Ancestral processes with selection

Ellen Baake

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

\[+ \quad d_i \quad i \quad b_i \quad i\]

\[\downarrow \quad m_{ij} \quad j\]

(Malthusian) fitness: $r_i := b_i - d_i$
\( y_i(t) \) abundance of type \( i \) at time \( t \) \( (i \in S) \)

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

or

\[
\dot{y}(t) = y(t) \left( \widehat{A} + \underbrace{\widehat{M} + \widehat{R}}_{M+R} \right)
\]

with solution

\[
y(t) = y(0) e^{tA}
\]
MuSe differential equation

relative frequencies:

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

\[ p_i(t) := \frac{y_i(t)}{\sum_j y_j(t)} \quad \bar{r}(t) := \sum_j r_j p_j(t) \]
\(\mathcal{M}\) irreducible \(\Rightarrow\) Perron-Frobenius: \(pA = \lambda p\) \((\langle p, 1 \rangle = 1)\)

\[p(t) \xrightarrow{t \to \infty} p\] (stationary type distribution)

\[\lambda = \sum_i r_i p_i = \langle p, r \rangle = \bar{r}\] (equilibrium mean fitness)
Multitype Branching

\[ \begin{align*}
  &+ \quad d_i \\
  &i \quad b_i \\
  &\quad m_{ij} \\
  &\quad j
\end{align*} \]

\( i \)-individual:

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

then: birth, death, mutation to \( j \) with probability \( \frac{b_i}{a_i}, \frac{d_i}{a_i}, \frac{m_{ij}}{a_i} \)
Multitype branching

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

\[
\mathbb{E}^i(Z_j(t)) = (e^{tA})_{i,j}
\]

assumption: \( \lambda > 0 \) \( \leadsto \) branching supercritical \( \text{(with asymptotic growth rate } \lambda \text{)} \)
Connections "branching" $\leftrightarrow$ "MuSe":

1. $\frac{Z(t)}{|Z(t)|} \xrightarrow{t \to \infty} p \xleftarrow{t} p(t)$
   
a.s., on \{non-extinction\} (Kesten–Stigum '66)

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

2. $\mathbb{E}^i (|Z(t)| e^{-\lambda t}) \xrightarrow{t \to \infty} h_i \xleftarrow{t} \frac{\sum_j (e^{tA})_{ij}}{|p(0)e^{tA}|}$

$h = (h_i)_{i \in S}$ 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)

\[
\tau \to \infty \quad \Rightarrow \quad \alpha_i := p_i h_i
\]

ancestral type distribution

Jagers, Nerman 1992 ...
Forward and backward

\[ p_0 = 0.5, \quad \alpha_0 = 0.5 \]

\[ s = 0.001, \quad \nu_0 = 0.005 \quad \text{and} \quad \nu_0 \to 0 \]

\[ \nu_0 + \nu_1 = 1 \]

\[ \nu_1 \gg \nu_0 \]
Forward and backward
Large deviations

mutation process \((M_t)\) on representative line \((M_t\) type at time \(t)\)
(generator \(M\), stationary distribution \(\pi)\)
Large deviations

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

LDP:
\[
P(L_t \sim \nu) \approx e^{-t I_M(\nu)} \quad \text{(large } t\text{)}
\]
\[
\left( \lim_{t \to \infty} \frac{1}{t} \log P(L_t \in A) = -\inf_{\nu \in A} I_M(\nu) \right)
\]

rate function:
\[
I_M(\nu) = \sup_{\nu > 0} \left( -\langle \nu, \frac{M\nu}{\nu} \rangle \right)
\]

\((M_t)\) reversible \(\leadsto I_M(\nu) = -\langle \sqrt{\frac{\nu}{\pi}}, \mathcal{M} \sqrt{\frac{\nu}{\pi}} \rangle_\pi\)
Variational principle

line with $L_t = \nu$ experiences

- mutation: changes $\nu$
  \[ \mathbb{P}(L_t \sim \nu) \text{ decays with } I_M(\nu) \quad (> 0 \text{ for } \nu \neq \pi \text{ (stat. distr.)}, \]
  \[ = 0 \text{ for } \nu = \pi) \]
- reproduction: duplicates $\nu$ at rate $r_M$ at time $t$
  mean rate $\langle \nu, r \rangle$

Theorem (EB & Georgii 2007)

\[ \langle p, r \rangle = \lambda = \sup_{\nu \in \mathcal{P}(S)} \left[ \langle \nu, r \rangle - I_M(\nu) \right] = \langle \alpha, r \rangle - I_M(\alpha) \]

↑

“present”

↑

energy

↑

entropy

↑

“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 $1 + s^N$, $s^N \geq 0$
- single offspring inherits parent's type and replaces uniformly chosen individual
- mutation at rate $u^N > 0$
- resulting type: 0 with probability $\nu_0$; 1 with probability $\nu_1$ ($\nu_0 + \nu_1 = 1$)
Interacting particle system: untyped

- \( \bullet \)
- \( \bullet \)
- \( \bullet \)
- \( \bullet \)
- \( \bullet \)

- Time \( t \)
- Population size \( N \)

- Type 0 (fit)
- Type 1 (unfit)

- Mutations to type 0: rate \( \nu^0_N \)
- Mutations to type 1: rate \( \nu^1_N \)

- Proportion of individuals of type 1 at time \( t \) in MoMo of size \( N \)

Ancestral processes with selection

Ellen Baake
Bielefeld University
Interacting particle system: untyped

neutral arrows: rate 1,
Interacting particle system: untyped

\[
N_1 \times \cdots \times N_k \times \text{type} 1 \quad (\text{unfit}) \\
\times \text{type} 0 \quad (\text{fit})
\]

\[
\text{mutation to type } 1:\quad \text{rate } u_1 N_\nu 1 \\
\text{mutation to type } 0:\quad \text{rate } u_0 N_\nu 0
\]

\[Y_N t := \text{proportion of individuals of type } 1 \text{ at time } t \text{ in MoMo of size } N\]

Ancestral processes with selection

Ellen Baake
Bielefeld University

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

Ancestral processes with selection

Ellen Baake  Bielefeld University

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

Ancestral processes with selection

Ellen Baake  Bielefeld University

neutral arrows: rate 1,
mutation to type 1: rate $u^N \nu_1$,
selective arrows: rate $s^N$,
movement to type 0: rate $u^N \nu_0$
Interacting particle system: typed

\[ Y_t^N := \text{proportion of individuals of type 1 at time } t \text{ in MoMo of size } N \]
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

\(\leadsto\) type at present together with true ancestral line
Diffusion limit

\[ N \to \infty \quad \text{s.t.} \quad Ns^N \to \sigma, \quad N\nu^N \to \theta, \quad Y_0^N \to y_0 \]
\[ \sim (Y_{tN}^N) \xrightarrow{d} \text{Wright–Fisher diffusion (} Y_t) \]
\[ dY_t = \sqrt{Y_t(1 - Y_t)} \, dW_t - \sigma Y_t(1 - Y_t) \, dt + (1 - Y_t)\theta\nu_1 \, dt - Y_t\theta\nu_0 \, dt, \]
\[ Y_0 = y_0, \quad (W_t) \text{ standard Brownian motion} \]

\[ \theta > 0, \ 0 < \nu_0 < 1, \ t \to \infty \sim Y_t \xrightarrow{d} \tilde{Y} \]
ASG in diffusion limit

- branching at rate $\sigma$ 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

\[ r \sim Y \]
Ancestral type

\[ I_r \in \{0, 1\} \]
\[ := \text{ancestral type at time } r \]

long-term success
bias towards type 0

\[ r \to \infty \sim I_r \overset{d}{\to} \tilde{I} \]

\[ \mathbb{P}(\tilde{I} = 0), \mathbb{P}(\tilde{I} = 1)? \]
Ordering the ASG (w/o mutation)

arrange lines according to pecking order (exchangeability!)

Diagram:

```
  ______
  ______
  ______
  ______
  ______
  ______
```

```
  ______
  ______
  ______
  ______
  ______
```

```
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!)

The ancestral line is the lowest type-0 line if there is one, otherwise the immune line.
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

The ancestral line is

- lowest type-0 line if there is one
- immune line otherwise
The line-counting process of the pruned ASG

$L_r = \text{number of lines at time } r$

$(L_r)$ Markov chain in continuous time with rates from

\[
\begin{array}{c}
\uparrow \\
\longrightarrow \\
\times \\
\longrightarrow \\
\downarrow \\
\end{array}
\]

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

\[
a_n := \mathbb{P}(\tilde{L} > n), \quad n \geq 0
\]

first-step analysis $\leadsto$ (Fearnhead’s) recursion:

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

\[
a_0 = 1, \quad \lim_{n \to \infty} a_n = 0.
\]
The type distribution of the ancestral line in the distant past is given by

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

\[ P(\tilde{I} = 1) = \sum_{n>0} (a_{n-1} - a_n) b_n, \quad P(\tilde{I} = 0) = \sum_{n \geq 0} a_n (b_n - b_{n+1}), \]

where \( b_n := \mathbb{E}(\tilde{Y}^n) \).
The bias towards type 0

distribution of $\widetilde{L}$

probability of fit ancestor

Ancestral processes with selection

Ellen Baake  Bielefeld University
Forward and backward

\[ F(1 - \tilde{Y}) \]

\[ P(\tilde{I} = 0) \]

\[
\begin{align*}
    s &= 0.001, \ \nu_0 = 0.005 \\
    \sigma &= N s, \ \vartheta = N u \text{ with } N = 10^4, N = 3 \cdot 10^4, N = 10^5
\end{align*}
\]
Large deviations and variational principle for (multitype) MoMo?