The Λ-Coalescent with Selection

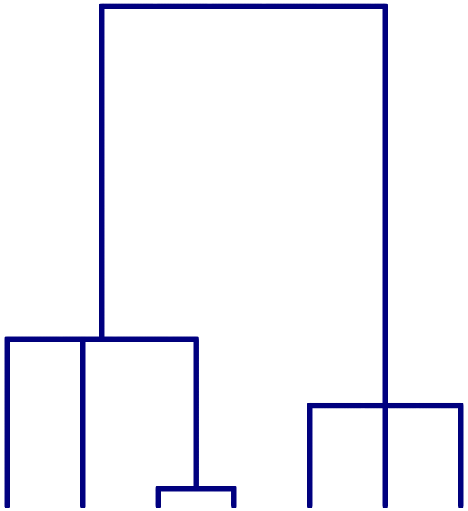
Bob Griffiths

University of Oxford

Joint research with
Alison Etheridge, Jay Taylor.

Neutral Λ coalescent

A random generalogical tree with multiple mergers.
Pitman (1999), Sagitov (1999).



Coalescence rates

The rate at which a specific m lineages coalesce when there are n lineages is

$$\lambda_{nm} = \int_0^1 x^m (1-x)^{n-m} \frac{\Lambda(dx)}{x^2},\ m \geq 2$$

where Λ is a probability measure.
Edges in the tree $n \to n - m + 1$.

Rates must be of this form for a consistent distribution of subtrees.

Examples

$$\lambda_{nm} = \int_0^1 x^m (1-x)^{n-m} \frac{\Lambda(dx)}{x^2},\ 2 \le m \le n$$

Kingman coalescent

$\Lambda(\{0\}) = 1$, $\lambda_{n2} = 1$.

Wakeley coalescent

$\Lambda(\{\psi\}) = 1$, $\lambda_{nm} = \psi^{m-2}(1-\psi)^{n-m},\ m \ge 2.$

Schweinsberg coalescent

Λ is a Beta $(2-\alpha, \alpha)$ measure, $0 < \alpha < 2$.

$$\lambda_{nm} = \frac{(2-\alpha)_{(m-2)}\alpha_{(n-m)}}{(n-1)!}$$

Bolhausen-Sznitman coalescent

Λ is a uniform measure on $(0,1)$. ($\alpha = 1$.)

A coalescent tree with an infinite number of leaves
Coming down from infinity

Kingman coalescent comes down from infinity, $\mathbb{E}(T_{\text{MRCA}}) = 2$.
The Λ-coalescent comes down from infinity if Λ has an atom at 0 or else it has no atom at 0 and

$$\sum_{n=2}^{\infty} \Big(\sum_{k=2}^{n} (k-1)\lambda_{nk} \Big)^{-1} < \infty$$

Otherwise it stays infinite for all times.
Wakeley coalescent $\Lambda(\{\psi\}) = 1$.
NO
Schweinsberg coalescent Λ is a Beta $(2-\alpha, \alpha)$ measure.
YES iff $1 < \alpha < 2$.
Bolhausen-Sznitman coalescent Λ is a uniform measure on $(0, 1)$.
NO, Borderline.

Population Frequencies, d-alleles, $x = (x_1, \ldots, x_d)$

Generator

$$\mathcal{L}g(x) = \int_0^1 \sum_{i=1}^d x_i \big(g(x(1-y) + ye_i) - g(x)\big)\frac{F(dy)}{y^2} + (\theta/2d)\sum_{i=1}^d (1 - dx_i)\frac{\partial}{\partial x_i} g(x)$$

Local Behaviour

Choose a type i gene with probability x_i, then

$$\begin{aligned} x_i &\to (1-y)x_i + y \\ x_j &\to (1-y)x_j, \ \ j \neq i \end{aligned}$$

where the jump rate of y is $y^{-2}F(dy)$.

Mutations occur at random to individuals.

A Neutral Moran-Cannings model

N individuals and type space $[d] = \{1, 2, \ldots, d\}$.
Reproduction at an overall rate of λ.
At a reproduction event a single individual is chosen to reproduce.
Distribution of the number of offspring is $\{r_a, 1 \leq a \leq N - 1\}$.
Λ-coalescent

$$\lambda \sum_{a > Nx} r_a \to \int_{(x,1]} \frac{1}{y^2} F(dy)$$

where F is a finite measure on $[0, 1]$.
A single individual's offspring can replace a large proportion of the population.

A Moran-Cannings model with selection

N individuals and type space $[d] = \{1, 2, \dots, d\}$.
Reproduction at an overall rate of λ.
At a reproduction event a single individual is chosen to reproduce.
Distribution of the number of offspring is $\{r_a, 1 \leq a \leq N-1\}$.
There is viability selection of offspring, if a type $i \in [d]$ individual has a offspring, then b are viable and $b-a$ are not viable with probability v_{iab}.
Viable offspring distribution

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

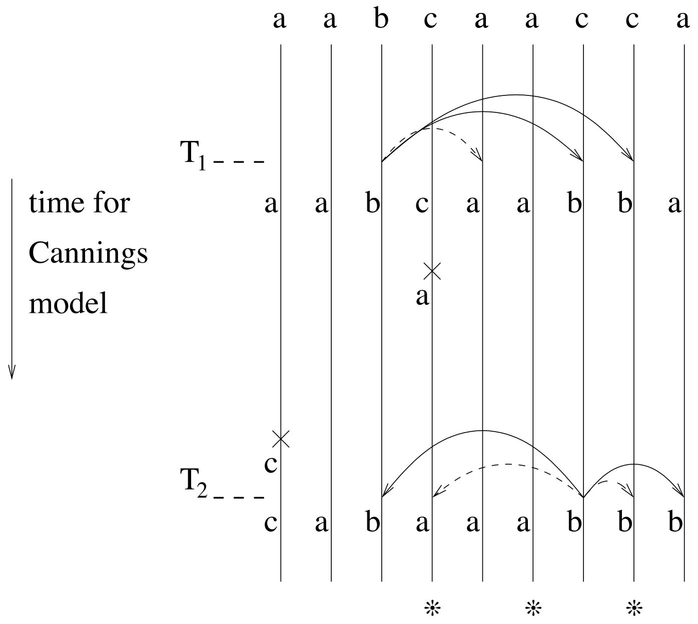
Viable offspring replace individuals at random keeping the population size at N.

Example of a viability offspring distribution

$$v_{iab} = \int_{[0,1]} \binom{a}{b} p^b (1-p)^{a-b} \nu_i(dp)$$

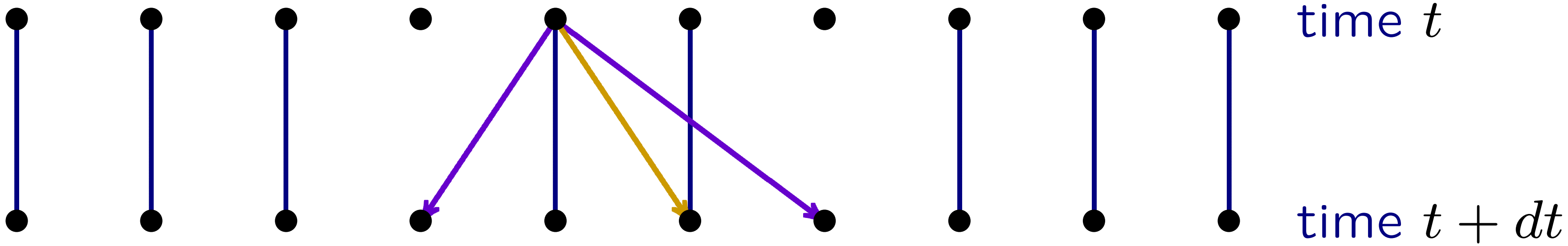
where $\{\nu_i(\cdot), i \in [d]\}$ is a collection of probability measures on $[0, 1]$.

Graphical representation



The population consists of three types, labelled a, b and c. At time T_1 a reproduction event takes place in which a type b parent produces three potential offspring of which two are viable. In the reproduction event at time T_2, two of the four potential offspring are viable. Mutations are denoted by crosses.

Reproduction



Viable arrows are purple.

Non-viable arrows are orange.

λ 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

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

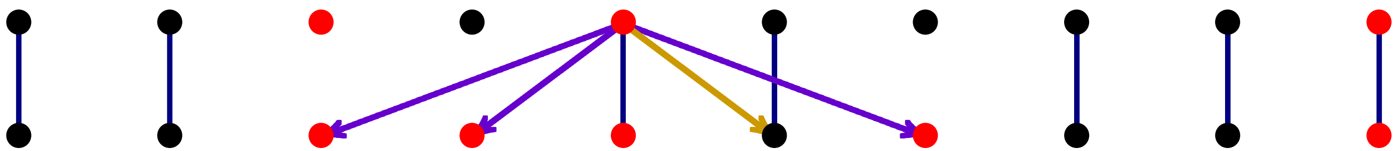
$z \longrightarrow z + |\gamma| e_i - \gamma$ at rate

$$\phi(z; z + |\gamma| e_i - \gamma) = \lambda \frac{z_i}{N} p_{i|\gamma|} \mathcal{H}(\gamma \mid z - e_i)$$

where $\mathcal{H}$ is the hypergeometric distribution

$$\mathcal{H}(\gamma \mid z - e_i) = \frac{\prod_{j \in [d]} \binom{z_j - \delta_{ij}}{\gamma_j}}{\binom{N-1}{|\gamma|}}$$

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



• $z_1 = 7$, • $z_2 = 3$, $|\gamma| = 3$, $\gamma_1 = 2$, $\gamma_2 = 1$

Moran model as a special case

$$\phi(z; z + e_i - e_j) = \lambda p_{i1} \frac{z_i}{N} \frac{z_j}{N-1}$$

Single offspring are viable or non-viable.

Mutation occurs at rate μ_{ij} from an individual of type i to type j, with $\sum_j \mu_{ij} = \mu$.

(μ_{ij}/μ) is assumed to be a transition matrix with a unique stationary distribution.

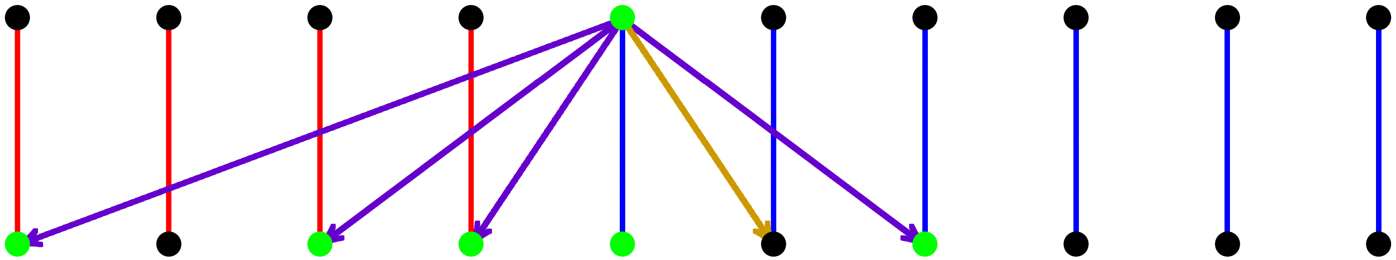
Then there is a stationary distribution φ of $\{Z(t), t \geq 0\}$, the number of individuals of the d-types in the population.

Sampling distribution

$$\mathcal{H}(\gamma) = \mathbb{E}^{\varphi}\left[\frac{\prod_{j\in[d]}\binom{z_j}{\gamma_j}}{\binom{N}{|\gamma|}}\right]$$

Looking back in time

Coalescence



Red lines are ancestral back in time.

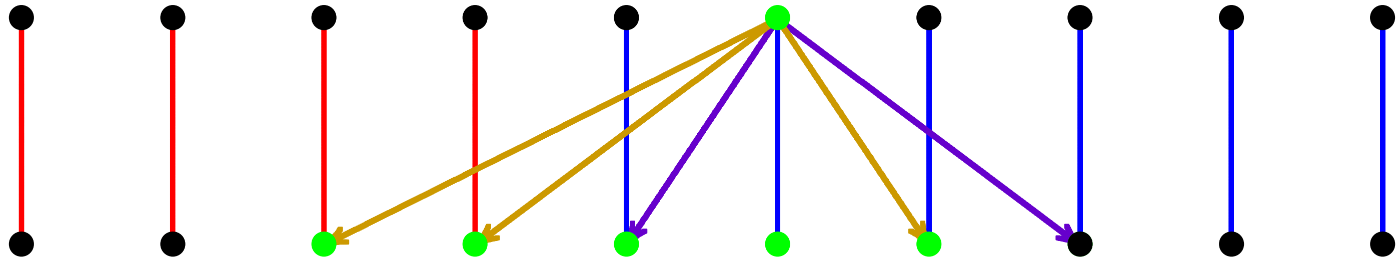
Blue lines are non-ancestral back in time.

Parent can be inside or outside the ancestral lines.

At least two from the parent and viable offspring are inside the ancestral lineages for coalescence.

Looking back in time

Selection from outside ancestral lines



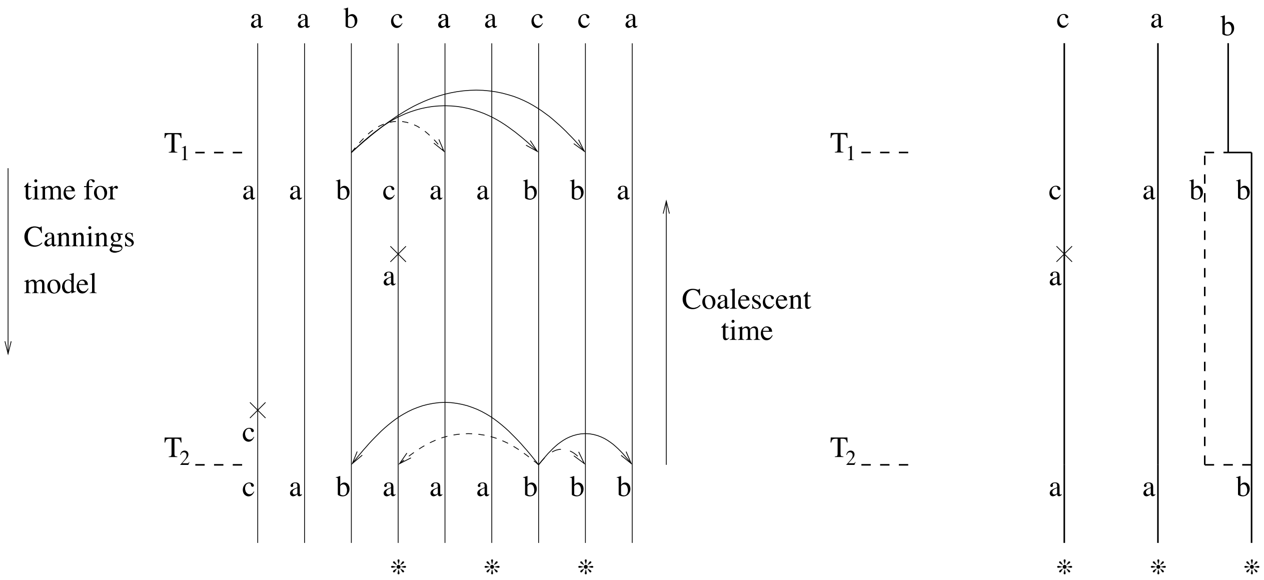
Red lines are ancestral back in time.

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.

Dual process



The dual process of real (solid lines) and virtual (dashed lines) lineages corresponding to sampling the individuals marked with an asterisk from the population in the last figure. At time T_2 we must add a virtual lineage since the sample is hit by two virtual arrows but no real arrows. Coalescence of real and virtual lineages results in a real lineage.

Dual process rates, $\xi = (\xi_i, i \in [d])$ is the configuration of types in ancestral lines back in time $(l \geq 2)$.

$$
\begin{aligned}
q(\xi, \xi - e_i(l-1)) &= \lambda \sum_{|\gamma| \in [N-1]} p_{i|\gamma|} \frac{\binom{|\xi|}{l}\binom{N-|\xi|}{|\gamma|+1-l}}{\binom{N}{|\gamma|+1}} \\
&\quad \times \frac{\xi_i + 1 - l}{|\xi| + 1 - l} \frac{\mathcal{H}(\xi - e_i(l-1))}{\mathcal{H}(\xi)} \\
q(\xi, \xi + e_i) &= \lambda \frac{N - |\xi|}{N} p^*_{i,|\xi|} \\
&\quad \times \frac{\xi_i + 1}{|\xi| + 1} \frac{\mathcal{H}(\xi + e_i)}{\mathcal{H}(\xi)} \\
q(\xi, \xi + e_i - e_j) &= |\xi| \mu_{ij} \times \frac{(\xi_i + 1 - \delta_{ij})}{|\xi|} \frac{\mathcal{H}(\xi + e_i - e_j)}{\mathcal{H}(\xi)}
\end{aligned}
$$

Dual process

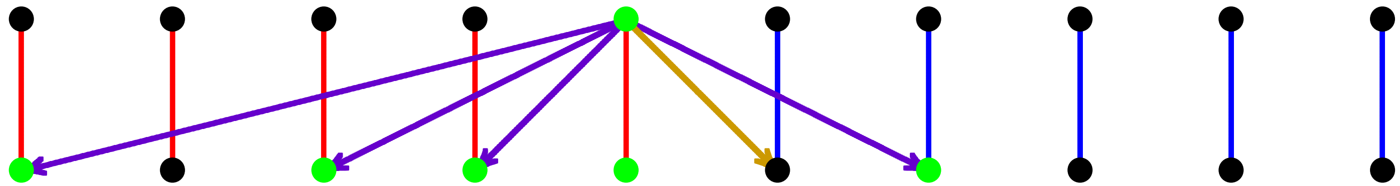
Coalescence rates $l \geq 2$

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

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 ξ.

Coalescence

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



$|\xi| = 5,\ |\gamma| = 4,\ l = 4,\ |\xi| - l + 1 = 5 - 4 + 1 = 2$

Dual process
Selection from outside lines

$$
\begin{aligned}
q(\xi, \xi + e_i) &= \lambda \frac{N - |\xi|}{N} p^*_{i,|\xi|} \\
&\quad \times \frac{\xi_i + 1}{|\xi| + 1} \frac{\mathcal{H}(\xi + e_i)}{\mathcal{H}(\xi)}
\end{aligned}
$$

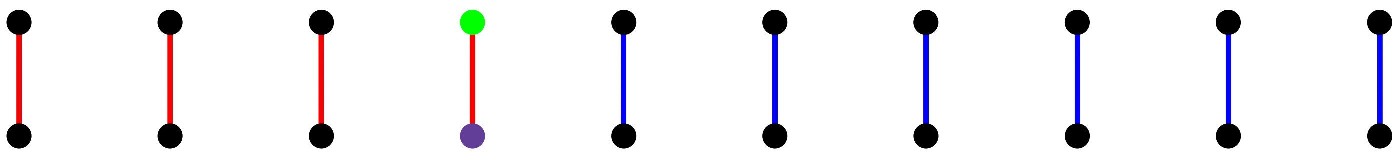
Rate at which the last parent is outside the ancestral lines with all viable offspring outside the ancestral lines and at least one non-viable offspring inside the ancestral lines × the conditional probability that the parent is type i given a configuration ξ.

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

Mutation along ancestral lines

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

The red part of the equation is the conditional probability that the last parent is type i given a configuration ξ.



The collection of $\bullet$ pairs represents different types in $[d]$.

Dual generator approach

Generator test functions

$$g_\xi(z) = \prod_{i\in[d]} z_{i[\xi_i]}/m_\xi$$

where $a_{[b]} = a(a-1)\ldots a-b+1$, $\xi \in \Delta_N$
and

$$m_\xi = \mathbb{E}^\varphi[f_\xi(Z)].$$

$\mathbb{E}[g_\xi(Z)] = 1$ in the stationary distribution φ of Z.

Rewrite the generator as acting on ξ with respect to the functions g_ξ.

$$\begin{aligned}
\mathcal{L}g_\xi(z) &= \sum_{i\in[d],\gamma\in\Delta_{N-1}} \phi(z; z+|\gamma|e_i-\gamma)[g_\xi(z+|\gamma|e_i-\gamma)-g_\xi(z)] \\
&\quad + \sum_{i,j\in[d]} z_i\mu_{ij}[g_\xi(z-e_i+e_j)-g_\xi(z)] \\
&= \sum_\chi q(\xi,\chi)\left[g_\chi(z)-g_\xi(z)\right],
\end{aligned}$$

where $\{q(\xi,\chi),\xi,\chi\in\Delta_N\}$ is a rate matrix for transitions of ξ.

Sampling distribution

$$g_\xi(z) = \frac{\mathcal{H}(\xi \mid z)}{\mathcal{H}(\xi)},$$

where

$$\mathcal{H}(\xi) = \mathbb{E}^\varphi\big[\mathcal{H}(\xi \mid Z)\big]$$

is the unconditional sampling distribution of a sample of $|\xi|$ individuals in a stationary distribution φ.

$$\widetilde{\mathcal{H}}(z \mid \xi) = \frac{\mathcal{H}(\xi \mid z)}{\mathcal{H}(\xi)}\varphi(z)$$

is the posterior distribution of the frequency of types in a stationary population, conditional on a sample configuration of ξ 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 ξ.

A Markov process $\{L(t), t \geq 0\}$ in Δ_N with rate matrix Q is dual to $\{Z(t), t \geq 0\}$ with duality equation

$$\mathbb{E}_{Z(0)}\Big[\frac{\mathcal{H}\big(L(0) \mid Z(t)\big)}{\mathcal{H}\big(L(0)\big)}\Big] = \mathbb{E}_{L(0)}\Big[\frac{\mathcal{H}\big(L(t) \mid Z(0)\big)}{\mathcal{H}\big(L(t)\big)}\Big]$$

where expectation on the left is with respect to the distribution of $Z(t)$ and on the right with respect to the distribution of $L(t)$.

Neutral Λ-Fleming-Viot process

Generator

$$\mathcal{L}g(x) = \int_0^1 \sum_{i=1}^{d} x_i\big(g(x(1-y)+ye_i) - g(x)\big)\frac{F(dy)}{y^2} + (\theta/2d)\sum_{i=1}^{d}(1-dx_i)\frac{\partial}{\partial x_i}g(x)$$

Local Jumps

$$x \to x(1-y)+ye_i \text{ at rate } x_i\frac{F(dy)}{y^2}$$

Λ-Fleming-Viot process with viability selection

Cannings model $\{r_a\}_{a\in\mathbb{N}}$
FV process - a probability measure F on $[0,1]$. We assume that $F(0)=0$.

Cannings model viability functions v_{iab}
FV process - substochastic measures $V_i(x,\cdot)$ supported on $[0,x]$ with no atoms at zero.

Offspring measures $\{G_i(\cdot); i\in[d]\}$ have a decomposition

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

Measuring selective measures relative to a neutral measure F

Define signed measures

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

Assume that $|K([0,1])| < \infty$ and set $\sigma_i = K([0,1])$.

Λ-Fleming-Viot process with viability selection

Generator

$$\mathcal{L}_\Lambda g(x) = \int_{[0,1]} \sum_{i \in [d]} x_i \left(g(x(1-y) + y e_i) - g(x) \right) \frac{G_i(dy)}{y^2} + \sum_{i,j \in [d]} \left(x_j \mu_{ji} - x_i \mu_{ij} \right) \frac{\partial}{\partial x_i} g(x)$$

Two types, $x = x_1$

$$\begin{aligned}\mathcal{L}_\Lambda g(x) &= \int_{[0,1]} \Big[x \left(g(x(1-y)+y) - g(x) \right) \frac{G_1(dy)}{y^2} \\ &\qquad + (1-x) \left(g(x(1-y)) - g(x) \right) \frac{G_2(dy)}{y^2} \Big] \\ &\qquad + \frac{1}{2} \left(-ax + b(1-x) \right) \frac{\partial}{\partial x} g(x)\end{aligned}$$

Local Jumps

$x \to x(1-y)+y$ at rate $xG_1(dy)/y^2$

$x \to x(1-y)$ at rate $(1-x)G_2(dy)/y^2$

General generator form for $\mathcal{L}_\Lambda g(x)$

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

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

$$\int_{[0,1]} - \sum_{i\in[d]} x_i\Big(g(x(1-y)+ye_i) - g(x)$$

$$-y \sum_{j\in[d]} (\delta_{ij} - x_j)\frac{\partial}{\partial x_j} g(x)\Big)\frac{K_i(dy)}{y}$$

$$- \sum_{j\in[d]} x_j\Big(\sigma_j - \sum_{k\in[d]} \sigma_k x_k\Big)\frac{\partial}{\partial x_j} g(x) + \sum_{i,j\in[d]} \Big(x_i\mu_{ij} - x_j\mu_{ji}\Big)\frac{\partial}{\partial x_j} g(x)$$

Dual Lambda coalescent rates $(l \geq 2)$

$$
\begin{aligned}
q_\Lambda(\xi, \xi - e_i(l-1)) &= \int_{[0,1]} \binom{|\xi|}{l} y^l (1-y)^{|\xi|-l} \frac{G_i(dy)}{y^2} \\
&\quad \times \frac{\xi_i + 1 - l}{|\xi| + 1 - l} \cdot \frac{\mathcal{M}(\xi - e_i(l-1))}{\mathcal{M}(\xi)} \\
q_\Lambda(\xi, \xi + e_i) &= \int_{[0,1]} \left(1 - (1-x)^{|\xi|}\right) \frac{K_i(dx)}{x} \\
&\quad \times \frac{\xi_i + 1}{|\xi| + 1} \cdot \frac{\mathcal{M}(\xi + e_i)}{\mathcal{M}(\xi)} \\
q_\Lambda(\xi, \xi + e_i - e_j) &= \mu_{ij}(\xi_i + 1 - \delta_{ij}) \cdot \frac{\mathcal{M}(\xi + e_i - e_j)}{\mathcal{M}(\xi)}
\end{aligned}
$$

In a viability model $\int_{[0,1]} \left(1 - (1-x)^{|\xi|}\right) \frac{K_i(dx)}{x} \geq 0$

Weak selection

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

If $|\sigma_i| < \infty$ for $i \in [d]$ without loss of generality take $\sigma_i \geq 0$ for all $i \in [d]$ because of shift invariance under the (forward) distribution.

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

$$K_i(\cdot) = \sigma_i \delta_\epsilon(\cdot)$$

Generator in a model with weak selection

$$
\begin{aligned}
\mathcal{L}^{\sigma}_{\Lambda} g(x) \;=\; & \int_{(0,1]} \sum_{i\in[d]} x_i \left(g(x(1-y)+ye_i) - g(x)\right) \frac{F(dy)}{y^2} \\
& - \sum_{j\in[d]} x_j \left(\sigma_j - \sum_{k\in[d]} \sigma_k x_k\right) \frac{\partial}{\partial x_j} g(x) \\
& + \sum_{i,j\in[d]} \left(x_i \mu_{ij} - x_j \mu_{ji}\right) \frac{\partial}{\partial x_j} g(x).
\end{aligned}
$$

Eldon-Wakeley example

$$r_1 = 1 - \frac{1}{N\gamma}, r_{\lfloor N\psi \rfloor - 1]} = \frac{1}{N\gamma}$$

Simple selective viability distributions.

$$\begin{aligned} v_{i11} &= 1 \\ v_{i\lfloor N\psi - 1\rfloor k} &= \begin{cases} 1 - \beta_i & \text{if } k = 1 \\ \beta_i & \text{if } k = \lfloor N\psi_i - 1 \rfloor \end{cases} \end{aligned}$$

$\psi_i \leq \psi$, and $0 \leq \beta_i \leq 1$ for $i \in [d]$

No selection acts on single offspring and there is a two point distribution if the number of offspring is greater than 1.

$$p_{i,1} = 1 - \frac{\beta_i}{N\gamma},\ p_{i,\lfloor N\psi_i - 1\rfloor} = \frac{\beta_i}{N\gamma}$$

The limit measures in the generator $\mathcal{L}_\Lambda$ and in the dual process rates are:

	F	G_i	K_i
$\gamma > 2$	δ_0	δ_0	Null
$\gamma = 2$	$\frac{2}{2+\psi^2}\delta_0 + \frac{\psi^2}{2+\psi^2}\delta_\psi$	$\frac{2}{2+\psi^2}\delta_0 + \frac{\beta_i\psi_i^2}{2+\psi^2}\delta_{\psi_i}$	$-\frac{\beta_i\psi_i^2}{2+\psi^2}\delta_{\psi_i}$
$\gamma < 2$	δ_ψ	$\frac{\beta_i\psi_i^2}{\psi^2}\delta_{\psi_i}$	$\frac{1}{\psi}\delta_\psi - \frac{\beta_i\psi_i}{\psi^2}\delta_{\psi_i}$

If $\gamma \leq 2$, the measures $V_i(x,\cdot)$ are atomic, with

$$V_i(\psi, \{\psi_i\}) = \frac{\beta_i\psi_i^2}{\psi^2},$$

and if $\gamma = 2$,

$$V_i(0,0) = 1.$$

References

Etheridge, A. M.(2011) Some mathematical models from population genetics. Saint-Flour. Lecture notes in Mathematics v 2012, Springer.

Etheridge, A. M., Griffiths (2010) A Coalescent Dual Process in a Moran model with Genic Selection. Theor. Popul. Biol. 75, 320–330.

Etheridge, A. M., Griffiths, R. C., Taylor, J. E. (2010) A Coalescent Dual Process in a Moran model with Genic Selection, and the Lambda coalescent limit. Theor. Popul. Biol. 78, 77–92.

Donnelly, P., Kurtz, T. G. (1996) A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Appl. Prob.* 24, 698–742.

Donnelly, P.J., Kurtz, T.G. (1999) Particle representations for measure-valued population models. *Ann. Prob.* 24, (1999), 166–205.

Kingman, J.F.C. (1982) The coalescent. *Stochastic Proc. Appl.* 13, 235–248.

Pitman, J. (1999) Coalescents with multiple collisions. *Ann. Probab.* 27, 1870–1902.

Pitman, J. (2006) Combinatorial Stochastic Processes Lecture Notes in Mathematics 1875 Springer-Verlag Berlin Heidelberg

Sagitov, S. (1999) The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* 26, 1116–1125.