arXiv:1903.06731v1 [math.PR] 15 Mar 2019

GENERAL SELECTION MODELS: BERNSTEIN DUALITY AND MINIMAL ANCESTRAL STRUCTURES

F. CORDERO, S. HUMMEL, AND E. SCHERTZER

Abstract. The Λ-Wright–Fisher process describes the type-frequency evolution of an infinite population. We model frequency-dependent selection pressure with a general polynomial drift vanishing at the boundary. An appropriate decomposition of the drift allows us to construct a series of Moran-type models that converge under suitable conditions to the solution of the associated stochastic differential equation. The genealogical structure inherent in the graphical representation of these finite population models can be seen in the large population limit as a generalisation of the ancestral selection graph of Krone and Neuhauser. We introduce an ancestral process that keeps track of the sampling distribution along the ancestral structures and that satisfies a duality relation with the type-frequency process. We refer to it as Bernstein coefficient process and to the relation as Bernstein duality. The latter is a generalisation of the classic moment duality. Many classic results in the restricted setting of a moment duality generalise into our framework. In particular, we derive criteria for the accessibility of the boundary and determine the time to absorption. It turns out that multiple ancestral processes are associated to the same forward dynamics. We characterise the set of optimal ancestral structures and provide a recipe to construct them from the drift. In particular, this allows us to recover well-known ancestral structures of the literature.

MSC 2010. Primary: 82C22, 92D15 Secondary: 60J25, 60J27

Keywords. duality, frequency-dependent selection, branching-coalescing system, ancestral selection graph, fixation probability, absorption probability, coming down from infinity

Contents

1. Introduction

Forward and backward neutral models. The analysis of models in population genetics is based on two perspectives. One traces the evolution of type-frequencies in the population forward in time; the other one the corresponding ancestral structures backward in time. The understanding of the interplay between the two perspectives is of great significance for the development of inference methods.

A neutral setting typically leads to a driftless stochastic differential equation (SDE) forward in time; whereas coalescent processes describe the genealogies of samples of the population. The most classic example is the Wright–Fisher diffusion with two types and its ancestral counterpart, the Kingman coalescent (see [40]). Forward in time, the evolution of the frequency of one type (in an essentially infinite

Date: March 19, 2019.

population) is described by the standard Wright–Fisher diffusion, i.e.

$$dX_t \;=\; \sqrt{X_t(1-X_t)}\,dW_t, \quad X_0 = x \in [0,1],$$

where $(W_t; t \geq 0)$ is a standard Brownian motion. Backward in time, the block-counting process of the Kingman coalescent, denoted by $(L_t; t \geq 0)$, describes the evolution of the number of ancestors of a given sample in the population. The formal link between the two processes is well-known to be a moment duality (see, e.g. [38] for a survey on duality methods), i. e.

$$\mathbb{E}_x[X_t^n] \;=\; \mathbb{E}_n[x^{L_t}] \qquad \forall x \in [0,1], \forall n \in \mathbb{N}. \tag{1.1}$$

There is an alternative way relating the backward and forward perspectives. Finite population models allow for a graphical representation as an interactive particle system that unites the two perspectives into the same picture. For example, in the neutral Moran model with constant population size N (see [18]) reproduction events that involve two individuals (the parent and the replaced individual) translate into binary coalescence events (two individuals sharing a common ancestor) in the backward picture. This leads to a natural coupling (a pathwise duality) between the type-frequency process and the N-Kingman coalescent, which traces back the genealogy of the entire population. As the size of the population tends to infinity, one asymptotically recovers the duality relation (1.1) between the Kingman coalescent and the Wright–Fisher diffusion from the pathwise duality at the level of the finite population approximation.

The convergence to the Wright–Fisher diffusion, for the forward evolution, and to the Kingman coalescent, for their genealogies, holds for a large class of population models (see [44, 45]). However, if the variance of the number of offspring per individual is asymptotically infinite, these approximations are inappropriate. Backward in time, this leads to consider more general exchangeable coalescents with multiple mergers of ancestral lines, called Λ-coalescents, which were introduced independently in [16], [50], and [51], and have been subject to extensive research in the past decades (see [8] for a review on the topic). They describe the genealogy of a sample from a forward in time population model in which the type-frequency process has jumps. In the two-type case, the type-frequency process $(X_t; t \geq 0)$ is called the Λ-Wright–Fisher process and evolves according to the following SDE

$$dX_t \;=\; \sqrt{\Lambda(\{0\})X_t(1-X_t)}\,dW_t + \int\limits_{(0,1]\times[0,1]} r\big(\mathbf{1}_{\{u\leq X_{t-}\}}(1-X_{t-}) - \mathbf{1}_{\{u>X_{t-}\}}X_{t-}\big)\tilde{N}(dt,dr,du),$$

with $X_0 = x \in [0,1]$, where $(W_t; t \geq 0)$ is a standard Brownian motion, $\tilde{N}(dt,dr,du)$ is an independent compensated Poisson measure on $[0,\infty) \times (0,1] \times [0,1]$ with intensity $dt \times r^{-2}\Lambda(dr) \times du$ (see [9] for more details). As in the Kingman case, the forward model is related to its genealogy by the moment duality (1.1), where $(L_t; t \geq 0)$ is now the block counting process of the underlying Λ-coalescent (see, e.g. [18]). Alternatively, one can relate the forward and backward models at the finite population level, as in the Wright–Fisher diffusion, via an extension of the Moran model where the offspring of one individual may replace a positive fraction of the population (see, e.g. [10]).

Forward and backward models with selection. Many tools employed in the analysis of the aforementioned models rely on the neutrality assumption. Forward in time, selection usually leads to the inclusion of a drift term of the form $x(1-x)\,s(x)$, for some function $s(\cdot)$, to the SDE, i.e.

$$\begin{aligned} dX_t \;&=\; X_t(1-X_t)\,s(X_t)dt \;+\; \sqrt{\Lambda(\{0\})X_t(1-X_t)}\,dW_t \\ &\quad + \int\limits_{(0,1]\times[0,1]} r\Big(\mathbf{1}_{\{u\leq X_{t-}\}}(1-X_{t-}) - \mathbf{1}_{\{u>X_{t-}\}}X_{t-}\Big)\tilde{N}(dt,dr,du), \quad X_0 = x \in [0,1], \end{aligned} \tag{1.2}$$

where W and $\tilde{N}$ are as above. The resulting process is called the Λ-*Wright–Fisher process with (frequency-dependent) selection*. Selection that is not frequency-dependent (i.e. when s is a constant function) is called genic selection. In this case, the genealogy of (1.2) was first described in the seminal work of Krone and Neuhauser [41][47] for the Wright–Fisher diffusion, and later complemented by [3, 22, 26, 33] to the Λ-Wright–Fisher case. They all rely on the ancestral selection graph (ASG), which is the graph in which lineages of potential ancestors augment the associated coalescent process. The result is a branching-coalescing process together with a rule that prescribes at each branching event the true parent (among a given set of potential ancestors).

Little is known beyond the case of genic selection, i.e. when s is a trivial function of x. In [46], Neuhauser described an extension of the ASG in the case of balancing selection, i.e. for $s(x) = 1 - 2x$. González Casanova and Spanò [32] constructed another extension of the ASG when s is a power series with negative, non-decreasing coefficients. Moreover, they prove that the moment duality (1.1) then also holds between $(X_t;\, t \geq 0)$ and the block-counting process of the associated ASG (see also [26] for the case where s is a negative constant).

Although frequency-dependent selection plays a central role in ecology and evolution [2], a general framework to treat models with general selection term is, to the best of our knowledge, still missing. The present article is a first step to fill this gap. We consider the SDE (1.2) for a general polynomial s and address the following questions.

(Q1) Can we associate a natural genealogy to the SDE (1.2)?

(Q2) Is there an extension of the moment duality (1.1)?

For those models that can be approximated via a sequence of Moran-type models, the answer to (Q1) is intuitive. Analogously to the neutral setting, for a Moran-type model, one can interpret selection and neutral reproduction mechanisms at the level of individuals. Forward and backward models are then embedded into the same graphical representation, which leads to a natural (sampling) duality between the two perspectives. This leads to the following reformulation of (Q1).

(Q1') Can we construct a family of Moran models converging to the solution of the SDE (1.2)?

Indeed, it turns out that for any polynomial s such a construction is possible, and this construction provides a natural dual backward process describing the genealogy of potential ancestors. Our answer to (Q1) generalises the ASG of Krone and Neuhauser [41][47] to models with general frequency-dependent selection term. We also show that there are many ways to approximate (1.2) by Moran-type models. One important (and puzzling) consequence is that many different ASGs can be associated to the same model. This plurality of ASGs will be addressed in more detail in (Q3) and (Q4) below.

We now turn to (Q2). For a given ASG, we formalise the aforementioned sampling duality at the infinite population level in the spirit of [14, 15, 52]. The duality relation (1.1) then extends to general selection models as

$$\mathbb{E}_x\big[X_t^n\big] \;=\; \mathbb{E}_{e_{n+1}}\Bigg[\sum_{\ell=0}^{L_t} V_t(\ell) b_{\ell,L_t}(x)\Bigg], \tag{1.3}$$

where L_t counts the number of potential ancestors in the ASG of a sample of size n, $b_{i,L_t}(x)$ is the i-th Bernstein polynomial in the basis of degree L_t, and the coefficient process $V := (V_t; t \geq 0)$ is an explicit Markov chain valued in $\cup_{n\in\mathbb{N}_0}\mathbb{R}^n$ that is started in $e_{n+1} = (0, \ldots, 0, 1)^T$, i.e. the $(n+1)$-st unit vector. We refer to this duality as Bernstein duality. After the formulation of the duality, the following question arises.

(Q3) What can we say about the solution of the SDE (1.2) from (one of) its genealogical processes and the duality relation among them?

The Bernstein duality allows us to relate the absorption (fixation/extinction) probabilities and the time to absorption of (1.2) to properties of the Bernstein coefficient process V, which appears in (1.3). In particular, we show that the fixation probabilities relate to the invariant measure of V (when it exists) and the time to fixation relates to the entrance law of V at ∞ (when it is non-trivial). [5, 26, 32] use the moment duality to characterise the absorption probabilities of the process $(X_t; t \geq 0)$; but this approach was until now only feasible in the restricted case in which a moment duality is available. We generalise the method to polynomial s for which this is not the case.

Next, since different ASGs associated to the same model can have substantially different behaviour (for example, transience or recurrence of $(L_t; t \geq 0)$), the ambiguity in the choice of the ancestral processes (i.e. the fact that we can construct different ASGs for a given forward model) makes it natural to ask:

(Q4) Are some ASGs better than others? Is there an optimal one? Is there a unique optimal one?

The aforementioned ambiguity in the choice of ancestral models resolves with the introduction of the notion of a minimal ASG. Loosely speaking, these ASGs are the ones that minimise the number of potential ancestors. Via a restriction to those minimal ancestral structures, one recovers classical cases from the literature [41, 46, 47].

Let us close this introduction with a reference to an independent work by González Casanova and Smadi [31], who answer question (Q1') in a multidimensional setting (i.e with more than two types) and with mutations. They design a fixed-size Wright–Fisher population model whose asymptotic type frequencies converge to a multi-dimensional version of (1.2). In this framework, they study fixation and extinction properties in some classical ecological models (such as the rock-paper-scissor and food-web models). Given the intriguing biological applications presented in [31], it would be interesting to investigate the extension of our duality result with regard to (Q2–Q4) in higher dimensions.

Outline. The article is organised as follows. Section 2 provides an outline of the paper and contains all our main results. The proofs and more in-depth analyses are shifted to the subsequent sections. Section 3 contains the proof of the convergence of appropriate Moran models to the SDE (1.2). A detailed discussion of the ancestral process and the proofs of its properties can be found in Section 4. In particular, it contains the proof of the Bernstein duality. The process that keeps track of the number of potential ancestors is analysed in Section 5. Section 6 is devoted to applications of the new processes and of the duality. In Section 7, we treat the problem of minimality among genealogies from two different perspectives. One that seeks to avoid superfluous branches and another that minimises the effective branching rate.

2. Summary of main results

In this section we provide a detailed outline of the paper and state the main results. We start with some notation that will be used throughout. The positive integers are denoted by $\mathbb{N}$ and we set $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. For $m \in \mathbb{N}$,

$$[m] := \{1, \ldots, m\} \quad [m]_0 := [m] \cup \{0\}, \quad \text{and} \quad]m] := [m] \setminus \{1\}.$$

Furthermore, define for $m \in \mathbb{N} \setminus \{1\}$

$$\mathcal{P}_m := \prod_{\ell=2}^{m} \{0\} \times [0,1]^{\ell-1} \times \{1\} \quad \text{and} \quad E^m := \left\{ (\beta, p) : \beta = (\beta_\ell)_{\ell=2}^m \in \mathbb{R}_+^{m-1},\ p = (p_{\cdot,\ell})_{\ell=2}^m \in \mathcal{P}_m \right\}.$$

For $m \in \mathbb{N}$ and $i \in [m]_0$, let $b_{i,m}$ be the i-th polynomial in the Bernstein basis of degree m, i.e. $b_{i,m}(x) := \binom{m}{i} x^i (1-x)^{m-i}$, $x \in [0,1]$. In addition, define the m-th Bernstein function B_m via $B_m(x) := (b_{i,m}(x))_{i=0}^m$, $x \in [0,1]$. Since $\{b_{i,m} : i \in [m]_0\}$ forms a basis of the polynomials of degree at most m, for any polynomial d with $\deg(d) \leq m$, there exists a unique vector $v \in \mathbb{R}^{m+1}$ such that

$$d(\cdot) = \langle B_m(\cdot), v \rangle,$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in $\mathbb{R}^{m+1}$. We call the vector v the Bernstein coefficient vector (BCV) of d. If in addition d is such that $d(0) = d(1) = 0$, we define $\rho(d) \in \mathbb{R}^{m-1}$ as the unique vector such that $d(\cdot) = \left\langle B_m(\cdot), (0, v^T, 0)^T \right\rangle$. For any Borel set $S \subset \mathbb{R}$, denote by $\mathcal{M}_f(S)$ (resp. $\mathcal{M}_1(S)$) the set of finite (resp. probability) measures on S. We use $\xrightarrow{(d)}$ to denote convergence in distribution of random variables and $\overset{(d)}{\Longrightarrow}$ for convergence in distribution of càdlàg process, where we endow the space of càdlàg functions with the Skorokhod topology. Let $n, m, k \in \mathbb{N}_0$ with $n \geq m \vee k$, $i \leq k \wedge m$. For a random variable K, we write $K \sim \mathrm{Hyp}(n, m, k)$ if K has a hypergeoemtric distribution with parameter n, m, and k. In particular, for $i \in \mathbb{N}_0$,

$$\mathbb{P}(K = i) = \frac{\binom{n-m}{k-i}\binom{m}{i}}{\binom{n}{k}}.$$

Furthermore, let $x \in [0,1]$ and $n \in \mathbb{N}$. For a random variable B, we write $B \sim \mathrm{Bin}(n, x)$ if B has a binomial distribution with parameter n and x. In particular, for $i \in [n]_0$, $\mathbb{P}(B = i) = \binom{n}{i} x^i (1-x)^{n-1}$.

2.1. **Moran models with frequency-dependent selection and large neutral offspring.** To answer (Q1'), we incorporate selection by means of selective replacement events; guided by an idea that is already present in [32]. The basic principle is that at any selection event, an individual samples a set of *potential parents*. One of them is chosen, according to a criterion that depends on the sample composition, to pass on its type to the individual that initiated the selection event.

We now spell out our continuous-time finite population model in more detail. Consider a haploid population of fixed size N. Each individual in this population has a colour-type, which is either type a or type A. The population dynamics is driven by frequency-dependent selection and neutral offspring events that may lead to the replacement of a considerable fraction of the population. More specifically, our model is characterised by two sets of parameters

- a selection mechanism that is a pair $(\beta, p) \in E^m$ for some $m \in]N]$,
- a neutral reproduction mechanism that is $\mu \in \mathcal{M}_f([N]_0)$.

The dynamics of selective reproductions is as follows. For each $\ell \in]m]$, each individual experiences an ℓ-replacement at rate β_ℓ, independently of each other. This means that the selected individual joins a group of $\ell - 1$ potential parents chosen uniformly at random (without replacement) among the other $N-1$ individuals. If j of the potential parents (including the selected individual) are of type a, with probability $p_{j,\ell}$ (resp. $1 - p_{j,\ell}$) one individual chosen uniformly at random among the type a (resp. type A) potential parents reproduces and its single offspring, which has the same type as the parent, replaces the selected individual, see Fig. 1 (left).

FIGURE 1. Solid (resp. dotted) lines correspond to a type a (resp. type A). Time runs form left to right. Left: the type A individual indicated by a black square is affected by a 4-replacement event; the other potential parents are depicted by white squares; the individual is replaced by a type a (which occurs with probability $p_{2,4}$). Right: The individual with a black circle initiates a 3-reproduction event, all individuals indicated by a white circle are replaced by its offspring.

Neutral reproduction is driven by the measure μ. For each $r \in [N]$, each individual at rate $\mu(\{r\}) + \mathbf{1}_{\{r=1\}}\mu(\{0\})/2$, independently of each other, gives birth to r individuals. They inherit the parent's type and replace r uniformly chosen individuals present in the population before the reproduction event, see Fig. 1 (right). We call this event an r-reproduction. By construction, ℓ-replacements and r-reproductions keep the population size constant.

We refer to the previously described model as the (β, p, μ)-Moran model. The description of its large population behaviour requires some more notation. Define the operator $T^N : \mathcal{M}_f([N]_0) \to \mathcal{M}_1([0,1])$ via

$$T^N \mu := \frac{1}{M_\mu}\left(\mu(\{0\})\delta_0 + \sum_{k=1}^{N} \delta_{\frac{k}{N}}\, \mu(\{k\})k^2\right),$$

where $M_\mu := \mu(\{0\}) + \sum_{k=1}^N \mu(\{k\})k^2$ and δ_y is the Dirac mass at y.

Theorem 2.1 (Large population limit)**.** *Fix $m \in \mathbb{N} \setminus \{1\}$. Let $(\beta, p) \in E^m$ and $\Lambda \in \mathcal{M}_f([0,1])$. For each $N \in \mathbb{N}$ with $N \geq m$, let $(\beta^N, p) \in E^m$ and $\mu_N \in \mathcal{M}_f([N]_0)$. Let $X^{(N)} := (X_t^{(N)}; t \geq 0)$ be the type-a frequency process in a (β^N, p, μ_N)-Moran model of size N. Assume that*

(1) $N\beta^N \xrightarrow[N\to\infty]{} \beta$,

(2) $\mu_N(\{0\}) \xrightarrow[N\to\infty]{} \Lambda(\{0\}), \quad M_{\mu_N} \xrightarrow[N\to\infty]{} \Lambda([0,1]), \quad$ *and* $\quad T^N\mu_N \xrightarrow[N\to\infty]{(d)} \Lambda/\Lambda([0,1]).$

If in addition, $X_0^{(N)} \xrightarrow[N\to\infty]{} x \in [0,1]$, *then* $(X_{Nt}^{(N)}; t \geq 0) \xRightarrow[N\to\infty]{(d)} X := (X_t; t \geq 0)$, *where X is the pathwise unique strong solution of the SDE*

$$\begin{aligned} dX_t =& \sum_{\ell=2}^{m} \beta_\ell \sum_{i=0}^{\ell} b_{i,\ell}(X_t)\Big(p_{i,\ell} - \frac{i}{\ell}\Big)dt + \sqrt{\Lambda(\{0\})X_t(1-X_t)}dW_t \\ &+ \int\limits_{(0,1]\times[0,1]} r\Big(\mathbf{1}_{\{u\leq X_{t-}\}}(1-X_{t-}) - \mathbf{1}_{\{u>X_{t-}\}}X_{t-}\Big)\tilde{N}(dt,dr,du), \quad X_0 = x. \end{aligned} \tag{2.1}$$

Remark 2.1. Note that condition (2) in Theorem 2.1 is equivalent to

$$\mu_N(\{0\}) \xrightarrow[N\to\infty]{} \Lambda(\{0\}) \quad \text{and} \quad N^2 \sum_{k=1}^{N} f\left(\frac{k}{N}\right)\mu_N(\{k\}) \xrightarrow[N\to\infty]{} \int\limits_{(0,1]} f(r)\frac{\Lambda(dr)}{r^2},$$

for every $f \in \mathcal{C}([0,1])$ such that $x \in [0,1] \mapsto f(x)/x^2 \in \mathcal{C}([0,1])$ (cf. [22, Condition (4.6)]).

The previous theorem provides conditions under which a given sequence of Moran models converges to the SDE (2.1). The next result states an explicit sequence of Moran models having this feature.

Corollary 2.2. *Let* $(\beta, p, \Lambda) \in E^m \times \mathcal{M}_f([0,1])$. *Define* $(\beta^N, p, \mu_N) \in E^m \times \mathcal{M}_f([N]_0)$ *via*

$$\beta^N := \frac{\beta}{N} \quad \textit{and} \quad \mu_N := \Lambda(\{0\})\,\delta_0 + \frac{1}{N^2}\sum_{k=1}^{N-1}\binom{N}{k+1}\lambda^0_{N,k+1}\,\delta_k,$$

where $\lambda^0_{n,k} := \int_{(0,1]} r^{k-2}(1-r)^{n-k}\Lambda(dr)$, $n \geq k \geq 2$. *Let* $X^{(N)} := (X_t^{(N)}; t \geq 0)$ *be the type-a frequency process in a* (β^N, p, μ_N)*-Moran model of size* N. *If in addition,* $X_0^{(N)} \xrightarrow[N\to\infty]{} x \in [0,1]$, *then*

$$(X_{Nt}^{(N)}; t \geq 0) \xRightarrow[N\to\infty]{(d)} X := (X_t; t \geq 0),$$

where X is the pathwise unique strong solution of (2.1).

Remark 2.2. For $n \geq k > 2$, $\lambda^0_{n,k}$ corresponds to the rate at which any given tuple of k blocks in the Λ-coalescent merges, when there are n blocks in total. The choice of μ_N in the previous corollary is used and studied in [10, Eq. (1.31)] (up to a factor N, because we consider rates per individual instead of total rates).

Remark 2.3. Typical choices for $p \in \mathcal{P}_m$ are

(1) $p_{i,\ell} = \mathbf{1}_{\{i \geq \lceil \ell/2 \rceil\}}$ - majority rule (e.g. [19]), (3) $p_{i,\ell} = \mathbf{1}_{\{i=\ell\}}$ - fittest type wins (e.g. [41]),
(2) $p_{i,\ell} = \mathbf{1}_{\{i \leq \lfloor \ell/2 \rfloor\}}$ - minority rule, (4) $p_{i,\ell} = i/\ell$ - uniform rule.

In general, if $p_{i,\ell} \in \{0,1\}$ for all $\ell \in]m]$, $i \in [\ell]_0$, we call the rule deterministic. The reason for this name is that in such replacement events, the type of the descendent is a deterministic function of the types of the potential parents. In particular, (1) − (3) are deterministic.

2.2. **Selection decomposition.** In order to give a complete answer to (Q1'), it remains to identify the polynomials s for which the SDE (1.2) can be expressed as (2.1). This leads to the notion of selection decomposition.

Definition 2.3 (Selection decomposition)**.** For $m \in \mathbb{N} \setminus \{1\}$, define the mapping $\mathcal{B}:\ E^m \to \mathbb{R}^{m-1}$

$$(\beta, p) \mapsto (\mathcal{B}_i(\beta, p))_{i=1}^{m-1},$$

where $\mathcal{B}_i(\beta, p)$ is the i-th Bernstein coefficient in the Bernstein basis of degree m of the polynomial

$$\sum_{\ell=2}^{m} \beta_\ell \left\langle B_\ell(\cdot), p_{\cdot,\ell} - u_\ell \right\rangle,$$

where $u_\ell := (i/\ell)_{i=0}^{\ell}$. Any element of $\mathcal{E}_d := \mathcal{B}^{-1}(\rho(d))$ is called a selection decomposition (SD) of the polynomial d with $\deg(d) = m$. Similarly, we say that $(\beta, p) \in E^m$ is a selection decomposition of $v \in \mathbb{R}^{m-1}$ if $v = \mathcal{B}(\beta, p)$.

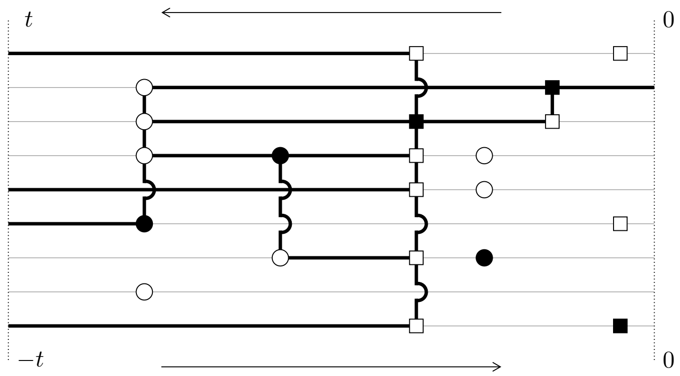

FIGURE 2. A realisation of the Moran interacting particle system (thin lines) for a population of size $N = 9$ and the embedded ASG (bold lines) for a sample of size 1. Time runs forward in the Moran model ($\rightarrow$) and backward in the ASG ($\leftarrow$). Backward time t corresponds to forward time $-t$. Circle represent (neutral) reproduction events. Squares represent (selective) replacement events. Backward in time, the potential parents that are involved in a r-reproduction event merge into a single ancestor (black circle). In contrast, in a ℓ-replacement event, the single potential parent (black square) branches into ℓ (not necessarily new) potential parents.

Theorem 2.4. *Let d be a non zero polynomial such that $d(0) = d(1) = 0$. The set $\mathcal{E}_d$ is infinite.*

Note that the condition $d(0) = d(1) = 0$ is equivalent to say that d can be expressed as $d(x) = x(1-x)s(x)$ for some polynomial s of degree $m-2$. Hence, the previous theorem implies that for any polynomial s there are infinitely many ways of expressing (1.2) in the form of (2.1). In particular, there are infinitely many (substantially different) sequences of Moran models converging to the solution of (1.2).

A proof for Theorem 2.4 is provided in Section 3.1. The proofs of Theorem 2.1 and Corollary 2.2 as well as some other complementary results are given in Section 3.2.

2.3. **Ancestry in finite populations.** In this section, we answer question (Q1) at the level of finite populations. We proceed in a heuristic manner. The hope is that this provides some intuition for the definitions at the infinite population level, see Definition 2.5 and 2.7 below.

For $(\beta, p) \in E^m$ and $\mu \in \mathcal{M}_f([N]_0)$, the (β, p, μ)-Moran model admits a natural graphical representation as an interactive particle system, see Fig. 2. Here, individuals are represented by pieces of horizontal lines. Forward time runs from left to right. Squares indicate which lines are involved in an ℓ-replacement ($\ell \in]m]$): a black square marks the individual that is replaced, and white squares mark the other potential parents. Circles indicate which lines are involved in a r-reproduction (for $r \in [N]$): a black circle marks the individual that reproduces and white circles mark the individuals that are replaced by its offspring. These graphical elements arise in the picture according to the arrival times of independent Poisson processes (the rates can be worked out easily from the definition of the Moran model).

So far, this procedure provides a construction of an untyped particle system. Given an initial type configuration, types propagate forward in time along the untyped particle system according to the (random) colouring procedure described in Section 2.1.

Genealogical structures are extracted from the particle picture as follows. Start with a sample of n individuals that is chosen at time $t = 0$ and trace back the set of their potential ancestors by reading the graphical picture from right to left, see Fig. 2. Assume there are currently n potential ancestors in the graph.

Then a k-reproduction has the following effect backward in time. If k potential ancestors simultaneously encounter circles, they merge into one and take the place of the individual marked with a black circle. In particular, the number of potential ancestors decreases to $n - k + 1$. On the other hand, the effect of a ℓ-replacement backward in time is as follows ($\ell \in]m]$). If a potential ancestor encounters a black square, we add all the lines marked with a white square (in this case $\ell - 1$) to the set of potential ancestors. In particular, if all the $\ell - 1$ white squares are outside the set, the number increases to $n + \ell - 1$.

Applying this procedure up to backward time t leads to a generalisation of the ancestral selection graph (ASG) of Krone and Neuhauser [41], see Fig. 2. In order to extract the true genealogy and the types of the individuals in the sample at time 0, we assign types to the lines in the ASG at backward time t in an exchangeable manner according to the initial type distribution and propagate them along the lines forward up to time 0 subject to the colouring rule p, see Fig. 3.

2.4. **Ancestral selection graph.** Fix a measure $\Lambda \in \mathcal{M}_f([0,1])$ and a polynomial s of degree $m-2$, together with a SD $(\beta, p) \in \mathcal{E}_d$ of $x \mapsto d(x) := x(1-x)s(x)$. Consider the sequence of (β^N, p, μ_N)-Moran models defined in Corollary 2.2. We are interested in the behaviour of the ASG described in the previous section for large N and time sped up by N. First, note that the probability that a branching event involves more than one line already present in the ASG is small (it vanishes as $N \to \infty$). Moreover, in this situation, it is not hard to see that: (1) each line branches into ℓ lines at a rate which is close to β_ℓ, and (2) a given group of k lines merges into one at a rate close to

$$\lambda_{n,k} := \int_{[0,1]} r^{k-2}(1-r)^{n-k}\Lambda(dr) = \Lambda(\{0\})\mathbf{1}_{\{k=2\}} + \lambda^0_{n,k}, \quad n \geq k \geq 2.$$

The previous discussion leads us to the following definition.

Definition 2.5 (ASG)**.** Let $(\beta, p) \in \mathcal{E}_d$ and $\Lambda \in \mathcal{M}_f([0,1])$. Let $n \in \mathbb{N}$ and consider the branching-coalescing particle system $\mathcal{G} = (\mathcal{G}_t)_{t\geq 0}$ starting with n particles at time 0 and the following dynamic.

- For $\ell \in]m]$, each particle branches at rate β_ℓ into ℓ particles.
- If the current number of particles is $n \geq 2$, then for $k \in]n]$, every k-tuple of particles coalesce into a single particle at rate $\lambda_{n,k}$.

Then the (β, p, Λ)-ASG in $[0,t]$ starting from a sample of size n is the pair $(\mathcal{G}_t, p)$, where $\mathcal{G}_t$ is the realisation of the particle system in $[0,t]$ and p is the colouring rule (see also Definition 4.1 for a more detailed description of the branching-coalescing system).

Remark 2.4. A realisation $\mathcal{G}_t$ of the ASG in $[0,t]$ equipped with the genealogical order can be understood as a directed acyclic graph (DAG) (see Definition 4.1). We refer to the particles in $\mathcal{G}_t$ that are present at time 0 as roots (sometimes called sources in graph theory) and to the ones at time t as leaves (or sinks).

Remark 2.5. For any $(\beta, p, \Lambda) \in \mathcal{E}_d \times \mathcal{M}_f([0,1])$, the (β, p, Λ)-ASG provides a natural genealogy to the SDE (1.2). From Theorem 2.4, we can conclude that, as announced in the introduction, one can associate infinitely many different genealogies to the same forward model.

The number of lines in the ASG plays a crucial role in the analysis of the type-frequency process in the cases in which a moment duality is available. Indeed, also in our framework this quantity will be important.

Definition 2.6 (Leaf process)**.** For each $t \geq 0$, denote by L_t the number of leaves in $\mathcal{G}_t$. We refer to $L := (L_t; t \geq 0)$ as the leaf process. It is a continuous-time Markov chain with the following transition rates.

(1) For $n \in \mathbb{N}$ and every $\ell \in [m]$

$$n \to n + \ell - 1 \quad \text{at rate} \quad n\beta_\ell.$$

(2) For $n \in \mathbb{N} \setminus \{1\}$ and every $k \in]n]$,

$$n \to n - k + 1 \quad \text{at rate} \quad \binom{n}{k}\lambda_{n,k}.$$

2.5. **Ancestral selection polynomial.** The ASG introduced in the previous section is a rather cumbersome object. We condense the information carried by the ASG to a minimum by tracking only the sample composition(s) as one goes backward in time in the ancestral structure. In order to do so, we introduce the ancestral selection polynomial, which will be the cornerstone of Bernstein duality exposed in the next section.

Definition 2.7 (Ancestral selection polynomial)**.** The (β, p, Λ)-*ancestral selection polynomial (ASP)* is the function $x \in [0,1] \mapsto P_t(x)$, where $P_t(x)$ is the probability that all the roots of $\mathcal{G}_t$ are of type a if each leaf of $\mathcal{G}_t$ is of type a with probability x (resp. of type A with probability $1-x$), conditional

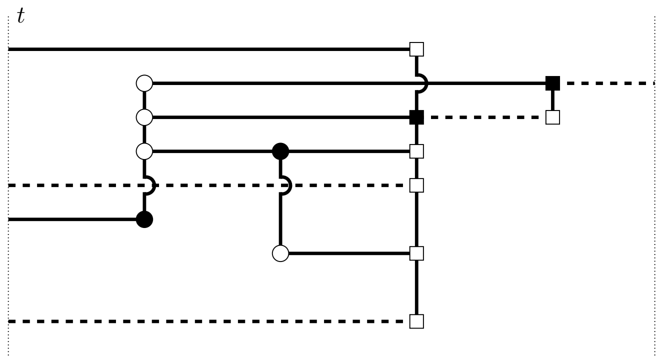

FIGURE 3. Colouring of the ASG of Fig 2 under the minority rule, i.e. $p_{i,\ell} = \mathbf{1}_{\{i\leq \lfloor \ell/2\rfloor\}}$. Assign colours to the lines at (backward) time t and propagate them through the graph up to time 0 by recursively applying the minority rule. At every replacement event, the colour of the replaced individual (black square) is determined by the colour of the potential parents (black and white squares). At every reproduction event, the offspring (white circles) inherit the colour of the parent (black circle).

on the observation of the leaf process $(L_s; s \in [0,t])$. It is assumed that the initial type assignment is independent for each leaf and that $\mathcal{G}_t$ is typed from the leaves to the roots applying the colouring rule p. (See also Definition 4.3 for a more precise definition.)

The following linear operators describe the effect of branching and coalescence events to the Bernstein coefficient vector associated to the ASP.

Definition 2.8 (Selection and coagulation matrices)**.** Fix $(\beta, p) \in E^m$. For every $n \in \mathbb{N}_0 := \mathbb{N}\cup\{0\}$, define the following linear operators.

(1) For every $\ell \in]m]$, let $S^{n,\ell} : \mathbb{R}^{n+1} \to \mathbb{R}^{n+\ell}$ with

$$S^{n,\ell} v \ := \ \left(\mathbb{E}\left[\, p_{K_i,\ell}\, v_{i+1-K_i} \ + \ (1-p_{K_i,\ell})\, v_{i-K_i}\right]\right)_{i=0}^{n+\ell-1},$$

where $K_i \sim \mathrm{Hyp}(n+\ell-1, i, \ell)$ and $v = (v_i)_{i=0}^n \in \mathbb{R}^{n+1}$

(2) For $k \in]n]$, let $C^{n,k} : \mathbb{R}^{n+1} \to \mathbb{R}^{n-k+2}$ with

$$C^{n,k} v \ := \ \left(\frac{i}{n-k+1} v_{i+k-1} + \left(1 - \frac{i}{n-k+1}\right) v_i\right)_{i=0}^{n-k+1},$$

where $v = (v_i)_{i=0}^n \in \mathbb{R}^{n+1}$.

The above operators define the transitions of a Markov process on $\mathbb{R}^\infty := \cup_{n\in\mathbb{N}}\mathbb{R}^n$ that codes codes the evolution of the Bernstein coefficient vector of the ASP. (See Remark 4.4 for details on the topology.)

Definition 2.9 (Bernstein coefficient process)**.** The Bernstein coefficient process is the Markov process $V := (V_t;\, t\geq 0)$ on $\mathbb{R}^\infty$ with the following transition rates.

(1) For $v \in \mathbb{R}^{n+1}$ and for every $\ell \in]m]$,

$$v \to S^{n,\ell} v \quad \text{at rate} \quad n\beta_\ell.$$

(2) For $v \in \mathbb{R}^{n+1}$ and for every $k \in]n]$,

$$v \to C^{n,k} v \quad \text{at rate} \quad \binom{n}{k}\lambda_{n,k}.$$

For any $t \geq 0$, set $L_t := \dim(V_t) - 1$ and denote by $V_t(i)$, for $i \in [L_t]_0$, the i-th coordinate of the vector V_t.

Remark 2.6. Note that if we start the process V at $V_0 = e_{n+1} := (0,\ldots,0,1)^T \in \mathbb{R}^{n+1}$, then the process L from the previous definition is by construction equal in law to the leaf process with parameters (β, Λ) started at n. This fact legitimises the abuse of notation in the use of L for both processes. Furthermore, note that for any $t > 0$, the value V_t is a deterministic function of the initial value V_0 and the realisation of $(L_s; s\in[0,t])$.

A precise relation between the Bernstein coefficient process and the ASP is provided by the following proposition.

Proposition 2.10. *For all $x \in [0,1]$ and $n \in \mathbb{N}$, the ASP $(P_t(x);\, t \geq 0)$ with initial condition x^n (corresponding to the ASG with n roots of type a) is identical in law to $(\langle B_{L_t}(x), V_t\rangle\,;\, t \geq 0)$ with initial condition $V_0 = e_{n+1}$.*

2.6. **Bernstein duality.** In this section, we address question (Q2). In particular, we extend the moment duality to what we call a *Bernstein duality.* Let us first explain the main idea behind this type of duality. Consider the solution of the SDE (1.2) starting at $X_0 = x$, with s being a polynomial of degree $m-2$ and $\Lambda \in \mathcal{M}_f([0,1])$. Consider the evolution of the underlying population model up to (forward) time t and independently sample at this point n individuals. The probability that they are all of type a, conditional on the observation of X_t, is X_t^n. Now, in order to approach the problem from a backward perspective, consider a SD $(\beta, p) \in \mathcal{E}_d$ of $x \mapsto d(x) = x(1-x)s(x)$, and run the corresponding branching-coalescing system starting with the n sampled individuals up to (forward) time 0. Assign types independently to the leaves of $\mathcal{G}_t$ according to the type distribution $(x, 1-x)$. The probability that the n sampled individuals are of type a, conditional on the observation of the leaf process in $[0,t]$, is by definition $P_t(x)$. After averaging over all possible observations, this intuitive argument suggests that $\mathbb{E}_x[X_t^n] = \mathbb{E}_{x^n}[P_t(x)]$. The next theorem makes this heuristic argument precise.

Theorem 2.11 (Bernstein duality)**.** *The processes $(X_t; t \geq 0)$ and $(V_t; t \geq 0)$ are dual with respect to the duality function $(x,v) \mapsto \langle B_{\dim(v)-1}(x), v\rangle$, i.e. for all $t > 0$,*

$$\mathbb{E}_x\left[\langle B_n(X_t), v\rangle\right] \;=\; \mathbb{E}_v\left[\langle B_{L_t}(x), V_t\rangle\right], \qquad \forall x \in [0,1],\, \forall n \in \mathbb{N},\, \forall v \in \mathbb{R}^{n+1}. \tag{2.2}$$

The following corollary illustrates that the Bernstein duality generalises the moment duality (1.1) to arbitrary polynomial selection term.

Corollary 2.12. *Consider the process V started at $V_0 = e_{n+1}$.*

(1) For all $x \in [0,1]$ and $t \geq 0$

$$\mathbb{E}_x\left[X_t^n\right] \;=\; \mathbb{E}_{e_{n+1}}\left[\langle B_{L_t}(x), V_t\rangle\right]. \tag{2.3}$$

(2) Let $m \geq 2$ and assume that d is of the form

$$d(x) \;=\; -x(1-x)\sum_{i=0}^{m-2} s_i x^i, \quad x \in [0,1], \tag{2.4}$$

where $(s_i)_{i=0}^{m-2}$ is a decreasing sequence valued in $\mathbb{R}_+$. There is a SD $(\beta,p) \in \mathcal{E}_d$ such that $\langle B_{L_t}(x), V_t\rangle = x^{L_t}$, for all $t \geq 0$. In particular, the Bernstein duality coincides with moment duality, i.e.

$$\mathbb{E}_x\left[X_t^n\right] \;=\; \mathbb{E}_n\left[x^{L_t}\right].$$

Remark 2.7. The specific class of selection terms exposed in (2.4) was already studied in [32] and contains the classical case of genic selection.

Remark 2.8. Dualities for diffusion processes in population genetics have been studied by Shiga [52] (see his Lem. 2.1, Lem. 2.2). His approach leads to a moment duality penalised by a Feynmann–Kac term. In contrast, our dual process has a larger state space, a more involved duality function, but no Feynman-Kac correction.

Remark 2.9 (Mutation-selection models)*.* The particular form of our drift term excludes models with mutations. In particular cases of selection, [6, 21, 22] obtain a weighted moment duality. [34] further extend this to models with recombination. In the diploid mutations-selection equation (i.e. if $\Lambda = 0$ in (1.2) and s is a specific linear polynomial), [4] formulate an ASG-based dual process that takes value in the weighted ternary trees. We believe that by introducing suitable operators that reflect mutations in the ancestral structures, the Bernstein duality translates to this framework.

The proof of Proposition 2.10, connecting the ancestral polynomial to the Bernstein coefficient process, is given in Section 4.2. The proof of the Bernstein duality (Theorem 2.11) and Corollary 2.12 are provided in Section 4.3.

2.7. **Properties of the Bernstein coefficient process and its leaf process.** In this section, we expose some of the most relevant properties of the Bernstein coefficient process and its leaf process. Recall that the leaf process depends only on the branching rates β and on the Λ-measure, and not on the colouring rule p. It turns out that the following two quantities play a crucial role in the analysis of X and V.

Definition 2.13 (Effective branching rate and coalescence impact)**.** For a SD $(\beta, p) \in \mathcal{E}_d$, define the effective branching rate as

$$b(\beta) \ := \ \sum_{\ell=2}^{m} \beta_\ell(\ell - 1).$$

For $\Lambda \in \mathcal{M}_f([0,1])$, define its coalescence impact as

$$c(\Lambda) \ = \int_{[0,1]} \log(1-r) \frac{\Lambda(dr)}{r^2}.$$

Remark 2.10. The quantity $c(\Lambda)$ was introduced in [36] as $\lim_{k\to\infty} \log(k)/\mathbb{E}_k[T_1']$, where T_1' is the absorption time of the Λ-coalescent in 1.

2.7.1. *Long time behaviour: invariant distributions.* In this part, we are interested in the long time behaviour of the processes V and L. We start with the following simple criteria for positive recurrence or transience of the leaf process.

Theorem 2.14 (Classification)**.** *Assume $\Lambda \neq \delta_1$. The leaf process L with parameters (β, Λ) is*

- *positive recurrent if $b(\beta) < c(\Lambda)$,*
- *transient if $b(\beta) > c(\Lambda)$.*

Remark 2.11. If $\Lambda = \delta_1$, the communication class of 1 is always positive recurrent, see Corollary 5.5. If $\beta_2 > 0$, this communication class is $\mathbb{N}$.

Remark 2.12. Note that $c(\Lambda) = \infty$ for the Kingman model ($\Lambda = \delta_0$) and the Bolthausen-Szitman model ($\Lambda = U[0,1]$, see [11]). Therefore, for these models the leaf process is always positive recurrent. In contrast, for the Eldon-Wakeley coalescent ($\Lambda = \delta_c$ for some $c \in (0,1)$, see [17]), we have $c(\Lambda) < \infty$.

Remark 2.13. The first part of Theorem 2.14, under the assumption $c(\Lambda) < \infty$, is already present in [32, Thm. 4.6]. Furthermore, they show that if $b(\beta) > c(\Lambda)$, then the process L is not positive recurrent.

A consequence of the previous result is that if $b(\beta) < c(\Lambda)$, then the leaf process admits a unique stationary distribution, which is characterised as the solution of a linear system of equations (see Eq. (5.4)). The latter is a generalisation of the well-known Fearnhead recursions, which were introduced in [24] (see also [53]) for a Wright–Fisher diffusion model and later extended to the Λ-Wright–Fisher model in [3].
The next result tells us that the condition $b(\beta) < c(\Lambda)$ also assures the existence of a stationary distribution for the Bernstein coefficient process.

Proposition 2.15 (Invariant distributions)**.** *The Bernstein coefficient process V keeps the entries $V_t(0)$ and $V_t(L_t)$ constant along time. Moreover, if $b(\beta) < c(\Lambda)$, then the following assertions hold.*

1. *For every a, b, the Bernstein coefficient process V has a unique invariant probability measure $\mu^{a,b}$ with support included in $\{v \in \mathbb{R}^\infty : \ v_0 = a, \, v_{\dim(v)-1} = b\}$.*
2. *Let $V_\infty^{a,b}$ be a random variable with law $\mu^{a,b}$. If $V_0 = v$ with $v_0 = a$ and $v_L = b$, then*

$$V_t \ \xrightarrow[t\to\infty]{(d)} \ V_\infty^{a,b}.$$

The proofs of Theorem 2.14 and Proposition 2.15 are given in Section 5.1. Other complementary results on the stationary distribution of the leaf process are given in Section 5.2.

2.7.2. *Small time behaviour: coming down from infinity.* Now, we turn our attention to the small time behaviour of the leaf process as the initial number of particles tends to infinity. More precisely, let $(L_t^n;\, t \geq 0)$ denote the leaf process started at n. It is not hard to see that this process is monotone in n, i.e. there exists a coupling such that for every $n > m$ and for all $t \geq 0$,

$$L_t^n \geq L_t^m \quad \text{a.s..}$$

As a consequence, we can always define a process $(L_t^\infty;\, t > 0)$ valued in $\mathbb{N}_0 \cup \{\infty\}$ as the monotone limit of the sequence processes $\{(L_t^n;\, t > 0)\}_{n \in \mathbb{N}_0}$. Note that $\lim_{t \to 0^+} L_t^\infty = \infty$.

Definition 2.16 (Coming down from infinity)**.** We say the the leaf process comes down from infinity (c.d.i.) if and only if for every $\varepsilon > 0$, $L_\varepsilon^\infty < \infty$ a.s.. We say that it stays infinite if for every $\varepsilon > 0$, $L_\varepsilon^\infty = \infty$ a.s..

It follows from Remark 2.10 that if the Λ-coalescent c.d.i., then $c(\Lambda) = \infty$. In particular, the corresponding leaf process is then positive recurrent. In fact, the following generalisation of a result of Pitman [50, Prop. 23] holds.

Theorem 2.17 (Criterion for c.d.i.)**.** *Assume that Λ has no mass at* 1*. Then the leaf process L either c.d.i. or stays infinite. Furthermore, L c.d.i. if and only if the underlying Λ-coalescent c.d.i..*

Moreover, if the leaf process c.d.i., then $+\infty$ is an entrance point for the Bernstein coefficient process. This is the content of the following proposition.

Proposition 2.18 (Entrance law at ∞)**.** *Assume that the leaf process L c.d.i.. Let $(V_t^n;\, t \geq 0)$ be the Bernstein coefficient process with initial condition $V_0^n = e_{n+1}$.*

(1) There exists $(V_t^\infty;\, t > 0)$ such that

$$(V_t^n;\, t > 0) \xrightarrow[n \to \infty]{(d)} (V_t^\infty;\, t > 0),$$

where $\mathbb{D}([0,\infty); \mathbb{R}^\infty)$ is endowed with the Skorokhod topology on any closed time interval $[t_1, t_2] \subset (0, \infty)$.

(2) Furthermore, for every $t_0 > 0$, conditioned on $V_{t_0}^\infty$, $(V_{t+t_0}^\infty;\, t \geq 0)$ is distributed as $(V_t; t \geq 0)$ with initial condition $V_{t_0}^\infty$.

The proofs of the results of the this section are presented in Section 5.4

2.8. **Absorption probability and time to absorption.** If the leaf process and the type-a frequency process are in moment duality, one can typically translate the long time behaviour of L into (time) asymptotic properties of X, see [26, 32]. This method extends to the Bernstein duality and allows us the derivation of results on absorption probabilities and on the time to fixation.
Define $T_i := \inf\{t > 0 : X_t = i\}$ for $i \in \{0, 1\}$. Furthermore, define $h(x) := \mathbb{P}_x(T_0 < T_1)$. Using the notation of Proposition 2.15, set $V_\infty := V_\infty^{0,1}$ and $L_\infty := L_\infty^{0,1}$.

Proposition 2.19 (Absorption probabilities)**.** *Assume $b(\beta) < c(\Lambda)$.*

(1) $T_0 \wedge T_1 < \infty$ almost surely.

(2) For all $x \in [0,1]$,

$$h(x) \;=\; \mathbb{E}\left[\langle B_{L_\infty}(x), V_\infty \rangle\right]. \tag{2.5}$$

(3) The boundary points 0 *and* 1 *are accessible from any point $x \in (0,1)$, i.e. $h(x) \in (0,1)$.*

Remark 2.14. Note that (2.5) can be expressed as

$$h(x) \;=\; \sum_{\ell=0}^{\infty} \mathbb{P}(L_\infty = \ell) \sum_{i=0}^{\ell} d_{i,\ell}\, b_{i,\ell}(x),$$

where $d_{i,\ell} := \mathbb{E}[\, V_\infty(i) \mid L_\infty = \ell]$. Moreover, under the extra assumption that L_∞ admits exponential moments, we obtain a series expansion of h around 0. See Proposition 6.2 for more details. This is similar in spirit to Baake et al. [3, Thm. 2.4].

Given a colouring rule $p \in \mathcal{P}_m$, define the colouring rule $\bar{p} \in \mathcal{P}_m$ with $\bar{p}_{i,\ell} := 1 - p_{\ell-i,\ell}$. Consider the Bernstein coefficient processes $(V_t; t \geq 0)$ and $(W_t; t \geq 0)$ with parameters (β, p, Λ) and $(\beta, \bar{p}, \Lambda)$, respectively. Both processes can be constructed on the basis of the same realisation of the leaf process L. The distribution of the fixation time of X can be related to the limiting ASP with initial condition x^n for $n \to \infty$ via the duality.

Proposition 2.20 (Absorption time)**.** *Assume the leaf process L c.d.i.. Define the random polynomial*

$$Q_t^\infty(x) \;:=\; \left\langle B_{L_t^\infty}(x), V_t^\infty \right\rangle + \left\langle B_{L_t^\infty}(1-x), W_t^\infty \right\rangle,$$

with V^∞ and W^∞ defined as in Proposition (2.18)*. Let $T := T_0 \wedge T_1$ be the absorption time of X and $\tau^{(\infty)} = \inf\{t > 0 \;:\; L_t^\infty = 1\}$. Then*

$$\begin{aligned} \mathbb{P}_x(T \leq t) \;&=\; \mathbb{E}\left[Q_t^\infty(x)\right], \\ \mathbb{E}_x[T] \;&=\; \mathbb{E}\left[\int_0^{\tau^{(\infty)}} (1 - Q_t^\infty(x)) dt\right]. \end{aligned}$$

Complementary properties and the proofs of the results of this section can be found in Section 6.

2.9. **Minimality.** It follows from the results of Section 2.1 and 2.4, that infinitely many genealogies are associated to the SDE (1.2) with polynomial drift term of the form $x(1-x)s(x)$. Among the possible genealogies, we want to identify the ones with good properties. In view of Theorem 2.14 and Proposition 2.15, it seems natural to distinguish genealogies according to their effective branching rate.

Definition 2.21 (λ-decomposability)**.** Let $m \in \mathbb{N} \setminus \{1\}$ and $\lambda > 0$. We say that $v \in \mathbb{R}^{m-1}$ is λ-decomposable if it admits a SD with effective branching rate λ. Denote by $\mathcal{S}_\lambda \subset \mathbb{R}^{m-1}$ the set of λ-decomposable vectors, i.e.

$$\mathcal{S}_\lambda \;:=\; \mathcal{B}\left(\{(\beta, p) \in E^m \;:\; b(\beta) = \lambda\}\right).$$

In other words, a vector $v \in \mathbb{R}^{m-1}$ is λ-decomposable if there is a SD with effective branching rate λ for the polynomial $\left\langle B_m(\cdot), (0, v^T, 0)^T \right\rangle$.

Definition 2.22 (Minimal selection decomposition)**.** Let d be a polynomial with $d(0) = d(1) = 0$. We say a SD $(\beta, p) \in \mathcal{E}_d$ is minimal (for d) if

$$b(\beta) = \inf_{(\beta', p') \in \mathcal{E}_d} b(\beta') =: b_\star(d).$$

We call $b_\star(d)$ the minimal effective branching rate of d.

Proposition 2.23. *For every polynomial d with $d(0) = d(1) = 0$, there exists a minimal SD and the minimal effective branching rate satisfies the following relation*

$$b_\star(d) \;=\; \inf\{\lambda > 0 : \rho(d) \in \lambda \mathcal{S}_1\}. \tag{2.6}$$

In order to prove this result, it will be crucial to understand the structure of the set $\mathcal{S}_\lambda$. It turns out that $\mathcal{S}_\lambda$ is a polytope with the property that $\mathcal{S}_\lambda = \lambda \mathcal{S}_1$ (see Proposition 3.1). An in-depth analysis of $\mathcal{S}_1$ allows us to prove the existence of a minimal SD and provides a geometrical recipe to find it (see Algorithm 7.8). Fig. 4 illustrates $\mathcal{S}_1$ for $\deg(d) = 3$. In this case, the faces of the polygon $\mathcal{S}_\lambda$ have a natural biological interpretation. More specifically, we show that if d corresponds to the selection term in a diploid Wright–Fisher diploid model with dominance (see 7.3), then the face of $\mathcal{S}_{b_\star(d)}$ that contains $\rho(d)$ depends on the nature of the dominance at hands (under- or overdominance). Section 7.2 exposes more details.

A classification of SDs according to their effective branching rates only provides a partial understanding of the connection between different ASGs. One is inclined to say that an ASG $\mathcal{G}$ is better than $\tilde{\mathcal{G}}$ if one recovers $\mathcal{G}$ by erasing superfluous branches from $\tilde{\mathcal{G}}$. This motivates the following definitions.

Definition 2.24 (Thinning)**.** A lower-triangular stochastic matrix $\mathbb{T} := (\mathbb{T}_{k,i})_{k,i=1}^m$ is called a thinning mechanism. For $\beta \in \mathbb{R}^{m-1}$, define $\mathbb{T}\beta \in \mathbb{R}^{m-1}$ as

$$(\mathbb{T}\beta)_\ell := \sum_{k=\ell}^m \beta_k \mathbb{T}_{k,\ell}.$$

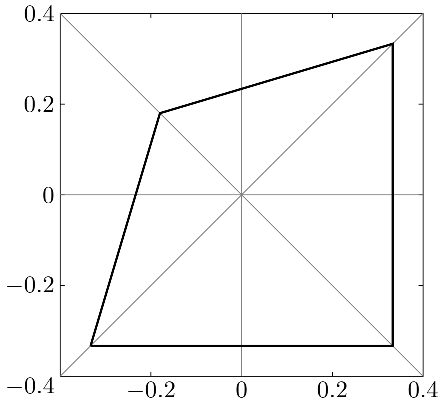

FIGURE 4. The region delimited by the bold lines corresponds to $\mathcal{S}_1$ for $\deg(d) = 3$.

Note that $\mathbb{T}_{k,\cdot}$ gives rise to a probability measure on $[k]$.

Definition 2.25 (Graph minimal)**.** Let $(\beta, p) \in \mathcal{E}_d$. We say that (β, p) is graph-minimal if and only if there is no thinning mechanism $\mathbb{T}$ and no colouring rule p' such that $(\mathbb{T}\beta, p') \in \mathcal{E}_d$.

We are now ready to state the main result of this section

Theorem 2.26. *Any minimal SD $(\beta, p) \in \mathcal{E}_d$ is graph-minimal.*

For $\deg(d) = 3$, we prove that the converse is also true. In particular, this shows that in this case the minimal genealogies are the only relevant ones, i.e. they are the only ones that do not contain any superfluous branches.

Proposition 2.27. *Assume that* $\deg(d) = 3$*. For every $(\beta, p) \in \mathcal{E}_d$ with $b(\beta) > b_\star(d)$, there is a thinning mechanism $\mathbb{T}$ and a colouring rule p' such that $(\mathbb{T}\beta, p') \in \mathcal{E}_d$ is a minimal SD.*

Remark 2.15. If we consider only genealogies that minimise the effective branching rate, we recover classical cases of the literature. For instance, the ASG of Krone and Neuhauser as the only minimal dual to the Wright–Fisher diffusion with genic selection [41, 47]; and finally the minority rule of Neuhauser as the only minimal dual of balancing selection [46].

2.10. **Open questions.** We list several open questions that stem from our work.

(1) The present article only deals with a drift term d of polynomial form that vanishes at the boundary. We conjecture that for any continuous function d with $d(0) = d(1) = 0$, there exist an infinite SD (β, p) such that
$$d(x) = \sum_{\ell=2}^{\infty} \beta_\ell \sum_{i=0}^{\ell} \left(p_{i,\ell} - \frac{i}{\ell} \right) b_{i,\ell}(x).$$
For such a decomposition, the results of the present paper would easily extend to ASGs with arbitrary large branching events.

(2) We showed that if the leaf process is positive recurrent, then both 0 and 1 are accessible. Is the reverse true, i.e. assuming that the (minimal) leaf process is transient, are 0 or 1 not accessible? This was answered positively by [32], when the coefficients of s are negative and non-increasing. The general case seems more involved.

(3) Along the same lines, we showed that if the leaf process c.d.i. then the fixation time has finite expected value. What about the converse ?

(4) We constructed the ASG of a selection model by tracing the set of potential parents backward in time. The true genealogy of the process can be recovered by only keeping track of the actual genealogical lines as in [53] (i.e. by removing all the "unused" potential parents in the ASG). As discussed earlier, there are infinitely many ASGs for a given selection model. However, are the actual genealogies embedded in those ASGs identical?

(5) If $\deg(d) = 2, 3$, we prove that the minimal ASGs are also graph-minimal. What about higher dimension?
(6) Mutation-selection models usually have a drift term that does not vanish at the boundary. We conjecture that by including a suitable transition of the Bernstein coefficient process, the Bernstein duality translates also to this setting.
(7) Barbour et al. [6] use a moment duality to derive a transition function expansion for a Wright–Fisher model with mutation and selection (see also [22]). If the the Bernstein duality translates also to this setting, does it allow us to generalise such a transition function expansion to general selection models?
(8) Duality methods have been successfully used in the study of spatial models with selection (see, e.g., [19, 20]). It would be interesting to see if the concepts of selection decomposition and Bernstein duality translate to this setting, and if so, whether they can be used to treat more general selection mechanisms.

3. From selection decompositions to selection mechanisms

3.1. **Existence and non-uniqueness of selection decompositions.** In this section, we show that every SDE (1.2) can be expressed in the form of the SDE (2.1). This boils down to show that every drift term in Eq. (1.2) admits a SD (see Definition 2.3). In fact, Theorem 2.4 states that every drift term d of the form $d(x) := x(1-x)s(x)$, $x \in [0,1]$, with s being a polynomial, admits infinitely many SDs. The following proposition is useful for its proof.

Proposition 3.1 (Scaling property). *For any $\lambda > 0$ the set $\mathcal{S}_\lambda$ is a polytope with the property that $\mathcal{S}_\lambda = \lambda\mathcal{S}_1$. In particular, the minimal effective branching rate satisfy the relation* (2.6).

Proof. The sets $G_\lambda := \{\beta \in \mathbb{R}_+^{m-1} : b(\beta) = \lambda\}$ and $\mathcal{P}_m$ are polytopes. Thus, $G_\lambda \times \mathcal{P}_m$ is a polytope, since it is the Cartesian product of polytopes. For every $p \in \mathcal{P}_m$, the map $\beta \mapsto \mathcal{B}(\beta, p)$ is linear and for every $\beta \in \mathbb{R}_+^{m-1}$, the map $p \mapsto \mathcal{B}(\beta, p)$ is affine. It follows that $\mathcal{S}_\lambda$ is a polytope. The property $\mathcal{S}_\lambda = \lambda\mathcal{S}_1$ is a straightforward consequence of the fact that $G_\lambda = \lambda G_1$. Finally, using the definitions of $\mathcal{S}_\lambda$, $\lambda > 0$, and of $b_\star(d)$, we obtain

$$b_\star(d) = \inf\{\lambda > 0 : \rho(d) \in \mathcal{S}_\lambda\}.$$

Hence, (2.6) follows from the scaling property $\mathcal{S}_\lambda = \lambda\mathcal{S}_1$. □

The previous result provides a geometric framework to prove Theorem 2.4 and Proposition 2.23, i.e. the existence of infinitely many SDs and the existence of a minimal decomposition, for any given polynomial drift.

Proof of Theorem 2.4. Fix an arbitrary colouring rule $p \in \mathcal{P}_m$. For every $w \in \{0,1\}^{m-1}$, define the colouring rule p^w by replacing $(p_{i,m})_{i=1}^{m-1}$ with w; keep the other entries unchanged. Furthermore, set $a := (a_\ell)_{\ell=2}^m \in \mathbb{R}^{m-1}$ with $a_m := 1/(m-1)$ and $a_\ell := 0$ for $\ell \neq m$. A straightforward calculation yields

$$\mathcal{B}(a, p^w) = \frac{w}{m-1} - \frac{c}{m(m-1)},$$

where $c := (i)_{i=1}^{m-1}$. Since $b(a) = 1$ and $\mathcal{S}_1$ is a polytope,

$$\operatorname{conv}\left(\left\{0, \frac{1}{m-1}\right\}^{m-1}\right) - \frac{c}{m(m-1)} \subset \mathcal{S}_1.$$

It follows that $\mathcal{S}_1$ contains an open neighbourhood of the origin of $\mathbb{R}^{m-1}$. From the scaling relation $\mathcal{S}_\lambda = \lambda\mathcal{S}_1$, we infer that

$$\mathbb{R}^{m-1} \;=\; \cup_{\lambda>0}\mathcal{S}_\lambda. \tag{3.1}$$

Now, let d be a polynomial of degree m with $d(0) = d(1) = 0$. We know that d admits a SD with effective branching rate λ if and only if $\rho(d) \in \mathcal{S}_\lambda$. From (3.1) we deduce that there is λ_0 such that $\rho(d) \in \mathcal{S}_{\lambda_0}$. Using again the scaling relation, we deduce that $\rho(d) \in \mathcal{S}_\lambda$ for all $\lambda \geq \lambda_0$. The result follows. □

Proof of Proposition 2.23. One consequence of (3.1) is that

$$\mathbb{R}^{m-1} = \cup_{\lambda>0} \partial \mathcal{S}_\lambda. \tag{3.2}$$

Thus, for any polynomial d of degree m with $d(0) = d(1) = 0$, there is $\lambda_\star(d)$ such that $\rho(d) \in \partial \mathcal{S}_{\lambda_\star(d)}$. It follows that there is $(\beta, p) \in \mathcal{E}_d$ with $b(\beta) = \lambda_\star(d)$. Moreover, by construction $\lambda_\star(d) = b_\star(d)$. Hence, the first part of the statement follows. The second part was already proved in Proposition 3.1. □

3.2. **Existence, Uniqueness and Convergence.** We start this section with a proof of the well-posedness of the SDE (1.2) for every polynomial s. Afterwards, we prove Theorem 2.1, which provides conditions for a sequence of Moran models with frequency-dependent selection and large neutral offspring to converge to the solution of the SDE (2.1). The asymptotic (properly scaled) selection mechanism consequently corresponds to a SD of the drift term. Corollary 2.2 states that the converse is also true, i.e. for a SDE of the form (2.1) and a SD (β, p) of its drift, one can always construct a sequence of Moran models converging (in the sense of Theorem 2.1) to the SDE (2.1).

Lemma 3.2 (Existence and uniqueness)**.** *Let $s : \mathbb{R} \to \mathbb{R}$ be a polynomial and let Λ be a finite measure on $[0, 1]$. Let W be a standard Brownian motion and let $\tilde{N}$ be an independent compensated Poisson measure on $[0, \infty) \times (0, 1] \times [0, 1]$ with intensity $dt \times r^{-2}\Lambda(dr) \times du$. Then for any $x_0 \in [0, 1]$, there is a pathwise unique strong solution $X := (X_t;\, t \geq 0)$ to the SDE* (1.2) *such that $X_0 = x_0$ and $X_t \in [0, 1]$ for all $t \geq 0$.*

Proof. We first introduce some notation. Set $b(x) := x(1-x)s(x)$, $\sigma(x) := \sqrt{\Lambda(\{0\})x(1-x)}$ and $g_0(x, r, u) := r((1-x)\mathbf{1}_{\{u\leq x\}} - x\mathbf{1}_{\{u>x\}})$ for $x \in [0, 1]$, $(r, u) \in (0, 1] \times [0, 1]$, complemented by $b(x) := \sigma(x) := g_0(x, r, u) = 0$ whenever $x \notin [0, 1]$. Clearly b and σ are continuous and g_0 is measurable. Moreover, for $x \in [0, 1]$ and $(r, u) \in (0, 1] \times [0, 1]$, we have $-x \leq g_0(x, r, u) \leq 1 - x$. It follows from [28, Prop. 2.1] that any solution of (1.2) starting at a point $x_0 \in [0, 1]$ remains in $[0, 1]$. In order to prove existence and pathwise uniqueness of strong solutions of the SDE (1.2), we use [42, Thm. 5.1]. We need to verify conditions (3a), (3b) and (5a) of that paper. Note first that b is Lipschitz (it is a polynomial in $[0, 1]$), and hence Condition (3a) is satisfied. Condition (3b) concerns only σ and g_0. In [32, Lem. 3.6] it was proved that there a constant $c > 0$ such that $|\sigma(x) - \sigma(y)|^2 \leq c|x - y|$. In addition, a straightforward calculation shows that

$$\int_{(0,1]\times[0,1]} |g_0(x, r, u) - g_0(y, r, u)|^2 \,\frac{\Lambda(dr)}{r^2}\, du \leq \Lambda((0, 1])|x - y|.$$

Hence, condition (3b) is satisfied. It remains to verify condition (5a). Since b and σ are bounded, this amounts to prove that $x \mapsto \int_{(0,1]\times[0,1]} g_0(x, r, x)^2 r^{-2}\Lambda(dr)du$ is bounded. This directly follows from the fact that $g_0(x, r, u)^2 \leq 2r^2$. □

Lemma 3.3 (Operator core)**.** *The solution of the SDE* (1.2) *is a Feller process with generator A satisfying*

$$Af(x) = A_s f(x) + A_{\mathrm{wf}} f(x) + A_\Lambda f(x), \quad \text{for any} \quad f \in \mathcal{C}^2([0, 1]), x \in [0, 1], \tag{3.3}$$

where

$$A_s f(x) = x(1-x)s(x)f'(x), \qquad A_{\mathrm{wf}} f(x) = \frac{\Lambda(\{0\})}{2}\, x(1-x)\, f''(x),$$
$$A_\Lambda f(x) = \int_{(0,1]} x\big[f(x + r(1-x)) - f(x)\big] + (1-x)\big[f(x - rx) - f(x)\big]\, \frac{\Lambda(dr)}{r^2}.$$

Moreover, $\mathcal{C}^\infty([0, 1])$ is a core for A.

Proof. Let X be the unique strong solution of the SDE (1.2). It follows from standard theory of SDEs that X is a strong Markov process with generator A satisfying (3.3). Since pathwise uniqueness implies weak uniqueness (see [7, Thm. 1]), we deduce from [37, Thm. 2.16] that the martingale problem associated to A in $\mathcal{C}^\infty([0, 1])$ is well-posed. Using [54, Prop. 2.2], we infer that X is Feller. The fact that $\mathcal{C}^\infty([0, 1])$ is a core follows then from [54, Thm. 2.5]. □

We now prove the convergence of the Moran-type models to the solution of the SDE (2.1).

Proof of Theorem 2.1. Let X be the solution of (2.1) with $X_0 = x_0 \in [0,1]$. Denote by A and A^N the generator of X and $\tilde{X}^{(N)} := (X^{(N)}_{Nt};\, t \geq 0)$, respectively. The main ingredient of the proof is to show that for every $f \in \mathcal{C}^\infty([0,1])$

$$\sup_{x \in \square_N} |A^N f|_{\square_N}(x) - Af(x)| \xrightarrow[N\to\infty]{} 0, \tag{3.4}$$

where $\square_N := \{k/N : k \in [N]_0\}$ and $f|_{\square_N}$ denotes the restriction of f to $\square_N$. Let us first assume that this is true. Since $\mathcal{C}^\infty([0,1])$ is a core for A (see Lemma 3.3), we can apply [23, Thm. 1.6.1 and Thm. 4.2.11] to deduce the convergence in distribution of $\tilde{X}^{(N)}$ to X. It remains to prove (3.4).
Let $K^n_{m,k} \sim \mathrm{Hyp}(n,m,k)$ for $n \geq m \vee k$, $i \leq k \wedge m$. Define the discrete differential operators

$$D_h f(x) := \frac{f(x+h) - f(x)}{h} \quad \text{and} \quad D^{(2)}_h f(x) := \frac{f(x+h) + f(x-h) - 2f(x)}{h^2}.$$

Then, the generator A^N takes the form

$$A^N f(x) = \sum_{\ell=2}^m A^N_{\mathrm{s},\ell} f(x) + A^N_{\mu_N} f(x) + A^N_0(x), \qquad f : \square_N \to \mathbb{R},\, x \in \square_N,$$

where

$$\begin{aligned}
A^N_0 f(x) &= \frac{\mu_N(\{0\})}{2}\, x(1-x)\, D^{(2)}_{\frac{1}{N}} f(x),\\
A^N_{\mathrm{s},\ell} f(x) &= N\beta^N_\ell \left\{(1-x)\,\mathbb{E}\left[p_{K^{N-1}_{Nx,\ell-1},\ell}\right] D_{\frac{1}{N}} f(x) - x\,\mathbb{E}\left[1 - p_{K^{N-1}_{Nx-1,\ell-1}+1,\ell}\right] D_{-\frac{1}{N}} f(x)\right\},\\
A^N_{\mu_N} f(x) &= N^2 \sum_{r=1}^N \mu_N(r) \left\{x\mathbb{E}\left[f\left(x + \frac{r - K^N_{Nx,r}}{N}\right) - f(x)\right] + (1-x)\mathbb{E}\left[f\left(x - \frac{K^N_{Nx,r}}{N}\right) - f(x)\right]\right\}.
\end{aligned}$$

Similarly, consider the generator A from (3.3) with

$$s(x) := \frac{1}{x(1-x)} \sum_{\ell=2}^m \beta_\ell \sum_{i=0}^\ell b_{i,\ell}(x)\Big(p_{i,\ell} - \frac{i}{\ell}\Big).$$

In particular, the selective term admits a decomposition as $A_{\mathrm{s}} = \sum_{\ell=2}^m A_{\mathrm{s},\ell}$, where

$$A_{\mathrm{s},\ell} f(x) := \beta_\ell \left(\mathbb{E}\left[p_{B_{\ell,x},\ell}\right] - x\right) f'(x) \quad \text{and} \quad B_{\ell,x} \sim \mathrm{Bin}(\ell, x).$$

Appropriate Taylor expansions and the triangular inequality yield for $f \in \mathcal{C}^\infty([0,1])$

$$\sup_{x \in \square_N} |A^N_0 f|_{\square_N}(x) - A_{\mathrm{wf}} f(x)| \leq \mu_N(\{0\}) \frac{\|f'''\|_\infty}{6N} + |\mu_N(\{0\}) - \Lambda(\{0\})| \frac{\|f''\|_\infty}{2} \xrightarrow[N\to\infty]{} 0. \tag{3.5}$$

For the term associated to $\ell-$replacements, first note that

$$(1-x)\mathbb{E}\left[p_{K^{N-1}_{Nx,\ell-1},\ell}\right] + x\mathbb{E}\left[p_{K^{N-1}_{Nx-1,\ell-1}+1,\ell}\right] = \mathbb{E}\left[p_{K^N_{Nx,\ell},\ell}\right].$$

As a consequence, appropriate Taylor expansions and the triangular inequality lead to

$$\sup_{x \in \square_N} |A^N_{\mathrm{s},\ell} f|_{\square_N}(x) - A_{\mathrm{s},\ell} f(x)| \leq |N\beta^N_\ell - \beta_\ell| \|f'\|_\infty + \beta_\ell \left(\frac{\|f''\|_\infty}{2N} + \|f'\|_\infty R^\ell_N\right),$$

where

$$R^\ell_N := \sup_{x \in \square_N} \left|\mathbb{E}\left[p_{K^N_{Nx,\ell},\ell}\right] - \mathbb{E}\left[p_{B_{\ell,x},\ell}\right]\right| \leq \binom{\ell}{\lfloor \frac{\ell}{2} \rfloor} \frac{2\ell^3}{N - \ell}.$$

The previous inequality follows from Lemma 3.4. We conclude that

$$\lim_{N\to\infty} \sup_{x \in \square_N} |A^N_{\mathrm{s},\ell} f|_{\square_N}(x) - A_{\mathrm{s},\ell} f(x)| = 0. \tag{3.6}$$

For the last term, note that

$$|A^N_{\mu_N} f|_{\square_N}(x) - A_\Lambda f(x)| \leq \varepsilon_{N,1}(x) + \varepsilon_{N,2}(x) + \varepsilon_{N,3}(x),$$

where

$$\varepsilon_{N,1}(x) := \left| N^2 \sum_{r=1}^{N} \mu_N(r) x \mathbb{E}\left[f\left(x + \frac{r - K^N_{Nx,r}}{N}\right) - f\left(x + \frac{r(1-x)}{N}\right)\right]\right|,$$
$$\varepsilon_{N,2}(x) := \left| N^2 \sum_{r=1}^{N} \mu_N(r)(1-x)\mathbb{E}\left[f\left(x - \frac{K^N_{Nx,r}}{N}\right) - f\left(x - \frac{rx}{N}\right)\right]\right|,$$
$$\varepsilon_{N,3}(x) := \left| N^2 \sum_{r=1}^{N} \mu_N(r)\left\{ x f\left(x + \frac{r(1-x)}{N}\right) + (1-x) f\left(x - \frac{rx}{N}\right) - f(x)\right\} - A_\Lambda f(x)\right|.$$

Second order Taylor expansions with integral remainder around the points $x + r(1-x)/N$ and $x - rx/N$ and standard properties of the hypergeometric distribution lead to

$$\sup_{x\in\square_N} \big(\varepsilon_{N,1}(x) + \varepsilon_{N,2}(x)\big) \leq \frac{\|f''\|_\infty}{2}\sum_{r=1}^{N}\mu_N(r) r.$$

Moreover, for any $\gamma \in (0,1)$, we have

$$\sum_{r=1}^{N}\mu_N(r)r = \sum_{1\leq r\leq \gamma N}\mu_N(r) r + \sum_{N\gamma < r \leq N}\mu_N(r) r \leq M_{\mu_N}\left(T^N\mu_N([0,\gamma]) - \frac{\mu_N(\{0\})}{M_{\mu_N}} + \frac{1}{\gamma N} T^N \mu_N([\gamma,1])\right).$$

Hence, the portmanteau theorem yields

$$\limsup_{N\to\infty}\sum_{r=1}^{N}\mu_N(r) r \leq \Lambda((0,\gamma]).$$

Since this holds for any $\gamma\in(0,1)$, we conclude that the previous lim sup is 0. Hence,

$$\lim_{N\to\infty}\sup_{x\in\square_N}\big(\varepsilon_{N,1}(x) + \varepsilon_{N,2}(x)\big) = 0. \tag{3.7}$$

Now, we set $f_x(r) := r^{-2}(x[f(x+r(1-x)) - f(x)] + (1-x)[f(x-rx) - f(x)])$. Note that $|f_x(r)| \leq \|f''\|_\infty/2$. Using this and the definition of $T^N\mu_N$, we get

$$\varepsilon_{N,3}(x) \leq |M_{\mu_N} - \Lambda([0,1])|\frac{\|f''\|_\infty}{2} + \Lambda([0,1])\left|\int_{(0,1]} f_x(r) T^N\mu_N(dr) - \int_{(0,1]} f_x(r)\frac{\Lambda(dr)}{\Lambda([0,1])}\right|. \tag{3.8}$$

For any $\gamma\in(0,1)$

$$\left|\int_{(0,\gamma]} f_x(r) T^N\mu_N(dr) - \int_{(0,\gamma]} f_x(r)\frac{\Lambda(dr)}{\Lambda([0,1])}\right| \leq \frac{\|f''\|_\infty}{2}\left(T^N\mu_N((0,\gamma]) + \frac{\Lambda((0,\gamma])}{\Lambda([0,1])}\right). \tag{3.9}$$

For the remaining term, we have

$$\begin{aligned}\left|\int_{[\gamma,1]} f_x(r) T^N\mu_N(dr) - \int_{[\gamma,1]} f_x(r)\frac{\Lambda(dr)}{\Lambda([0,1])}\right| &\leq \left|\int_{[0,1]} \tilde f_x(r) T^N\mu_N(dr) - \int_{[0,1]} \tilde f_x(r)\frac{\Lambda(dr)}{\Lambda([0,1])}\right| \\ &\quad + \frac{\|f''\|_\infty}{2}\left|T^N\mu_N([0,\gamma]) - \frac{\Lambda([0,\gamma])}{\Lambda([0,1])}\right|, \end{aligned} \tag{3.10}$$

where $\tilde f_x$ is the continuous function coinciding with f_x on $[\gamma,1]$ and being constant in $[0,\gamma]$. Note that $\tilde f_x$ is Lipschitz and bounded. Moreover, $\|\tilde f_x\|_{\text{Lip}} \leq 4\|f\|_\infty/\gamma^3 + 2\|f'\|_\infty/\gamma^2 := C_\gamma(f)$, where $\|f\|_{\text{Lip}} := \|f\|_\infty \vee \sup_{x\neq y}|f(x) - f(y)|/|x-y|$ denotes the bounded Lipschitz norm of f. Thus,

$$\left|\int_{[0,1]} \tilde f_x(r) T^N\mu_N(dr) - \int_{[0,1]} \tilde f_x(r)\frac{\Lambda(dr)}{\Lambda([0,1])}\right| \leq C_\gamma(f) d_{\text{Lip}}\left(T^N\mu_N, \frac{\Lambda}{\Lambda([0,1])}\right), \tag{3.11}$$

where $d_{\text{Lip}}(\nu_1,\nu_2) := \sup\{|\int f d\nu_1 - \int f d\nu_2| : \|f\|_{\text{Lip}} \leq 1\}$ denotes the bounded Lipschitz metric in the space of probability measures on $[0,1]$. Assume that γ is a continuity point of Λ. Combining (3.8), (3.9), (3.10), (3.11), and letting $N\to\infty$, we deduce that

$$\limsup_{N\to\infty}\sup_{x\in\square_N}\varepsilon_{N,3}(x) \leq \|f''\|_\infty\frac{\Lambda((0,\gamma])}{\Lambda([0,1])}. \tag{3.12}$$

This holds for any continuity point $\gamma \in (0,1)$. Hence, the previous limit exists and equals 0. Together with (3.7), this implies that

$$\lim_{N\to\infty} \sup_{x\in\square_N} |A^N_{\mu_N} f|_{\square_N}(x) - A_\Lambda f(x)| = 0. \tag{3.13}$$

This ends the proof. $\square$

At last, we prove Corollary 2.2, which provides a particular Moran model that converges to the solution of (2.1) for a given (β, p, Λ).

Proof of Corollary 2.2. It is enough to show that conditions (1) and (2) of Theorem 2.1 are satisfied. Since $N\beta^N = \beta$ and $\mu_N(\{0\}) = \Lambda(\{0\})$, conditions (1) and the first part of condition (2) are satisfied. In order to prove the other two parts of condition (2), it suffices to show that for every $f \in \mathcal{C}([0,1])$

$$\sum_{k=1}^{N-1} f\left(\frac{k}{N}\right)\frac{k^2}{N^2}\lambda^0_{N,k+1}\binom{N}{k+1} \xrightarrow[N\to\infty]{} \int_{(0,1]} f(r)\Lambda(dr).$$

By the definition of $\lambda^0_{N,k+1}$ and a straightforward calculation,

$$\sum_{k=1}^{N-1} f\left(\frac{k}{N}\right)\frac{k^2}{N^2}\lambda^0_{N,k+1}\binom{N}{k+1} = \frac{N-1}{N}\int_{(0,1]} \mathbb{E}\left[f\left(\frac{B_{N-2,r}+1}{N}\right)\frac{B_{N-2,r}+1}{B_{N-2,r}+2}\right]\Lambda(dr),$$

where $B_{N-2,r} \sim \mathrm{Bin}(N-2, r)$. The result follows then as an application of the law of large numbers and the dominated convergence theorem. $\square$

We end this section with an auxiliary result, which is used in the proof of Theorem 2.1. For $N \in \mathbb{N}$, $m_N \in [N]_0$ and $\ell \in [N]$, consider $K^N_{m_N,\ell} \sim \mathrm{Hyp}(N, m_N, \ell)$. It is well-known that

$$\lim_{N\to\infty}\frac{m_N}{N} = p \in [0,1] \quad\Rightarrow\quad K^N_{m_N,\ell} \xrightarrow[N\to\infty]{(d)} B_{\ell,p}.$$

The next lemma provides a uniform version of this result.

Lemma 3.4. *For $N \in \mathbb{N}$, $\ell \in [N]$ and $x \in [0,1]$, we have*

$$\sup_{x\in\square_N, i\in[\ell]_0} \left|\mathbb{P}(K^N_{Nx,\ell} = i) - \mathbb{P}(B_{\ell,x} = i)\right| \le \binom{\ell}{\lfloor\frac{\ell}{2}\rfloor}\frac{2\ell^2}{N-\ell+1}.$$

Proof. Note first that

$$\left|\mathbb{P}(K^N_{Nx,\ell} = i) - \mathbb{P}(B_{\ell,x} = i)\right| = \binom{\ell}{i}\left|\prod_{k=0}^{i-1}\frac{Nx-k}{N-k}\prod_{k=0}^{\ell-i-1}\frac{N(1-x)-k}{N-i-k} - x^i(1-x)^{\ell-i}\right|.$$

Hence, the result is a direct consequence of the following fact: if $(a_n)_{n=0}^m$ and $(b_n)_{n=0}^m$ are two sequences of numbers in $[0,1]$, then

$$\left|\prod_{k=0}^m a_k - \prod_{k=0}^m b_k\right| \le \sum_{k=0}^m |a_k - b_k|.$$

$\square$

4. Ancestral structures and Bernstein duality

4.1. **Ancestral selection graph.** This section provides a more detailed description of the ancestral selection graph. The branching-coalescing system that underlies the ASG arises from the following (marked) particle system. Let $\beta \in \mathbb{R}_+^{m-1}$ and $\Lambda \in \mathcal{M}_f$. Each particle is equipped with a label in $[0,1]$. Start at time $t = 0$ with n particles. The i-th particle carries label U_i, where $(U_i)_{i=0}^n$ are independent random variables that are uniformly distributed in $[0,1]$. If there are n particles present at time t, then

(1) at rate $\beta_\ell n$ for every $\ell \in]m]$, mark one of the existing particles chosen uniformly at random and generate $\ell - 1$ new particles, where each new particle carries a new label that is independent from the other labels and uniformly distributed in $[0,1]$,
(2) at rate $\binom{n}{k}\lambda_{n,k}$ for every $k \in]n]$, eliminate k particles and give birth to a new particle with a new label that is uniformly distributed in $[0,1]$ and independent of the previous labels.

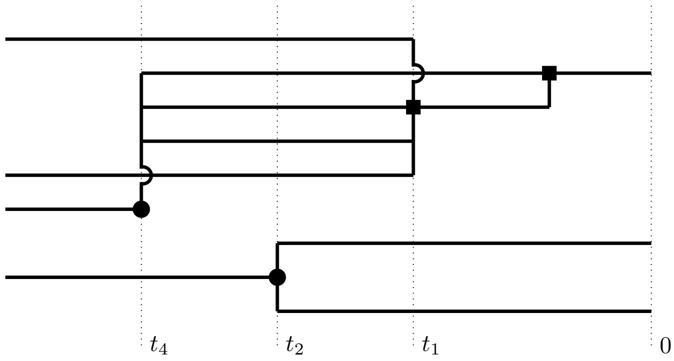

FIGURE 5. A realisation of the ASG. Each particle is represented by a labelled line. The vertical coordinate corresponds to the label (on $[0, 1]$). The right and left extremity of a line corresponds to the particle's birth and death time, respectively. Black elements indicate the marked particle at a replacement (square) or a newly born particle at a reproduction event (circle).

Note that by construction a particle with label u has an associated birth and death time (b_u, d_u). See Fig. 5. Let $\mathcal{U}$ be the set of labels assigned from $t = 0$ to $t = \infty$. The branching-coalescing system with parameters (β, Λ), which is denoted by $\mathcal{G}$, is the following (infinite and uncountable) directed acyclic graph (DAG) with distinct horizontal and vertical edges.

- The set of vertices is given by $\{(u, s) \ : \ u \in \mathcal{U}, s \in [b_u, d_u]\}$.
- The set of *horizontal* (directed) edges is given by
$$\big\{\big((u, s), (u, t)\big) \ : \ s < t \in [b_u, d_u]\big\}$$
- Let $\{D_i\}$ (resp. $\{I_i\}$) be the set of times at which the number of particle decreases (resp. increases). The set of (directed) *vertical* edges is given by the union of the two sets
$$\bigcup_i \big\{\big((u, D_i), (v, D_i)\big) : \ u \text{ killed at time } t, \ v \text{ born at time } t\big\}$$
$$\bigcup_i \big\{\big((u, I_i), (v, I_i)\big) : \ u \text{ marked at time } t, \ v \text{ born at time } t\big\}$$
- The set of edges is the union of horizontal and vertical edges.

$\mathcal{G}$ is a DAG with distinct vertical and horizontal edges. In the following, we denote this type of graph by vhDAG.

Definition 4.1 (ASG)**.** Let $(\beta, p) \in \mathcal{E}_d$, $\Lambda \in \mathcal{M}_f([0, 1])$ and $n \in \mathbb{N}$. The (β, p, Λ)-ASG is the pair $(\mathcal{G}, p)$, where $\mathcal{G}$ is the branching-coalescing system with parameters (β, Λ), as defined above, starting with n particles. Finally, the graph $\mathcal{G}_t$ denotes the induced graph on the set of vertices with time coordinates less than t. In particular, the process $(L_t; t \geq 0)$, where L_t counts the number of leaves in $\mathcal{G}_t$, evolves according to the rates in Definition 2.6

4.2. **Ancestral selection polynomial and the Bernstein coefficient process.** Let us start formalising the idea of propagation of types (colouring) in a vhDAG G. We denote by $R(G)$ and $L(G)$ the set of roots (sources) and leaves (sinks) of G, respectively.

Definition 4.2 (Colouring of a directed acyclic graph)**.** Let G be a vhDAG having a maximal vertical outdegree smaller than or equal to m and finitely many vertical edges. We call a vector $t := (t_\ell)_{\ell \in L(G)} \in \{a, A\}^{L(G)}$ a leaves colouring. For two vertices x, y of G, we say that y transmits the colour to $x+$, if there is at least one directed path from x to y and there is no directed path from x to y that contains a marked point, different from x and y, with vertical outdegree larger than 0. We say that $x+$ has colour a (resp. A) if there is some vertex y with colour a (resp. A) that transmits the colour to $x+$. Given a leaves colouring t and a colouring rule $p \in \mathcal{P}_m$, we randomly colour the vertices of G according to the following rules:

- A vertex x that is not marked gets the same colour as vertex y, if y transmits the colour to $x+$.

- For a marked vertex x with vertical outdegree $k-1>0$, consider the colours of the $k-1$ vertical neighbours together with the colour of $x+$. If i of them are colour a and $k-i$ are colour A, then colour x with type a (resp. A) with probability $p_{i,k}$ (resp. $1-p_{i,k}$).

Now let U_k, $k\in[L(G)]$, be a uniform random variable on the leaves colourings of G that have k leaves of colour a and $L(G)-k$ leaves of colour A. Define $\mathcal{R}(G,p,k)$ to be the probability that all the roots of G get colour a if G is coloured according to U_k and p.

The previous notion of colouring of a vhDAG leads to a more precise definition of the ancestral selection polynomial.

Definition 4.3 (Ancestral selection polynomial)**.** Let $(\beta,p)\in\mathcal{E}_d$, $\Lambda\in\mathcal{M}_f([0,1])$, and let $(\mathcal{G}_t,L_t)$ be as in Definition 4.1. For every time $t>0$, define $V_t\in\mathbb{R}^{L_t+1}$ via

$$V_t(i):=\mathbb{E}\big[\mathcal{R}(\mathcal{G}_t,p,i)\mid(L_s;s\in[0,t])\big],\quad i\in[L_t]. \tag{4.1}$$

The *ancestral selection polynomial (ASP)* associated to the triplet (β,p,Λ) is defined as

$$P_t(x):=\sum_{i=0}^{L_t}V_t(i)b_{i,L_t}(x),\quad x\in[0,1]. \tag{4.2}$$

Remark 4.1. We abuse the notation by using the same symbol V_t for (1) the coefficients of the ASP in the Bernstein basis of degree L_t, and (2) the Bernstein coefficient process defined in Theorem 2.11. This abuse of notation is motivated by the fact that the two objects are identical in distribution as announced in Proposition 2.10.

In words, conditional on the leaf process $(L_s;s\in[0,t])$ up to time t, $P_t(x)$ is the probability that every root of the graph $\mathcal{G}_t$ is of type a if we first assign type a (resp. A) with probability x (resp. $1-x$) to each leaf of the graph $\mathcal{G}_t$, independently for every leaf, and then propagate types from the leaves to the roots as in Definition 4.2. We stress that P_t is a probability conditional on the leaf process up to time t, and thus a random object arising by the removal of all details of the graph $\mathcal{G}_t$ except for the leaf process.

Remark 4.2. For fixed t, the graph $\{(x,P_t(x))\,;x\in[0,1]\}$ is the the Bézier curve associated to the set of points $\{(i/n,V_t(i))\}_{i=0}^{L_t}$ (see, e.g. [30]).

Remark 4.3. It is plain from the definition that $L_t\geq\deg(P_t)$. A natural conjecture is that $L_t=\deg(P_t)$. However, it is easily seen to be false. For instance, consider the case $\deg(d)=3$ with $\beta_2=0$, $\beta_3=1$ and $p_3=(0,1,0,1)$, which corresponds to the selection term

$$d(x)=x(1-x)(1-2x).$$

This colouring rule corresponds to a *minority rule*, which means that the minority type in the set of potential parents provides the true parent. In this particular case, one can check that

$$C^{3,2}\circ S^{3,3}(Id)=Id$$

so that one can envision a scenario (branching followed by coalescence of a pair) where $L_t=1+2-1=2$ and $\deg(P_t)=1$.

Remark 4.4 (Topology on $\mathbb{R}^\infty$)*.* Let us briefly comment on the topology considered on the state space $\mathbb{R}^\infty:=\cup_{n\in\mathbb{N}}\mathbb{R}^n$ of the Bernstein coefficient process. One is inclined to embed $\mathbb{R}^\infty$ in the set of infinite sequences (adding infinite zeros at the end of every vector) equipped with the supremum norm. The main reason not to do so is that in order to define the Bernstein coefficient process, one needs to know the dimension of the current state of the process, and the latter does not necessarily coincides with the last non-zero coordinate of the process. Instead, the following metric is more appropriate. For $u\in\mathbb{R}^n$ and $v\in\mathbb{R}^m$, with $n\leq m$, define

$$d(u,v):=d(v,u):=\max_{i\in[n]_0}|u_i-v_i|+\max_{i\in[m]_0\setminus[n]_0}|v_i|+m-n.$$

Note that the restriction to $\mathbb{R}^n$ of this metric coincides with the metric induced by then supremum norm, and that the distance between two vectors with different dimensions is at least 1. Hence, a function

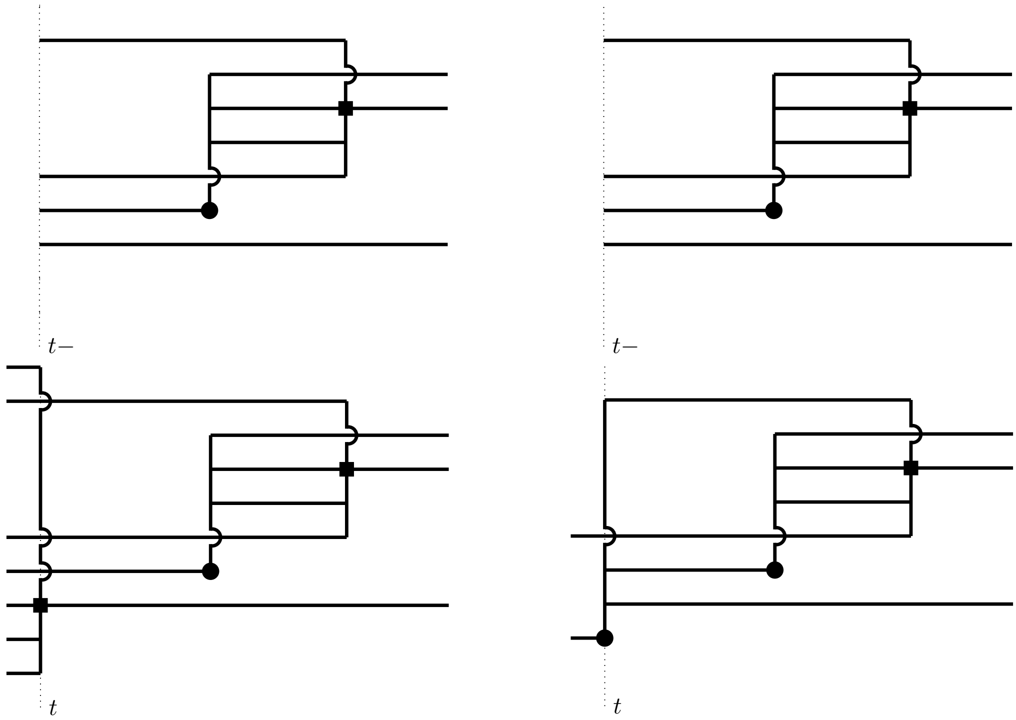

FIGURE 6. Left: At a replacement event, a branching point is grafted at a leaf chosen uniformly at random at time $t-$. In this example, it is a 4-replacement, $L_{t+} = 7$, and $L_{t-} = 4$. Right: In a coalescence event, a subset of leaves at time t_- collapse into a single line. Here, 3 leaves merge, $L_{t+} = 2$, and $L_{t-} = 4$.

$f : \mathbb{R}^\infty \to \mathbb{R}$ is continuous if and only, for any $n \in \mathbb{N}$, its restriction to $\mathbb{R}^n$ is continuous. In particular, the duality function in Theorem 2.11 is continuous under this topology. Moreover, one can easily prove that $(\mathbb{R}^\infty, d)$ is a Polish space, which is a suitable setting for stochastic processes.

We now prove that the Bernstein coefficient process and the Bernstein coefficient vector of the ancestral selection polynomial are identical in distribution.

Proof of Proposition 2.10. Let $(V_t^\star, t \geq 0)$ be the Bernstein coefficient process with initial condition $V_0^\star = e_{n+1}$ so that $\langle B_{L_0^\star}(x), V_0^\star \rangle = x^n$, where $L_t^\star := \dim(V_t^\star) - 1$. Let $(V_t;\, t \geq 0)$ be the Bernstein coefficients of the ASP associated with the ASG started with n particles (see Eq. (4.1)). Our aim is to prove that the two objects are identical in law.
First note that L and $L^\star := (L_t^\star;\, t \geq 0)$ are by definition identical in law. Thus, in order to prove that if $V_0 = V_0^\star = e_{n+1}$, then $(V_t;\, t \geq 0) \overset{(d)}{=} (V_t^\star;\, t \geq 0)$, it is enough to show that

(i) conditional on a positive jump of size $\ell - 1$ of the leaf process (corresponding to a ℓ-branching event) at time t, we have $V_{t+} = S^{L_{t-},\ell} V_{t-}$ (where $S^{L,\ell}$ is the selection operator of Definition 2.8).
(ii) conditional on a negative jump of size k (corresponding to a coalescence event of k leaves) at time t, we have $V_{t+} = C^{L_{t-},k} V_{t-}$ (where $C^{L,k}$ is the coagulation operator of Definition 2.8).

(i) *Selection event.* Let $A(t, \ell)$ be the event there is an ℓ-branching event at time t. In this case, $L_{t+} = L_{t-} + \ell - 1$. For each $i \in [L_{t+}]_0$, we need to determine $\mathbb{E}[\mathcal{R}(\mathcal{G}_{t+}, p, i) \mid \mathcal{G}_{t-},\, A(\ell, t)]$. First note that, conditional on $(\mathcal{G}_{t-}, A(\ell, t))$, the ℓ-branching point, call it u, is chosen uniformly at random among the L_{t-} leaves at time $t-$. In other words, the ℓ-branching point is grafted uniformly at random on a leaf available at time $t-$. In order to determine $\mathcal{R}(\mathcal{G}_{t+}, p, i)$, colour i leaves at time $t+$ with type a (uniformly at random on the set of L_{t+} leaves). Starting from time $t+$, there is a subset of leaves Γ consisting of exactly ℓ leaves that collapses into a single leaf u at time $t-$ (see Fig. 6). Let K_i be the number of type a in the subset of leaves Γ, so that $K_i \sim \mathrm{Hyp}(L_{t+}, i, \ell)$. According to the colouring rule p, conditional on K_i, the leaf u is of type a (resp. type A) with probability $p_{K_i,\ell}$ (resp. $1 - p_{K_i,\ell}$). Furthermore, at time $t-$ there are $i - K_i$ leaves different from u that carry type a. They are distributed uniformly at random among the remaining leaves. Since u is chosen uniformly at random at time $t-$ (as argued above),

$$\mathbb{E}\big[\mathcal{R}(\mathcal{G}_{t+},p,i)\mid \mathcal{G}_{t-},A(\ell,t)\big] \;=\; \mathbb{E}\big[\,p_{K_i,\ell}\,\mathcal{R}(\mathcal{G}_{t-},i+1-K_i)+(1-p_{K_i,\ell})\,\mathcal{R}(\mathcal{G}_{t-},i-K_i)\mid \mathcal{G}_{t-}\big],$$

and, since $\sigma((L_s;s\in[0,t]))\subset\sigma(\mathcal{G}_{t-},A(\ell,t))$,

$$\begin{aligned}\mathbb{E}\big[\mathcal{R}(\mathcal{G}_{t+},p,i)\mid (L_s;s\in[0,t-]),A(\ell,t)\big] &= \mathbb{E}\big[\,p_{K_i,\ell}\,V_{t-}\,(i+1-K_i)+(1-p_{K_i,\ell})\;\;V_{t-}\,(i-K_i)\,\big]\\ &= (S^{L_{t-},\ell}V_{t-})(i),\end{aligned}$$

which is the desired result.

(ii) *Coalescence event.* Consider a k-coalescence event with $k\geq 2$ and fix $i\leq L_{t+}=L_{t-}-k+1$. Colour i leaves at time $t+$ uniformly at random. Starting from time $t+$, one leaf splits into k leaves at time $t-$. Seen from $t-$, this corresponds to k leaves merging into one leaf at time $t+$ (see Fig. 6). If this leaf at time $t+$ is of type a (which occurs with probability $i/L_{t+}=i/(L_{t-}-k+1)$), then there are $i+k-1$ leaves of type a at time $t-$. Otherwise, there are i leaves of type a at time $t-$. This translates into

$$V_{t+}(i)=\frac{i}{L_{t-}-k+1}V_{t-}(i+k-1)\;+\;\left(1-\frac{i}{L_{t-}-k+1}\right)V_{t-}(i)=(C^{L_{t-},k}V_{t-})(i).$$

The combination of (i) and (ii) yields that V is a Markov process with the desired transition rates.

□

4.3. **Bernstein duality (proofs).** In this section, we prove Theorem 2.11, i.e. the Bernstein duality between the type-a frequency process and the Bernstein coefficient process, both associated to the same triplet (β,p,Λ). We also prove Corollary 2.12, which exhibits the connection between Bernstein and moment dualities.

Let us start with the following result about the selection and coagulation matrices of Definition 2.8.

Lemma 4.4. *Fix $n\in\mathbb{N}_0$ and $v\in\mathbb{R}^{n+1}$. Then*

$$(S^{n,\ell}v)_0\;=\;v_0\;=\;(C^{n,k}v)_0,\quad\textit{and}\quad(S^{n,\ell}v)_{n+\ell-1}\;=\;v_\ell\;=\;(C^{n,k}v)_{n-k+1}.$$

Furthermore, $\|S^{n,\ell}v\|_\infty\leq\|v\|_\infty$ and $\|C^{n,\ell}v\|_\infty\leq\|v\|_\infty$.

Proof. This is plain from the definition of the fragmentation and coagulation operators. □

Proof of Theorem 2.11. We want to prove that X and V are dual with respect to the duality function

$$H:[0,1]\times\mathbb{R}^\infty\to\mathbb{R}_+,\qquad \big(x,v\big)\mapsto\big\langle B_{dim(v)-1}(x),v\big\rangle\,.$$

First note that the generator B of the process V can be expressed as

$$Bf(v):=B_sf(v)+B_{\mathrm{wf}}f(v)+B_\Lambda f(v),\quad v\in\mathbb{R}^\infty,f\in\mathcal{C}(\mathbb{R}^\infty),$$

where, for v with $\dim(v)=n+1$,

$$B_sf(v):=\sum_{\ell=2}^m n\beta_\ell\left(f(S^{n,\ell}v)-f(v)\right),\quad B_{\mathrm{wf}}f(v):=\frac{\Lambda(\{0\})}{2}n(n-1)\left(f(C^{n,2}v)-f(v)\right),$$

$$B_\Lambda f(v):=\sum_{k=2}^n\binom{n}{k}\lambda^0_{n,k}\left(f(C^{n,k}v)-f(v)\right).$$

In addition, for any $v\in\mathbb{R}^\infty$, the function $x\mapsto H(x,v)$ is $\mathcal{C}^\infty([0,1])$, and therefore, belongs to the domain of the generator A of X given in (3.3). Similarly, for any $x\in[0,1]$, $v\mapsto H(x,v)$ is continuous (with respect to the metric d defined in Remark 4.4), and hence, belongs to the domain of B. Now, we proceed to show that for $v\in\mathbb{R}^\infty$ and $x\in[0,1]$ the following holds

$$AH(\cdot,v)(x)=BH(x,\cdot)(v).\tag{4.3}$$

For this we prove the intermediate identities $A_\kappa H(\cdot,v)(x)=B_\kappa H(x,\cdot)(v)$, for $\kappa\in\{s,\mathrm{wf},\Lambda\}$. Let $(Y_\ell^x)_{\ell\geq0}$, $(W_\ell^x)_{\ell\geq0}$, $(K_{\ell,i}^n)_{0\leq i\leq\ell<n}$ be sequences of independent random variables, with $Y_\ell^x,W_\ell^x\sim\mathrm{Bin}(\ell,x)$ and $K_{\ell,i}\sim\mathrm{Hyp}(n-1,\ell,i)$. For any $v=(v_i)_{i=0}^n\in\mathbb{R}^{n+1}$,

$$\frac{\partial}{\partial x}\left\langle B_n(\cdot),v\right\rangle(x)=n\mathbb{E}[v_{Y^x_{n-1}+1}-v_{Y^x_{n-1}}]\quad\text{and}\quad\frac{\partial^2}{\partial x^2}\left\langle B_n(\cdot),v\right\rangle(x)=n(n-1)\mathbb{E}[v_{Y^x_{n-2}+2}-2v_{Y^x_{n-2}+1}+v_{Y^x_{n-2}}].$$

Then a straightforward calculation yields

$$\begin{aligned}
A_sH(\cdot,v)(x) &= \sum_{\ell=2}^{m} n\beta_\ell(\mathbb{E}\big[\,p_{Y_\ell^x,\ell}\big]-x)\mathbb{E}\big[v_{W_{n-1}^x+1}-v_{W_n^x}\big] \\
&= \sum_{\ell=2}^{m} n\beta_\ell\Big(\mathbb{E}\big[\,p_{Y_\ell^x,\ell}\,v_{W_{n-1}^x+1}+(1-p_{Y_\ell^x,\ell})\,v_{W_{n-1}^x}\big]-H(x,v)\Big) \\
&= \sum_{\ell=2}^{m} n\beta_\ell\Big(\sum_{i=0}^{n+\ell-1}\mathbb{E}\big[\,p_{K_{\ell,i}^n,\ell}\,v_{i-K_{\ell,i}^n+1}+(1-p_{K_{\ell,i}^n,\ell})\,v_{i-K_{\ell,i}^n}\big]b_{i,n+\ell-1}(x)-H(x,v)\Big) \\
&= B_sH(x,\cdot)(v).
\end{aligned}$$

This proves identity for the part that corresponds to selection. For the Wright–Fisher part, we have

$$\begin{aligned}
&A_{\mathrm{wf}}H(\cdot,v)(x) \\
&= \frac{\Lambda(\{0\})}{2}n\sum_{i=1}^{n}v_ib_{i-1,n-1}(x)\big(i-1-x(n-1)\big)+n\sum_{i=0}^{n-1}v_ib_{i,n-1}(x)\big(n-i-1-(1-x)(n-1)\big) \\
&= \frac{\Lambda(\{0\})}{2}n(n-1)\left(\sum_{i=0}^{n-1}\Big(\frac{i}{n-1}v_{i+1}+\frac{n-1-i}{n-1}v_i\Big)b_{i,n-1}(x)-H(x,v)\right) \\
&= B_{\mathrm{wf}}H(x,\cdot)(v).
\end{aligned}$$

For the Λ-part, it is convenient to realise that

$$x+r(1-x)=r+x(1-r),\ 1-(x+r(1-x))=(1-x)(1-r),\ 1-x+rx=r+(1-x)(1-r).$$

Then a straightforward calculation yields

$$\begin{aligned}
&A_\Lambda H(\cdot,v)(x) \\
&= \int_{(0,1]} x\left\langle B_n\big(r+(1-r)x\big),v\right\rangle+(1-x)\left\langle B_n\big(x(1-r)\big),v\right\rangle-\langle B_n(x),v\rangle\,\frac{\Lambda(dr)}{r^2} \\
&= \sum_{k=2}^{n}\binom{n}{k}\lambda_{n,k}^0\left(\sum_{i=1}^{n-k+1}v_{i+k-1}\frac{i}{n-k+1}b_{i,n-k+1}(x)+\sum_{i=0}^{n-k}v_i\frac{n-k+1-i}{n-k+1}b_{i,n-k+1}(x)-H(x,v)\right) \\
&= B_\Lambda H(x,\cdot)(v)
\end{aligned}$$

which ends the proof of (4.3). Now, assume that the process V starts at $V_0=v$. Using the definition of H and Lemma 4.4, we obtain

$$\sup_{x\in[0,1],s\in[0,T]}|H(x,V_s)|\le\|v\|_\infty. \tag{4.4}$$

Similarly,

$$\sup_{x\in[0,1],s\in[0,T]}|B_sH(x,\cdot)(V_s)|\le 2|\beta|\|v\|_\infty\sup_{s\in[0,T]}L_s, \tag{4.5}$$

where $|\beta|:=\sum_{\ell=2}^m\beta_\ell$. Moreover, using [36, Lem. 3.3], we deduce that

$$\sup_{x\in[0,1],s\in[0,T]}|(B_\Lambda+B_{\mathrm{wf}})H(x,\cdot)(V_s)|\le 2\|v\|_\infty\Lambda([0,1])\left(\sup_{s\in[0,T]}L_s\right)^2, \tag{4.6}$$

Finally, note that one can couple V to a birth process $(\Gamma_t;t\ge0)$ such that: (a) $\Gamma_0=\dim(V_0)-1$, (b) each particle split into m particles at rate $|\beta|$, and (c) $\sup_{s\in[0,T]}L_s\le\Gamma_T$. Since, Γ_T and Γ_T^2 are integrable, the result follows by Ethier and Kurtz [23, Cor. 4.4.13]. ☐

Proof of Corollary 2.12. (1) The proof of the first statement follows directly from the Bernstein duality and the fact that $\langle B_n(x),e_{n+1}\rangle=x^n$.
(2) For the second statement, fix $m\ge2$ and $(s_i)_{i=0}^{m-2}\in\mathbb{R}_+^m$ with $s_i\le s_{i-1}$ for all $i\in\,]m]$. Set

$$\beta_m:=s_{m-2}\ge0,\quad \beta_i:=s_{i-2}-s_{i-1}\ge0,\ i\in\,]m-1],\qquad p_{i,\ell}:=\mathbf{1}_{\{i=\ell\}},\ i\in[\ell]_0,\ell\in\,]m].$$

Note that

$$-x(1-x)\sum_{i=0}^{m} s_i x^i = \sum_{\ell=2}^{m} \beta_\ell \left\langle B_\ell(x), \left(p_{i,\ell} - \frac{i}{\ell}\right)_{i=0}^{\ell} \right\rangle,$$

which means that (β, p) is a SD of the polynomial on the left hand side of the previous identity. In general, $C^{n,k} e_{n+1} = e_{n-k+2}$. A straightforward calculation yields that for our particular choice of p, we have $S^{n,\ell} e_{n+1} = e_{n+\ell}$. In particular, if $V_0 = e_{n+1}$, then $V_t = e_{L_t+1}$ for all $t \geq 0$. Hence, the Bernstein duality yields

$$\mathbb{E}_x[X_t^n] = \mathbb{E}_x[\langle B_n(X_t), e_{n+1}\rangle] = \mathbb{E}_{e_{n+1}}[\langle B_{L_t}(x), e_{L_t+1}\rangle] = \mathbb{E}_n[x^{L_t}],$$

which proves the claim. □

5. Properties of the Bernstein coefficient process and its leaf process

In this section, we study properties of the Bernstein coefficient process and its leaf process. The section begins with the proof of the condition for transience and recurrence of the leaf process. This allows us to derive and characterise the invariant measures for L and V. In particular, we obtain a recursion for the tail probabilities of the stationary measure of the leaf process. At last, we study the property of coming down from infinity for the leaf process.

5.1. **Recurrence and transience of leaf process.** The leaf process L with parameter (β, p) (see Definition 2.6) take values in $\mathbb{N}$. Its infinitesimal generator is given by

$$\mathcal{L}f(n) := \mathcal{L}_\beta f(n) + \mathcal{L}_\Lambda f(n), \tag{5.1}$$

for $f : \mathbb{N} \to \mathbb{R}$, where

$$\mathcal{L}_\beta f(n) := \sum_{\ell=2}^{m} n\beta_\ell[f(n+\ell-1) - f(n)], \qquad \mathcal{L}_\Lambda f(n) := \sum_{k=2}^{n} \binom{n}{k} \lambda_{n,k}[f(n-k+1) - f(n)].$$

We want to prove Theorem 2.14, which provides conditions for positive recurrence and transience of the leaf process. Let us stress again that the first part of Theorem 2.14 is already present in González Casanova and Spanò [32, Thm. 4.6] in the case $c(\Lambda) < \infty$. The latter also states that L is not positive recurrent for $b(\beta) > c(\Lambda)$. We use here a different approach that allows us to show the transience of L for $b(\beta) > c(\Lambda)$. We closely follow the lines of Foucart [26] and adapt his arguments to fit our setting. For completeness, we also prove the first part in such a way. Let us define $f : \mathbb{N} \to \mathbb{R}$ by

$$f(\ell) := \sum_{k=2}^{\ell} \frac{k}{\delta(k)} \log\left(\frac{k}{k-1}\right),$$

where $\delta : \mathbb{N} \to \mathbb{R}$ with

$$\delta(n) := -n \int_0^1 \log\left(1 - \frac{1}{n}(nr - 1 + (1-r)^n)\right) \frac{\Lambda(dr)}{r^2}.$$

It will be convenient to collect some known properties of these functions.

Lemma 5.1.

- *$n \mapsto \delta(n)$ is non-decreasing.*
- *$n \mapsto \delta(n)/n$ is non-decreasing.*
- *$\delta(n)/n \to c(\Lambda)$ for $n \to \infty$.*

For a proof, see Herriger and Möhle [36, Lem. 3.1, Cor. 4.2].

Lemma 5.2. *The function f is non-negative and increasing. Further, the Λ-coalescent c.d.i. if and only if $\lim_{n\to\infty} f(n) < \infty$.*

Proof. The first part of the statement follows from the fact that $\delta(n) > 0$ for $n > 1$ (this is easy to see once one observes that this is true for $n = 2$ and $\delta(n)$ is non-decreasing by Lemma 5.1). In [36], it is shown that the Λ-coalescent c.d.i. if and only if $\sum_{k\geq 2} \delta(k) < \infty$. Since $k\log(k/(k-1))/\delta(k) \sim 1/\delta(k)$ as $k \to \infty$, this completes the proof of the second part of the proposition. □

The next lemma is a generalisation of Foucart [26, Lem. 2.3], which corresponds to the case $\beta_2 > 0$, $\beta_\ell = 0$ for $\ell \neq 2$, and $p_{1,2} = 0$.

Lemma 5.3. *We have*

$$\mathcal{L}f(n) \leq -1 + \sum_{\ell=2}^{m} \beta_\ell \sum_{j=n+1}^{n+\ell-1} \frac{j}{\delta(j)}. \tag{5.2}$$

Furthermore, if $b(\beta) < c(\Lambda)$, then there exists $n_0 \in \mathbb{N}$ and $\varepsilon > 0$ such that,

$$\mathcal{L}f(n) \leq -1 + \frac{b(\beta)}{c(\Lambda)} + \varepsilon b(\beta) < 0, \qquad \forall n \geq n_0,$$

with the usual convention that $1/\infty = 0$.

Proof. Foucart [26, Proof Lem. 2.3] proves that $\mathcal{L}_\Lambda f(n) \leq -1$. Hence, for the first claim it suffices to prove that $\mathcal{L}_\beta f(n) \leq \sum_{\ell=2}^{m} \beta_\ell \sum_{j=n+1}^{n+\ell-1} j/\delta(j)$. Note

$$n[f(n+\ell-1) - f(n)] = n \sum_{j=n+1}^{n+\ell-1} \frac{j}{\delta(j)} \log\Big(1 + \frac{1}{j-1}\Big) \leq \sum_{j=n+1}^{n+\ell-1} \frac{j}{\delta(j)} \log\Big(\Big(1 + \frac{1}{n}\Big)^n\Big) \leq \sum_{j=n+1}^{n+\ell-1} \frac{j}{\delta(j)},$$

where we use that $(1 + 1/n)^n$ is monotonically increasing to Euler's number. The first claim follows in a straightforward way. For the second claim, note that Lemma 5.1 implies that $n/\delta(n)$ is non-increasing and $n/\delta(n) \to 1/c(\Lambda)$. In particular, for all $\varepsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$,

$$\frac{n}{\delta(n)} \leq \frac{1}{c(\Lambda)} + \varepsilon.$$

For $n \geq n_0$, we can now estimate the right-hand side of (5.2) by

$$-1 + \sum_{\ell=2}^{m} \beta_\ell \sum_{j=n+1}^{n+\ell-1} \frac{j}{\delta(j)} \leq -1 + \frac{b(\beta)}{c(\Lambda)} + b(\beta)\varepsilon,$$

which, for $b(\beta) < c(\Lambda)$ and $\varepsilon > 0$ small enough, is negative. □

The following lemma gives a condition for the positive recurrence of the leaf process. In the case of genic selection, this result again agrees with Foucart [26, Lem. 2.4]. Define for $n \in \mathbb{N}$,

$$T^n := \inf\{s \geq 0 : L_s < n\} \quad \text{and} \quad \mathcal{T}_n := \inf\{s \geq J_1 : L_s = n\},$$

where J_1 is the time of the first jump of L.

Lemma 5.4. *Assume $b(\beta) < c(\Lambda)$. Then there exists n_0 and a constant $\tilde{c}$ such that for all $n \geq n_0$,*

$$\mathbb{E}_n\big[T^{n_0}\big] \ < \ \tilde{c} f(n).$$

Remark 5.1. In the analogous statement to Lemma 5.4, Foucart [26] has the additional condition $\sum_{k=2}^{\infty} 1/\delta(k) = \infty$ to assure that the process is non-explosive. With our argument, we get rid of that condition. This was already noted in [3, p. 4].

Proof of Lemma 5.4. We mimic the proof of Foucart [26, Lem. 2.4]. Note that, L is dominated by a process that has only m-branchings at rate $\sum_{\ell=2}^{m} \beta_\ell$ per existing line and no coalescences. Clearly, this process is non-explosive. Hence, also L is non-explosive. Next, define for $N \in \mathbb{N}$, $f_N(n) := f(n)\mathbf{1}_{\{n \leq N+m\}}$. By Dynkin's formula,

$$\Big(f_N(L_t) - \int_0^t \mathcal{L}f_N(L_s)ds\Big)_{t\geq 0}$$

is a martingale. By the previous lemma, there is $n_0 \in \mathbb{N}$ and $\varepsilon > 0$ such that

$$\mathcal{L}f(n) \leq -1 + b(\beta)/c(\Lambda) + \varepsilon b(\beta) < 0,$$

for all $n \geq n_0$. Set $C := 1 - b(\beta)/c(\Lambda) - \varepsilon b(\beta)$ and let $n_0 \leq n \leq N$. Define $S_N := \inf\{s \geq 0 : L_s \geq N+1\}$, i.e. the first time L is above N. Applying the optional stopping theorem to $f_N(L_{T^{n_0}\wedge S_N \wedge k}) - \int_0^{T^{n_0}\wedge S_N\wedge k} \mathcal{L}f_N(L_s)ds$ leads to

$$\mathbb{E}_n\big[f_N(L_{T^{n_0}\wedge S_N\wedge k})\big] = f_N(n) + \mathbb{E}_n\left[\int_0^{T^{n_0}\wedge S_N\wedge k} \mathcal{L}f_N(L_s)ds\right].$$

Clearly, $\mathcal{L}f_N(n) = \mathcal{L}f(n)$ for $n \leq N$ and using Lemma 5.3, yields

$$\mathbb{E}_n\big[f_N(L_{T^{n_0}\wedge S_N\wedge k})\big] \leq f_N(n) - C\mathbb{E}_n\big[T^{n_0} \wedge S_N \wedge k\big].$$

Hence,

$$C\,\mathbb{E}_n\big[T^{n_0} \wedge S_N \wedge k\big] \leq f_N(n) - \mathbb{E}_n\big[f_N(L_{T^{n_0}\wedge S_N\wedge k})\big] \leq f_N(n),$$

where the last inequality is a consequence of the fact that $f_N(n) \geq 0$. Since L is non-explosive, $S_N \to \infty$ almost surely as $N \to \infty$ such that

$$C\,\mathbb{E}_n\big[\,T^{n_0} \wedge k\,\big] \leq f(n).$$

Letting $k \to \infty$ yields the result. ☐

Corollary 5.5. *Assume $b(\beta) < c(\Lambda)$.*

- *If $\Lambda \neq \delta_1$, then L is positive recurrent.*
- *If $\Lambda = \delta_1$, then n is positive recurrent for L if and only if there exists $k \in \mathbb{N}$ such that $n = n_1 + \ldots + n_k$ for $n_\ell \in \{i \in \mathbb{N} : \beta_{i+1} > 0\}$, $\ell \in [k]$.*

In particular, if $\beta_2 > 0$, then L is positive recurrent.

Proof. Let L be the leaf process with parameter (β, Λ) with $\Lambda \neq \delta_1$ and let $q(n,j)$ be the corresponding transition rates. Consider n_0 from Lemma 5.4. Define another Markov chain $\hat{L}$ with transition rates

$$\hat{q}(n,j) := \begin{cases} 0, & \text{if} \quad n < n_0,\ j \in \{n_0, \ldots, n_0 + m - 1\}, \\ \sum_{k=n_0}^{n_0+m-1} q(n,k), & \text{if} \quad n < n_0,\ j = n_0 + m, \\ q(n,j), & \text{otherwise.} \end{cases}$$

Define $\hat{T}^n$ and $\hat{\mathcal{T}}_n$ for $\hat{L}$ as the analogue of T^n and $\mathcal{T}_n$ for L. Since the transition rates of L and $\hat{L}$ agree for $n \geq n_0$, then we have $\mathbb{E}_n[\hat{T}^{n_0}] = \mathbb{E}_n[T^{n_0}] < \tilde{c}f(n)$. Let $T(n_0) := \inf\{t \geq 0 : L_t \geq n_0\}$. Analogously, define $\hat{T}(n_0)$ for $\hat{L}$. Clearly, $\mathbb{E}_k[T(n_0)] = \mathbb{E}_k[\hat{T}(n_0)] < \infty$ for all $k < n_0$. Hence,

$$\mathbb{E}_{n_0+m}[\hat{\mathcal{T}}_{n_0+m}] \leq \mathbb{E}_{n_0+m}[\hat{T}^{n_0}] + \sum_{k=1}^{n_0-1} \mathbb{E}_k[\hat{T}(n_0)] < \infty$$

and so, $n_0 + m$ is positive recurrent for $\hat{L}$. By irreducibility (since $\Lambda \neq \delta_1$), 1 is positive recurrent for $\hat{L}$. Assume that $L_0 = \hat{L}_0 = n$ for some $n > n_0$. Clearly, there exists a coupling such that $L_t \leq \hat{L}_t$ for all $t \geq 0$. In particular, since state 1 is positive recurrent for $\hat{L}$, 1 is also positive recurrent for L. By irreducibility, all states of L are positive recurrent.
If $\Lambda = \delta_1$, then $\mathbb{E}_n[\mathcal{T}_1] < \infty$. Hence, 1 is positive recurrent for L. The result follows, since the communication class of 1 consists of the numbers described in the second part of the statement. ☐

The next results agrees with Foucart [26, Lem.2.5] in the case of genic selection.

Lemma 5.6. *If $b(\beta) > c(\Lambda)$, then L is transient.*

Proof. Again, we mimic the proof of Foucart [27, Lem. 0.1]. Assume that there is $n_0 \in \mathbb{N}$ and a bounded strictly decreasing function g that is chosen such that $\mathcal{L}g(n) < 0$ for all $n \geq n_0$. The process $(g(L_{t\wedge T^{n_0}}))_{t\geq 0}$ started from $n \geq n_0$ is a supermartingale. By the martingale convergence theorem, $\lim_{t\to\infty} \mathbb{E}_n[g(L_{t\wedge T^{n_0}})] \leq g(n) < g(n_0)$ and so $\mathbb{P}_n(T^{n_0} < \infty) < 1$. Decompose

$$\mathbb{P}_{n_0-1}(T(n_0-1) < \infty) = \sum_{n<n_0-1} \mathbb{P}_n(T(n_0-1) < \infty)\mathbb{P}(L_{J_1} = n) + \sum_{n\geq n_0}^{\infty} \mathbb{P}_n(T(n_0-1) < \infty)\mathbb{P}(L_{J_1} = n).$$

Since for $n \geq n_0$, $\mathbb{P}_n(T(n_0-1) < \infty) < \mathbb{P}_n(T^{n_0} < \infty) < 1$, it follows that $\mathbb{P}_{n_0-1}(T(n_0-1) < \infty) < 1$ and so L is transient [48, Thm. 4.3.2]. As we will show, the conditions are indeed satisfied for the function

$$g(n) := \frac{1}{\log(n+1)}.$$

It is proven in Foucart [27, p.2] that

$$\mathcal{L}_\Lambda g(n) = \frac{1}{\log(n+1)\log(n+2)}\big(c(\Lambda) + o(1)\big).$$

Our claim is that $\mathcal{L}_\beta g(n) \leq \frac{1}{\log(n+1)\log(n+2)}(-b(\beta)+o(1))$. It then follows together with the just mentioned result of Foucart [27], that

$$\mathcal{L}g(n) = \frac{1}{\log(n+2)\log(n+1)}\big(c(\Lambda) - b(\beta) + o(1)\big) < 0,$$

for n large enough. It remains to prove the claim. Note that,

$$\begin{aligned}\mathcal{L}_\beta g(n) &= \sum_{\ell=2}^m \beta_\ell \frac{1}{\log(n+\ell)\log(n+1)} n \log\Big(\frac{n+1}{n+\ell}\Big)\\ &= -\sum_{\ell=2}^m \beta_\ell \frac{1}{\log(n+\ell)\log(n+1)}\big((\ell-1)+o(1)\big)\\ &\leq -\frac{1}{\log(n+m)\log(n+1)}\big(b(\beta)+o(1)\big),\end{aligned}$$

where in the second equality, we use that $n\log(\frac{n+1}{n+\ell}) = -((\ell-1)+o(1))$. At last, since $\frac{\log(n+2)}{\log(n+m)} = 1-o(1)$, we have

$$-\frac{1}{\log(n+2)\log(n+1)}\frac{\log(n+2)}{\log(n+m)}\big(b(\beta)+o(1)\big) \leq \frac{1}{\log(n+2)\log(n+1)}\big(-b(\beta)+o(1)\big),$$

which proves the claim. □

Proof of Theorem 2.14. The result follows from Corollary 5.5 and Lemma 5.6. □

5.2. **Siegmund duality and Fearnhead-type recursions for the leaf process.** It is well-known that in the neutral case and in the case of genic selection, the leaf process L is in Siegmund duality with the fixation line (shifted by -1) [3, Rem. 4.6]. Loosely speaking, the fixation line codes how a new most recent common ancestor establishes itself in the population, see [1, 29, 35, 49] for more details. In this section, we extend the duality to our general setup and use it as an analytical tool to derive recursions for the stationary tail-probabilities of the process L (in the positive recurrent case). We subsequently use these recursions to deduce that the leaf process has exponential moments if $\Lambda(\{0\}) > 0$. For (β, p, Λ) with $(\beta, p) \in \mathbb{E}^m$, we define the process $D := (D_t; t \geq 0)$ on $\mathbb{N} \cup \{\infty\}$ with the infinitesimal generator $Gg(d) := G_\Lambda g(d) + G_\beta g(d)$, where

$$\begin{aligned}G_\Lambda g(d) &:= \sum_{c\geq 1}\binom{d+c-1}{c+1}\lambda_{c+d,c+1}[g(c+d)-g(d)] + \mathbf{1}_{\{d\geq 2\}}\Lambda(\{1\})[g(\infty)-g(d)],\\ G_\beta g(d) &:= \sum_{r=1}^{(m\wedge d)-1}\Big((d-r)\beta_{r+1} + \sum_{k=r+1}^{m-1}\beta_{k+1}\Big)[g(d-r)-g(d)].\end{aligned}$$

Note that 1 is an absorbing state of D.

Lemma 5.7 (Siegmund duality)**.** *The process $(D_t; t \geq 0)$ and $(L_t; t \geq 0)$ are Siegmund dual, i.e. for $t \geq 0$*

$$\mathbb{P}_\ell(d \leq L_t) = \mathbb{P}_d(D_t \leq \ell), \qquad \forall\, d, \ell \in \mathbb{N}, \tag{5.3}$$

Proof. Set $\hat{H}(d,\ell) = \mathbf{1}_{\{d\leq \ell\}}$ for $d, \ell \in \mathbb{N}$. We will show that $\mathcal{L}\hat{H}(d,\cdot)(\ell) = G\hat{H}(\cdot,\ell)(d)$. The result then follows by [38, Prop. 1.2]. First, note that the result in the neutral case (see [3, Eq. (4.1), Lem. 4.5]) implies that the processes generated by $\mathcal{L}_\Lambda$ and G_Λ are dual with respect to $\hat{H}$. Hence, $\mathcal{L}_\Lambda\hat{H}(d,\cdot)(\ell) = G_\Lambda\hat{H}(\cdot,\ell)(d)$. Thus, it remains to prove $\mathcal{L}_\beta\hat{H}(d,\cdot)(\ell) = G_\beta\hat{H}(\cdot,\ell)(d)$. Clearly, for $d \leq \ell$ we have $\mathcal{L}_\beta\hat{H}(d,\cdot)(\ell) = 0 = G_\beta\hat{H}(\cdot,\ell)(d)$. In addition, for $d > \ell$, a straightforward calculation yields

$$\begin{aligned}&G_\beta\hat{H}(\cdot,\ell)(d) - \mathcal{L}_\beta\hat{H}(d,\cdot)(\ell)\\ &\quad= \sum_{r=2}^{(d\wedge m)-1}(d-r)\beta_{r+1}\mathbf{1}_{\{d\leq \ell+r\}} + \sum_{r=2}^{m-1}\beta_{r+1}(\ell+d\wedge r-d)\mathbf{1}_{\{d<\ell+d\wedge r\}} - \ell\sum_{r=2}^{m-1}\beta_{r+1}\mathbf{1}_{\{d\leq \ell+r\}}.\end{aligned}$$

The result follows by simple inspection of the cases $m \leq d$ and $m > d$. □

For the remainder of this section, we assume $b(\beta) < c(\lambda)$ such that L is positive recurrent (in the communication class of 1). Let L_∞ be a random variable distributed according to the stationary distribution of L and define $a_n := \mathbb{P}(L_\infty > n)$, $n \in \mathbb{N}$. By the Siegmund duality, we deduce that $a_n = \mathbb{P}_{n+1}(D \text{ absorbs in } 1)$. This relation is exploited in [3] in order to obtain a generalisation of the Fearnhead recursion [24] for a Λ-Wright–Fisher process with genic selection. We further generalise this recursions to our setting.

Proposition 5.8. *The tail-probabilities are the unique solution to the system of equations*

$$\sum_{c\geq 2}\binom{n+c-1}{c}\lambda_{c+n,c}[a_n-a_{c+n-1}]+\Lambda(\{1\})a_n=\sum_{r=1}^{(m-1)\wedge n}\Big((n+1-r)\beta_{r+1}+\sum_{k=r+1}^{m-1}\beta_{k+1}\Big)[a_{n-r}-a_n]. \quad (5.4)$$

with boundary conditions $a_0 = 1$ and $\lim_{n\to\infty} a_n = 0$.

Proof. Clearly, the boundary conditions hold because a_n is a tail probability. Set $h(n) := a_{n-1}$, $n \in \mathbb{N}$. Lemma 5.7 implies that $h(n) = \mathbb{P}_n(D \text{ absorbs in } 1)$. Hence h is harmonic for G, i.e. $Gh(n) = 0$, and Eq. (5.4) follows. For the uniqueness, we follow the proof of [3, Thm. 2.4]. Denote by $a' = (a'_n)_{n\geq 0}$ another solution of the recursion and set $g(n) := a_{n-1} - a'_{n-1}$ for $n \in \mathbb{N}$. Hence, $g(1) = 0$, and $\lim_{n\to\infty} g(n) = 0$. Since $n \mapsto a_{n-1}$ and $n \mapsto a'_{n-1}$ are both harmonic for G, also g is harmonic for G. In particular, $g(D_t) - \int_0^t Gg(D_s)ds = g(D_t)$, $t \geq 0$. Hence, $(g(D_t); t \geq 0)$ is a bounded martingale. Let $T_{1,k} := \inf\{t \geq 0 : D_t \in \{1, k, k+1, \dots\}\}$. Then $T_{1,k}$ is finite almost surely for every $k \in \mathbb{N}$. If $D_0 = d$, the optional stopping theorem yields $g(d) = \mathbb{E}_d[g(D_{T_{1,k}})]$ for all $k \in \mathbb{N}$. But $g(D_{T_{1,k}}) \to 0$ as $k \to \infty$. Hence, by the dominated convergence theorem, $g(d) = 0$ for all $d \in \mathbb{N}$ and so $a' = a$. □

Remark 5.2. In the case of genic selection, [12] give a general approach to solve (5.4). Moreover, they provide explicit solutions for the Kingman case, the start-shaped case and the Bolthausen-Sznitman case.

Corollary 5.9. *If $\Lambda(\{0\}) > 0$, then L_∞ has exponential moments of all orders, i.e. $\mathbb{E}[\exp(xL_\infty)] < \infty$ for all $x \in \mathbb{R}$.*

Proof. Define $q_n := \mathbb{P}(L_\infty = n) = a_{n-1} - a_n$. A straightforward manipulation of Eq. (5.4) yields that $q_0 = 0$, $\sum_{n\geq 1} q_n = 1$, and

$$\sum_{k\geq n+1} q_k\left(c_{n,k}+\frac{\Lambda(\{1\})}{n}\right)=\sum_{k=(n-m+2)\vee 1}^{n} q_k b_{n,k}, \qquad n\in\mathbb{N} \quad (5.5)$$

where

$$c_{n,k} := \frac{1}{n}\sum_{\ell\geq k}\binom{\ell}{\ell-n+1}\int_{[0,1)} r^{\ell-n-1}(1-r)^n\Lambda(dr), \quad b_{n,k} := \sum_{r=n-k+1}^{m-1}\frac{k\wedge(k-r+n+1)}{n}\beta_{r+1}.$$

Note first that

$$\sum_{k\geq n+1} q_k\left(c_{n,k}+\frac{\Lambda(\{1\})}{n}\right) \geq q_{n+1}c_{n,n+1} \geq q_{n+1}\,\frac{(n+1)}{2}\,\Lambda(\{0\}).$$

In order to get an upper bound for the right-hand side of (5.5), we set some notation. For $n \in \mathbb{N}$, define

$$\hat{q}_n := \max\{q_k : k \in [n-1]\setminus[n-m]\} \quad \text{and} \quad r(n) := \min\{k \in [n-1]\setminus[n-m] : q_k = \hat{q}_n\},$$

with the convention $[k] = \emptyset$ for $k \leq 1$. Write r^i for the i-th composition of r, i.e. $r^i(n) = r^{i-1}(r(n))$ and $r^0(n) = n$. By construction, we have $r(n) \leq n-1$ for $n > 1$, and hence, $r^{n-1}(n) = 1$. Set

$$\hat{d}(n) := \min\{i \in [n-1] : r^i(n) = 1\} \leq n-1.$$

Clearly,

$$\sum_{k=(n-m+2)\vee 1}^{n} q_k b_{n,k} \leq m|\beta| q_{r(n+1)},$$

where $|\beta| := \sum_{\ell=2}^m \beta_\ell$. Therefore, we obtain

$$q_n \leq \frac{1}{n}\frac{2m|\beta|}{\Lambda(\{0\})}q_{r(n)} \leq \frac{1}{\prod_{i=0}^{\hat{d}(n)-1} r^i(n)}\left(\frac{2m|\beta|}{\Lambda(\{0\})}\right)^{\hat{d}(n)} q_1.$$

We claim that $r^i(n) > \hat{d}(n) - i$ for all $i \in [\hat{d}(n)]_0$ and $n > 1$. We proceed by induction. Since $\hat{d}(n) \leq n-1$, the claim is true for $i = 0$. Now, let us assume the result is true for $i-1$ and let $n > 1$ such that $i \in [\hat{d}(n)]_0$. Since $\hat{d}(r(n)) = \hat{d}(n) - 1$, we obtain

$$r^i(n) = r^{i-1}(r(n)) > \hat{d}(r(n)) - i + 1 = \hat{d}(n) - i,$$

which proves the claim. As a consequence, we obtain that $\prod_{i=0}^{\hat{d}(n)-1} r^i(n) \geq \hat{d}(n)!$, and hence,

$$q_n \leq \frac{1}{\hat{d}(n)!} \left(\frac{2m|\beta|}{\Lambda(\{0\})} \right)^{\hat{d}(n)} q_1.$$

Since, $n - r(n) \leq m$, we deduce that $n/m - 1 \leq \hat{d}(n) \leq n - 1$. The result follows.

□

Remark 5.3. The system of equations (5.5), complemented by the boundary conditions $q_0 = 0$ and $\sum_n q_n = 1$, is equivalent to the infinite system $0 = q\mathcal{L}$, in the sense that both characterise the stationary probabilities $(q_n)_{n \in \mathbb{N}_0}$. However, in many situations is easier to deal with (5.5). For the sake of illustration, let us consider the case where $\Lambda = \delta_0$ and $\beta = 0$, which corresponds to the Kingman coalescent. In this particular setting, the leaf process is absorbed at 1 after a finite time so that $q_n = \mathbf{1}_{\{n=1\}}$. On the one hand, the typical condition $0 = q\mathcal{L}$ yields $\frac{n(n+1)}{2} q_{n+1} - \frac{n(n-1)}{2} q_n = 0$ for every $n \geq 2$. On the other hand, (5.5) reads $q_{n+1} = 0$ for every $n \in \mathbb{N}$, so it directly yields the solution to the infinite system.

5.3. **Invariant measure of the Bernstein coefficient process.** Let us now turn to the analysis of the asymptotic behaviour of the Bernstein coefficient process. Before we begin let us make the following remark. Even though the state space $\mathbb{R}^\infty$ is non countable, V is morally a Markov chain. To see this, note first that the space $\mathbb{M}_V$ of finite matrices that can be obtained as the product of a finite number of (compatible) selection and coagulation matrices is countable. Moreover, one can use the rates of the process V to define a continuous-time Markov chain $M := (M_t; t \geq 0)$ on $\mathbb{M}_V$, such that if M_0 is the identity matrix of size $\dim(V_0)$, then $V_t = M_t V_0$ for all $t \geq 0$. Hence, the dynamics of the process V is completely determined by the Markov chain M.

Proof of Proposition 2.15. It directly follows from Lemma 4.4 that $V_t(0)$ and $V_t(L_t)$ are constant along time. We now prove the other parts.

(1) Fix $a, b \in \mathbb{R}$. Denote by

$$C_V(a,b) := \left\{ M \begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^\infty : \ M \in \mathbb{M}_V \right\},$$

the set of points that can be reached by V starting from $(a,b)^T$. From Lemma 4.4, the set $C_V(a,b)$ is invariant for V. Moreover, by definition, $C_V(a,b)$ forms a communication class of V, which contains $(a,b)^T$. Indeed, for $w, w' \in C_V(a,b)$, in order to go from w to w', first go from w to $(a,b)^T$ by successive coagulation operations. Then go from $(a,b)^T$ to w' in a finite number of successive selection and coagulation operations. By Theorem 2.14, the assumption $b(\beta) < c(\Lambda)$ implies that L is positive recurrent. In particular, the state $(a,b)^T$ is positive recurrent for V. Furthermore, the restriction of V to $C_V(a,b)$ is irreducible and positive recurrent. Hence, there exists a unique invariant distribution $\mu^{a,b}$ [48, Thm. 3.5.2, Thm. 3.5.3] for V restricted to $C_V(a,b)$. It remains to see that $\mu^{a,b}$ is the unique stationary distribution of V with support included in $\widehat{C}_V(a,b) := \{v \in \mathbb{R}^\infty : v_0 = a, \ v_{\dim(v)-1} = b\}$. But this follows directly noting that, since L is positive recurrent, the process V starting in $V_0 = v \in \widehat{C}_V(a,b)$ enters $C_V(a,b)$ in finite time.

(2) Let $V_\infty^{a,b} \sim \mu^{a,b}$. If $V_0 = (a,b)^T$, then $V_t \xrightarrow[t \to \infty]{(d)} V_\infty^{a,b}$ in law by classic Markov chain theory [48, Thm. 3.6.2]. On the other hand, for $n \in \mathbb{N}_0$ and $v \in \mathbb{R}^{n+1}$ with $v_0 = a$ and $v_n = b$, as remarked above, V enters $C_V(a,b)$ in finite time. Hence, the result follows.

□

Remark 5.4. Let us provide more details on the invariant measures of Proposition 2.15. We start with two remarks. Let $t_1 > t_0$. From the definition of the dynamics, V_{t_1} can be recovered deterministically from V_{t_0} and the trajectory of the process $(L_t; [t_0, t_1])$, i.e. by observing the succession of coagulation and selection events. The second remark is that Lemma 4.4 implies that for every t such that $L_t = 1$, we have $V_t = (a, b)^T$. Since the process $(L_s; s \geq 0)$ is positive recurrent, it can be decomposed into its finite successive excursions away from state 1. Let us denote by e_t the excursion straddling time t. More precisely, for $i \geq 1$, let $\tilde{T}_i$ be the i-th return time to state 1 of L and define

$$\mathrm{Ex}_i \;:=\; \{L_{t+\tilde{T}_i} : t \in [0, \tilde{T}_{i+1} - \tilde{T}_i)\},$$

the i-th excursion. Finally, whenever $\tilde{T}_1 < t$, define

$$\tau_t \;:=\; \inf\{i \geq 1 : \tilde{T}_i \geq t\} \;\text{ and }\; e_t \;:=\; \mathrm{Ex}_{\tau_t},$$

i.e. at time t, the leaf process is in excursion τ_t. From the two previous remarks, it is clear that there exists a measurable map G (which only depends on the values of a and b) such that $V_t = G(e_t(t - \tau_t))$. Next, define $\tilde{e}_0$ the duration-biased excursion (from state a) of the leaf process

$$\mathbb{E}\left[f(\tilde{e}_0)\right] \;:=\; \frac{1}{\mathbb{E}_0[d_0]}\mathbb{E}\left[f(e_0)d_0\right],$$

where e_0 is an excursion of the leaf process and d_0 its duration (i.e. d_0 is the first returning time to 0). Since L is positive recurrent, the duration d_0 has a finite first moment. From standard renewal theory and the previous lemma, we have

$$(e_t, \tau_t - t) \Longrightarrow (e_\infty, U_\infty),$$

where e_∞ is the duration-biased excursion, and conditional on e_∞, the random variable U_∞ is a uniform random variable on $[0, d(e_\infty)]$. Since U_∞ is a.s. not a jumping time for e_∞, we get

$$e_t(\tau_t - t) \Longrightarrow e_\infty(U_\infty),$$

and

$$V_t \Longrightarrow V_\infty^{a,b} = G(e_\infty(U_\infty)).$$

5.4. **Coming down from infinity.** Let us consider the leaf process $(L_t^n;\, t \geq 0)$ started at n. It is not hard to see that the leaf process $(L_t^n;\, t \geq 0)$ is monotone in n, i.e. there exists a coupling such that for every $n > m$,

$$\forall t \geq 0, L_t^n \geq L_t^m \;\text{ a.s.}$$

As a consequence, we can always define a process $(L_t^\infty;\, t > 0)$ valued in $\mathbb{N}_0 \cup \{\infty\}$ as the monotone limit of the sequence processes $\{(L_t^n;\, t > 0)\}_{n \in \mathbb{N}_0}$. Note that $\lim_{t\to 0^+} L_t^\infty = +\infty$.
Next, we prove the generalisation of a result of [50, Prop. 23].

Proof of Theorem 2.17. Recall from [50] that if $\Lambda(\{1\}) = 0$, then the block counting of Λ-coalescent counting process either stays infinite with probability 1, or c.d.i.with probability 1. The same arguments that are used in the proof of Theorem 4 and Proposition 23 of [50] can be extended to our coalescing-branching system. In particular, the same dichotomy also holds here.
Let us now show that L c.d.i. if and only if the underlying Λ-coalescent c.d.i.. Since the leaf process stochastically dominates the underlying Λ-coalescent, it suffices to show that if the underlying Λ-coalescent c.d.i., so does the leaf process. This directly follows from Lemma 5.2 and by letting $n \to \infty$ in the inequality in Lemma 5.4. □

The next proposition states that the coming down from infinity property is a stronger property than the positive recurrence of the leaf process.

Proposition 5.10. *If L comes down from ∞, then L is positive recurrent*

Proof. Assume that L c.d.i.. From Theorem 2.17, the underlying Λ-coalescent also c.d.i.. From [36] (see also [26, Thm. 2.2]), this implies $\sum_{k=2}^\infty \delta(k)^{-1} < \infty$. Since $\delta(k)/k \to c(\Lambda)$, we must have $c(\Lambda) = \infty$ and the proposition is then a direct application of Theorem 2.14. □

Let us now prove the result about the entrance law at ∞.

Proof of Proposition 2.18. The second item directly follows from the first one since the processes $\{(V_t^n;\, t \geq 0)\}$ have different initial conditions but evolve according to the same dynamics. We now turn to the proof of the first item of the proposition. For every $n \in \mathbb{N}$, let $\mathcal{G}_t^n$ be the ASG with n roots at time t. Let V_t^n be the random vector (4.1) constructed from the graph $\mathcal{G}_t^n$. Namely, for every $i \leq L_t^n$, the i-th coordinate $V_t^n(i)$ is the probability of having all the roots of type a (resp, A) if we assign uniformly i leaves of type a (resp., $L_t^n - i$ leaves of type A) for the ancestral selection graph $\mathcal{G}_t^n$. The consistency of the branching-coalescing system of particles allows the construction of the graphs $\{\mathcal{G}_t^n\}_{n=1}^{\infty}$ on the same probability space so that that for every $m < n$ the graph $\mathcal{G}_t^m$ is a subgraph of $\mathcal{G}_t^n$, with the set of leaves $\mathcal{G}_t^m$ included in the set of leaves of $\mathcal{G}_t^n$. (In particular $L_t^n \geq L_t^m$ for every $t \geq 0$.) It is then clear from our colouring algorithm that

$$0 \leq V_t^n(i) \leq V_t^m(i), \quad \forall t \in \mathbb{R}_+, \forall i \leq L_t^n. \tag{5.6}$$

(Indeed, on top of the coupling between the graphs $\mathcal{G}_t^m$ and $\mathcal{G}_t^n$, one can couple the colouring in such a way that the trace of colouring of $\mathcal{G}_t^n$ on the subgraph $\mathcal{G}_t^m$ is identical in law with the colouring of $\mathcal{G}_t^m$.) Since the leaf process L c.d.i., and $\{L_t^n\}_{i=1}^{\infty}$ is increasing in n, for every $0 < t_1 < t_2$, the process $(L_t^n;\, t \in [t_1, t_2])$ coincides with $(L_t^\infty;\, t \in [t_1, t_2])$ after a certain rank. As a consequence, the number of jumps of the process $(V_t^n;\, t \in [t_1, t_2])$ is controlled by the ones of L_∞. Finally, the monotonicity property (5.6) ensures that there exists a process $(V_t^\infty;\, t > 0)$ such that

$$(V_t^n;\, t \in [t_1, t_2]) \;\to\; (V_t^\infty;\, t \in [t_1, t_2]) \;\text{ a.s.}$$

in the Skorohod topology. (The monotonicity property ensures the a.s. convergence for every fixed t, the control on the number of jumps then ensures convergence in the Skorohod topology.)

□

6. Applications: Absorption probabilities and absorption time

In this section, we illustrate how one can make use of the results from the previous sections. The main results provide conditions for the accessibility of both boundary points, expressions for the absorption probabilities and times.

6.1. **Absorption probabilities.**

In order to prove Proposition 2.19, we start with a useful lemma.

Lemma 6.1. *The following three statements are equivalent*

1. *For all $x \in [0,1]$*
$$\lim_{t\to\infty} \mathbb{E}_{e_2}\big[\,\langle B_{L_t}(x), V_t\rangle\,\big] = \lim_{t\to\infty} \mathbb{E}_{e_3}\big[\,\langle B_{L_t}(x), V_t\rangle\,\big] = p(x), \quad \textit{for some} \quad p(x) \in [0,1]. \tag{6.1}$$
2. *Let $x \in [0,1]$. For $X_0 = x$, the limit $X_\infty := \lim_{t\to\infty} X_t$ exists almost surely and $X_\infty \sim \mathrm{Ber}(p(x))$ for some $p(x) \in [0,1]$.*
3. *For all $x \in [0,1]$ and $v \in \mathbb{R}^{n+1}$ with $n \geq 2$,*
$$\lim_{t\to\infty} \mathbb{E}_v\big[\,\langle B_{L_t}(x), V_t\rangle\,\big] = \big(1 - p(x)\big)\, v_n + p(x)\, v_0,$$
for some $p(x) \in [0,1]$.

Proof.
(1) $\to$ (2) Let $x \in [0,1]$. The duality in combination with (6.1) yields

$$\lim_{t\to\infty} \mathbb{E}_x\big[X_t\big] = \lim_{t\to\infty} \mathbb{E}_{e_2}\big[\,\langle B_{L_t}(x), V_t\rangle\,\big] = p(x) = \lim_{t\to\infty} \mathbb{E}_{e_3}\big[\,\langle B_{L_t}(x), V_t\rangle\,\big] = \lim_{t\to\infty} \mathbb{E}_x\big[X_t^2\big]. \tag{6.2}$$

Denote by μ_t the law of X_t. Then, $\mu_t \in \mathcal{M}_1[0,1]$. Every collection of probability measures on $[0,1]$ is tight, since $[0,1]$ is compact. Therefore, by Prokhorov's Theorem, for every sequence of probability measures $(\mu_{t_k})_{k\geq 0}$ with $t_k \nearrow \infty$ as $k \to \infty$, there exists a weakly convergent subsequence $(\mu_{t_{k_l}})_{l\geq 0}$ with $\mu_{t_{k_l}} \to \mu$ as $l \to \infty$. Let X be a random variable distributed according to μ. Then,

$$\mathbb{E}_\mu[X(1-X)] = \lim_{k\to\infty} \mathbb{E}_x[X_{t_k}(1 - X_{t_k})] = 0,$$

due to (6.2). Furthermore, $\mu \sim \mathrm{Ber}(p(x))$ and this is independent of the choice of the subsequence. Hence, $\lim_{t\to\infty} X_t$ exists almost surely and has distribution μ.

(2) $\to$ (3) Assume that $X_0 = x$, $X_\infty = \lim_{t\to\infty} X_t$ exists almost surely, and $X_\infty \sim \mathrm{Ber}(p(x))$ for some $p(x) \in [0,1]$. Let $v \in \mathbb{R}^{n+1}$, with $n \geq 2$. Then,

$$(1-p(x))v_n + p(x)v_0 = \mathbb{E}_x\big[\,\langle B_n(X_\infty), v\rangle\,\big] = \lim_{t\to\infty} \mathbb{E}_x\big[\,\langle B_n(X_t), v\rangle\,\big] = \lim_{t\to\infty} \mathbb{E}_v\big[\,\langle B_{L_t}(x), V_t\rangle\,\big].$$

(3) $\to$ (1) Fix $x \in [0,1]$ and choose $v = e_2$ and $v' = e_3$. Since $v_2 = v_3' = 1$ and $v_0 = v_0' = 0$, the result follows. $\square$

Proof of Proposition 2.19. Assume $b(\beta) < c(\Lambda)$. Let $V_0 = e_2$. By Proposition 2.15, $V_t \xrightarrow{(d)} V_\infty$, where $V_\infty \sim \mu^{0,1}$. Denote by $p(x) = \mathbb{E}[\langle B_{L_\infty}(x), V_\infty\rangle]$. Then,

$$\lim_{t\to\infty} \mathbb{E}[X_t] = \lim_{t\to\infty} \mathbb{E}_{e_2}[\langle B_{L_t}(x), V_t\rangle] =: p(x). \tag{6.3}$$

Also by Proposition 2.15, $\lim_{t\to\infty} \mathbb{E}_{e_3}[\langle B_{L_t}(x), V_t\rangle] = p(x)$. Furthermore,

$$p(x) = \mathbb{E}\Big[\sum_{i=0}^{L_\infty} V_\infty(i) b_{i,L_\infty}(x)\Big] = \sum_{\ell=1}^{\infty} \mathbb{P}(L_\infty = \ell) \sum_{i=0}^{\ell} d_{i,\ell} b_{i,\ell}(x), \tag{6.4}$$

where $d_{i,\ell} := \mathbb{E}[V_\infty(i) \mid L_\infty = \ell]$. Since $\|V_t\| \leq 1$ for all $t \geq 0$, we have $d_{i,\ell} \leq 1$, by Lemma 4.4, such that also $p(x) \in [0,1]$. Also by Lemma 4.4, $d_{0,\ell} = 0$ and $d_{\ell,\ell} = 1$. Hence, if $x \in (0,1)$,

$$0 < \sum_{\ell=1}^{\infty} \mathbb{P}(L_\infty = \ell) x^\ell \leq p(x) \leq \sum_{\ell=1}^{\infty} \mathbb{P}(L_\infty = \ell)(1-(1-x)^\ell) < 1. \tag{6.5}$$

In particular, we are in the framework of Proposition 6.1. Hence, $X_\infty := \lim_{t\to\infty} X_t$ exists almost surely and $X_\infty \sim \mathrm{Ber}(p(x))$. Consequently, (1) holds. Furthermore,

$$h(x) = \mathbb{P}_x(T_0 < T_1) = \lim_{t\to\infty} \mathbb{E}[X_t] = p(x),$$

such that (2) follows by (6.3). Then, (3) follows by (6.5). $\square$

For $\ell \in \mathbb{N}$ and $v \in \mathbb{R}^{\ell+1}$, let $\bar{c}_{k,\ell}(v)$ be the k-th coefficient in the monomial basis of $\langle B_\ell(x), v\rangle$, where $k \in [\ell]_0$. Inspired by Proposition 2.19, a naive guess is that $h(x) = \sum_{k=0}^{\infty} x^k \mathbb{E}[\bar{c}_{k,L_\infty}(V_\infty)]$. We make this precise in the next lemma.

Proposition 6.2. *Assume $b(\beta) < c(\Lambda)$. If L_∞ admits exponential moments of order* $\ln(3)$*, then h is analytic and has the series representation $h(x) = \sum_{k=1}^{\infty} c_k x^k$, where $c_k = \sum_{\ell=k}^{\infty} \mathbb{E}\left[\bar{c}_{k,\ell}(V_\infty)\mathbf{1}_{\{L_\infty=\ell\}}\right]$. Further,*

$$\forall x \in [0,1], \quad h(x) \leq \mathbb{E}[(1+2x)^{L_\infty}] - 1$$

Remark 6.1. If $\Lambda(\{0\}) > 0$, then L_∞ has exponential moments of all orders (see Corollary 5.9). In this case h is also harmonic for the infinitesimal generator and hence, it is then possible to derive a system of equations for the c_k.

Proof of Proposition 6.2. First note that for $\ell \in \mathbb{N}$ and $v \in \mathbb{R}^{\ell+1}$, a straightforward computation yields that for v with $v_0 = 0$, we have for $k \geq 1$

$$\bar{c}_{k,\ell}(v) = \sum_{i=1}^{k} \binom{\ell}{k}\binom{k}{i}(-1)^{k-i} v_i,$$

complemented by $\bar{c}_{0,\ell}(v) = 0$. In particular, $\langle B_\ell(x), v\rangle = \sum_{k=1}^{\ell} \bar{c}_{k,\ell}(v) x^k$. If $b(\beta) < c(\Lambda)$, the leaf process is positive recurrent. Hence, we can start from (6.4). Provided that one can exchange the following two infinite sums, a straightforward formal calculation yields

$$h(x) \;=\; \sum_{\ell=1}^{\infty}\sum_{k=1}^{\ell} x^k \mathbb{E}[\bar{c}_{k,\ell}(V_\infty)\mathbf{1}_{\{L_\infty=\ell\}}] \;=\; \sum_{k=1}^{\infty} x^k \sum_{\ell=k}^{\infty} \mathbb{E}[\bar{c}_{k,\ell}(V_\infty)\mathbf{1}_{\{L_\infty=\ell\}}].$$

Let us justify the interchange of the two sums in the last equality. This requires the absolute convergence of the series. Note that $\|V_\infty\|_\infty \leq 1$ (by means of Lemma 4.4). As a consequence $|\bar{c}_{k,\ell}(V_\infty)| \leq \sum_{i=0}^k \binom{\ell}{k}\binom{k}{i} = \binom{\ell}{k}2^k$ so that

$$\sum_{k=1}^\infty x^k \sum_{\ell=1}^\infty \mathbb{E}[|\bar{c}_{k,\ell}(V_\infty)|\mathbf{1}_{\{L_\infty=\ell\}}] \;\leq\; \sum_{\ell=1}^\infty \mathbb{P}(L_\infty=\ell)\sum_{k=1}^\ell \binom{\ell}{k}(2x)^k \;=\; \mathbb{E}[(1+2x)^{L_\infty}]-1.$$

In particular the series is absolutely convergent for all x if we have exponential moments of order $\ln(1+2x)$. This is indeed the case under our assumption, and hence the interchange of the summation is justified. □

6.2. **Absorption time.** Recall from Section 2.8, $\bar{p}_{i,\ell} = 1-p_{\ell-i,\ell}$. Let $V^n := (V_t^n;\, t>0)$ and $V^\infty := (V_t^\infty;\, t>0)$ be defined as in Proposition 2.18 for (β, p, Λ). Let $W^n := (W_t^n;\, t>0)$ and $W^\infty := (W_t^\infty;\, t>0)$ be the analogue, but defined w.r.t. $(\beta,\bar{p},\Lambda)$. Since the leaf process only depends on the pair β and Λ, the ASG and the leaf process are identical for the two sets of parameters (only the colouring algorithm for the graphs are different) so that the processes V^n and W^n are naturally coupled.

Lemma 6.3. *For every $t>0$ and $n\in\mathbb{N}$,*

$$\mathbb{E}_x\big[(1-X_t)^n\big] \;=\; \mathbb{E}\big[\big\langle B_{L_t^n}(1-x), W_t^n\big\rangle\big].$$

Proof. Let $Y_t = 1-X_t$. Then $Y:=(Y_t;\, t\geq 0)$ is identical in law to the solution of the SDE

$$dY_t \;:= -d(1-Y_t)\,dt+\sqrt{\Lambda(\{0\})\,Y_t(1-Y_t)}\,dW_t+\int\limits_{(0,1]\times[0,1]}\big(\mathbf{1}_{\{u\leq Y_{t-}\}}r(\mathbf{1}-Y_{t-})-\mathbf{1}_{\{u>Y_{t-}\}}rY_{t-}\big)\,\tilde{N}(dt,dr,du).$$

Note that

$$-d(1-x)=\sum_{\ell=2}^m\beta_\ell\sum_{i=0}^\ell b_{\ell-i,\ell}(x)\Big(-p_{i,\ell}+\frac{i}{\ell}\Big)=\sum_{\ell=2}^m\beta_\ell\sum_{i=0}^\ell b_{i,\ell}(x)\Big(\bar{p}_{i,\ell}-\frac{i}{\ell}\Big).$$

By the duality Theorem 2.11, Y is dual to the Bernstein coefficient process with parameters $(\beta,\bar{p},\Lambda)$, which completes the proof of the lemma. □

Proof of Proposition 2.20. First,

$$\begin{aligned}\mathbb{E}_x\big[X_t^n+(1-X_t)^n\big]&=\mathbb{E}_x\big[\mathbf{1}_{\{T\leq t\}}\;(X_t^n+(1-X_t)^n)\big]\;+\;\mathbb{E}_x\big[\mathbf{1}_{\{T>t\}}\;(X_t^n+(1-X_t)^n)\big]\\&=\mathbb{P}_x(T\leq t)\;+\mathbb{E}_x\big[\mathbf{1}_{\{T>t\}}\;(X_t^n+(1-X_t)^n)\big].\end{aligned}$$

By the monotone convergence theorem, the second term on the right goes to 0 as $n\to\infty$. On the other hand, by duality, we have

$$\mathbb{E}_x\big[X_t^n\big]\;=\;\mathbb{E}[\langle B_{L_t^n}(x),V_t^n\rangle],\quad \mathbb{E}_x\big[(1-X_t)^n\big]=\mathbb{E}[\langle B_{L_t^n}(1-x),W_t^n\rangle].$$

The first identity of the proposition then follows by letting $n\to\infty$. For the second identity, first note that $V^\infty_{\tau^{(\infty)}}=W^\infty_{\tau^{(\infty)}}=e_2$. This follows by first applying Lemma 4.4 to V^n and W^n and then in a second step considering $n\to\infty$. Hence, for $t=\tau^{(\infty)}$,

$$Q_t^\infty(x)\;=\;\big\langle B_{L_t^\infty}(x),V_t^\infty\big\rangle+\big\langle B_{L_t^\infty}(1-x),W_t^\infty\big\rangle\;=\;x+(1-x)\;=\;1.$$

Further, from the definition of the coagulation and selection operators, it follows that the latter identity extends to any $t\geq\tau^{(\infty)}$. The second identity is then obtained from the first one by writing $\mathbb{E}_x[T]$ as $\int_0^\infty\mathbb{P}_x(T\geq t)dt$.

□

Corollary 6.4. *If the leaf process c.d.i., then for every $x\in[0,1]$*

$$\mathbb{E}_x\left[T_0\wedge T_1\right]\;\leq\;\mathbb{E}[\tau^{(\infty)}].$$

Proof. By Propositon 2.20,

$$\mathbb{E}_x[T_0\wedge T_1]\;=\;\mathbb{E}\left[\int_0^{\tau^{(\infty)}}(1-Q_t^\infty(x))dt\right]\leq\mathbb{E}[\tau^{(\infty)}],$$

since Q_t^∞ is the sum of two non-negative polynomials. □

7. Minimal ancestral structures

In this section, study minimal ancestral structures. The geometrical characterisation of the set of minimal SDs is derived in Section 7.1. For the sake of illustration, we examine the case $\deg(d) = 2, 3$ in Section 7.2 more closely. The drift term in this case is of the form $d(x) = x(1 - x)(Ax + B)$ with $A, B \in \mathbb{R}$. We explain in more detail the classic diploid selection model with dominance (see Eq. (7.3)), which fits in this framework. In particular, the faces of the polygon $\mathcal{S}_1$ get a natural biological interpretation in terms of under- and overdominance.

In Section 7.3, we show that a minimal selection decomposition $(\beta, p) \in \mathcal{E}_d$ is also graph-minimal, i.e. the branching-coalescing system associated to (β, Λ) (for any finite measure Λ) can not be obtained by thinning the branching-coalescing system of a different pair (β', Λ) with $(\beta', p') \in \mathcal{E}_d$. Loosely speaking, the ASG associated to a minimal SD does not contain "dummy branches".

Finally, in Section 7.4, we show that for $\deg(d) = 3$, the set of graph-minimal SDs coincides with the set of minimal SDs. In particular, the two notions of minimality agree in this case. In higher dimensions, i.e. $\deg(d) > 3$, the situation is more involved, and the question of the equivalence between the two notions of minimality remains as an open problem. Our conjecture is that they are indeed equivalent in any dimension.

In this entire section, d is a polynomial with $\deg(d) = m$ for some $m \in \mathbb{N} \setminus \{1\}$ and $d(0) = d(1) = 0$. The latter property motivates the following abuse of notation. We refer to $\rho(d)$ (instead of $(0, \rho(d)^T, 0)^T$) as the BCV of d.

The following notations will be used in the remainder of the paper. For any compact set $K \subset \mathbb{R}^{m-1}$, $\operatorname{conv}(K)$ denotes its convex hull. For $m \geq 2$, the $m-2$-simplex is defined as

$$\Delta_{m-2} := \left\{ \alpha := (\alpha_\ell)_{\ell=2}^m \in [0,1]^m : \sum_{\ell=2}^m \alpha_\ell = 1 \right\}.$$

7.1. **Finding minimal selection decompositions.** In this section, we derive a geometric characterisation of the set of minimal SDs. The starting point is the geometrical representation of the minimal effective branching rate $b_\star(d)$ as $\inf\{\lambda > 0 : \rho(d) \in \mathcal{S}_\lambda\}$ in Proposition 2.23, where $\mathcal{S}_\lambda \subset \mathbb{R}^{m-1}$ is the set of λ-decomposable vectors, i.e. $\mathcal{S}_\lambda = \mathcal{B}\left(\{(\beta, p) \in E^m : \ b(\beta) = \lambda\}\right)$.

The representation of the function $\mathcal{B}$ given in the next lemma provides insight into the structure of the set $\mathcal{S}_\lambda$.

Lemma 7.1. *For every $(\beta, p) \in E^m$, and for every $i \in]m-1]$, we have*

$$\mathcal{B}_i(\beta, p) = \sum_{\ell=2}^m \beta_\ell \left(\mathbb{E}\left[p_{K_{\ell,i},\ell}\right] - \frac{i}{m} \right),$$

where $K_{\ell,i} \sim Hyp(m, \ell, i)$.

Proof. By the definition of $\mathcal{B}$,

$$\mathcal{B}_i(\beta, p) = \sum_{\ell=2}^m \beta_\ell \sum_{j=0}^{\ell} \left(p_{j,\ell} - \frac{j}{\ell} \right) b_{j,\ell}(x).$$

Since, in addition

$$b_{j,\ell}(x) = \sum_{i=j}^{m-\ell+j} \frac{\binom{\ell}{j}\binom{m-\ell}{i-j}}{\binom{m}{i}} b_{i,m}(x), \tag{7.1}$$

applying Fubini's theorem yields

$$\mathcal{B}_i(\beta, p) = \sum_{\ell=2}^m \beta_\ell \sum_{j=0 \vee i-(m-\ell)}^{i \wedge \ell} \frac{\binom{\ell}{j}\binom{m-\ell}{i-j}}{\binom{m}{i}} \left(p_{j,\ell} - \frac{j}{\ell} \right).$$

The result follows by using classical properties of the hypergeometric distribution. □

Remark 7.1. The operation in (7.1) is often referred to as the degree elevation of the Bernstein basis [39].

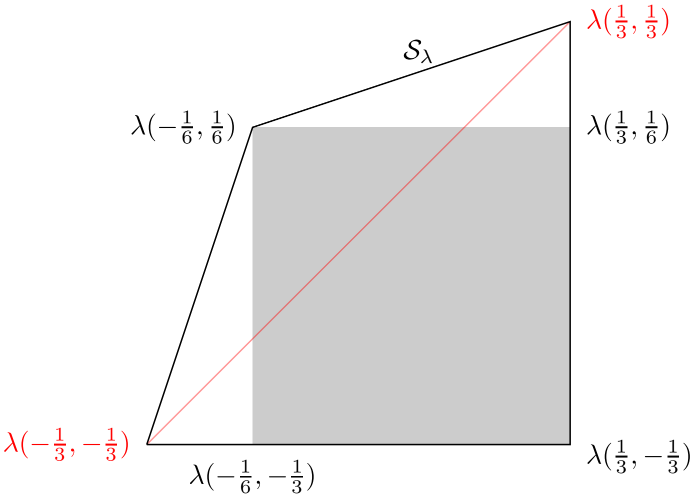

FIGURE 7. Points with red coordinates (resp. black coordinates) represent the extreme points of $\mathcal{S}_\lambda^2$ (resp. $\mathcal{S}_\lambda^3$), and correspond to deterministic 2-colouring rules (resp. deterministic 3-colouring rules). $\mathcal{S}_\lambda^2$ is the red line, $\mathcal{S}_\lambda^3$ is the grey square, and $\mathcal{S}_\lambda$ is their convex hull.

The previous result motivates the introduction of the function $\theta_\ell : \{0\} \times [0,1]^{\ell-1} \times \{1\} \to [0,1]^{m-1}$, $\ell \in]m]$, defined via

$$\theta_\ell(p) := \Big(\mathbb{E}\big[p_{K_{\ell,i}}\big]\Big)_{i=1}^{m-1}.$$

Set $\mathcal{P}_{0,1}^\ell := \{0\} \times \{0,1\}^{\ell-1} \times \{1\}$ the elements of which we refer to as deterministic ℓ-colouring rules (cf. Remark 2.3).

Proposition 7.2 (Characterisation of $\mathcal{S}_\lambda$)**.** *We have*

$$\mathcal{S}_\lambda \;=\; conv\left(\left\{\frac{\lambda}{\ell-1}\big(\theta_\ell(p) - u_m\big) : \ell \in]m], p \in \mathcal{P}_{0,1}^\ell\right\}\right),$$

where $u_m := (i/m)_{i=1}^{m-1}$. Further, define for every $\ell \in]m]$,

$$\mathcal{S}_\lambda^\ell \;:=\; \mathcal{B}\big((\beta, p) \in E^m : \; b(\beta) = \lambda, \;\forall i \in]m] \setminus \{\ell\}, \quad \beta_i = 0\big).$$

Then

$$\mathcal{S}_\lambda^\ell \;=\; conv\left(\left\{\frac{\lambda}{\ell-1}\big(\theta_\ell(p) - u_m\big) : p \in \mathcal{P}_{0,1}^\ell\right\}\right).$$

Proof. We use the notation of Proposition 3.1. Recall that $\mathcal{S}_\lambda$ is the image of $G_\lambda \times \mathcal{P}_m$ under the linear-affine map $\mathcal{B}$. As a consequence, $\mathcal{S}_\lambda$ is the convex hull of the image of the extreme points of $G_\lambda \times \mathcal{P}_m$. These extreme points are the points (β, p) such that

- there is $\ell \in]m]$ with $\beta_\ell = \lambda/(\ell-1)$, and for all $i \in]m] \setminus \{\ell\}, \beta_i = 0$.
- $p_{\cdot,k} \in \mathcal{P}_{0,1}^k$ for every $k \in]m]$.

By Proposition 7.1, for (β, p) of this form, $\mathcal{B}(\beta, p) = \lambda\,(\theta_\ell(p_{\cdot,\ell}) - u_m)\,/(\ell-1)$. This proves the first identity. The identity for $\mathcal{S}_\lambda^\ell$ is proved along the same lines. □

Lemma 7.3. *θ_ℓ is an injective map.*

Proof. Note that θ_ℓ can be extended to a linear map on $\{0\} \times \mathbb{R}^\ell$. Hence, it suffices to show that if $\theta_\ell(p) = 0$, then $p = 0$. Assume $p \in \{0\} \times \mathbb{R}^\ell$ is such that $\theta_\ell(p) = 0$. By assumption, $p_0 = 0$. Since $\mathbb{E}[p_{K_{\ell,1}}] = p_1 \ell!/m! = 0$, it follows that $p_1 = 0$. We proceed by induction. Assume $p_i = 0$ for all $i \leq k$ for some $k < \ell$. Since $K_{\ell,k+1}$ is supported on $[k+1]_0$ and by the induction hypothesis

$$0 = \mathbb{E}[p_{K_{\ell,k+1}}] = \sum_{j=0}^{k+1} \frac{\binom{\ell}{j}\binom{m-\ell}{k+1-j}}{\binom{m}{k+1}} p_j = \frac{\binom{\ell}{k+1}}{\binom{m}{k+1}} p_{k+1}.$$

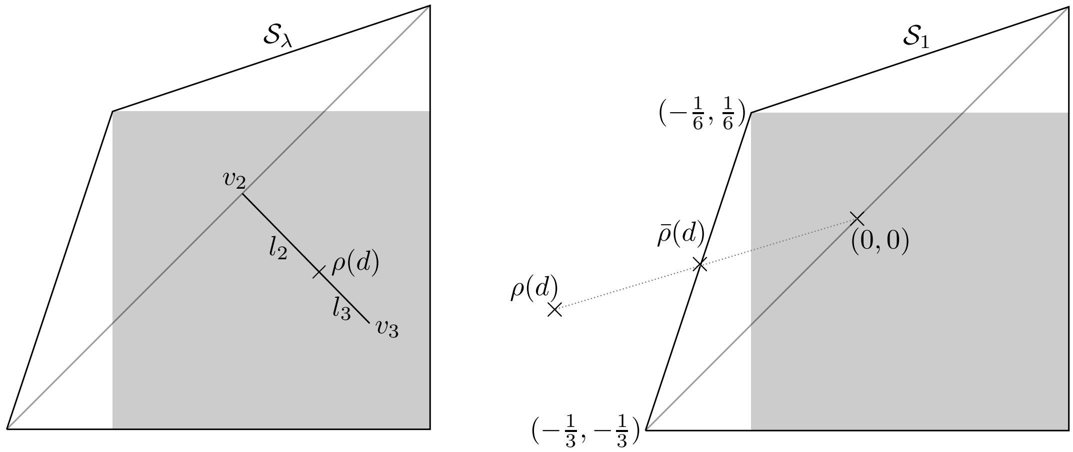

FIGURE 8. Left: Representation of some $(\lambda, \vec{v}, \alpha) \in \mathcal{C}_d$ for $m = 3$. Here, $l_2 = \|v_2 - \rho\|_2$ and $l_3 = \|v_3 - \rho\|_2$. $\alpha_2 = l_3/(l_2 + l_3)$ and $\alpha_3 = l_2/(l_2 + l_3)$. Right: L_d and $\mathcal{S}_1$ intersect on the west face of the polygon at the point $\bar\rho(d)$. Since 1-convex decompositions must involve a point in $\mathcal{S}_1^2$ (the diagonal segment inside $\mathcal{S}_1$), the unique 1-convex decomposition of $\bar\rho(d)$ is obtained by taking the 1-convex decomposition of $\rho(d)$ involving the two extreme points of the west face.

It follows that also $p_{k+1} = 0$. Altogether, $p = 0$. □

Definition 7.4 (λ-convex decompositions)**.** Let $\lambda > 0$. We call $\mathcal{C}^\lambda := \prod_{\ell=2}^m \mathcal{S}_\lambda^\ell \times \Delta_{m-2}$ the *set of λ-convex decompositions*. For $\rho \in \mathbb{R}^{m-1}$, we call $(\vec{v}, \alpha) \in \mathcal{C}^\lambda$ a *λ-convex decomposition* of ρ if $\rho = \sum_{\ell=2}^m \alpha_\ell v_\ell$. Define

$$\mathcal{C} := \left\{ (\lambda, \vec{v}, \alpha) \in \mathbb{R}_+ \times \prod_{\ell=2}^m \mathbb{R}^{m-1} \times \Delta_{m-1} : (\vec{v}, \alpha) \in \mathcal{C}^\lambda \right\}. \tag{7.2}$$

For a polynomial d with $\deg(d) = m$ and $d(0) = d(1) = 0$, define

$$\mathcal{C}_d = \{(\lambda, \vec{v}, \alpha) \in \mathcal{C} \ : \ (\vec{v}, \alpha) \text{ is a } \lambda\text{-convex combination of } \rho(d)\}.$$

Furthermore, define $\varphi : \mathcal{C} \to E^m$ as $\varphi(\lambda, \vec{v}, \alpha) = (\beta^{(\lambda,\vec{v},\alpha)}, p^{(\lambda,\vec{v},\alpha)})$, where

$$\beta_\ell^{(\lambda,\vec{v},\alpha)} := \frac{\lambda}{\ell - 1}\alpha_\ell \qquad \text{and} \qquad p_{\cdot,\ell}^{(\lambda,\vec{v},\alpha)} = \theta_\ell^{-1}\left(\frac{\ell - 1}{\lambda} v_\ell + u_m\right).$$

The next result states that any λ-convex decomposition of a BCV can be associated to a SD.

Proposition 7.5 (Embedding)**.** *φ is a bijection from $\mathcal{C}$ to E^m*

Proof. The injectivity of φ follows from the injectivity of θ_ℓ. For the surjectivity, consider $(\beta, p) \in E^m$. Set $\lambda := b(\beta)$, $\alpha_\ell := \beta_\ell(\ell - 1)/\lambda$, and $v_\ell := \lambda(\theta_\ell(p_{\cdot,\ell}) - u_m)/(\ell - 1)$. Clearly, $\lambda \in \mathbb{R}_+$ and $\alpha \in \Delta_{m-1}$. We claim that $\vec{v} := (v_\ell)_{\ell=2}^m \in \prod_{\ell=2}^m \mathcal{S}_\lambda^\ell$. To see this, first note that $p_{\cdot,\ell}$ can be written as a convex combination of elements in the extreme set $\mathcal{P}_{0,1}^\ell$. Since $\mathcal{B}$ is affine in the second argument (see proof Proposition 3.1), the claim follows by the characterisation of $\mathcal{S}_\lambda^\ell$ given in the second part of Proposition 7.2. At last, note that indeed $\varphi(\lambda, \vec{v}, \alpha) = (\beta, p)$. □

Corollary 7.6 (Geometric characterisation of $\mathcal{E}_d$)**.** *$\mathcal{E}_d$ coincides with $\varphi(\mathcal{C}_d)$.*

Proof. $(\beta, p) \in \mathcal{E}_d$ if and only if $\rho(d) = \mathcal{B}(\beta, p)$. Moreover, for $(\lambda, \vec{v}, \alpha) \in \mathcal{C}_d$,

$$\mathcal{B}(\beta^{(\lambda,\vec{v},\alpha)}, p^{(\lambda,\vec{v},\alpha)}) = \sum_{\ell=2}^m \alpha_\ell v_\ell = \rho(d),$$

where the last equality holds, since $(\vec{v}, \alpha)$ is a λ-convex decomposition of $\rho(d)$. Since φ is bijective, the result follows. □

Remark 7.2. The previous result states that any SD of d can be identified with (1) an effective branching rate λ and (2) a λ-convex decomposition of the BCV $\rho(d)$.

Remark 7.3. Let us explain the case $m = 3$ more explicitly. The objects of Definition 7.4 admit a clear graphical interpretation. Consider a SD $(\beta, p) \in \mathcal{E}_d$. We can read of $\varphi^{-1}(\beta, p) = (\lambda, \vec{v}, \alpha)$ directly from the Figs. 7 and 8. The first entry fixes the effective branching rate, the diagonal $\mathcal{S}^2_\lambda$ (grey), and the square (grey) $\mathcal{S}^3_\lambda$. For $\vec{v} = (v_2, v_3)$, the BCVs v_2 and v_3 with

$$v_\ell = \beta_\ell \left(\theta_\ell(p_{\cdot,\ell}) - u_3\right), \quad \ell = 2, 3,$$

correspond to points in $\mathcal{S}^2_\lambda$ and $\mathcal{S}^3_\lambda$, respectively. If we set $l_2 = \|v_2 - \rho(d)\|_2$ and $l_3 = \|v_3 - \rho(d)\|_2$, then $\alpha_2 = l_3/(l_2 + l_3)$ and $\alpha_3 = l_2/(l_2 + l_3)$.

Denote by O the origin in $\mathbb{R}^{m-1}$. Let $L_d := \{v \in \mathbb{R}^{m-1} : v = \lambda\rho(d) \text{ for some } \lambda > 0\}$, i.e. the half line passing through the origin and $\rho(d)$, with extremity O. Since $\mathcal{S}_1$ contains O, L_d intersects with $\mathcal{S}_1$ in a unique point, see also Fig. 8 (right) for an illustration if $m = 3$. Let $\bar{\rho}(d)$ be this point and let $\bar{d}$ be the polynomial with BCV $\bar{\rho}(d)$, i.e. $\bar{d}(x) = \left\langle B_m(x), (0, \bar{\rho}(d)^T, 0)^T \right\rangle$. Note that $\bar{\rho}(d)$ is on the boundary of $\mathcal{S}_1$, and hence, $b_\star(\bar{d}) = 1$.

Proposition 7.7 (Scaling)**.**

(i) Let $i \in]m-1]$ such that $\bar{\rho}_i(d) \neq 0$. Then

$$b_\star(d) = \frac{\rho_i(d)}{\bar{\rho}_i(d)}.$$

(ii) $(\beta, p) \in \mathcal{E}_{\bar{d}}$ if and only if $(b_\star(d)\beta, p) \in \mathcal{E}_d$.

Proof. This easily follows from the previous result and the scaling relation $\mathcal{S}_\lambda = \lambda\mathcal{S}_1$. □

In order to characterise the minimal SDs of d, by the previous result, it is enough to determine the minimal SDs of $\bar{d}$. This requires the knowledge of the intersection point $\bar{\rho}(d)$ of the line L_d and the polyhedron $\mathcal{S}_1$. Finding an intersection point of a line and a polyhedron is a classic problem in computational geometry, see e.g. [13, 25, 43]. We summarise this section with an algorithm that allows us to characterize the set of minimal SDs of $\mathcal{E}_d$.

Algorithm 7.8.

Step 1. Compute the intersection point $\bar{\rho}(d)$.
Step 2. Compute $b_\star(d)$ via Proposition 7.7-(i).
Step 3. Determine $\mathcal{C}^{\min}_{\bar{d}} := \{(\lambda, \vec{v}, \alpha) \in \mathcal{C}_{\bar{d}} : \lambda = 1\}$ and let $\mathcal{E}^{\min}_{\bar{d}} := \varphi(\mathcal{C}^{\min}_{\bar{d}})$.
Step 4. Finally $\mathcal{E}^{\min}_d := \{(b_\star(d)\beta, p) : (\beta, p) \in \varphi(\mathcal{S}^{\min}_{\bar{d}})\}$.

Then, $\mathcal{E}^{\min}_d$ is the set of minimal SDs.

7.2. **Minimal SDs if $m = 2, 3$.** For the sake of illustration, we examine the case $m = 3$ more closely. Here, d is of the form

$$d(x) = x(1 - x)(Ax + B).$$

In order to characterise $\mathcal{S}^2_1, \mathcal{S}^3_1$, and $\mathcal{S}_1$ in this low-dimensional case, set

$$v^{1,2} := \left(\tfrac{1}{3}, \tfrac{1}{3}\right), \quad v^{2,2} := \left(-\tfrac{1}{3}, -\tfrac{1}{3}\right), \quad \text{and} \quad v^{1,3} := \left(-\tfrac{1}{6}, \tfrac{1}{6}\right), \quad v^{2,3} := \left(\tfrac{1}{3}, -\tfrac{1}{3}\right), \quad v^{3,3} := \left(-\tfrac{1}{6}, -\tfrac{1}{3}\right), \quad v^{4,3} := \left(\tfrac{1}{3}, \tfrac{1}{6}\right).$$

Then,

$$\mathcal{S}^2_1 = \operatorname{conv}\left(v^{1,2}, v^{2,2}\right), \qquad \mathcal{S}^3_1 = \operatorname{conv}\left(v^{1,3}, v^{2,3}, v^{3,3}, v^{4,3}\right), \quad \text{and} \quad \mathcal{S}_1 = \operatorname{conv}\left(v^{1,2}, v^{2,2}, v^{1,3}, v^{2,3}\right).$$

See also Fig. 7. The faces of $\mathcal{S}_1$ are

$$\mathrm{F}_{\mathrm{UD}+} := \{\alpha v^{2,2} + (1-\alpha)v^{1,3} : \alpha \in [0,1]\}, \qquad \mathrm{F}_{\mathrm{OD}+} := \{\alpha v^{2,2} + (1-\alpha)v^{2,3} : \alpha \in [0,1]\},$$
$$\mathrm{F}_{\mathrm{UD}-} := \{\alpha v^{1,2} + (1-\alpha)v^{1,3} : \alpha \in [0,1]\}, \qquad \mathrm{F}_{\mathrm{OD}-} := \{\alpha v^{1,2} + (1-\alpha)v^{2,3} : \alpha \in [0,1]\}.$$

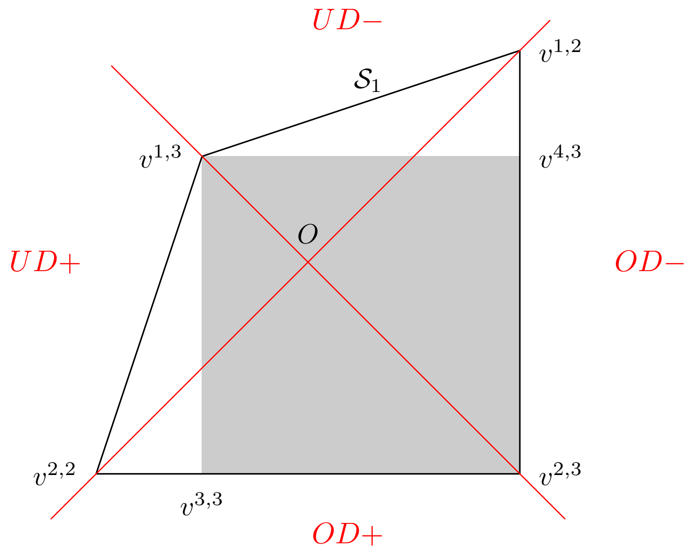

FIGURE 9. $\mathcal{D}_2$ and $\mathcal{D}_3$ delimits the plane into 4 regions. Those regions have a natural biological interpretation in terms of over- and underdominance.

It will be convenient to identify the following regions

$$\mathrm{UD}+ := \{(x,y) \in \mathbb{R}^2 : y > x \text{ and } y > -x\}, \qquad \mathrm{OD}+ := \{(x,y) \in \mathbb{R}^2 : y < x \text{ and } y < -x\},$$
$$\mathrm{UD}- := \{(x,y) \in \mathbb{R}^2 : y > x \text{ and } y < -x\}, \qquad \mathrm{OD}- := \{(x,y) \in \mathbb{R}^2 : y < x \text{ and } y > -x\}.$$

We refer to them as Underdominant+ (UD+), Underdominant− (UD−), Overdominant+ (OD+), and Overdominant− (OD−). The terminology will be justified at the end of the section. See also Fig. 9 for an illustration of these sets. A direct computation yields

$$\rho(d) = (a,b) \quad \text{with} \quad a = \frac{B}{3} \quad \text{and} \quad b = \frac{A+B}{3}.$$

The interpretation and analysis of the (minimal) SDs depends on the region of $\rho(d)$. We now distinguish the different cases.

7.2.1. *Region* UD+, UD−. Let us first assume that $(a,b) \in \mathrm{UD}+$. Then, $L_d \subset \mathrm{UD}+$ and L_d intersects $\mathrm{F}_{\mathrm{UD}+}$. In particular, $\bar{\rho}(d) \in \mathrm{F}_{\mathrm{UD}+}$. More specifically, a straightforward calculation yields

$$\bar{\rho}(d) = \frac{-2}{9a-3b}(a,b).$$

By Proposition 7.7 (alternatively Step 1 and Step 2 of Algorithm 7.8), the minimal effective branching rate is

$$b_\star(d) \;=\; \frac{3}{2}(b-3a).$$

Let us now proceed to Step 3. The only 1-convex decomposition $(\vec{v},\alpha)$ such that

$$\bar{\rho}(d) = \alpha_2 v_2 + (1-\alpha_2)v_3$$

for $v_\ell \in \mathcal{S}_1^\ell$, $\ell = 2,3$, is when v_2 and v_3 are extremal in $\mathrm{F}_{\mathrm{UD}+}$, i.e. $v_2 = v^{2,2}$ and $v_3 = v^{1,3}$. See again Fig. 8. Another computation yields

$$\alpha_2 = \frac{a+b}{3a-b}, \quad \alpha_3 = 1-\alpha_2.$$

By definition of $v^{2,2}$ and $v^{1,3}$, we have $v_2 = \frac{1}{2}(\theta_2(p_{\cdot,2}) - u_3)$ with $p_{\cdot,2} = (0,0,1)$ and $v_3 = \frac{1}{2}(\theta_3(p_{\cdot,3}) - u_3)$ with $p_{\cdot,3} = (0,0,0,1)$. In particular, $\mathcal{E}_{\vec{d}}^{\min} = \{(\beta^{(1,\vec{v},\alpha)}, p^{(1,\vec{v},\alpha)})\}$, where

$$\beta_\ell^{(1,\vec{v},\alpha)} = \frac{\alpha_\ell}{\ell-1} \quad \text{and} \quad p_{\cdot,\ell}^{(1,\vec{v},\alpha)} = p_{\cdot,\ell} \quad \text{for} \quad \ell \in]3].$$

FIGURE 10. Left: $\mathcal{S}_1^2, \mathcal{S}_1^3, \mathcal{S}_1^4$ when $m=4$. Right: $\mathcal{S}_1$ when $m=4$. The front triangular face (bottom, dark green) has two vertices corresponding to a deterministic colouring rule of order 4, i.e. in $\mathcal{P}_{0,1}^4$. As a consequence, if $\rho(d)$ belongs to the interior of this face, there is no minimal SD with deterministic colouring rules.

At last, by Algorithm 7.8, we have $\mathcal{E}_d^{\min} = \{(\beta, p)\}$ with

$$\beta_\ell = b_\star(d)\frac{\alpha_\ell}{\ell-1} \quad \text{and} \quad p_{\cdot,\ell} \quad \text{as above} \quad \text{for} \quad \ell \in]3].$$

If (a,b) belongs to region $\mathrm{F}_{\mathrm{UD}-}$, a symmetry argument exposes that in this case $b_\star(d)$ and α are obtained from the case of region $\mathrm{F}_{\mathrm{UD}+}$ by the transformation $(a,b) \mapsto (-b,-a)$ and by setting $p_{\cdot,2} = (0,1,0)$ and $p_{\cdot,3} = (0,0,1,1)$.

7.2.2. *Region* OD+, OD−. Assume that $(a,b) \in \mathrm{OD}+$. Then $L_d \subset \mathrm{OD}+$ and $\bar{\rho}(d) \in \mathrm{F}_{\mathrm{OD}+}$. More precisely,

$$\bar{\rho}(d) = -\frac{1}{3}\left(\frac{a}{b}, 1\right).$$

We use again Proposition 7.7 and obtain the minimal effective branching rate as

$$b_\star(d) \;=\; -\frac{b}{3}.$$

In contrast to $\rho(d) \in \mathrm{UD}$, here there are several 1-convex decompositions of $\bar{\rho}(d)$. The set of 1-convex decompositions of $\bar{\rho}(d)$ is the set of points of the form

$$v_2 = v^{2,2}, \quad v_3 \;=\; \left(x_0, -\frac{1}{3}\right) \quad \text{with} \quad x_0 \in \left[-\frac{1}{6} \vee -\frac{b}{3a}, \frac{1}{3}\right],$$

and

$$\alpha_2 = \frac{1}{a}\frac{b+3ax_0}{1+3x_0}.$$

For fixed x_0, $v_3 = \frac{1}{2}(\theta_3(p_{\cdot,3}) - u_3)$ with $p_{\cdot,3} = (0, 2x_0 + \frac{1}{3}, 0, 1)$. As a consequence, the corresponding SD is given by

$$\beta_\ell = b_\star(d)\frac{\alpha_\ell}{\ell-1}, \quad \text{for} \quad \ell \in]3], \quad \text{and} \quad p_2 = \Big(0,0,1\Big), \; p_3 = \Big(0, 2x_0 + \frac{1}{3}, 0, 1\Big).$$

For $(a,b) \in \mathrm{OD}-$, one easily recovers the minimal SD by symmetry.

Remark 7.4. In general there can be infinitely many minimal SDs of d (see regions OD+ and OD− above).

Remark 7.5. If $m=3$, there always is a minimal SD with a deterministic colouring rule. In regions UD+ and UD−, this is automatic since v_2 and v_3 correspond to deterministic colouring rules. In region OD+, this corresponds to choose one of the vertices of the cube as v_3, i.e. either $v_3 = v^{2,3}$ or $v_3 = v^{3,3}$ (if the latter is permitted, i.e. when $\bar{\rho}(d)$ lies in between $v^{2,2}$ and $v^{3,3}$). Then we have a deterministic colouring rule, i.e. the colour of the roots are a deterministic function of the ASG and the colouring of the leaves. It is tempting to conjecture that also in higher dimensions there is always a minimal SD with deterministic colouring rules. According to Step 3 of our algorithm, this equivalent to say that each face of $\mathcal{S}_1$ contains at most one vertex on each $\mathcal{S}_1^\ell$, $\ell \in]m]$ (i.e. each extreme point of a given face represents a deterministic ℓ-colouring rule, with a distinct ℓ). However, numerical simulations when $m = 4$ reveal that this is not the case, see Fig. 10.

7.2.3. *Classic examples for m=2,3.* Let us consider several examples from the literature that clarify the names of the various regions.

- The haploid Wright–Fisher diffusion with genic selection has the following drift term
$$d(x) = -\sigma x(1-x)$$
with $\sigma > 0$. Then $\rho(d) = -\sigma v^{2,2}$ belongs to the 1-dimensional subspace UD+ $\cap$ OD+. By the previous calculations, the minimal SD of $\bar{\rho}(d) = v^{2,2}$ is of the form $v_2 = v^{1,2}$ and $\alpha_2 = 1$. In particular, $b_\star(d) = \sigma$ and the minimal SD is $\beta = \sigma$ and $p_2 = (0, 0, 1)$. This is consistent with the classical ASG of Krone and Neuhauser [41][47].
- More generally, consider a diploid Wright–Fisher model with two alleles a and A and the following relative fitnesses
$$f_{aa} = -\sigma, \quad f_{aA} = -h\sigma, \quad f_{AA} = 0,$$
where $\sigma \in \mathbb{R}$ corresponds again to the selection strength and $h \in [0, 1]$ is the dominance coefficient, i.e. h quantifies the contribution of a to the fitness of an heterozygote. When $h = 1/2$, selection is said to be additive (and agrees with genic selection). When $h \in [0, 1/2)$, a is called recessive or underdominant. When $h \in (1/2, 1]$, a is said to be overdominant. The appropriate large population approximation in a weak-selection regime is the Wright–Fisher diffusion with drift term
$$d(x) = -\sigma x(1-x)\big(h - x(2h-1)\big). \tag{7.3}$$
Here, $\rho(d) = -\sigma(h, 1-h)/3$. A direct calculation shows that when $\sigma > 0$, $\rho(d)$ belongs to region UD+ if $h < 1/2$ (underdominance), and to region OD+ if $h > 1/2$ (overdominance). Note that the case $h = 1/2$ corresponds to the case of genic (additive) selection studied in the previous example. When $\sigma < 0$, $\rho(d)$ belongs to region UD− if $h < 1/2$ (underdominance) and it belongs to region OD− if $h > 1/2$ (overdominance). Again, $h = 1/2$ corresponds to the case of genic (additive) selection studied in the previous example. In conclusion, we find that the transition from underdominant to overdominant corresponds to a transition from one face of the polytope $\mathcal{S}_1$ to another face when considering the minimal SD of d. We also note that the previous analysis shows that in the underdominant regime, there is a unique minimal SD; whereas there are infinitely many possible choices in the overdominant case. It would be interesting to investigate if there is any biological meaning of this multiplicity.
- A model with balancing selection is a model in which the drift term is of the form
$$d(x) = x(1-x)(1-2x)$$
Here, $\rho(d) = v^{2,3}$. In this case, $v_3 = v^{2,3}$ and $\alpha_3 = 1$. This leads to $\beta = (0, 1)$ and $p_3 = (0, 1, 0, 1)$. This is consistent with the duality obtained in Neuhauser [46]. The particular form of the colouring rule p is also called the minority rule. We note that in [46], it is shown that the ASG generated from this SD has a natural genealogical interpretation in terms of the genealogy of a diploid population model.

7.3. **Graph-minimal selection decompositions.** Let us consider a thinning mechanism, i.e. a lower-triangular stochastic matrix $\mathbb{T} := \{\mathbb{T}_{k,i}\}_{k,i=1}^m$ (see Definition 2.24). Recall that a thinning mechanism acts on $\mathbb{R}_+^{m-1}$ as follows
$$(\beta)_{\ell=2}^m \mapsto (\mathbb{T}\beta)_{\ell=2}^m, \quad \text{with} \quad (\mathbb{T}\beta)_\ell = \sum_{k=\ell}^m \beta_k \mathbb{T}_{k,\ell}. \tag{7.4}$$

In what follows, we give a natural interpretation to this definition, by explaining how the thinning acts on a branching-coalescing system (see Section 4 for a definition of the branching-coalescing system).

Definition 7.9 (Thinning a branching-coalescing system)**.** Let $(\mathcal{G}_t; t \geq 0)$ be the branching-coalescing particle system with parameters (β, Λ), and let $\mathbb{T} := \{\mathbb{T}_{k,i}\}_{k,i=1}^m$ be a thinning mechanism. The branching-coalescing system $\mathcal{G}$ thinned by $\mathbb{T}$, denoted by $\mathbb{T}\mathcal{G} := (\mathbb{T}\mathcal{G}_t; t \geq 0)$, is defined dynamically according to the following random procedure. Independently at every k-branching in $\mathcal{G}$, with probability $\mathbb{T}_{k,i}$, remove $k - i$ particles chosen uniformly at random among the new ones (the marked particle giving rise to the branching event is never removed), and keep the remaining particles. We say that the thinning procedure is non-trivial if $\mathbb{T}$ is distinct from the identity matrix.

Figure 11 illustrates such a branching-coalescing system and a thinned version of it. It follows from the consistency of the rates of the Λ-coalescent that $\mathbb{T}\mathcal{G}$ is again distributed as a branching-coalescing system with (unchanged) coalescence mechanism Λ and branching mechanism $(\mathbb{T}\beta_\ell)_{\ell=2}^m$ (as in Eq. (7.4))
The next result is a straightforward consequence of this construction.

Proposition 7.10. *Let $(\mathcal{G}_t; t \geq 0)$ and $(\mathcal{G}_t^\star; t \geq 0)$ be the branching-coalescing particle systems with parameters (β, Λ) and $(\mathbb{T}\beta, \Lambda)$, respectively. If both branching-coalescing systems start with the same number of particles, then*

$$\forall t \geq 0, \quad \mathcal{G}_t^\star \overset{(d)}{=} \mathbb{T}\mathcal{G}_t \subseteq \mathcal{G}_t.$$

The notion of graph-minimality given in Definition (2.25) can be expressed at the level of the branching-coalescing systems as follows.

Definition 7.11. We say that $(\beta, p) \in \mathcal{E}_d$ is graph-minimal if and only if there is no (non-trivial) thinning mechanism $\mathbb{T}$ and $(\beta', p') \in \mathcal{E}_d$ such that $\mathcal{G}' \overset{(d)}{=} \mathbb{T}\mathcal{G}$, where $\mathcal{G}'$ (resp. $\mathcal{G}$) is the branching-coalescing system generated from the pair (β', Λ) (resp. (β, Λ)) starting with n particles at time $t = 0$.

In other words, the SD $(\beta, p) \in \mathcal{E}_d$ is minimal if and only if there are no superfluous (dummy) branches in the ASG.
In the remainder of this section, we prove Theorem 2.26, which states that minimal SDs are graph-minimal. For this, we introduce a partial ordering $\preceq$ on $\mathbb{R}_+^{m-1}$ as follows. For any $\beta, \beta' \in \mathbb{R}_+^{m-1}$, write $\beta' \preceq \beta$ if and only if for all $k \in]m]$

$$\sum_{j=k}^m \beta_j' \;\leq\; \sum_{j=k}^m \beta_j. \tag{7.5}$$

Moreover, we write $\beta' \prec \beta$ if and only if $\beta' \preceq \beta$ and $\beta' \neq \beta$. The relation between this partial ordering and the thinning is the content of the following proposition.

Proposition 7.12. *Let $\beta', \beta \in \mathbb{R}_+^{m-1}$, Λ be a finite measure, and $n \in \mathbb{N}$. Let $\mathcal{G}'$ and $\mathcal{G}$ be the branching-coalescing particle systems constructed from the pairs (β', Λ) and (β, Λ), both system starting with n particles. Then $\beta' \prec \beta$ if and only if there exists a non-trivial thinning mechanism $\mathbb{T}$ such that $\mathcal{G}' \overset{(d)}{=} \mathbb{T}\mathcal{G}$.*

Proof. First, it is immediate to check from (7.4) that $\mathbb{T}\beta \prec \beta$ if $\mathbb{T}$ is non trivial. Let us now show the converse, i.e. assuming that $\beta' \prec \beta$, we construct a non-trivial thinning of $\mathcal{G}$ distributed as $\mathcal{G}'$. First

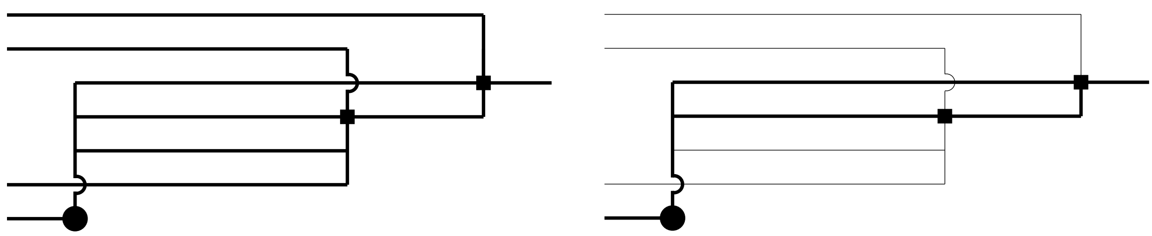

FIGURE 11. A branching-coalescing system (left) and its thinned version (right).

define the thinning mechanism $\mathbb{T}^{(m)}$ via

$$\begin{aligned}
\mathbb{T}^{(m)}_{m,m} &:= \frac{\beta'_m}{\beta_m},\\
\mathbb{T}^{(m)}_{m,m-1} &:= \left(1-\frac{\beta'_m}{\beta_m}\right)\left(\frac{\beta'_{m-1}}{\beta_m-\beta'_m}\wedge 1\right)\\
\mathbb{T}^{(m)}_{m,m-2} &:= \left(1-\frac{\beta'_m}{\beta_m}\right)\left(1-\frac{\beta'_{m-1}}{\beta_m-\beta'_m}\right)^+\\
\mathbb{T}^{(m)}_{m,m-k} &:= 0 \quad \text{for } k>2,
\end{aligned}$$

and $\mathbb{T}^{(m)}_{k,i} := \mathbf{1}_{\{i=k\}}$ for $k<m$ so that it is only thinned at m-branching events. Set $\beta^{(m)} := \mathbb{T}^{(m)}\beta$. Clearly, $\beta^{(m)}_m = \beta'_m$. Furthermore, from (7.5), it is direct to check that $\mathbb{T}^{(m)}$ has been chosen in such a way that for all $k \in]m-1]$

$$\sum_{j=k}^{m-1}\beta'_j \le \sum_{j=k}^{m-1}\beta^{(m)}_j.$$

Next, thin the $(m-1)$-branching events in the (thinned) system $\mathcal{G}^{(m)} := \mathbb{T}^{(m)}\mathcal{G}$ via the thinning $\mathbb{T}^{(m-1)}$ given by

$$\begin{aligned}
\mathbb{T}^{(m-1)}_{m-1,m-1} &:= \frac{\beta'_{m-1}}{\beta^{(m)}_{m-1}},\\
\mathbb{T}^{(m-1)}_{m-1,m-2} &:= \left(1-\frac{\beta'_{m-1}}{\beta^{(m)}_{m-1}}\right)\left(\frac{\beta'_{m-2}}{\beta^{(m)}_{m-1}-\beta'_{m-1}}\wedge 1\right)\\
\mathbb{T}^{(m-1)}_{m-1,m-3} &:= \left(1-\frac{\beta'_{m-1}}{\beta^{(m)}_{m-1}}\right)\left(1-\frac{\beta'_{m-2}}{\beta^{(m)}_{m-1}-\beta'_{m-1}}\right)^+\\
\mathbb{T}^{(m-1)}_{m-1,m-k} &:= 0 \quad \text{for } k>3,
\end{aligned}$$

and $\mathbb{T}^{(m-1)}_{k,i} := \mathbf{1}_{\{i=k\}}$ for $k \neq m-1$. The thinned system $\mathcal{G}^{(m-1)} := \mathbb{T}^{(m-1)}\mathcal{G}^{(m)}$ has branching rates $\beta^{(m-1)} := \mathbb{T}^{(m-1)}\beta^{(m)}$ such that $\beta^{(m-1)}_m = \beta'_m$, $\beta^{(m-1)}_{m-1} = \beta'_{m-1}$ and for all $k \in]m-2]$

$$\sum_{j=k}^{m-2}\beta'_j \le \sum_{j=k}^{m-2}\beta^{(m-1)}_j.$$

Iterating this procedure, we construct successive thinning mechanisms $\mathbb{T}^{(k)}$ until reaching a branching-coalescing system $\mathcal{G}^{(2)}$ with rates $\beta^{(2)} = \beta'$, i.e. so that $\mathcal{G}^{(2)}$ is distributed as $\mathcal{G}'$. Further, by (7.4), $\mathcal{G}^{(2)}$ is identical in law to $\mathbb{T}\mathcal{G}$, where $\mathbb{T} := \mathbb{T}^{(2)}\cdots\mathbb{T}^{(m)}$. This completes the proof of the proposition. □

Proof of Theorem 2.26. By Proposition 7.12, it is sufficient to show that $\beta' \prec \beta$ implies that $b(\beta') < b(\beta)$. This is easily seen by summing the inequalities in (7.5). □

7.4. **Equivalence of minimality if m=3.** Theorem 2.26 states that every minimal SD is graph-minimal. In dimension $m=2$, i.e. when $\deg(d)=2$, both notions of minimality clearly agree. In higher dimensions, the question is more involved. In this section, we prove the equivalence of the two notions in dimension $m=3$.

The next result is valid in any dimension and describes an operation that improves the effective branching rate and at the same time leads to a thinner ancestral structure.

Proposition 7.13 (Shrinking the polygon)**.** *Consider* $(\lambda,\vec{v},\alpha)\in\mathcal{C}_d$. *Set*

$$\lambda^\star(\vec{v}) := \inf\{\gamma\ge 0 : v_\ell \in \mathcal{S}^\ell_\gamma,\ \forall \ell\in]m]\}.$$

Then $(\lambda^\star(\vec{v}),\vec{v},\alpha)\in\mathcal{C}_d$ *and* $\beta^{(\lambda^\star(\vec{v}),\vec{v},\alpha)} \preceq \beta^{(\lambda,\vec{v},\alpha)}$.

Proof. Assume $\lambda\neq\lambda^\star(\vec{v})$. Since, $v_\ell\in\mathcal{S}^\ell_{\lambda^\star(\vec{v})}$ and $\rho=\sum_{\ell=2}^m \alpha_\ell v_\ell$ the first statement follows. The second statement follows by the definition of $\beta^{(\lambda,\vec{v},\alpha)}$. □

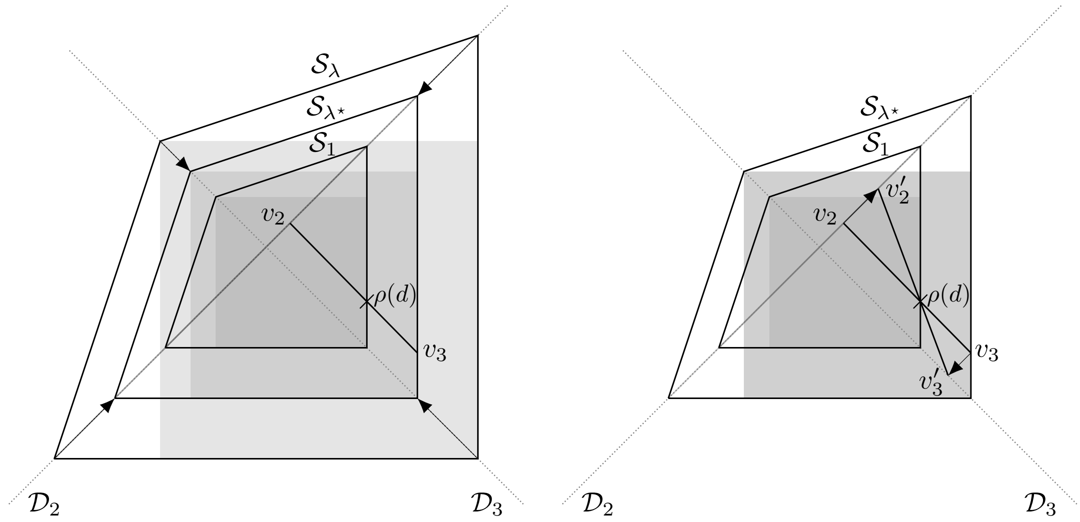

FIGURE 12. Left: Shrinking the Polygon $\mathcal{S}_\lambda$ to $\mathcal{S}_{\lambda^\star}$ while keeping $\vec{v}$ constant. Right: Moving $\vec{v}$ into $\operatorname{int}(\mathcal{S}_{\lambda^\star})$ along the direction of $\mathcal{D}_2$. By the intercept theorem, this keeps the relative distances to $\rho(d)$ constant.

Remark 7.6. The idea of this shrinking procedure is to consider v_ℓ not as a point in $\mathcal{S}_\lambda$, but in (the smaller) $\mathcal{S}_{\lambda^\star(\vec{v})}$, see also Fig. 12 (left). If $\vec{v}$ is expressed in terms of the extremal points of the smaller polygon, the total effective branching rate decreases, but the relative branching rates remain constant.

In the remainder of the section, assume that $\deg(d) = 3$. We continue to use the notation of Section 7.2. In particular, $v^{1,2}, v^{2,2} \in \mathcal{S}_1^2$ and $v^{1,3}, v^{2,3} \in \mathcal{S}_1^3$ are the extremal points of $\mathcal{S}_1$. Denote by

$$\mathcal{D}_2 := \{av^{1,2} : a \in \mathbb{R}\} \quad \text{and} \quad \mathcal{D}_3 := \{av^{1,3} : a \in \mathbb{R}\}.$$

Proposition 7.14 (Shifting $\vec{v}$)**.** *Consider $(\lambda, \vec{v}, \alpha) \in \mathcal{C}_d$. Assume that $b_\star(d) = 1$ and that $\lambda^\star := \lambda^\star(\vec{v}) > 1$. If $v_3 \notin \{\lambda^\star v^{1,3}, \lambda^\star v^{2,3}\}$, then there exists $\lambda' < \lambda^\star$ and $\vec{v}'$ such that*

- $(\lambda', \vec{v}', \alpha) \in \mathcal{C}_d$,
- $\beta^{(\lambda', \vec{v}', \alpha)} \preceq \beta^{(\lambda^\star, \vec{v}, \alpha)}$,
- *either $\vec{v}' \in \{\lambda' v^{1,3}, \lambda' v^{2,3}\}$ or $v_i' \in F_\rho$ for $i \in]3]$,*

where F_ρ is the face of $\mathcal{S}_1$ containing $\rho(d)$.

Remark 7.7. $v_3 \notin \{\lambda^\star v^{1,3}, \lambda^\star v^{2,3}\}$ means that $v_3 \in \mathcal{S}^3_{\lambda^\star}$ is not a corner point $\mathcal{S}_{\lambda^\star}$.

Proof of Proposition 7.14. Since $\lambda^\star > 1$, it follows that $\alpha_2 \in (0,1)$. For $a \in \mathbb{R}$, define $\vec{v}(a) := (v_2(a), v_3(a))$ via

$$v_3(a) := v_3 + av^{1,2} \quad \text{and} \quad v_2(a) := v_2 - \frac{\alpha_3}{\alpha_2} av^{1,2}.$$

Note that $\alpha_2 v_2(a) + \alpha_3 v_3(a) = \rho(d)$. Hence, $(\lambda^\star, \vec{v}(a), \alpha) \in \mathcal{C}_d$ if and only if $\vec{v}(a) \in \mathcal{S}^2_{\lambda^\star} \times \mathcal{S}^3_{\lambda^\star}$. Since $\mathcal{D}_2 \perp \mathcal{D}_3$ and $v_3 \notin \{\lambda^\star v^{1,3}, \lambda^\star v^{2,3}\}$, there is a unique $a_0 \in \mathbb{R} \setminus \{0\}$ such that $v_3(a_0) \in \mathcal{D}_3 \cap \operatorname{int}(\mathcal{S}^3_{\lambda^\star})$. We split now the analysis in two cases: (i) $v_3(a_0) \notin \mathcal{S}_1^3$ and (ii) $v_3(a_0) \in \mathcal{S}_1^3$. In case (i), since any line from $\mathcal{D}_3 \cap (\mathcal{S}_1^3)^C$ to $(\mathcal{S}_1^2)^C$ lies outside of $\mathcal{S}_1$ and $\rho(d) \in \mathcal{S}_1$ is a convex combination of $v_2(a_0)$ and $v_3(a_0)$, we infer that $v_2(a_0) \in \mathcal{S}_1^2$. In particular, $\vec{v}(a_0) \in \mathcal{S}^2_{\lambda^\star} \times \mathcal{S}^3_{\lambda^\star}$. Hence, $(\lambda^\star, \vec{v}(a_0), \alpha) \in \mathcal{C}_d$ and $\beta^{(\lambda^\star, \vec{v}(a_0), \alpha)} = \beta^{(\lambda^\star, \vec{v}, \alpha)}$. Moreover, $v_i(a_0) \in \operatorname{int}(\mathcal{S}^i_{\lambda^\star})$ for $i \in]3]$. Therefore, $\lambda' := \lambda^\star(\vec{v}(a_0)) < \lambda^\star$, and then, setting $\vec{v}' := \vec{v}(a_0)$, the result follows from Proposition 7.13. In case (ii), there is a unique $a_1 \in \mathbb{R}$ such that $v_3(a_1) \in F_\rho$. Since F_ρ is the face that contains $\rho(d)$, we conclude that also $v_2(a_1) \in F_\rho$. As before, we obtain that $(\lambda^\star, \vec{v}(a_1), \alpha) \in \mathcal{C}_d$ and $\beta^{(\lambda^\star, \vec{v}(a_1), \alpha)} = \beta^{(\lambda^\star, \vec{v}, \alpha)}$. The result follows from Proposition 7.13 by setting $\lambda' := \lambda^\star(\vec{v}(a_1)) < \lambda^\star$ and $\vec{v}' := \vec{v}(a_1)$. □

Figure 12 (right) illustrates the idea behind the proof of Proposition 7.14.

Proposition 7.15. *Assume $b_\star(d) = 1$. Let $(\beta, p) \in \mathcal{E}_d$ be a SD with $b(\beta) = \lambda > 1$. Then there exists $(\beta', p') \in \mathcal{E}_d$ such that $\beta' \preceq \beta$ and $b(\beta') = 1$.*

Proof. Lemma 7.5 allows us to identify (β, p) with $(\lambda, \vec{v}, \alpha) \in \mathcal{C}_d$. We make the following case distinctions: (1) $v_\ell \in \operatorname{int}(\mathcal{S}_\lambda^\ell)$ for all $\ell \in]3]$, (2) there is ℓ such that $v_\ell \in \partial \mathcal{S}_\lambda^\ell$, but $v_3 \notin \{\lambda v^{1,3}, \lambda v^{2,3}\}$, and (3) $v_3 \in \{\lambda v^{1,3}, \lambda v^{2,3}\}$. In case (1), apply the shrinking operation of Proposition 7.13. This leads to a new triple $(\lambda^\star, \vec{v}, \alpha)$ with $\beta^{(\lambda^\star, \vec{v}, \alpha)} \preceq \beta^{(\lambda, \vec{v}, \alpha)}$. In particular, $(\lambda^\star, \vec{v}, \alpha)$ falls now into case (2) or (3). In case (2), apply the shift-operation of Proposition 7.14. This leads to a new triple $(\lambda', \vec{v}', \alpha)$ with $\beta^{(\lambda', \vec{v}', \alpha)} \preceq \beta^{(\lambda^\star, \vec{v}, \alpha)}$. Moreover, either $\vec{v}' \in \{\lambda' v^{1,3}, \lambda' v^{2,3}\}$ or $v_i' \in F_\rho$ for $i = 2, 3$. The first case falls into case (3), and in the second case, the result directly follows by setting $(\beta', p') = \varphi(\lambda', \vec{v}', \alpha)$. Hence, the proof reduces to prove (3). Assume $v_3 \in \{\lambda v^{1,3}, \lambda v^{2,3}\}$. In particular, $u_3 := v_3/\lambda$ has to be an extremal point of the face F_ρ and $v_2 \in \mathcal{S}_1^2$ (otherwise the connection v_3 to $\mathcal{D}_2$ does not intersect F_ρ). Similarly, if $u_2 \in \mathcal{S}_1^2$ is the other extremal point of F_ρ, then $v_2 = \gamma u_2$ for some $\gamma \in [0, 1]$. Therefore, $\rho(d) = \alpha_2 \gamma u_2 + \alpha_3 \lambda u_3$. Since u_2 and u_3 are the extremal points of F_ρ, there is $\hat{\alpha} \in \Delta_1$ such that $\rho(d) = \hat{\alpha}_2 u_2 + \hat{\alpha}_3 u_3$, i.e. $(1, (u_2, u_3), \hat{\alpha}) \in \mathcal{C}_d$. Since u_2 and u_3 are linearly independent, we conclude that $\alpha_2 \gamma = \hat{\alpha}_2$ and $\alpha_3 \lambda = \hat{\alpha}_3$. It follows that $\beta^{(1, (u_2, u_3), \hat{\alpha})} \preceq \beta^{(\lambda, \vec{v}, \alpha)}$, which proves the result.

□

Proof of Proposition 2.27. The proof follows by Proposition 7.15 together with the scaling property, which is described in Proposition 7.7. □

There is a third notion of minimality we have not considered so far, namely, the one induced by the component-wise ordering, i.e. $\beta \preceq_{\mathrm{cw}} \beta'$ if and only if $\beta_\ell \leq \beta_\ell'$ for all $\ell \in]m]$. Clearly,

$$\beta \preceq_{\mathrm{cw}} \beta' \Rightarrow \beta \preceq \beta' \Rightarrow b(\beta) \leq b(\beta').$$

The three notions are equivalent for $m = 2$. For $m \geq 3$, this is not anymore the case. However, by inspection of the proofs of Propositions 7.13, 7.14 and 7.15, we see that for $m = 3$, we have proved the following stronger version of Proposition 2.27.

Proposition 7.16. *Assume that* $\deg(d) = 3$. *For any* $(\beta, p) \in \mathcal{E}_d$ *with* $b(\beta) > b_\star(d)$, *there is* $(\beta', p') \in \mathcal{E}_d$ *such that* $b(\beta') = b_\star(d)$ *and* $\beta' \preceq_{cw} \beta$.

Acknowledgements

We are grateful to A. González Casanova for many interesting discussions. F. Cordero and S. Hummel received financial support from Deutsche Forschungsgemeinschaft (CRC 1283 "Taming Uncertainty", Project C1).

References

[1] G. Achaz, A. Lambert, and E. Schertzer. The sequential loss of allelic diversity. *Adv. Appl. Probab.*, 50:13–29, 2018.

[2] F. J. Ayala and C. A. Campbell. Frequency-dependent selection. *Ann. Rev. Ecol. Syst.*, 5:115–138, 1974.

[3] E. Baake, U. Lenz, and A. Wakolbinger. The common ancestor type distribution of a Λ-Wright–Fisher process with selection and mutation. *Electron. Commun. Probab.*, 21:no. 59, 16 pp., 2016.

[4] E. Baake, F. Cordero, and S. Hummel. Lines of descent in the deterministic mutation-selection model with pairwise interaction. *arXiv e-prints*, 2018.

[5] B. Bah and E. Pardoux. The Λ-lookdown model with selection. *Stochastic Process. Appl.*, 125: 1089–1126, 2015.

[6] A. D. Barbour, S. N. Ethier, and R. C. Griffiths. A transition function expansion for a diffusion model with selection. *Ann. Appl. Probab.*, 10:123–162, 2000.

[7] M. Barczy, Z. Li, and G. Pap. Yamada-Watanabe results for stochastic differential equations with jumps. *Int. J. Stoch. Anal.*, pages Art. ID 460472, 23, 2015.

[8] N. Berestycki. *Recent progress in coalescent theory*, volume 16 of *Ensaios Matemáticos*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009.

[9] J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Related Fields*, 126:261–288, 2003.

[10] M. Birkner and J. Blath. Measure-valued diffusions, coalescents and genetic inference. In J. Blath, P. Mörters, and M. Scheutzow, editors, *Trends in Stochastic Analysis*, pages 329–364, Cambridge, 2009. Cambridge University Press.

[11] E. Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.*, 197:247–276, 1998.

[12] F. Cordero and M. Möhle. On the stationary distribution of the block counting process for population models with mutation and selection. *J. Math. Anal. Appl.*, 474:1049–1081, 2019.

[13] M. Cyrus and J. Beck. Generalized two-and three-dimensional clipping. *Computers & Graphics*, 3: 23–28, 1978.

[14] D. A. Dawson and A. Greven. Duality for spatially interacting Fleming-Viot processes with mutation and selection. *ArXiv e-prints*, 2011.

[15] D. A. Dawson and A. Greven. *Spatial Fleming-Viot Models with Selection and Mutation.* Springer, Cham, 2014.

[16] P. Donnelly and T. G. Kurtz. Particle representations for measure-valued population models. *Ann. Probab.*, 27:166–205, 1999.

[17] B. Eldon and J. Wakeley. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, 172:2621–2633, 2006.

[18] A. Etheridge. *Some Mathematical Models from Population Genetics.* Springer, Heidelberg, 2011.

[19] A. Etheridge, N. Freeman, and S. Penington. Branching brownian motion, mean curvature flow and the motion of hybrid zones. *Electron. J. Probab.*, 22:no. 103, 40 pp., 2017.

[20] A. Etheridge, N. Freeman, S. Penington, and D. Straulino. Branching Brownian motion and selection in the spatial Λ-Fleming-Viot process. *Ann. Appl. Probab.*, 27(5):2605–2645, 2017.

[21] A. M. Etheridge and R. C. Griffiths. A coalescent dual process in a Moran model with genic selection. *Theor. Popul. Biol.*, 75:320 – 330, 2009.

[22] A. M. Etheridge, R. C. Griffiths, and J. E. Taylor. A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theor. Popul. Biol.*, 78:77–92, 2010.

[23] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence.* Wiley, New York, 1986.

[24] P. Fearnhead. The common ancestor at a nonneutral locus. *J. Appl. Probab.*, 39:38–54, 2002.

[25] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes. *Computer Graphics: Principles and Practice.* Addison-Wesley, Reading, MA, second edition, 1996.

[26] C. Foucart. The impact of selection in the Λ-Wright–Fisher model. *Electron. Commun. Probab.*, 18: no. 72, 10 pp., 2013.

[27] C. Foucart. Erratum: The impact of selection in the Λ-Wright–Fisher model. *Electron. Commun. Probab.*, 19:no. 15, 3 pp., 2014.

[28] Z. Fu and Z. Li. Stochastic equations of non-negative processes with jumps. *Stochastic Process. Appl.*, 120:306–330, 2010.

[29] F. Gaiser and M. Möhle. On the block counting process and the fixation line of exchangeable coalescents. *ALEA Lat. Am. J. Probab. Math. Stat.*, 13:809–833, 2016.

[30] J. Gallier. *Curves and Surfaces in Geometric Modeling: Theory and Algorithms.* Morgan Kaufmann, San Francisco, CA, 1999.

[31] A. González Casanova and C. Smadi. Multidimensional Λ-Wright–Fisher processes with general frequency-dependent selection. *in preparation*, 2019.

[32] A. González Casanova and D. Spanò. Duality and fixation in Ξ-Wright–Fisher processes with frequency-dependent selection. *Ann. Appl. Probab.*, 28:250–284, 2018.

[33] R. C. Griffiths. The Λ-Fleming–Viot process and a connection with Wright–Fisher diffusion. *Adv. Appl. Probab.*, 46:1009–1035, 2014.

[34] R. C. Griffiths, P. A. Jenkins, and S. Lessard. A coalescent dual process for a Wright–Fisher diffusion with recombination and its application to haplotype partitioning. *Theor. Popul. Biol.*, 112:126 – 138, 2016.

[35] O. Hénard. The fixation line in the Λ-coalescent. *Ann. Appl. Probab.*, 25:3007–3032, 2015.

[36] P. Herriger and M. Möhle. Conditions for exchangeable coalescents to come down from infinity. *ALEA Lat. Am. J. Probab. Math. Stat*, 9:637–665, 2012.
[37] P. Imkeller and N. Willrich. Solutions of martingale problems for Lévy-type operators with discontinuous coefficients and related SDEs. *Stochastic Process. Appl.*, 126:703–734, 2016.
[38] S. Jansen and N. Kurt. On the notion(s) of duality for Markov processes. *Probab. Surveys*, 11: 59–120, 2014.
[39] K. I. Joy. Bernstein polynomials. *On-Line Geometric Modeling Notes*, 13, 2000.
[40] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13:235–248, 1982.
[41] S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theor. Popul. Biol.*, 51:210–237, 1997.
[42] Z. Li and F. Pu. Strong solutions of jump-type stochastic equations. *Electron. Commun. Probab.*, 17:no. 33, 13 pp., 2012.
[43] Y.-D. Liang and B. A. Barsky. A new concept and method for line clipping. *ACM Trans. Graph.*, 3:1–22, 1984.
[44] M. Möhle. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.*, 32:983–993, 2000.
[45] M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.*, 29:1547–1562, 2001.
[46] C. Neuhauser. The ancestral graph and gene genealogy under frequency-dependent selection. *Theor. Popul. Biol.*, 56:203–214, 1999.
[47] C. Neuhauser and S. M. Krone. The genealogy of samples in models with selection. *Genetics*, 145: 519–534, 1997.
[48] J. R. Norris. *Markov Chains, 2nd ed.* Cambridge Univ. Press, Cambridge, 1998.
[49] P. Pfaffelhuber and A. Wakolbinger. The process of most recent common ancestors in an evolving coalescent. *Stochastic Process. Appl.*, 116:1836–1859, 2006.
[50] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27:1870–1902, 1999.
[51] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36:1116–1125, 1999.
[52] T. Shiga. Diffusion processes in population genetics. *J. Math. Kyoto Univ.*, 21:133–151, 1981.
[53] J. E. Taylor. The common ancestor process for a Wright-Fisher diffusion. *Electron. J. Probab.*, 12: no. 28, 808–847, 2007.
[54] J. A. van Casteren. On martingales and Feller semigroups. *Results in Mathematics*, 21:274–288, 1992.

F. Cordero, S. Hummel: Faculty of Technology, Bielefeld University, Box 100131, 33501 Bielefeld, Germany
Email: fcordero@techfak.uni-bielefeld.de, shummel@techfak.uni-bielefeld.de

E. Schertzer: Laboratoire de Probabilités, Statistiques et Modélisation (LPSM), Sorbonne Université, CNRS UMR 8001, Paris, France
Center for Interdisciplinary Research in Biology (CIRB), Collège de France, CNRS UMR 7241, PSL Research University, Paris, France
Email: emmanuel.schertzer@upmc.fr