THE Λ-FLEMING–VIOT PROCESS AND A CONNECTION WITH WRIGHT–FISHER DIFFUSION

ROBERT C. GRIFFITHS,* *Oxford University*

Abstract

The d-dimensional Λ-Fleming–Viot generator acting on functions $g(x)$, with x being a vector of d allele frequencies, can be written as a Wright–Fisher generator acting on functions g with a modified random linear argument of x induced by partitioning occurring in the Λ-Fleming–Viot process. The eigenvalues and right polynomial eigenvectors are easy to see from this representation. The two-dimensional process, which has a one-dimensional generator, is considered in detail. A nonlinear equation is found for the Green's function. In a model with genic selection a proof is given that there is a critical selection value such that if the selection coefficient is greater than or equal to the critical value then fixation, when the boundary 1 is hit, has probability 1 beginning from any nonzero frequency. This is an analytic proof different from the proofs of Der, Epstein and Plotkin (2011) and Foucart (2013). An application in the infinitely-many-alleles Λ-Fleming–Viot process is finding an interesting identity for the frequency spectrum of alleles that is based on size biasing. The moment dual process in the Fleming–Viot process is the usual Λ-coalescent tree back in time. The Wright–Fisher representation using a different set of polynomials $g_n(x)$ as test functions produces a dual death process which has a similarity to the Kingman coalescent and decreases by units of one. The eigenvalues of the process are analogous to the Jacobi polynomials when expressed in terms of $g_n(x)$, playing the role of x^n. Under the stationary distribution when there is mutation, $\mathbb{E}[g_n(X)]$ is analogous to the nth moment in a beta distribution. There is a d-dimensional version $g_n(X)$, and even an intriguing Ewens' sampling formula analogy when $d \to \infty$.

Keywords: Λ-coalescent Fleming–Viot process; Wright–Fisher diffusion process

2010 Mathematics Subject Classification: Primary 60G99
Secondary 92D15

1. Introduction

The d-dimensional Λ-Fleming–Viot process $\{X_t\}_{t\geq 0}$ representing frequencies of d types of individuals in a population has state space $\Delta = \{x \in [0,1]^d : \sum_{i\in[d]} x_i \leq 1\}$ with generator $\mathcal{L}$ acting on functions in $C^2(\Delta)$ described by

$$\mathcal{L}g(x) = \int_0^1 \sum_{i=1}^d x_i (g(x(1-y) + ye_i) - g(x)) \frac{\Lambda(dy)}{y^2}. \tag{1}$$

In general, Λ is a nonnegative finite measure on $[0,1]$. We take a time scale so that $\Lambda \equiv F$ is a probability measure on $[0,1]$. Informally, the population is partitioned at events of change by choosing type $i \in \{1, 2, \ldots, d\}$ to reproduce with probability x_i, then rescaling the population

with additional offspring y of type i so that the frequencies are $\boldsymbol{x}(1-y)+y\boldsymbol{e}_i$, at rate $y^{-2}F(\mathrm{d}y)$. If F has a single atom at 0 then $\{\boldsymbol{X}_t\}_{t\geq 0}$ is the d-dimensional Wright–Fisher diffusion process in Δ with generator

$$\mathcal{L} = \frac{1}{2}\sum_{i,j=1}^{d} x_i(\delta_{ij}-x_j)\frac{\partial^2}{\partial x_i \partial x_j} \qquad (2)$$

acting on functions in $C^2(\Delta)$. The general process $\{\boldsymbol{X}_t\}_{t\geq 0}$ with generator (1) has a Wright–Fisher diffusive component if $F(0) > 0$, and discontinuous sample paths from jumps where the frequencies are changed by adding mass y from the points of F in $(0, 1]$ to the frequency of a type and rescaling the frequencies. Eventually, the process becomes absorbed into one state in $\{\boldsymbol{e}_i\}_{i=1}^{d}$. Eldon and Wakeley (2006) introduced a model where F has a single point of increase in $(0, 1]$ with a possible atom at 0 as well. A natural class that arises from discrete models are beta coalescents, particularly when F has a Beta$(2-\alpha, \alpha)$ density coming from a discrete model where the offspring distribution tails are asymptotic to a power law of index α. This beta-coalescent model is studied in Schweinsberg (2003) and Birkner *et al.* (2005). Birkner and Blath (2009) describe the Λ-Fleming–Viot process and discrete models whose limit gives rise to the process. Möhle and Sagitov (2001) also consider limits from a discrete population with an exchangeable reproduction structure.

The Λ-coalescent is a random tree back in time which has multiple merger rates for a specific $2 \leq k \leq n$ edges merging from n edges in the tree of

$$\lambda_{nk} = \int_0^1 x^k(1-x)^{n-k}\frac{\Lambda(\mathrm{d}x)}{x^2}, \qquad k \geq 2. \qquad (3)$$

After coalescence there are $n - k + 1$ edges in the tree. The process is often regarded as having a state space on the set of partitions Π_∞ of the positive integers. The leaves of an infinite leaf Λ-coalescent tree at time $t = 0$ are labelled with singleton sets $\{1\}, \{2\}, \ldots$ and edges at time t are labelled by sets in $\Pi_\infty(t)$. The number of blocks at time t is the number of sets in the partition $\Pi_\infty(t)$, denoted by $|\Pi_\infty(t)|$, which is the same as the number of edges in the tree at time t. If there are n edges at time t, and k merge at t^{+0}, then a new partition is formed by taking the union of the k partition blocks in the merger for the parent block at t^{+0}. This occurs at rate λ_{nk}. The Λ-coalescent is said to *come down from infinity* if, for all $t > 0$, $\mathbb{P}(|\Pi_\infty(t)| < \infty) = 1$, which is equivalent to an infinite-leaf Λ-coalescent tree at $t = 0$ having a finite number of edges at any time $t > 0$ back with probability 1.

The Λ-coalescent process was introduced in Donnelly and Kurtz (1999), Pitman (1999), and Sagitov (1999), and has been extensively studied (see Pitman (2006) and Berestycki (2009)). The coalescent process is a moment dual to the Λ-Fleming–Viot process. See, for example, Etheridge (2011). There is a distinction between an untyped coalescent process and a typed process; see Etheridge *et al.* (2010).

There is a connection between continuous-state branching processes and the Λ-coalescent. For example, see Bertoin and Le Gall (2003), (2006), Birkner *et al.* (2005), and Berestycki *et al.* (2014a), (2014b). The connection is through the Laplace exponent

$$\psi(q) = \int_0^1 (\mathrm{e}^{-qy} - 1 + qy)y^{-2}\Lambda(\mathrm{d}y).$$

Bertoin and Le Gall (2006) showed that the Λ-coalescent comes down from infinity under the same condition that the continuous-state branching process becomes extinct in finite time,

that is, when

$$\int_1^\infty \frac{\mathrm{d}q}{\psi(q)} < \infty.\tag{4}$$

Schweinsberg (2000) proved earlier that coming down from infinity was equivalent to

$$\sum_{n=2}^\infty \left[\sum_{k=2}^n (k-1)\binom{n}{k}\lambda_{nk}\right]^{-1} < \infty.$$

In this paper we express the Λ-Fleming–Viot generator acting on functions as a Wright–Fisher diffusion generator where the argument of the function is replaced by a random linear transformation. For example, if $d = 2$, the generator acting on functions of $x_1 = x$ in $\mathcal{C}^2([0, 1])$ is specified by

$$\mathcal{L}g(x) = \int_0^1 [x(g(x(1-y)+y) - g(x)) + (1-x)(g(x(1-y)) - g(x))]\frac{F(\mathrm{d}y)}{y^2},\tag{5}$$

where $\Lambda = F$, a probability measure. A Wright–Fisher generator equation, identical to (5), is

$$\mathcal{L}g(x) = \tfrac{1}{2}x(1-x)\mathbb{E}[g''(x(1-W)+VW)],\tag{6}$$

where $W = UY$, Y has distribution F, U has a density $2u$, $u \in (0, 1)$, V is uniform on $(0, 1)$, and U, V, and Y are independent. If $W = 0$, the usual Wright–Fisher generator is obtained. Equation (6) is very suggestive of a strong representation between the Λ-Fleming–Viot and Wright–Fisher processes.

The d-dimensional generator has polynomial eigenvectors and eigenvalues which are analogues of those in the Wright–Fisher generator. The eigenvalues are

$$\tfrac{1}{2}n(n-1)\mathbb{E}[(1-W)^{n-2}], \qquad n = 2, 3, \ldots,$$

which are equal to the Λ-coalescent total merger rates from n blocks. If $d = 2$, the polynomial eigenvectors are analogues of the Jacobi polynomials.

The two-dimensional process is considered in detail in this paper. An integral equation is found for the stationary distribution when there is mutation. This leads to an interesting equation for the frequency spectrum in the infinitely-many-alleles Λ-Fleming–Viot model when the Λ-coalescent comes down from infinity. If frequencies of the alleles are denoted by $x_{(1)} \geq x_{(2)} \geq \cdots$ and $\mathbb{E}$ denotes the expectation in the stationary distribution, then the (one-dimensional) frequency spectrum $\beta(x)$ is defined by

$$\mathbb{E}\left[\sum_{k=1}^\infty f(x_{(k)})\right] = \int_0^1 f(x)\beta(x)\,\mathrm{d}x,\tag{7}$$

where $f \in C([0, 1])$ and $f(x)/x$ is bounded as $x \to 0$. The one-dimensional frequency spectrum is the same as the first factorial moment measure for the allele frequencies $\{x_{(i)}\}$ regarded as a point process. Equation (7) follows from general point processes theory (see Daley and Vere-Jones (2003)). From definition (7), it follows that $z\beta(z)$, $0 < z < 1$, is a probability density. Let Z be a random variable with this density, let Z_* be a random variable size biased with respect to Z, let Z^* be a random variable size biased with respect to $1 - Z$, and let V be a uniform random variable on $[0, 1]$. Then

$$VZ_* \stackrel{\mathrm{D}}{=} (1-W)Z^* + VW,\tag{8}$$

where the random variables are independent of each other. The left-hand side is the limit distribution of excess life in a renewal process with increments distributed as Z (see Cox (1970)), so the equation suggests a renewal process. We do not have a probabilistic solution of (8) which would possibly lead to knowing $\beta(z)$.

In a two-dimensional process with no mutation and genic selection a proof is given that there is a critical selection value such that if the selection coefficient is greater than or equal to the critical value then fixation, when the boundary 1 is hit, has probability 1 beginning from any nonzero frequency. This is an analytic proof different from the proofs of Der *et al.* (2011) and Foucart (2013) which uses our particular representation of the generator. A computational solution for the probability of fixation, when fixation is not certain, is found which is analogous to that in the Wright–Fisher model. Bah and Pardoux (2013) constructed a lookdown process (see Donnelly and Kurtz (1996)) in this model.

The moment dual process in the Fleming–Viot process is the usual Λ-coalescent back in time. In a model with two types, generator (5), and $X(t)$ the frequency of the first type at time t we have the dual equation

$$\mathbb{E}_{X(0)=x}[X(t)^n] = \mathbb{E}_{L(0)=n}[x^{L(t)}].$$

In this equation $\{L(t)\}_{t \geq 0}$ is a Λ-coalescent process back in time with transition rates λ_{nk}. The expectation on the left is with respect to $X(t)$, and on the right with respect to $L(t)$.

In the Wright–Fisher representation using a different set of polynomials $g_n(x)$ which mimic x^n in the usual Wright–Fisher diffusion as test functions produces a dual death process which has a similarity to the Kingman coalescent and decreases by units of one. The d-dimensional version $g_n(x)$ analogous to x^n has an expectation in the stationary distribution of a model with parent-independent mutation that is similar to the Dirichlet moment

$$\mathbb{E}[g_n(x)] = \frac{\prod_{i=1}^d [\prod_{j=1}^{n_i} ((j-1)\mathbb{E}[(1-W)^{j-2}] + \theta_i)]}{\prod_{j=1}^n ((j-1)\mathbb{E}[(1-W)^{j-2}] + \theta)}.$$

Bold face notation will be used for d-dimension vectors in the paper, and the shorthand notation $x^n \equiv \prod_{i=1}^d x_i^{n_i}$. There is even an analogue of Ewens' sampling formula in the Poisson Dirichlet process of

$$\frac{n! \theta^k}{n_1 \cdots n_k} \cdot \frac{\prod_{i=1}^k [\prod_{j=2}^{n_i} \mathbb{E}[(1-W)^{j-2}]]}{\prod_{j=1}^n ((j-1)\mathbb{E}[(1-W)^{j-2}] + \theta)}.$$

There are many intriguing analogues between the Λ-Fleming–Viot process and the Wright–Fisher diffusion process which come from the generator representation.

Exact calculations are always likely to be difficult because of the jump process nature of the Λ-Fleming–Viot process. A first step in this direction, for certain classes of Fleming–Viot processes where stationary distributions are characterized, can be found in Handa (2014).

2. A Wright–Fisher generator connection

The Λ-Fleming–Viot generator has an interesting connection with a Wright–Fisher diffusion generator that we now develop.

Theorem 1. *Let $\mathcal{L}$ be the Λ-Fleming–Viot generator (1), let V be a uniform random variable on $[0, 1]$, let U be a random variable on $[0, 1]$ with density $2u$, $0 < u < 1$, and let $W = YU$,*

where Y has distribution F and V, U, and Y are independent. Denote the first and second derivatives of a function g(x) in $C^2(\Delta)$ by

$$g_i(\mathbf{x}) = \frac{\partial}{\partial x_i} g(\mathbf{x}), \qquad g_{ij}(\mathbf{x}) = \frac{\partial^2}{\partial x_i \partial x_j} g(\mathbf{x}).$$

Then

$$\mathcal{L}g(\mathbf{x}) = \frac{1}{2} \sum_{i,j=1}^{d} x_i(\delta_{ij} - x_j)\mathbb{E}[g_{ij}(\mathbf{x}(1-W) + WV\mathbf{e}_i)], \tag{9}$$

where the expectation $\mathbb{E}$ is taken over V and W.

Proof. Taking the expectation with respect to V, the right-hand side of (9) is equal to

$$\frac{1}{2} \sum_{i,j=1}^{d} x_i(\delta_{ij} - x_j)\mathbb{E}\left[\frac{g_j(\mathbf{x}(1-W) + W\mathbf{e}_i) - g_j(\mathbf{x}(1-W))}{W}\right]$$

$$= \int_0^1 \left[\sum_{i,j=1}^{d} x_i(\delta_{ij} - x_j)\int_0^1 g_j(\mathbf{x}(1-uy) + uy\mathbf{e}_i)\,\mathrm{d}u\right]\frac{F(\mathrm{d}y)}{y}$$

$$- \int_0^1 \left[\sum_{i,j=1}^{d} x_i(\delta_{ij} - x_j)\int_0^1 g_j(\mathbf{x}(1-uy))\,\mathrm{d}u\right]\frac{F(\mathrm{d}y)}{y}. \tag{10}$$

To simplify (10), note that

$$\frac{\partial}{\partial u} g(\mathbf{x}(1-uy)) = -y \sum_{j=1}^{d} x_j g_j(\mathbf{x}(1-uy)),$$

$$\frac{\partial}{\partial u} g(\mathbf{x}(1-uy) + uy\mathbf{e}_i) = y \sum_{j=1}^{d} (\delta_{ij} - x_j) g_j(\mathbf{x}(1-uy) + uy\mathbf{e}_i).$$

Therefore, (10) is equal to

$$\int_0^1 \left[\sum_{i=1}^{d} x_i \int_0^1 \frac{\partial}{\partial u} g(\mathbf{x}(1-uy) + uy\mathbf{e}_i)\,\mathrm{d}u\right]\frac{F(\mathrm{d}y)}{y^2} \tag{11}$$

$$- \int_0^1 \left[\int_0^1 \left(1 - \sum_{i=1}^{d} x_i\right)\frac{\partial}{\partial u} g(\mathbf{x}(1-uy))\,\mathrm{d}u\right]\frac{F(\mathrm{d}y)}{y^2} \tag{12}$$

$$= \int_0^1 \left[\sum_{i=1}^{d} x_i(g(\mathbf{x}(1-y) + y\mathbf{e}_i) - g(\mathbf{x}))\right]\frac{F(\mathrm{d}y)}{y^2}. \tag{13}$$

In the calculation the term (11) is equal to (13), and the term (12) vanishes.

The Wright–Fisher generator (2) is included in (9) when $W \equiv 0$.

Corollary 1. $\{X_1(t)\}_{t\geq 0}$ *is a Markov process with generator acting on functions in $C^2([0,1])$ specified by*

$$\mathcal{L}g(x) = \tfrac{1}{2}x(1-x)\mathbb{E}[g''(x(1-W) + WV)]. \tag{14}$$

Proof. Let $g(\boldsymbol{x})$ in (9) be a function of the first coordinate only. Then (14) follows easily, with $x \equiv x_1$.

The random variable W possibly has an atom at 0, $\mathbb{P}(W = 0) = \mathbb{P}(Y = 0)$, and is continuous for $W > 0$ with a density

$$f_W(w) = 2w F^+(w), \tag{15}$$

where

$$F^+(w) = \int_w^1 y^{-2} F(\mathrm{d}y).$$

There is a correspondence between F and the distribution of W. Given a random variable W with a possible atom at 0 and a density $f_W(w)$, $0 < w \leq 1$, there exists independent random variables U and Y, where U has density $2u$, $0 < u < 1$, such that $W = YU$ if and only if $f_W(1) = 0$ and $f_W(w)/w$ is decreasing in $(0, 1]$. Possible densities for the continuous component of W are proportional to the Beta(a, b) densities with $a \leq 2$ and $b \geq 1$. In particular, if Y has a Beta$(2 - \alpha, \alpha)$ distribution then W has a Beta$(2 - \alpha, 1 + \alpha)$ distribution.

The next theorem gives a connection between W, the Λ-coalescent rates, and the Laplace exponent.

Theorem 2. *The total coalescent rate away from n can be expressed as*

$$\sum_{k=2}^n \binom{n}{k} \lambda_{nk} = \int_0^1 [1 - (1 - y)^n - ny(1 - y)^{n-1}] \frac{F(\mathrm{d}y)}{y^2}$$

$$= \tfrac{1}{2} n(n - 1) \mathbb{E}[(1 - W)^{n-2}] \quad \text{for } n \geq 2.$$

The individual rates (3) can be expressed for $2 \leq k \leq n$ as

$$\binom{n}{k} \lambda_{nk} = \binom{n}{k} \int_0^1 y^k (1 - y)^{n-k} \frac{F(\mathrm{d}y)}{y^2} = \frac{n}{2} \mathbb{E}[P_{k-1}(n, W) - P_k(n, W)], \tag{16}$$

where

$$P_k(n, w) = \binom{n-1}{k} (1 - w)^{n-k-1} w^{k-1}.$$

The Laplace exponent

$$\psi(q) = \frac{q}{2} \mathbb{E}\left[\frac{1 - \mathrm{e}^{-qW}}{W} \right]. \tag{17}$$

Proof. We have

$$\frac{1}{2} n(n - 1) \mathbb{E}[(1 - W)^{n-2}] = \frac{1}{2} n(n - 1) \int_0^1 \int_0^1 (1 - uy)^{n-2} 2u \, \mathrm{d}u \, F(\mathrm{d}y)$$

$$= \int_0^1 \int_0^1 u \frac{\partial^2}{\partial u^2} (1 - uy)^n \, \mathrm{d}u \frac{F(\mathrm{d}y)}{y^2}$$

$$= \int_0^1 [1 - (1 - y)^n - ny(1 - y)^{n-1}] \frac{F(\mathrm{d}y)}{y^2}$$

$$= \sum_{k=2}^n \binom{n}{k} \lambda_{nk}.$$

For the individual rates, showing (16) is an exercise in integration by parts which follows from

$$\frac{n}{2}\mathbb{E}[P_k(n, W)] = \frac{n}{2}\binom{n-1}{k}\int_0^1\int_0^1 (1-uy)^{n-k-1}(uy)^{k-1}2u\,\mathrm{d}u\,F(\mathrm{d}y)$$

$$= -n\binom{n-1}{k}(n-k)^{-1}\int_0^1 y^{k-2}\int_0^1 u^k\frac{\partial(1-uy)^{n-k}}{\partial u}\,\mathrm{d}u\,F(\mathrm{d}y)$$

$$= -\binom{n}{k}\lambda_{nk} + \frac{n}{2}\mathbb{E}[P_{k-1}(n, W)].$$

Note that $P_n(n, w) \equiv 0$, so $\lambda_{nn} = (n/2)\mathbb{E}[P_{n-1}(n, W)]$. To show (17),

$$\psi(q) = \int_0^1 (\mathrm{e}^{-qy} - 1 + qy)\frac{F(\mathrm{d}y)}{y^2}$$

$$= \int_0^1 \sum_{k=2}^\infty (-1)^k \frac{q^k y^{k-2}}{k!} F(\mathrm{d}y)$$

$$= \frac{1}{2}\int_0^1 \sum_{k=2}^\infty (-1)^k \frac{q^k}{(k-1)!}\int_0^1 (uy)^{k-2}2u\,\mathrm{d}u\,F(\mathrm{d}y)$$

$$= \frac{q}{2}\mathbb{E}\left[\frac{1 - \mathrm{e}^{-qW}}{W}\right].$$

The random variables Y, W, and V from Theorem 1 are used frequently in the paper, so their definition will be assumed.

2.1. Mutation and selection

Mutation can be added to the model by assuming that mutations occur at rate $\theta/2$ and changes of type i to type j are made according to a transition matrix P. This is equivalent to mutations occurring at rate $\theta/2$ on the dual Λ-coalescent tree. The generator (1) then has an additional term added of

$$\frac{\theta}{2}\sum_{i=1}^d\left(\sum_{j=1}^d p_{ji}x_j - x_i\right)\frac{\partial}{\partial x_i}. \tag{18}$$

If mutation is parent independent, $\theta p_{ji} = \theta_i$, not depending on j, and the additional term simplifies to

$$\frac{1}{2}\sum_{i=1}^d (\theta_i - \theta x_i)\frac{\partial}{\partial x_i}.$$

If $d = 2$ and $x_1 = x$, $x_2 = 1 - x$, then the generator acting on functions $g(x)$ in $C^2([0, 1])$ is specified by

$$\mathcal{L}g(x) = \int_0^1 [x(g(x(1-y)+y) - g(x)) + (1-x)(g(x(1-y)) - g(x))]\frac{F(\mathrm{d}y)}{y^2}$$

$$+ \frac{1}{2}(\theta_1 - \theta x)g'(x). \tag{19}$$

Handa (2014) finds the stationary distribution in a process with generator specified by

$$
\mathcal{L}_H g(x) = \int_0^1 [x(g(x(1 - y) + y) - g(x)) + (1 - x)(g(x(1 - y)) - g(x))] \frac{B_{1-\alpha, 1+\alpha}(\mathrm{d}y)}{y^2}
$$
$$
+ \int_0^1 [\theta_1 g(x(1 - y) + y) + \theta_2 g(x(1 - y)) - \theta g(x)] \frac{B_{1-\alpha, \alpha}(\mathrm{d}y)}{(\alpha + 1)y}, \tag{20}
$$

where $0 < \alpha < 1$ and $B_{a,b}(\mathrm{d}y)$ denotes a Beta(a, b) density. In his model there is simultaneous mutation, where, at rate $\theta_1 B_{1-\alpha, \alpha}(\mathrm{d}y)/((\alpha + 1)y)$, a proportion y of the population is replaced by type-1 individuals, and similarly, at rate $\theta_2 B_{1-\alpha, \alpha}(\mathrm{d}y)/((\alpha + 1)y)$, a proportion y of the population is replaced by type-2 individuals. This is an unusual mutation mechanism and the generators (19) and (20) are different even when $F = B_{1-\alpha, 1+\alpha}$.

Etheridge *et al.* (2010) studied a Λ-Fleming–Viot process with viability selection whose generator acting on functions in $C^2(\Delta)$ takes the form

$$
\mathcal{L}g(x) = \int_0^1 \sum_{i=1}^d x_i (g(\boldsymbol{x}(1 - y) + y\boldsymbol{e}_i) - g(\boldsymbol{x})) \frac{F(\mathrm{d}y)}{y^2}
$$
$$
- \int_0^1 \sum_{i=1}^d x_i (g(\boldsymbol{x}(1 - y) + y\boldsymbol{e}_i) - g(\boldsymbol{x})) \frac{K_i(\mathrm{d}y)}{y}
$$
$$
+ \frac{\theta}{2} \sum_{i=1}^d \left(\sum_{j=1}^d p_{ji} x_j - x_i \right) \frac{\partial}{\partial x_i} g(\boldsymbol{x}). \tag{21}
$$

To describe the measures in (21), let G_i, $i \in [d]$, be the Λ-measures for the individual types, which are positive measures on $[0, 1]$, and let F be a reference measure such that

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

are bounded signed measures on $[0, 1]$. A selection model analogous to the Wright–Fisher model with genic selection (see, for example, Ewens (2004)) is obtained by taking

$$
K_i(\cdot) = \sigma_i \delta_\varepsilon(\cdot),
$$

and letting $\varepsilon \to 0+$. Selection is very weak in this limit in the sense that a limit is taken where all the measures approach F, whereas there is a much larger effect when the measures G_i are different. The corresponding sequence of generators converges to

$$
\mathcal{L}^\sigma g(x) = \int_0^1 \sum_{i=1}^d x_i (g(\boldsymbol{x}(1 - y) + y\boldsymbol{e}_i) - g(\boldsymbol{x})) \frac{F(\mathrm{d}y)}{y^2} - \sum_{i=1}^d x_i \left(\sigma_i - \sum_{k=1}^d \sigma_k x_k \right) \frac{\partial}{\partial x_i} g(\boldsymbol{x})
$$
$$
+ \frac{\theta}{2} \sum_{i=1}^d \left(\sum_{j=1}^d p_{ji} x_j - x_i \right) \frac{\partial}{\partial x_i} g(\boldsymbol{x}). \tag{22}
$$

Etheridge *et al.* (2010) found the dual lambda coalescent corresponding to (21) and (22).

2.1.1. *Fixation probability with selection when there are $d = 2$ types.* If there are $d = 2$ types, no mutation, $X = X_1, \sigma_1 \leq 0$, and $\sigma_2 = 0$, then, with the notation $\beta = -\sigma_1 \geq 0$, the generator equation (22) reduces to

$$\mathcal{L}^\beta g(x) = \tfrac{1}{2}x(1 - x)\mathbb{E}[g''(x(1 - W) + WV)] + \beta x(1 - x)g'(x).$$

Let $P(x)$ be the probability that the first type fixes, starting from an initial frequency of x. Then $P(0) = 0$, $P(1) = 1$, and $P(x)$ is the solution of

$$\mathcal{L}^\beta P(x) = 0.$$

That is,

$$\mathbb{E}[P''(x(1 - W) + WV)] + 2\beta P'(x) = 0, \tag{23}$$

and taking the expectation with respect to V,

$$\mathbb{E}\left[\frac{P'(x(1 - W) + W) - P'(x(1 - W))}{W}\right] + 2\beta P'(x) = 0. \tag{24}$$

Integrating and taking care of a possible discontinuity $P(0+)$ at $x = 0$,

$$\mathbb{E}\left[\frac{P(x(1 - W) + W) - P(x(1 - W)) - P(W) + P(0+)}{W(1 - W)}\right] + 2\beta[P(x) - P(0+)]$$
$$= 0. \tag{25}$$

Alison Etheridge and Jay Taylor have obtained equivalent formulae to (24) and (25) for the beta coalescent using integration by parts (private communication (2008)). Der *et al.* (2011), (2012) studied fixation probabilities in the Λ-coalescent. An interesting feature is that, for some Λ-measures and β, it can happen that $P(x) = 1$ or $P(x) = 0$ for all $x \in (0, 1)$. They showed that fixation is certain (that is, $P(x) = 1$, $x \in (0, 1]$) if and only if

$$\beta \geq \beta^* = -\int_0^1 \frac{\log(1 - y)}{y^2} F(\mathrm{d}y) \tag{26}$$

under the assumption that $\beta^* < \infty$. If $\beta^* = \infty$ then fixation is not certain. Their proof is for the Eldon–Wakeley coalescent where F has a single point of increase in $(0, 1]$. The general formula (26) is mentioned in the paper and has an analogous proof to the Eldon–Wakeley case (private communication (2013)). They used a clever comparison of $P(x)$ with subharmonic and superharmonic functions. If $u(x)$ is such that $u(0) = 0$ and $u(1) = 1$, then, if $\mathcal{L}^\beta u(x) \leq 0$ for all $x \in (0, 1)$, they showed that $P(x) \leq u(x)$ for all $x \in (0, 1)$. Similarly, if $\mathcal{L}^\beta u(x) \geq 0$ for all $x \in (0, 1)$, $P(x) \geq u(x)$ for all $x \in (0, 1)$. Comparison functions used are $u(x) = x^p$, and $u(x) = Cx^p + (1 - C)x$, $0 < p < 1$ and $C > 1$. Foucart (2013) gave an elegant martingale proof based on a dual process showing that (26) is necessary and sufficient for $P(x) = 1$, $x \in (0, 1]$, but did not include the critical case when $\beta = \beta^*$ in his proof. Another

way to express (26) is

$$2\beta \geq 2\beta^* = \mathbb{E}\left[\frac{1}{W(1-W)}\right],$$

$$\frac{1}{2}\mathbb{E}[\frac{1}{W(1-W)}] = \frac{1}{2}\int_0^1\int_0^1 \frac{1}{uy(1-uy)}2u\,\mathrm{d}u\,F(\mathrm{d}y)$$

$$= \int_0^1\int_0^1 \frac{\mathrm{d}u}{1-uy}\frac{F(\mathrm{d}y)}{y}$$

$$= \int_0^1 \frac{-\log(1-y)}{y^2}F(\mathrm{d}y).$$

For interest, we show how our representation can be used to give a proof when $\beta^* < \infty$.

Theorem 3. (Der *et al.* (2011), (2012) and Foucart (2013).) *Let $\beta^* < \infty$. Then $P(x) = 1$ for all $x \in (0, 1]$ if and only $\beta \geq \beta^*$.*

Proof. If. Let $\beta = \beta^*$. For $x \in (0, 1]$, from (25),

$$0 = \mathbb{E}\left[\frac{P(x(1-W)+W) - P(x(1-W)) - P(W) + P(x)}{W(1-W)}\right], \qquad (27)$$

where $P(x)$ is a nondecreasing function of x, and since the right-hand side of (27) must be 0, with probability 1,

$$P(x(1-W)+W) - P(W) = 0 \quad \text{and} \quad P(x) - P(x(1-W)) = 0.$$

This can only be true if $P(x) = 1$ for all $x \in (0, 1]$, since $P(1) = 1$. Now take $\beta \geq \beta^*$. For fixed x, $P_\beta(x) \equiv P(x)$ is a nondecreasing function of β because a higher selective parameter produces a higher probability of fixation. Thus, $P_\beta(x) \geq P_{\beta^*}(x) = 1$ for all $x \in (0, 1]$ and it must be that $P_\beta(x) = 1$.

Only if. Let $\beta < \beta^* < \infty$, and suppose that $P(x) = 1$ for $x \in (0, 1]$. We show that this assumption is contradictory. Consider the test function

$$v(x) = \log(x) + K(1-x),$$

where $K > 0$ is a constant. A generator equation is

$$\mathbb{E}_x[v(X(t))] - v(x) = \int_0^t \mathbb{E}_x[\mathcal{L}^\beta v(X(u))]\,\mathrm{d}u. \qquad (28)$$

Equation (28) evaluates to

$$\mathbb{E}_x[v(X(t))] - v(x) = \frac{1}{2}\int_0^t \mathbb{E}_x[(1-X(u))A(u)]\,\mathrm{d}u, \qquad (29)$$

where

$$A(u) = \mathbb{E}\left[X(u)\frac{(X(u)(1-W)+W)^{-1} - (X(u)(1-W))^{-1}}{W}\right.$$

$$\left. + X(u)2\beta X(u)^{-1} - 2K\beta X(u)\right]$$

$$= \mathbb{E}\left[-\frac{1}{(X(u)(1-W)+W)(1-W)} - 2K\beta X(u)\right] + 2\beta.$$

Choose K large enough so that the minimum value over $x \in [0, 1]$ of

$$\mathbb{E}\left[\frac{1}{(x(1 - W) + W)(1 - W)} + 2K\beta x\right]$$

is attained when $x = 0$. Then $A(u) \leq -2\beta^* + 2\beta < 0$. Let $t \to \infty$ in (29). With probability 1, $X(t) \to 1$ so, $\mathbb{E}_x[\log(X(t)) + K(1 - X(t))] \to 0$ and the limit equation is

$$-\log x - K(1 - x) = \frac{1}{2} \int_0^\infty \mathbb{E}_x[(1 - X(u))A(u)] \, du$$

$$\leq (\beta - \beta^*) \int_0^\infty \mathbb{E}_x[(1 - X(u))] \, du$$

$$< 0. \tag{30}$$

Choose x small enough so that the left-hand side of (30) is positive. Then the signs of both sides of (30) are contradictory. Therefore, the assumption that $P(x) = 1$ for all $x \in (0, 1]$ is contradictory. Let x_0 be the maximal point where $P(x) < 1$ for $0 < x < 1$. It cannot happen that $P(x) = 1$ for $x_0 \leq x < 1$. Suppose that this does occur. Let $X(0) = x_0$, and consider the local exit behaviour of X in $(x_0 - \varepsilon, x_0 + \varepsilon)$ for small $\varepsilon > 0$. For small enough ε, there is positive probability that there is a path where X first exits the interval at less than or equal to $x_0 - \varepsilon$. The strong Markov property of X then implies that $P(x_0 - \varepsilon) = 1$, which is contradictory. Therefore, $P(x) < 1$ for all $x \in [0, 1)$.

In the Kingman coalescent (23) becomes

$$P''(x) + 2\beta P'(x) = 0,$$

with a solution

$$P(x) = \frac{1 - e^{-2\beta x}}{1 - e^{-2\beta}}. \tag{31}$$

We provide a computational solution for $P(x)$ in the Λ-Fleming–Viot model that imitates (31) when fixation or loss is not certain from $x \in (0, 1)$. A sequence of polynomials $\{h_n(x)\}_{n=0}^\infty$ that is used in the proof is defined as the solutions of

$$\mathbb{E}\left[\frac{h_n(x(1 - W) + W) - h_n(x(1 - W))}{W}\right] = nh_{n-1}(x), \tag{32}$$

where the leading coefficient in $h_n(x)$ is

$$\frac{1}{\prod_{j=1}^{n-1} \mathbb{E}[(1 - W)^j]}. \tag{33}$$

This choice makes the coefficients of x^{n-1} in (32) agree. The argument in the expectation in (32) is interpreted as $h_n'(x)$ at $W = 0$. There is a family of polynomial solutions to (32) depending on an arbitrary recursive choice of constant coefficients. The constant coefficients in the polynomials are chosen carefully to obtain a solution for the fixation probability. The polynomials $h_n(x)$ imitate x^n and are equal if $W \equiv 0$. Let $h_0(x) = 1$ and

$$h_n(x) = \sum_{r=0}^n a_{nr} x^r.$$

Then, from (32) for $j = n - 2, \ldots, 0$,

$$\sum_{j=0}^{n-1} \sum_{r=j+1}^{n} \binom{r}{j} \mathbb{E}[(1 - W)^j W^{r-j-1}] a_{nr} x^j = n \sum_{j=0}^{n-1} a_{n-1 j} x^j;$$

so equating coefficients of x^j on both sides,

$$\sum_{r=j+1}^{n} \binom{r}{j} \mathbb{E}[(1 - W)^j W^{r-j-1}] a_{nr} = n a_{n-1 j}. \tag{34}$$

Given the coefficients $\{a_{n-1 j}\}_{j=0}^{n-1}$ of $h_{n-1}(x)$, the coefficients of $h_n(x)$ and $\{a_{nj}\}_{j=1}^{n}$ are recursively determined by choosing a_{nn} from (33), then taking $j = n - 1, \ldots, 0$ in (34). There is an arbitrary choice of a_{n0} that needs to be made at this stage to progress with the recursion.

Theorem 4. *Let $0 < \beta < \beta^*$. The fixation probability*

$$P(x) = (1 - e^{-2\beta})^{-1} \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(2\beta)^n}{n!} H_n(x),$$

where $\{H_n(x)\}$ are polynomials derived from

$$H_n(x) = \int_0^x n h_{n-1}(\xi) \, d\xi$$

with the constants $\{h_n(0)\}$ chosen so that

$$\int_0^1 n h_{n-1}(\xi) \, d\xi = 1. \tag{35}$$

   Proof. Try the series solution

$$P'(x) = B(\beta) \sum_{n=1}^{\infty} (-1)^{n-1} (2\beta)^n c_n h_{n-1}(x), \tag{36}$$

where $\{h_n(x)\}$ satisfies (32), $B(\beta)$ is a constant, and $\{c_n\}$ are constants not depending on β. Then, substituting into (24),

$$\sum_{n=2}^{\infty} (-1)^{n-1} (2\beta)^n c_n (n - 1) h_{n-2}(x) + 2\beta \sum_{n=1}^{\infty} (-1)^{n-1} (2\beta)^n c_n h_{n-1}(x) = 0.$$

This identity is satisfied if $c_1 = -1$, without loss of generality, and

$$c_n = -\frac{1}{(n - 1)!}, \qquad n = 2, 3, \ldots.$$

Integrating (36),

$$P(x) = B(\beta) \sum_{n=1}^{\infty} (-1)^{n-1} \frac{(2\beta)^n}{n!} \int_0^x n h_{n-1}(\xi) \, d\xi.$$

Choosing (35) to hold and knowing $P(1) = 1$ shows that

$$B(\beta) = (1 - e^{-2\beta})^{-1}.$$

Corollary 2. *A computational solution for $P(x)$ is found from evaluating the polynomials*

$$H_n(x) = \sum_{r=1}^{n} b_{nr} x^r,$$

where $H_1(x) = x$ and the coefficients $\{b_{nr}\}$ are defined recursively from

$$\sum_{r=j}^{n} \binom{r}{j-1} \mathbb{E}[(1-W)^{j-1} W^{r-j}](r+1) b_{n+1\,r+1} = (n+1) j b_{nj} \tag{37}$$

with

$$b_{n+1\,1} = 1 - \sum_{j=2}^{n+1} b_{n+1\,j}$$

for $n = 1, 2, \ldots$ and $j = n-1, \ldots, 1$. Equation (37) is equivalent to

$$2 \sum_{r=j+1}^{n+1} \left[\sum_{k=r-j+1}^{r} \binom{r}{k} \lambda_{rk} \right] b_{n+1\,r} = (n+1) j b_{nj}. \tag{38}$$

Proof. Relating the coefficients of $H_n(x)$ to those of $h_{n-1}(x)$, we obtain

$$b_{nj} = \frac{n}{j} a_{n-1\,j-1}, \quad j = 2, \ldots, n, \quad \text{and} \quad b_{n1} = 1 - \sum_{j=2}^{n} b_{nj}.$$

Substituting into (34) and shifting the index $j \to j+1$ completes the proof of (37). The alternative form (38) is found by noting that

$$\frac{r+1}{2} \mathbb{E}[P_{r-j+1}(r+1, W)] = \frac{r+1}{2} \binom{r}{j-1} \mathbb{E}[(1-W)^{j-1} W^{r-j}]$$

$$= \sum_{k=r-j+2}^{r+1} \binom{r+1}{k} \lambda_{r+1\,k}$$

from (16), substituting, then shifting the index of summation $r \to r+1$.

2.2. Eigenstructure of the Λ-Fleming–Viot process

The generator of the Λ-Fleming–Viot process (9) with mutation term (18),

$$\mathcal{L} g(\boldsymbol{x}) = \frac{1}{2} \sum_{i,j=1}^{d} x_i (\delta_{ij} - x_j) \mathbb{E}[g_{ij}(\boldsymbol{x}(1-W) + W V \boldsymbol{e}_i)]$$

$$+ \frac{\theta}{2} \sum_{i=1}^{d} \left(\sum_{j=1}^{d} p_{ji} x_j - x_i \right) g_i(\boldsymbol{x}), \tag{39}$$

acting on functions in $\mathcal{C}^2(\Delta)$ maps d-dimensional polynomials into polynomials of the same degree, so the right eigenvectors $\{P_{\boldsymbol{n}}(\boldsymbol{x})\}$ with eigenvalues $-\lambda_{\boldsymbol{n}}$ are polynomials of the same degree that satisfy

$$\mathcal{L} P_{\boldsymbol{n}}(\boldsymbol{x}) = -\lambda_{\boldsymbol{n}} P_{\boldsymbol{n}}(\boldsymbol{x}). \tag{40}$$

The index $\boldsymbol{n}$ is $d-1$ dimensional because of the constraint that $\sum_1^d x_j = 1$. The eigenvalues $\lambda_{\boldsymbol{n}}$, given in (41) below, have a linear form in the $d-1$ nonunit eigenvalues of $I - P$ with coefficients $n_1, \ldots, n_{d-1}$, which define $\boldsymbol{n}$.

Theorem 5. *Let $\{\lambda_{\boldsymbol{n}}\}$ and $\{P_{\boldsymbol{n}}(\boldsymbol{x})\}$ be the eigenvalues and right eigenvectors of $\mathcal{L}$, (39), satisfying (40). Denote the $d-1$ eigenvalues of P which have modulus less than 1 by $\{\phi_k\}_{k=1}^{d-1}$, which correspond to eigenvectors that are rows of a $d-1 \times d$ matrix R satisfying*

$$\sum_{i=1}^{d} r_{ki} p_{ji} = \phi_k r_{kj}, \qquad k = 1, \ldots, d-1.$$

Define a $(d-1)$-dimensional vector $\boldsymbol{\xi} = R\boldsymbol{x}$. Then the polynomials $P_{\boldsymbol{n}}(\boldsymbol{x})$ are polynomials in $\boldsymbol{\xi}$ whose only leading term of degree n is $\boldsymbol{\xi^n}$ and

$$\lambda_{\boldsymbol{n}} = \frac{1}{2} n(n-1) \mathbb{E}[(1-W)^{n-2}] + \frac{\theta}{2} \sum_{k=1}^{d-1} (1-\phi_k) n_k. \tag{41}$$

Proof. The second-order derivative term in $\mathcal{L}$ acting on $\boldsymbol{x^m}$ is

$$-\frac{1}{2} \sum_{i,j=1}^{d} x_i x_j \mathbb{E}[(1-W)^{m-2}] m_i (m_j - \delta_{ij}) \boldsymbol{x^{m-e_i-e_j}} + \text{lower-order terms}$$

$$= -\frac{1}{2} m(m-1) \mathbb{E}[(1-W)^{m-2}] \boldsymbol{x^m} + \text{lower-order terms}.$$

Therefore, the same term acting on $\boldsymbol{\xi^n}$ with $m = n$ is

$$-\tfrac{1}{2} n(n-1) \mathbb{E}[(1-W)^{n-2}] \boldsymbol{\xi^n} + \text{lower-order terms in } \boldsymbol{\xi}. \tag{42}$$

The linear differential term acting on $\boldsymbol{\xi^n}$ is

$$\frac{\theta}{2} \sum_{i=1}^{d} \left(\sum_{j=1}^{d} p_{ji} x_j - x_i \right) \frac{\partial}{\partial x_i} \boldsymbol{\xi^n} = \frac{\theta}{2} \sum_{k=1}^{d-1} \sum_{i=1}^{d} \left(\sum_{j=1}^{d} p_{ji} x_j - x_i \right) n_k r_{ki} \boldsymbol{\xi^{n-e_k}}$$

$$= -\frac{\theta}{2} \sum_{k=1}^{d-1} (1-\phi_k) n_k \boldsymbol{\xi^n}. \tag{43}$$

Equations (42) and (43) are enough to complete the proof of (41). Suppose that we have constructed $\{P_{\boldsymbol{m}}(\boldsymbol{x})\}_{\boldsymbol{m}<\boldsymbol{n}}$. Then take

$$P_{\boldsymbol{n}}(\boldsymbol{x}) = \boldsymbol{\xi^n} - \sum_{\{\boldsymbol{m}\,:\,\boldsymbol{m}<\boldsymbol{n}\}} a_{\boldsymbol{nm}} P_{\boldsymbol{n}}(\boldsymbol{x}),$$

where the coefficients are to be determined. We have

$$\mathcal{L} P_{\boldsymbol{n}}(\boldsymbol{x}) = -\lambda_{\boldsymbol{n}} P_{\boldsymbol{n}}(\boldsymbol{x}) + \sum_{\{\boldsymbol{m}\,:\,\boldsymbol{m}<\boldsymbol{n}\}} b_{\boldsymbol{nm}} P_{\boldsymbol{m}}(\boldsymbol{x}) - \sum_{\{\boldsymbol{m}\,:\,\boldsymbol{m}<\boldsymbol{n}\}} a_{\boldsymbol{nm}} \lambda_{\boldsymbol{m}} P_{\boldsymbol{m}}(\boldsymbol{x})$$

for determined constants $b_{\boldsymbol{nm}}$. Choosing $a_{\boldsymbol{nm}} \lambda_{\boldsymbol{m}} = b_{\boldsymbol{nm}}$ completes the construction.

Corollary 3. *The generator (14) with no mutation term has eigenvalues*

$$\lambda_{\boldsymbol{n}} = \lambda_n = \tfrac{1}{2}n(n-1)\mathbb{E}[(1-W)^{n-2}]$$

repeated $\binom{n+d-2}{n}$ *times and eigenfunctions* $\{P_{\boldsymbol{n}}(\boldsymbol{x})\}_{n\geq 2}$.

Corollary 4. *In the parent-independent model of mutation the generator has eigenvalues*

$$\lambda_{\boldsymbol{n}} = \lambda_n = \tfrac{1}{2}n\{(n-1)\mathbb{E}[(1-W)^{n-2}] + \theta\} \tag{44}$$

repeated $\binom{n+d-2}{n}$ *times and eigenfunctions* $\{P_{\boldsymbol{n}}(\boldsymbol{x})\}_{n\geq 1}$.

Proof. The transition matrix P has rows $(\theta_1/\theta, \ldots, \theta_d/\theta)$. The right eigenvectors of P comprise one vector of units with eigenvalue 1, and $d-1$ other vectors such that $\sum_{i=1}^{d} r_{ki}\theta_i/\theta = 0$. Thus, $\phi_k = 0$, $k = 1, \ldots, d-1$, and $\lambda_{\boldsymbol{n}}$ is equal to (44).

In two dimensions the generator is specified by

$$\mathcal{L}g(x) = \tfrac{1}{2}x(1-x)\mathbb{E}[g''(x(1-W) + WV)] + \tfrac{1}{2}(\theta_1 - \theta x)g'(x). \tag{45}$$

The eigenvalues are

$$\lambda_n = \tfrac{1}{2}n\{(n-1)\mathbb{E}[(1-W)^{n-2}] + \theta\}$$

and the eigenvectors are polynomials satisfying

$$\mathcal{L}P_n(x) = -\lambda_n P_n(x), \qquad n \geq 1.$$

The eigenvalues and polynomials do not depend on W and V for $n = 1, 2$. Writing the eigenvalue equation as

$$x(1-x)\mathbb{E}[P_n''(x(1-W) + VW)] + (\theta_1 - \theta x)P_n'(x)$$
$$+ n\{(n-1)\mathbb{E}[(1-W)^{n-2}] + \theta\}P_n(x)$$
$$= 0,$$

there is a similarity to the hypergeometric equation for the Jacobi polynomials which are the eigenvectors when $W \equiv 0$ (see Kimura (1964)). Writing the nth Jacobi polynomial with index parameters (θ_1, θ_2), orthogonal on the beta distribution with the same parameters, as $\widetilde{P}_n^{(\theta_1, \theta_2)}(x) \equiv z$ for ease of notation, the hypergeometric equation is

$$x(1-x)z'' + (\theta_1 - \theta x)z' + n((n-1) + \theta)z = 0;$$

see, for example, Ismail (2005). Usually, the Jacobi polynomials $P_n^{(\alpha_1, \alpha_2)}(x)$ are defined as orthogonal on the weight function

$$(1-x)^{\alpha_1}(1+x)^{\alpha_2}, \qquad -1 < x < 1,$$

so the translation to orthogonal polynomials on the Beta(θ_1, θ_2) distribution is given by

$$\widetilde{P}_n^{(\theta_1, \theta_2)}(x) = P_n^{(\theta_2-1, \theta_1-1)}(2x - 1).$$

2.3. Stationary distributions

If the mutation matrix P is recurrent then there is a stationary distribution for the process with generator (39). The first- and second-order moments do not depend on W because they can be found from the generator equations

$$\mathbb{E}[\mathcal{L}X_i] = 0, \qquad \mathbb{E}[\mathcal{L}X_i X_j] = 0,$$

which do not depend on W as the second derivatives of X_i and $X_i X_j$ are constant.

In particular, for the parent-independent model of mutation, comparing moments with those of the Dirichlet($\boldsymbol{\theta}$) distribution, which is the stationary distribution for the Wright–Fisher diffusion, we have, for $i, j = 0, 1, \ldots, d$ and any F,

$$\mathbb{E}[X_i] = \frac{\theta_i}{\theta} \quad \text{and} \quad \mathbb{E}[X_i X_j] = \frac{\theta_i(\theta_j + \delta_{ij})}{\theta(\theta + 1)},$$

with expectation in the stationary distribution (see, for example, Ewens (1972)). Now consider the simplest case, the stationary distribution in two dimensions when the generator is (45). An interesting recurrence for the moments of X, the frequency of the first allele, is found in terms of size-biased versions of X.

Theorem 6. *Let Z be a random variable with the size-biased distribution of X, let Z_* be a size-biased Z random variable; let Z^* be a size-biased random variable with respect to $1 - Z$; let $W = UY$, where Y has distribution F and U has a density $2u$, $u \in (0, 1)$; let V be a uniform random variable on $(0, 1)$; and let B be a Bernoulli random variable such that $\mathbb{P}(B = 1) = \theta_2/(\theta(\theta_1 + 1))$ with U, V, Y, Z^*, Z_*, and B independent. Then*

$$V Z_* \overset{\mathcal{D}}{=} (1 - B)V Z + B(Z^*(1 - W) + W V). \tag{46}$$

Proof. Let $g(x) = x^{n+2}$. Then, since $\mathbb{E}[\mathcal{L}g(X)] = 0$ with expectation in the stationary distribution,

$$\frac{(n+2)(n+1)}{2}\mathbb{E}[(X(1 - W) + W V)^n X(1 - X)] + \frac{n+2}{2}\mathbb{E}[X^n X(\theta_1 - \theta X)] = 0$$

or

$$\frac{\theta}{n+1}\mathbb{E}[X^n X^2] = \frac{\theta_1}{n+1}\mathbb{E}[X^n X] + \mathbb{E}[(X(1 - W) + W V)^n X(1 - X)]. \tag{47}$$

Let Z be a random variable with the size-biased distribution of X, let Z_* be a size-biased Z random variable, and let Z^* be a size-biased random variable with respect to $1 - Z$. The distribution of Z is reweighted by Z and divided by $\mathbb{E}[Z]$ to obtain the distribution of Z_*, and similarly the distribution is weighted by $1 - Z$ and divided by $\mathbb{E}[1 - Z]$ to obtain the distribution of Z^*. Then, knowing that

$$\mathbb{E}[X] = \frac{\theta_1}{\theta}, \qquad \mathbb{E}[X^2] = \frac{\theta_1(\theta_1 + 1)}{\theta(\theta + 1)}, \qquad \mathbb{E}[X(1 - X)] = \frac{\theta_1(\theta - \theta_1)}{\theta(\theta + 1)},$$

(47) can be written as

$$\mathbb{E}[(V Z_*)^n] = \frac{\theta_1(\theta + 1)}{\theta(\theta_1 + 1)}\mathbb{E}[(V Z)^n] + \frac{\theta - \theta_1}{\theta(\theta_1 + 1)}\mathbb{E}[(Z^*(1 - W) + W V)^n]. \tag{48}$$

Recall that

$$\mathbb{P}(B = 1) = \frac{\theta_2}{\theta(\theta_1 + 1)}.$$

Then (48) implies the distributional identity (46).

This equation may be related to a renewal process, because the distribution of excess life γ_t in a renewal process with increments distributed as Z satisfies

$$\lim_{t \to \infty} \mathbb{P}(\gamma_t > \eta) = \mathbb{P}(V Z_* > \eta) = \int_\eta^1 \frac{\mathbb{P}(Z > z)}{\mathbb{E}[Z]} \, dz,$$

where $\mathbb{E}[Z] = \theta_1/\theta$ (see Cox (1970)).

Identity (46) implies an integral equation for the stationary distribution in the two-dimensional model.

Theorem 7. *Let $f_X(u)$, $0 < u < 1$, be the stationary density in the diffusion process with generator (14), and let $f_W(w)$ be the density of W. Suppose that F has no atom at 0. Then $f_X(u)$ satisfies the integral equations*

$$(\theta_1 - \theta u) f_X(u) = -\int_0^u \frac{1}{u-z} f_W\left(1 - \frac{1-u}{1-z}\right) z(1-z) f_X(z) \, dz$$
$$+ \int_u^1 \frac{1}{z-u} f_W\left(1 - \frac{u}{z}\right) z(1-z) f_X(z) \, dz \qquad (49)$$

and

$$(\theta_2 - \theta(1-u)) f_X(u) = \int_0^u \frac{1}{u-z} f_W\left(1 - \frac{1-u}{1-z}\right) z(1-z) f_X(z) \, dz$$
$$- \int_u^1 \frac{1}{z-u} f_W\left(1 - \frac{u}{z}\right) z(1-z) f_X(z) \, dz. \qquad (50)$$

These equations are equivalent to

$$(\theta_1 - \theta u) f_X(u) = -\int_0^u 2F^+\left(1 - \frac{1-u}{1-z}\right) z f_X(z) \, dz$$
$$+ \int_u^1 2F^+\left(1 - \frac{u}{z}\right)(1-z) f_X(z) \, dz \qquad (51)$$

and

$$(\theta_2 - \theta(1-u)) f_X(u) = \int_0^u 2F^+\left(1 - \frac{1-u}{1-z}\right) z f_X(z) \, dz$$
$$- \int_u^1 2F^+\left(1 - \frac{u}{z}\right)(1-z) f_X(z) \, dz. \qquad (52)$$

Proof. Let the random line $L = Z^*(1 - W) + WV$ as a function of W. The line segment L varies from $\min(Z^*, V)$ to $\max(Z^*, V)$ as W varies. The density of the line L conditional on $(Z^*, V) = (z, v)$ is, for $\min(z, v) < u < \max(z, v)$,

$$f_{L \mid (z,v)}(u) = \frac{1}{|z - v|} f_W\left(\frac{z - u}{z - v}\right)$$

and there is a possible atom

$$\mathbb{P}(L = z \mid (z, v)) = \mathbb{P}(W = 0).$$

Splitting the region into $v < z$ and $v > z$, the unconditional density of L is

$$f_L(u) = \mathbb{P}(W = 0) f_{Z^*}(u) + \int_{0 < v < u < z < 1} \frac{1}{z - v} f_W\left(\frac{z - u}{z - v}\right) f_{Z^*}(z) \, dz \, dv$$

$$+ \int_{0 < z < u < v < 1} \frac{1}{v - z} f_W\left(\frac{u - z}{v - z}\right) f_{Z^*}(z) \, dz \, dv.$$

Changing variables in the integral we obtain

$$f_L(u) = \mathbb{P}(W = 0) f_{Z^*}(u) + \int_0^u \int_{1 - (1-u)/(1-z)}^1 \frac{1}{\xi} f_W(\xi) \, d\xi \, f_{Z^*}(z) \, dz$$

$$+ \int_u^1 \int_{1 - u/z}^1 \frac{1}{\xi} f_W(\xi) \, d\xi \, f_{Z^*}(z) \, dz. \tag{53}$$

The density identity equivalent to identity (46) is therefore

$$f_{VZ_*}(u) = \mathbb{P}(B = 0) f_{VZ}(u) + \mathbb{P}(B = 1) f_L(u). \tag{54}$$

Note that if ζ is a random variable on $(0, 1)$ with density $f_\zeta(y)$ then the density of $V\zeta$, where V is independent of ζ and uniform on $(0, 1)$, is

$$f_{V\zeta}(u) = \int_u^1 y^{-1} f_\zeta(y) \, dy.$$

Therefore, (54) is equivalent to

$$\frac{\int_u^1 y f_X(y) \, dy}{\mathbb{E}[X^2]} = \mathbb{P}(B = 0) \frac{\int_u^1 f_X(y) \, dy}{\mathbb{E}[X]} + \mathbb{P}(B = 1) f_L(u). \tag{55}$$

Differentiating (55), the density $f_X(u)$ satisfies the integral equation

$$u f_X(u) = \frac{\theta_1}{\theta} f_X(u) - \frac{1}{\theta} f'_{\diamond L}(u), \tag{56}$$

where

$$f_{\diamond L}(u) = \mathbb{E}[X(1 - X)] f_L(u).$$

The density $f_{\diamond L}(u)$ is similar to (53) with $f_{Z_*}(z)$ replaced by $z(1 - z) f_X(z)$. When W has no atom at 0, a straightforward calculation gives

$$f'_{\diamond L}(u) = -\int_0^u \frac{1}{u - z} f_W\left(1 - \frac{1 - u}{1 - z}\right) z(1 - z) f_X(z) \, dz$$

$$+ \int_u^1 \frac{1}{z - u} f_W\left(1 - \frac{u}{z}\right) z(1 - z) f_X(z) \, dz. \tag{57}$$

Recalling (15), another form is

$$f'_{\diamond L}(u) = -\int_0^u 2F^+\left(1 - \frac{1 - u}{1 - z}\right) z f_X(z) \, dz$$

$$+ \int_u^1 2F^+\left(1 - \frac{u}{z}\right)(1 - z) f_X(z) \, dz. \tag{58}$$

Considering $1 - X$, a second integral equation is

$$(1 - u) f_X(u) = \frac{\theta_2}{\theta} f_X(u) + \frac{1}{\theta} f'_{\diamond L}(u). \tag{59}$$

Substituting the expression for $f'_{\diamond L}(u)$ given in (57) into (56) and (59) gives (49) and (50). The alternative form (58) gives (51) and (52).

Another approach that imitates the usual way of finding the stationary distribution in a diffusion process is to consider the equation

$$\int_0^1 [\mathcal{L}g(x)] f_X(x) \, dx = 0, \tag{60}$$

where $g(x)$ is a test function in $C^2([0, 1])$. Define $\sigma^2(x) = x(1 - x)$ and $\mu(x) = \theta_1 - \theta x - \sigma x(1 - x)$, and let

$$k(x) = \mathbb{E}[(1 - W)^{-2} g(x(1 - W) + VW)].$$

Equation (60) is equivalent to

$$\int_0^1 \left[\frac{1}{2} \sigma^2(x) \frac{d^2}{dx^2} k(x) + \mu(x) \frac{d}{dx} g(x) \right] f_X(x) \, dx = 0. \tag{61}$$

Integrating by parts in (61) and taking care with boundary conditions gives

$$0 = \int_0^1 \left[k(x) \frac{1}{2} \frac{d^2}{dx^2} [\sigma^2(x) f_X(x)] - g(x) \frac{d}{dx} [\mu(x) f_X(x)] \right] dx$$
$$+ \left[\left[\frac{d}{dx} k(x) \right] \left[\frac{1}{2} \sigma^2(x) f_X(x) \right] - k(x) \frac{d}{dx} \left[\frac{1}{2} \sigma^2(x) f_X(x) \right] + g(x) \mu(x) f_X(x) \right]_0^1.$$

If $W \equiv 0$ then $k(x) = g(x)$ and we can conclude that $f_X(x)$ satisfies the forward equation

$$\frac{1}{2} \frac{d^2}{dx^2} [\sigma^2(x) f_X(x)] - \frac{d}{dx} [\mu(x) f_X(x)] = 0.$$

An equivalent approach seems difficult when $k(x) \neq g(x)$.

2.4. Green's function

Green's function $G(x, \xi)$, whether or not there is mutation and selection, is obtained via a standard approach of solving, for $\gamma(x)$, the differential equation

$$\mathcal{L}\gamma(x) = -g(x), \qquad \gamma(0) = \gamma(1) = 0. \tag{62}$$

Then

$$\gamma(x) = \int_0^1 G(x, \xi) g(\xi) \, d\xi.$$

Consider the model with no selection. Equation (62) is nonlinear and equivalent to

$$\tfrac{1}{2} x(1 - x) \mathbb{E}[\gamma''(x(1 - W) + VW)] + \tfrac{1}{2} (\theta_1 - \theta x) \gamma'(x) = -g(x)$$

or

$$\tfrac{1}{2} x(1 - x) k''(x) + \tfrac{1}{2} (\theta_1 - \theta x) \gamma'(x) = -g(x), \tag{63}$$

where
$$k(x) = \mathbb{E}[(1-W)^{-2}\gamma(x(1-W)+VW)].$$

In the simplest case when $\theta = 0$, (63) becomes
$$k''(x) = -2\frac{g(x)}{x(1-x)}.$$

Taking a standard Green's function approach, with care that $k(0)$ and $k(1)$ are not 0,
$$k(x) = k(0)(1-x) + k(1)x + (1-x)\int_0^x \frac{2g(\eta)}{1-\eta}\,d\eta + x\int_x^1 \frac{2g(\eta)}{\eta}\,d\eta.$$

If $g(x) = 1$, $x \in (0, 1)$, then $\gamma(x)$ is the mean time to absorption at 0 or 1 when $X(0) = x$. There is a nonlinear equation to solve for $\gamma(x)$ from $k(x)$, knowing that
$$k(x) = k(0)(1-x) + k(1)x - 2(1-x)\log(1-x) - 2x\log x.$$

It is possible that $\gamma(x) = \infty$ if the Λ-coalescent does not come down from infinity.

2.5. The frequency spectrum in the infinitely-many-alleles model

We consider the infinitely-many-alleles model as a limit from a d-allele model with $\theta_i = \theta/d$, $i = 1, \ldots, d$. The limit is thought of as a limit from d points $X_1^d, \ldots, X_d^d$ to points of a point process $\{X_i\}_{i=1}^\infty$. The one-dimensional frequency spectrum μ is a nonnegative measure such that, for functions f in $C([0, 1])$ such that $f(x)/x$ is bounded as $x \to 0$, with expectation in the stationary distribution,
$$\mathbb{E}\left[\sum_{i=1}^\infty f(X_i)\right] = \int_0^1 f(x)\mu(dx). \tag{64}$$

There is an assumption that the point process does not have multiple points at any single position for (64) to hold. Symmetry in the d-allele model shows that
$$\int_0^1 f(x)\mu(dx) = \lim_{d\to\infty} d\mathbb{E}[f(X_1)].$$

If the Λ-coalescent does not come down from infinity then there may be an accumulation of points at 0 as $d \to \infty$ and $\int_0^1 x\mu(x)\,dx < 1$. We do not consider this case in the next theorem. The classical Wright–Fisher diffusion gives rise to the Poisson–Dirichlet process with a frequency spectrum of
$$\mu(dx) = \theta x^{-1}(1-x)^{\theta-1}\,dx, \qquad 0 < x < 1.$$

Theorem 8. *Let $\mu(dz)$ be the frequency spectrum measure in an infinitely-many-alleles Λ-Fleming–Viot process which comes down from infinity, and let Z be a random variable with probability measure $z\mu(dz)$. Let Z_* be a random variable with a size-biased distribution of Z, and let Z^* be a random variable with a size-biased distribution of Z with respect to $1 - Z$. The random variable Z_* has a measure $(\theta + 1)z^2\mu(dz)$ and Z^* has a measure $\theta^{-1}(\theta + 1)z(1 - z)\mu(dz)$, $0 < z < 1$. Then*
$$VZ_* \overset{\mathcal{D}}{=} Z^*(1 - W) + WV, \tag{65}$$

where V, Z_*, Z^*, and W are independent. Let $\mu(\mathrm{d}z) = \beta(z)\,\mathrm{d}z$. Suppose that Y has no atom at 0. Then an integral equation for $\beta(x)$ is

$$\theta u \beta(u) = \int_0^u \frac{1}{u-z} f_W\left(1 - \frac{1-u}{1-z}\right) z(1-z)\beta(z)\,\mathrm{d}z$$
$$- \int_u^1 \frac{1}{z-u} f_W\left(1 - \frac{u}{z}\right) z(1-z)\beta(z)\,\mathrm{d}z,$$

which is equivalent to

$$\theta u \beta(u) = \int_0^u 2F^+\left(1 - \frac{1-u}{1-z}\right) z\beta(z)\,\mathrm{d}z - \int_u^1 2F^+\left(1 - \frac{u}{z}\right)(1-z)\beta(z)\,\mathrm{d}z.$$

Proof. To obtain a limit in the Λ-Fleming–Viot process, let $\theta_1 = \theta/d$ in generator (45). In identity (46) the density of Z is $d\,z f_{X_1}(z), 0 < z < 1$, by symmetry. Let $d \to \infty$ in identity (46). Then the identity becomes (65). The integral equations for the stationary distribution when there are two types imply an integral equation for $\beta(x)$. In view of (56),

$$x\beta(x) = -\frac{1}{\theta} f'_{\bullet L}(u),$$

where $f'_{\bullet L}(u)$ is similar to $f'_{\diamond L}(u)$ with $f_X(z)$ replaced by $\beta(z)$.

In the Wright–Fisher diffusion $W \equiv 0$ and identity (65) is

$$V Z_* \overset{\mathcal{D}}{=} Z^*. \tag{66}$$

It is straightforward to verify that if Z has density $\theta(1-z)^{\theta-1}$ then (65) is satisfied. A direct solution can be found in the following way. From (66),

$$\theta \int_z^1 y\beta(y)\,\mathrm{d}y = z(1-z)\beta(z), \tag{67}$$

where θ is defined by

$$\theta = \frac{\int_0^1 z^2 \beta(z)\,\mathrm{d}z}{\int_0^1 z(1-z)\beta(z)\,\mathrm{d}z}.$$

Write (67) as

$$\frac{\mathrm{d}}{\mathrm{d}z} \log \int_z^1 y\beta(y)\,\mathrm{d}y = -\theta(1-z)^{-1}.$$

Solving this differential equation,

$$\log \int_z^1 y\beta(y)\,\mathrm{d}y = \theta \log(1-z) + A$$

for a constant A. Therefore,

$$\int_z^1 y\beta(y)\,\mathrm{d}y = (1-z)^\theta$$

because $\int_0^1 yh(y)\,\mathrm{d}y = 1$, and, since (67) holds,

$$\beta(z) = \theta z^{-1}(1-z)^{\theta-1}, \qquad 0 < z < 1.$$

2.6. A different dual process

The typed Λ-coalescent tree process is a moment dual in the Fleming–Viot process; see, for example, Etheridge *et al.* (2010). We work through a different type of dual process which is a death process decreasing in steps of 1. Let $d = 2$ for simplicity. The generator $\mathcal{L}$ is specified by (45). Let $\{g_n(x)\}$ be a sequence of monic polynomials that are defined below and satisfy the generator equation

$$
\begin{aligned}
\mathcal{L}g_n &= \tfrac{1}{2}x(1-x)\mathbb{E}g_n''(x(1-W)+VW) + \tfrac{1}{2}(\theta_1 - \theta x)g_n'(x) \\
&= \binom{n}{2}\mathbb{E}(1-W)^{n-2}[g_{n-1}(x) - g_n(x)] + n\tfrac{1}{2}[\theta_1 g_{n-1}(x) - \theta g_n(x)]
\end{aligned} \tag{68}
$$

with $g_0(x) = 1$. Equation (68) is an analogue of the Wright–Fisher diffusion when we look at $g_n(x) = x^n$, with the second line chosen to mimic the Wright–Fisher case. Rearrange the equation to define $g_n(x)$ in terms of $g_{n-1}(x)$ as

$$
\begin{aligned}
&\tfrac{1}{2}x(1-x)\mathbb{E}g_n''(x(1-W)+VW) + \tfrac{1}{2}(\theta_1 - \theta x)g_n'(x) + \lambda_n g_n(x) \\
&\quad = \frac{n}{2}[(n-1)\mathbb{E}(1-W)^{n-2} + \theta_1]g_{n-1}(x).
\end{aligned} \tag{69}
$$

The polynomials $\{g_n(x)\}$ are well defined by (69) by recursively calculating the coefficients of x^r in $g_n(x)$ from $r = n-1, n-2, \ldots, 0$.

Theorem 9. *Let $\{g_n(x)\}_{n=0}^{\infty}$ be defined by (69). If X has a stationary distribution then*

$$
\mathbb{E}[g_n(X)] = \frac{\prod_{j=1}^{n}((j-1)\mathbb{E}(1-W)^{j-2} + \theta_1)}{\prod_{j=1}^{n}((j-1)\mathbb{E}(1-W)^{j-2} + \theta)}. \tag{70}
$$

Let $h_n(x) = g_n(x)/\mathbb{E}[g_n(X)]$. There is a dual process $\{N(t)\}_{t\geq 0}$ to $\{X(t)\}_{t\geq 0}$ based on the test functions $\{h_n(x)\}_{n=0}^{\infty}$ which is a death process with rates for $n \to n-1$, $n \geq 1$, of

$$
\lambda_n = \frac{n}{2}[(n-1)\mathbb{E}(1-W)^{n-2} + \theta].
$$

The dual equation is

$$
\mathbb{E}_{X(0)=x}[h_n(X(t))] = \mathbb{E}_{N(0)=n}[h_{N(t)}(x)]. \tag{71}
$$

The transition functions for the dual process are

$$
\begin{aligned}
&P(N(t) = j \mid N(0) = i) \\
&= \sum_{k=j}^{i} e^{-\lambda_k t} r_i^{(k)} l_j^{(k)} \\
&= \sum_{k=j}^{i} e^{-\lambda_k t}(-1)^{k-j} \frac{\lambda_{j+1}\cdots\lambda_i}{(\lambda_j - \lambda_k)\cdots(\lambda_{k-1}-\lambda_k)(\lambda_{k+1}-\lambda_k)\cdots(\lambda_i - \lambda_k)}.
\end{aligned} \tag{72}
$$

The process $\{N(t)\}_{t\geq 0}$ comes down from infinity if and only if

$$
\int_1^{\infty} \frac{\mathrm{d}q}{q^2 \mathbb{E}[(1-W)^q]} < \infty, \tag{73}
$$

which implies that the Λ-coalescent comes down from infinity. The distribution of $N(t)$ given an entrance boundary at infinity,

$$P(N(t) = j \mid N(0) = \infty) = \sum_{k=j}^{\infty} e^{-\lambda_k t} r_\infty^{(k)} l_j^{(k)}, \tag{74}$$

where

$$r_\infty^{(k)} = \prod_{l=k}^{\infty} \left(1 - \frac{\lambda_k}{\lambda_l}\right)^{-1},$$

is well defined assuming that condition (73) holds when the coalescent comes down from infinity.

Proof. Equation (70) follows directly from $\mathbb{E}[g_n(X)]$ in (68). Note that

$$\mathcal{L}h_n = \lambda_n[h_{n-1} - h_n],$$

which is correctly set up as a dual generator equation of the death process $\{N(t)\}_{t \geq 0}$. The dual equation is then (71).

The process $\{N(t), \ t \geq 0\}$ comes down from infinity if and only if

$$\sum_{2}^{\infty} \lambda_n^{-1} < \infty, \tag{75}$$

which implies that the Λ-coalescent comes down from infinity because

$$\lambda_n = \sum_{k=2}^{n} \binom{n}{k} \lambda_{nk} + n\theta;$$

so (75) is equivalent to

$$\sum_{n=2}^{\infty} \left[\sum_{k=2}^{n} \binom{n}{k} \lambda_{nk}\right]^{-1} < \infty, \tag{76}$$

and

$$\sum_{n=2}^{\infty} \left[\sum_{k=2}^{n} (k-1)\binom{n}{k} \lambda_{nk}\right]^{-1} < \sum_{n=2}^{\infty} \left[\sum_{k=2}^{n} \binom{n}{k} \lambda_{nk}\right]^{-1} < \infty. \tag{77}$$

Recalling that

$$\sum_{k=2}^{n} \binom{n}{k} \lambda_{nk} = \frac{1}{2} n(n-1)\mathbb{E}[(1 - W)^{n-2}],$$

by the integral comparison test, (76) is equivalent to (73).

For example, if W has a Beta(α, β) distribution for $\alpha, \beta > 0$ then $\mathbb{E}[(1 - W)^n] \sim Cn^{-\alpha}$, where C is a constant, so if $\alpha < 1$ then $\sum_{n=2}^{\infty} \lambda_n^{-1} < \infty$, because the nth term is asymptotic to $(C/2)n^{2-\alpha}$. Coming down from infinity does not necessarily imply that (75) or (76) holds. In general, the tail of the series (77) satisfies

$$\sum_{n=N}^{\infty} \left[\sum_{k=2}^{n} \binom{n}{k} \lambda_{nk}\right]^{-1} \approx \frac{1}{2} \int_N^{\infty} \frac{1}{q^2} \frac{dq}{\mathbb{E}[(1 - W)^q]} = \frac{1}{2} \int_0^{N^{-1}} \frac{dz}{\mathbb{E}[(1 - W)^{z^{-1}}]}.$$

Convergence of the integral depends on $\mathbb{E}[(1-W)^{z^{-1}}]$ being large enough as $z \to 0$. It is very likely that there are connections with the speed of coming down from infinity studied in Berestycki *et al.* (2014b), but the exact connections are not clear.

The transition functions for the process $\{N(t), \, t \geq 0\}$ are easily found from an eigenfunction analysis of the Q matrix, where $q_{jj} = -\lambda_j$ and $q_{jj-1} = \lambda_j$. The approach used to find the eigenfunction expansion for the transition distribution in the Kingman coalescent is that given in Tavaré (1984) (see also Griffiths (1980)). The left and right eigenvectors $l_j^{(k)}$ and $r_i^{(k)}$ are triangular in form with $l_j^{(k)} = 0, \, j > k$, and $r_i^{(k)} = 0, \, i < k$. Explicit formulae are $l_k^{(k)} = r_k^{(k)} = 1$ and

$$l_j^{(k)} = \frac{(-1)^{k-j}\lambda_{j+1}\cdots\lambda_k}{(\lambda_j - \lambda_k)\cdots(\lambda_{k-1} - \lambda_k)}, \qquad j < k,$$

$$r_i^{(k)} = \frac{\lambda_i \cdots \lambda_{k+1}}{(\lambda_i - \lambda_k)\cdots(\lambda_{k+1} - \lambda_k)}, \qquad i > k.$$

The transition functions are then given by (72).

The distribution of $N(t)$ given an entrance boundary at ∞ is the distribution as $i \to \infty$, which is (74), and is well defined assuming that condition (75) holds when the coalescent comes down from infinity.

The condition of Bertoin and Le Gall (2006), (4), for coming down from infinity is equivalent to

$$\int_1^\infty \frac{\mathrm{d}q}{q\mathbb{E}[(1-\mathrm{e}^{-qW})/W]} < \infty. \tag{78}$$

There can be a gap where the Λ-coalescent comes down from infinity but $\{N(t)\}_{t \geq 0}$ does not come down from infinity because (73) and (78) are not equivalent.

2.6.1. *Eigenfunctions $P_n(x)$ and polynomials $g_n(x)$.* It is extremely interesting that the polynomials $\{P_n(x)\}$ are analogous to the monic Jacobi polynomials distribution with $\{g_n(x)\}$ analogous to $\{x^n\}$.

Express

$$g_n(x) = P_n(x) + \sum_{r=0}^{n-1} b_{nr} P_r(x),$$

where $P_n(x)$ are the eigenfunctions of $\mathcal{L}$. Define

$$\lambda_n^\circ = \frac{n}{2}[(n-1)\mathbb{E}(1-W)^{n-2} + \theta_1].$$

From (69) and noting that

$$\mathcal{L}P_n = -\lambda_n P_n, \qquad \mathcal{L}g_n = -\lambda_n g_n + \lambda_n^\circ g_{n-1}, \tag{79}$$

it follows that

$$\sum_{r=0}^{n-1} b_{nr}[-\lambda_r + \lambda_n]P_r(x) = \lambda_n^\circ \sum_{r=0}^{n-1} b_{n-1r} P_r(x). \tag{80}$$

Note that $g_l(x)$ being a monic polynomial means that $b_{ll} = 1, \, l = 0, 1, \ldots$. Calculating coefficients from (80) we obtain

$$b_{nr} = \frac{\lambda_n^\circ}{\lambda_n - \lambda_r} b_{n-1r} = \frac{\lambda_n^\circ \lambda_{n-1}^\circ \cdots \lambda_{r+1}^\circ}{(\lambda_n - \lambda_r)(\lambda_{n-1} - \lambda_r)\cdots(\lambda_{r+1} - \lambda_r)}.$$

The eigenfunctions $\{P_n(x)\}$ also have an expansion in terms of the polynomials $\{g_r(x)\}$. Let

$$P_n(x) = g_n(x) + \sum_{r=0}^{n-1} c_{nr} g_r(x). \tag{81}$$

From (79),

$$-\lambda_n P_n(x) = -\lambda_n g_n(x) + \lambda_n^\circ g_{n-1}(x) + \sum_{r=0}^{n-1} [-\lambda_r c_{nr} + \lambda_{r+1}^\circ c_{nr+1}] g_r(x).$$

Expressing the left-hand side by the expansion (81) and equating coefficients of $g_r(x)$ we obtain

$$-\lambda_n c_{nr} = -\lambda_r c_{nr} + \lambda_{r+1}^\circ c_{nr+1}.$$

The coefficients are therefore

$$c_{nr} = \frac{\lambda_{r+1}^\circ \cdots \lambda_n^\circ}{(\lambda_r - \lambda_n) \cdots (\lambda_{n-1} - \lambda_n)}. \tag{82}$$

Scale (68) by taking

$$g_n(x) = \frac{\lambda_n^\circ \cdots \lambda_1^\circ}{\lambda_n \cdots \lambda_1} h_n(x) = \frac{\prod_{j=1}^n ((j-1)\mathbb{E}(1-W)^{j-2} + \theta_1)}{\prod_{j=1}^n ((j-1)\mathbb{E}(1-W)^{j-2} + \theta)} h_n(x).$$

Defining ω_n as a beta moment analog,

$$\omega_n = \frac{\prod_{j=1}^n ((j-1)\mathbb{E}(1-W)^{j-2} + \theta_1)}{\prod_{j=1}^n ((j-1)\mathbb{E}(1-W)^{j-2} + \theta)},$$

we obtain

$$g_n(x) = \omega_n h_n(x).$$

Note that if X has a stationary distribution then

$$\mathbb{E}[g_n(X)] = \omega_n.$$

The polynomials $\{P_n(x)\}$ are analogous to the monic Jacobi polynomials orthogonal on the Beta(θ_1, θ_2) distribution with $\{g_n(x)\}$ analogous to $\{x^n\}$. If $W \equiv 0$ then they are identical in the analogy. In the Jacobi polynomial case (82) simplifies to

$$c_{nr} = (-1)^{n-r} \frac{(n-r-1)!}{r!} \frac{\theta_{1(n)}}{\theta_{1(r)}} \frac{(n+\theta)_{(r)}}{(n+\theta)_{(n-1)}}.$$

The process is not reversible, so the polynomials are not orthogonal on any measure unless they are the Jacobi polynomials.

2.6.2. *Higher dimensions.* Let $\mathcal{L}$ be the d-dimensional Λ-Fleming–Viot generator with mutation, and define polynomials $\{g_n(x)\}$ with $g_0(x) = 1$ by

$$\mathcal{L}g_n(x) = -\lambda_n g_n(x) + \frac{1}{2} \sum_{i=1}^d \frac{n_i}{n} n((n_i - 1)\mathbb{E}[(1-W)^{n_i-2}] + \theta_i) g_{n-e_i}(x).$$

This is an analogy with the Wright–Fisher generator acting on x^n. The polynomials are well defined by recursion on their coefficients. In a similar calculation to the two-dimensional case there is a Dirichlet moment analogue:

$$\mathbb{E}[g_{\boldsymbol{n}}(\boldsymbol{X})] = \frac{\prod_{i=1}^{d}[\prod_{j=1}^{n_i}((j-1)\mathbb{E}[(1-W)^{j-2}] + \theta_i)]}{\prod_{j=1}^{n}((j-1)\mathbb{E}[(1-W)^{j-2}] + \theta)}.$$

The dual process constructed from test functions $g_{\boldsymbol{n}}(\boldsymbol{X})/\mathbb{E}[g_{\boldsymbol{n}}(\boldsymbol{X})]$ has transitions

$$\boldsymbol{n} \to \boldsymbol{n} - \boldsymbol{e}_i \quad \text{at rate} \quad \frac{n_i}{n}((n-1)\mathbb{E}[(1-W)^{n-2}] + \theta).$$

The dual equation is similar to (71). Let

$$h_{\boldsymbol{n}}(\boldsymbol{x}) = \frac{g_{\boldsymbol{n}}(\boldsymbol{x})}{\mathbb{E}[g_{\boldsymbol{n}}(\boldsymbol{X})]}.$$

Then

$$\mathbb{E}_{X(0)=\boldsymbol{x}}[h_{\boldsymbol{n}}(\boldsymbol{X}(t)] = \mathbb{E}_{N(0)=\boldsymbol{n}}[h_{\boldsymbol{N}(t)}(\boldsymbol{x})].$$

The multitype death process has transition probabilities which are easy to describe from the sum of the entries $|\boldsymbol{N}(t)|$ and (72):

$$\mathbb{P}(\boldsymbol{N}(t) = \boldsymbol{m} \mid \boldsymbol{N}(0) = \boldsymbol{n}) = \frac{\prod_{j=1}^{d}\binom{n_j}{m_j}}{\binom{n}{m}}\mathbb{P}(|\boldsymbol{N}(t)| = m \mid |\boldsymbol{N}(0)| = n).$$

An equation analogous to the k-dimensional Ewens sampling formula in the Poisson Dirichlet process is to let $\theta_i = \theta/d$, $i = 1, \dots, d$, then the (labelled) sampling formula is

$$\lim_{d \to \infty} d_{[k]} \binom{n}{\boldsymbol{n}} \mathbb{E}[g_{\boldsymbol{n}}(\boldsymbol{X})],$$

where $\boldsymbol{n} = (n_1, \dots, n_k, 0, \dots, 0)$. The sampling formula limit is

$$\frac{n!\,\theta^k}{n_1 \cdots n_k} \frac{\prod_{i=1}^{k}[\prod_{j=2}^{n_i} \mathbb{E}[(1-W)^{j-2}]]}{\prod_{j=1}^{n}((j-1)\mathbb{E}[(1-W)^{j-2}] + \theta)}.$$

Möhle (2006) and Lessard (2010) studied recursive equations leading to the Λ-coalescent sampling formula. The familiar Ewens sampling formula is obtained by taking $W = 0$.

References

BAH, B. AND PARDOUX, E. (2013). Lambda-lookdown model with selection. Preprint. Available at http://uk.arxiv.org/abs/1303.1953.

BERESTYCKI, N. (2009). *Recent Progress in Coalescent Theory* (Ensaios Matemáticos **16**). Sociedade Brasileira de Matemática, Rio de Janeiro.

BERESTYCKI, J., BERESTYCKI, N. AND LIMIC, V. (2014a). A small-time coupling between Λ-coalescents and branching processes. *Ann. Appl. Prob.* **24,** 449–475.

BERESTYCKI, J., BERESTYCKI, N. AND LIMIC, V. (2014b). Asymptotic sampling formulae for Λ-coalescents. *Ann. Inst. H. Poincaré Prob. Statist.* **50,** 715–731

BERTOIN, J. AND LE GALL, J.-F. (2003). Stochastic flows associated to coalescent processes. *Prob. Theory Relat. Fields* **126,** 261–288.

BERTOIN, J. AND LE GALL, J.-F. (2006). Stochastic flows associated to coalescent processes. III. Limit theorems. *Illinois J. Math.* **50,** 147–181.

BIRKNER, M. AND BLATH, J. (2009). Measure-valued diffusions, general coalescents and population genetic inference. In *Trends in Stochastic Analysis*, Cambridge University Press, pp. 329–363.

BIRKNER, M. *et al.* (2005). Alpha-stable branching and beta-coalescents. *Electron. J. Prob.* **10,** 303–325.

COX, D. R. (1970). *Renewal Theory*. Methuen, London.

DALEY, D. J. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. I, 2nd edn. Springer, New York.

DER, R., EPSTEIN, C. L. AND PLOTKIN, J. B. (2011). Generalized population models and the nature of genetic drift. *Theoret. Pop. Biol.* **80,** 80–99.

DER, R., EPSTEIN, C. AND PLOTKIN, J. B. (2012). Dynamics of neutral and selected alleles when the offspring distribution is skewed. *Genetics* **191,** 1331–1344.

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

DONNELLY, P. AND KURTZ, T. G. (1999). Particle representations for measure-valued population models. *Ann. Prob.* **27,** 166–205.

ELDON, B. AND WAKELEY, J. (2006). Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172,** 2621–2633.

ETHERIDGE, A. (2011). *Some Mathematical Models from Population Genetics*. Springer, Heidelberg.

ETHERIDGE, A. M., GRIFFITHS, R. C. AND TAYLOR, J. E. (2010). A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theoret. Pop. Biol.* **78,** 77–92.

EWENS, W. J. (1972). The sampling theory of selectively neutral alleles. *Theoret. Pop. Biol.* **3,** 87–112.

EWENS, W. J. (2004). *Mathematical Population Genetics. I. Theoretical Introduction*, 2nd edn. Springer, New York.

FOUCART, C. (2013). The impact of selection in the Λ–Wright–Fisher model. *Electron. Commun. Prob.* 18, no. 72. (Erratum: **19** (2014), no. 15.)

GRIFFITHS, R. C. (1980). Lines of descent in the diffusion approximation of neutral Wright–Fisher models. *Theoret. Pop. Biol.* **17,** 37–50.

HANDA, K. (2014). Stationary distributions for a class of generalized Fleming–Viot processes. *Ann. Prob.* **42,** 1257–1284.

ISMAIL, M. E. H. (2005). *Classical and Quantum Orthogonal Polynomials in One variable* (Encyclopedia Math. Appl. **98**). Cambridge University Press.

KIMURA, M. (1964). Diffusion models in population genetics. *J. Appl. Prob.* **1,** 177–232.

LESSARD, S. (2010). Recurrence equations for the probability distribution of sample configurations in exact population genetics models. *J. Appl. Prob.* **47,** 732–751.

MÖHLE, M. (2006). On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12,** 35–53.

MÖHLE, M. AND SAGITOV, S. (2001). A classification of coalescent processes for haploid exchangeable population models. *Ann. Prob.* **29,** 1547–1562.

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

PITMAN, J. (2006). *Combinatorial Stochastic Processes*. Springer, Berlin.

SAGITOV, S. (1999). The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Prob.* **36,** 1116–1125.

SCHWEINSBERG, J. (2000). A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Prob.* **5,** 1–11.

SCHWEINSBERG, J. (2003). Coalescent processes obtained from supercritical Galton–Watson processes. *Stoch. Process. Appl.* **106,** 107–139.

TAVARÉ, S. (1984). Line-of-descent and genealogical processes, and their application in population genetics models. *Theoret. Pop. Biol.* **26,** 119–164.