A $\Lambda$-Coalescent Dual Process
in a Cannings model with Genic Selection

Bob Griffiths
University of Oxford

Joint research with
Alison Etheridge, Jay Taylor.
A Cannings model with selection

$N$ individuals and type space $E$.
Reproduction at an overall rate of $\lambda$.
At a reproduction event a single individual is chosen to reproduce (Birkner and Blath 2009).
Distribution of the number of offspring is $\{r_a, 1 \leq a \leq N - 1\}$.
There is viability selection of offspring, if a type $i \in E$ individual has $a$ offspring, then $b$ are viable and $b - a$ are not viable with probability $v_{iab}$.
Viable offspring distribution

$$p_{ib} = \sum_{a=b}^{N-1} r_a v_{iab}, \ 0 \leq b \leq N - 1.$$
Reproduction

Viable arrows are purple.
Non-viable arrows are orange.

$\lambda$ Event rate
$r_a$ Total offspring distribution
$v_{iab}$ Viability distribution of a type $i$ parent with $a$ offspring
Transition rates for numbers of types 

\[ z = (z_i, i \in E) \] are numbers of individuals of types in \( E \).

\[ z \to z + |\gamma| e_i - \gamma \] at rate

\[ \phi(z; z + |\gamma| e_i - \gamma) = \frac{\lambda z_i}{N} p_i |\gamma| H(\gamma | z - e_i) \]

where \( H \) is the hypergeometric distribution

\[ H(\gamma | z - e_i) = \frac{\prod_{j \in E} (z_j - \delta_{ij})}{\left(\begin{array}{c} N - 1 \end{array}\right)} \]

with \( \gamma, z \in \mathbb{Z}_+^E \) with \( |\gamma| \) and \( |z| = N \) held fixed.

\[ z_1 = 7, \quad z_2 = 3, \quad |\gamma| = 3, \quad \gamma_1 = 2, \quad \gamma_2 = 1 \]
Looking back in time
Coalescence

Red lines are ancestral back in time.
Blue lines are non-ancestral back in time.

Parent can be inside or outside the ancestral lines.
At least two from the parent and viable offspring are inside the ancestral lines.
Looking back in time
Selection from outside ancestral lines

Red lines are ancestral back in time.
Blue lines are non-ancestral back in time.

Parent is outside the ancestral lines.
At least one non-viable offspring is inside the ancestral lines and all viable offspring are outside.
\[ \xi = (\xi_i, i \in E) \] configuration of types in dual ancestral lines.

\[ E = \{ \bullet, \cdot \} \]

- Parent arrow tail

Purple arrows are viable, Orange arrows are non-viable.

Initial first six individuals \( \xi_\bullet = 3, \xi_\cdot = 3 \).

Red ancestral lineages, Blue non-ancestral lineages.

Five ancestors at time zero \( \xi_\bullet = 3, \xi_\cdot = 2 \).
Dual process rates, $\xi = (\xi_i, i \in E)$ is the configuration of types in ancestral lines back in time.

$$q(\xi, \xi - e_i(l - 1)) = \lambda \sum_{|\gamma| \in [N-1]} p_{i|\gamma|} \frac{(|\xi|)}{l} \frac{(N-|\xi|)}{(|\gamma|+1-l)}$$

$$q(\xi, \xi + e_i) = \lambda \frac{N - |\xi|}{N} p_{i,|\xi|}$$

$$q(\xi, \xi + e_i - e_j) = |\xi| \mu_{ij} \frac{(\xi_i + 1 - \delta_{ij})}{|\xi|} \frac{\mathcal{H}(\xi + e_i - e_j)}{\mathcal{H}(\xi)}$$
Dual process
Coalescence rates \( l \geq 2 \)

\[
q(\xi, \xi - e_i(l - 1)) = \lambda \sum_{|\gamma|\in[N-1]} p_{i|\gamma} \left( \frac{|\xi|}{l} \right) \left( \frac{N-|\xi|}{|\gamma|+1-l} \right) \left( \frac{N}{|\gamma|+1} \right) \times \frac{\xi_i + 1 - l}{|\xi| + 1 - l} \frac{H(\xi - e_i(l - 1))}{H(\xi)}
\]

Rate at which there are \( l \) from the parent and viable offspring in the ancestral lines \( \times \) the conditional probability that the last parent is type \( i \) given a configuration \( \xi \).
Coalescence

$$\sum_{|\gamma| \in [N-1]} p_{i|\gamma|} \frac{\binom{|\xi|}{l} \binom{N-|\xi|}{|\gamma|+1-l}}{\binom{N}{|\gamma|+1}}$$

$|\xi| = 5, \quad |\gamma| = 4, \quad l = 4, \quad |\xi| - l + 1 = 5 - 4 + 1 = 2$
Dual process
Selection from outside lines

\[ q(\xi, \xi + e_i) = \lambda \frac{N - |\xi|}{N} p^*_{i,|\xi|} \]
\[ \times \frac{\xi_i + 1}{|\xi| + 1} \frac{\mathcal{H}(\xi + e_i)}{\mathcal{H}(\xi)} \]

Rate at which the last parent is outside the ancestral lines with all viable offspring outside the ancestral lines and at least one non-viable offspring inside the ancestral lines \( \times \) the conditional probability that the parent is type \( i \) given a configuration \( \xi \).

\[
p^*_{i,|\xi|} = \sum_{|\gamma| \in [N-1]} \sum_{k \in [N-1]} r_k v_{ik} |\gamma| \left\{ \frac{(N-1-|\xi|)}{|\gamma|} \right\} - \frac{\left( \frac{N-1-|\xi|}{k} \right)}{\left( \frac{N-1}{k} \right)} \]

\[
\sum_{|\gamma| \in [N-1]} \sum_{k \in [N-1]} r_k v_{ik} |\gamma| \left\{ \frac{(N-1-|\xi|)}{|\gamma|} \right\} - \frac{\left( \frac{N-1-|\xi|}{k} \right)}{\left( \frac{N-1}{k} \right)} \]
Mutation along ancestral lines

\[ q(\xi, \xi + e_i - e_j) = |\xi|\mu_{ij} \times \frac{(\xi_i + 1 - \delta_{ij})H(\xi + e_i - e_j)}{|\xi| H(\xi)} \]

The red part of the equation is the conditional probability that the last parent is type \(i\) given a configuration \(\xi\).

The collection of \(\bullet\) pairs can be different types in \(E\).
Dual generator approach
Generator test functions

\[ f_\xi(z) = \prod_{i \in E} z_i^{\xi_i} \]

where \( a_{[b]} = a(a - 1) \ldots a - b + 1 \).

To identify the dual, write

\[
\mathcal{L} f_\xi(z) = \sum_{i \in E, \gamma \in \Delta_{N-1}} \phi(z; z + |\gamma|e_i - \gamma)[f_\xi(z + |\gamma|e_i - \gamma) - f_\xi(z)] \\
+ \sum_{i, j \in E} z_i \mu_{ij} [f_\xi(z - e_i + e_j) - f_\xi(z)] \\
= \sum_{\chi} c(\xi, \chi) [f_\chi(z) - f_\xi(z)],
\]

where \( \{c(\xi, \chi), \xi, \chi \in \Delta_N\} \) are non-negative constants.
\[ \mathbb{E}[\mathcal{L} f_{\xi}(Z)] = 0 \text{ implies} \]
\[ 0 = \sum_{\chi} c(\xi, \chi) [m_\chi - m_\xi] \]

Rate matrix
\[ q(\xi, \chi) = \begin{cases} 
 c(\xi, \chi)m_\chi/m_\xi, & \xi \neq \chi \\
 -\sum_{\chi} c(\xi, \chi), & \xi = \chi 
\end{cases} \]

Generator acting on \( g_{\xi}(z) = f_{\xi}(z)/m_\xi \) as a function of \( \xi \)
\[ \mathcal{L} g_{\xi}(z) = \sum_{\chi} q(\xi, \chi) [g_\chi(z) - g_{\xi}(z)]. \]
Sampling distribution

\[ g_\xi(z) = \frac{\mathcal{H}(\xi | z)}{\mathcal{H}(\xi)}, \]

where

\[ \mathcal{H}(\xi) = \mathbb{E}_\varphi[H(\xi | Z)] \]

is the unconditional sampling distribution of a sample of \(|\xi|\) individuals in a stationary distribution \(\varphi\).

\[ \tilde{\mathcal{H}}(z | \xi) = g_\xi(z)\varphi(z) \]

is the posterior distribution of the frequency of types in a stationary population, conditional on a sample configuration of \(\xi\) in \(|\xi|\) distinct individuals.
The Dual Equation

\[ \mathcal{L} \frac{\mathcal{H}(\xi \mid z)}{\mathcal{H}(\xi)} = \sum_{\chi \in \Delta_N} q(\xi, \chi) \left[ \frac{\mathcal{H}(\chi \mid z)}{\mathcal{H}(\chi)} - \frac{\mathcal{H}(\xi \mid z)}{\mathcal{H}(\xi)} \right] \]

where \( \mathcal{L} \) is now regarded as acting on the index \( \xi \).

A Markov process \( \{L(t), t \geq 0\} \) in \( \mathbb{Z}_+^E \), \( |L(t)| \leq N \) with rate matrix \( Q \) is dual to \( \{Z(t), t \geq 0\} \) with duality equation

\[ \mathbb{E}_{Z(0)} \left[ \frac{\mathcal{H}(L(0) \mid Z(t))}{\mathcal{H}(L(0))} \right] = \mathbb{E}_{L(0)} \left[ \frac{\mathcal{H}(L(t) \mid Z(0))}{\mathcal{H}(L(t))} \right] \]

where expectation on the left is with respect to the distribution of \( Z(t) \) and on the right with respect to the distribution of \( L(t) \).
Transition function expansion
\[ \{ s_{yz}(t), y, z \in \mathbb{Z}_+^E, |y| = |z| = N, t \geq 0 \} \] are transition functions of \( \{ Z(t), t \geq 0 \} \)
\[ \{ h_{\xi\chi}(t), \xi, \chi \in \Delta_N, t \geq 0 \} \] are transition functions of the dual process \( \{ L(t), t \geq 0 \} \).

Dual expansion
\[
s_{yz}(t) = \varphi(z) \sum_{\xi \in \Delta_N} h_{z\xi}(t) \frac{\tilde{H}(y \mid \xi)}{\varphi(y)}
\]
for \( y, z \in \mathbb{Z}_+^E, |y| = |z| = N \).
\( \varphi \) is the stationary distribution of the process.
\( \tilde{H}(y \mid \xi) \) is the posterior distribution of the frequency of types in a stationary population, conditional on a sample configuration of \( \xi \) in \( |\xi| \) distinct individuals.
Theorem. For \( y, z \in \mathbb{Z}^E_+ \), \(|y| = |z| = N\),
\[
s_{yz}(t) = \sum_{\alpha \in \mathbb{Z}^E_+, |\alpha| \leq N} h_{y\alpha}(t) \pi(z - \alpha; \theta + \alpha, N - |\alpha|),
\]
where \( \{s_{yz}(t), t \geq 0\} \) are the transition functions in the Moran model with parent independent mutation and genic selection. \( \{Z(t), t \geq 0\} \) and \( \{h_{y\alpha}(t), t \geq 0\} \) are transition functions of multi-type birth and death process \( \{L(t), t \geq 0\} \).
\[
\pi(z; \theta, N) = \prod_{i \in E} \lambda_i^z \binom{N}{z} \frac{\prod_{i \in E} \theta_i(z_i)}{\theta(N)}
\]
is the \( \lambda \)-weighted Multinomial-Dirichlet distribution. Etheridge and Griffiths (2009).
Lambda coalescent: General generator form for $\mathcal{L}_\Lambda g(x)$

$$
\frac{1}{2} F(\{0\}) \sum_{j,k \in E} x_j (\delta_{jk} - x_k) \frac{\partial^2}{\partial x_j \partial x_k} g(x)
$$

$$
+ \sum_{i \in E} \int_{[0,1]} x_i (g(x(1 - y) + y e_i) - g(x)) \frac{F(dy)}{y^2}
$$

$$
+ \sum_{i \in E} \int_{[0,1]} x_i \left( g(x(1 - y) + y e_i) - g(x) - y \sum_{j \in E} (\delta_{ij} - x_j) \frac{\partial}{\partial x_j} g(x) \right) \frac{K_i(dy)}{y}
$$

$$
+ \sum_{j \in E} x_j \left( \sigma_j - \sum_{k \in E} \sigma_k x_k \right) \frac{\partial}{\partial x_j} g(x) + \sum_{i,j \in E} \left( x_i \mu_{ij} - x_j \mu_{ji} \right) \frac{\partial}{\partial x_j} g(x)
$$
Offspring measures \( \{G_i(\cdot); i \in E\} \) have a decomposition

\[
\frac{G_i(dy)}{y^2} = \int_{(y,1]} V_i(x, dy) \frac{F(dx)}{x^2}
\]

Define signed measures

\[
K_i(dy) = \frac{G_i(dy) - F(dy)}{y}
\]

with \( \sigma_i = K([0, 1]) \).

Assume that \( |K([0, 1])| < \infty \).
Dual Lambda coalescent rates

\[ q_\Lambda(\xi, \xi - e_i(l - 1)) = \int_{[0,1]} \left( \frac{|\xi|}{l} \right) y^l (1 - y) |\xi| - l \frac{G_i(dy)}{y^2} \]

\[ \times \frac{\xi_i + 1 - l}{|\xi| + 1 - l} \cdot \frac{\mathcal{M}(\xi - e_i(l - 1))}{\mathcal{M}(\xi)} \]

\[ q_\Lambda(\xi, \xi + e_i) = \int_{[0,1]} \left( 1 - (1 - x)|\xi| \right) \frac{-K_i(dx)}{x} \]

\[ \times \frac{\xi_i + 1}{|\xi| + 1} \cdot \frac{\mathcal{M}(\xi + e_i)}{\mathcal{M}(\xi)} \]

\[ q_\Lambda(\xi, \xi + e_i - e_j) = \mu_{ij}(\xi_i + 1 - \delta_{ij}) \cdot \frac{\mathcal{M}(\xi + e_i - e_j)}{\mathcal{M}(\xi)} \]

In a viability model

\[ \int_{[0,1]} \left( 1 - (1 - x)|\xi| \right) \frac{-K_i(dx)}{x} \geq 0 \]
Weak selection

\[ q_\Lambda(\xi, \xi + e_i) = \frac{|\xi|}{|\xi| + 1}(\xi_i + 1)(-\sigma_i) \cdot \frac{\mathcal{M}(\xi + e_i)}{\mathcal{M}(\xi)} \]

If \(|\sigma_i| < \infty\) for \(i \in E\) without loss of generality take \(\sigma_i \leq 0\) for all \(i \in E\) because of shift invariance under the (forward) distribution.

Limit as \(\epsilon \to 0\) from the general generator by choosing

\[ K_i(\cdot) = \sigma_i \delta_\epsilon(\cdot) \]