A probabilistic view on the deterministic mutation-selection equation

Ellen Baake

joint work with Fernando Cordero and Sebastian Hummel

Bielefeld University

PROBABILISTIC STRUCTURES IN EVOLUTION

DFG SPP 1590

Universität Bielefeld

A probabilistic view on the deterministic mutation-selection equation

A probabilistic view on the deterministic mutation-selection equation

1. 2-type Moran model and its deterministic limit
   - 2-type Moran model
   - Deterministic limit: haploid mutation-selection equation
   - Properties of deterministic solution

2. Ancestries in the Moran model and in the deterministic limit
   - Ancestral selection graph
   - Killed ancestral selection graph
   - Pruned lookdown ancestral selection graph

Moran model with 2 types

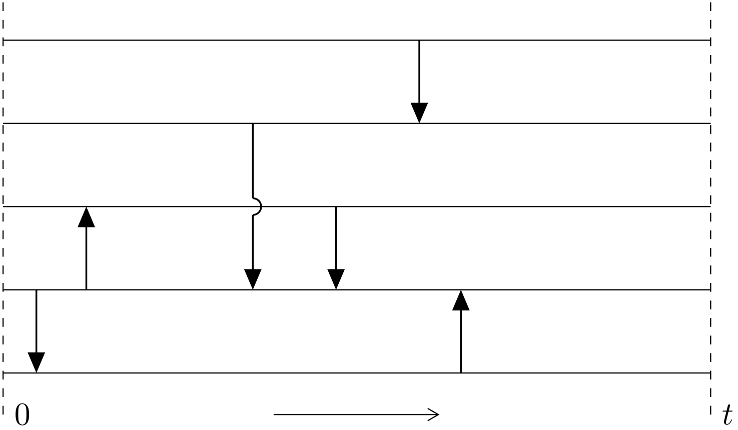
- Haploid population of fixed size N
- Types: 0 ('fit') and 1 ('unfit')
- Individuals of type 1 reproduce at rate 1
- Individuals of type 0 reproduce at rate $1+s, \quad s \geq 0$
- Single offspring inherits parent's type and replaces uniformly chosen individual
- Mutation at rate $u > 0$
- Resulting type: 0 with probability ν_0; 1 with probability ν_1 $(\nu_0 + \nu_1 = 1)$

Graphical representation (interacting particle system)

○ Mutation to type 0 ⟶▷ Selective arrow

× Mutation to type 1 ⟶▶ Neutral arrow

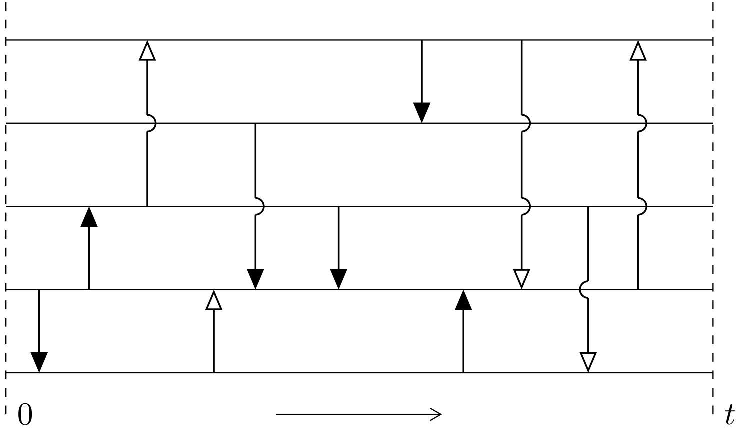
Graphical representation (interacting particle system)



○ Mutation to type 0 —▷ Selective arrow

× Mutation to type 1 —▶ Neutral arrow

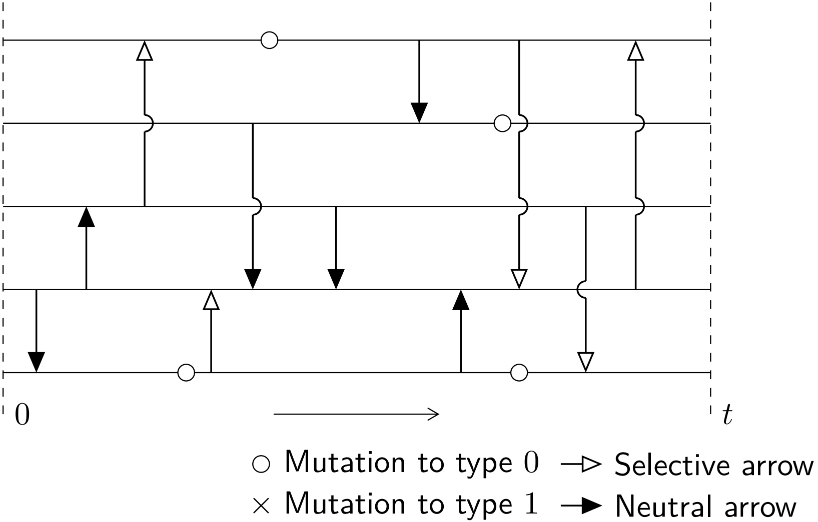
Graphical representation (interacting particle system)



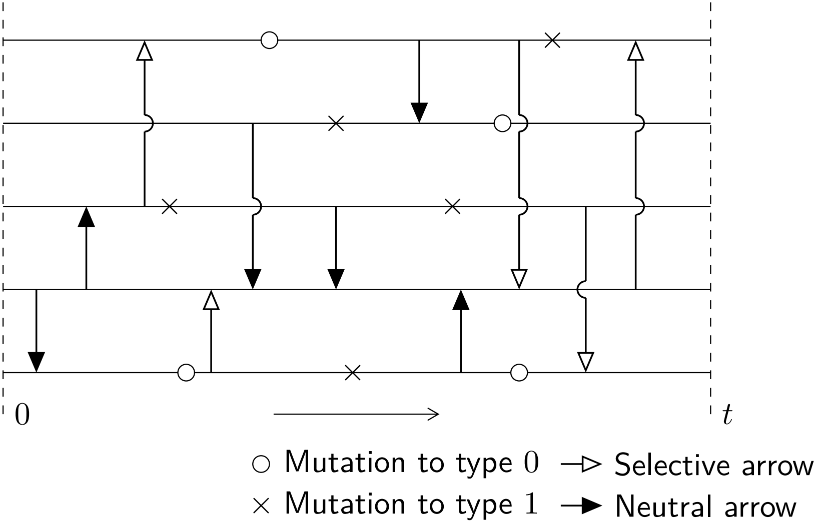
○ Mutation to type 0 —▷ Selective arrow

× Mutation to type 1 —▶ Neutral arrow

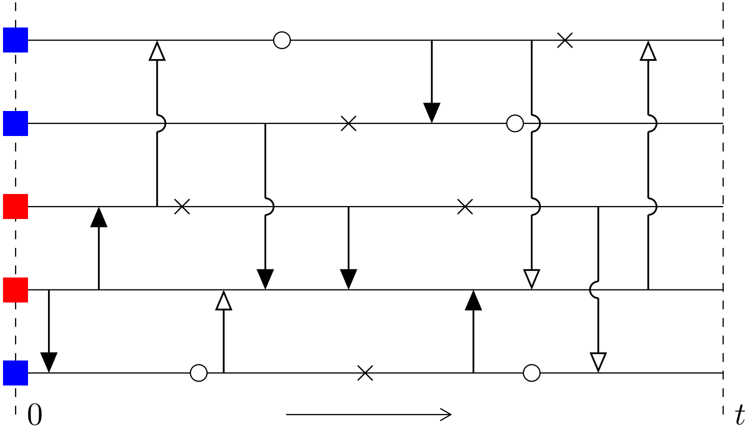
Graphical representation (interacting particle system)



Graphical representation (interacting particle system)

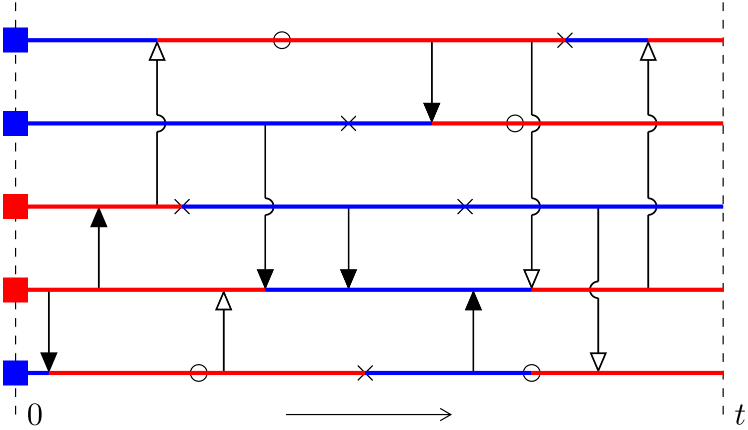


Graphical representation (interacting particle system)



Type 0 ('fit') ○ Mutation to type 0 ▷ Selective arrow
Type 1 ('unfit') × Mutation to type 1 ▶ Neutral arrow

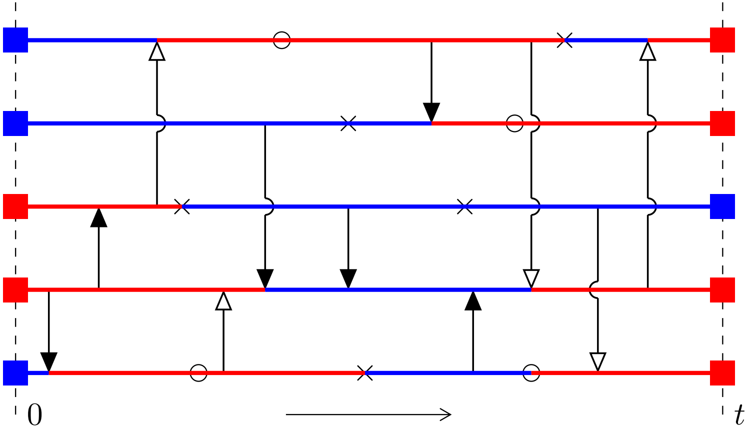
Graphical representation (interacting particle system)



Type 0 ('fit') ○ Mutation to type 0 ▷ Selective arrow

Type 1 ('unfit') × Mutation to type 1 ▶ Neutral arrow

Graphical representation (interacting particle system)



Type 0 ('fit') ○ Mutation to type 0 —▷ Selective arrow

Type 1 ('unfit') × Mutation to type 1 —▶ Neutral arrow

Deterministic limit of the Moran model

- $Y_t^{(N)}$ proportion of type 1 at time t in MoMo
  (Markov process on $\frac{1}{N}\{0,1,\dots,N\}$)
- $y(t;y_0)$ proportion of type 1 at time t in inf. pop. under MuSe
  solution of haploid mutation-selection equation

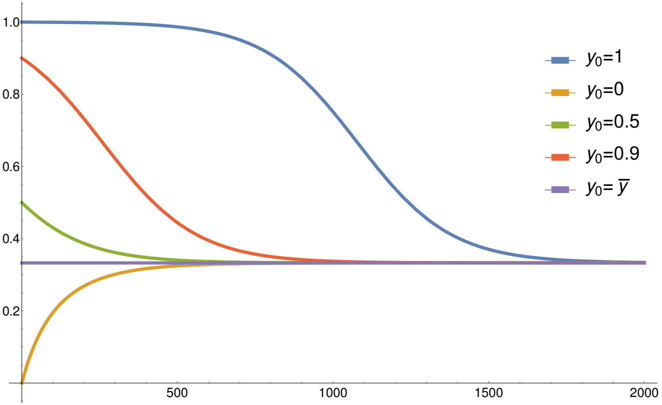
$$\dot{y} = -sy(1-y) - u\nu_0 y + u\nu_1(1-y)$$
$$y(0) = y_0 \qquad \text{for } y_0 \in [0,1]$$

- If $\lim_{N\to\infty} Y_0^{(N)} = y_0$, then $\forall \varepsilon, T > 0$,

$$\lim_{N\to\infty} P\Big(\sup_{t\le T} |Y_t^{(N)} - y(t;y_0)| > \varepsilon\Big) = 0$$

- Convergence carries over to the stationary state $(t \to \infty)$
- No rescaling of time or parameters

Properties of the deterministic solution $y(t;y_0)$



$y(t;y_0)$ as a function of t

Parameters: $s = \frac{1}{100}, \; u = \frac{1}{300}, \; \nu_0 = \frac{1}{1000}, \; \nu_1 = 1 - \frac{1}{1000}.$

Properties of the deterministic solution $y(t; y_0)$

- $s = 0$: unique equilibrium $\bar{y} = \nu_1$ (stable)
- $s > 0$: two equilibria, $\bar{y}$ (stable) and $y^\star$ (unstable)

$$\bar{y} = \frac{1}{2}\left(1 + \frac{u}{s} - \sqrt{\left(1 - \frac{u}{s}\right)^2 + 4\frac{u}{s}\nu_0}\right) \qquad \in [0,1]$$

$$y^\star = \frac{1}{2}\left(1 + \frac{u}{s} + \sqrt{\left(1 - \frac{u}{s}\right)^2 + 4\frac{u}{s}\nu_0}\right) \qquad \geq 1$$

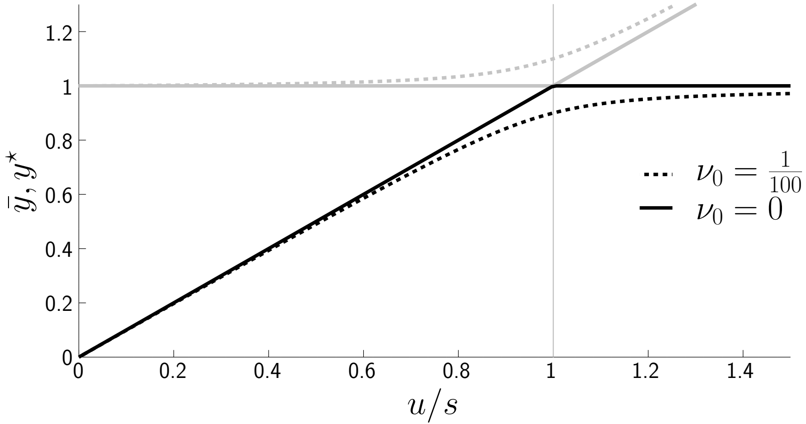
- $\nu_0 > 0$:

$$\bar{y} \in [0,1), \quad y^\star > 1$$

- $\nu_0 = 0$:

$$\bar{y} = \min\left\{\frac{u}{s}, 1\right\}, \quad y^\star = \max\left\{1, \frac{u}{s}\right\}$$

The equilibria for $s > 0$



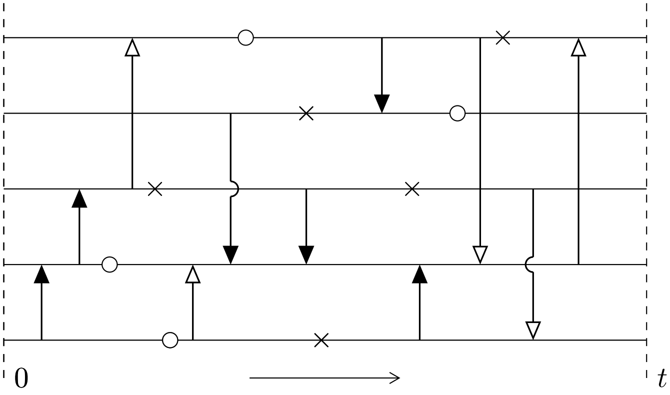
Black: $\bar{y}$ (stable); grey: $y^\star$ (unstable)

Transcritical bifurcation ('error threshold', 'phase transition')

A probabilistic view on the deterministic mutation-selection equation

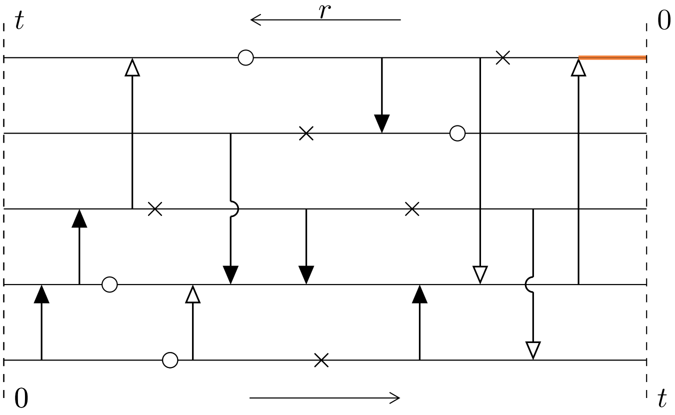
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



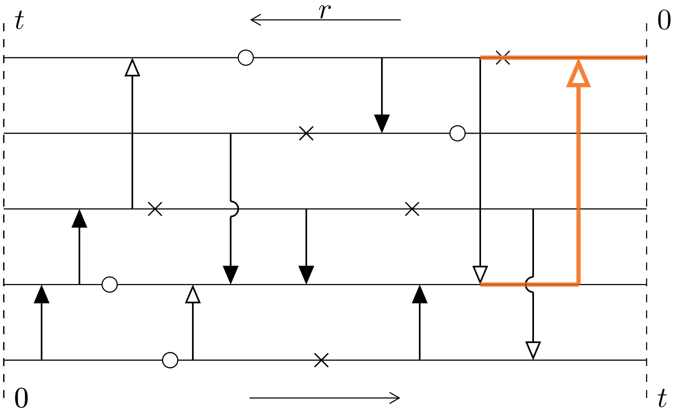
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



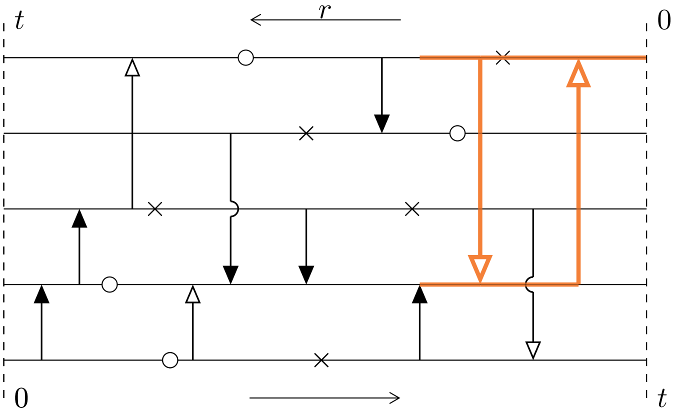
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



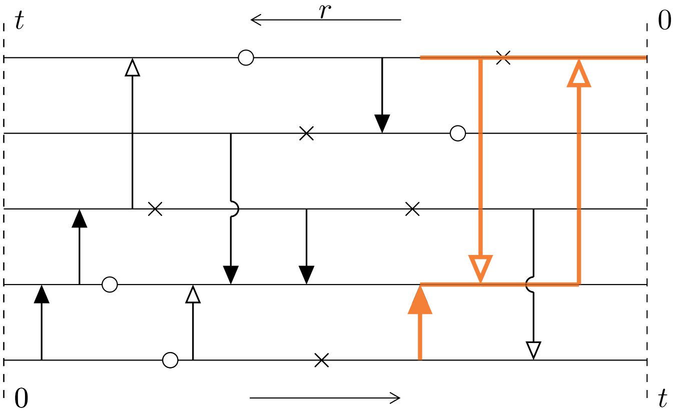
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



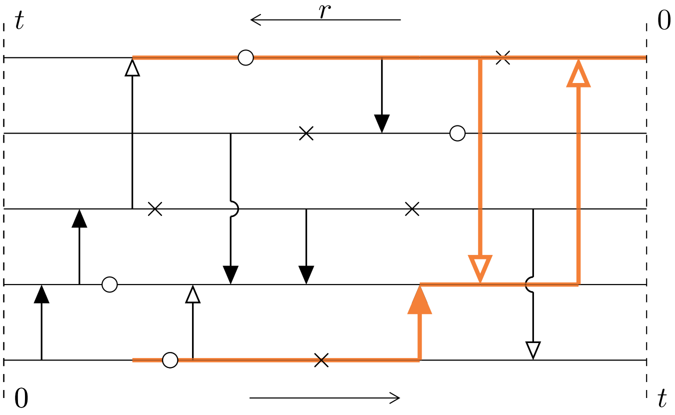
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



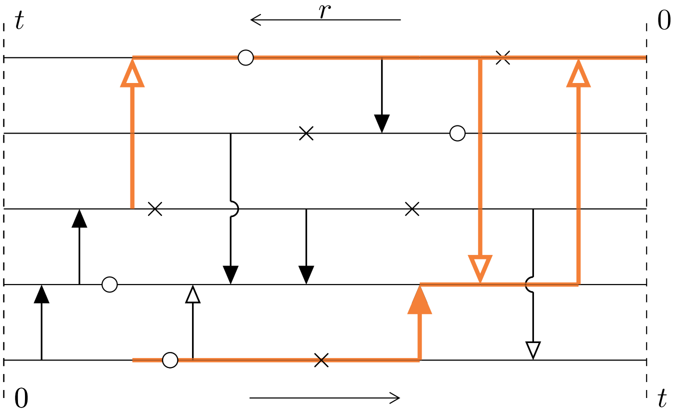
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



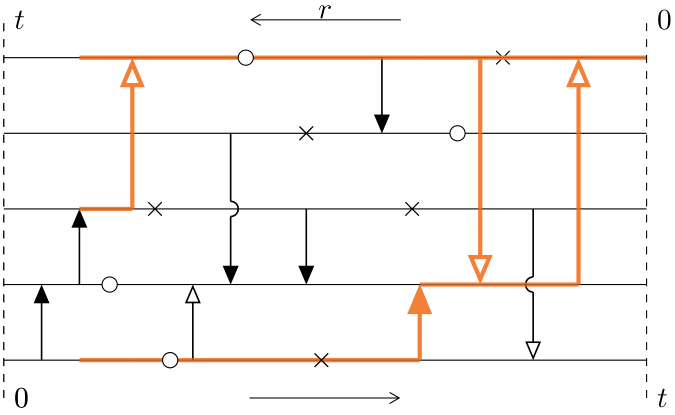
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



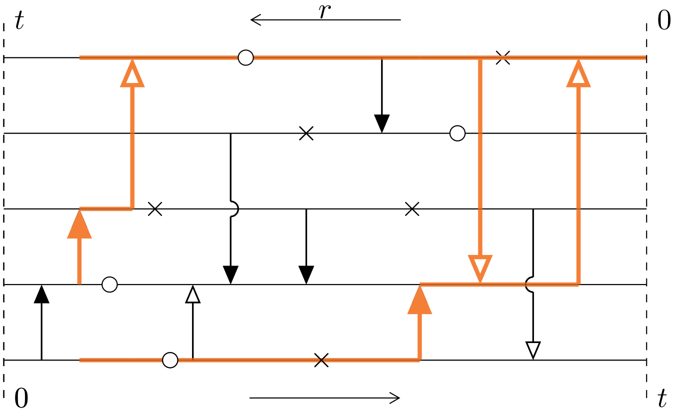
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



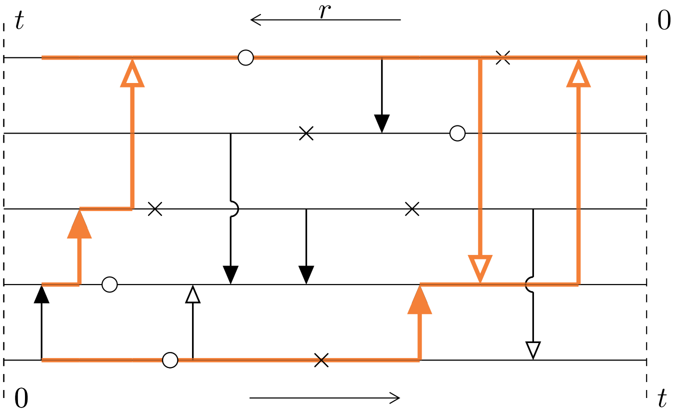
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



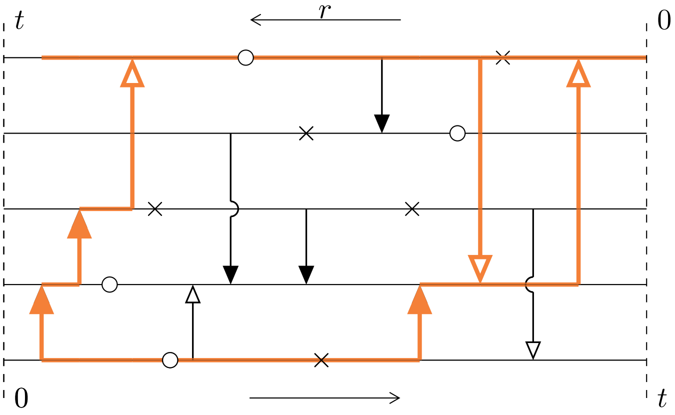
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



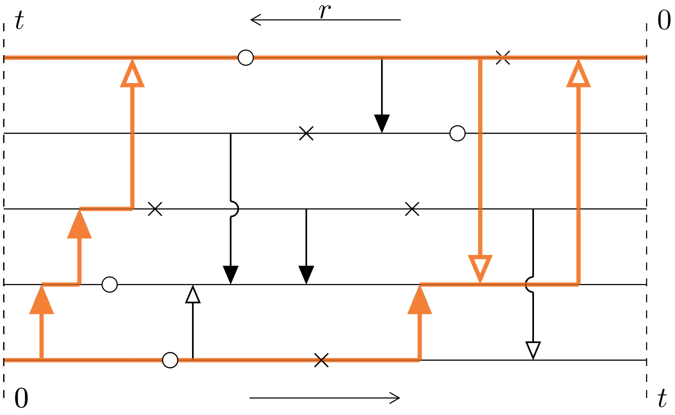
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



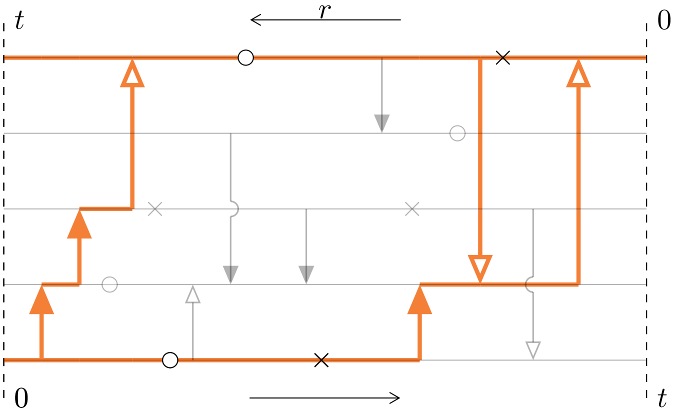
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



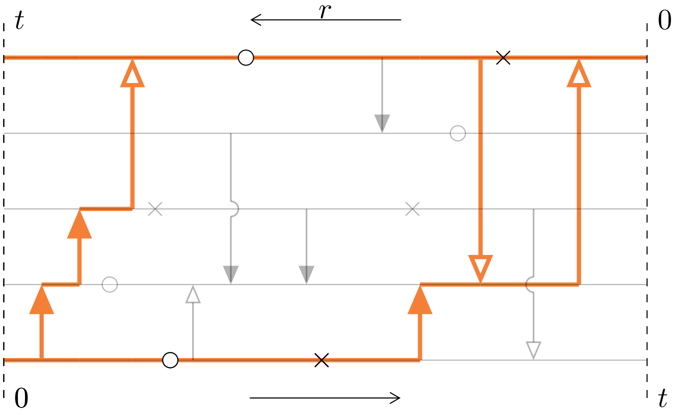
Ancestral selection graph (ASG)

(Krone and Neuhauser 1997)



Ancestral selection graph (ASG)

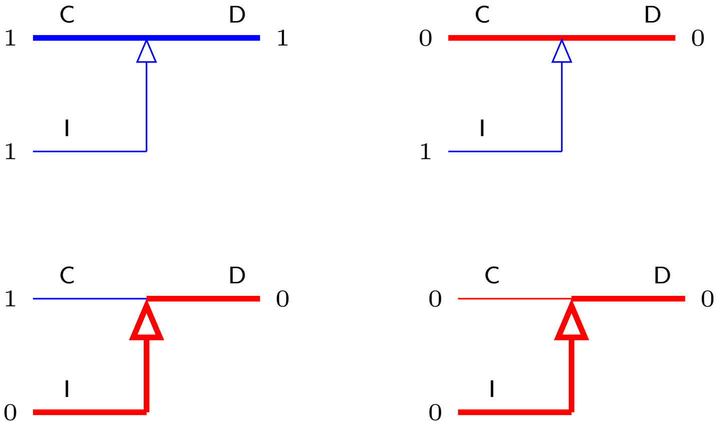
(Krone and Neuhauser 1997)



draw types at time 0 from $Y_0^{(N)}$ and propagate them forward

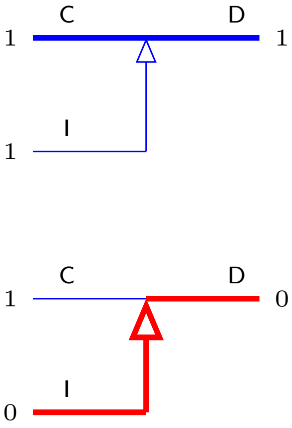
Pecking order

D=Descendant C=Continuing I=Incoming



Pecking order

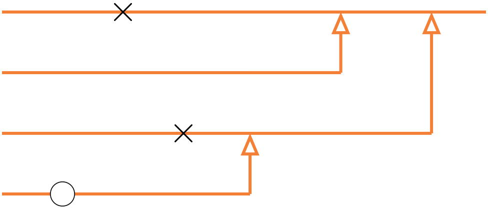
D=Descendant C=Continuing I=Incoming



Descendant type $1 \Leftrightarrow$ all potential ancestors type 1

ASG in the deterministic limit

- No coalescence events, no collisions (rate $O(\frac{1}{N})$ per pair)
- Branching at rate s per line
- Mutation to type 0 at rate $u\nu_0$ per line
- Mutation to type 1 at rate $u\nu_1$ per line



The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

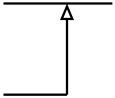
The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)

$$q_R(k, k+1) = ks$$



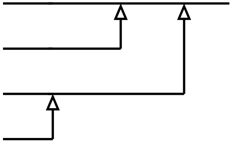
The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)

$$q_R(k, k+1) = ks$$



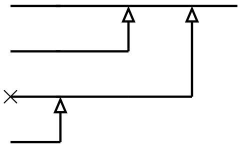
The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)

$$q_R(k, k+1) = ks$$



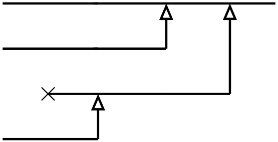
The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)
- Pruning on deleterious mutation (rate $u\nu_1$ per line)

$$q_R(k, k+1) = ks$$
$$q_R(k, k-1) = ku\nu_1$$

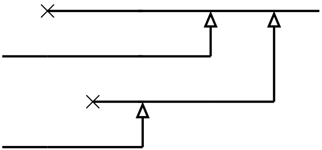


The killed ASG

Type of randomly chosen individual at time t?
$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)
- Pruning on deleterious mutation (rate $u\nu_1$ per line)

$$q_R(k, k+1) = ks$$
$$q_R(k, k-1) = ku\nu_1$$



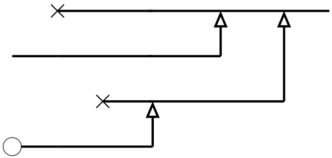
The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)
- Pruning on deleterious mutation (rate $u\nu_1$ per line)

$$q_R(k, k+1) = ks$$
$$q_R(k, k-1) = ku\nu_1$$



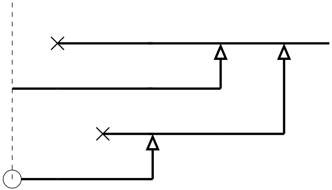
The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)
- Pruning on deleterious mutation (rate $u\nu_1$ per line)
- Killing on beneficial mutation (rate $u\nu_0$ per line)

$$q_R(k, k+1) = ks$$
$$q_R(k, k-1) = ku\nu_1$$
$$q_R(k, \Delta) = ku\nu_0$$



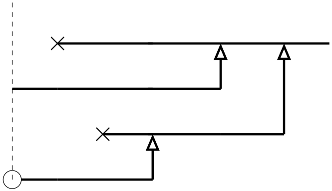
The killed ASG

Type of randomly chosen individual at time t?

$\rightsquigarrow$ R_r number of potential ancestors at time r that can influence the type at present

- Branching (rate s per line)
- Pruning on deleterious mutation (rate $u\nu_1$ per line)
- Killing on beneficial mutation (rate $u\nu_0$ per line)

$$q_R(k, k+1) = ks$$
$$q_R(k, k-1) = ku\nu_1$$
$$q_R(k, \Delta) = ku\nu_0$$



Absorbing states: 0 ($\rightsquigarrow$ type 1) and Δ ($\rightsquigarrow$ type 0)

Theorem (Duality)

For $t \geq 0$,

$$y(t; y_0)^n = \mathbb{E}\Big[y_0^{R_t} \mid R_0 = n\Big] \qquad \forall n \in \mathbb{N}_0 \cup \{\Delta\},\ y_0 \in [0,1],$$

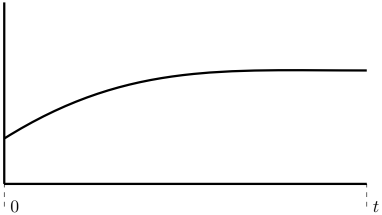
where $y^{\Delta} := 0$.

Theorem (Duality)

For $t \geq 0$,

$$y(t; y_0)^n = \mathbb{E}\Big[y_0^{R_t} \mid R_0 = n\Big] \qquad \forall n \in \mathbb{N}_0 \cup \{\Delta\},\ y_0 \in [0, 1],$$

where $y^{\Delta} := 0$.

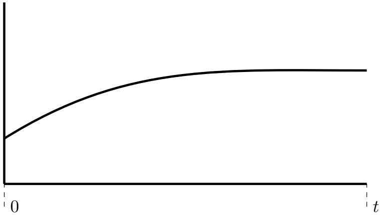


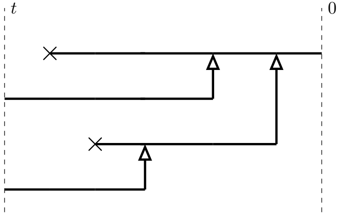
Theorem (Duality)

For $t \geq 0$,

$$y(t;y_0)^n = \mathbb{E}\Big[y_0^{R_t} \mid R_0 = n\Big] \qquad \forall n \in \mathbb{N}_0 \cup \{\Delta\},\ y_0 \in [0,1],$$

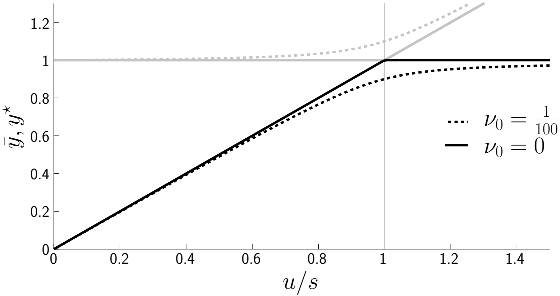
where $y^\Delta := 0$.





Stochastic representation of deterministic solution

Equilibrium frequency and absorption probability



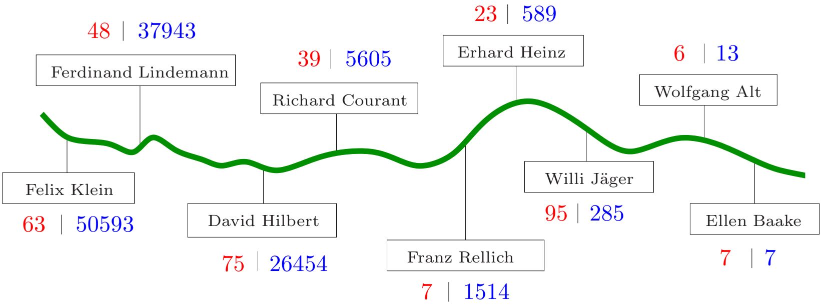
$$\bar{y} = \lim_{t\to\infty} \mathbb{E}\left[y_0^{R_t} \mid R_0 = 1\right] = \mathbb{P}(R \text{ absorbs in } 0 \mid R_0 = 1)$$

$\nu_0 = 0$: R birth-death process (birth rate s, death rate u per line)

$u \geqslant s$: dies out w.p. 1 ($\rightsquigarrow$ absorbs in 0)

$u < s$: dies out w.p. $\frac{u}{s} < 1$; survives with probability $1 - \frac{u}{s} > 0$, then grows to ∞ almost surely

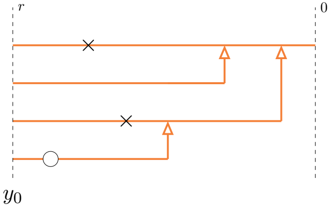
Mathematics genealogy project



Ancestral type

Definition

The ancestral type at backward time r, denoted by $I_r \in \{0,1\}$, is the type of the ancestor at backward time r of an individual uniformly chosen at time 0.



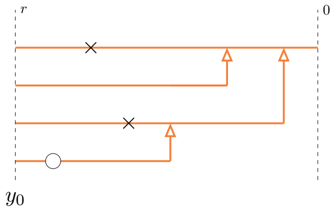
Ancestral type

Definition

The ancestral type at backward time r, denoted by $I_r \in \{0,1\}$, is the type of the ancestor at backward time r of an individual uniformly chosen at time 0.

Quantities of interest:

- $g(y_0, r) := P(I_r = 1 \mid Y_0 = y_0)$
- $g_\infty(y_0) := \lim_{r\to\infty} g(y_0, r)$
  asymptotic ancestral type distribution
- $g_\infty(\bar{y})$
  ancestral type distribution in equilibrium



Pruned lookdown ASG (p-LD-ASG)

Idea:

- Count potential ancestors of a single individual
- Arrange ancestral lines in hierarchy according to pecking order
- Prune lines upon mutation

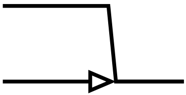
p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r ($L_r > 0$)

p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r ($L_r > 0$)

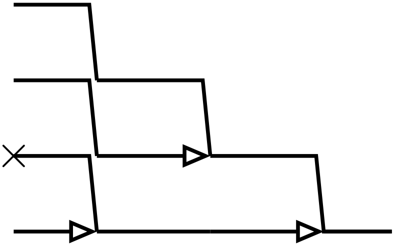
$$q_L(k, k+1) = ks$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r ($L_r > 0$)

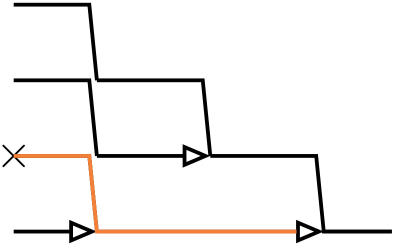
$$q_L(k, k+1) = ks$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r $(L_r > 0)$

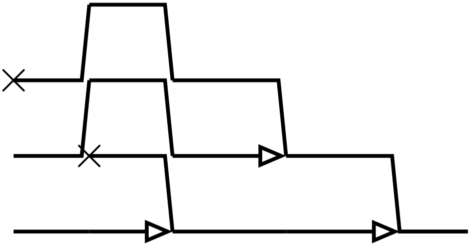
$$q_L(k, k+1) = ks$$
$$q_L(k, k-1) = (k \qquad)u\nu_1$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r $(L_r > 0)$

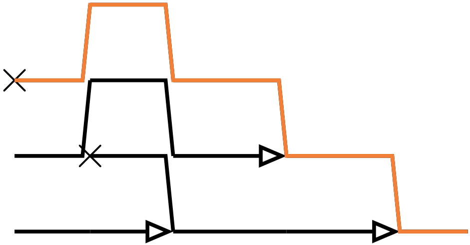
$$q_L(k, k+1) = ks$$
$$q_L(k, k-1) = (k \qquad)u\nu_1$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r $(L_r > 0)$

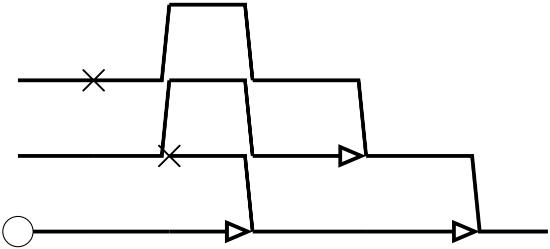
$$q_L(k, k+1) = ks$$
$$q_L(k, k-1) = (k-1)u\nu_1$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r ($L_r > 0$)

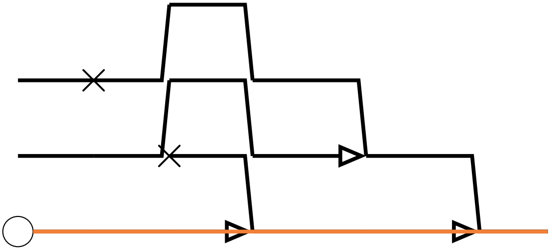
$$q_L(k, k+1) = ks$$
$$q_L(k, k-1) = (k-1)u\nu_1$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r $(L_r > 0)$

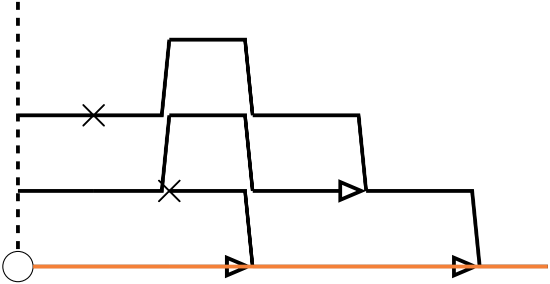
$$q_L(k, k+1) = ks$$
$$q_L(k, k-1) = (k-1)u\nu_1$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r ($L_r > 0$)

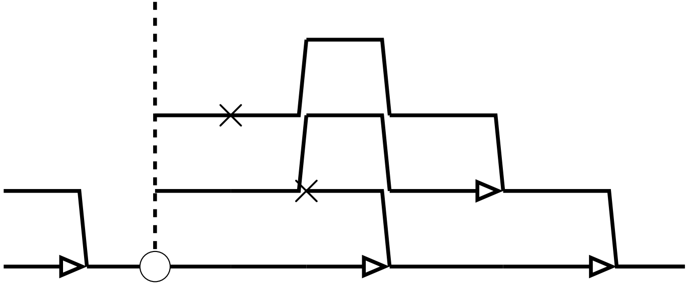
$$q_L(k, k+1) = ks$$
$$q_L(k, k-1) = (k-1)u\nu_1 + \mathbb{1}_{\{k>1\}}u\nu_0$$
$$q_L(k, \ell) = u\nu_0, \quad \ell \in \{1, \dots, k-2\}$$



p-LD-ASG in deterministic limit

L_r number of potential ancestors at backward time r $(L_r > 0)$

$$q_L(k, k+1) = ks$$
$$q_L(k, k-1) = (k-1)u\nu_1 + \mathbb{1}_{\{k>1\}} u\nu_0$$
$$q_L(k, \ell) = u\nu_0, \quad \ell \in \{1, \ldots, k-2\}$$



$\rightsquigarrow$ ancestor of type $1 \Leftrightarrow$ all lines of type $1 \rightsquigarrow$ bias towards type 0

p-LD-ASG - exploiting the hierarchy

- $\rightsquigarrow$
$$\begin{aligned} g(y_0, r) &= P(I_r = 1 \mid Y_0 = y_0) \\ &= \mathbb{E}[y_0^{L_r} \mid L_0 = 1] \end{aligned}$$
- $g_\infty(y_0) := \lim_{r \to \infty} g(y_0, r)$??

Asymptotic behaviour of L

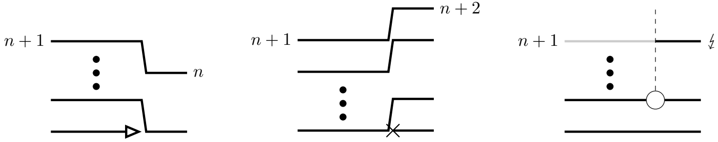
Proposition

1. *If* $s = 0$, $L_r \equiv 1$, *so, in particular,* $L_\infty := \lim_{r\to\infty} L_r = 1$.
2. *If* $u \leqslant s$ *and* $\nu_0 = 0$, $L_\infty = \infty$ *(a.s. for* $u < s$, *in probability for* $u = s$*).*
3. *For* $s > 0$ *and either* $u > s$ *or* $\nu_0 > 0$, L_∞ *has the stationary distribution Geo*$(1-p)$ *with*

$$p = \begin{cases} \frac{s}{u\nu_1}\bar{y}, & \text{if } \nu_1 > 0, \\ \frac{s}{u+s}, & \text{if } \nu_1 = 0. \end{cases}$$

Intuition behind the geometric law

- On $\{L_\infty > n\}$: first event (from left to right) on the first n lines



- First-step decomposition:

$$P(L_\infty > n) = \frac{s}{u+s} P(L_\infty > n-1) + \frac{u\nu_1}{u+s} P(L_\infty > n+1)$$

- $\rightsquigarrow$ lack of memory property, $p = \dots$

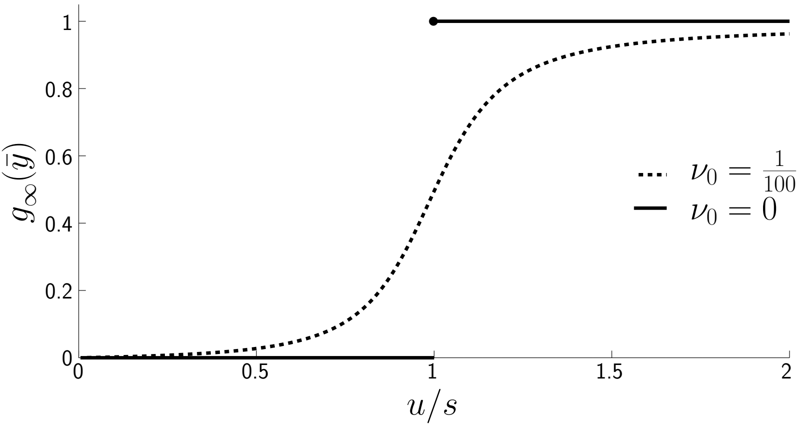
Asymptotic ancestral type distribution

$$g_\infty(y_0, r) = \lim_{r\to\infty} P(I_r = 1 \mid Y_0 = y_0) = \mathbb{E}[y_0^{L_\infty} \mid L_0 = 1]$$

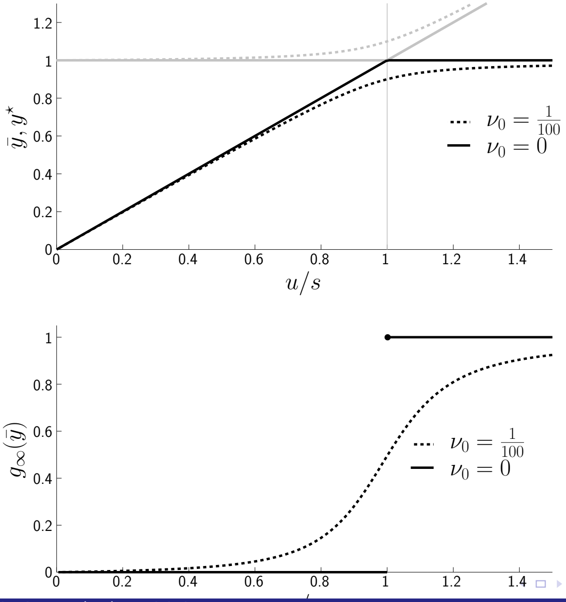
Theorem

1. *If* $s = 0$, $g_\infty(y_0) = y_0$ *for all* $y_0 \in [0,1]$.
2. *If* $u \leq s$ *and* $\nu_0 = 0$, $g_\infty(y_0) = \begin{cases} 0, & \textit{if } y_0 \in [0,1), \\ 1, & \textit{if } y_0 = 1. \end{cases}$
3. *If* $s > 0$ *and* $(u > s$ *or* $\nu_0 > 0)$, $g_\infty(y_0) = \frac{1-p}{1-py_0} y_0$.

Ancestral type distribution in equilibrium



The present and the past



Original article:

E. Baake, F. Cordero, and S. Hummel, A probabilistic view on the deterministic mutation-selection equation: dynamics, equilibria, and ancestry via individual lines of descent
J. Math. Biol. 2018

Review article:

E. Baake and A. Wakolbinger, Lines of descent under selection,
J. Stat. Phys. 2018

Special issue on 'Statistical Theory of Biological Evolution'
edited by K. Jain and L. Peliti