Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Measure-valued diffusions, general coalescents and population genetic inference

Matthias Birkner[1], Jochen Blath[2]

submitted: 2nd July 2007

[1] Weierstraß-Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39
10117 Berlin
Germany
E-mail: `birkner@wias-berlin.de`

[2] Institut für Mathematik, Technische Universität Berlin
Straße des 17. Juni 136
10623 Berlin
Germany
E-mail: `blath@math.tu-berlin.de`

No. 1243
Berlin 2007

2000 *Mathematics Subject Classification.* Primary: 92D15; Secondary: 60J70, 60G09, 60G57.

Key words and phrases. Generalised Fleming-Viot process, Λ-coalescent, lookdown construction, mathematical population genetics, Monte-Carlo simulation.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We review recent progress in the understanding of the interplay between population models, measure-valued diffusions, general coalescent processes and inference methods for evolutionary parameters in population genetics. Along the way, we will discuss the powerful and intuitive (modified) lookdown construction of Donnelly and Kurtz, Pitman's and Sagitov's Λ-coalescents as well as recursions and Monte Carlo schemes for likelihood-based inference of evolutionary parameters based on observed genetic types.

1 Introduction

We discuss mathematical models for an effect which in population genetics jargon, somewhat orthogonal to diffusion process nomenclature, is called "genetic drift", namely the phenomenon that the distribution of genetic types in a population changes in the course of time simply due to stochasticity in the individuals' reproductive success and the finiteness of all real populations. We will only consider "neutral" genetic types. This contrasts and complements the notion of selection, which refers to scenarios in which one or some of the types confer a direct or indirect reproductive advantage to their bearers. Thus, in the absence of demographic stochasticity, the proportion of a selectively advantageous type would increase in the population, whereas that of neutral types would remain constant. The interplay between small fitness differences among types and the stochasticity due to finiteness of populations leads to many interesting and challenging problems, see e.g. the article by A. Etheridge, P. Pfaffelhuber and A. Wakolbinger [EPW07].

Genetic drift can be studied using two complementary approaches, which are in many senses, some of which we will discuss below, dual to each other. Looking "forwards" in time, the evolution of the type distribution can be approximately described by Markov processes taking values in the probability measures on the space of possible types. Looking "backwards", one describes the random genealogy of a sample from the population. Given the genealogical tree, one can then superimpose the mutation process in a second step. The article by P. Mörters [M07b] studies asymptotic properties of these genealogical trees in the limit of a large sample sizes as an example of the use of the multifractal spectrum.

The classical model for genetic drift is the so-called Wright-Fisher diffusion, which is appropriate when the variability of the reproductive success among individuals is small. Recently, there has been mathematical and biological interest in situations where the variance of the number of offspring per individual is (asymptotically) infinite, and detailed descriptions of the possible limiting objects have been obtained. We review these developments, giving particular emphasis to the interplay between the forwards models, generalised Fleming-Viot processes, and their dual backwards models, Λ-coalescents. We use this opportunity to advertise the lookdown construction of Donnelly and Kurtz (in its [DK99] "flavour"), which provides a realisation-wise coupling for this duality. Furthermore, we show how these approaches can be used to derive recursions for the probabilities of observed types in a sample from a stationary population. These recursions can usually not be solved in closed form and can be difficult to implement exactly, in particular if the space of possible types or the sample size is large. We describe a Monte-Carlo method which allows an approximate solution.

Many important and interesting aspects of mathematical population genetic models are omitted in our review, in particular the possibilities of varying population sizes, selective effects, spatial or other population substructure, multi-locus dynamics and recombination. We also focus on haploid models, meaning that our individuals have only one parent. For an introduction to coalescents with emphasis on biology, see e.g. [H90], [N01], [HSW05], [W06], for background on (classical and generalised) Fleming-Viot processes and variations of Kingman's coalescent, see e.g. [EK86], [D93], [EK93] and [DK99].

2 Population genetic models with neutral types

Cannings-models. In neutral population models, the main (and only) sources of stochasticity are due to random genetic drift and mutation. The first feature is captured in a basic class of population models, namely the so-called *Cannings-models* ([C74, C75]). We will subsequently extend these by adding mutations.

Consider a (haploid) population of constant size (e.g. due to a fixed amount of resources) consisting of, say, N individuals. Suppose the population is undergoing "random mating" with fixed non-overlapping generations and ideally has evolved for a long time, so that it can be considered "in equilibrium". In each generation $t \in \mathbb{Z}$, the distribution of the offspring numbers is given by a non-trivial random vector

$$(\nu_1^{(t)}, \dots, \nu_N^{(t)}) \quad \text{with} \quad \sum_{i=1}^{N} \nu_i^{(t)} = N, \tag{1}$$

where $\nu_k^{(t)}$ is the number of children of individual k. The vectors $\nu^{(t)}$, $t \in \mathbb{Z}$ are assumed i.d.d.

Neutrality means that we additionally suppose that the distribution of each such random vector is *exchangeable*, i.e. for each permutation $\sigma \in S_N$, we have that

$$(\nu_{\sigma(1)}, \dots, \nu_{\sigma(N)}) = (\nu_1, \dots, \nu_N) \quad \text{in law.}$$

If these conditions are met, we speak of a Cannings-model.

To explain the notion of random genetic drift, imagine that each individual has a certain genetic type. For example, at the genetic locus under consideration, each individual is of one of the types (or alleles) $\{a, A\}$. Each type is passed on unchanged from parent to offspring (we will introduce mutation to this model later).

For each generation t, let X_t denote the number of individuals which carry the "a"-allele. By the symmetries of the model, $\{X_t\}$ is a finite Markov-chain on $\{0, \dots N\}$ as well as a martingale. In particular, we may represent its dynamics as

$$X_{t+1} = \sum_{i=1}^{X_t} \nu_i^{(t)}. \tag{2}$$

Note that although $\mathbb{E}[X_t] = X_0$ for all t (due to the martingale property), the chain will almost surely be absorbed in either 0 or N in finite time. In fact, the probability that type a will be fixed in the population equals its initial frequency X_0/N. This is a simple example of the power of genetic drift: although in this model there is no evolutionary advantage of one of the types over the other, one type will eventually get fixed (this force will later be balanced by mutation, which introduces new genetic variation).

2.1 "Classical" limit results in the finite variance regime

Two-type neutral Wright-Fisher model. The classical example from this class is the famous *Wright-Fisher model* ([F22], [W31]). Informally, one can think of the following reproduction mechanism. At generation t, each individual picks one parent uniformly at random from the population alive at time $t-1$ and copies its genetic type (i.e. either a or A). Denoting by $p_{t-1} = X_{t-1}/N$ the proportion of alleles of type a in generation $t-1$, the number X_t of a-alleles in generation t is binomial, that is,

$$\mathbb{P}\{X_t = k | X_{t-1}\} = \binom{N}{k} p_{t-1}^k (1-p_{t-1})^{N-k}.$$

Compliant with (1), the offspring vector $(\nu_1, \ldots, \nu_N)$ would be multinomial with N trials and success probabilities $1/N, \ldots, 1/N$.

The Wright-Fisher diffusion as a limit of "many" Cannings-models. For large populations, it is often useful to pass to a diffusion limit. To this end, denote by

$$Y^N(t) := \frac{1}{N} X_{\lfloor t/c_N \rfloor}, \quad t \geq 0,$$

where $\lfloor t/c_N \rfloor$ is the integer part of t/c_N, and the time scaling factor is

$$c_N := \frac{\mathbb{E}[\nu_1(\nu_1-1)]}{N-1} = \frac{\mathbb{V}[\nu_1]}{N-1}, \tag{3}$$

the (scaled) "offspring variance". Note that c_N can also be interpreted as the probability that two randomly sampled individuals from the population have the same ancestor one generation ago (this will be important in Section 3). The following exact conditions for convergence follow from the conditions given by [MS01] and a straightforward application of duality, which we will discuss below [see (42)]: If

$$c_N \to 0 \quad \text{and} \quad \frac{\mathbb{E}[\nu_1(\nu_1-1)(\nu_1-2)]}{N^2 c_N} \to 0 \qquad \text{as} \quad N \to \infty, \tag{4}$$

$\{Y_t^N\}$ converges weakly to a diffusion process $\{Y_t\}$ in $[0,1]$, which is the unique strong solution of

$$dY_t = \sqrt{Y_t(1-Y_t)}\, dB_t, \quad Y_0 := x \in [0,1],$$

where $\{B_t\}$ is a standard Brownian motion. Equivalently, $\{Y_t\}$ is characterised as a (strong) Markov process with generator

$$Lf(y) = \frac{1}{2} y(1-y) \frac{d^2}{dy^2} f(y), \quad y \in [0,1], \quad f \in C^2([0,1]). \tag{5}$$

To this continuous model, the machinery of one-dimensional diffusion theory may be applied. For example, it is easy to compute the mean time to fixation, if $Y_0 = x$, which is

$$m(x) = -2(x \log x + (1-x)\log(1-x)).$$

In terms of the original discrete model, if $X_0/N = 1/2$, one obtains

$$(2\log 2)\frac{N}{\sigma^2} \approx 1.39\frac{N}{\sigma^2} \text{ generations}$$

(assuming that asymptotically, $\mathbb{V}[\nu_1] \approx \sigma^2$).

Moran's model. An equally famous model for a discrete population, living in *continuous* time, customarily attributed to P. A. P. Moran, works as follows: Each of the N individuals carries an independent exponential clock (with rate 1). If a bell rings, the corresponding individual (dies and) copies the type of a uniformly at random chosen individual from the current population (including itself). Another way to think about this is to pick the jumps times according to a Poisson-process with rate N and then independently choose a particle which dies and another particle which gives birth.

Note that even though this model does not literally fit into the Cannings class, its "skeleton chain" *is* a Cannings model with ν uniformly distributed on all the permutations of

$$(2, 0, 1, 1, \ldots, 1).$$

The fraction of type a-individuals in both models, suitably rescaled, converge to the Wright-Fisher diffusion: the continuous-time variant has to be sped up by a factor of N, the skeleton chain by a factor of N^2, as N skeleton steps roughly correspond to one "generation".

Remark (several types and higher-dimensional Wright-Fisher diffusions). It is straightforward to extend the discussion above to a situation with finitely-many (say k) genetic types, and obtain analogous limit theorems. Under the same assumptions, the fraction of type i in generation $\lfloor t/c_N \rfloor$ is approximately described by Y_t^i, where

$$Y_t = (Y_t^1, \ldots, Y_t^k) \in \{(y_1, \ldots, y_k) \,:\, y_i \geq 0, \, \textstyle\sum_i y_i = 1\}$$

is a diffusion with generator $L^{(k)}$, acting on $f \in C^2(\mathbb{R}^k)$ as

$$L^{(k)} f(y) = \frac{1}{2}\sum_{i,j=1}^{k} y_i(\delta_{ij} - y_j)\frac{\partial^2}{\partial y_i \partial y_j} f(y). \tag{6}$$

□

Fleming-Viot processes and infinitely many types. To incorporate scenarios with infinitely many possible types, it is most convenient to work with measure-valued processes. For simplicity and definiteness, we choose here $E = [0, 1]$ as the space of possible types, and consider random processes on $\mathcal{M}_1([0, 1])$. For example, let $\tilde{X}(t, i)$ (with values in E) be the type of individual i in generation t in a Cannings model, and let

$$Z_t^N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\tilde{X}(t,i)} \tag{7}$$

be the empirical type distribution in generation t. Then, under Assumptions (4), if $Z_0^N \Rightarrow \mu \in \mathcal{M}_1([0,1])$, the rescaled processes $\{Z^N_{\lfloor t/c_N \rfloor}\}$ converge weakly towards a measure-valued diffusion $\{Z_t\}$, which uniquely solves the (well-posed) martingale problem with respect to the generator

$$\mathcal{L}\Phi(\mu) = \sum_{J \subseteq \{1,\dots,p\}, |J|=2} \int \mu(da_1)\cdots\mu(da_p)\left(\phi(a_1^J,\dots,a_p^J) - \phi(a_1,\dots,a_p)\right) \quad (8)$$

for $\mu \in \mathcal{M}_1([0,1])$ and test functions

$$\Phi(\mu) = \int \phi(a_1,\dots,a_p)\,\mu(da_1)\cdots\mu(da_p), \quad (9)$$

where $p \in \mathbb{N}$ and $\phi : [0,1]^p \to \mathbb{R}$ is measurable and bounded, and for $a = (a_1,\dots,a_p) \in [0,1]^p$ and $J \subseteq \{1,\dots,p\}$, we put

$$a_i^J = a_{\min J} \text{ if } i \in J, \text{ and } a_i^J = a_i \text{ if } i \notin J,\ i = 1,\dots,p, \quad (10)$$

see e.g. [EK86], Ch. 10, Thm 4.1. Thinking of a as the types of a sample of size p drawn from μ, passage from a to a^J means a coalescence of a_i, $i \in J$.

In particular, if $\mu = \sum_{i=1}^k y_i\delta_{a_i}$ for k different points $a_i \in [0,1]$, then

$$Z_t = \sum_{i=1}^k Y_t^i \delta_{a_i},$$

where $\{Y_t^i : i = 1,\dots,k,\ t \geq 0\}$ is the k-dimensional Wright-Fisher diffusion with generator (6).

2.2 Beyond finite variance: occasional extreme reproduction events

Since the end of the 1990ies, more general reproduction mechanisms and their infinite population limits have been studied in the mathematical community ([S99], [P99], [DK99], [MS01], [S00]).

Although the motivation for this came from considerations about the genealogy of population resp. coalescent processes, we describe the corresponding population models forward in time first. Many of the technical assumption here will become more clear after the next Section has been studied.

Implicit in (4) is the assumption that each family size ν_i is small compared to the total population size N. A natural generalisation, motivated by considering species with potentially very many offspring, is to consider scenarios where occasionally, a single family is of appreciable size when compared to N. In this spirit, Eldon and Wakeley ([EW06]) proposed a family of Cannings models, where in a population of size N, ν is a (uniform) permutation of

$$\left(2,0,1,\dots,1\right) \quad \text{or of} \quad \left(\lfloor \psi N \rfloor, \underbrace{0,0,\dots,0}_{\lfloor \psi N \rfloor \text{ times}}, 1,\dots,1\right) \quad (11)$$

with probability $1 - N^{-\gamma}$ resp. $N^{-\gamma}$ for some fixed parameter $\psi \in (0,1]$ and $\gamma > 0$. The idea is of course that from time to time, an exceptionally large family is produced, which recruits a (non-negligible) fraction ψ of the next generation.

This is appealing as being presumably the conceptually simplest model of this phenomenon. On the other hand, while one may be willing to accept the assumption that in a species with high reproductive potential and variability, such extreme reproductive events can occur, the stipulation that these generate *always the same* fraction ψ is certainly an over-simplification.

A more realistic model would allow *"random"* ψ, where the parameter ψ is chosen according to some (probability) measure F. So far, the question which F are "natural" for which biological applications is largely open.

A plausible class of Cannings models for scenarios with (asymptotically) heavy-tailed offspring distributions has been introduced and studied by Schweinsberg ([S03]): In each generation, individuals generate *potential* offspring as in a supercritical Galton-Watson process, where the tail of the offspring distribution varies regularly with index α, more precisely the probability to have more than k children decays like Const. $\times k^{-\alpha}$. Among these, N are sampled without replacement to survive and form the next generation. The parameter $\alpha \in (1, 2]$ governing the tail of individual litter sizes characterises the limit process, and intuitively smaller α corresponds to more extreme variability among offspring numbers.

Mathematically, the situation is well understood (see [S99], [MS01]): For the discussion of limit processes, we first specialise to the situation of two types only. Consider the Markov chain (2) on the time scale $1/c_N$, where c_N is defined in (3). If $c_N \to 0$, for some probability measure F on $[0, 1]$

$$\frac{N}{c_N}\mathbb{P}\{\nu_1 > Nx\} \longrightarrow \int_{(x,1]} \frac{1}{y^2}\, F(dy) \tag{12}$$

for all $x \in (0, 1)$ with $F(\{x\}) = 0$ and

$$\frac{\mathbb{E}[\nu_1(\nu_1 - 1)\, \nu_2(\nu_2 - 1)]}{N^2} \cdot \frac{1}{c_N} \longrightarrow 0\,, \quad \text{as } N \to \infty, \tag{13}$$

then the processes $\{X^N_{\lfloor t/c_N \rfloor}/N\}$ converge weakly to a Markov process $\{Y_t\}$ in $[0, 1]$ with generator

$$\begin{aligned} Lf(y) = {} & \frac{F(\{0\})}{2} y(1 - y)\frac{d^2}{dy^2} f(y) \\ & + \int_{(0,1]} \Big(yf\big((1 - r)y + r\big) + (1 - y)f\big((1 - r)y\big) - f(y)\Big)\frac{1}{r^2}F(dr) \end{aligned} \tag{14}$$

for $f \in C^2([0, 1])$. The moment condition (13) has a natural interpretation in terms of the underlying genealogy, see the remark about simultaneous multiple collisions on page 15. Alternatively, $\{Y_t\}$ can be described as the solution of

$$\begin{aligned} dY_t = {} & \sqrt{F(\{0\})Y_{t-}(1 - Y_{t-})}\, dB_t \\ & + \int\limits_{(0,t]\times(0,1]\times[0,1]} \Big(1_{\{u\le Y(t-)\}}r(1 - Y_{t-}) - 1_{\{u>Y(t-)\}}rY_{t-}\Big)\, N\big(ds\, dr\, du\big), \end{aligned} \tag{15}$$

where $\{B_t\}$ is a standard Brownian motion and N is an independent Poisson process on $[0, \infty) \times (0, 1] \times [0, 1]$ with intensity measure $dt \otimes r^{-2}F_0(dr) \otimes du$ with $F_0 = F - F(\{0\})\delta_0$. Here, $r^{-2}F_0(dr)$ is the intensity with which exceptional reproductive events replacing a fraction r of the total population occur in the limiting process.

The class considered by Eldon and Wakeley ([EW06]) leads to $F = \delta_0$ for $\gamma > 2$,

$$F = \frac{2}{2 + \psi^2}\delta_0 + \frac{\psi^2}{2 + \psi^2}\delta_\psi \quad \text{for } \gamma = 2 \tag{16}$$

and δ_ψ for $1 < \gamma < 2$. The models considered by Schweinsberg in [S03] yield Beta measures, namely

$$F(dr) = \frac{\Gamma(2)}{\Gamma(2-\alpha)\Gamma(\alpha)} r^{1-\alpha}(1-r)^{\alpha-1}\, dr. \tag{17}$$

In [BBC05], these processes have been characterised as time-changes of α-stable continuous-mass branching processes renormalised to have total mass 1 at any time.

For the situation with infinitely many possible types, the corresponding limiting generalised Fleming-Viot process can be considered as a measure-valued diffusion with càdlàg paths whose generator, on test functions of the form (9) with ϕ two times continuously differentiable, is

$$F(\{0\})\mathcal{L}\Phi(\mu) + \int_E \int_{(0,1]} \Big(\Phi\big((1-r)\mu + r\delta_a\big) - \Phi(\mu)\Big) \frac{F_0(dr)}{r^2}\, \mu(da), \tag{18}$$

where $\mathcal{L}$ is defined in (8).

2.3 Introducing mutation.

We now introduce another major evolutionary "player", which counteracts the levelling force of random genetic drift. Indeed, when on the right scale, see (22) below, mutation continuously introduces new types to a population, leading to reasonable levels of genetic variability.

Example: The two alleles case. For our pre-limiting Cannings-models, imagine the following simple mechanism. At each reproduction event, particles retain the type of their parents with high probability. However, with a small probability, the type can change according to some mutation mechanism. In the situation of the two-allele model given by the types $\{a, A\}$, suppose that independently for each child, a mutation from parental type a to A happens with probability $p^{(N)}_{a\to A}$, and denote by $p^{(N)}_{A\to a}$ the corresponding probability for a mutation from A to a.

Let c_N, as defined in (3), tend to zero. If the assume, in addition to (12), (13), that

$$\frac{p^{(N)}_{a\to A}}{c_N} \to \mu_{a\to A} \quad \text{ and } \quad \frac{p^{(N)}_{A\to a}}{c_N} \to \mu_{A\to a}, \tag{19}$$

then, the process describing the fraction of the a-population, converges to a limit which has generator, for a suitable test-function $f \in C^2$, given by

$$Lf(y) + \Big(- y\mu_{a\to A} + (1-y)\mu_{A\to a}\Big) \frac{d}{dy} f(y), \tag{20}$$

where L is given by (14).

General mutation mechanisms. Here, we come back to consider measure-valued diffusions on some type space E. Let E be a compact metric space (we will later usually assume $E = [0,1]^{\mathbb{Z}}$ or $[0,1]$). To describe a mutation mechanism, let $q(x, dy)$ be a Feller transition function on $E \times \mathcal{B}(E)$, and define the bounded linear operator B on the set of bounded function on E by

$$Bf(x) = \int_E \big(f(y) - f(x)\big)\, q(x, dy). \tag{21}$$

Denote the individual mutation probability per individual in the N-th stage of the population approximation by r_N and assume that

$$\frac{r_N}{c_N} \to r \in [0, \infty), \tag{22}$$

where c_N is defined in (3). Note that the scaling depends on the class of Cannings models considered. For example, for the models in the domain of attraction of a Beta-coalescent [see the considerations leading to (17)], the choice of α fixes the scaling of the individual mutation probability μ per generation: in a population of size N, this translates to a rate

$$r = C_\alpha N^{\alpha-1}\mu \tag{23}$$

with which mutations appear in the limit. In the case $\alpha = 2$, this is the familiar formula $r\ (=\theta/2) = 2N\mu$.

Then, the empirical process $\{Z_t^N\}$, describing the distribution of types on E and defined in analogy to (7), converges to a limiting Markov process Z, whose evolution is described by the generator [using the notation from (9)]

$$\mathcal{L}_{B,F}\Phi(\mu) = r\sum_{i=1}^{p}\int_{E^p} B_i(\phi(a_1,\dots,a_p))\mu^{\otimes p}(da_1\dots da_p) + \mathcal{L}_F\Phi(\mu), \tag{24}$$

where $\mathcal{L}_F$ is defined by (18), and $B_i\phi$ is the operator B, defined in (21), acting on the i-th coordinate of ϕ. This process is called the F-generalised Fleming-Viot process with individual mutation process B. Note that in the nomenclature of [BLG03], this would be a ν-generalised FV process with $\nu(dr) = F(dr)/r^2$.

General Moran model with mutation. While the Cannings class uses discrete generations, the phenomena discussed above can also be expressed in terms of a continuous time model, which is a natural generalisation of the classical Moran model. For a given (fixed) total population size N let $\mathcal{B}_N$ be a Poisson process on $[0,\infty)\times\{1,2\dots,N-1\}$ with intensity measure $dt\otimes\mu_N$, where μ_N is some finite measure. If (t,k) is an atom of $\mathcal{B}_N$, then at time t, a "k-birth event" takes place: k uniformly chosen individuals die and are immediately replaced by the offspring of another individual, which is picked uniformly among the remaining $N-k$. "Extreme" reproductive events can thus be included by allowing μ_N to have suitable mass on ks comparable to N. The classical Moran model, in which only single birth events occur, corresponds to $\mu_N = N\delta_1$.

Additionally, assume that individuals have a type in E, and each particle mutates during its lifetime independently at rate $r_N \geq 0$ according to the jump process with generator B given by (21). Write $X_i^{(N)}(t)$ for the type of individual i at time t.

Let us denote the empirical process for the N-particle system by

$$Z_N(t) := \frac{1}{N}\sum_{i=1}^{N}\delta_{X_i^{(N)}(t)}. \tag{25}$$

We will further on assume that $X_i^{(N)}(0) = X_i$, $i = 1,\dots,N$, where the X_i are exchangeable and independent of $\mathcal{B}_N$, so in particular $\lim_{N\to\infty} Z_N(0)$ exists a.s. by de Finetti's Theorem.

For a reasonable large population limit, one obviously has to impose assumptions on μ_N and r_N. To connect to the formulation in [DK99], note that $\mathcal{B}_N$ can be equivalently described by the "accumulated births" process

$$A_N(t) := \sum_{(s,k)\in \mathrm{supp}(\mathcal{B}_N),\, s\le t} k, \qquad t \ge 0, \tag{26}$$

which is of course simply a compound Poisson process. We write $[A_N](t) = \sum_{s\le t} \big(\Delta A_N(s)\big)^2$ for the quadratic variation of A_N. Then, if

$$N r_N \to r \tag{27}$$

and

$$\frac{[A_N](Nt) + A_N(Nt)}{N^2} =: U_N(t) \;\Rightarrow\; U(t), \tag{28}$$

where the limit process U must necessarily be a subordinator with generator

$$G_U f(x) = \int_{[0,1]} \big(f(x+u) - f(x)\big)\tilde{\nu}(du) + af'(x), \tag{29}$$

the time-rescaled empirical processes

$$\big\{Z_N(Nt), t \ge 0\big\} \;\Rightarrow\; Z, \quad \text{as } N \to \infty,$$

where $\{Z(t)\}$ is the solution of the well-posed martingale problem corresponding to (24), see [DK99], Theorems 3.2 and 1.1. The relation between G_U and F appearing in (24) is as follows:

$$a = 2F(\{0\}), \quad \tilde{\nu} \text{ is the image measure of } \frac{1}{r^2}F(dr) \text{ under } r \mapsto \sqrt{r}. \tag{30}$$

The latter is owed to the fact that "substantial" birth events, where k is of order N, appear with their squared relative size as jumps of U_N.

While the Assumption (28) is quite general, it is instructive (and will be useful later) to specialise to a particular class of approximating birth event rates $\mu_N(\{k\})$ which is closely related to the limiting operators (18): For a given $F \in \mathcal{M}_1([0,1])$, put

$$\begin{aligned}\mu_N(\{k\}) =& NF(\{0\})1_{\{k=1\}}\\ &+ \frac{1}{N}\int_{(0,1]} \binom{N}{k+1} r^{k+1}(1-r)^{N-k-1}\frac{1}{r^2}F(dr), \quad k = 1,\dots,N-1.\end{aligned} \tag{31}$$

Then, (28) is fulfilled and the limiting U is described by (29) and (30). This is the (randomised) "Moran equivalent" of the "random ψ" discussed in Subsection 2.2, and will turn out to be the natural mechanism of the first N levels of the lookdown construction, see below. A way to think about the second term in (31) is that particles participate in an "r-extreme birth event" independently with probability r. Note that (31) implies, for any $x \in (0,1)$ with $F(\{x\}) = 0$,

$$N \sum_{k\ge \lfloor xN\rfloor}^{N-1} \mu_N(\{k\}) \underset{N\to\infty}{\longrightarrow} \int_x^1 \frac{1}{r^2}F(dr), \tag{32}$$

so in the limiting process, "x-reproductive events" occur at rate $dt \otimes x^{-2}F(dx)$. As in a k-birth event, the probability for a given particle to die is k/N, (31) implicitly defines the average lifetime of an individual in the N-th approximating model. The individual death rate of a "typical" particle in the N-particle model is

$$d_N = \sum_{k=1}^{N-1} \frac{k}{N} \mu_N(\{k\}) = F(\{0\}) + \int_{(0,1]} \frac{1-(1-r)^{N-1}}{r} F(dr). \tag{33}$$

If $1/r$ is not in $L_1(F)$, this will diverge as $N \to \infty$. In the last paragraph of the remark about "coming down from infinity" on page 16, we will see a relation to structural properties of the corresponding coalescents. Also note that (27) and (33) implicitly determine the mutation rate per "lifetime unit" in the N-th model, similarly as in (23).

Popular mutation models. Having the full generator (24) at hand, it is now easy to specialise to the following classical mutation models.

1) *Finitely-many alleles.* In this model, we assume a general finite type space, say, $E = \{1, \dots, d\}$. Then, the mutation mechanism can always be written as a stochastic transition matrix $P = (P_{ij})$ times the overall mutation rate $r \in (0, \infty)$. That is,

$$Bf(i) = r \sum_{j=1}^{d} P_{ij} \big(f(j) - f(i)\big).$$

2) *Infinitely-many alleles.* Here, one assume that each mutation leads to an entirely new type. Technically, one simply assumes that $E = [0,1]$ and that each mutation, occuring at rate $r > 0$, independently picks a new type $x \in [0,1]$, according to the uniform distribution on $[0,1]$, i.e.

$$Bf(y) = r \int_{[0,1]} \big(f(x) - f(y)\big) dx.$$

Note that this the paradigm example of a parent-independent mutation model. After one mutation step, all information about the ancestral type is lost.

3) *Infinitely-many sites model.* One thinks of a long part of a DNA sequence, so that each new mutation occurs at a different site. Hence in principle, the information about the ancestral type is retained. Moreover, it is possible to speak about the "distance" between two types (e.g. by counting the pairwise differences).

As a rule of thumb, if the number of mutations observed is small compared to the square-root of the length of the sequence, this assumption is reasonable. For a mathematical formulation, one may set $E = [0,1]^{\mathbb{N}}$ and define the mutation operator by

$$Bf(x_1, x_2, \dots) = \int_{[0,1]} f(u, x_1, x_2, \dots) - f(x_1, x_2, x_3, \dots)\, du.$$

For a type vector $\bar{x} = (x_1, x_2, \dots)$, one can interpret x_1 as the most recently mutated site, x_2 as the second most recently mutated site and so on. This additional information about the temporal order of mutations, which is usually not present in real sequence data, is "factored out" afterwards by considering appropriate equivalence classes.

For a sufficiently "old" population, which can be assumed to be in equilibrium, it is an interesting question whether for each pair of types $\bar{x}, \bar{y}$ visible in the population,

$$\text{there exist indices } i, j, \text{ such that } x_{i+k} = y_{j+k} \text{ for each } k \in N. \tag{34}$$

The condition means that there is a most recent common ancestor for all the types. This question is a prototype of a question for which the evolution of a population should be studied *backwards* in time. We will come back to this in the Section 3, see page 17.

The infinitely-many sites model has an interesting combinatorial structure, see, e.g. [GT95] or [BB07], Section 2. For example, in practice, one frequently does not know which of the bases visible at a segregating site is the mutant. This can be handled by considering appropriate equivalence classes.

2.4 Lookdown

The famous lookdown construction of Donnelly and Kurtz (see [DK99]) provides a unified approach to all the limiting population models which we have discussed so far, providing a clever nested coupling of approximating generalised Moran models in such a way that the measure-valued limit process is recovered as the empirical distribution process of an exchangeable system of countably many particles. However, its full power will only become clear when we consider genealogies of samples and follow history backwards in time in the next section. We present here a version suitable for populations of fixed total size. The construction is very flexible and works for many scenarios, including (continuous-mass) branching processes.

Note that [DK99] call what follows the 'modified' lookdown construction, in order to distinguish it from the construction of the classical Fleming-Viot superprocess introduced by the same authors in [DK96]. Here we drop the prefix 'modified'.

Let $F \in \mathcal{M}_1([0,1])$. The lookdown-construction leading to an empirical process with generator (24) works as follows:

We consider a countably infinite system of individuals, each particle being identified by a level $j \in \mathbb{N}$. We equip the levels with types ξ_t^j, $j \in \mathbb{N}$ in some type space E (and we think of E being an element of $\{\{1,\dots,d\},[0,1],[0,1]^{\mathbb{N}}\}$ depending on our choice of mutation model). Initially, we require the types $\xi_0 = (\xi_0^j)_{j\in\mathbb{N}}$ to be an exchangeable random vector, so that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N} \delta_{\xi_0^j} = \mu,$$

for some finite measure μ on E. The point is that the construction will preserve exchangeability.

There are two "sets of ingredients" for the reproduction mechanism of these particles, one corresponding to the "finite variance" part $F(\{0\})$, and the other to the "extreme reproductive events" described by $F_0 = F - F(\{0\})\delta_0$. Restricted to the first N levels, the dynamics is that of a very particular permutation of the generalised Moran model described by (31), with the property that always that particle with the highest level is the next to die.

For the first part, let $\{L_{ij}(t)\}$, $1 \le i < j < \infty$ be independent Poisson processes with rate $F(\{0\})$. Intuitively, at jump times t of L_{ij}, the particle at level j "looks down" at level i and copies the type there, corresponding to a single birth event in a(n approximating) Moran model. Types on levels above j are shifted accordingly, in formulas

$$\xi_k(t) = \begin{cases} \xi_k(t-), & \text{if } k < j, \\ \xi_i(t-), & \text{if } k = j, \\ \xi_{k-1}(t-), & \text{if } k > j, \end{cases} \tag{35}$$

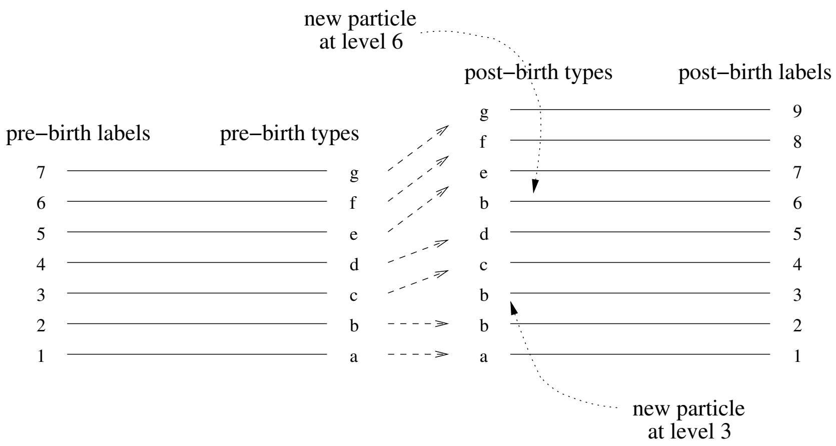

Figure 1: Relabelling after a birth event involving levels 2, 3 and 6.

if $\Delta L_{ij}(t) = 1$. This mechanism is well defined because for each k, there are only finitely many processes L_{ij}, $i < j \le k$ at whose jump times ξ_k has to be modified.

For the second part, which corresponds to multiple birth events, let $\mathcal{B}$ be Poisson point process on $\mathbb{R}^+ \times (0,1]$ with intensity measure $dt \otimes r^{-2} F_0(dr)$. Note that for almost all realisations $\{(t_i, y_i)\}$ of $\mathcal{B}$, we have

$$\sum_{i\,:\,t_i \le t} y_i^2 < \infty \quad \text{for all } t \ge 0. \tag{36}$$

The jump times t_i in our point configuration $\mathcal{B}$ correspond to reproduction events. Let U_{ij}, $i, j \in \mathbb{N}$, be i.i.d. uniform($[0,1]$). Define for $J \subset \{1, \ldots, l\}$ with $|J| \ge 2$,

$$L^l_J(t) := \sum_{i\,:\,t_i \le t} \prod_{j \in J} 1_{U_{ij} \le y_i} \prod_{j \in \{1,\ldots,l\} - J} 1_{U_{ij} > y_i}. \tag{37}$$

$L^l_J(t)$ counts how many times, among the levels in $\{1, \ldots, l\}$, exactly those in J were involved in a birth event up to time t. Note that for any configuration $\mathcal{B}$ satisfying (36), since $|J| \ge 2$, we have

$$\mathbb{E}\big[L^l_J(t) \,\big|\, \mathcal{B}\big] = \sum_{i\,:\,t_i \le t} y_i^{|J|}(1 - y_i)^{l-|J|} \le \sum_{i\,:\,t_i \le t} y_i^2 < \infty,$$

so that $L^l_J(t)$ is a.s. finite.

Intuitively, at a jump t_i, each level tosses a uniform coin, and all the levels j with $U_{ij} \le y_i$ participate in this birth event. Each participating level adopts the type of the smallest level involved. All the other individuals are shifted upwards accordingly, keeping their original order with respect to their levels (see Figure 1). More formally, if $t = t_i$ is a jump time and j is the smallest level involved, i.e. $U_{ij} \le y_i$ and $U_{ik} > y_i$ for $k < j$, we put

$$\xi^k_t = \begin{cases} \xi^k_{t-}, & \text{for } k \le j, \\ \xi^j_{t-}, & \text{for } k > j \text{ with } U_{ik} \le y_i, \\ \xi^{k-J^k_t}_{t-}, & \text{otherwise}, \end{cases} \tag{38}$$

where $J^k_{t_i} = \#\{m < k : U_{im} \le y_i\} - 1$.

So far, we have treated the reproductive mechanism of the particle system. We now turn our attention to the third ingredient, the mutation steps.

For a given mutation rate r and mutation operator B, as defined in (21), define for each level $\in \mathbb{N}$ an independent Poisson processes M_i with rate r, so that if process M_i jumps, and the current type at level i is x, then a new type is being chosen according to the kernel $q(x, \cdot)$. For a rigorous formulation, all three mechanisms together can be cast into a countable system of Poisson process-driven stochastic differential equations, see [DK99], Section 6.

Then ([DK99]), for each $t > 0$, $(\xi_t^1, \xi_t^2, \dots)$ is an exchangeable random vector, so that

$$Z_t = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \delta_{\xi_t^j} \tag{39}$$

exists almost surely by de Finetti's Theorem, and is the Markov process with generator (24) and initial condition $Z_0 = \mu$.

Remark. An alternative and very elegant way to encode the genealogy of a Fleming-Viot process with generator (18) is via a flow of bridges, as described in [BLG03]. However, unlike the situation for the lookdown construction, it seems unclear how to incorporate mutation in this approach. □

3 Neutral genealogies: beyond Kingman's coalescent

After having spent a considerable amount of pages on models for the evolution of the type distribution of a population forwards in time, we now turn to the fruitful approach of looking backwards in time by analysing the genealogies of samples drawn at present. An important advantage of this approach is that in a neutral situation, this allows one to think of a stochastic two-step procedure, first simulating a genealogy, and then independently superimposing the mutation events on the given genealogical tree. This point of view has many computational and conceptual advantages. We will see below how the lookdown construction, introduced in Section 2.4, provides a unified framework by simultaneously describing the forwards evolution and all the genealogical trees of the approximating particle systems.

3.1 Genealogies and coalescent processes

A way to describe the genealogy of a sample of size n from a (haploid) population is to introduce a family of partitions of $\{1, \dots, n\}$ as follows:

$$i \sim_t j \text{ iff } i \text{ and } j \text{ have the same ancestor time } t \text{ before present.} \tag{40}$$

Obviously, if $t \geq t'$, then $i \sim_{t'} j$ implies $i \sim_t j$, i.e. the ancestral partition becomes coarser as t increases.

For neutral population models of fixed population size in the domain of attraction of the classical Fleming-Viot process, such as the Wright-Fisher- and the Moran model, the (random) genealogy of a finite sample can be (approximately) described by the now classical Kingman-coalescent, which we introduce briefly, followed by the more recently discovered and much more general Λ-coalescents.

Kingman's coalescent. Let $\mathcal{P}_n$ be the set of partitions of $\{1,\ldots,n\}$ and let $\mathcal{P}$ denote the set of partitions of $\mathbb{N}$. For each $n \in \mathbb{N}$, Kingman [K82] introduced the so-called *n-coalescent*, which is a $\mathcal{P}_n$-valued continuous time Markov process $\{\Pi_t^{(n)}, t \geq 0\}$, such that $\Pi_0^{(n)}$ is the partition of $\{1,\ldots,n\}$ into singleton blocks, and then each pair of blocks merges at rate one. Given that there are b blocks at present, this means that the overall rate to see a merger between blocks is $\binom{b}{2}$. Note that only *binary mergers* are allowed. Kingman [K82] also showed that there exists a $\mathcal{P}$-valued Markov process $\{\Pi_t, t \geq 0\}$, which is now called the (standard) *Kingman-coalescent*, and whose restriction to the first n positive integers is the n-coalescent. To see this, note that the restriction of any n-coalescent to $\{1,\ldots,m\}$, where $1 \leq m \leq n$, is an m-coalescent. Hence the process can be constructed by an application of the standard extension theorem.

Λ-coalescents. Pitman [P99] and Sagitov [S99] introduced and discussed coalescents which allow *multiple collisions*, i.e. more than just two blocks may merge at a time. Again, such a coalescent with multiple collisions (which will be later called a Λ-*coalescent*) is a $\mathcal{P}$-valued Markov-process $\{\Pi_t, t \geq 0\}$, such that for each n, its restriction to the first n positive integers is a $\mathcal{P}_n$-valued Markov process (the "n-Λ-coalescent") with the following transition rates. Whenever there are b blocks in the partition at present, each k-tuple of blocks (where $2 \leq k \leq b \leq n$) is merging to form a single block at rate $\lambda_{b,k}$, and no other transitions are possible. The rates $\lambda_{b,k}$ do not depend on either n or on the structure of the blocks. Pitman showed that in order to be consistent, which means that for all $2 \leq k \leq b$,

$$\lambda_{b,k} = \lambda_{b+1,k} + \lambda_{b+1,k+1},$$

such transition rates must necessarily satisfy

$$\lambda_{b,k} = \int_0^1 x^k (1-x)^{b-k} \frac{1}{x^2} \Lambda(dx), \tag{41}$$

for some finite measure Λ on the unit interval. We exclude the (trivial) case $\Lambda = 0$. By a trivial time transformation, one can always assume that Λ is a probability measure. In [S99], the corresponding measure is termed $F\ (= \Lambda/\Lambda([0,1])$, and this is the F appearing throughout Section 2.2.

Note that (41) sets up a one-to-one correspondence between coalescents with multiple collisions and finite measures Λ. Indeed, it is easy to see that the $\lambda_{b,k}$ determine Λ by an application of Hausdorff's moment problem, which has a unique solution in this case.

Due to the restriction property, the Λ-coalescent on $\mathcal{P}$ (with rates obtained from the measure Λ as described above) can be constructed from the corresponding n-Λ-coalescents via extension.

Approximation of genealogies in finite population models. Consider a sample of size n from a (stationary) Cannings model of size $N \gg n$, without mutation, and define an ancestral relation process $\{R_k^{(N,n)} : k = 0, 1, \ldots\}$ via (40). Recalling that c_N, as defined in (3), is the probability for a randomly picked pair of individuals to have the same ancestor one generation ago, it seems reasonable to rescale time and define

$$\Pi_t^{(N,n)} := R_{\lfloor t/c_N \rfloor}^{(N,n)}, \quad t \geq 0, \tag{42}$$

as then (if $c_N \to 0$) for a sample of size two, the time to the most recent common ancestor is approximately exponentially distributed with rate 1.

Indeed, [S99] and [MS01] have shown that if $c_N \to 0$ and (4) holds true, then $\{\Pi_t^{(N,n)} : t \geq 0\}$ converges weakly to Kingman's n-coalescent, while (12) and (13) imply that the limit is a Λ-coalescent with transition rates given by (41), where $\Lambda = F$, with F from the right-hand side of (12).

Obviously, there is a close relation between multiple merger events in the genealogy of the sample and "extreme" reproductive events in the population, in which a non-negligible proportion, say $x \in (0, 1]$, of the population alive in the next generation goes back to a single ancestor in the current generation. In fact, the integrand in (41) can be interpreted as follows: When following b lineages backwards, in such an event, each of them flips a coin with success probability x and all the successful lineages subsequently merge.

On the other hand, although individuals *can* have more than two offspring, the moment condition (4) ensures that families are typically small compared to the total population size and thus implies that in the limit, only binary mergers are visible in the genealogy.

Remark (Simultaneous multiple collisions). It should be pointed out that Möhle and Sagitov [MS01] provide a complete classification of possible limits of genealogies in Cannings-models, in particular if the condition (13) is violated. In this case, the resulting genealogies contain *simultaneous* multiple collisions, which have been studied independently and termed "Ξ-coalescents" by Schweinsberg in [S00], in which several groups of lineages can merge at exactly the same time. Note that the first factor in (13) is the probability to observe two simultaneous mergers in one generation in a sample of size four, whereas the second factor is the inverse of the pair coalescence probability.

Since a corresponding theory of forward population models in the spirit of Section 2.2 is not yet completely established and our space is limited, we restrict ourselves here to the "Λ-world". □

Analytic Duality. Consider an F-generalised Fleming-Viot process $\{Z_t\}$ with generator (18) starting from $Z_0 = \mu \in \mathcal{M}_1(E)$. The idea that the type distribution in an n-sample from the population at time t can be obtained by "colouring" t-ancestral partitions independently according to Z_0 has the following explicit analytical incarnation: For bounded measurable $f : E^n \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}\Big[\int_E \cdots \int_E f_{\Pi_0}(a_1, \ldots, a_{|\Pi_0|})\, Z_t(da_1) \cdots Z_t(da_p)\Big]&\\ = \mathbb{E}\left[\int_E \cdots \int_E f_{\Pi_t}(b_1, ..., b_{|\Pi_t|})\, Z_0(db_1) \cdots Z_0(db_{|\Pi_t|})\right],& \qquad (43)\end{aligned}$$

where Π is the n-F-coalescent starting at $\pi_0 = \{\{1\}, ..., \{n\}\}$, and, for any partition $\pi = \{C_1, ..., C_q\}$ of $\{1, ..., n\}$,

$$f_\pi(b_1, ..., b_q) := f(a_1, ..., a_p)$$

with $a_i := b_k$ if $i \in C_k$. This is classical for the Kingman case, and has first been explicitly formulated in [BLG03] for the Λ-case. Note that specialising (43) in the case $F = \delta_0$ to a two-point space yields the well-known moment duality between the Wright-Fisher diffusion (5) and the block-counting process of Kingman's coalescent, which is a pure death process with death rate $\binom{n}{2}$.

Remarks ("Coming down from infinity"). 1. Not all Λ-coalescents seem to be reasonable as models for biological populations, since some do not allow for a finite "time to the most recent common ancestor" of the entire population (T_{MRCA}) in the sense of "coming down from infinity in finite time". The latter means that any initial partition in $\mathcal{P}$, and for all $\varepsilon > 0$, the partition Π_ε) a.s. consists of finitely many blocks only. Schweinsberg [S00] proves that if either Λ has an atom at 0 or Λ has no atom at zero and

$$\lambda^* := \sum_{b=2}^{\infty} \left(\sum_{k=2}^{b} (k-1) \binom{b}{k} \lambda_{b,k} \right)^{-1} < \infty, \tag{44}$$

where $\lambda_{b,k}$ is given by (41), then the corresponding coalescent does come down from infinity (and if so, the time to come down to only one block has finite expectation). For the corresponding generalised $(\Lambda/\Lambda([0,1]))$-Fleming-Viot process $\{Z_t\}$ without mutation, (44) means that the size of the support of Z_t becomes one in finite time – the process fixes on the type of the population's "eve".

2. An important example for a coalescent, which (only just) does not come down from infinity is the Bolthausen-Sznitman coalescent, where $\Lambda(dx) = dx$ is the uniform distribution on $[0,1]$. This is the Beta$(2-\alpha, \alpha)$-coalescent with $\alpha = 1$, and it plays an important role in statistical mechanics models for disordered systems (see e.g. [Bo06] for an introduction).

3. However, it should be observed that all n-Λ-coalescents (for finite n) do have an a.s. finite T_{MRCA}.

4. Note that by Kingman's theory of exchangeable partitions, for each $t > 0$, asymptotic frequencies of the classes exists. If a Λ-coalescent does *not* come down from infinity, it may or may not be the case that these frequencies sum to one ("proper frequencies"). [P99] showed that the latter holds iff $\int_{0+} r^{-1}\Lambda(dr) = \infty$. Note that if $\int_{[0,1]} r^{-1}\,\Lambda(dr) < \infty$, we see from (33) that $\lim_{N\to\infty} d_N < \infty$. Hence in the lookdown construction, at each time $t \geq 0$ there is a positive fraction of levels which have not yet participated in any lookdown event. These correspond to "dust". □

Examples for coalescents which satisfy (44) are Kingman's coalescent, the process considered in [EW06], corresponding to (16), (but note that [EW06] also considers $F = \delta_\psi$ with $\psi \in (0,1)$, for which (44) fails), and the so-called Beta$(2-\alpha, \alpha)$-coalescents with $\alpha \in (1,2)$, with $\Lambda = F$ given by (17). Note that even though (17) makes no sense for $\alpha = 2$, Kingman's coalescent can be included in this family as the weak limit Beta$(2-\alpha, \alpha) \to \delta_0$ as $\alpha \to 2$).

Coalescents and the modified lookdown construction. We now make use of the explicit description of the modified construction to determine the coalescent process embedded in it. Fix a (probability) measure F on $[0,1]$. Recall the Poisson processes L_{ij} and L_K^l from (37) in Section 2.4 above. For each $t \geq 0$ and $k = 1, 2, \ldots$, let $N_k^t(s), 0 \leq s \leq t$, be the level at time s of the ancestor of the individual at level k at time t. In terms of the L_K^l and L_{ij}, the process $N_k^t(\cdot)$ solves, for

$0 \le s \le t$,

$$
\begin{aligned}
N_k^t(s) = k &- \sum_{1 \le i < j < k} \int_{s-}^{t} 1_{\{N_k^t(u) > j\}} \, dL_{ij}(u) \\
&- \sum_{1 \le i < j < k} \int_{s-}^{t} (j-i) 1_{\{N_k^t(u) = j\}} \, dL_{ij}(u) \\
&- \sum_{K \subset \{1,\ldots,k\}} \int_{s-}^{t} (N_k^t(u) - \min(K)) 1_{\{N_k^t(u) \in K\}} \, dL_K^k(u) \\
&- \sum_{K \subset \{1,\ldots,k\}} \int_{s-}^{t} (|K \cap \{1,\ldots,N_k^t(u)\}| - 1) \\
&\qquad \times 1_{\{N_k^t(u) > \min(K),\, N_k^t(u) \notin K\}} \, dL_K^k(u), \qquad (45)
\end{aligned}
$$

Fix $0 \le T$ and, for $t \le T$, define a partition $\Pi^T(t)$ of $\mathbb{N}$ such that k and l are in the same block of $\Pi^T(t)$ if and only if $N_k^T(T-t) = N_l^T(T-t)$. Thus, k and l are in the same block if and only if the two levels k and l at time T have the same ancestor at time $T-t$. Then ([DK99], Section 5),

$$\text{the process } \left\{\Pi_t^T : 0 \le t \le T\right\} \text{ is an } F\text{-coalescent run for time } T.$$

Note that by employing a natural generalisation of the lookdown construction using driving Poisson processes on $\mathbb{R}$ and e.g. using $T = 0$ above, one can use the same construction to find an F-coalescent with time set $\mathbb{R}_+$. We would like to emphasise that in contrast to (43), which can be proved using the reasoning above, the lookdown construction provides a realisation-wise coupling of the type distribution process $\{Z_t\}$ and the coalescent describing the genealogy of a sample.

Superimposing mutations. Consider now an F-generalised Fleming-Viot process $\{Z_t\}$ with "individual" mutation operator rB, described by the generator $\mathcal{L}_{B,F}$ given by (24), starting from $Z_0 = \mu$. The lookdown construction easily allows to prove that for each t, the distribution of a sample of size n from Z_t can be equivalently described as follows: Run an n-F-coalescent for time t, interpret this as a forest with labelled leaves. "Colour" each root independently according to μ, then run the Markov process with generator rB independently along the branches of each tree, and finally read off the types at the leaves.

Remark. If (44) is fulfilled and the individual mutation process with generator B has a unique equilibrium, one can let $t \to \infty$ in the above argument to see that $\{Z_t\}$ has a unique equilibrium, and the distribution of an n-sample from this equilibrium can be obtained by running an n-F-coalescent until it hits the trivial partition. Then colour this most recent common ancestor randomly according to the stationary distribution of B, and run the mutation process along the branches as above.

This approach is very fruitful in population genetics applications. For example, under condition (44), (34) will be satisfied for t large enough, irrespective of the initial condition.

4 Population genetic inference

Populations with extreme reproductive behaviour. Recently, biologists have studied the genetic variation of certain marine species with rather extreme reproductive behaviour, see, e.g., Árnason [A04] (Atlantic Cod) and [BBB94] (Pacific Oyster). In this situation, one would like to decide which coalescent is suitable, based upon observed genetic types in a sample from the population.

Eldon and Wakeley [EW06] analysed the sample described in [BBB94] and proposed a one-parameter family of Λ-coalescents, which comprises Kingman's coalescent as a boundary case, namely those described by (16), as models for their genealogy. Inference is then based on a simple *summary statistic*, the number of *segregating sites* and *singleton polymorphisms*. They conclude that ([EW06], p. 2622):

> *For many species, the coalescent with multiple mergers might be a better null model than Kingman's coalescent.*

In this section, we obtain recursions for the type probabilities of an n-sample from a general Λ-coalescent under with a general finite alleles model. We present two approaches, one based on the lookdown construction, the other on direct manipulations with the generator $\mathcal{L}_{B,F}$. We discuss how this recursion can then be used to derive a Monte-Carlo scheme to compute likelihoods of model parameters in Λ-coalescent scenarios given the observed types, in the spirit of [GT94b], see also [BB07] for the infinite-sites case. These can be used e.g. for maximum likelihood estimation.

Remark. Analogous recursions for the probability of configurations in the infinite-alleles model have been obtained in [M06b]. Exact asymptotic expressions for certain summary statistics for the infinite-alleles and infinite-sites models under Beta-coalescents [recall (17)] have been obtained in [BBS06]. □

4.1 Finite-alleles recursion I: Using the lookdown construction

Recall that in the finite alleles model, type changes, or mutations, occurr at rate r, and $P = (P_{ij})$ is an irreducible stochastic transition matrix on the finite type space E. Note that silent mutations are allowed (i.e. $P_{jj} \geq 0$), denote the unique equilibrium of P by μ. We assume that the reproduction mechanism is described by some $F = \Lambda \in \mathcal{M}_1([0,1])$.

Suppose the system, described by the lookdown construction, is in equilibrium. Consider the first n levels at time 0 and let τ_{-1} be the last instant before 0 when at least one of the types at levels $1,\ldots,n$ changes. Then, $-\tau_{-1}$ is exponentially distributed with rate

$$r_n = nr + \sum_{k=2}^{n} \binom{n}{k} \lambda_{n,k}. \tag{46}$$

Denote by p the distribution of the types of the first n levels in the stationary lookdown construction, say, at time 0. Later, due to exchangeability, we will merely be interested in the type frequency probability $p^0(\mathbf{n})$. Decomposing according to which event occurred at time τ_{-1}, we obtain

$$\begin{aligned} p\big((y_1,\ldots,y_n)\big) = {} & \frac{r}{r_n}\sum_{i=1}^{n}\sum_{z\in E} p\big((y_1,\ldots,y_{i-1},z,y_{i+1},\ldots,y_n)\big)P_{zy_i} \\ & + \frac{1}{r_n}\sum_{\substack{K\subset\{1,\ldots,n\}\\ |K|\geq 2}} \lambda_{n,|K|}1_{\{\text{all } y_j \text{ equal for } j\,\in\,K\}}\,p\big(\gamma_K(y_1,\ldots,y_n)\big), \end{aligned} \tag{47}$$

where $\gamma_K(y_1,\dots,y_n)\in E^{n-|K|+1}$ is that vector of types of length $n-|K|+1$ which $\big(\xi_1(\tau_{-1}-),\dots,\xi_{n-|K|+1}(\tau_{-1}-)\big)$ must be in order that a resampling event involving exactly the levels in K among levels $1,\dots,n$ generates $\big(\xi_1(\tau_{-1}),\dots,\xi_n(\tau_{-1})\big)=(y_1,\dots,y_n)$. Formally,

$$\gamma_K(y_1,\dots,y_n)_i = y_{i+\#((K\setminus\{\min K\})\cap\{1,\dots,i\})},\ 1\le i\le n-|K|+1.$$

As the type at level 1 is the stationary Markov process with generator rB, we have the boundary condition $p\big((y_1)\big)=\mu(y_1),\ y_1\in E$. Note that, by exchangeability,

$$p\big((y_1,\dots,y_n)\big)=p\big((y_{\pi(1)},\dots,y_{\pi(n)})\big)$$

for any permutation π of $\{1,\dots,n\}$. So, the only relevant information is (of course) how many samples were of which type. For $\mathbf{n}=(n_1,\dots,n_d)\in\mathbb{Z}_+^d$ we write $\#\mathbf{n}:=n_1+\cdots+n_d$ for the 'length', and

$$\kappa(\mathbf{n})=\big(\underbrace{1,1,\dots,1}_{n_1},\underbrace{2,\dots,2}_{n_2},\dots,\underbrace{d,\dots,d}_{n_d}\big)\in E^{\#\mathbf{n}}$$

for a 'canonical representative' of the (absolute) type frequency vector $\mathbf{n}$. Let

$$p^0(\mathbf{n}):=\binom{\#\mathbf{n}}{n_1,n_2,\dots,n_d}p\big(\kappa(\mathbf{n})\big) \tag{48}$$

be the probability that in a sample of size $\#\mathbf{n}$, there are exactly n_j of type j, $j=1,\dots,d$. We abbreviate $n:=\#\mathbf{n}$, and write e_k for the k-th canonical unit vector of $\mathbb{Z}^d$. Noting that

$$n_j\binom{\#\mathbf{n}}{n_1,n_2,\dots,n_d}p(\mathbf{n}-\mathbf{e}_j+\mathbf{e}_i)=\big(n_i+1-\delta_{ij}\big)p^0(\mathbf{n}-\mathbf{e}_j+\mathbf{e}_i)$$

and that (for $n_j\ge k$, otherwise the term is 0)

$$\binom{n_j}{k}\binom{\#\mathbf{n}}{n_1,n_2,\dots,n_d}p(\mathbf{n}-(k-1)\mathbf{e}_j)=\binom{n}{k}\frac{n_j-k+1}{n-k+1}p^0(\mathbf{n}-(k-1)\mathbf{e}_j),$$

(47) translates into the following recursion for p^0:

$$\begin{aligned} p^0(\mathbf{n}) &= \frac{r}{r_n}\sum_{j=1}^d\sum_{i=1}^d(n_i+1-\delta_{ij})P_{ij}p^0(\mathbf{n}-\mathbf{e}_j+\mathbf{e}_i)\\ &\quad+\frac{1}{r_n}\sum_{\substack{j=1\\ n_j\ge 2}}^d\sum_{k=2}^{n_j}\binom{n}{k}\lambda_{n,k}\frac{n_j-k+1}{n-k+1}p^0(\mathbf{n}-(k-1)\mathbf{e}_j) \end{aligned}\tag{49}$$

with boundary conditions $p^0(\mathbf{e}_j)=\mu_j$.

Remark. In the Kingman-case, we have $\lambda_{n,k}=1(n\ge 2=k)$, $r_n=n\theta/2+n(n-1)/2=n(n-1+\theta)/2$ (and we assume $r=\theta/2$ as "usual"), hence (49) becomes the well-known

$$\begin{aligned} p^0(\mathbf{n}) &= \frac{\theta}{n-1+\theta}\sum_{j=1}^d\sum_{i=1}^d\frac{n_i+1-\delta_{ij}}{n}P_{ij}p^0(\mathbf{n}-\mathbf{e}_j+\mathbf{e}_i)\\ &\quad+\frac{n-1}{n-1+\theta}\sum_{\substack{j=1\\ n_j\ge 2}}^d\frac{n_j-1}{n-1}p^0(\mathbf{n}-\mathbf{e}_j). \end{aligned}\tag{50}$$

4.2 Finite-alleles recursion II: Generator approach

An alternative method to obtain the recursion for the type probabilities in the finite-alleles case is by using a generator approach, see [DIG04a]. Let $f \in C_2$ and $\Delta_d = \{(x_1, \dots, x_d) : x_i \geq 0, x_1 + \dots + x_d = 1\}$ and consider the mutation operator

$$\widetilde{B}f(x_1, \dots, x_d) = r \sum_{i=1}^{d} \Big(\sum_{j=1}^{d} x_j P_{ji} - x_i P_{ij} \Big) \frac{\partial f}{\partial x_i}(x_1, \dots, x_d).$$

For the resampling operator, we distinguish the Kingman- and non-Kingman components. First, assume $\Lambda(\{0\}) = 0$ (non-Kingman). Consider

$$\begin{aligned} R_1 f(x_1, \dots, x_d) = \sum_{i=1}^{d} \int x_i \Big(& f(\bar{r}x_1, \dots, \bar{r}x_{i-1}, \bar{r}x_i + r, \bar{r}x_{i+1}, \dots, \bar{r}x_d) \\ & - f(x_1, \dots, x_d) \Big) r^{-2} \Lambda(dr), \end{aligned} \tag{51}$$

where $\bar{r} = 1 - r$. For the Kingman-part ($\Lambda = \delta_0$) of the resampling operator, we have

$$R_2 f(x_1, \dots, x_d) = \frac{1}{2} \sum_{i,j=1}^{d} x_i(\delta_{ij} - x_j) \frac{\partial^2 f}{\partial x_i \partial x_j}(x_1, \dots, x_d).$$

Finally, for general Λ and $a \geq 0$, write $R = R_1 + aR_2$, where R_1 uses $\Lambda_0 = \Lambda - \Lambda(\{0\}\delta_0$. Now, let $X(t) = (X_1(t), \dots, X_d(t))$ be the stationary process with generator $L = \widetilde{B} + R$ [note that $X_i(t) = Z_t(\{i\})$, where $\{Z_t\}$ is the stationary process with generator (24)]. Write $X = X(0)$. Let $\mathbf{n} = (n_1, \dots, n_d)$, $n = n_1 + \dots + n_d$. Then,

$$\mathbb{E}\Big[\prod_{i=1}^{d} X_i^{n_i} \Big]$$

is the probability of observing in a sample of size n from the equilibrium population type i precisely n_i times in a particular order (e.g. first n_1 samples of type 1, next n_2 samples of type 2, etc.). Put

$$f_{\mathbf{n}}(\mathbf{x}) := \mathbf{x}^{\mathbf{n}} := \prod_{i=1}^{d} x_i^{n_i}.$$

Then,

$$g(\mathbf{n}) := \binom{n}{n_1 \dots n_d} \mathbb{E}\big[f_{\mathbf{n}}(X)\big]$$

is the probability of observing type i exactly n_i times, $i = 1, \dots, d$, without regard of the order. Note that

$$\begin{aligned} \widetilde{B} f_{\mathbf{n}}(x_1, \dots, x_d) &= r \sum_{i=1}^{d} \Big(\sum_{j=1}^{d} x_j P_{ji} - x_i P_{ij} \Big) n_i f_{\mathbf{n}-\mathbf{e}_i}(x_1, \dots, x_d) \\ &= r \sum_{i,j=1}^{d} n_i P_{ji} f_{\mathbf{n}-\mathbf{e}_i+\mathbf{e}_j}(\mathbf{x}) - rn f_{\mathbf{n}}(\mathbf{x}) \end{aligned}$$

and

$$
\begin{aligned}
f_{\mathbf{n}}((1-r)\mathbf{x}+r\mathbf{e_i}) &= (1-r)^{n-n_i}\prod_{j\neq i}^{d} x_j^{n_j} \times \big((1-r)x_i+r\big)^{n_i} \\
&= (1-r)^{n-n_i}\prod_{j\neq i}^{d} x_j^{n_j} \times \sum_{k=0}^{n_i}\binom{n_i}{k} r^k(1-r)^{n_i-k}x_i^{n_i-k} \\
&= \sum_{k=0}^{n_i}\binom{n_i}{k} r^k(1-r)^{n-k}\Big(x_i^{n_i-k}\prod_{j\neq i}^{d} x_j^{n_j}\Big),
\end{aligned}
$$

so the term inside the integral in the expression (51) for R_1 can be written as

$$
\begin{aligned}
&\sum_{i=1}^{d}\sum_{k=0}^{n_i}\binom{n_i}{k} r^k(1-r)^{n-k}x_i^{n_i-k+1}\prod_{j\neq i}^{d} x_j^{n_j} - \sum_{k=0}^{n}\binom{n}{k} r^k(1-r)^{n-k}\prod_{\ell=1}^{d} x_\ell^{n_\ell} \\
&= \sum_{i:n_i\geq 2}^{d}\sum_{k=2}^{n_i}\binom{n_i}{k} r^k(1-r)^{n-k}x_i^{n_i-k+1}\prod_{j\neq i}^{d} x_j^{n_j} - \sum_{k=2}^{n}\binom{n}{k} r^k(1-r)^{n-k}\prod_{\ell=1}^{d} x_\ell^{n_\ell},
\end{aligned}
$$

observing that the terms with $k=0$ and $k=1$ cancel since $x_1+\cdots+x_d=1$ and $n_1+\cdots+n_d=n$. Recalling the definition of $\lambda_{n,k}$ from (41), we obtain

$$
R_1 f_{\mathbf{n}}(\mathbf{x}) = \sum_{i:n_i\geq 2}^{d}\sum_{k=2}^{n_i}\binom{n_i}{k}\lambda_{n,k} f_{\mathbf{n}-(k-1)\mathbf{e}_i}(\mathbf{x}) - \sum_{k=2}^{n}\binom{n}{k}\lambda_{n,k} f_{\mathbf{n}}(\mathbf{x}). \tag{52}
$$

Furthermore

$$
\begin{aligned}
R_2 f_{\mathbf{n}}(\mathbf{x}) &= \frac{1}{2}\sum_{i,j=1}^{d} x_i(\delta_{ij}-x_j)n_i(n_j-\delta_{ij})f_{\mathbf{n}-\mathbf{e}_i-\mathbf{e}_j}(\mathbf{x}) \\
&= \sum_{i=1}^{d}\frac{n_i(n_i-1)}{2}f_{\mathbf{n}-\mathbf{e}_i}(\mathbf{x}) - \sum_{i,j=1}^{d}\frac{n_i(n_j-\delta_{ij})}{2}f_{\mathbf{n}}(\mathbf{x}) \\
&= \sum_{i=1}^{d}\frac{n_i(n_i-1)}{2}f_{\mathbf{n}-\mathbf{e}_i}(\mathbf{x}) - \frac{n(n-1)}{2}f_{\mathbf{n}}(\mathbf{x}).
\end{aligned} \tag{53}
$$

Combining the terms from R_1 and R_2 (using (52) and (53) above, and replacing Λ by Λ_0 in (51)), we have

$$
R f_{\mathbf{n}}(\mathbf{x}) = \sum_{i:n_i\geq 2}^{d}\sum_{k=2}^{n_i}\binom{n_i}{k}\lambda_{n,k} f_{\mathbf{n}-(k-1)\mathbf{e}_i}(\mathbf{x}) - \sum_{k=2}^{n}\binom{n}{k}\lambda_{n,k} f_{\mathbf{n}}(\mathbf{x}).
$$

Thus we obtain from the stationarity condition $\mathbb{E}\, L f_{\mathbf{n}}(X)=0$ that

$$
r_n\mathbb{E}f_{\mathbf{n}}(X) = r\sum_{i,j=1}^{d} n_i P_{ji}\mathbb{E}f_{\mathbf{n}-\mathbf{e}_i+\mathbf{e}_j}(X) + \sum_{i:n_i\geq 2}^{d}\sum_{k=2}^{n_i}\binom{n_i}{k}\lambda_{n,k}\mathbb{E}f_{\mathbf{n}-(k-1)\mathbf{e}_i}(X),
$$

where r_n is defined in (46). Multiplying with $\binom{n}{n_1\ldots n_d}/r_n$ and some algebra gives

$$
\begin{aligned}
g(\mathbf{n}) &= \frac{r}{r_n}\sum_{i,j=1}^{d}(n_j+1-\delta_{ij})P_{ji}g(\mathbf{n}-\mathbf{e}_i+\mathbf{e}_j) \\
&\quad + \frac{1}{r_n}\sum_{i:n_i\geq 2}^{d}\sum_{k=2}^{n_i}\binom{n}{k}\lambda_{n,k}\frac{n_i-k+1}{n-k+1}\,g(\mathbf{n}-(k-1)\mathbf{e}_i),
\end{aligned}
$$

which agrees with (49).

4.3 A Monte Carlo Scheme for sampling probabilities

Recursion (49) can be used to estimate $p^0(\mathbf{n})$ for a given $\mathbf{n} \in \mathbb{Z}_+^d$ using a Markov chain, in the spirit of [GT94b], as follows:

Let $\{X_k\}$ be a Markov chain on $\mathbb{Z}_+^d$ with transitions

$$\mathbf{n} \to \begin{cases} \mathbf{n} - \mathbf{e}_j + \mathbf{e}_i & \text{w. p.} \quad \frac{r}{r_n f(\mathbf{n})}(n_i + 1 - \delta_{ij})P_{ij} \quad \text{if} \quad n_j > 0, \\ \mathbf{n} - (k-1)\mathbf{e}_i & \text{w. p.} \quad \frac{1}{r_n f(\mathbf{t},\mathbf{n})}\binom{n}{k}\lambda_{n,k}\frac{n_i - k + 1}{n-k+1} \quad \text{if} \quad 2 \le k \le n_i, \end{cases}$$

where [with r_n defined in (46)]

$$f(\mathbf{n}) = \frac{1}{r_n}\left(\sum_{\substack{i,j=1 \\ n_j > 0}}^{d} r(n_i + 1 - \delta_{ij})P_{ij} + \sum_{\substack{1 \le i \le d \\ n_i \ge 2}} \sum_{k=2}^{n_i} \binom{n}{k}\lambda_{n,k}\frac{n_i - k + 1}{n - k + 1} \right). \tag{54}$$

Then,

$$p^0(\mathbf{n}) = \mathbb{E}_{(\mathbf{n})} \prod_{l=0}^{\tau} f(\mathbf{t}(l), \mathbf{n}(l)). \tag{55}$$

Remark (Inference for Kingman's coalescent). Likelihood-based inference methods for Kingman's coalescent, some solving recursion (50) approximately via Monte Carlo methods, others using MCMC, have been developed since the beginning of the 1990ies, see [EG87], [GT94a], [GT94b], [GT94c], [GT96a], [GT96b], [GT97], [FKY99], [DIG04a], [SD00]. In [SD00], Stephens and Donnelly provide proposal distributions for importance sampling, which are optimal in some sense, and compare them to various other methods. Their importance sampling scheme seems, at present, to be the most efficient tool for inference for relatively large datasets, but heavily uses the fact that Kingman's coalescent allows only binary mergers. It is at present unclear what an analogous strategy in the general Λ-case ought to be. □

4.4 Simulating samples

Let $E, (P_{ij}), \mu, r$ be the parameters of a finite-alleles model. Then, one may obtain the type configuration in an n-sample as follows:

Let $\{Y_t^{(n)}\}_{t \ge 0}$ be the *block counting process* corresponding to an n-Λ-coalescent, i.e. $Y_t^{(n)} = \#\{\text{blocks of } \Pi_t\}$ is a continuous-time Markov chain on $\mathbb{N}$ with jump rates

$$q_{ij} = \binom{i}{i-j+1}\lambda_{i,i-j+1}, \quad i > j \ge 1$$

starting from $Y_0^{(n)} = n$. Its Green function is

$$g(n,m) := \mathbb{E}\left[\int_0^\infty \mathbf{1}_{\{Y_s^{(n)} = m\}} \, ds \right] \quad \text{for } n \ge m \ge 2, \tag{56}$$

which can easily be computed recursively, see [BB07], Section 7.1. Denoting by $\tau := \inf\{t : Y_t^{(n)} = 1\}$ be the time required to come down to only one class and by

∂ a "cemetery state", it follows from Nagasawa's Formula [see, e.g., [RW87], (42.4)] that the time-reversed path

$$\tilde{Y}_t^{(n)} := \begin{cases} Y_{(\tau-t)-}^{(n)}, & 0 \le t < \tau, \\ \partial, & \tau \le t, \end{cases} \tag{57}$$

is a continuous-time Markov chain on $\{2,\dots,n\}\cup\{\partial\}$ with jump rate matrix

$$\tilde{q}_{ji}^{(n)} = \frac{g(n,i)}{g(n,j)} q_{ij},\ j < i \le n, \quad -\tilde{q}_{jj}^{(n)} = \sum_{i=j+1} \tilde{q}_{ji}^{(n)} = \sum_{\ell=1}^{j-1} q_{j\ell}, \quad \tilde{q}_{n\partial}^{(n)} = -q_{nn}$$

and initial distribution $\mathbb{P}\{\tilde{Y}_0^{(n)} = k\} = g(n,k)q_{k1},\ k = 2,3,\dots,n$. Note that unless Λ is concentrated on $\{0\}$, the dynamics does depend on n. We write $\tilde{p}_{ji}^{(n)} := \tilde{q}_{ji}^{(n)}/(-\tilde{q}_{jj}^{(n)}),\ j < i \le n$ for the transition matrix of the skeleton chain of $Y^{(n)}$.

In view of the remark on page 17, it is clear that he following algorithm generates an n-sample from the stationary distribution of the process with generator $\mathcal{L}_{B,F}$ given by (24):

Algorithm (generating samples).

(i) Generate K with $\mathbb{P}\{K = k\} = g(n,k)q_{k1}$, $k = 2,\dots,n$, begin with $\eta = K\delta_X$, where $X \sim \mu$.

(ii) Draw $U \sim \mathrm{Unif}([0,1])$.

If $U \le \frac{kr}{kr+(-\tilde{q}_{kk}^{(n)})}$:

Replace one of the present types by a P-step from it, i.e. replace $\eta := \eta - \delta_x + \delta_y$ with probability $\frac{\eta_x}{\#\eta} P_{xy}$ (for $x \neq y$), where $\#\eta$ is the total mass of η.

Otherwise:

If $\#\eta = n$: Output η and stop.

Else, pick $J \in \{\#\eta,\dots,n\}$ with $\mathbb{P}\{J = j\} = \tilde{p}_{\#\eta,j}^{(n)}$. Choose one of the present types (according to their present frequency), and add $J - \#\eta$ copies of this type, i.e. replace $\eta := \eta + (J - \#\eta)\delta_x$ with probability $\frac{\eta_x}{\#\eta}$.

(iii) Repeat (ii).

Remark. Ordered samples can be obtained from a realization of η by random reordering. In the case of parent-independent mutation, i.e. if $P_{ij} = P_j$ for all i,j, it is possible to simplify the procedure by simulating "backwards in time". "Active"ancestral lineages are lost either by (possibly multiple) coalescence or when hitting their "defining" mutation, in which case one simply assigns a random type drawn accroding to P_j. □

References

[A04] ÁRNASON, E.: Mitochondrial Cytochrome b DNA Variation in the High-Fecundity Atlantic Cod: Trans-Atlantic Clines and Shallow Gene Genealogy, *Genetics* **166**, 1871–1885 (2004).

[BBS05] BERESTYCKI, N.; BERESTYCKI, J.; SCHWEINSBERG, J.: Small time behaviour of Beta-coalescents. Preprint, (2005).

[BBS06] BERESTYCKI, N.; BERESTYCKI, J.; SCHWEINSBERG, J.: Beta-coalescents and continuous stable random trees. *to appear in: Ann. Probab.*, (2006).

[BLG03] BERTOIN, J.; LE GALL, J.-F.: Bertoin, J.; Le Gall, J.-F.: Stochastic flows associated to coalescent processes. Probab. Theory Related Fields 126 (2003), no. 2, 261–288.

[BBC05] BIRKNER, M.; BLATH, J.; CAPALDO, M.; ETHERIDGE, A.; MÖHLE, M.; SCHWEINSBERG, J.; WAKOLBINGER, A.: Alpha-stable branching and Beta-coalescents. *Electron. J. Probab.* **10**, 303–325, (2005).

[BB07] BIRKNER, M; BLATH, J: Computing likelihoods for coalescents with multiple collisions in the infinitely-many-sites model, WIAS preprint 1237 (2007).

[BBB94] BOOM, J. D. G.; BOULDING, E. G.; BECKENBACH, A. T.: Mitochondrial DNA variation in introduced populations of Pacific oyster, Crassostrea gigas, in British Columbia. *Can. J. Fish. Aquat. Sci.* 51:16081614, (1994).

[Bo06] BOVIER, A.: *Statistical mechanics of disordered system. A mathematical perspective.* Cambridge University Press, (2006).

[C74] CANNINGS, C.: The latent roots of certain Markov chains arising in genetics: a new approach, I. Haploid models. *Adv. Appl. Prob.* **6**, 260–290, (1974).

[C75] CANNINGS, C.: The latent roots of certain Markov chains arising in genetics: a new approach, II. Further haploid models. *Adv. Appl. Prob.* **7**, 264–282, (1975).

[D93] DAWSON, D.: Lecture Notes, Ecole d'Eté de Probabilités de Saint-Flour XXI, Berlin, Springer, (1993).

[DIG04a] DE IORIO, M. AND GRIFFITHS, R. C.: Importance sampling on coalescent histories I. *Adv. Appl. Prob.* **36**, 417-433, (2004).

[DIG04b] DE IORIO, M. AND GRIFFITHS, R. C.: Importance sampling on coalescent histories II: Subdivided population models. *Adv. Appl. Prob.* **36**, 434-454, (2004).

[DK96] DONNELLY, P.; KURTZ, T.: A countable representation of the Fleming-Viot measure-valued diffusion. Ann. Probab. 24, no. 2, 698–742, (1996)

[DK99] DONNELLY, P.; KURTZ, T.: Particle representations for measure-valued population models. Ann. Probab. 27, no. 1, 166–205, (1999)

[DS05] DURRETT, R.; SCHWEINSBERG, J.: A coalescent model for the effect of advantageous mutations on the genealogy of a population. *Stoch. Proc. Appl.* **115**, 1628–1657 (2005)

[EW06] ELDON, B.; WAKELEY, J.: Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **172**, 2621–2633, (2006).

[EPW07] ETHERIDGE, A.; PFAFFELHUBER, P.; WAKOLBINGER, A.: Muller's ratchet: An intriguing problem in mathematical population genetics. To appear in *Trends in Stochastic Analysis*, Cambridge University Press, 2007.

[EG87] ETHIER, S.; GRIFFITHS, R.C.: The infinitely-many-sites model as a measure-valued diffusion. *Ann. Probab.* **15**, no. 2, 515–545, (1987).

[EK86] ETHIER, S.; KURTZ, T.: Markov Processes: Characterization and Convergence. Wiley, (1986).

[EK93] ETHIER, S.; KURTZ, T.: Fleming-Viot processes in population genetics, SIAM J. Control Optim. 31, no. 2, 345–386, (1993).

[F22] FISHER, R.A.: On the dominance ratio. *Proc. Roy. Soc. Edin.*, **42**:321–431, (1922).

[FKY99] FELSENSTEIN, J., KUHNER, M. K., YAMATO, J. AND BEERLI, P.: Likelihoods on coalescents: a Monte Carlo sampling approach to inferring parameters from population samples of molecular data. IMS Lecture Notes - Monograph Series, **33**, 163–185, (1999).

[GT94a] GRIFFITHS, R. C. AND TAVARÉ, S.: Simulating probability distributions in the coalescent. *Theor. Pop. Biol.* **46**, 131–159, (1994).

[GT94b] GRIFFITHS, R. C. AND TAVARÉ, S.: Ancestral Inference in population genetics. *Statistical Science* **9**, 307–319, (1994).

[GT94c] GRIFFITHS, R. C. AND TAVARÉ, S.: Sampling theory for neutral alleles in a varying environment. *Philosophical Transactions of the Royal Society London*, Series B, **344**, 403–410, (1994).

[GT95] GRIFFITHS, R. C. AND TAVARÉ, S.: Unrooted genealogical tree probabilities in the infinitely-many-sites model. *Mathematical Biosciences* **127**, 77-98 (1995).

[GT96a] GRIFFITHS, R. C. AND TAVARÉ, S.: Monte Carlo inference methods in population genetics. Monte Carlo and quasi-Monte Carlo methods. *Math. Comput. Modeling* **23** (1996), no. 8-9, 141–158.

[GT96b] GRIFFITHS, R. C. AND TAVARÉ, S.: Markov chain inference methods in population genetics. *Math. Comput. Modeling* **23**, 8/9, 141-158, (1996).

[GT96c] GRIFFITHS, R. C. AND TAVARÉ, S.: Markov chain inference methods in population genetics. *Math. Comput. Modeling* **23**, 8/9, 141-158.

[GT97] GRIFFITHS, R. C. AND TAVARÉ, S.: Computational Methods for the coalescent. Progress in Population Genetics and Human Evolution, 165–182, Springer, (1997).

[GT98] GRIFFITHS, R. C.; TAVARÉ, S.: The age of a mutation in a general coalescent tree. *Comm. Statist. Stochastic Models* **14**, 273–29, (1998).

[GT99] GRIFFITHS, R. C.; TAVARÉ, S.: The ages of mutations in gene trees. *Ann. Appl. Probab.* **9**, no. 3, 567–590, (1999).

[H90] HUDSON, R.R.: Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*, **7**, 1–44, (1990).

[HSW05] HEIN, J.; SCHIERUP, M. H.; WIUF, C.: *Gene Genealogies, Variation and Evolution – A Primer in Coalescent Theory.* Oxford University Press, 2005.

[K82] KINGMAN, J. F. C.: The coalescent. *Stoch. Proc. Appl.* **13**, 235–248, (1982).

[KYF95] KUHNER, M. K., YAMATO, J. AND FELSENSTEIN, J.: Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics* **140**, 1421–1430, (1995).

[KYF98] KUHNER, M. K., YAMATO, J. AND FELSENSTEIN, J.: Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**, 429–434, (1998).

[M05] MÖHLE, M.: Simulation algorithms for integrals of a class of sampling distributions arising in population genetics. *J. Stat. Comp. Simul.* **75**, 731-749 (2005)

[M06a] MÖHLE, M.: On the number of segregating sites for populations with large family sizes. *J. Appl. Prob.* “in press”, (2006).

[M06b] MÖHLE, M.: On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12**, “in press”, (2006).

[M07] MÖHLE, M.: On a class of non-regenerative sampling distributions. *To appear in: Combin. Probab. Comput.* **16**, (2007)

[MS01] MÖHLE, M.; SAGITOV, S.: A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, 1547–1562, (2001).

[M07b] MÖRTERS, P.: Why study multifractal spectra? To appear in *Trends in Stochastic Analysis*, Cambridge University Press, 2007.

[N01] NORDBORG, M.: Coalescent Theory. In Balding, Bishop, and Cannings, eds. *Handbook of Statistical genetics,* 179–208. Wiley, (2001).

[P99] PITMAN, J.: Coalescents with multiple collisions. *Ann. Probab.* **27** (4), 1870-1902, (1999).

[RW87] ROGERS, L.C.G.; WILLIAMS, D.: Diffusions, Markov Processes and Martingales. Vol. 1, 2nd ed., Wiley, (1994)

[S99] SAGITOV, S.: The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36** (4) 1116–1125, (1999).

[S00] SCHWEINSBERG, J.: A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Comm. Probab.* **5**, 1–11, (2000).

[S00b] SCHWEINSBERG, J.: Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, Paper no. 12, 50 pp., (2000).

[S03] SCHWEINSBERG, J.: Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch. Proc. Appl.* **106**, 107–139, (2003).

[SD00] Stephens, M.; Donnelly, P.: Inference in molecular population genetics. *J. Roy. Stat. Soc.* B. **62**, 605–655 (2000).

[T01] Tavaré, S.: Ancestral Inference in Population Genetics. Springer Lecture Notes in Mathematics **1837**, 2001.

[W06] Wakeley, J.: *Coalescent Theory.* To appear (2007).

[W31] Wright, S.: Evolution in Mendelian populations. *Genetics,* **16**:97–159, (1931).