Theoretical Population Biology 156 (2024) 40–45

Contents lists available at ScienceDirect

Theoretical Population Biology

journal homepage: www.elsevier.com/locate/tpb

Bernoulli factories and duality in Wright–Fisher and Allen–Cahn models of population genetics

Jere Koskela [a,b,*], Krzysztof Łatuszyński [b], Dario Spanò [b]

[a] School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK
[b] Department of Statistics, University of Warwick, Coventry CV4 7AL, UK

ARTICLE INFO

MSC:
35C99
60J70
60J90
92D10

Keywords:
Allen–Cahn equation
Bernoulli factory
Duality
Frequency-dependent selection
Wright–Fisher diffusion

ABSTRACT

Mathematical models of genetic evolution often come in pairs, connected by a so-called duality relation. The most seminal example are the Wright–Fisher diffusion and the Kingman coalescent, where the former describes the stochastic evolution of neutral allele frequencies in a large population forwards in time, and the latter describes the genetic ancestry of randomly sampled individuals from the population backwards in time. As well as providing a richer description than either model in isolation, duality often yields equations satisfied by quantities of interest. We employ the so-called Bernoulli factory – a celebrated tool in simulation-based computing – to derive duality relations for broad classes of genetics models. As concrete examples, we present Wright–Fisher diffusions with general drift functions, and Allen–Cahn equations with general, nonlinear forcing terms. The drift and forcing functions can be interpreted as the action of frequency-dependent selection. To our knowledge, this work is the first time a connection has been drawn between Bernoulli factories and duality in models of population genetics.

1. Introduction

The Bernoulli factory problem is to construct a realisation of a Bernoulli($f(p)$) random variable (or an $f(p)$-coin) using an almost surely finite number of independent p-coins, where $f : [0,1] \mapsto [0,1]$ is a known function but $p \in [0,1]$ is unknown. The special case $f(p) = 1/2$ was formulated and solved by John von Neumann in 1951 (von Neumann, 1951). Later, Keane and O'Brien provided a necessary and sufficient condition for a given function f to have a Bernoulli factory (Keane and O'Brien, 1994). In brief, f has a Bernoulli factory if and only if it is identically equal to zero or one, or it is continuous and *polynomially bounded*, i.e.

$$\min\{f(p), 1-f(p)\} \geq \min\{p, 1-p\}^n \tag{1}$$

for all $p \in [0,1]$ and some $n \in \mathbb{N}$. However, the proof of Keane and O'Brien is only partly constructive: it relies on a recursively defined sequence whose explicit solution is intractable. Constructions of algorithms have relied on approximation of f by Bernstein polynomials, which are naturally associated with p-coins (Holtz et al., 2011; Łatuszyński et al., 2011; Nacu and Peres, 2005), or on other series expansions of f with non-negative coefficients (Mendo, 2019).

Many seminal models of population genetics rely on the random propagation of alleles from one generation to the next. A prototypical example is the Wright–Fisher model, in which a population of fixed size $N \in \mathbb{N}$ evolves in discrete generations. Individuals carry one of two alleles, a or A, and each individual inherits the allele of a parent which it samples independently and uniformly from the previous generation. The mechanism of sampling alleles by sampling parents ensures that the model carries information of the forward-in-time evolution of allele frequencies, as well as the backward-in-time genealogies of samples of individuals. Frequently, these two modelling perspectives satisfy a duality relation which renders both models more tractable than they would be in isolation. We direct interested readers to e.g. Durrett (2008) for an introduction to Wright–Fisher and genealogical models in population genetics.

In the absence of mutation, inheriting an allele from a uniformly sampled parent models *neutral* evolution, where the conditional mean allele frequency in a generation equals that in the previous generation. Non-neutral models in which the mean is not constant can be obtained by sampling offspring alleles as $f(p)$-coins when the allele frequency in the previous generation is p, and f models so-called *frequency-dependent selection*. In order to retain the aforementioned backward-in-time genealogical picture and its associated duality, it is desirable to generate the $f(p)$-coins in a given generation by sampling parental alleles from the previous generation. Since the allele frequency in the parental generation is p, parental alleles can be thought of as p-coins, motivating a connection to Bernoulli factories. Our contribution is to use a

* Corresponding author at: School of Mathematics, Statistics and Physics, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.
E-mail addresses: jere.koskela@newcastle.ac.uk (J. Koskela), k.g.latuszynski@warwick.ac.uk (K. Łatuszyński), d.spano@warwick.ac.uk (D. Spanò).

https://doi.org/10.1016/j.tpb.2024.01.002
Received 31 May 2023
Available online 30 January 2024

Bernoulli factory to extend two standard models of population genetics to more general settings than has been done previously. They are (i) the non-neutral Wright–Fisher diffusion and its ancestral selection graph dual (Cordero et al., 2022; Etheridge et al., 2010; Gonzalez Casanova and Spanò, 2018; Krone and Neuhauser, 1997), for which the function f models frequency-dependent selection as described above, and (ii) the Allen–Cahn PDE which models stationary allele frequencies in a spatial continuum, in which f appears as external forcing (Etheridge et al., 2022).

The remainder of the manuscript is organised as follows. In Section 2, we review the Bernoulli factory of Keane and O'Brien (1994) which turns out to be convenient for our purposes. In Sections 3 and 4 we introduce the Wright–Fisher diffusion with frequency–dependent selection and the Allen–Cahn equation. In each case, we also demonstrate how Bernoulli factories facilitate the construction of very large classes of these models. Section 5 concludes with a discussion on connections to earlier results, as well as some potential extensions.

2. The Keane–O'Brien factory

Let $f : [0,1] \mapsto [0,1]$ be continuous and polynomially bounded as in (1). Let $\bar{X}_n(p)$ be the sample mean of n independent p-coins. Define sequences of functions f_k and integers $\eta(f,k)$ as follows: set $f_1(p) := f(p)$, and

$$f_{k+1}(p) := \frac{4}{3}\left(f_k(p) - \frac{1}{4}\mathbb{P}\left(f_k(\bar{X}_{\eta(f,k)}(p)) \geq \frac{1}{2} \right) \right),$$

where each $\eta(f,k)$ is finite, independent of p, and large enough that

$$f_k(p) - \frac{1}{4}\mathbb{P}\left(f_k(\bar{X}_{\eta(f,k)}(p)) \geq \frac{1}{2} \right) \in [0, 3/4].$$

By Keane and O'Brien (1994, page 218) such an $\eta(f,k)$ exists, and moreover, each f_k is polynomially bounded.

Algorithm 1 Bernoulli factory for continuous $f : [0,1] \mapsto [0,1]$ satisfying (1).

Require: Sequences $\{f_k\}_{k\geq 1}$, $\{\eta(f,k)\}_{k\geq 1}$.
1: Sample $L \sim \text{Geo}(1/4)$.
2: Sample $X_{\eta(f,L)}(p) \sim \text{Bin}(\eta(f,L), p)$.
3: **if** $f_L(X_{\eta(f,L)}(p)/L) \geq 1/2$ **then**
4: Return 1.
5: **else**
6: Return 0.
7: **end if**

Let $L \sim \text{Geo}(1/4)$ be independent of all p-coins. Then

$$\mathbb{1}\left\{ f_L(\bar{X}_{\eta(f,L)}(p)) \geq \frac{1}{2} \right\} \sim \text{Ber}(f(p))$$

is a Bernoulli factory for f (Keane and O'Brien, 1994, pages 217–219) based on the series expansion

$$\begin{aligned} f(p) &= \sum_{k=1}^{\infty} \left(\frac{3}{4}\right)^{k-1} \frac{1}{4}\mathbb{P}\left(f_k(\bar{X}_{\eta(f,k)}(p)) \geq \frac{1}{2} \right) \\ &= \sum_{k=1}^{\infty} \left(\frac{3}{4}\right)^{k-1} \frac{1}{4} \sum_{j=0}^{\eta(f,k)} \binom{\eta(f,k)}{j} p^j (1-p)^{\eta(f,k)-j} \mathbb{1}\left\{ f_k\left(\frac{j}{\eta(f,k)}\right) \geq \frac{1}{2} \right\}, \end{aligned} \quad (2)$$

which converges uniformly in p. Pseudocode for this Bernoulli factory is shown in Algorithm 1. We will refer to Bernoulli factories based on (2) as Keane–O'Brien factories. They have two features which will turn out to be essential.

Remark 1. The Keane–O'Brien factory is defined for all $p \in [0,1]$ and $f(p) \in [0,1]$. Factories based on Bernstein polynomial approximation typically have to constrain the range (and sometimes the domain) of f to (subsets of) $(0,1)$ to avoid degenerate distributions (Holtz et al., 2011; Łatuszyński et al., 2011; Nacu and Peres, 2005). In population genetics applications, it is essential to allow the whole range $p \in [0,1]$, and desirable to allow $f(0) = 0$ and $f(1) = 1$ to model fixation.

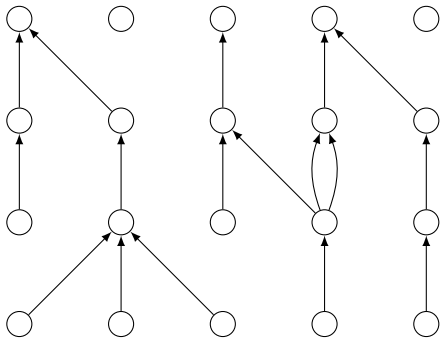

Fig. 1. Illustration of $\{\mathbf{Z}_k^{(N)}\}_{k\geq 0}$ with $N = 5$ and 4 generations. The second generation from the bottom includes a merger of three lineages, as well as an individual which samples three parents, two of which ended up overlapping. The generation at the top includes two binary mergers. Alleles for all individuals in all generations could be generated given alleles for the generation at the top, as well as a rule for deciding the allele of the three-parent child.

Remark 2. The random variable L, and hence the number of p-coins $\eta(f,L)$ needed to determine the allele of a child, is independent of the realisations of those coins. This will facilitate the construction of ancestral graphs containing all possible ancestors before the alleles carried by any of those ancestors are known.

Any Bernoulli factory for which Remark 2 holds can be thought of as a mixture of finite factories, that is, a mixture of Bernoulli factories with a deterministic number of coins each.

3. The frequency-dependent Wright–Fisher diffusion

Consider a population of N individuals $\{\mathbf{Z}_k^{(N)}\}_{k\geq 0}$ evolving in discrete time, where the state of the system at time k is $\mathbf{Z}_k^{(N)} := (Z_1^{(N)}(k), \dots, Z_N^{(N)}(k))$ with $Z_k^{(N)} \in \{a, A\}$. Let

$$\tilde{Y}_k^{(N)} := \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{\{A\}}(Z_i^{(N)}(k))$$

be the frequency of the A allele in generation k. Each individual in generation $k+1$ samples its allele conditionally independently given $\tilde{Y}_k^{(N)}$, where the probability of an A allele is $f^{(N)}(Y_k^{(N)})$ with

$$f^{(N)}(p) := \left(1 - \frac{\sigma}{N}\right)p + \frac{\sigma}{N} f(p)$$

for a fixed constant $\sigma > 0$, a continuous function $f : [0,1] \mapsto [0,1]$ satisfying (1), and where we assume the population size is large enough that $N \geq \sigma$.

Concretely, the realisation of each allele is implemented as follows. With probability $1 - \sigma/N$, a given individual samples one parent and inherits its allele. With the complementary probability, the individual samples an independent copy of L, followed by $\eta(f,L)$ parents chosen uniformly at random with replacement. Conditional on L and $\tilde{Y}_k^{(N)}$, the allele of the individual is A if event $\{f_L(\bar{X}_{\eta(f,L)}(\tilde{Y}_k^{(N)})) \geq 1/2\}$ happens. By (2), the marginal probability of an A allele is $f^{(N)}(\tilde{Y}_k^{(N)})$. A realisation of the process of children sampling parents is depicted in Fig. 1.

Sampling parents with replacement is essential for the connection with Bernoulli factories, as each parent can then be thought of as an i.i.d. coin on the two-point space of alleles $\{a, A\}$. However, sampling with replacement also permits a child to choose the same individual as a parent many times, which is unnatural from a biological point of view. Formally, the $\eta(f,L)$ parents are merely a mechanism for generating alleles for children in a parent-driven way which turns out to facilitate duality with a genealogical process. A biologically interpretable genealogy with one parent per child can be recovered by selecting, for each multi-parent child, one of

the child-parent connections to denote the "real" parent, with all other child-parent connections designated "virtual" and hence non-biological. The way in which the real parent is chosen can be arbitrary as long as it has no impact on the distribution of the child allele, and is independent of the child-parent connections of other individuals. However, retention of virtual parents is essential for constructing both the genealogical and allele frequency processes simultaneously, and hence obtaining the duality result in Theorem 3.

Our aim is to verify convergence of the allele frequency process $\{\tilde{Y}_k^{(N)}\}_{k\geq 0}$ in the infinite population limit. To that end, we define the continuous-time process $(Y_t^{(N)})_{t\geq 0}$ via

$$Y_t^{(N)} := \tilde{Y}_{\lfloor Nt\rfloor}^{(N)}.$$

Theorem 1. *Let f be polynomially bounded and Lipschitz continuous on $[0,1]$. Then, in the Skorokhod topology, $(Y_t^{(N)})_{t\geq 0} \to (Y_t)_{t\geq 0}$ weakly as $N\to\infty$, where Y_t solves*

$$\mathrm{d}Y_t = \sigma(f(Y_t)-Y_t)\mathrm{d}t + \sqrt{Y_t(1-Y_t)}\mathrm{d}W_t \tag{3}$$

subject to appropriate initial conditions, where $(W_t)_{t\geq 0}$ is a Brownian motion.

Proof. Existence, uniqueness, and the Feller property of the putative limiting process all follow from Ethier and Kurtz (1986, Theorem 2.8) because the drift function $\sigma(f(y)-y)$ is Lipschitz continuous.

The discrete-time process $\tilde{Y}_k^{(N)}$ has discrete generator

$$\begin{aligned} L^N h(y) := &\sum_{k=0}^{N}\binom{N}{k}\Big(\frac{\sigma}{N}\Big)^k\Big(1-\frac{\sigma}{N}\Big)^{N-k}\sum_{x=0}^{N-k}\binom{N-k}{x}y^x(1-y)^{N-k-x}\\ &\times\sum_{z=0}^{k}\binom{k}{z}f(y)^z(1-f(y))^{k-z}h\Big(\frac{x+z}{N}\Big)-h(y), \end{aligned}$$

which is well-defined for any bounded, measurable function $h:[0,1]\to\mathbb{R}$. To show the claimed convergence, let h be a $C^2([0,1])$ function with bounded third derivatives, which is a convergence-determining class. Expanding h around y on the right-hand side yields

$$\begin{aligned} L^N h(y) = &\sum_{k=0}^{N}\binom{N}{k}\Big(\frac{\sigma}{N}\Big)^k\Big(1-\frac{\sigma}{N}\Big)^{N-k}\sum_{x=0}^{N-k}\binom{N-k}{x}y^x(1-y)^{N-k-x}\\ &\times\sum_{z=0}^{k}\binom{k}{z}f(y)^z(1-f(y))^{k-z}\\ &\times\left[\Big(\frac{x+z}{N}-y\Big)h'(y)+\frac{1}{2}\Big(\frac{x+z}{N}-y\Big)^2h''(y)\right], \end{aligned}$$

up to terms which are of lower order since h''' is bounded. Evaluating the binomial expectations yields

$$NL^N h(y) = \sigma(f(y)-y)h'(y)+\frac{1}{2}y(1-y)h''(y)+o(1),$$

as $N\to\infty$. The proof of weak convergence is completed by using the pseudo-Poisson approximation of Kallenberg (2002, Theorem 19.28) with time step $1/N$, because the limiting diffusion is Feller. □

Remark 3. The requirement of a Lipschitz drift could be slightly relaxed by using the more cumbersome conditions for a Feller semigroup given in Xi et al. (2019, Theorem 3.3). They cover drifts satisfying a condition akin to

$$|f(y)-f(z)| \leq -C|y-z|\log(|y-z|),$$

or small variations thereof, where $C>0$ is a constant. See Xi et al. (2019, equations (14) and (15), as well as Remark 2.3) for a precise class of non-Lipschitz functions which can be handled.

Next we consider a sample of $n\in\mathbb{N}$ individuals from a given generation (which we arbitrarily set as generation 0) in the pre-limiting Wright–Fisher model $\{\mathbf{Z}_k^{(N)}\}_{k\geq 0}$. We define the *ancestral process* $A_k^{(N)}$ as the number of lineages which are ancestral to the sample k generations in the past. The number of lineages decreases by one when two lineages find a common ancestor, decreases by more than one if multiple lineages find common ancestors (which will happen with negligible probability when $N\to\infty$), or increase by $\eta(f,L)-1$ whenever a lineage samples $\eta(f,L)$ ancestors. In the pre-limiting particle system, any number of these events can co-occur in one generation, particularly when $n\geq 3$. But transitions other than isolated binary mergers and single multifurcations turn out to vanish in a suitably rescaled infinite population limit. To that end, we define the continuous time Markov jump process

$$A_t := \lim_{N\to\infty} A_{\lfloor Nt\rfloor}^{(N)} \tag{4}$$

whose existence we prove next.

Theorem 2. *The limit in (4) exists, and $(A_t)_{t\geq 0}$ has generator*

$$Gh(n) = \binom{n}{2}[h(n-1)-h(n)] + \sigma n\sum_{k=1}^{\infty}\Big(\frac{3}{4}\Big)^{k-1}\frac{1}{4}[h(n+\eta(f,k)-1)-h(n)].$$

Proof. The pre-limiting, discrete-time process $\{A_k^{(N)}\}_{k\geq 0}$ undergoes a myriad of transitions involving subsets of individuals finding common ancestors, as well as individuals branching into many potential ancestors, potentially in the same generation. However, only transitions with a per-generation probability $\Theta(1/N)$ will contribute to the time-rescaled limit with a finite rate. Events with probability $o(1/N)$ will not occur in the limit at all, while events with probability $\omega(1/N)$ will appear at a dense set of times. However, it will turn out that the latter only result in identity transitions in A_t, and hence do not affect the limit.

The probability of two lineages originating from a common ancestor one generation earlier, with neither being involved in a branching event, is

$$\Big(1-\frac{\sigma}{N}\Big)^2\frac{1}{N} = \frac{1}{N}+o(1/N), \tag{5}$$

implying that two lineages will merge to a common ancestor at rate 1 in the limit. A triple merger, or more than one simultaneous merger, has probability at most

$$\Big(1-\frac{\sigma}{N}\Big)^3\frac{1}{N^2} = o(1/N),$$

and hence will not appear in the limit.

A single individual branches into $\eta(f,k)$ ancestral lineages with probability

$$\frac{\sigma}{N}\Big(\frac{3}{4}\Big)^{k-1}\frac{1}{4}, \tag{6}$$

while any event involving more than two lineages branching in one generation has probability at most

$$\Big(\frac{\sigma}{N}\Big)^2 = o(1/N).$$

Hence, only isolated branching events appear in the limit. It is also clear from (5) and (6) that the probability of at least one merger and branching event in one generation is $o(1/N)$.

All other transitions involve no mergers or branching events, and hence do not affect the limiting ancestral process. Noting that there are $\binom{n}{2}$ pairs of individuals to merge with probability (5) and n individuals to branch with probabilities (6) yields the claimed generator G. □

Duality between the Wright–Fisher diffusion (3) and the ancestral process (4) is a relation

$$\mathbb{E}_y[h(Y_t,n)|Y_0=y] = \mathbb{E}_n[h(y,A_t)|A_0=n],$$

where $y\in[0,1]$ and $n\in\mathbb{N}$ are respective initial conditions, and h is a *duality function*, the specification of which will require some exposition.

We follow Cordero et al. (2022) and define the random function $P_t(y)$ as the conditional probability that all n leaves at time zero in the

ancestral process carry allele A, given that the A allele frequency at time t in the past is $y \in [0,1]$, and given the realisation of the ancestral process $A_{0:t} := (A_s)_{s\in[0,t]}$ started from the n lineages. For example, in the absence of branching events we have $P_t(y) = y^{A_t}$, while when $n = 1$, a single branching event into $\eta(f,k)$ ancestors and no mergers in the history $A_{0:t}$ yields

$$P_t(y) := \sum_{j=0}^{\eta(f,k)} \binom{\eta(f,k)}{j} y^j (1-y)^{\eta(f,k)-j} \mathbb{1}\Big\{ f_k\Big(\frac{j}{\eta(f,k)}\Big) \geq \frac{1}{2} \Big\},$$

where f_k is as specified in Section 2. Other patterns of merger and branching events will result in a more complicated Bernstein polynomial of degree A_t,

$$P_t(y) := \sum_{j=0}^{A_t} V_t(j) \binom{A_t}{j} y^j (1-y)^{A_t-j},$$

where $V_t(j)$ is the random Bernstein coefficient which equals the probability that the n leaves all carry allele A, given that j of A_t roots do. As in Cordero et al. (2022, Definition 2.12 and Proposition 2.13), the vector $(\boldsymbol{V}_t)_{t\geq 0} := (V_t(0), \dots, V_t(n))_{t\geq 0}$ is also a Markov jump process with transitions

$$\boldsymbol{v} \mapsto \Bigg(\sum_{j=0}^{i\wedge\eta(f,k)} \frac{\binom{i}{j}\binom{n+\eta(f,k)-1-j}{\eta(f,k)-j}}{\binom{n+\eta(f,k)-1}{i}} [\mathbb{1}\{f_h(j/\eta(f,k)) \geq 1/2\} v_{i+1-j} + \mathbb{1}\{f_k(j/\eta(f,k)) < 1/2\} v_{i-j}] \Bigg)_{i=0}^{n+\eta(f,k)-1} \tag{7}$$

at rate $n\sigma\mathbb{P}(L = k)$, and

$$\boldsymbol{v} \mapsto \Big(\frac{i}{n-1} v_{i+1} + \frac{n-1-i}{n-1} v_i \Big)_{i=0}^{n-1} \tag{8}$$

at rate $\binom{n}{2}$. Also, let

$$\boldsymbol{B}_n(x) := \left(\binom{n}{i} x^i (1-x)^{n-i} \right)_{i=0}^{n}$$

be the vector of order n Bernstein polynomials, and for vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ of common length $n+1$, let

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle := \sum_{i=0}^{n} u_i v_i$$

be the usual inner product.

Theorem 3. *The processes $(Y_t)_{t\geq 0}$ and $(V_t)_{t\geq 0}$ are dual with duality function $H : (y, \boldsymbol{v}) \mapsto \langle \boldsymbol{B}_{dim(\boldsymbol{v})-1}(y), \boldsymbol{v} \rangle$, in that*

$$\mathbb{E}[\langle \boldsymbol{B}_n(Y_t), \boldsymbol{v} \rangle | Y_0 = y] = \mathbb{E}[\langle \boldsymbol{B}_{dim(\boldsymbol{V}_t)-1}(y), \boldsymbol{V}_t \rangle | \boldsymbol{V}_0 = \boldsymbol{v}] \tag{9}$$

for all $t \geq 0$, each $y \in [0,1]$, every $n \in \mathbb{N}$, and each $\boldsymbol{v} \in \mathbb{R}^{n+1}$.

Proof. The proof is a small adaptation of that of Cordero et al. (2022, Theorem 2.14) to our more general setting. Duality between the coalescing mechanism of the ancestral process (8) and the diffusion coefficient of the Wright–Fisher diffusion (3) is standard, and we omit it to focus on establishing the same relation for the branching mechanism (7) and the drift term in (3).

Following Cordero et al. (2022, Section 4.3), let $Y_k^y \sim \mathrm{Bin}(k, y)$ and $K_{k,i}^n \sim \mathrm{Hyp}(n+k-1, k, i)$ be independent random variables, where $\mathrm{Hyp}(a,b,c)$ denotes the hypergeometric distribution with b draws without replacement from a population of size a containing c successes. Then, for $\boldsymbol{v} \in \mathbb{R}^{n+1}$ we have

$$\begin{aligned}
\partial_y H(y,\boldsymbol{v}) &= n\mathbb{E}[v_{Y_{n-1}^y+1} - v_{Y_{n-1}^y}], \\
H(y,\boldsymbol{v}) &= \mathbb{E}[v_{Y_n^y}] = \mathbb{E}[(1-y)v_{Y_{n-1}^y} + y v_{Y_{n-1}^y+1}], \\
f(y) &= \mathbb{E}[\mathbb{1}\{f_L \bar{X}_{\eta(f,L)}(y) \geq 1/2\}],
\end{aligned}$$

where $L \sim \mathrm{Geo}(1/4)$ and the third equality is due to (2). We also have

$$(\eta(f,L)\bar{X}_{\eta(f,L)}(y), Y_{n-1}^y) \overset{d}{=} (K^n_{\eta(f,L),Y^y_{n+\eta(f,L)-1}}, Y^y_{n+\eta(f,L)-1} - K^n_{\eta(f,L),Y^y_{n+\eta(f,L)-1}}), \tag{10}$$

as can be seen by considering the event that the left-hand side takes value (j,k), which requires $K^n_{\eta(f,L),Y^y_{n+\eta(f,L)-1}} = j$ and $Y^y_{n+\eta(f,L)-1} = k+j$ on the right-hand side. Using the shorthand $\eta = \eta(f,L)$, the probability of the event on the left can be written as

$$\begin{aligned}
\binom{\eta}{j} y^j (1-y)^{\eta-j} \binom{n-1}{k} y^k (1-y)^{n-1-k} &= \frac{\binom{\eta}{j}\binom{n+\eta-1-\eta}{k+j-j}}{\binom{n+\eta-1}{k+j}} \\
&\quad \times \binom{n+\eta-1}{k+j} y^{k+j} (1-y)^{n+\eta-1-k-j},
\end{aligned}$$

which is the claimed probability of the event on the right. Intuitively, (10) expresses the fact that flipping η coins and $n-1$ coins in separate groups is statistically identical to flipping $n+\eta-1$ coins and randomly allocating them into groups of size η and $n-1$ afterwards. Using (10), the drift of the Wright–Fisher diffusion (3) satisfies

$$\begin{aligned}
&\sigma(f(y)-y)\partial_y H(y,\boldsymbol{v}) \\
&= \sigma\mathbb{E}[\mathbb{1}\{f_L(\bar{X}_{\eta(f,L)}(y)) \geq 1/2\} - y] n\mathbb{E}[v_{Y^y_{n-1}+1} - v_{Y^y_{n-1}}] \\
&= n\sigma\mathbb{E}[\mathbb{1}\{f_L(\bar{X}_{\eta(f,L)}(y)) \geq 1/2\} v_{Y^y_{n-1}+1} - \mathbb{1}\{f_L(\bar{X}_{\eta(f,L)}(y)) \geq 1/2\} v_{Y^y_{n-1}} \\
&\quad - y v_{Y^y_{n-1}+1} + y v_{Y^y_{n-1}} + v_{Y^y_{n-1}} - v_{Y^y_{n-1}}] \\
&= n\sigma(\mathbb{E}[\mathbb{1}\{f_L(\bar{X}_{\eta(f,L)}(y)) \geq 1/2\} v_{Y^y_{n-1}+1} + \mathbb{1}\{f_L(\bar{X}_{\eta(f,L)}(y)) < 1/2\} v_{Y^y_{n-1}}] \\
&\quad - H(y,\boldsymbol{v})) \\
&= n\sigma\Bigg(\sum_{k=1}^{\infty} \Big(\frac{3}{4}\Big)^{k-1} \frac{1}{4} \sum_{j=0}^{n+\eta(f,k)-1} \mathbb{E}\Bigg[\mathbb{1}\Big\{ f_k\Big(\frac{K^n_{\eta(f,k),j}}{\eta(f,k)}\Big) \geq 1/2 \Big\} v_{j-K^n_{\eta(f,k),j}+1} \\
&\qquad + \mathbb{1}\Big\{ f_k\Big(\frac{K^n_{\eta(f,k),j}}{\eta(f,k)}\Big) < 1/2 \Big\} v_{j-K^n_{\eta(f,k),j}} \Bigg] \binom{n+\eta(f,k)-1}{j} \\
&\quad \times y^j (1-y)^{n+\eta(f,k)-1-j} - H(y,\boldsymbol{v}) \Bigg),
\end{aligned}$$

which is precisely (7) applied to the $\boldsymbol{v}$-argument of $H(y,\boldsymbol{v}) = \langle \boldsymbol{B}_{\dim(\boldsymbol{v})-1}(y), \boldsymbol{v} \rangle$. □

4. The Allen–Cahn model of spatial genetics

The Allen–Cahn equation on a Lipschitz domain $\Omega \subseteq \mathbb{R}^d$ is

$$\partial_t u - \Delta u = \lambda f(u), \tag{11}$$

for a given $\lambda > 0$ and $f : [0,1] \mapsto [0,1]$, and subject to suitable initial and boundary conditions. In Etheridge et al. (2022), the authors consider a model of the spatially structured frequency $u(x,t)$ of an allele A governed by

$$\begin{aligned}
\partial_t u - \Delta u &= \frac{1}{\varepsilon^2} u(1-u)(2u-1+\nu\varepsilon) \text{ for } x \in \Omega, t > 0, \\
\partial_n u &= 0 \text{ for } x \in \partial\Omega, t > 0, \\
u(x,0) &= u_0(x) \text{ for } x \in \Omega,
\end{aligned} \tag{12}$$

where $\varepsilon > 0$, $\nu > 0$, ∂_n denotes the normal derivative at the boundary, and $u_0 : \Omega \to [0,1]$ (see also Gooding (2018)). They construct a solution to (12) by using a particle system in which a single particle started at a location $x \in \Omega$ undergoes a Brownian motion with speed 2, branches into three independent copies at rate $(1+\varepsilon\nu)/\varepsilon^2$, and particles are reflected from the boundary. At a given end time $t > 0$, a leaf particle at location $z \in \Omega$ samples one of two alleles, a and A, with respective probabilities $(1-u_0(z), u_0(z))$. All particles sample their alleles independently. Then, particles propagate their alleles rootwards along the realisation of the Brownian tree. The allele of an internal branch is decided by a majority vote among its three children unless exactly one child carries allele A, in which case the parent branch carries A

with probability $2v\varepsilon/(3+3v\varepsilon)$, and a otherwise. The probability that the root particle $x \in \Omega$ carries allele A under these dynamics solves (12) (Etheridge et al., 2022, Proposition 2.4).

The ingredients of the particle system can be read off from (12). The Laplacian Δ is the generator of Brownian motion run at speed 2, the branching rate $(1+\varepsilon v)/\varepsilon^2$ is an upper bound for the right-hand side, and

$$\frac{1}{\varepsilon^2}u(1-u)(2u-1+v\varepsilon) = \frac{1+v\varepsilon}{\varepsilon^2}\left(u^3+3u^2(1-u)+\frac{2v\varepsilon}{3+3v\varepsilon}3u(1-u)^2-u\right),$$

which demonstrates that the nonlinearity in (12) has an interpretation as the voting system described above. It would be straightforward to adapt the proof of Etheridge et al. (2022, Proposition 2.4) to any other polynomial right-hand side of the form

$$\lambda\sum_{j=0}^{m} p_j\binom{m}{j}u^j(1-u)^{m-j}-u,$$

for some $m \in \mathbb{N}$, $\lambda > 0$, and coefficients $p_j \in [0,1]$ by setting λ as the branching rate into m particles, and suitably adapting the voting scheme. Our contribution is to prove an analogous result for (11), subject to the same initial and boundary conditions as (12), when f is merely continuous and polynomially bounded.

Consider a branching Brownian motion reflected off the boundary $\partial\Omega$, with branching rate λ, started from a single particle at $x \in \Omega$ at time 0. At a branching event, the number of offspring is given by $\eta(f, L)$, where $L \sim \mathrm{Geo}(1/4)$. At a terminal time $t > 0$, a leaf particle at $z \in \Omega$ samples allele a (resp. A) with probability $1-u_0(z)$ (resp. $u_0(z)$), and each leaf carries out this choice independently. Alleles are propagated rootwards along the tree: a branch with $\eta(f,k)$ offspring, j of whom carry allele A, carries allele A if $f_k(j/\eta(f,k)) \geq 1/2$, and carries allele a otherwise. Let $F(x,t)$ be the allele carried by the root particle at position $x \in \Omega$ when the branching process is run until time $t > 0$.

Theorem 4. *Suppose f is continuous and polynomially bounded as in (1). Viewed as a function of $x \in \Omega$, the probability $\mathbb{P}(F(x,t)=A)$ solves (11) subject to the initial and boundary conditions in (12).*

Proof. The proof is an adaptation of that of Etheridge et al. (2022, Proposition 2.4) to our more general setting. To verify that $q(x,t) := \mathbb{P}(F(x,t)=A)$ solves (11) on the interior of Ω, let S denote the first branching time of the initial particle and $(W_t)_{t\geq 0}$ be a Brownian motion. For small $h > 0$,

$$\begin{aligned}
q(x,t+h) &= \mathbb{P}(F(x,t+h)=A|S>h)\mathbb{P}(S>h)+\mathbb{P}(F(x,t+h)\\
&= A|S\leq h)\mathbb{P}(S\leq h)\\
&= \mathbb{E}_x[q(W_h,t)|S>h]\mathbb{P}(S>h)+\mathbb{P}(S\leq h)\sum_{k=1}^{\infty}\left(\frac{3}{4}\right)^{k-1}\frac{1}{4}\sum_{j=0}^{\eta(f,k)}\\
&\quad\times\mathbb{1}\left\{f_k\left(\frac{j}{\eta(f,k)}\right)\geq\frac{1}{2}\right\}\binom{\eta(f,k)}{j}\\
&\qquad\times\mathbb{E}_x[q(W_S,t+h-S)^j\\
&\quad\times(1-q(W_S,t+h-S))^{\eta(f,k)-j}|S\leq h]\\
&= \mathbb{E}_x[q(W_h,t)|S>h]e^{-\lambda h}+\mathbb{E}_x[f(q(W_S,t+h-S))|S\leq h](1-e^{-\lambda h}),
\end{aligned}$$

where the subscript in $\mathbb{E}_x$ denotes the starting point of $(W_t)_{t\geq 0}$ and the last equality follows via (2). Since f is continuous, regularity of the heat semigroup and the tower law (viewing $q(x,t)$ as the expectation of an indicator function) yield

$$\begin{aligned}
\mathbb{E}_x[f(q(W_S,t+h-S))|S\leq h] &= f(\mathbb{E}_x[q(W_S,t+h-S)|S\leq h])+O(h)\\
&= f(q(x,t))+O(h).
\end{aligned}$$

Hence,

$$\begin{aligned}
\partial_t\mathbb{P}(F(x,t)=A) &= \lim_{h\to 0}\frac{\mathbb{E}_x[\mathbb{P}(F(W_h,t)=A)|S>h]-\mathbb{P}(F(x,t)=A)}{h}e^{-\lambda h}\\
&\quad+\lim_{h\to 0}\frac{f(\mathbb{P}(F(x,t)=A))-\mathbb{P}(F(x,t)=A)}{h}(1-e^{-\lambda h})\\
&= \Delta\mathbb{P}(F(x,t)=A)+\lambda[f(\mathbb{P}(F(x,t)=A))-\mathbb{P}(F(x,t)=A)].
\end{aligned}$$

The boundary condition is inherited from reflecting Brownian motion (Bass and Hsu, 1991) since branching events occur at a finite rate and hence will not take place on the boundary. □

5. Discussion

In Cordero et al. (2022), the authors prove analogues of Theorems 1–3 for the case

$$f(x)-x := \sum_{\ell=2}^{m}\beta_\ell\sum_{i=0}^{\ell}\binom{\ell}{i}x^i(1-x)^{\ell-i}\left(p_{i,\ell}-\frac{i}{\ell}\right), \tag{13}$$

for a fixed $m \in \mathbb{N}$, positive coefficients $\{\beta_\ell\}_{\ell=2}^m$, and a sequence of $[0,1]$-valued coefficients $\{\{p_{i,\ell}\}_{i=0}^{\ell}\}_{\ell=2}^m$. Earlier work by González Casanova and Spanò also covered the case

$$f(x) := \sum_{j=0}^{\infty}\pi_j x^j,$$

where $\{\pi_j\}_{j=0}^{\infty}$ is a probability mass function (Gonzalez Casanova and Spanò, 2018). Our results cover all Lipschitz continuous, polynomially bounded functions $f : [0,1] \mapsto [0,1]$, which includes both of these classes as special cases. Furthermore, Cordero et al. (2022, Section 2.10) mentions that their approach should extend to the $m = \infty$ case. Our Bernoulli factory approach demonstrates that this is true, and also that there is no further difficulty in handling our more general class of drift functions.

The works of Gonzalez Casanova and Spanò (2018) and Cordero et al. (2022) are more general than our results in two ways: they tackle sequences of drift functions $f_N \to f$ as $N \to \infty$, and they incorporate jumps in the limiting forward-in-time Wright–Fisher diffusion, along with multiple mergers into the reverse-time ancestral process. Both of these generalisations could be incorporated into our model at the cost of increased technicality. We have chosen to omit them to focus on the class of drift functions, which is our main interest.

The link our work establishes between individual-based models and diffusive scaling limits provides rigorous justification for a range of diffusive approximations which have been obtained for non-neutral finite-population models (Taylor and Nowak, 2006; Lessard and Ladret, 2007). In addition, we provide an associated genealogical description and a formal duality relation between the two processes. Genealogical processes of Wright–Fisher diffusions with frequency-dependent selection are also considered in Coop and Griffiths (2004), though their approach is to condition on the allele frequency trajectory and hence avoid the need for branching events in the ancestral process. Their setting is formulated for generic drift functions of the form $\beta(x)x(1-x)$, though in practice they focus on Bernstein polynomials of degree no more than two.

Approaches based on moment duality, with duality function $h(x,n) = x^n$, have been used to obtain series expansions of Wright–Fisher diffusion transition functions (Barbour et al., 2000), and Cordero et al. (2022) uses the Bernstein duality in (9) and the Bernstein coefficient process $(\mathbf{V}_t)_{t\geq 0}$ to study fixation for drifts of the form (13). It is unclear whether similar results can be usefully obtained in our setting in practice because the sequences $\{f_k\}_{k\geq 1}$ and $\{\eta(f,k)\}_{k\geq 1}$ are difficult to compute. Hence, so are the Bernstein coefficients.

The construction of the solution to the Allen–Cahn Eq. (11) from a branching Brownian motion is also an example of duality between these two processes. This result could also be generalised by replacing the Laplacian with a more general second order differential operator, provided that it has suitable regularity and generates a diffusion with tractable reflecting behaviour at boundaries.

The key advantage of the Keane–O'Brien factory is that it covers the whole class of continuous, polynomially bounded functions f. Other

factories could be used to obtain different ancestral processes, Wright–Fisher diffusions, and constructions of the solution to the Allen–Cahn equation. However, it is essential that the number of coins needed by the factory is independent of the realisations of those coins, or at least that there is a almost surely finite upper bound on the number of coins regardless of realisations. Otherwise, the trick of keeping track of all resulting branches to fill in alleles (or votes) later cannot work, because the number of branches cannot be determined until earlier alleles have been resolved. As far as we are aware, the only other somewhat general Bernoulli factory with this independence property is that of Mendo (2019), which is identical to the branching mechanism used in Gonzalez Casanova and Spanò (2018) and applies to exactly the same functions. Seminal Bernoulli factories, such as that of Nacu and Peres (2005) based on Bernstein polynomial approximation of f, inherently link the decision to keep flipping coins to the outcomes of earlier coins.

Finally, there are multivariate analogues of Bernoulli factories, in which independent, m-sided dice with probability mass function $(p_1, \dots, p_m)$ are used to construct a v-sided die with mass function $f(p_1, \dots, p_m)$. To date, attention has focused on domains which exclude boundaries, and where the coordinates of f are rational functions (Morina, 2021; Morina et al., 2022), or where $v = 1$ (Paes Leme and Schneider, 2023). We believe a suitable extension of such *dice enterprises* to cases including boundaries could be used to obtain analogues of the convergence and duality results presented here in the case with more than two non-neutral alleles.

Funding and data sharing statement

This work was supported by the Engineering and Physical Sciences Research Council, United Kingdom grant number EP/V049208/1, and by the Royal Society through the Royal Society University Research Fellowship. Data sharing is not applicable to this article as no new data were generated or analysed.

CRediT authorship contribution statement

Jere Koskela: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Krzysztof Łatuszyński:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing. **Dario Spanò:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing.

Data availability

No data was used for the research described in the article.

References

Barbour, A.D., Ethier, S.N., Griffiths, R.C., 2000. A transition function expansion for a diffusion model with selection. Ann. Appl. Probab. 10, 123–162.
Bass, R.F., Hsu, P., 1991. Some potential theory for reflecting brownian motion in Holder and Lipschitz domains. Ann. Probab. 19 (2), 486–508.
Coop, G., Griffiths, R.C., 2004. Ancestral inference on gene trees under selection. Theor. Popul. Biol. 66, 219–232.
Cordero, F., Hummel, S., Schertzer, E., 2022. General selection models: Bernstein duality and minimal ancestral structures. Ann. Appl. Probab. 32 (3), 1499–1556.
Durrett, R., 2008. Probability Models for DNA Sequence Evolution. Springer.
Etheridge, A.M., Gooding, M.D., Letter, I., 2022. On the effects of a wide opening in the domain of the (stochastic) Allen–Cahn equation and the motion of hybrid zones. Electron. J. Probab. 27, Article 1.
Etheridge, A.M., Griffiths, R.C., Taylor, J.E., 2010. A coalescent dual process in a moran model with genic selection, and the Lambda coalescent limit. Theor. Popln Biol. 78 (2), 77–92.
Ethier, S.N., Kurtz, T.G., 1986. Markov Processes: Characterization and Convergence. John Wiley & Sons Inc..
Gonzalez Casanova, A., Spanò, D., 2018. Duality and fixation in Ξ-Wright–Fisher processes in frequency-dependent selection. Ann. Appl. Probab. 28 (1), 250–284.
Gooding, M., 2018. Long Term Behaviour of Spatial Population Models with Heterozygous or Asymmetric Homozygous Selection (Ph.D. thesis). Oxford University.
Holtz, O., Nazarov, F., Peres, Y., 2011. New coins from old, smoothly. Constr. Approx. 33, 331–363.
Kallenberg, O., 2002. Foundations of Modern Probability, 2nd Edition Springer, New York, NY.
Keane, M.S., O'Brien, G.L., 1994. A Bernoulli factory. AMC Trans. Model. Comput. Simul. 4, 213–219.
Krone, S.M., Neuhauser, C., 1997. Ancestral processes with selection. Theor. Popln Biol. 51 (3), 210–237.
Łatuszyński, K., Kosmidis, I., Papaspiliopoulos, O., Roberts, G.O., 2011. Simulating events of unknown probabilities via reverse-time martingales. Random Struct. Algorithms 38 (4), 441–452.
Lessard, S., Ladret, V., 2007. The probabilty of fixation of a single mutant in an exchangeable selection model. J. Math. Biol. 54, 721–744.
Mendo, L., 2019. An asymptotically optimal Bernoulli factory for certain functions that can be expressed as power series. Stochastic Process. Appl. 129, 4366–4384.
Morina, G., 2021. Extending the Bernoulli Factory to a Dice Enterprise (Ph.D. thesis). University of Warwick.
Morina, G., Łatuszyński, K., Nayar, P., Wendland, A., 2022. From the Bernoulli factory to a dice enterprise via perfect sampling of Markov chains. Ann. Appl. Probab. 32 (1), 327–359.
Nacu, Ş., Peres, Y., 2005. Fast simulation of new coins from old. Ann. Appl. Probab. 15, 93–115.
Paes Leme, R., Schneider, J., 2023. Multiparameter Bernoulli factories. Ann. Appl. Probab. 33 (5), 3987–4007.
Taylor, C., Nowak, M.A., 2006. Evolutionary game dynamics with non-uniform interaction rates. Theor. Popul. Biol. 69, 243–252.
von Neumann, J., 1951. Various techniques used in connection with random digits. Appl. Math. Ser. 12, 36–38, reprinted in von Neumann's Collected Works 5, Oxford University Press, 1963, 768–770.
Xi, F., Yin, G., Zhu, C., 2019. Regime-switching jump diffusions with non-Lipschitz coefficients and countably many switching states: existence and uniqueness, Feller, and strong Feller properties. In: Yin, G., Zhang, Q. (Eds.), Modelling, Stochastic Control, Optimization, and Applications. Springer, Cham.