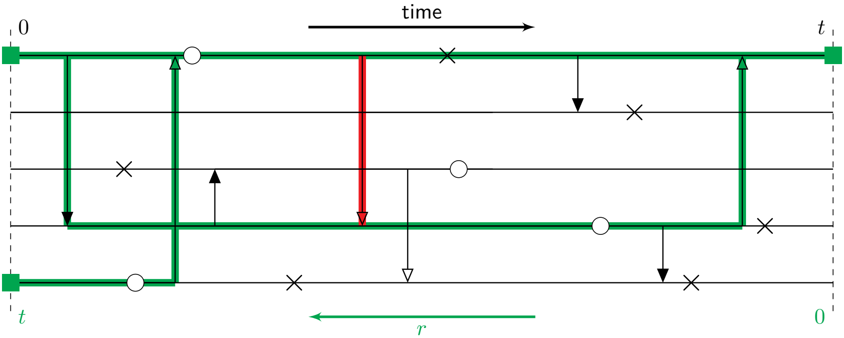
Krone and Neuhauser 1997



- *branching:* rate $s^N(N-n)/N$ per line (n current number of lines)

Ancestral selection graph (ASG)

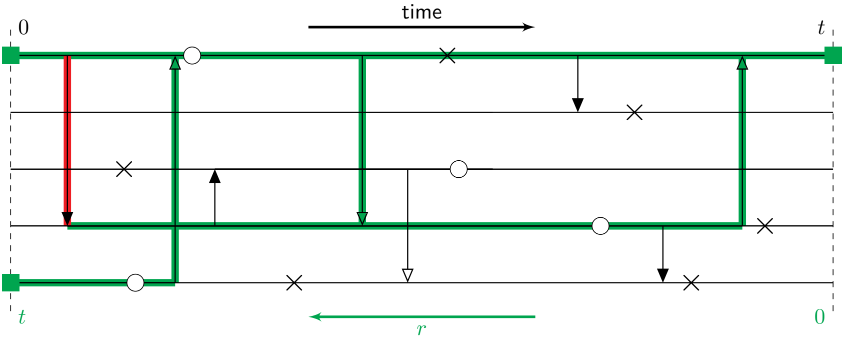
Krone and Neuhauser 1997



- *branching:* rate $s^N(N-n)/N$ per line (n current number of lines)

Ancestral selection graph (ASG)

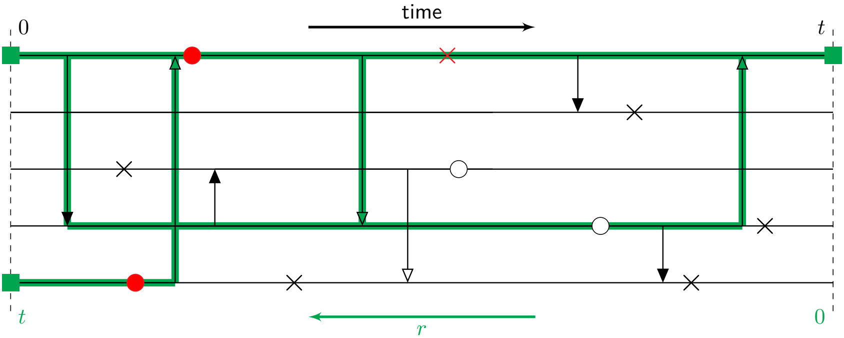
Krone and Neuhauser 1997



- *branching:* rate $s^N(N-n)/N$ per line (n current number of lines)
- *coalescence:* rate $1/N$ per pair

Ancestral selection graph (ASG)

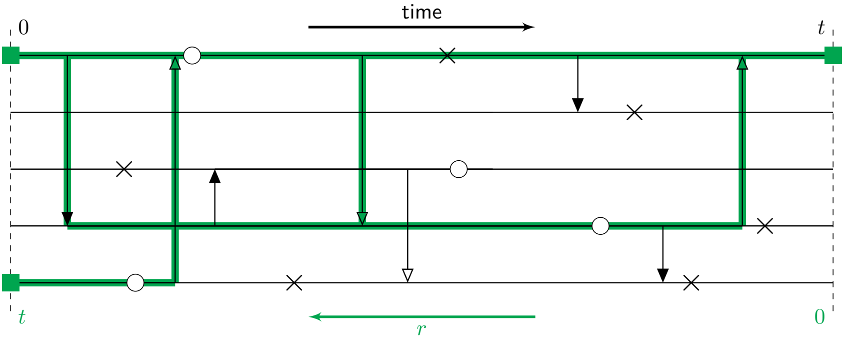
Krone and Neuhauser 1997



- *branching:* rate $s^N(N-n)/N$ per line (n current number of lines)
- *coalescence:* rate $1/N$ per pair
- *mutation:* rates $u^N\nu_0$, $u^N\nu_1$ per line

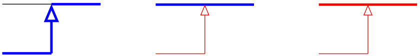
Ancestral selection graph (ASG)

Krone and Neuhauser 1997



draw types at time 0 from $(1 - Y_0^N, Y_0^N)$ and propagate them forward

respect pecking order



$\rightsquigarrow$ type at present together with true ancestral line

Diffusion limit

$N \to \infty \qquad$ s.t. $\quad Ns^N \to \sigma, \qquad Nu^N \to \theta, \quad Y_0^N \to y_0$

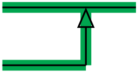
$\rightsquigarrow (Y_{tN}^N) \xrightarrow{d}$ Wright–Fisher diffusion (Y_t)

$$
\begin{aligned}
\mathrm{d}Y_t &= \sqrt{Y_t(1-Y_t)}\,\mathrm{d}W_t - \sigma Y_t(1-Y_t)\,\mathrm{d}t + (1-Y_t)\theta\nu_1\,\mathrm{d}t - Y_t\theta\nu_0\,\mathrm{d}t,\\
Y_0 &= y_0, \quad (W_t) \text{ standard Brownian motion}
\end{aligned}
$$

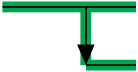
$\theta > 0,\ 0 < \nu_0 < 1, t \to \infty \rightsquigarrow\ Y_t \xrightarrow{d} \widetilde{Y}$

ASG in diffusion limit

- branching at rate σ per line

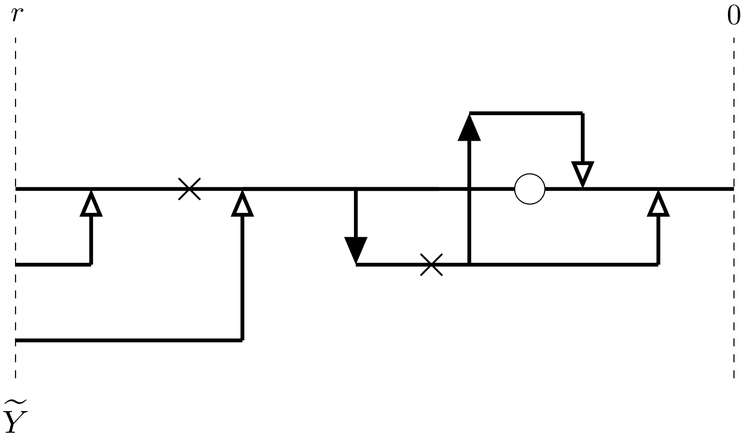


- coalescence events at rate 1 per pair

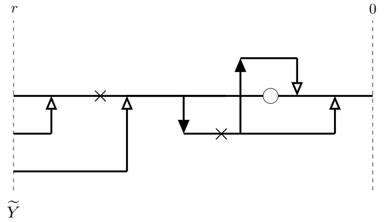


- mutation superimposed on lines at rate $\theta\nu_0$ and $\theta\nu_1$

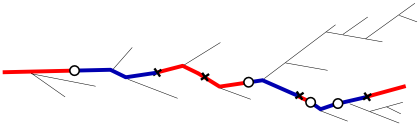
ASG in diffusion limit



Ancestral type



$I_r \in \{0, 1\}$
$:=$ ancestral type at time r



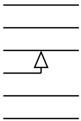
long-term success
bias towards type 0

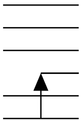
$$r \to \infty \rightsquigarrow I_r \xrightarrow{d} \tilde{I}$$

$$\mathbb{P}(\tilde{I} = 0),\ \mathbb{P}(\tilde{I} = 1)?$$

Ordering the ASG (w/o mutation)

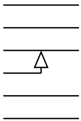
arrange lines according to pecking order (exchangeability!)

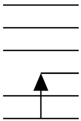


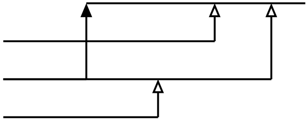


Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)





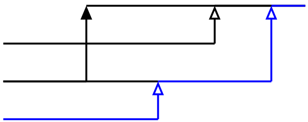


ancestral line is

- lowest type-0 line if there is one
- immune line otherwise

Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)

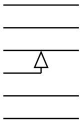


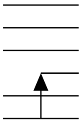
ancestral line is

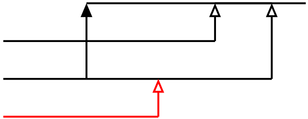
- lowest type-0 line if there is one
- immune line otherwise

Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)





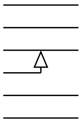


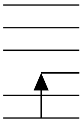
ancestral line is

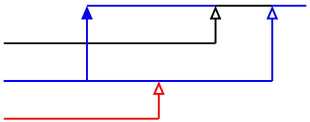
- lowest type-0 line if there is one
- immune line otherwise

Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)





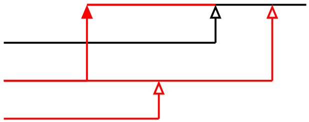


ancestral line is

- lowest type-0 line if there is one
- immune line otherwise

Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)

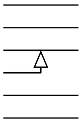


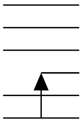
ancestral line is

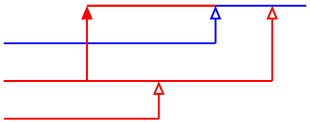
- lowest type-0 line if there is one
- immune line otherwise

Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)





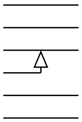


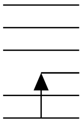
ancestral line is

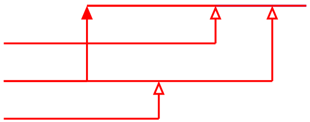
- lowest type-0 line if there is one
- immune line otherwise

Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)





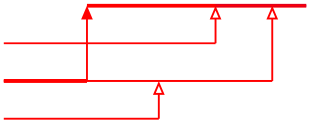


ancestral line is

- lowest type-0 line if there is one
- immune line otherwise

Ordering the ASG (w/o mutation)

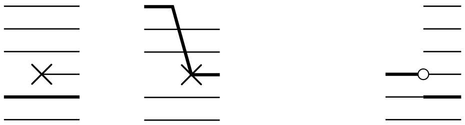
arrange lines according to pecking order (exchangeability!)

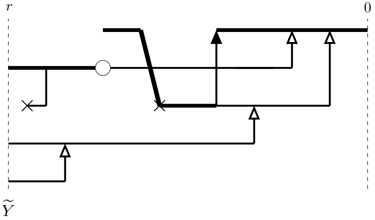


ancestral line is

- lowest type-0 line if there is one
- immune line otherwise

Pruning upon mutation





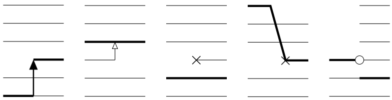
ancestral line is

- lowest type-0 line if there is one
- immune line otherwise

The line-counting process of the pruned ASG

L_r = number of lines at time r

(L_r) Markov chain in continuous time with rates from



stationary distribution $(r \to \infty)$:

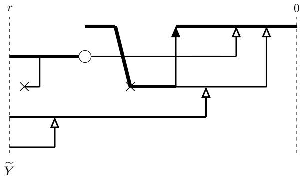
$$a_n := \mathbb{P}(\widetilde{L} > n), \quad n \geqslant 0$$

first-step analysis $\rightsquigarrow$ (Fearnhead's) recursion:

$$(n+1+\theta+\sigma)a_n = (n+1+\theta\nu_1)a_{n+1} + \sigma a_{n-1}, \quad n > 0,$$

$$a_0 = 1, \quad \lim_{n\to\infty} a_n = 0.$$

Type of ancestral line



Theorem (Lenz, Kluth, EB, Wakolbinger 2015)

The type distribution of the ancestral line in the distant past is given by

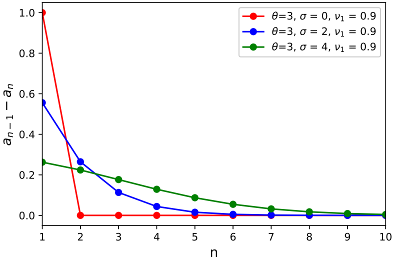
$$\mathbb{P}(\tilde{I} = 1 \mid \widetilde{Y}) = \sum_{n>0} (a_{n-1} - a_n) \widetilde{Y}^n$$

$$\mathbb{P}(\tilde{I} = 1) = \sum_{n>0} (a_{n-1} - a_n) b_n, \quad \mathbb{P}(\tilde{I} = 0) = \sum_{n \geqslant 0} a_n (b_n - b_{n+1}),$$

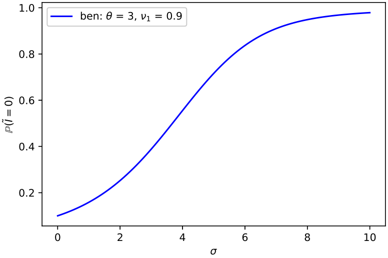
where $\quad b_n := \mathbb{E}(\widetilde{Y}^n).$

The bias towards type 0

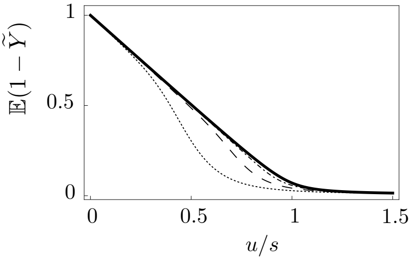
distribution of $\widetilde{L}$

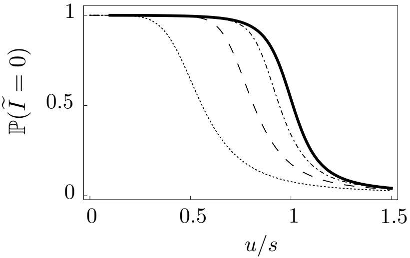


probability of fit ancestor



Forward and backward





$$s = 0.001,\ \nu_0 = 0.005$$

$\sigma = Ns$, $\vartheta = Nu$ with $N = 10^4$, $N = 3 \cdot 10^4$, $N = 10^5$

ToDo

Large devations and variational principle for (multitype) MoMo?