Stochastic models from population biology
lecture notes for a course at TU Berlin, Summer 2005

Matthias Birkner

Preliminary version, 30th June 2005

Chapter 1

Genealogical models for fixed-size populations

1.1 The Wright-Fisher model and the Wright-Fisher diffusion

Consider a population of fixed size $2N$ (the 2 is partly for historic reasons, partly because of the author's bad conscience to talk only about haploid models ...), which evolves in discrete generations. Assume furthermore that individuals can be of two different types (a and A, say) which are *neutral*, i.e. an individual's reproductive success does not depend on her type, and that reproduction is *random*. The latter quality is modelled as follows: Each individual of the daughter generation picks one uniformly (with replacement) from the previous generation, declares this to be her mother, and copies her type. Thus denoting

$$X_k^{(N)} := \#\text{type } a\text{-individuals in generation } k$$

we obtain

$$\mathscr{L}\big(X_{k+1}^{(N)}\big|X_k^{(N)} = j\big) = \text{Bin}(2N, j/2N), \quad j \in \{0, 1, \dots, 2N\}.$$

From this we immediately see that $(X_k^{(N)})_{k=0,1,\dots}$ is a (finite state space) Markov chain and a (bounded) martingale. In particular, $X_k^{(N)}$ converges almost surely as $k \to \infty$, and as 0 and $2N$ are absorbing, we will have $X_\infty^{(N)} \in \{0, 2N\}$, i.e. eventually, one of the types will be fixed. Furthermore, as bounded martingales also converge in L^1, we have

$$j = \mathbb{E}\big[X_0^{(N)}\big|X_0^{(N)} = j\big] = \mathbb{E}\big[X_\infty^{(N)}\big|X_0^{(N)} = j\big] = 2N\mathbb{P}\big(X_\infty^{(N)} = 2N\big|X_0^{(N)} = j\big) + 0,$$

hence the probability that type a fixates, given we start with j individuals of this type, is just its initial frequency $j/2N$.

So, the Wright-Fisher model demonstrates the principal possibility to lose genetic variability by pure chance (which is called *genetic drift* in the population genetics literature, even though the second half of this name is somewhat orthogonal to probabilistic nomenclature).

On the other hand, when N is large, by the central limit theorem, the distribution $\text{Bin}(2N, j/2N)$ is concentrated on $j \pm \text{const.}\sqrt{j(1 - j/2N)}$, which has quite some distance from both 0 and $2N$ when j is of the order N. So how long will we typically have to wait before one of the types is lost? Obviously, if all the living individuals in some future generation k descend from the same ancestor in generation 0 (and, as we assume here, there is no mutation), then everyone in the population will have the same type. We can ask this question also 'backwards in time':

$$\mathbb{P}(\text{two randomly chosen individuals have a different ancestor one generation ago}) = 1 - \frac{1}{2N},$$

so

$$\mathbb{P}(\text{two randomly chosen individuals have a different ancestor } k \text{ generations ago}) = \left(1 - \frac{1}{2N}\right)^k,$$

which is 'non-trivial' if $k \approx \text{const.} \times 2N$. Indeed

$$\mathbb{P}\left(\begin{array}{l}\text{the most recent ancestor of two randomly chosen}\\ \text{individuals lived more than } [2Nt] \text{ generations ago}\end{array}\right) = \left(1 - \frac{1}{2N}\right)^{[2Nt]} \xrightarrow[N \to \infty]{} e^{-t}. \qquad (1.1)$$

This suggests to consider a renormalised version of $X^{(N)}$:

$$Y_t^{(N)} := \frac{1}{2N} X_{[2Nt]}^{(N)}, \ t \geq 0, \qquad (1.2)$$

i.e. we consider the evolution of the fraction of type a-individuals in time units of the population size. One checks that

$$\mathbb{E}\big[(Y_{t+\Delta t}^{(N)} - Y_t^{(N)})|Y_t^{(N)}\big] = 0, \quad \mathbb{E}\big[(Y_{t+\Delta t}^{(N)} - Y_t^{(N)})^2|Y_t^{(N)}\big] \approx \frac{1}{2} Y_t^{(N)}(1 - Y_t^{(N)})\Delta t$$

(when $t \in \mathbb{Z}_+/(2N)$, $\Delta t = 1/(2N)$), so there is hope for convergence towards a diffusion process. Indeed:

Theorem 1 (Fisher, Wright, Kimura, ...). *If $Y_0^{(N)} \to^d Y_0$, then $\mathscr{L}(Y_\cdot^{(N)}) \to \mathscr{L}(Y_\cdot^{(N)})$ (weakly as probability measures on $D_{[0,1]}[0, \infty)$, the space of càdlàg paths on $[0, 1]$, where $Y_\cdot$ is the solution of the following well-posed martingale problem: For any $f \in C^{(3)}([0, 1])$, the process*

$$M_t(f) := f(Y_t) - f(Y_0) - \int_0^t \frac{1}{2} Y_s(1 - Y_s) f''(Y_s) \, ds \qquad (1.3)$$

is a (continuous) martingale (with respect to its own filtration).

Remark 1. There are well-known connections between martingale problems and stochastic analysis, which read in this simple case as follows: The solution of (1.3) can be alternatively described as the solution to a stochastic differential equation, namely

$$dY_t = \sqrt{Y_t(1 - Y_t)} \, dW_t, \qquad (1.4)$$

where (W_t) is a Brownian motion: Let Y be a solution of (1.4) and $f \in C^{(3)}([0, 1])$. Then Itô's formula yields

$$
\begin{aligned}
df(Y_t) &= f'(Y_t) \, dY_t + \frac{1}{2} f''(Y_t) \, d[Y]_t = f'(Y_t)\sqrt{Y_t(1 - Y_t)} \, dW_t + \frac{1}{2} Y_t(1 - Y_t) f''(Y_t) \, dt \\
&= \frac{1}{2} Y_t(1 - Y_t) f''(Y_t) \, dt + d\text{martingale}_t,
\end{aligned}
$$

hence Y also solves (1.3).

The other direction requires more technical work, we only give an impressionistic sketch: if Y solves the martingale problem (1.3), we apply (1.3) to $f(y) = y$ and $f(y) = y^2$ to obtain

$$Y \text{ is a cont. martingale with quadratic variation } [Y]_t = \int_0^t Y_s(1 - Y_s) \, ds.$$

So (modulo integrability and 'stopping on the boundary issues'), the process $\tilde{W}_t := \int_0^t (Y_t(1 - Y_t))^{-1/2} \, dY_t$, suitably augmented when $Y_\cdot$ hits the boundary, is a continuous martingale with $[\tilde{W}]_t = t$, hence is a Brownian motion (and of course $dY_t = (Y_t(1 - Y_t))^{1/2} \, d\tilde{W}_t$).

Note that (even) pathwise uniqueness holds for (1.4) (by the Yamada-Watanabe criterion, see e.g. [RW00], Thm. V.40.1), thus in particular, (1.4) characterises $\mathscr{L}(Y_\cdot|Y_0 = y)$ for any $y \in [0, 1]$, as does (1.3). In that sense, the two formulations are equivalent.

Proof of Thm. 1. There is a well-known three-step choreography which tends to govern the proofs of such limit theorems for Markov processes: 1) check that the sequence $\mathscr{L}(Y^{(N)})$ is tight, so that there are limit points, 2) prove that any limit point solves (1.3), and 3) check that (1.3) has a unique solution (i.e. prove that there is at most one, as one would have already exhibited at least one in step 2), so that all limit points have to agree.

In the following, we carry out this programme, sometimes consciously (and possibly sometimes unconsciously!) glossing over a few details.

3) Even though in the case of a one-dimensional diffusion, other methods to proving well-posedness of a martingale problem are available, let us use the approach "which, despite its strange, ad hoc appearance, has found widespread applicability" ([EK86, p. 188]), namely the use of duality.

Let (N_t) be a pure death process on $\mathbb{N}$ with transitions $n \mapsto n-1$ at rate $\binom{n}{2}$. Then for any solution $Y.$ of (1.3) starting from $Y_0 = y$ and $n \in N$ we have

$$\mathbb{E}\, Y_t^n = \mathbb{E}\big[y^{N_t}\big|N_0 = n\big], \quad t \geq 0. \tag{1.5}$$

In order to prove (1.5), put $g(n, y, t) := \mathbb{E}\, Y_t^n$. Applying (1.3) we see that

$$g(n, y, t) - y^n = \int_0^t \frac{n(n-1)}{2}\big(\mathbb{E}\, Y_s^{n-1} - \mathbb{E}\, Y_s^n\big)ds = \int_0^t \binom{n}{2}\big(g(n-1, y, s) - g(n, y, s)\big)ds.$$

On the other hand, we see from the form of the transition rates of (N_t) that $h(n, y, t) := \mathbb{E}\big[y^{N_t}\big|N_0 = n\big]$ solves the same set of equations. As this is a linear ode system (and hence uniquely solvable), we must have (1.5). Because the right-hand side of (1.5) only depends on Y_0 and polynomials are dense in $C([0, 1])$, we see immediately that the one-dimensional marginals of any two solutions of (1.3) with the same initial value agree. It is then standard to 'patch this up' to any finite-dimensional marginal distribution, details are e.g. in [EK86], Prop. 4.4.7.

The processes $Y.$ and $N.$ are said to be dual with respect to the function $H(y, n) = y^n$. Despite Ethier & Kurtz' above-cited verdict, there is a very natural interpretation of (1.5) in the particular case at hand: (Y_t) describes (at least approximately) the evolution of the type decomposition of a (large) population. Imagine we sample n individuals at time t. The left-hand side of (1.5) is the probability to see only type a in the sample. We can compute this probability differently: If we know that the n individuals in the sample are the descendants of N_t different ancestors at time 0 (when the fraction of type a was y), and then average over the random genealogy, we obtain the right-hand side of (1.5). We will see in Section 1.2.2 that (N_t) is indeed the number of families related from time t back in Kingman's n-coalescent (see (1.12)), which describes the genealogy of a sample in a selectively neutral scenario. An embryonic version of this statement for $n = 2$ is already contained in (1.1).

2) $y = k/2N$, $k \in \{0, 1, \ldots, 2N\}$, $f \in C^{(3)}([0, 1])$. Let $B^{(2N,y)}$ be a binomial$(2N, y)$-distributed random variable. By Taylor expansion,

$$\begin{aligned}
G^{(N)}f(y) &:= \mathbb{E}\Big[f\big(\tfrac{1}{2N}B^{(2N,y)}\big) - f(y)\Big] \\
&= \mathbb{E}\Big[\big(\tfrac{1}{2N}B^{(2N,y)} - y\big)f'(y) + \tfrac{1}{2}\big(\tfrac{1}{2N}B^{(2N,y)} - y\big)^2 f''(y) \\
&\qquad\qquad\qquad\qquad + \tfrac{1}{6}\big(\tfrac{1}{2N}B^{(2N,y)} - y\big)^3 f^{(3)}(\xi)\Big] \\
&= \frac{1}{2N} \cdot \frac{1}{2}y(1 - y)f''(y) + r_N(y),
\end{aligned}$$

where the error term satisfies

$$|r_N(y)| \leq \frac{1}{6}\|f^{(3)}\|_\infty \sup_{y \in [0,1]} \mathbb{E}\Big|\frac{1}{2N}B^{(2N,y)} - y\Big|^3 = O(N^{-3/2}).$$

3

Standard results for discrete time Markov chains yield that for $0 \leq t \leq T$,

$$
\begin{aligned}
M_t^{(N)} &:= f(Y_t^{(N)}) - f(Y_0^{(N)}) - \sum_{i=0}^{[2Nt]-1} G^{(N)} f(Y_{i/(2N)}^{(N)}) \\
&= f(Y_t^{(N)}) - f(Y_0^{(N)}) - \int_0^{[2Nt]/(2N)} \frac{1}{2} Y_s^{(N)}(1 - Y_s^{(N)}) f''(Y_s^{(N)}) + 2N r_N(Y_s^{(N)}) \, ds
\end{aligned}
$$

is a bounded martingale (w.r.t. the filtration generated by $Y^{(N)}$), so for any $0 \leq s < t \leq T$ and any bounded, continuous function $\varphi : D_{[0,1]}[0,s) \to \mathbb{R}$ we have

$$
\mathbb{E}\Big[M_t^{(N)} \varphi\big((Y_u^{(N)})_{0 \leq u \leq s}\big) \Big] = \mathbb{E}\Big[M_s^{(N)} \varphi\big((Y_u^{(N)})_{0 \leq u \leq s}\big) \Big]. \tag{1.6}
$$

Assume that $Y^{(N_k)} \to^d \tilde{Y}$ along some subsequence (N_k) as $k \to \infty$. Then by (1.6) and the Dominated Convergence Theorem, $\tilde{Y}$ solves (1.3).

1) A well-known criterion for tightness on path space (e.g. [EK86, Thm. 3.7.2]) requires that a) the one-dimensional marginals are tight, and b) 'big jumps cannot accumulate', technically this is captured by the requirement

$$
\forall T > 0, \epsilon > 0 \; \exists \delta > 0 \text{ such that } \sup_N \mathbb{P}\big(w(Y^{(N)}, \delta, T) \geq \epsilon\big) \leq \epsilon, \tag{1.7}
$$

where for $g \in D_{[0,1]}[0, \infty)$ the *modulus of continuity* is defined as

$$
w(g, \delta, T) := \inf_{(t_i)} \max_i \sup_{u, v \in [t_{i-1}, t_i)} |g(u) - g(v)|
$$

(the infimum in the definition ranges over all possible partitions $0 = t_0 < t_1 < \cdots < t_{n-1} \leq T < t_n$ $(n \in \mathbb{N})$ satisfying $t_i - t_{i-1} > \delta$; thus, big jumps are 'okay' as long as they are at least δ apart).

In our scenario, a) is trivial because $[0,1]$ is itself compact. In order to verify b), we will prove more, namely

$$
\forall T > 0, \epsilon > 0 \; \exists C_{T,\epsilon} < \infty \; \forall \delta > 0 \; : \; \sup_N \mathbb{P}\big(\exists 0 \leq s < t \leq T : t \leq s + \delta, |Y_t^{(N)} - Y_s^{(N)}| \geq \epsilon\big) \leq C_{T,\epsilon} \delta. \tag{1.8}
$$

Note that this implies (1.7). In fact, it even implies that any limit point concentrates on the set of continuous paths (see e.g. [EK86, Thm. 3.10.2]).

Let $B^{(2N,y)}$ be as above, and note that

$$
\sup_{y \in [0,1]} \mathbb{E}\big(B^{(2N,y)} - y\big)^2 \leq CN, \qquad \sup_{y \in [0,1]} \mathbb{E}\big(B^{(2N,y)} - y\big)^4 \leq CN^2 \tag{1.9}
$$

(e.g. decompose $B^{(2N,y)}$ into a sum of independent Bernoulli variables to prove this). Let $s = i/2N < t = j/2N$, $i, j \in \mathbb{N}_0$. We have

$$
\begin{aligned}
\mathbb{E}\big(Y_t^{(N)} - Y_s^{(N)}\big)^4 &= \frac{1}{(2N)^4} \mathbb{E}\bigg(\sum_{k=i}^{j-1} \big(X_{k+1}^{(N)} - X_k^{(N)}\big)\bigg)^4 \\
&= \frac{1}{(2N)^4} \mathbb{E} \sum_{k_1, \ldots, k_4 = i}^{j-1} \prod_{\ell=1}^4 \big(X_{k_\ell + 1}^{(N)} - X_{k_\ell}^{(N)}\big) \\
&\leq \frac{1}{(2N)^4} \bigg\{ (j-i) \sup_{y \in [0,1]} \mathbb{E}\big(B^{(2N,y)} - y\big)^2 + (j-i)^2 \sup_{y \in [0,1]} \mathbb{E}\big(B^{(2N,y)} - y\big)^4 \bigg\} \\
&\leq C\Big(\frac{j-i}{2N}\Big)^2 = C(t-s)^2
\end{aligned}
$$

4

because only $(k_1, \ldots, k_4)$ with $\#\{k_1, \ldots, k_4\} \in \{1, 2\}$ contribute to the expectation. Now consider $s \in \mathbb{Z}_+/2N$, and $\epsilon, \delta > 0$. From the above and Doob's L^p-inequality we obtain

$$
\begin{aligned}
\mathbb{P}\Big(\sup_{s \leq t \leq t+\delta} \big|Y_t^{(N)} - Y_s^{(N)}\big| \geq \epsilon/4\Big) &\leq \Big(\frac{4}{\epsilon}\Big)^4 \mathbb{E}\Big[\sup_{s \leq t \leq s+\delta} \big|Y_t^{(N)} - Y_s^{(N)}\big|^4\Big] \\
&\leq C' \mathbb{E}\Big[\big|Y_{s+\delta}^{(N)} - Y_s^{(N)}\big|^4\Big] \leq C'' \delta^2.
\end{aligned}
$$

Cutting $[0, T]$ into intervals of length δ, we see that

$$
\begin{aligned}
\mathbb{P}&\big(\exists\, 0 \leq s < t \leq T : t \leq s + \delta, |Y_t^{(N)} - Y_s^{(N)}| \geq \epsilon\big) \\
&\leq \sum_{i=0}^{[T/\delta]+1} \mathbb{P}\Big(\sup_{i\delta \leq t \leq (i+1)\delta} \big|Y_t^{(N)} - Y_{i\delta}^{(N)}\big| \geq \epsilon/4\Big) + \sum_{i=0}^{[T/\delta]+1} \mathbb{P}\Big(\big|Y_{(i+1)\delta}^{(N)} - Y_{i\delta}^{(N)}\big| \geq \epsilon/4\Big) \\
&\leq 2C'' \big([T/\delta] + 1\big) \delta^2 \leq C''' \delta,
\end{aligned}
$$

proving (1.8). □

1.2 Neutral genealogies

1.2.1 Cannings' model

Consider a population of fixed size N of (haploid, asexual) individuals which live and reproduce in discrete generations in a temporally stable environment such that each individual's chance to contribute to the next generation is 'the same' (i.e. neutrality). A way to formalise this scenario mathematically is due to Chris Cannings: Imagine that the individuals of the first generation are labelled in some arbitrary way, and let ν_i^r, $i = 1, \ldots, N$ be the number of offspring of the i-th individual in the r-th generation ($r = 0, 1, \ldots$). Necessarily, $\nu_1^r + \cdots + \nu_N^r = N$. To formally specify the model, we would have to define the labelling of individuals in generation r, $r \geq 1$. This will not matter very much for our following considerations, but for definiteness let us e.g. imagine that the labels are assigned randomly (but of course consistent with the prescribed offspring numbers).

Cannings imposed the following conditions on the vectors of offspring numbers:

$$\nu^r, r \in \mathbb{N}_0 \text{ are i.i.d. copies of } \nu, \tag{1.10}$$

$$\mathcal{L}(\nu) \text{ is exchangeable, i.e. } \big(\nu_1, \ldots, \nu_N\big) \overset{d}{=} \big(\nu_{\pi(1)}, \ldots, \nu_{\pi(N)}\big) \text{ for any permutation } \pi \in S_N \tag{1.11}$$

By exchangeability, we have $\mathbb{E}\,\nu_1 = 1$, and further on we will assume that $\mathrm{Var}(\nu_1) > 0$, so that the trivial case $\mathbb{P}(\nu_1 = \cdots = \nu_N = 1) = 1$ is excluded.

Note that this naturally generalises the Wright-Fisher model, in which ν is multinomial, i.e.

$$\mathbb{P}(\nu_1 = k_1, \ldots, \nu_N = k_n) = \binom{N}{k_1 \ldots k_N} N^{-N}.$$

1.2.2 Kingman's coalescent

Imagine we pick $n < N$ individuals at random without replacement from the 'present generation' of a population of size N which is governed by Cannings model (and which 'has been living forever'). Let us label the sampled individuals with $1, 2, \ldots, n$ and write $i \sim_r j$ ($i, j \in \{1, \ldots, n\}$, $r = 0, 1, \ldots$) if i and j have the same ancestor r generations back in the past. Let us denote the collection of such relations by $R_r^{(N,n)}$. Each $R_r^{(N,n)}$ is a (random) equivalence relation on $\{1, \ldots, n\}$, which we can equivalently encode by listing its equivalence classes. Obviously $i \sim_0 j \iff i = j$, i.e. $R_0^{(N,n)}$ is the trivial equivalence relation.

For $\xi, \eta \in \mathcal{E}_n := \{$equivalence relations on $\{1, \ldots, n\}\}$ let us write $\xi \subseteq \eta$ if $i \sim_\xi j \Rightarrow i \sim_\eta j$, i.e. if η can be constructed by lumping some classes of ξ together. For fixed N and n, $(R_r^{(N,n)})_{r=0,1,\ldots}$ is a Markov chain.

Let $\eta = \{C_1, \ldots, C_a\}$ consist of $|\eta| = a$ classes, and assume that it can be obtained from a $\xi = \{C_{\alpha\beta} : 1 \le \alpha \le a, 1 \le \beta \le b_\alpha\}$ with $|\xi| = b_1 + \cdots + b_a =: b$ classes by coarsening: $C_\alpha = \cup_{\beta=1}^{b_\alpha} C_{\alpha\beta}$. Assume that $R_r^{(N,n)} = \xi$. Then there are presently b 'active lineages' (one representative for each equivalence class). In order to have $R_{r+1}^{(N,n)} = \eta$, representatives $1, \ldots, b_1$ have to pick the same ancestor, $b_1 + 1, \ldots, b_1 + b_2$ have to pick the same ancestor (but different from the one chosen by $\{1, \ldots, b_1\}$), etc. Fix a possible offspring vector $(k_1, \ldots, k_n)$ with $k_1 + \cdots + k_N = N$ and a pairwise different indices $i_1, \ldots, i_a \in \{1, \ldots, N\}$. The probability that representatives $1, \ldots, b_1$ pick i_1 as their ancestor in the previous generations, representatives $b_1 + 1, \ldots, b_1 + b_2$ pick i_2, etc. is

$$\frac{k_{i_1}(k_{i_1} - 1) \cdots (k_{i_1} - b_1 + 1)}{N(N-1) \cdots (N - b_1 + 1)} \cdot \frac{k_{i_2}(k_{i_2} - 1) \cdots (k_{i_2} - b_2 + 1)}{(N - b_1)(N - b_1 - 1) \cdots (N - b_1 - b_2 + 1)} \cdots$$
$$\cdot \frac{k_{i_a}(k_{i_a} - 1) \cdots (k_{i_a} - b_a + 1)}{(N - b_1 - \cdots - b_{a-1})(N - b_1 - \cdots - b_{a-1} - 1) \cdots (N - b_1 - b_2 - \cdots - b_{a-} + 1)}$$
$$= \frac{1}{(N)_b} \prod_{j=1}^a (k_{i_j})_{b_j},$$

where $(x)_c := x(x-1) \cdots (x - c + 1)$ is the lower factorial. Thus the transition probabilities are

$$p^{(N,n)}(\xi, \eta) := \mathbb{P}(R_{r+1}^{(N,n)} = \eta \,|\, R_r^{(N,n)} = \xi) = \frac{1}{(N)_b} \sum_{\substack{i_1, \ldots, i_a \\ \text{pairwise distinct}}} \mathbb{E}\left[\prod_{j=1}^a (\nu_{i_j})_{b_j}\right]$$

$$= \frac{(N)_a}{(N)_b} \mathbb{E}\left[\prod_{j=1}^a (\nu_j)_{b_j}\right]$$

(the last identity uses exchangeability).

Let

$$c_N := \frac{1}{N(N-1)} \sum_{i=1}^N \mathbb{E}[\nu_i(\nu_i - 1)] = \frac{1}{N-1} \mathbb{E}[\nu_1(\nu_1 - 1)]$$

be the probability that two individuals chosen today have the same ancestor one generation back (or equivalently, that two lineages will coalesce in the next generation backwards). Alternatively, in this model $c_N = (N-1) \times$ variance of the number of offspring of one individual. Assume that as $N \to \infty$, $c_N \to 0$. Note that

$$\mathbb{P}(1 \not\sim_r 2) = (1 - c_N)^r \approx e^{-t} \quad \text{if } r = t/c_N,$$

which suggests to look at

$$\mathcal{R}_t^{(N,n)} := R_{[t/c_N]}^{(N,n)}, \ t \ge 0.$$

Let $(\mathcal{R}_t^{(n)})_{t \ge 0}$ be a continuous time Markov chain on $\mathcal{E}_n$ with rate matrix given by

$$q_{\xi\eta} = \begin{cases} 1 & \text{if } \eta \text{ arises from } \xi \text{ by merging exactly two classes,} \\ -\binom{|\xi|}{2} & \text{if } \xi = \eta, \\ 0 & \text{otherwise.} \end{cases} \tag{1.12}$$

This is Kingman's (n-)coalescent.

Theorem 2 (Kingman, Möhle-Sagitov). $(\mathcal{R}_t^{(N,n)})_{t \ge 0} \longrightarrow (\mathcal{R}_t^{(n)})_{t \ge 0}$ *in distribution (on $D_{[-,\infty)}(\mathcal{E}_n)$) if and only if $c_N \to 0$ as $N \to \infty$ and $\lim_{N \to \infty} d_N/c_N = 0$, where*

$$d_N := \frac{(N)_1}{(N)_3} \mathbb{E}[(\nu_1)_3] = \frac{1}{(N-1)(N-2)} \mathbb{E}[\nu_1(\nu_1 - 1)(\nu_2 - 2)] \tag{1.13}$$

6

is the probability to observe a triple merger within the sample in one generation.

Proof. Fix n for the moment. For sufficiency, we have to show that (1.13) implies

$$p^{(N,n)}(\xi,\eta) = \delta_{\xi,\eta} + c_N q_{\xi\eta} + o(c_N), \tag{1.14}$$

where the error term is uniformly small for all $\xi, \eta \in \mathcal{E}_n$. The claimed convergence follows from (1.14) by standard arguments, see e.g. Lemma 1 below.

Note that a useful estimate is

$$\mathbb{E}\big[\nu_2 f(\nu_1)\big] \leq \frac{N}{N-1}\mathbb{E}\big[f(\nu_1)\big] \quad \text{for any } f : \{0,1,\ldots,N\} \to \mathbb{R}_+, \tag{1.15}$$

which holds true because by exchangeability,

$$(N-1)\mathbb{E}\big[\nu_2 f(\nu_1)\big] = \sum_{j=2}^{N}\mathbb{E}\big[\nu_j f(\nu_1)\big] = \mathbb{E}\big[(N-\nu_1)f(\nu_1)\big] \leq N\mathbb{E}\big[f(\nu_1)\big].$$

Let $I_1, I_2, \ldots, I_n$ be random (ordered) picks from the (ordered) offspring generated by ν, i.e.

$$\mathbb{P}\big(I_1 = i_1, \ldots, I_n = i_n\big) = \frac{1}{(N)_n}\mathbb{E}\bigg[\prod_{i=1}^{n}\nu_{i_1}\big(\nu_{i_2} - \mathbf{1}_{i_2=i_1}\big)\cdots\big(\nu_{i_2} - \sum_{\ell=1}^{n-1}\mathbf{1}_{i_n=i_\ell}\big)\bigg].$$

We have $c_N = \mathbb{P}(I_1 = I_2)$, $d_N = \mathbb{P}(I_1 = I_2 = I_3)$, i.e. (1.13) means

$$\lim_{N\to\infty}\frac{\mathbb{P}(I_1 = I_2 = I_3)}{\mathbb{P}(I_1 = I_2)} = 0. \tag{1.16}$$

Furthermore, the Markov inequality and (1.13) imply for any $\epsilon > 0$

$$\mathbb{P}(\nu_1 > \epsilon N) \leq \frac{1}{(\epsilon N)_3}\mathbb{E}\big[(\nu_1)_3\big] = \frac{1}{\epsilon^3 N^3}o\big(N\mathbb{E}[(\nu_1)_2]\big) = \epsilon^{-3}o\big(c_N/N\big). \tag{1.17}$$

From this, we obtain

$$\begin{aligned}
\mathbb{E}\big[(\nu_1)_2(\nu_2)_2\big] &\leq& \epsilon N\mathbb{E}\big[(\nu_1)_2\nu_2\mathbf{1}(\nu_2 \leq \epsilon N)\big] + N^2\mathbb{E}\big[(\nu_1)_2\mathbf{1}(\nu_2 > \epsilon N)\big]\\
&\leq& \epsilon N\mathbb{E}\big[(\nu_1)_2\nu_2\big] + N^3\mathbb{E}\big[\nu_1\mathbf{1}(\nu_2 > \epsilon N)\big]\\
&\leq& \epsilon N\frac{N}{N-1}\mathbb{E}\big[(\nu_1)_2\big] + N^3\frac{N}{N-1}\mathbb{P}(\nu_2 > \epsilon N),
\end{aligned}$$

where we used (1.15) in the last inequality. Combing this with (1.17) we see that for any $\epsilon > 0$

$$\limsup_{N\to\infty}\frac{\mathbb{E}\big[(\nu_1)_2(\nu_2)_2\big]}{N\mathbb{E}\big[(\nu_1)_2\big]} \leq \epsilon + \limsup_{N\to\infty}\frac{N\mathbb{P}(\nu_1 > \epsilon N)}{c_N} = \epsilon,$$

which yields

$$\lim_{N\to\infty}\frac{\mathbb{P}(I_1 = I_2 \neq I_3 = I_4)}{\mathbb{P}(I_1 = I_2)} = \lim_{N\to\infty}\frac{(N)_2\mathbb{E}\big[(\nu_1)_2(\nu_2)_2\big]}{(N)_4}\cdot\frac{(N)_2}{N\mathbb{E}\big[(\nu_1)_2\big]} = 0. \tag{1.18}$$

Consider $\xi = \{C_{11}, C_{12}, C_2, \ldots, C_a\}$ and $\eta = \{C_1, \ldots, C_a\}$, where $C_1 = C_{11} \cup C_{12}$. We have

$$p^{(N,n)}(\xi,\eta) = \mathbb{P}\big(\{I_1 = I_2\} \cap \{I_m \neq I_1, m = 3, \ldots, a\} \cap \{I_\ell \neq I_m, 3 \leq \ell < m \leq a\}\big) \tag{1.19}$$

so $p^{(N,n)}(\xi,\eta) \leq \mathbb{P}(I_1 = I_2)$, on the other hand

$$\begin{aligned}
p^{(N,n)}(\xi,\eta) &\geq& \mathbb{P}\big(I_1 = I_2\big) - \mathbb{P}\big(\{I_1 = I_2\} \cap \{\exists\, 3 \leq m \leq a : I_m = I_1\}\big)\\
&& - \mathbb{P}\big(\{I_1 = I_2\} \cap \{\exists\, 3 \leq \ell < m \leq a : I_\ell = I_m \neq I_1\}\big).
\end{aligned}$$

7

Note that

$$\mathbb{P}\big(\{I_1 = I_2\} \cap \{\exists\, 3 \le m \le a : I_m = I_1\}\big) \;\le\; (a-2)\mathbb{P}\big(I_1 = I_2 = I_3\big) = o\big(\mathbb{P}(I_1 = I_2)\big)$$

by (1.13) and that

$$\mathbb{P}\big(\{I_1 = I_2\} \cap \{\exists\, 3 \le \ell < m \le a : I_\ell = I_m \ne I_1\}\big) \;\le\; \binom{a-2}{2}\mathbb{P}\big(I_1 = I_2 \ne I_3 = I_4\big)$$
$$= o\big(\mathbb{P}(I_1 = I_2)\big)$$

by (1.18), so that indeed $p(\xi,\eta) = c_N + o(c_N)$ in this case.

Similarly, if η' arises from ξ by merging more than two classes, we have $p(\xi,\eta') = o(c_N)$, as then there must be at least either a triple merger or two double mergers. Finally, as $\sum_{\eta \in \mathcal{E}_n} p(\xi,\eta) = 1$ and $|\mathcal{E}_n| < \infty$, we must also have $p(\xi,\xi) = 1 - \binom{|\xi|}{2}c_N + o(c_N)$. Combining, we obtain (1.14).

$\square$

Lemma 1. *Let E be a finite set and $q = (q_{xy})_{x,y \in E}$ a conservative rate matrix (i.e. $q_{xy} \ge 0$ for $x \ne y$, $q_{xx} = -\sum_{y \ne x} q_{xy}$), so that there is a unique continuous time Markov chain X on E with generator matrix q. Let furthermore $X^{(N)}$, $N \in \mathbb{N}$ be discrete time Markov chains with transition matrix $p^{(N)}$ satisfying*

$$p^{(N)}(x,y) = \delta_{x,y} + c_N q_{xy} + o(c_N), \quad x, y \in E,$$

where $c_N \to 0$ as $N \to \infty$. Then $X_0^{(N)} \to^d X_0$ implies that the processes $(X_{[t/c_N]}^{(N)})_{t \ge 0}$ converge in distribution (on $D_{[0,\infty)}(E)$) to $(X_t)_{t \ge 0}$.

Let $\mathcal{R}^{(n)}$ start from the trivial relation on $\mathcal{E}_n$, and denote by T_k the length of the time interval during which one sees k classes ($k = n, n-1, \ldots, 2$, $T_1 = \infty$ because the all-relation is absorbing). We see from (1.12) that the variables

$$T_n, T_{n-1}, \ldots, T_2 \quad \text{are independent, and } \mathscr{L}(T_k) = \mathrm{Exp}\Big(\frac{k(k-1)}{2}\Big),$$

so the expected time until one sees only one class (i.e. the time to the most recent common ancestor of all the individuals in the sample) is

$$\sum_{k=2}^n \frac{2}{k(k-1)} = 2\sum_{k=2}^n \frac{1}{k-1} - \frac{1}{k} = 2\Big(1 - \frac{1}{n}\Big). \tag{1.20}$$

Kingman's (n-)coalescents have a consistency property (which is natural from the viewpoint of modelling a sample from a very large population): Let $n' > n$, and denote by $d_n(R)$ the restriction of $R \in \mathcal{E}_{n'}$ to $\mathcal{E}_n$, then

$$d_n\big((\mathcal{R}_t^{(n')})_{t \ge 0}\big) \overset{d}{=} (\mathcal{R}_t^{(n)})_{t \ge 0}.$$

This allows to define Kingman's coalescent $(\mathcal{R}_t)$ as the Markov process on $\mathcal{E} := \{\text{equivalence relations on } \mathbb{N}\}$ with the property that $d_n\big((\mathcal{R}_t)\big) =^d (\mathcal{R}_t^{(n)})$ for each n. We see from (1.20) that $(\mathcal{R}_t)$ *comes down from infinity*, i.e. starting from $\mathcal{R}_0 = \{\{1\}, \{2\}, \ldots\}$, we will have $|\mathcal{R}_t| < \infty$ almost surely for any $t > 0$. Note that while coalescence is very quick at the beginning, the end does come comparatively slowly: (1.20) shows that the expected time to the MRCA of an infinite sample is 2, but the expected time to merge the last two classes into one is 1.

8

1.3 Infinitely many alleles

In the previous section, we have considered models for the *genealogy* connecting a sample from a population with random 'mating', but we have not really looked at types. Let us now introduce types, which can change through *mutation*. In this section, we consider a *neutral* scenario with *parent independent* mutation, i.e. an individual's type does not affect her (chances of) reproductive success, and given that someone is hit by a mutation, her new type will be just a random draw from a fixed mutation law, independent of the previous type. While these assumptions seem quite restrictive, there are some applications (e.g. two neutral types or the so-called infinitely-many-alleles model), and furthermore, they allow a nice decoupling of the genealogy and the types in a sample.

Let E be the set of possible types, $\mu \in \mathscr{M}_1(E)$. Consider a population of $2N$ (haploid) individuals, each with a type $\in E$, which evolves according to the Wright-Fisher model. Types are passed on from parent to offspring, but assume that for each offspring independently with probability $\theta/(4N)$ ($\theta \in \mathbb{R}_+$ is a parameter) there occurs a mutation, and the child gets a random type according to (an independent draw from) μ. If N is very large, the results above may render the following description of the genealogy and type history of a sample of size $n \ll N$, measured in units of $2N$ generations, plausible:

The genealogy is given by Kingman's n-coalescent. Additionally, on each lineage there is a(n independent) Poisson process with rate $\theta/2$ of mutation events, mutations are marked independently according to μ. The type of sample no. i is determined by following its lineage backwards until it hits the first mutation, from which we can read off the type. See e.g. the following picture.
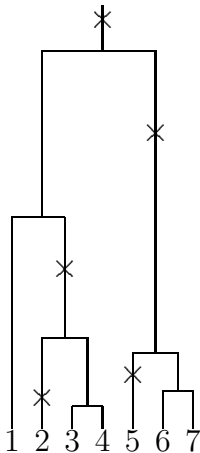

Figure 1.1: An example. Note that we have shamelessly drawn a rather unlikely realisation in which no 'crossing of lineages' occurs.

The infinitely-many-alleles model refers to such a situation where each mutation produces a novel type, realised e.g. via $E = [0, 1]$ and $\mu =$ Lebesgue measure. Essentially, one only records whether two given samples are of the same type or not. For example, in the picture above, the families are $\{1\}, \{2\}, \{3, 4\}, \{5\}, \{6, 7\}$, i.e. we observe 3 singletons and 2 families of size 2.

Note that, because a mutation completely erases the information about the previous type in the infinitely-many-alleles model, the part of the coalescent 'below' a mutation is irrelevant for the determination of the types in the sample. Thus, if we are only interested in the type distribution in the sample, we could as well discontinue (or 'kill') a lineage once it has hit a mutation, like in the following variant of the above picture
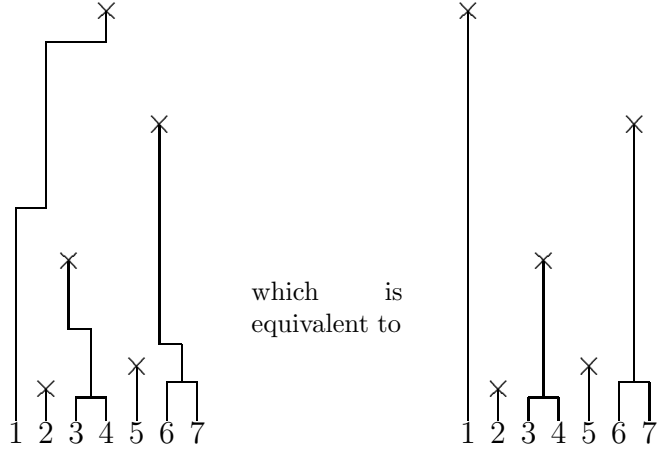
We can produce this killed coalescent as follows: Start with $\{\{1\}, \{2\}, \ldots, \{n\}\}$, i.e. initially only singleton classes, as usual.

$$\text{Any pair of active classes is merged at rate 1, and} \tag{1.21}$$
$$\text{any active class is killed at rate } \theta/2.$$

When a class is 'killed', all its elements are assigned the same (novel) type. Finish when there are no classes left.

1.3.1 Hoppe's urn

Assume that there are presently k active lineages in the killed coalescent described by (1.21). Then the probability that the next event (backwards in time) will be a coalescence is

$$\frac{\binom{k}{2}}{\binom{k}{2} + k\frac{\theta}{2}} = \frac{k-1}{k-1+\theta},$$

and with probability

$$\frac{k\frac{\theta}{2}}{\binom{k}{2} + k\frac{\theta}{2}} = \frac{\theta}{k-1+\theta}$$

it will be a mutation (corresponding to killing of a lineage). By the symmetry of the model, given this type, all possible coalescences resp. killings are equally likely. The history of a sample of size n corresponds to n events $e_n, e_{n-1}, \ldots, e_1 \in \{\texttt{coal}, \texttt{mut}\}$ (say, indexed by the number of lineages active before the event), and the probability to observe a particular sequence is thus

$$\frac{\prod_{k=1}^{n} \left(\theta \mathbf{1}(e_k = \texttt{mut}) + (k-1)\mathbf{1}(e_k = \texttt{coal})\right)}{\prod_{k=1}^{n} (k-1+\theta)}. \tag{1.22}$$

Fred Hoppe [Hop84] observed that one can also generate this sequence 'forwards in time' using an urn model (in the spirit of Pólya's urn):

Hoppe's urn model. Initially, the urn contains one black ball of mass θ. In each step, a ball is

10

drawn from the urn with probability proportional to its mass. If the drawn ball is black, return it together with a ball of mass 1 of a new, not previously used colour; if the chosen ball is coloured, return it together with another ball of mass 1 and the same colour.

In the k-th step, there are k balls (including the black or 'mutation' ball) in the urn, thus the probability to pick the mutation ball is $\theta/(k-1+\theta)$, and the probability to pick a coloured ball is $k/(k-1+\theta)$. Hence, if we denote

$$e_k = \begin{cases} \texttt{mut} & \text{if in the } k\text{-th step, the black ball is chosen,} \\ \texttt{coal} & \text{otherwise,} \end{cases}$$

we see that the probability to observe any particular sequence of length n is also given by (1.22). Furthermore, given that $e_k = \texttt{coal}$, each of the then present $k-1$ coloured balls is equally likely to be picked. Thus:

> The distribution of the family sizes generated by the n non-black balls in Hoppe's urn after n steps is the same as the one induced by the n-coalescent in the infinitely-many-alleles model. $\qquad(1.23)$

But note different labellings: Hoppe's urn produces 'age order', we had numbered the lineages in 'sample order'.

Immediate upshot: Let

$$K_n := \text{no. of different types observed in a sample of size } n,$$

then we see

$$\mathbb{E}\,K_n = \sum_{i=1}^{n} \frac{\theta}{\theta + i - 1} \sim \theta \log n, \quad \mathrm{Var}(K_n) = \sum_{i=1}^{n} \frac{\theta}{\theta + i - 1} \cdot \frac{i-1}{\theta + i - 1} \sim \theta \log n, \qquad (1.24)$$

$$\text{and} \quad \frac{K_n - \mathbb{E}\,K_n}{\sqrt{\mathrm{Var}(K_n)}} \xrightarrow{d} \text{standard normal.}$$

In order to see this it suffices to write $K_n = A_1 + \cdots + A_n$, where $A_i = \mathbf{1}$(black ball is drawn in the i-th step) and note that $A_1, \ldots, A_n$ are independent with $\mathbb{P}(A_i = 1) = \theta/(\theta + i - 1)$.

Remark 2. There is a nice relation between Hoppe's urn and random permuations, as noted by [JT87]: Think of the coloured balls as being numbered in the order of their appearance. The family decomposition generated by n draws from Hoppe's urn can then be uniquely represented as the cyclic decomposition of a permuation of $\{1, 2, \ldots, n\}$ which arises as follows: If in the k-th step we draw the black ball, start a new cycle consisting only of (k). Otherwise, add k (say, to the left) of the number of the drawn ball.

For example, the picture on the right encodes the permutation $(1)(52)(73)(4)(6)$, which arose sequentially as (1), $(1)(2)$, $(1)(2)$, $(1)(2)(3)$, $(1)(2)(3)(4)$, $(1)(2)(3)(4)$, $(1)(52)(3)(4)$, $(1)(52)(3)(4)(6)$, $(1)(52)(73)(4)(6)$.

11
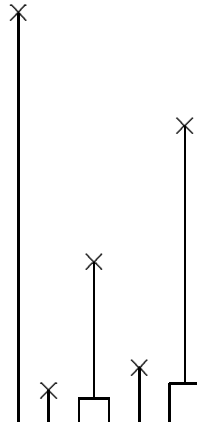
1.3.2 Ewens' sampling formula

Here, we derive a formula (first found by, and subsequently named after, Warren Ewens) for the probability distribution of the type spectrum in a sample of size n from the equilibrium distribution of the infinitely many alleles model.

Theorem 3 (Ewens' sampling formula). *Let $b_1, \ldots, b_n \in \mathbb{N}_0$ with $\sum_{j=1}^{n} b_j = k \leq n$, $\sum_{j=1}^{n} j b_j = n$ be given. The probability to observe b_j types with j representatives $(j = 1, \ldots, n)$ in a sample of size n is given by*

$$\frac{n!}{1^{b_1} 2^{b_2} \cdots n^{b_n}} \cdot \frac{1}{b_1! b_2! \cdots b_n!} \cdot \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)}. \tag{1.25}$$

Note that an alternative way to write (and possibly remember) (1.25) is

$$C(n, \theta) \times \prod_{j=1}^{n} e^{-\theta/j} \frac{(\theta/j)^{b_j}}{b_j!} \tag{1.26}$$

(where $C(n, \theta) = n! \exp\left(\theta \sum_{j=1}^{n} 1/j\right)/(\theta(\theta+1)\cdots(\theta+n-1)))$ which shows that the distribution of the type spectrum $(B_1, \ldots, B_n)$ in an n-sample is $\otimes_{j=1}^{n} \mathrm{Poi}(\theta/j)$, conditioned on $\sum_{j=1}^{n} j B_j = n$.

Proof. We follow [GL04]. Let us consider an n-coalescent in which the sampled individuals are (artificially) labelled by $1, 2, \ldots, n$. In the infinitely-many-alleles model, mutations (to novel types) 'rain down' at rate $\theta/2$ per lineage. Imagine we follow this coalescent backwards in time and sequentially record the changes we see. Once a lineage has hit a mutation, we do not follow it backwards any more: irrespectively of what happened before, the type is defined by this mutation. This *fine protocol* will be a sequence $e_1, e_2, \ldots, e_n$ of n *elementary events*, which will be either of type

$\mathtt{mut}(i)$, i.e. lineage i hits a mutation event, or

$\mathtt{coal}(i \to j)$, i.e. lineage i coalesces into lineage $j (\neq i)$

for some i (and possibly j) $\in \{1, \ldots, n\}$. Unlike the 'classical' coalescent thinking, we keep track of who *coalesces into* whom for the moment. Note that both $\mathtt{coal}(i \to j)$ and $\mathtt{mut}(i)$ render lineage no. i *inactive*, meaning that it cannot appear again in an elementary event later in the protocol. Obviously, the possible fine protocols have to satisfy this consistency condition. Furthermore, the last event e_n must necessarily be a mutation.

There is a coalescence rate of $1/2$ per ordered pair of (still active) lineages, and a mutation rate of $\theta/2$ per lineage. Before the m-th elementary event, there are $n - m + 1$ active lineages, so the probability of observing a particular e_m is

$$\begin{cases} \dfrac{1/2}{\frac{1}{2}(n-m)(n-m+1) + \frac{\theta}{2}(n-m+1)} = \dfrac{1}{(n-m+1)(n-m+\theta)} & \text{if } e_m \text{ is a coalescence,} \\[2ex] \dfrac{\theta/2}{\frac{1}{2}(n-m)(n-m+1) + \frac{\theta}{2}(n-m+1)} = \dfrac{\theta}{(n-m+1)(n-m+\theta)} & \text{if } e_m \text{ is a mutation.} \end{cases} \tag{1.27}$$

Consequently the probability to observe a given (consistent) fine protocol which contains $k (\leq n)$ mutation events, and thus describes the history of a sample in which we observe k distinct types, is given by

$$\frac{\theta^k}{\prod_{m=1}^{n}(n-m+1)(n-m+\theta)} = \frac{\theta^k}{n! \, \theta(\theta+1)\cdots(\theta+n-1)}. \tag{1.28}$$

For given type spectrum $b_1, \ldots, b_n$ with $\sum_{j=1}^{n} b_j = k$ (and, of course, $\sum_{j=1}^{n} j b_j = n$), we have to count how many possible fine protocols would yield exactly this type spectrum:

To facilitate our reasoning, let us assume for the moment that there are k *artificially labelled* types. Pick family sizes $n_\ell \in \mathbb{N}$, $\ell = 1, 2, \ldots, k$ subject to $\#\{\ell : n_\ell = j\} = b_j$, $j = 1, 2, \ldots, n$. Let us also enrich our notion of fine protocols with these labels, in the sense that a mutation event now also carries the information which type was produced, e.g. $\mathtt{mut}_\ell(i)$ means that (the presently active) lineage i hits a mutation event which produces type ℓ. Of course, these enriched fine protocols have to obey the additional condition that each of the types from $\{1, \ldots, k\}$ appears exactly once in a mutation event.

In order to produce a fine protocol *with labelled types* which describes a sample with n_ℓ representatives of type ℓ, $\ell = 1, \ldots, k$, we can proceed as follows:

1. Fix an order in which the lineages become inactive. There are $n!$ possibilities.

2. Assign a type to each lineage. There are $\binom{n}{n_1 n_2 \ldots n_k}$ possibilities: this amounts to putting n distinguishable balls (the lineages) into k distinguishable boxes (the types) such that n_ℓ balls land in box ℓ, $\ell = 1, \ldots, k$.

3. Prescribe 'coalescence targets' within each type. While the previous two steps were completely independent, we now have to respect the previous choices.

   Consider type ℓ, and assume that in Step 2, we have decided that samples $i_1, i_2, \ldots, i_{n_\ell}$ are of this type, and that our choice of ordering from Step 1 dictates that i_1 is lost first, i_2 second, etc. So we are free to chose into which one of the $n_\ell - 1$ other lineages $\{i_j, 2 \leq j \leq n_\ell\}$ lineage i_1 coalesces, etc, yielding $(n_\ell - 1)(n_\ell - 2) \cdots 2 = (n_\ell - 1)!$ possibilities. We can (and have to) prescribe these 'targets' within each of the k types, leading to altogether

   $$(n_1 - 1)! \times (n_2 - 1)! \times \cdots \times (n_\ell - 1)!$$

   possible choices.

We see that the choices from Steps 1–3 together determine a fine protocol *with labelled types*, and vice versa. Combining the number of possible choices in the three steps, we obtain

$$n! \times \binom{n}{n_1 \ldots n_\ell} \times (n_1 - 1)! \cdots (n_\ell - 1)! = \frac{(n!)^2}{\prod_{\ell=1}^{k} n_\ell} = \frac{(n!)^2}{\prod_{j=1}^{n} j^{b_j}} \tag{1.29}$$

different fine protocols with labelled types which produce n_ℓ representatives of type ℓ ($\ell = 1, \ldots, k$) in the sample.

Finally, we translate this back to unlabelled fine protocols: A fine protocol with labelled types immediately yields one without labels by simply ignoring the labels. We have to correct for over-counting: if we exchange two types ℓ and ℓ' with the same number of representatives $n_\ell = n_{\ell'}$ in the sample, the corresponding unlabelled fine protocols will be identical. Thus, $b_1! \cdot b_2! \cdots b_n!$ labelled protocols correspond to the same unlabelled one. Altogether we find

$$\frac{(n!)^2}{\prod_{j=1}^{n}(b_j! j^{b_j})} \tag{1.30}$$

different (unlabelled) fine protocols which produce the desired type spectrum. Note that (1.30) multiplied with (1.28) yields (1.25), which is the claim. $\square$

1.3.3 Estimating θ

We see from (1.24) that $K_n / \log n$ is an asymptotically normal (and consistent) estimate of θ, but we also see that the variance of this estimator decays only like $1/\log n$. Unfortunately, one cannot do really better:

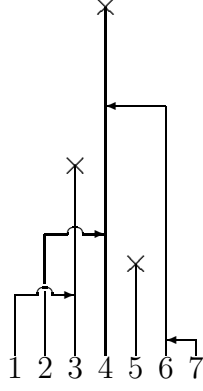$$K_n \text{ is a sufficient statistic for } \theta, \tag{1.31}$$

13

Figure 1.2: An example. The corresponding fine protocol is $(\texttt{coal}(7 \to 6), \texttt{coal}(1 \to 3), \texttt{mut}(5), \texttt{coal}(2 \to 4), \texttt{mut}(3), \texttt{coal}(6 \to 4), \texttt{mut}(4))$, the type spectrum produced is $b_1 = 1, b_2 = 1, b_3 = 0, b_4 = 1, b_5 = b_6 = b_7 = 0$.

i.e. for each k, the distribution of the types in the sample, given $K_n = k$, does not depend on θ. Let

$$C_{k,n} := {\sum}'_{(b_1,\ldots,b_n)} \prod_{i=1}^{n} \frac{n!}{i^{b_i} i!},$$

where the sum $\sum'$ extends over all $(b_1, \ldots, b_n) \in \mathbb{N}_0^n$ such that $\sum b_i = k$, $\sum i b_i = n$. Note that $C_{k,n}$ counts the number of permutations of $\{1, \ldots, n\}$ with exactly k cycles, it is known as (the modulus of) a Stirling number of the first kind in the literature. Then we see from (1.25) that the probability to observe a given type spectrum $(b_1, \ldots, b_n)$ with k types in a sample of size n from the infinitely-many-alleles model with mutation rate $\theta/2$, conditioned on seeing k different types, is

$$\frac{1}{C_{k,n}} \frac{n!}{1^{b_1} 2^{b_2} \cdots n^{b_n}} \cdot \frac{1}{b_1! b_2! \cdots b_n!}, \tag{1.32}$$

proving (1.31).

Denote by $\mathbb{P}_\theta(\cdot)$ the probability distribution governing the infinitely-many-alleles model with mutation rate $\theta/2$ per lineage. From (1.25) we obtain for $k \leq n$

$$\mathbb{P}_\theta(K_n = k) = C_{k,n} \frac{\theta^k}{\theta(\theta+1)\cdots(\theta+n-1)}.$$

For a given observation $K_n = k$, we want to maximise this likelihood $L_n(\theta, k) := \mathbb{P}_\theta(K_n = k)$ with respect to θ. Note that

$$\frac{\partial}{\partial \theta} \log L_n(\theta, k) = \frac{\partial}{\partial \theta} \left(k \log \theta - \sum_{i=0}^{n-1} \log(\theta + i) \right) = \frac{k}{\theta} - \sum_{i=0}^{n-1} \frac{1}{\theta + i} = \frac{1}{\theta} \left(k - \sum_{i=0}^{n-1} \frac{\theta}{\theta + i} \right),$$

so the maximum likelihood estimate $\hat{\theta}$ is the solution of

$$k = \sum_{i=0}^{n-1} \frac{\hat{\theta}}{\hat{\theta} + i} = \mathbb{E}_{\hat{\theta}} K_n, \tag{1.33}$$

i.e. it is that value of θ under which the expected number of types equals the observed number. Furthermore we see from this that the Fisher information is in this scenario

$$I(\theta) = \mathbb{E}_\theta \left(\frac{\partial}{\partial \theta} \log L_n(\theta, K_n) \right)^2 = \frac{1}{\theta^2} \mathbb{E}_\theta \left(K_n - \sum_{i=0}^{n-1} \frac{\theta}{\theta + i} \right)^2 = \frac{1}{\theta^2} \operatorname{Var}_\theta(K_n).$$

14

As $\mathrm{Var}_\theta(K_n) \sim \theta \log n$ by (1.24), we see that $I(\theta) \sim \theta^{-1} \log n$ as $n \to \infty$, and the variance of $\hat\theta$ decays (only) like $1/\log n$.

1.3.4 The GEM distribution

In this section, we study the asymptotic distribution of the frequencies we observe in a (large) sample from the infinitely many alleles model with mutation rate $\theta/2$ per lineage. In order to do this let us consider a Hoppe urn, label the types (or 'families') in the order of appearance ('age order') and denote

$$X_k(n) := \text{size of family } k \text{ after the } n\text{-th draw from Hoppe's urn.}$$

(obviously $X_1(1) = 1$, $X_k(n) = 0$ if $k > n$). Mathematically, our problem is to describe the distribution of

$$\left(\frac{1}{n}X_1(n), \frac{1}{n}X_2(n), \frac{1}{n}X_3(n), \dots\right) \tag{1.34}$$

for $n \to \infty$. Note that $(n+\theta)^{-1}X_1(n)$ $(n = 2, 3, \dots)$ is a bounded martingale:

$$\mathbb{E}\left[\frac{1}{n+1+\theta}X_1(n+1)\Big|\mathcal{F}_n\right] = \frac{X_1(n)+1}{n+1+\theta}\frac{X_1(n)}{n+\theta} + \frac{X_1(n)}{n+1+\theta}\frac{\theta+n-X_1(n)}{n+\theta} = \frac{X_1(n)}{n+\theta}$$

and thus converges almost surely. Similarly, if α_k denotes the time of the first appearance of type k, $(n+\alpha_k+\theta)^{-1}X_k(n+\alpha_k)$, $k = 1, 2, \dots$ is a martingale. As $n/(n+\theta) \to 1$, we see that (1.34) converges almost surely (at least in the sense that any finite initial piece converges).

The following representation of the GEM distribution appeared in [Tav87].

Theorem 4 (GEM distribution, named for Griffiths, Engen, McCloskey). *Let $B_1, B_2, \dots$ be i.i.d. Beta$(1,\theta)$, i.e. with density $\theta(1-b)^{\theta-1}$ on $[0,1]$. The distribution of the limit of (1.34) is given by*

$$\left(B_1, (1-B_1)B_2, (1-B_1)(1-B_2)B_3, (1-B_1)(1-B_2)(1-B_3)B_4, \dots\right).$$

It is convenient to represent (1.34) by means of a (continuous-time) Yule process with immigration. Let $0 < T_1 < T_2 < \cdots$ be the jump times of a homogeneous Poisson point process with rate θ on $\mathbb{R}_+$. At time T_i, the i-th immigrant appears and founds the i-th family, which from then on grows independently of everything else as a (rate 1) Yule process (i.e. a pure birth process on $\mathbb{N}$ with linear birth rate). Denote by $Z_i(t)$ the size of family i at time t. In particular, $Z_i(T_i) = 1$, $Z_i(t) = 0$ if $t < T_i$. $S(t) := \sum_{i=1}^\infty Z_i(t)$ is the total size of the population at time t. Let $\tau_n := \min\{t : S(t) = n\}$. Then we have

$$\left(\frac{1}{n}Z_1(\tau_n), \frac{1}{n}Z_2(\tau_n), \frac{1}{n}Z_3(\tau_n), \dots\right)_{n=1,2,\dots} \stackrel{d}{=} \left(\frac{1}{n}X_1(n), \frac{1}{n}X_2(n), \frac{1}{n}X_3(n), \dots\right)_{n=1,2,\dots} \tag{1.35}$$

In order to see this note that if there are presently k families of sizes $j_1, j_2, \dots, j_k$ with $j_1 + \cdots + j_k = n$, the jump rate of $S(t)$ is $n + \theta$, so the probability that the next event in the Yule process is an immigration is $\theta/(n+\theta)$, while it will be an increase of family i with probability $j_i/(n+\theta)$. In words, Hoppe's urn describes the skeleton chain of our Yule process with immigration.

We see from Lemma 2 that

$$\left(e^{-t}Z_1(t), e^{-t}Z_2(t), e^{-t}Z_3(t), \dots\right) \to \left(e^{-T_1}A_1, e^{-T_2}A_2, e^{-T_3}A_3, \dots\right) \quad \text{a.s.,} \tag{1.36}$$

where $A_1, A_2, \dots$ are i.i.d. Exp(1). Consequently,

$$e^{-t}S(t) \to \sum_{n=1}^\infty e^{-T_n}A_n \quad \text{a.s.} \tag{1.37}$$

15

(In order to justify the exchange of limit and infinite summation note e.g. that $M_i := \sup_{t\geq 0} e^{-t} Z_i(T_i+t)$ are i.i.d. with $\mathbb{E} M_1 < \infty$, hence $\limsup M_n/n = 0$ by the Borel-Cantelli Lemma. Additionally, $T_n/n \to \theta^{-1}$ by the strong law, which yields for $m \geq N_0$

$$\sup_{t\geq 0} \sum_{n=m}^{\infty} e^{-T_n} e^{-(t-T_n)} Z_n(t) \leq \sum_{n=m}^{\infty} e^{-T_n} M_n \leq \sum_{n=m}^{\infty} n e^{-2n/\theta}. \quad)$$

Furthermore,

$$\mathscr{L}\Big(\sum_{n=1}^{\infty} e^{-T_n} A_n \Big) = \mathrm{Gamma}(\theta). \tag{1.38}$$

In order to see this, note that $\sum_i \delta_{(A_i, T_i)}$ is a Poisson process on $\mathbb{R}_+ \times \mathbb{R}_+$ with intensity measure $\theta dt \otimes e^{-x} dx$. Then $\tilde{\Pi} := \sum_i \delta_{e^{-T_i} A_i}$ is a Poisson process on $\mathbb{R}_+$ with intensity measure $\frac{\theta}{y} e^{-y} dy$, which we recognise as the Lévy measure of the Gamma process, see e.g. [Ber96], p. 73. To check the claimed form of the intensity measure, let $h : (0,\infty) \to \mathbb{R}_+$ be, say continuous with compact support, and note that

$$\int_0^{\infty} \int_0^{\infty} h(e^{-t} a)\, \theta dt\, e^{-a} da = \int_0^{\infty} \int_0^a h(r) \theta \frac{dr}{r} e^{-a} da = \int_0^{\infty} h(r) \int_r^{\infty} e^{-a}\, da\, \theta \frac{dr}{r} = \int_0^{\infty} h(r) \theta e^{-r} \frac{dr}{r}.$$

The general observation behind this is of course that if $\Pi = \sum \delta_{a_i}$ is a PPP on E with intensity measure ν, $f : E \to E'$ is measurable, and $\tilde{\Pi} = \sum \delta_{f(a_i)}$, then $\tilde{\Pi}$ is a PPP on E' with intensity measure $\tilde{\nu} = \nu \circ f^{-1}$.

The argument that led to (1.38) shows also that

$$\mathscr{L}(G,T) = \mathrm{Gamma}(1+\theta) \otimes \mathrm{Exp}(1) \;\Rightarrow\; \mathscr{L}(e^{-T/\theta} G) = \mathrm{Gamma}(1+\theta), \tag{1.39}$$

because $\sum_{n=1}^{\infty} e^{-T_n} A_n = e^{-T_1}\big(A_1 + \sum_{n=2}^{\infty} e^{-(T_n - T_1)} A_n \big)$ (alternatively, one can note that $e^{-T/\theta} \sim \mathrm{Beta}(\theta, 1)$ and then use Lemma 3).

We see from the above that

$$\frac{Z_1(t)}{S(t)} = \frac{e^{T_1 - t} Z_1(t)}{e^{T_1 - t} Z_1(t) + \sum_{i=2}^{\infty} e^{T_1 - t} Z_i(t)} \to \frac{A_1}{A_1 + \sum_{i=2}^{\infty} e^{-(T_i - T_1)} A_i} =: B_1 \quad \text{a.s.,}$$

and by Lemma 3, $\mathscr{L}(B_1) = \mathrm{Beta}(1, \theta)$, and B_1 and $A_1 + \sum_{i=2}^{\infty} e^{-(T_i - T_1)} A_i$ are independent. Put

$$C_n := A_n + \sum_{i=n+1}^{\infty} e^{-(T_i - T_n)} A_i, \quad B_n := \frac{A_n}{C_n}.$$

We prove by induction that for any $n \in \mathbb{N}$,

$$\mathscr{L}\Big(C_1, B_1, B_2, \ldots, B_n \Big) = \mathrm{Gamma}(1+\theta) \otimes \mathrm{Beta}(1, \theta)^{\otimes n}. \tag{1.40}$$

By the argument above, (1.40) holds true for $n = 1$. Assume that it is satisfied for n, then the stationarity and independence properties of (A_i) and the increments of a Poisson process imply that also $\mathscr{L}(C_2, B_2, B_3, \ldots, B_{n+1}) = \mathrm{Gamma}(1+\theta) \otimes \mathrm{Beta}(1, \theta)^{\otimes n}$. Furthermore, $(C_2, B_2, B_3, \ldots, B_{n+1})$ and $(A_1, T_2 - T_1)$ are independent. We have

$$C_1 = A_1 + e^{-(T_2 - T_1)} C_2, \quad B_1 = \frac{A_1}{C_1} = \frac{A_1}{A_1 + e^{-(T_2 - T_1)} C_2}.$$

By (1.39), $e^{-(T_2 - T_1)} C_2 \sim \mathrm{Gamma}(\theta)$, so Lemma 3 shows that (1.40) is satisfied for $n + 1$. (We remark that of course, (1.40) can be also proved by direct calculation, cf. [Tav87], p. 167.)

16

Finally we observe that

$$\frac{e^{-t}Z_n(t)}{e^{-t}S(t)} \rightarrow \frac{e^{-T_n}A_n}{\sum_{i=1}^{\infty}e^{-T_i}A_i} = \frac{\sum_{i=2}^{\infty}e^{-T_i}A_i}{\sum_{i=1}^{\infty}e^{-T_i}A_i} \times \cdots \times \frac{\sum_{i=n}^{\infty}e^{-T_i}A_i}{\sum_{i=n-1}^{\infty}e^{-T_i}A_i} \times \frac{e^{-T_n}A_n}{\sum_{i=n}^{\infty}e^{-T_i}A_i}$$

$$= (1-B_1) \times \cdots \times (1-B_n) \times \frac{A_n}{A_n + \sum_{i=n+1}^{\infty}e^{-(T_i-T_n)}A_i}$$

$$= (1-B_1) \times \cdots \times (1-B_{n-1})B_n.$$

This concludes the proof of Theorem 4.

Lemma 2. *Let (Y_t) be a rate 1 Yule process starting from $Y_0 = 1$. Then $\mathscr{L}(Y_t) = \mathrm{geom}(e^{-t})$. $(e^{-t}Y_t)_{t\geq 0}$ is an L_2-bounded martingale, its limit is $\mathrm{Exp}(1)$-distributed.*

Proof. Let $T_i := |\{t : Y_t = i\}|$ be the time during which there are i individuals in the population. By the form of the transition rates, $T_1, T_2, \ldots$ are independent, and $\mathscr{L}(T_i) = \mathrm{Exp}(i)$. Consequently for $n = 0, 1, 2, \ldots$ we have

$$\mathbb{P}(Y_t > n) = \mathbb{P}(T_1 + \cdots + T_n < t) = \mathbb{P}\left(\max_{i=1,\ldots,n} \tau_i < t\right) = (1 - e^{-t})^n,$$

where the τ_i are i.i.d. $\mathrm{Exp}(1)$. This shows that $\mathscr{L}(Y_t) = \mathrm{geom}(e^{-t})$. Alternatively, one can note that the solution of the forward equation

$$\frac{\partial}{\partial t}\mathbb{P}_1(Y_t = n) = (n-1)\mathbb{P}_1(Y_t = n-1) - n\mathbb{P}_1(Y_t = n), \quad \mathbb{P}_1(Y_0 = n) = \delta_{1n}$$

is $\mathbb{P}_1(Y_t = n) = e^{-t}(1 - e^{-t})^{n-1}$, or just look up the answer e.g. in Athreya & Ney. This together with the branching property of the Yule process (i.e. $\mathscr{L}(Y_t|Y_0 = k + j) = \mathscr{L}(Y_t|Y_0 = k) * \mathscr{L}(Y_t|Y_0 = j)$, which stems from the linearity of the birth rates) yields $\mathbb{E}[Y_{t+h}|Y_t] = e^h Y_t$, proving that $(e^{-t}Y_t)$ is a martingale.

Elementary properties of the geometric distribution imply that $\sup_{t\geq 0}\mathbb{E}[(e^{-t}Y_t)^2] < \infty$ and $\mathscr{L}(e^{-t}Y_t) \rightarrow \mathrm{Exp}(1)$. $\square$

Lemma 3. *Let G_1 and G_2 be independent, G_i be Gamma(θ_i)-distributed (i.e. density $(\Gamma(\theta_i))^{-1}g^{\theta_i-1}e^{-g}$ on $\mathbb{R}_+$). Then*

$$\mathscr{L}\left(G_1 + G_2, \frac{G_1}{G_1 + G_2}\right) = \mathrm{Gamma}(\theta_1 + \theta_2) \otimes \mathrm{Beta}(\theta_1, \theta_2).$$

Proof. Put $G := G_1 + G_2$. The joint density of (G_1, G) is

$$f_{(G_1,G)}(g_1, g) = c\mathbf{1}(0 \leq g_1 \leq g)g_1^{\theta_1-1}e^{-g_1}(g-g_1)^{\theta_2-1}e^{-(g-g_1)} = ce^{-g}\mathbf{1}(0 \leq g_1 \leq g)g_1^{\theta_1-1}(g-g_1)^{\theta_2-1},$$

so the conditional density of G_1, given $G = g$ is

$$f_{G_1|G=g}(g_1) = c(g)\mathbf{1}(0 \leq g_1 \leq g)g_1^{\theta_1-1}(g - g_1)^{\theta_2-1},$$

and hence the conditional density of G_1/G, given $G = g$ is

$$f_{(G_1/G)|G=g}(b) = \tilde{c}(g)\mathbf{1}(0 \leq b \leq 1)b^{\theta_1-1}(1 - b)^{\theta_2-1}.$$

On the other hand, as $\int_0^1 f_{(G_1/G)|G=g}(b)\,db = 1$, we see that $\tilde{c}(g) = \Gamma(\theta_1 + \theta_2)/(\Gamma(\theta_1)\Gamma(\theta_2))$ is in fact independent of g. $\square$

The Poisson-Dirichlet distribution

The GEM distribution describes the frequency spectrum in a particular random order ('age order', or 'size-biased order' in random partition jargon). A possibly natural question is to look at the distribution of the *ranked* type frequencies. Let $1 \geq V_1 > V_2 > \cdots$ be the frequency of the most frequent, the second most frequent type, etc., in the stationary infinitely many alleles model. So $(V_1, V_2, \dots)$ is the order statistic of the distribution described in Thm. 4. By ranking and then normalising the right-hand side of (1.36), we obtain

Corollary 1. *Let $\Pi = \sum_i \delta_{X_i}$ be a Poisson point process on $\mathbb{R}_+$ with intensity measure $(\theta/x)e^{-x}dx$. Put $S := \sum X_i$, and let $X_{[1]} > X_{[2]} > \cdots$ be the ranked sequence of the X_is. Then*

$$\left(X_{[1]}/S, X_{[2]}/S, \dots \right) \overset{d}{=} \left(V_1, V_2, \dots \right).$$

In words, the ranked type frequencies have the distribution of the ranked jumps of a 'standard' gamma subordinator (up to time θ), normalised to sum to one. This distribution is called the Poisson-Dirichlet distribution .

Unfortunately, the Poisson-Dirichlet distribution is, as Kingman writes , "not very user-friendly".

Parent-independent mutation with finitely many types

Imagine a situation with k different neutral types, where mutations occur at rate $\theta/2$ per line and a mutation leads to type j with probability $\pi_j \, (> 0)$ (π is a probability measure on $\{1, 2, \dots, k\}$) independently of the type of the parent. We can now easily read off the stationary distribution:

$$\mathscr{L}\left(X_1(\infty), \dots, X_k(\infty) \right) = \text{Dirichlet}(\theta\pi_1, \dots, \theta\pi_k), \tag{1.41}$$

i.e. the joint density is

$$\frac{\Gamma(\theta)}{\Gamma(\theta\pi_1) \cdots \Gamma(\theta\pi_k)} x_1^{\theta\pi_1 - 1} \cdots x_k^{\theta\pi_k - 1}$$

with respect to Lebesgue measure on $\{(x_1, \dots, x_k) : 0 \leq x_i \leq 1, x_1 + \cdots + x_k = 1\}$.

In order to see this note that to generate the stationary distribution of the k-allele model with parent independent mutation, we can first generate a realisation of the type frequencies in the infinitely-many-alleles model and then 'colour' each of the types independently according to π. Corollary 1 tells us that the type frequencies are given (up to normalisation) by the jumps of a gamma subordinator. It is well-known that the jumps of a subordinator form a Poisson process, and if we colour the jumps independently, the coloured jumps will form independent Poisson processes. Thus if $Y_i \sim \text{Gamma}(\theta\pi_i)$ and the $Y_1, \dots, Y_k$ are independent, then

$$\left(X_1(\infty), \dots, X_k(\infty) \right) \overset{d}{=} \left(\frac{Y_1}{Y_1 + \cdots + Y_k}, \dots, \frac{Y_k}{Y_1 + \cdots + Y_k} \right),$$

and the right-hand side has the required Dirichlet distribution. This multivariate generalisation of Lemma 3 can be found e.g. in Chap. 40.5 of [JK72].

Two neutral types

Let us briefly remark that a two-type situation can always be reformulated to fit into the parent independent context: Assume that there are types a and A, and that mutations $a \to A$ occur at rate $\theta_0/2$ (on the underlying coalescent), similarly $A \to a$ at rate $\theta_1/2$. Put $\theta := \theta_0 + \theta_1$, $\pi = (\pi_a, \pi_A) := (\theta_1/\theta, \theta_0/\theta)$. Imagine following a line which is currently of type a. It will we be hit by mutations at rate $\theta/2$, but the mutation will be 'silent' (i.e. invisible to us because it changes $a \to a$) with probability π_a, so effectively the rate of mutation on this lineage is $\theta/2 \times \pi_A = \theta_0/2$, as required.

In order to describe the state of the population (with respect to the particular locus we are interested in) it suffices to record the fraction X_t of a-individuals (at time t). Specialising (1.41) to the first coordinate in the case $k = 2$, we see that then

$$\mathscr{L}(X_\infty) = \mathrm{Beta}(\theta\pi_a, \theta\pi_A) = \mathrm{Beta}(\theta_1, \theta_0). \tag{1.42}$$

Remark 3. One can derive this in a completely different way: The forwards in time evolution of the fraction of a-individuals is given by the Wright-Fisher diffusion with mutation, i.e. (X_t) solves

$$
\begin{aligned}
dX_t &= \Big(-\frac{\theta_0}{2}X_t + \frac{\theta_1}{2}(1 - X_t)\Big)dt + \sqrt{X_t(1 - X_t)}dW_t \\
&= -\frac{\theta}{2}\Big(X_t - \frac{\theta_1}{\theta}\Big)dt + \sqrt{X_t(1 - X_t)}dW_t = b(X_t) + \sigma(X_t)dW_t
\end{aligned}
$$

(where W is a BM, $b(\cdot)$ is the drift coefficient, $\sigma(\cdot)$ the diffusion coefficient). Let $\phi(x) := \exp\big(-\int_{1/2}^x \frac{2b(z)}{\sigma^2(z)}\,dz\big)$, $x \in [0, 1]$. (Note that $s(x) := \int^x \phi(z)\,dz$ is a scale function for the diffusion (X_t): e.g. Itô's Formula shows that $s(X_t)$ is a martingale.) Then by standard results on one-dimensional diffusion processes, the equilibrium density of X is given by $\rho(x) = C/(\phi(x)\sigma^2(x))$ (where $1/C = \int_0^1 1/(\phi(x)\sigma^2(x))dx$). See e.g. Satz 3 in § 23 of I.I. Gichman and A.W. Skorochod, *Stochastische Differentialgleichungen*, Akademie-Verlag, Berlin, 1971.
A 'quick and dirty' way to check this at least heuristically is as follows: Assuming that an equilibrium density $\rho(x)$ exists, we must have for f in the domain of the generator

$$0 = \int \rho(x)\big[b(x)f'(x) + \tfrac{1}{2}\sigma^2(x)f''(x)\big]dx = \int \big[-(\rho b)'(x) + \tfrac{1}{2}(\rho\sigma^2)''(x)\big]f(x)\,dx,$$

by integration by parts (assuming that the boundary terms vanish). As this must hold for 'many' fs, we have $(\rho b)' = \frac{1}{2}(\rho\sigma^2)''$, for which a sufficient condition is $\rho b = \frac{1}{2}(\rho\sigma^2)'$, or equivalently $\big(\log(\rho\sigma^2)\big)' = (\rho\sigma^2)'/(\rho\sigma^2) = 2b/\sigma^2$. Hence $\rho = \mathrm{const}/(\sigma^2\phi)$.

In the situation at hand,

$$\int_{1/2}^x \frac{2b(z)}{\sigma^2(z)}\,dz = \int_{1/2}^x -\frac{\theta(z - \frac{\theta_1}{\theta})}{z(1 - z)}\,dz = \int_{1/2}^x \frac{\theta_1 - \theta}{1 - z} + \frac{\theta_1}{z}\,dz = \theta_0\log(1 - x) + \theta_1\log(x) + \mathrm{const.},$$

yielding

$$\rho(x) = C\exp\Big(\theta_0\log(1 - x) + \theta_1\log(x)\Big)\frac{1}{x(1 - x)} = C\,x^{\theta_1 - 1}(1 - x)^{\theta_0 - 1}.$$

1.3.5 The infinitely-many-sites model

If in a genetic study, information on (a certain piece of) the, say DNA or polypeptide, sequences of the sampled individuals is available, then it seems that the infinitely-many-alleles model is too 'rough': With sequence information, one can not only decide whether two samples have exactly the same type; given that they are different, we can now ask by which mutation(s) they differ. A model that allows to take this information into account is the *infinitely-many-sites* model, introduced by M. Kimura in 1971.

We model a sequence as the unit interval $[0, 1]$, and we assume that in the course of evolutions, mutations 'rain down' one the sequence (as it is passed through the generations) at rate $\theta/2$. Given that a mutation occurs, it hits a uniformly (and independently of everything else) chosen position in $[0, 1]$. In particular, each mutation hits a new site (thence the name of the model). So the type of an individual is described by a simple counting measure on $[0, 1]$, which describes the pattern of mutations that we see in this individual compared to some hypothetical reference type. We still assume that mutations are neutral with respect to an individuals fitness.

Note that there is no true stationary distribution for the actual types in the population, because as time increases, more and more mutations occur. On the other hand, a mutation

that is carried by everyone in the population is invisible to us because we do not know the founding reference type. Thus, if we agree that mutation patterns are defined only relative to the (unknown) type of the most recent common ancestor, we can generate a sample of size n as follows:

First, grow an n-coalescent. Second, throw down mutations at the lineages at rate $\theta/2$ per lineage, and mark each mutation with an independent, unif$([0,1])$ label. The type of a leaf is the point measure on $[0,1]$ whose atoms are the labels of the mutations which occurred on the lineage which connects this leaf to the MRCA (a rigorous definition of the model is in [EG87]). See the picture below.
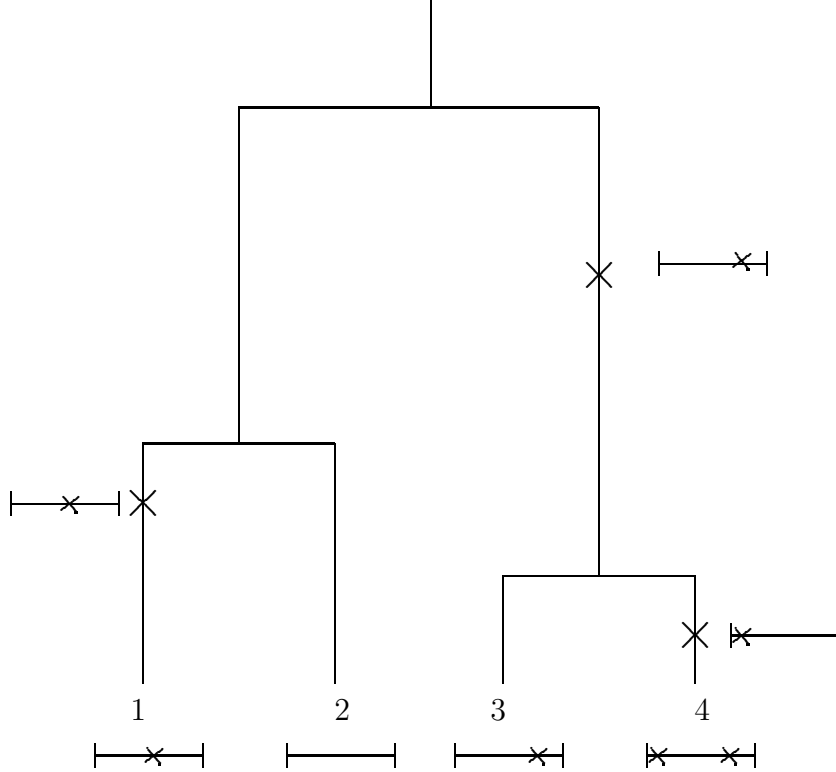
This is still a bit more than the information that is actually contained in the sampled sequences: if we observe that sequence i and j differ at a certain site, we do not know which of them carries the mutation relative to the type of the MRCA. A site is called *segregating* if at least one pair of samples differs at this site. So, the information from an n-sample under the infinitely-many-sites model is

$$S_n = \#\text{ segregating sites, and } D_m(i,j) = \begin{cases} 1 & \text{if samples } i \text{ and } j \text{ differ at segregating site } m \\ 0 & \text{otherwise} \end{cases}$$

for $m = 1, \ldots, S_n$, $1 \le i < j \le n$. Note that S_n = total number of mutations on the coalescent connecting the sample up to the MRCA, so given the total length T_{tot} of the coalescent, S_n is Poi$\big((\theta/2)T_{\text{tot}}\big)$-distributed. As $T_{\text{tot}} =^d \sum_{j=2}^n jT_j$, where $T_2, \ldots, T_n$ are independent, $\mathscr{L}(T_j) = \text{Exp}\big(\binom{j}{2}\big)$, we see that

$$\mathbb{E}_\theta S_n = \frac{\theta}{2} \sum_{j=2}^n j\binom{j}{2} = \theta \sum_{i=1}^{n-1} \frac{1}{i}, \tag{1.43}$$

$$\text{Var}_\theta(S_n) = \mathbb{E}_\theta\big[\text{Var}_\theta(S_n|T_{\text{tot}})\big] + \text{Var}_\theta\big(\mathbb{E}_\theta[S_n|T_{\text{tot}}]\big)$$

$$= \mathbb{E}_\theta\big[\frac{\theta}{2}T_{\text{tot}}\big] + \text{Var}_\theta\big(\frac{\theta}{2}T_{\text{tot}}\big) = \theta \sum_{i=1}^{n-1} \frac{1}{i} + \theta^2 \sum_{i=1}^{n-1} \frac{1}{i^2}. \tag{1.44}$$

20

Thus, $S_n / \sum_{i=1}^{n-1} (1/i)$ is an unbiased estimate of θ, and its variance decays like $1/\log n$ as the sample size tends to infinity. No closed-form expression for the likelihood of a given pattern, say as a function of θ, under the infinitely-many-sites model is known, so for inference questions, one usually has to rely on computer-intensive methods. See e.g. [SD00], Section 5.5.

Note that if we ignore the fine structure of the sequences and just record which samples were identical, the model reduces to the infinitely-many-alleles model. Also note that, unlike the case of the latter, it is possible that data explicitly contradict the infinitely-many-sites model. Consider the following example:
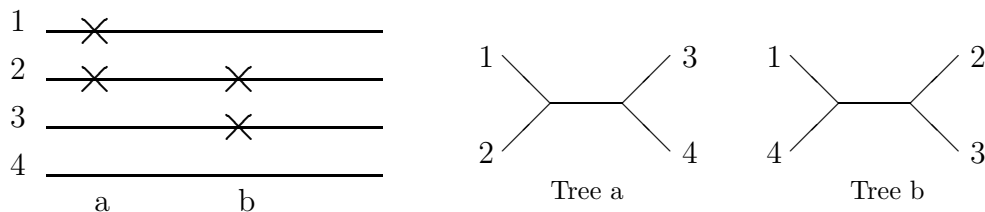
Here, the mutation pattern at site a requires that the topology of the genealogical tree connecting the sampled individuals 1,2,3,4 must be tree a, whereas the observations from site b dictate that it must be tree b. So, if there was a genealogical tree at all, one of the sites must have been hit twice.

1.3.6 'Tests for neutrality'

The quotation marks around the title of this subsection refer to the following general problem: if in a statistical test, we use some neutral model as the null hypothesis and find that the data allow to reject this null hypothesis at a certain level of significance, this does not necessarily mean that we have found evidence for selection acting in the population. Other violations of the model, like spatial substructure, assortative mating, non-constant population size, recombination, non-appropriateness of the mutation model etc. could have a similar effect.

Watterson's test

Assume that we observe a $K_n = k$ certain type frequency spectrum $(B_1, B_2, \ldots, B_n)$ in a sample of size n (i.e. B_j types with j representatives, $\sum_j B_j = K_n$) and we would like to test whether this is plausible under the infinitely-many-alleles model. Watterson (1977) suggested to consider

$$F := \sum_{j=1}^{n} B_j \left(\frac{j}{n}\right)^2, \quad \text{the 'sample homozygosity'.} \tag{1.45}$$

If the observed value of F is less than the $\alpha/2$-quantile or larger than the $1 - \alpha/2$ of f under the infinitely-many-alleles model, we can reject it at the α-significance level. An enjoyable property of the infinitely-many-alleles model is that by (1.31), $\mathscr{L}(F|K_n = k)$ does not depend on θ. No closed form for of distribution of F, given $K_n = k$ types in a sample of size n is known. In principle (at least for moderate n, where moderate depends on the available computing power), (1.32) allows to explicitly compute this distribution numerically by brute force. On the other hand, it is also easy using a Hoppe urn to simulate $\mathscr{L}(F|K_n = k)$. A table of values can be found e.g. in Appendix B of [Ewe04].

Tajima's D

In the situation of the infinitely-many-sites model, F. Tajima (1989) noted that one can build a statistical test out of two different estimators for θ. Let $T_{ij} := \sum_{m=1}^{S_n} D_m(i, j)$ be the number

of differences we observe between the sampled sequences i and j. As the expected coalescent distance between any two samples is 2, we have $\mathbb{E}_\theta\, T_{ij} = \theta$, so

$$\hat{\theta}_T := \frac{1}{\binom{n}{2}} \sum_{1 \le i < j \le n} T_{ij}$$

is an unbiased estimate of θ. One can show by a lengthy coalescent calculation that $\mathrm{Var}_\theta(\hat{\theta}_T) = \frac{n+1}{3(n-1)}\theta + \frac{2(n^2+n+3)}{9n(n-1)}\theta^2$, see e.g. [Dur02], p. 44–50. Note that this converges to $\theta/3 + 2\theta^2/9 > 0$ as $n \to \infty$, which makes it a rather unreliable estimator, but this does not prevent its use in a statistical test. We have seen above that $\hat{\theta}_S := S_n/\sum_{i=1}^{n-1}(1/i)$ is also an unbiased estimator for θ, so if the infinitely-many-sites models applies, $\hat{\theta}_S - \hat{\theta}_T$ should be zero up to statistical fluctuations. Another lengthy coalescent calculation (cf. e.g. [Dur02], Section 4.1) yields

$$\mathrm{Var}_\theta(\hat{\theta}_S - \hat{\theta}_T) = \left(b_1 - \frac{1}{a_1}\right)\theta + \left(b_2 - \frac{n+2}{a_1 n} + \frac{a_2}{a_1}\right)\theta^2, \quad \text{where}$$

$$a_1 = \sum_{i=1}^{n-1} \frac{1}{i}, \; a_2 = \sum_{i=1}^{n-1} \frac{1}{i^2}, \; b_1 = \frac{n+1}{3(n-1)}, \; b_2 = \frac{2(n^2+n+3)}{9n(n-1)}.$$

Using this one can check that

$$\hat{V} := \left(\frac{b_1}{a_1} - \frac{1}{(a_1)^2}\right)S_n + \frac{b_2 - \frac{n+2}{a_1 n} + \frac{a_2}{a_1}}{(a_1)^2 + a_2}S_n(S_n - 1)$$

is an unbiased estimate of $\mathrm{Var}_\theta(\hat{\theta}_S - \hat{\theta}_T)$. This suggests to use

$$D := \frac{\hat{\theta}_S - \hat{\theta}_T}{\sqrt{\hat{V}}}, \quad \text{Tajima's } D,$$

as a test statistic. Note that the distribution does depend on θ and is not explicitly known. Based on simulation results, Tajima suggested to approximate the distribution of D by a scaled and translated beta distribution and based a test on this. Cf. e.g. [Ewe04], Section 11.3.3 for details.

1.4 Selection, the modified look down construction, and some of that

1.4.1 Preliminaries

Moran model

A popular time continuous variant of the Wright-Fisher model is Moran's model (we introduce this model here, because e.g. the 'look down' construction considered below is easier to work with in a time-continuous setting): N (haploid) individuals, each with an independent, $\mathrm{Exp}(1)$-distributed lifetime. Upon her death, a particle is immediately replaced by the offspring of an individual sampled randomly from the current population. Assume that individuals have types in E (and for the moment, that types are neutral and there is no mutation). Let $X_i^{(N)}(t)$ be the type of individual i at time t, then $X^{(N)}$ is an E^N-valued continuous time Markov chain with generator

$$Gf(x_1, \dots, x_N) = \sum_{\substack{1 \le i,j \le N \\ i \ne j}} \frac{1}{N-1}\Big(f(\eta_i(x; x_j)) - f(x_1, \dots, x_N)\Big),$$

where $\eta_i(x; x_j)$ is x with the value of the i-th coordinate replaced x_j.

If we assume that $E = \{a, A\}$ and we record only the fraction of type a-individuals, and (as we have found reasonable for the Wright-Fisher model) speed up time by a factor of N, then $Y^{(N)}(t) := \#\{i : X_i^{(N)}(Nt) = a\}/N$ is 'almost a diffusion': it is a continuous time Markov chain on $\{0, 1/N, 2/N, \ldots, 1\}$ (in fact, a scaled birth and death process) with generator

$$L\phi(x) = N^2 x(1-x)\big(\phi(x+1/N) + \phi(x-1/N) - 2\phi(x)\big) \approx x(x-1)\phi''(x).$$

This makes it at least plausible that $Y^{(N)} \to Y$, where $dY(t) = \sqrt{2Y(t)(1-Y(t))}dW(t)$, which is a (scaled) Wright-Fisher diffusion.

Why 'weak' selection

Assume $E = \{a, A\}$, and type a has a *selective advantage* over type A. To be specific, we will consider *fecundity selection* here: type A-individuals produce offspring at rate 1, while type a-individuals reproduce at rate $1 + s > 1$. Then in a population of fixed size N, the number of type a-individuals is a birth-death chain with jump rates

$$n \mapsto \begin{cases} n+1 & \text{at rate } n(1+s)(1-n/N) \\ n-1 & \text{at rate } (N-n) \times (n/N) \end{cases}$$

So, denoting the fraction of a-individuals at time t by Z_t, we have (for $\phi \in C^2([0,1])$)

$$\mathbb{E}\big[\phi(Z_{t+\Delta t}) - \phi(Z_t)\big|Z_t\big]$$
$$\approx \Delta t \times \Big(N Z_t(1+s)(1-Z_t)\big(\phi(Z_t+1/N) - \phi(Z_t)\big) + N(1-Z_t)Z_t\big(\phi(Z_t-1/N) - \phi(Z_t)\big)\Big)$$
$$\approx \Delta t \times Z_t(1-Z_t)\Big(s\phi'(Z_t) + \frac{1+s/2}{N}\phi''(Z_t)\Big),$$

which suggests that for fixed $s > 0$, as N tends to infinity, the process (Z_t) — at least as long as it is inside $(\epsilon, 1-\epsilon)$, say — should look similar to the solution of the *deterministic* ODE $(d/dt)z(t) = z(t)(1-z(t))$. Note that we have not speeded up time in this argument, whereas in a population of size N, as we have seen above, we would have to wait time of the order N before we can observe type frequency changes which are due resampling fluctuations. Thus, selection with fixed selective advantage in a very large population operates on a different time scale than random genetic drift, and in such a situation, a stochastic model may not be appropriate (or at least not necessary). . To sum up, in this scenario, selection is the dominating force determining the evolution of the type distribution in the population, and random genetic drift is too slow to compete with it (this is why such a regime is sometimes called 'strong selection').

Interesting (and stochastic) things happen if N is big and s is small 'in the right way': Assume that $N \gg 1$ and $s = \sigma/N$. This is called a *weak selection* regime; of course, it will depend on the particular application if it applies. Denoting again the fraction of type a-individuals at time t by Z_t, we see now that

$$\mathbb{E}\big[\phi(Z_{t+\Delta t}) - \phi(Z_t)\big|Z_t\big]$$
$$\approx \Delta t \times Z_t(1-Z_t)\Big(\frac{\sigma}{N}\phi'(Z_t) + \frac{1+\sigma/2N}{N}\phi''(Z_t)\Big) \approx \frac{\Delta t}{N} Z_t(1-Z_t)\Big(\sigma\phi'(Z_t) + \phi''(Z_t)\Big).$$

This suggests that if we now speed up time by N again, (Z_t) will converge to the solution of

$$dZ_t = \sigma Z_t(1-Z_t)dt + \sqrt{2Z_t(1-Z_t)}dW_t,$$

a Wright-Fisher diffusion with (directional) selection. We will see a general version of this fact below.

1.4.2 Donnelly & Kurtz' modified 'look down' construction for the Fleming-Viot process with mutation and selection

We follow [DK99] (in a sometimes overview-like manner). The key point will be to construct versions of the Moran model in such a way that in the limit of infinite particle number, there are still individual particles to talk about.

Moran model with mutation and selection

The N-particle model, where we have already implicitly speeded up time by N, looks as follows:

Individuals have a type in E (where $E \subset \mathbb{R}$, say). Mutation is governed by a Markov process in E with bounded generator B, we assume that it can be written in the form

$$Bf(x) = \alpha \int_0^1 \big(f(h(x,u)) - f(x) \big) du \qquad (1.46)$$

for some $\alpha \geq 0$ and $h : E \times [0,1] \to E$. Any pure jump Markov process with globally bounded jump rates can written in this form, which will be convenient for us later on.

Let $\tilde{X}_i^{(N)}(t)$ be the type of the i-th individual, and $\tilde{Z}_t^{(N)} := N^{-1} \sum_{i=1}^N \delta_{\tilde{X}_i^{(N)}(t)}$ the empirical type distribution at time t.

There are two types of reproductive or resampling events: 1) Any individual reproduces at rate $N/2$, in which case it replaces an individual chosen at random by a copy of its present type. These are the so-called neutral resamplings. Note that a 'silent' event where an individual replaces itself is also allowed. 2) Let $\sigma : E \times \mathcal{M}_1(E) \to \mathbb{R}_+$ be a bounded measurable function. $\sigma(x, \mu)$ models the the fitness advantage of a type x-individual if the present type composition is given by μ. In addition to 1), each individual i produces additional offspring at instantaneous rate $\sigma(\tilde{X}_i^{(N)}(t-), \tilde{Z}_{t-}^{(N)})$, which replaces another individual chosen at random. We will call these events 'selective resamplings'. We are considering 'fecundity selection' here: the number of offspring depends on the type. Note that another popular type of selection is 'viability selection', where a particle's death rate is a function of the type.

Remark 4. There are two important special cases of this type of selection:

1.) $\sigma(x, \mu) = \tilde{\sigma}(x)$ independent of μ: genic selection (sometimes also: directional selection).

2.) The 'usual diploid model': $\sigma(x, \mu) = \int \tilde{\sigma}(x, y) \mu(dy)$ for some $\tilde{\sigma} : E \times E \to \mathbb{R}_+$. Rationale is as follows: fitness depends on a diploid type, but we are following gametes in a Hardy-Weinberg situation. The probability that a gametes of type x produces an individual of type (x, y) if the present type configuration in the (gamete pool of the) population is μ is given by $\mu(dy)$.

Note: a certain form of balancing selection (i.e. selection which favors a less frequent type) can be fits into this framework: $E = \{0, 1\}$, $\tilde{\sigma}(x, y) = 1 - \delta_{xy}$, so if the present frequency of type 0 is a, the fitness of type 0 is $1 - a$.

For fixed N, the vector of types $\big(\tilde{X}_1^{(N)}(t), \dots, \tilde{X}_N^{(N)}(t) \big)$ evolves according to a Markov process on E^N, which we can for example describe as the solution to the martingale problem with generator ($\underline{x} = (x_1, \dots, x_N)$, $\mu = N^{-1} \sum_{i=1}^N \delta_{x_i}$)

$$A_0^N f(\underline{x}) = \sum_{i=1}^N B_i f(\underline{x}) + \frac{1}{2} \sum_{1 \leq i,j \leq N} \Big(1 + \frac{2}{N} \sigma(x_i, \mu) \Big) \Big(f(\eta_j(\underline{x}; x_i)) - f(\underline{x}) \Big), \qquad (1.47)$$

where $B_i f(\underline{x})$ denotes B acting on the i-th coordinate of f, and $\eta_j(\underline{x}; x_i)$ is $\underline{x}$ with the value of the j-th coordinate replaced by x_i. Existence and uniqueness of the martingale problem with generator (1.47), and domain $\{ f : E^N \to \mathbb{R} \text{ cont., bounded} \}$, say, are standard: A_0^N is a bounded operator.
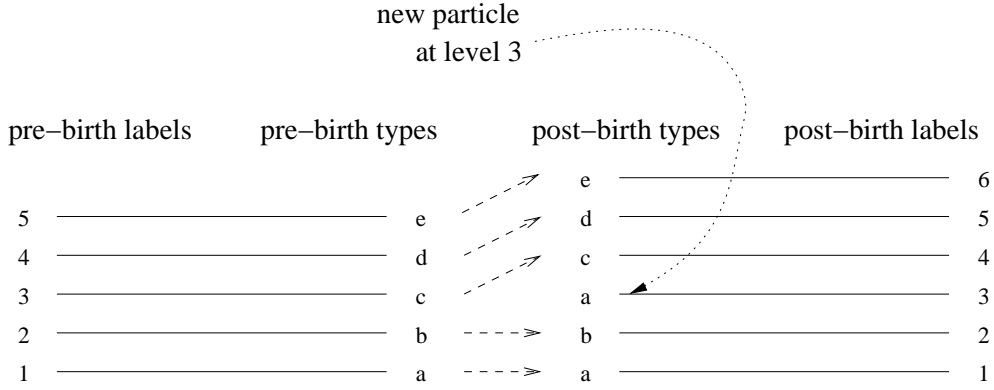
Figure 1.3: The particle at level 1 produces an offspring at level 3, the other types are shifted accordingly.

An ordered version

N, B, σ as above. Let $(X_i^{(N)}(t), i = 1, \ldots, N)_{t \geq 0}$ be the solution of the martingale problem with generator

$$
\begin{aligned}
A^N f(\underline{x}) \;=\; & \sum_{i=1}^{N} B_i f(\underline{x}) + \sum_{1 \leq i < j \leq N} \Big(f(\theta_j(\underline{x}; x_i)) - f(\underline{x}) \Big) \\
& + \sum_{i=1}^{n} \sigma(x_i, \mu) \frac{1}{N} \sum_{j=1}^{n} \Big(f(\eta_j(\underline{x}; x_i)) - f(\underline{x}) \Big),
\end{aligned}
\tag{1.48}
$$

where $\theta_j(\underline{x}; z) \in E^N$ is defined as follows:

$$
\big(\theta_j(\underline{x}; z)\big)_k = \begin{cases} x_k & k < j \\ z & k = j \\ x_{k-1} & k > j. \end{cases}
\tag{1.49}
$$

See Figure 1.4.2. The labelling is such that always the individual with the highest index dies next. This may seem strange (or perfectly adequate for Western civilisation with its broken relation to death ;-), the rationale behind this numbering is as follows: 1) the empirical distribution of $X^{(N)}$ and $\tilde{X}^{(N)}$ is the same (see Proposition 1 below), and 2) as $N \to \infty$, $X^{(N)}$ converges to an E^∞-valued limit process, whereas the process $\tilde{X}^{(N)}$ has no limit (only its empirical process $\tilde{Z}$ converges). Let $Z_t^{(N)} := (1/N) \sum_{i=1}^{N} \delta_{X_i^{(N)}(t)}$.

Proposition 1. *Assume that $X^{(N)}(0)$ and $\tilde{X}^{(N)}(0)$ have the same,* exchangeable *distribution. Then $(Z_t)_{t \geq 0}$ and $(\tilde{Z}_t)_{t \geq 0}$ have the same distribution. Furthermore, for each $t \geq 0$, $X^{(N)}(t)$ and $\tilde{X}^{(N)}(t)$ have the same exchangeable distribution.*

Remark 5. In general, $X^{(N)}$ and $\tilde{X}^{(N)}$ as processes have different distributions.

Proof. We fix N and drop the superscript N throughout the proof. We will define a permutation-valued process $(\theta(t))_{t \geq 0}$ such that $\theta(t)$ is uniform on S_N for each t,

$$
\big(Y_1(t), \ldots, Y_N(t)\big) := \big(X_{\theta_1(t)}, \ldots, X_{\theta_N(t)}\big)
\tag{1.50}
$$

is a version of $\tilde{X}$, and for each $t \geq 0$, $\theta(t)$ is independent from $\sigma(Y(s) : s \leq t)$.

Let $(I_m, J_m)_{m=1,2,\ldots}$ be i.i.d., uniform on $\{(i,j) : 1 \leq i < j \leq N\}$, $A_1, A_2, \ldots$ an independent sequence of fair coin tosses, and π_0 independent, uniform on S_N. (Interpretation: (I_m, J_m)

governs the indices involved in the m-th look-down event.) Define inductively

$$\tilde{I}_m := \pi_{m-1}^{-1}(I_m), \ \tilde{J}_m := \pi_{m-1}^{-1}(N) \tag{1.51}$$

and set

$$
\begin{aligned}
\pi_m(\tilde{I}_m) &:= \mathbf{1}_{A_m=1}I_m + \mathbf{1}_{A_m=0}J_m, \quad \pi_m(\tilde{J}_m) := \mathbf{1}_{A_m=1}J_m + \mathbf{1}_{A_m=0}I_m, \quad \text{for } k \neq \tilde{I}_m, \tilde{J}_m \\
\pi_m(k) &:= \begin{cases} \pi_{m-1}(k) & \text{if } \pi_{m-1}(k) < J_m, \\ \pi_{m-1}(k)+1 & \text{if } \pi_{m-1}(k) \geq J_m. \end{cases}
\end{aligned}
$$
$$\tag{1.52}$$

In words, π_m maps $\{\tilde{I}_m, \tilde{J}_m\}$ to $\{I_m, J_m\}$ in randomised order, and on $\{1, 2, \ldots, N\} \setminus \{\tilde{I}_m, \tilde{J}_m\}$, it retains the ordering of π_{m-1}. We claim that for each m

$$(\tilde{I}_1, \tilde{J}_1), \ldots, (\tilde{I}_m, \tilde{J}_m), \pi_m \text{ are independent and uniform on their corresponding set of values.} \tag{1.53}$$

As π_m depends only on π_{m-1} and (I_m, J_m, A_m), it suffices to prove (1.53) for $m = 1$. Let $\rho \in S_N$, $i, j \in \{1, \ldots, N\}$, $i \neq j$. If $\rho(i) < \rho(j)$, we have

$$
\begin{aligned}
\mathbb{P}(\tilde{I}_1 = i, \tilde{J}_1 = j, \pi_1 = \rho) &= \frac{1}{2}\mathbb{P}\big(\ I_1 = \rho(i), J_1 = \rho(j), \pi_0(j) = n, \text{ for } k \neq j: \\
&\qquad\qquad \pi_0(k) = \rho(k) \text{ if } \rho(k) < \rho(j), \ \pi_0(k) = \rho(k) - 1 \text{ if } \rho(k) > \rho(j)\big) \\
&= \frac{1}{2}\frac{2}{N(N-1)}\frac{1}{N!} = \frac{1}{N(N-1)}\frac{1}{N!}.
\end{aligned}
$$

Similarly, if $\rho(i) > \rho(j)$

$$
\begin{aligned}
\mathbb{P}(\tilde{I}_1 = i, \tilde{J}_1 = j, \pi_1 = \rho) &= \frac{1}{2}\mathbb{P}\big(\ I_1 = \rho(j), J_1 = \rho(j), \pi_0(j) = n, \text{ for } k \neq j: \\
&\qquad\qquad \pi_0(k) = \rho(k) \text{ if } \rho(k) < \rho(i), \ \pi_0(k) = \rho(k) - 1 \text{ if } \rho(k) > \rho(i)\big) \\
&= \frac{1}{2}\frac{2}{N(N-1)}\frac{1}{N!} = \frac{1}{N(N-1)}\frac{1}{N!}.
\end{aligned}
$$

Independent ingredients for coupling: $T_1 < T_2 < \cdots$ jump times of a rate $N(N-1) + N\bar{\sigma}$ Poisson process on $\mathbb{R}_+$. $U_1, U_2, \ldots,\ V_1, V_2, \ldots,$ i.i.d. uniform$([0,1])$. $(I_1, J_1), (I_2, J_2), \ldots$ independent, uniform on $\{(i, j) : 1 \leq i < j \leq N\}$. $(I_1', J_1'), (I_2', J_2'), \ldots$ independent, uniform on $\{(i, j) : 1 \leq i \neq j \leq N\}$. $M_1, \ldots, M_N$ independent Poisson processes on $[0, 1] \times \mathbb{R}_+$. with intensity measure $\alpha \times$Lebesgue measure. π_0 uniform on S_N. $(X_1(0), \ldots, X_N(0))$ exchangeably distributed.

We put $Z(t) := N^{-1}\sum_{i=1}^N \delta_{X_i(t)}$. Set $\theta(t) := \pi_0$ for $t \in [0, T_1)$. Between the jump times T_i, each coordinate evolves according to the mutation process, i.e. for $t \in [T_{m-1}, T_m)$

$$X_i(t) = X_i(T_m) + \int_{(T_m, t] \times [0,1]} h(X_i(s-), u) - X_i(s-) \, M_i(du\,ds).$$

The jump times T_m correspond to (potential) resampling events. If $U_m \leq N(N-1)/(N(N-1+\bar{\sigma}))$, the m-th event is a neutral resampling event, otherwise it is a selective resampling event.

Assume that at time T_m, a neutral resampling event takes place. Then in the ordered model, individual I_m has an offspring which is placed at position J_m, and the individual with label N dies, i.e.

$$X_i(T_m) := \begin{cases} X_i(T_m-) & \text{if } i < J_m, \\ X_{I_m}(T_m-) & \text{if } i = J_m, \\ X_{i-1}(T_m-) & \text{if } i < J_m. \end{cases}$$

In order to keep track of the unordered model too, define $\tilde{I}_m, \tilde{J}_m, \pi_m$ using (π_{m-1}, I_m, J_m) as in (1.51), (1.52), and put $\theta(t) := \pi_m$ for $t \in [T_m, T_{m+1})$.

Note that by definition, this resampling event looks as follows for the process $Y(t)$: individual $\tilde{J}_m$ dies and is replaced by an offspring of individual $\tilde{I}_m$. Also note that by (1.53), $(\tilde{I}_m, \tilde{J}_m)$ is uniform on $\{(i,j) : 1 \le i \ne j \le N\}$.

If the m-th event is a selective resampling, we put $\theta(t) := \pi_m$ for $t \in [T_m, T_{m+1})$ (the selective mechanism is the same in the ordered and in the unordered model). If $V_m \le \sigma(X_{I'_m}(T_m-), Z(T_m-))/\bar{\sigma}$, the potential resampling is successful, and an offspring of individual I'_m is placed at level J'_m:

$$X_i(T_m) := \begin{cases} X_i(T_m-) & \text{if } i \ne J'_m, \\ X_{I'_m}(T_m) & \text{if } i = J'_m. \end{cases}$$

Otherwise, we leave X unchanged: $X_i(T_m) = X_i(T_m-)$.

By construction, $(Y(t))_{t \ge 0}$ evolves according to the unordered model. As $Y(t)$ is a (random) permutation of $X(t)$, the empirical distribution processes agree. Furthermore, the evolution of Y can be simulated if one knows $Y(0)$, the $(\tilde{I}_m, \tilde{J}_m)_m$ and the driving Poisson processes and the auxiliary coin tosses, in particular without knowing π_0 and the $(I_m, J_m)_m$. Thus for a given t, by considering m such that $t \in [T_m, T_{m+1})$, we see from (1.53) that $\theta(t)$ is independent from $\sigma(Y(s), s \le t)$. As $Y(t)$ is exchangeable because of the symmetries of the unordered model and $\theta^{-1}(t)$ is also independent of $Y(t)$, we see that

$$\big(X_i(t), i = 1, \dots, N\big) = \big(Y_{\theta_i^{-1}(t)}(t), i = 1, \dots, N\big) \stackrel{d}{=} \big(Y_i(t), i = 1, \dots, N\big),$$

in particular $X(t)$ is exchangeable. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In the following, we assume that the selection function $\sigma(x, \mu)$ is given by

$$\sigma(x, \mu) = \int_E \sigma(x, y) \, \mu(dy) \qquad\qquad (1.54)$$

for some bounded, measurable function $\sigma : E \times E \to \mathbb{R}_+$ (with a small abuse of notation). This allows to rewrite $\sigma(X_i^N(t), Z^N(t)) = N^{-1} \sum_{k=1}^N \sigma(X_i^N(t), X_k^N(t))$.

Let us consider an infinite system of Poisson-process driven SDEs which corresponds to the limit $N \to \infty$ of the ordered model (the gist of the ordering is that such a limit system exists): Let $\tilde{\rho} : \mathcal{M}_1(E) \times [0,1] \to E$ be such that $\mathscr{L}(\tilde{\rho}(\mu, U)) = \mu$ if $U \sim \text{unif}([0,1])$. Let $0 \le \sigma(\cdot, \cdot) \le \bar{\sigma}$. $L_{ij}(t)$, $1 \le i < j$ independent rate 1 Poisson processes on $\mathbb{R}_+$, M_i, $i = 1, 2, \dots$ independent rate α Poisson processes on $[0,1] \times \mathbb{R}_+$, K_j independent Poisson processes on $[0,1]^3 \times \mathbb{R}_+$ with intensity $\bar{\sigma} \times (\text{Lebesgue measure})^4$.

$$\begin{aligned}
X_j(t) \;=\; & X_j(0) + \int_{[0,1] \times (0,t]} h(X_j(s-), u) - X(s-) \, M_j(du, ds) && (1.55) \\
& + \sum_{i<j} \int_{(0,t]} X_i(s-) - X_j(s-) \, L_{ij}(ds) + \sum_{1 \le i < k < j} \int_{(0,t]} X_{j-1}(s-) - X_j(s-) \, L_{ij}(ds) \\
& + \int_{[0,1]^3 \times (0,t]} \big(\tilde{\rho}(Z_{s-}, u_1) - X_j(s-)\big) \mathbf{1}\big(\sigma(\tilde{\rho}(Z_{s-}, u_1), \tilde{\rho}(Z_{s-}, u_2)) \ge \bar{\sigma} u_3\big) K_j(du_1, du_2, du_3, ds).
\end{aligned}$$

Interpretation of terms: initial value, individual mutation, birth of an offspring of i at level j, birth at $k < j$ effecting a shift of one level of the type at j, selective birth event replacing the type at level j.

Note: in the neutral case ($\sigma(\cdot) \equiv 0$), if X is a solution of (1.55), then for any $N \in \mathbb{N}$, the E^N-valued process $(X_1(t), \dots, X_N(t))_{t \ge 0}$ is a version of the ordered model with N particles. In particular, by Proposition 1, $X(t)$ is exchangeable for each $t \ge 0$.

An immediate consequence of exchangeability of an E^∞-valued random vector X is that by de Finetti's Theorem (cf. e.g. [Kal02], Thm. 11.10), the empirical distribution $Z := \lim_{m \to \infty} (1/m) \sum_{i+1}^m \delta_{X_i}$ exists almost surely. The following will make heavy use of this fact.

The infinite system of SDEs

In the selective resamplings, we need to be able to generate a sample of a type according to the current population. While we can do this for a finite population by simply picking a label uniformly between 1 and N and looking at the type there, we need something else in the situation $N = \infty$. Furthermore, we would like to be able make sense out of the ancestry of the specimen sampled 'uniformly from $1, 2, \ldots$'. For this, we introduce the *neutral marker process*.

Let $E_0 := [0, 1]^\infty$. The (individual) neutral marker process evolves as follows: each coordinate is, independently of the others, after Exp(1)-distributed waiting times, replaced by a uniform draw from $[0, 1]$, so the generator is given by

$$B_M f(x) = \sum_{k=1}^\infty \int_0^1 f(\eta_k(x; u)) - f(x) \, du$$

for $f : E_0 \to \mathbb{R}$ (say, bounded, depending on finitely many coordinates). E_0 is a totally ordered set via lexicographical ordering: for $x, y \in E_0$ write $x \le y$ if there is a $k \in \mathbb{N}$ such that $x_1 = y_1, \ldots, x_{k-1} = y_{k-1}, x_k \le y_k$.

Write $\xi_j(t) = (\xi_{j1}(t), \xi_{j2}(t), \ldots)$ for the neutral marker at level j at time t. For definiteness, assume that $\xi_j(0)$, $j = 1, 2, \ldots$ are independent i.i.d. uniform($[0, 1]$) sequences. Let Ξ_{jl} be independent Poisson processes on $[0, 1] \times \mathbb{R}_+$ (with Lebesgue intensity), and let $\xi(t)$, $t \ge 0$ be the unique solution of

$$
\begin{aligned}
\xi_j(t) \;=\; & \xi_j(0) + \sum_l \int_{[0,1] \times (0,t]} (u - \xi_{jl}(s-)) e_l \, \Xi_{jl}(du \, ds) \\
& + \sum_{i=1}^{j-1} \int_{(0,t]} \big(\xi_i(s-) - \xi_j(s-)\big) L_{ij}(ds) \\
& + \sum_{i<k<j} \int_{(0,t]} \big(\xi_{j-1}(s-) - \xi_j(s-)\big) L_{ik}(ds)
\end{aligned}
\tag{1.56}
$$

($e_l \in E_\infty$ has 1 at the l-th coordinate, and 0 elsewhere). Note: system for ξ_j, $j \le N$, observed at finitely many coordinates of the marker process is trivially uniquely solvable. Then extend to $j \in \mathbb{N}$ and full marker process.

Observe that the Poisson point processes L_{ij} determine a 'neutral genealogy': For $t \ge 0$, let $N_k^t(s)$, $k \in \mathbb{N}$, $0 \le s \le t$ be the level of the (neutral) ancestor at time s of the individual at level k at time t. N_k^t is the (unique) solution of

$$
\begin{aligned}
N_k^t(s) \;=\; & k - \sum_{1 \le i < j < k} \int_{(s,t]} \mathbf{1}(N_k^t(u) > j) \, L_{ij}(du) \\
& - \sum_{1 \le i < j \le k} \int_{(s,t]} (j - i) \mathbf{1}(N_k^t(u) = j) \, L_{ij}(du).
\end{aligned}
\tag{1.57}
$$

Inspection of the transition rates shows that for any N and t, the equivalence relation valued process $R^N(v) := \{(k, l) : 1 \le k, l \le N, N_k^t(t - v) = N_l^t(t - v)\}$, $0 \le v \le t$ is a (left continuous) version of Kingman's N-coalescent. In fact, the same can be defined for $N = \infty$. In particular, because Kingman's coalescent comes down from infinity, the set of all neutral ancestors at time s of the particles alive at time $t > s$,

$$\Gamma(s, t) = \{N_k^t(s) : k \in \mathbb{N}\} \quad \text{is finite a.s.} \tag{1.58}$$

The total order on E_0 and the fact that (almost surely) $\xi_i(t) \neq \xi_j(t)$ for all $i \neq j$ allows to uniquely define the (N-th approximate) p-quantiles ($p \in (0,1)$): $k_N(p,t) \in \{1, \ldots, N\}$ with the property

$$\begin{aligned} \#\big\{1 \leq k \leq N : \xi_k(t) \leq \xi_{k_N(p,t)}(t)\big\} &\geq [Np] + 1 \quad \text{and} \\ \#\big\{1 \leq k \leq N : \xi_k(t) \geq \xi_{k_N(p,t)}(t)\big\} &\geq N - [Np]. \end{aligned}$$

The following lemma shows how to use this to generate a uniform sample from the infinite population at time t in such a way that the ancestry of the sampled individual is identifiable for any $s < t$:

Lemma 4. *Let V be* uniform$([0,1])$-*distributed, independent of ξ. Then*

$$\xi_\infty(V,t) := \lim_{N \to \infty} \xi_{k_N(V,t)}(t) \quad \text{exists a.s.}$$

(in the sense that for any l there exists $n_l < \infty$ such that $\xi_{k_N(V,t),l}(t) = \xi_{k_{N'}(V,t),l}(t)$ for all $N, N' > n_l$). Furthermore, for $s < t$,

$$N^t(V,s) := \lim_{N \to \infty} N^t_{k_N(V,t)}(s) \quad \text{exists a.s.}$$

(and defines a path through the neutral genealogy, ending at $N^t(V,t-) = \infty$).

Proof. Note that $Z_\xi(t) := \lim_{N \to \infty} (1/N) \sum_{i=1}^N \delta_{\xi_i(t)}$ exists almost surely by exchangeability. Because $|\Gamma(0,t)| < \infty$ by (1.58) and the set of atoms of $\cup_{jl} \Xi_{jl}$ in $[0,t]$ is countable, the empirical measure of the first m coordinates of the marker process,

$$Z_\xi^m(t) := \lim_{N \to \infty} (1/N) \sum_{i=1}^N \delta_{(\xi_{i1}(t), \ldots, \xi_{im}(t))}$$

is almost surely purely atomic. As the lexicographic order on E_0 induces the lexicographic order on $[0,1]^m$, for N sufficiently large, $\big(\xi_{k_N(V,t),1}(t), \ldots, \xi_{k_N(V,t),m}(t)\big)$ will fix on one of the atoms of $Z_\xi^m(t)$. As m is arbitrary, this proves that $\xi_\infty(V,t)$ exists.

In order to identify the genealogy of the sampled 'individual', consider a marker coordinate l and let $s_l(V)$ be the time when the value $\xi_{\infty,l}(V)$ appeared first: $s_l(V) = 0$ if this value was already present in the population at time zero, if $s_l(V) > 0$, this value was created in a mutation (of the marker process) which took place at time $s_l(V)$ at some level $k_l(V)$. Note that if $s_l(V) > 0$ and N is so large that $\xi_{k_N(V,t),l}(t) = \xi_{\infty,l}(V,t)$, then necessarily $N^t_{k_N(V,t)}(s_l(V)) = k_l(V)$ and hence $N^t_{k_N(V,t)}(s) = N^{s_l(V)}_{k_l(V)}(s)$ for all $s \in [0, s_l(V))$. In words, if the value of $\xi_{\infty,l}(V,t)$ was created at time $s_l(V) > 0$, we can identify the genealogy up to that time. So to prove the second claim, we wish to show that

$$\sup_{l \in \mathbb{N}} s_l(V) = t \quad \text{a.s.} \tag{1.59}$$

Let $s_l^N(V)$ be the time when the value $\xi_{k_N(V,t),l}(t)$ was created. $k_N(V,t)$ is uniform on $\{1, \ldots, N\}$, independent of L_{ij} and Ξ_{jl}, so the random variables $t - s_1^N(V), t - s_2^N(V), \ldots$ are i.i.d. with distribution $e^{-r}\mathbf{1}(0 < r < t)dr + e^{-t}\delta_t$. As $\lim_N s_l^N(V) = s_l(V)$ a.s. the same is true for $t - s_1(V), t - s_2(V), \ldots$, proving (1.59). $\square$

For an E^∞-valued (càdlàg) process Y, $v \in (0,1)$ put

$$\rho(Y,t,v) := \begin{cases} \lim_{s \nearrow t} Y_{N^t(v,s)}(s) & \text{if the limit exists} \\ x_0 & \text{otherwise} \end{cases} \tag{1.60}$$

(where $x_0 \in E$ is some arbitrary value).

Let $0 \leq \sigma(\cdot, \cdot) \leq \bar{\sigma}$. $L_{ij}(t)$, $1 \leq i < j$ independent rate 1 Poisson processes on $\mathbb{R}_+$, M_i, $i = 1, 2, \ldots$ independent rate α Poisson processes on $[0,1] \times \mathbb{R}_+$, K_j independent Poisson processes

on $[0,1]^3 \times \mathbb{R}_+$ with intensity $\bar{\sigma} \times (\text{Lebesgue measure})^4$ (all of them adapted to a given filtration $\mathscr{F} = (\mathscr{F}_t)$). We will sometimes write $\mathscr{F}^X$ for the filtration generated by a process X.

The following infinite system of Poisson-process driven SDEs is our 'target object' in this section, it is a way to describe the infinite population size limit of the Moran models from the previous subsection which retains individual particles:

$$
\begin{aligned}
X_j(t) \;=\; & X_j(0) + \int_{[0,1]\times(0,t]} h(X_j(s-),u) - X(s-)\, M_j(du,ds) \qquad\qquad (1.61) \\
& + \sum_{i<j} \int_{(0,t]} X_i(s-) - X_j(s-)\, L_{ij}(ds) + \sum_{1\le i<k<j} \int_{(0,t]} X_{j-1}(s-) - X_j(s-)\, L_{ij}(ds) \\
& + \int_{[0,1]^3\times(0,t]} \big(\rho(X,s,u_1) - X_j(s-)\big)\mathbf{1}\big(\bar{\sigma}u_3 \le \sigma(\rho(X,s,u_1),\rho(X,s,u_2))\big) K_j(du_1,du_2,du_3,ds).
\end{aligned}
$$

The terms appearing on the right-hand side can be interpreted as follows: initial value, individual mutation, birth of an offspring of i at level j, birth at $k < j$ effecting a shift of one level of the type at j, selective birth event replacing the type at level j.

Theorem 5. *Assume that $X(0)$ is exchangeable, and independent of the driving Poisson processes in (1.61). There is a unique solution X of (1.61). For each $t \ge 0$, $X(t)$ is exchangeable, thus $Z(t) = \lim_{m\to\infty}(1/m)\sum_{i=1}^m \delta_{X_i(t)}$ exists a.s., and for bounded, measurable $f : E^m \to \mathbb{R}$ we have*

$$
\mathbb{E}\big[f(X_1(t),\dots,X_m(t))\big|\mathscr{F}_t^Z\big] = \langle f, Z(t)^{\otimes m}\rangle \qquad a.s.
$$

Remark 6. Note that we have not claimed (and will not prove) path properties of $Z(t)$. It is known that there is a continuous version of $(Z(t))_{t\ge 0}$.

For the proof of Thm. 5, we will need the following lemma.

Lemma 5. *Let a (measurable) function $\tilde{\rho} : \mathcal{M}_1(E) \times [0,1] \to E$ be given with the property that $\mathscr{L}(\tilde{\rho}(\mu,V)) = \mu$ for any $\mu \in \mathcal{M}_1(E)$ if $V \sim \text{uniform}([0,1])$. Let $V(t)$ be an (adapted, càdlàg) E^∞-valued process such that $\gamma(V(t-)) := \lim_m (1/m)\sum_{i=1}^m \delta_{V_i(t-)}$ exists a.s. for any $t \ge 0$. Let $Y(0)$ be exchangeable, and Y (the unique) solution of*

$$
\begin{aligned}
Y_j(t) \;=\; & Y_j(0) + \int_{[0,1]\times(0,t]} h(Y_j(s-),u) - Y(s-)\, M_j(du,ds) \qquad\qquad (1.62) \\
& + \sum_{i<j} \int_{(0,t]} Y_i(s-) - Y_j(s-)\, L_{ij}(ds) + \sum_{1\le i<k<j} \int_{(0,t]} Y_{j-1}(s-) - Y_j(s-)\, L_{ij}(ds) \\
& + \int_{[0,1]^3\times(0,t]} \big(\tilde{\rho}(\gamma(V(s-)),u_1) - Y_j(s-)\big) \\
& \qquad\qquad \times \mathbf{1}\big(\bar{\sigma}u_3 \le \sigma(\tilde{\rho}(\gamma(V(s-)),u_1),\tilde{\rho}(\gamma(V(s-))u_2))\big) K_j(du_1,du_2,du_3,ds).
\end{aligned}
$$

Then for $t \ge 0$, $Y(t)$ is exchangeable, and for bounded, measurable $f : E^m \to \mathbb{R}$,

$$
\mathbb{E}\big[f(Y_1(t),\dots,Y_m(t))\big|\mathscr{F}_t^{\gamma(Y),\gamma(V)}\big] = \gamma(Y(t))^m. \qquad\qquad (1.63)
$$

Furthermore, for any $t > 0$, the limit in (1.60) defining $\rho(Y,t,v)$ exists for almost all $v \in [0,1]$. If $U \sim \text{unif}([0,1])$, independent of the other ingredients, then

$$
\mathscr{L}(\rho(Y,t,U)|\mathscr{F}_t) = \lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^m \delta_{Y_i(t-)} \qquad a.s.
$$

(in particular, the limit on the right-hand side exists a.s.)

Proof. If $K_j \equiv 0$, this is the exchangeability of the neutral model, which follows e.g. from Proposition 1, because then $(Y_1(t), \ldots, Y_N(t))$ has the same distribution as the corresponding unordered system of size N. Consider $\epsilon > 0$. K_j^ϵ is 'discretised K_j':

$$K_j^\epsilon(du_1\, du_2\, du_3 \times \{k\epsilon\}) = K_j(du_1\, du_2\, du_3 \times (k\epsilon, (k+1)\epsilon])\mathbf{1}(K_j([0,1]^3 \times (k\epsilon, (k+1)\epsilon]) \le 1).$$

Y^ϵ solution of (1.62) with K replaced by K^ϵ. $Y^\epsilon(t)$ is exchangeable for $t \in [0, \epsilon)$, the equivalent of (1.63) holds for such t. At time $t = \epsilon$: Each level j throws an independent coin with success probability $\mathbb{P}(K_j([0,1]^3 \times (k\epsilon, (k+1)\epsilon] = 1)$. Each 'successful' level j draws U_{j1}, U_{j2}, U_{j3} (indep., uniform($[0,1]$)). If $\sigma(\rho(V(\epsilon-), U_1, \epsilon), \rho(V(\epsilon-), U_1, \epsilon)) \ge \bar{\sigma}U_3$, we (have to) set $Y_j(\epsilon) := \rho(V(\epsilon-), U_1, \epsilon)$, otherwise $Y_j(\epsilon) = Y_j(\epsilon-)$. This mechanism preserves exchangeability and shows that $Y^\epsilon(\epsilon)$ again satisfies (1.63). Repeat at $t = 2\epsilon$, etc. Furthermore $Y^\epsilon \to Y$ (at least in the sense that $(Y_1^\epsilon(t), \ldots, Y_m^\epsilon(t)) \to_{\epsilon \searrow 0} (Y_1(t), \ldots, Y_m(t))$), so $Y(t)$ is also exchangeable.

It remains to prove that the limit in (1.60) defining $\rho(Y, t, U)$ exists and has the correct conditional distribution. Fix $t > 0$ for the moment. Let

$$\tilde{Y}_v^t(s) := Y_{N^t(v,s)}(s), \ 0 \le s < t$$

be the (neutral) ancestral type process of the 'individual' corresponding to the v-quantile of the neutral markers at time t (which is well defined for almost all $v \in (0,1)$ by Lemma 4). For $C, C_i \subset [0,1]$, $r < t$ put

$$\tilde{M}_v^t(C \times (0,r]) := \sum_{j=1}^\infty \int_{[0,1] \times (0,r]} \mathbf{1}_C(u) \mathbf{1}(N^t(v,s) = j) M_j(du\, ds),$$

$$\tilde{K}_v^t(C_1 \times C_2 \times C_3 \times (0,r]) := \sum_{j=1}^\infty \int_{[0,1]^3 \times (0,r]} \mathbf{1}_{C_1}(u_1) \mathbf{1}_{C_2}(u_2) \mathbf{1}_{C_3}(u_3) \mathbf{1}(N^t(v,s) = j) K_j(du_1\, du_2\, du_3\, ds).$$

Then $\tilde{M}_v^t$ is a Poisson process on $[0,1] \times [0,t]$ with the same distribution as the M_js (but of course not independent), and similarly $\tilde{K}_v^t$ is a Poisson process on $[0,1]^3 \times [0,t]$ with the same characteristics as K_j: We are 'piecing' together the M_j resp. K_j according to $N^t(v, \cdot)$, which was read off from the L_{ij} (and $(L_{..})$ is independent from $M_., K_.$). Furthermore we have for $r < t$ by construction

$$\tilde{Y}_v^t(r) = \tilde{Y}_v^t(0) + \int_{[0,1] \times (0,r]} h(\tilde{Y}_v^t(s-), u) - \tilde{Y}_v^t(s-)\, \tilde{M}_v^t(du\, ds)$$

$$+ \int_{[0,1]^3 \times (0,r]} \left(\tilde{\rho}(\gamma(V(s-)), u_1) - \tilde{Y}_v^t(s-)\right)$$
$$\times \mathbf{1}\left(\bar{\sigma}u_3 \le \sigma(\tilde{\rho}(\gamma(V(s-)), u_1), \tilde{\rho}(\gamma(V(s-))u_2))\right) K_j(du_1, du_2, du_3, ds). \tag{1.64}$$

Let
$$\tilde{\tau}_v^t := \sup\{0 \le s < t : \tilde{M}_v^t([0,1] \times \{s\}) > 0 \text{ or } \tilde{K}_v^t([0,1]^3 \times \{s\}) > 0\}$$
be the time of the last jump of $\tilde{M}_v^t$ or $\tilde{K}_v^t$ before t, then $\tilde{\tau}_v^t < t$ a.s. We see from (1.64) that $\tilde{Y}_v^t(r)$ is constant for $r \in [\tilde{\tau}_v^t, t)$, so that $\lim_{r \nearrow t} \tilde{Y}_v^t(r)$ exists.

Let V be uniform($[0,1]$)-distributed, independent of everything else. To check the claim about the distribution of $\rho(Y, t, V)$ put

$$\gamma_V^{t,N} := \sup\{s \ge 0 : N_{k_N(V,t)}^t(s) = N^t(V, s)\}.$$

Note that $\gamma_v^{t,N} \to t$ as $N \to \infty$ by Lemma 4. Furthermore,

$$\mathbb{P}\left(Y_{N_{k_N(V,t)}^t}(t)(t) \ne \lim_{r \nearrow t} \tilde{Y}_V^t(r)\right) \le \mathbb{P}(\gamma_v^{t,N} < \tilde{\tau}_v^t)$$

$$+ \mathbb{P}\left(\begin{array}{l}\text{there is a mutation or selective resampling on} \\ \text{the ancestral line of } Y_{N_{k_N(V,t)}^t(\cdot)}(\cdot) \text{ in } (\gamma_v^{t,N}, t]\end{array}\right)$$

$$\to \quad 0 \quad \text{as } N \to \infty.$$

As $N^t_{k_N(V,t)}(t) = k_N(V,t)$ is uniform($\{1,2,\ldots,N\}$) by construction, we obtain

$$\mathbb{P}(\lim_{r\nearrow t}\tilde{Y}^t_V(r) = a|\mathscr{F}_t) = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \mathbf{1}(Y_i(t) = a) \quad \text{a.s.}$$

as required. $\qquad\square$

Remark 7. Note that both mutations and (potential) selective resamplings are 'perturbations' of the 'pure look-down process' in that each level is hit only by finitely many such events in any finite time interval. Furthermore, these events have the potential to create a 'new' type which differs from the one we would see of only the look down-dynamics were acting. Let

$$\Lambda := \big\{(j,s) \;:\; j \in \mathbb{N},\, s \text{ is the 'time' coordinate of an atom of } M_j \text{ or } K_j\big\}$$

(note: these are the points of a Poisson process on $\mathbb{N} \times \mathbb{R}_+$ with intensity $\alpha\bar{\sigma}$ Lebesgue $\otimes$ counting measure). Fix $t > 0$ for the moment. Put

$$\tau^t_k = \sup\{s < t : (N^t_k(s), s) \in \Lambda\} \vee 0,$$

this is the time point $u \in [0,t)$ when the type which is at level k at time t was (potentially) created, either by a mutation or by a selective resampling event. Note that $X_k(t) = X_{N^t_k(\tau^t_k)}$ ($= X_{N^t_k(u)}$ for $u \in [\tau^t_k, t]$) if X is a solution of (1.61). Analogously, $Y_k(t) = Y_{N^t_k(\tau^t_k)}$ for the solution Y of (1.62).

We may view the points on Λ as the mutation events in (the lookdown construction of) an infinitely-many-alleles model, thus for any $(j,s) \in \Lambda$,

$$a(j,s,t) = \lim_{m\to\infty}\frac{1}{m}\sum_{i=1}^m \mathbf{1}\big((N^t_i(\tau^t_i), \tau^t_i) = (j,s)\big) \tag{1.65}$$

exists almost surely. $a(j,s,t)$ is the fraction of that type in population at time t which was founded at time s (if there exists a j such that $(j,s) \in \Lambda$. Note that this allows to express the empirical measure at time t as

$$Z(t) = \sum_{\substack{(j,s)\in\Lambda \\ s<t}} a(j,s,t)\delta_{X_j(s)}.$$

Proof of Thm. 5. We use Picard iteration to show existence. Let $X^{(0)}$ be the neutral model, i.e. the solution of (1.61) with $K_j \equiv 0$. For $n \in \mathbb{Z}_+$, let $X^{(n+1)}$ be the solution of

$$
\begin{aligned}
X^{(n+1)}_j(t) \;=\;\; & X_j(0) + \int_{[0,1]\times(0,t]} h(X^{(n+1)}_j(s-), u) - X^{(n+1)}(s-)\,M_j(du, ds) \\
& + \sum_{i<j}\int_{(0,t]} X^{(n+1)}_i(s-) - X^{(n+1)}_j(s-)\,L_{ij}(ds) \\
& + \sum_{1\le i<k<j}\int_{(0,t]} X^{(n+1)}_{j-1}(s-) - X^{(n+1)}_j(s-)\,L_{ij}(ds) \\
& + \int_{[0,1]^3\times(0,t]} \big(\rho(X^{(n)}, s, u_1) - X^{(n+1)}_j(s-)\big) \\
& \qquad \times \mathbf{1}\big(\bar{\sigma}u_3 \le \sigma(\rho(X^{(n)}, s, u_1), \rho(X^{(n)}, s, u_2))\big)K_j(du_1, du_2, du_3, ds).
\end{aligned}
\tag{1.66}
$$

Existence and uniqueness of $X^{(n+1)}$, given $X^{(n)}$ is standard: we can solve 'level by level'. Note that $\rho(X^{(n)}, s, u)$ exists almost surely by Lemma 5.

Ancestral type process: Fix $T > 0$ for the moment. Let $\tilde{X}_j^{T,n}(t) := X_{N_j^T(t)}^{(n)}(t)$ be the type of the neutral ancestor at time t of the particle at level j at time $T \, (\geq t)$. Define

$$\tilde{M}_j^T(C \times [0,t]) \quad := \quad \sum_{i=1}^{j} \int_{C \times [0,t]} \mathbf{1}(N_j^T(s) = i) M_i(du\, ds) \qquad (1.67)$$

$$\tilde{K}_j^T(C_1 \times C_2 \times C_3 \times [0,t]) \quad := \quad \sum_{i=1}^{j} \int_{C_1 \times C_2 \times C_3 \times [0,t]} \mathbf{1}(N_j^T(s) = i) K_i(du_1\, du_2\, du_3\, ds) (1.68)$$

These are the driving mutation resp. selective resampling process observed 'along' the neutral ancestral line of the particle at level j at time T. By construction, $\tilde{M}_j^T$ and $\tilde{K}_j^T$ are Poisson processes (with the same characteristics as M_j resp. K_j, but not independent for different j). Then (1.66) and the definition of $N_j^T(s)$ imply that for $0 \leq t \leq T$, $\tilde{X}_j^{T,n+1}(t)$ is the solution of

$$\tilde{X}_j^{T,n+1}(t) \quad = \quad \tilde{X}_j^{T,n+1}(0) + \int_{[0,1] \times (0,t]} h(\tilde{X}_j^{T,n+1}(s-), u) - \tilde{X}_j^{T,n+1}(s-)\, \tilde{M}_j(du, ds) \qquad (1.69)$$

$$+ \int_{[0,1]^3 \times (0,t]} \left(\rho(X^{(n)}, s, u_1) - \tilde{X}_j^{T,n+1}(s-) \right)$$

$$\times \mathbf{1}\left(\bar{\sigma} u_3 \leq \sigma(\rho(X^{(n)}, s, u_1), \rho(X^{(n)}, s, u_2)) \right) \tilde{K}_j(du_1, du_2, du_3, ds).$$

Put $V_j^{T,n+1}(t) := \mathbf{1}\left(\tilde{X}_j^{T,n+1}(t) \neq \tilde{X}_j^{T,n}(t) \right)$. Note: $\tilde{M}, \tilde{K}$ do not depend on n, so $\tilde{X}_j^{T,n+1}(t) = \tilde{X}_j^{T,n}(t)$ as long as $\rho(X^{(n)}, s, u_1) = \rho(X^{(n-1)}, s, u_2)$ for all (u_1, u_2, u_3, s) in the support of $\tilde{K}_j$ with $s \leq t$. Hence

$$V_j^{T,n+1}(t) \quad \leq \quad \int_{[0,1]^3 \times (0,t]} \left(1 - V_j^{T,n+1}(s-) \right) \mathbf{1}\left(\rho(X^{(n)}, s, u_1) \neq \rho(X^{(n-1)}, s, u_1) \right.$$

$$\left. \text{or } \rho(X^{(n)}, s, u_2) \neq \rho(X^{(n-1)}, s, u_2) \right) \tilde{K}_j(du_1, du_2, du_3, ds)$$

$$= \quad \int_{[0,1]^3 \times (0,t]} (\dots) \left(\tilde{K}_j(du_1, du_2, du_3, ds) - \bar{\sigma} du_1\, du_2\, du_3\, ds \right)$$

$$+ \bar{\sigma} \int_{[0,1]^3 \times (0,t]} (\dots)\, du_1\, du_2\, du_3\, ds$$

$$\leq \quad \text{martingale}(t) + 2\bar{\sigma} \int_0^t \int_{[0,1]} \mathbf{1}(\rho(X^{(n)}, s, u) \neq \rho(X^{(n-1)}, s, u))\, du\, ds,$$

where the term 'martingale(t)' denotes an integral with respect to the compensated process, which is by well-known properties of Poisson processes in fact a bounded martingale. Noting that

$$R^{(n)}(s) \quad := \quad \int_{[0,1]} \mathbf{1}(\rho(X^{(n)}, s, u) \neq \rho(X^{(n-1)}, s, u))\, du$$

$$= \quad \sum_{\substack{(j,r) \in \Lambda \\ r < s}} a(j, r, s) \mathbf{1}(X_j^{(n)}(r) \neq X_j^{(n-1)}(r)) = \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}(X_i^{(n)}(s) \neq X_i^{(n-1)}(s))$$

(cf Remark 7) we see that

$$\mathbb{E}\, \mathbf{1}(\rho(X^{(n)}, s, U) \neq \rho(X^{(n-1)}, s, U)) = \mathbb{E} \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}(X_i^{(n+1)}(s) \neq X_i^{(n)}(s)) = \mathbb{E}\, R^{(n)}(s)$$

So $t = T$ and taking expectations above yields $\mathbb{P}(X_i^{(n+1)}(T) \neq X_i^{(n)}(T)) \leq 2\bar{\sigma} \int_0^T \mathbb{E}\, R^{(n)}(s)\, ds$, hence

$$\mathbb{E}\, R^{(n+1)}(T) = \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^{m} \mathbb{P}(X_i^{(n+1)}(T) \neq X_i^{(n)}(T)) \leq 2\bar{\sigma} \int_0^T \mathbb{E}\, R^{(n+1)}(s)\, ds,$$

which proves that $\mathbb{E}\,R^{(n+1)}(T) \to 0$ as $n \to \infty$.

This shows that for any $T > 0$ and $j \in \mathbb{N}$, $\mathbb{P}(X_j^{(n+1)}(T) = X_j^{(n)}(T)) \to 1$ as $n \to \infty$. As for any n, the value of the process $X_j^{(n)}$ can only change at the jump times of M_j, K_j or L_{ik} ($i < k \leq j$) and there are only finitely many such jumps in any interval $[0, t]$, we obtain for any $t > 0$, $m \in \mathbb{N}$

$$\mathbb{P}\Big(X_j^{(n+1)}(s) = X_j^{(n)}(s) \text{ for } 0 \leq s \leq t,\, j = 1, 2, \ldots, m\Big) \xrightarrow[N \to \infty]{} 1.$$

This proves that $X^{(n)}$ converges. Furthermore, for fixed t, the limit $X(t)$ is exchangeably distributed, because this is true for any $X^{(n)}$ by Lemma 5.

In order to convince ourselves that the limit process X is in fact a solution of (1.61) it suffices to consider the jump events of the selective resampling process: assume that T is a jump time of K_j, let $V \sim \mathrm{unif}([0, 1])$ and $\tau_V^T := \sup\{s < T : (N^T(V, s), s) \in \Lambda\}$ be the time when the type $\rho(X, T, V)$ was (potentially) created. As $N^T(V, \tau_V^T)$ is a finite level, we have $\rho(X, T, V) = X_{N^T(V, \tau_V^T)}(\tau_V^T) = X_{N^T(V, \tau_V^T)}^{(n)}(\tau_V^T) = \rho(X^{(n)}, T, V)$ for n sufficiently large.

Finally, we check uniqueness. Assume that X, X' are two solutions of (1.61) with $X(0) = X'(0)$ exchangeable. For $T > 0$, $j \in \mathbb{N}$, define corresponding ancestral type processes $X_j^T(t)$, $X_j'^T(t)$, $0 \leq t \leq T$ as above These will at least agree up to the first time point when a $\tilde{K}$-jump produces a difference. Denoting $\tilde{R}(T) := \lim_m m^{-1} \sum_{i=1}^m \mathbf{1}(X_i(T) \neq X_i'(T))$ we can argue as above the estimate $\mathbb{E}\,\tilde{R}(T) \leq 2\bar{\sigma} \int_0^T \mathbb{E}\,\tilde{R}(t)\,dt$, which proves $\mathbb{E}\,\tilde{R}(T) \equiv 0$, e.g. by Gronwall's Lemma. $\qquad\square$

The empirical process

Let X be the solution of (1.61). By exchangeability, the empirical process

$$Z(t) := \lim_{m \to \infty} \frac{1}{m} \sum_{i=1}^m \delta_{X_i(t)}$$

(with values in $\mathcal{M}_1(E)$) exists a.s.) and satisfies ($f \in B_b(E^m)$)

$$\mathbb{E}\big[\langle f, Z(t)^m \rangle\big] = \mathbb{E}\big[f(X_1(t), \ldots, X_m(t))\big].$$

We also obtain from (1.61) that

$$f\big(X_1(t), \ldots, X_m(t)\big) - \int_0^t Af\big(X(s), Z(s)\big)ds \quad \text{is an } \mathscr{F}^X\text{-martingale,} \qquad (1.70)$$

where for $\underline{x} = (x_1, \ldots, x_m) \in E^m$, $\mu \in \mathcal{M}_1(E)$

$$
\begin{aligned}
Af(x, \mu) \;=\; & \sum_{i=1}^m B_i f(\underline{x}) + \sum_{1 \leq i < j \leq m} f\big(\theta_j(\underline{x}; x_i)\big) - f(\underline{x}) \\
& + \sum_{j=1}^m \int_{E \times E} \sigma(y_1, y_2)\Big(f\big(\eta_j(\underline{x}; x_i)\big) - f(\underline{x})\Big)\mu(dy_1)\mu(dy_2).
\end{aligned}
$$

(Recall that if L is a Poisson process on $\mathbb{R}_+ \times A$ with intensity measure $dt \otimes \zeta(da)$, Z is a previsible process with values in E and $f : E \times \mathbb{R}_+ \times A \to \mathbb{R}$ is a bounded, measurable function, then the compensated process

$$\int_{(0,t] \times A} f(Z_s, s, a)\, L(ds\, da) - \int_{(0,t] \times A} f(Z_s, s, a)\, ds\, \zeta(da)$$

is a martingale. Compensate the Poisson processes in (1.61) to obtain (1.70).)

34

For $f : E^m \to \mathbb{R}_+$ bd., meas. define $F_f : \mathcal{M}_1(E) \to \mathbb{R}_+$ via $F_f(\mu) = \langle f, \mu^{\otimes m} \rangle$. Define $\mathbb{A}$ on functions of the type F_f as

$$\mathbb{A} F_f(\mu) := \langle Af(\cdot, \mu), \mu^{\otimes m} \rangle. \tag{1.71}$$

A straightforward computation shows that

$$\mathbb{A} F_f(\mu) \;=\; \sum_{i=1}^m \langle B_i f, \mu^{\otimes m} \rangle + \sum_{i<j}^m \langle \Phi_{ij} f, \mu^{\otimes m} \rangle + \sum_{j=1}^m \Big(\langle \sigma_j f, \mu^{\otimes (m+1)} \rangle - \langle \sigma \otimes f, \mu^{\otimes (m+2)} \rangle \Big),$$

where B_i is B acting on the i-th coordinate of f, $\Phi_{ij} f$ is f with the j-th input set equal to the i-th, $\sigma_j f(y_1, \ldots, y_{m+1}) = \sigma(y_j, y_{m+1}) f(y_1, \ldots, y_m)$ and $\sigma \otimes f(y_1, \ldots, y_{m+2}) = \sigma(y_{m+1}, y_{m+2}) f(y_1, \ldots, y_m)$.

Proposition 2. *The empirical process $(Z(t))$ corresponding to X solves the martingale problem for $\mathbb{A}$.*

Z is called a *Fleming-Viot process* with mutation and selection. It is known that the martingale problem with generator (1.71) is uniquely solvable, cf e.g. [EK93], Thm. 3.2. Uniqueness is usually proved through a duality argument. Let us remark that we have – in principle – the ingredients for this before our eyes, cf the paragraph on the ancestral selection graph below for an embryonic version.

Proof. We obtain from the above and the fact that $\mathscr{F}_t^Z \subset \mathscr{F}_t^X$ that

$$\mathbb{E}\big[f(X_1(t), \ldots, X_m(t)) \big| \mathscr{F}_t^Z \big] - \int_0^t \mathbb{E}\big[Af(X_1(t), \ldots, X_m(t), Z(s)) \big| \mathscr{F}_s^Z \big]\, ds$$

$$= \;\; \langle f, Z(t)^{\otimes m} \rangle - \int_0^t \langle Af(\cdot, Z(s)), Z(t)^{\otimes m} \rangle\, ds = F_f(Z(t)) - \int_0^t \mathbb{A} F_f(Z(s))\, ds$$

is an $\mathscr{F}^Z$-martingale. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Box$

Remark 8. In the situation $|E| < \infty$, say $E = \{1, 2, \ldots, d\}$, the process $(Z(T)) = (Z_1(t), \ldots, Z_d(t))$ is a diffusion on the $(d-1)$-dimensional simplex $\{(z_1, \ldots, z_d) : z_i \geq 0, z_1 + \cdots + z_d = 1\}$. Its generator acts on $\varphi \in C^2(\mathbb{R}_d)$ as

$$\frac{1}{2} \sum_{i,j=1}^d z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} \varphi + \sum_{j=1}^d \Big(\sum_{i=1}^d q_{ij} z_i \Big) \frac{\partial}{\partial z_j} \varphi + \sum_{j=1}^d z_j \Big(\sum_{i,k=1}^d \big(\sigma(j,i) - \sigma(k,i) \big) z_i z_k \Big) \frac{\partial}{\partial z_j} \varphi, \tag{1.72}$$

where (q_{ij}) is the q-matrix of the mutation process. Note that the last term describes selection: the fraction z_j of type j-individuals experiences a drift which is the difference between the fitness of type j in the present population and the present total fitness.

In order to check this claim: a straightforward calculation shows that it holds true for monomials $\varphi(z_1, \ldots, z_d) = z_1^{n_1} \cdots z_d^{n_d}$ (which have an obvious sampling interpretation: pick $n_1 + \cdots + n_d$ individuals, require the first n_1 to be of type 1, the next n_2 to be of type 2, etc.). Approximate more general φ uniformly by polynomials.

Remark 9. The simplest non-trivial case involving selection is a two-type scenario with directional selection, e.g. $E = \{1, 2\}$, $\sigma(i, j) = s\mathbf{1}_{i=1}$ for some $s > 0$ (i.e. type 1 has an absolute fitness advantage over type 2). Assume that the mutation process has generator $Bf(i) = \theta_1 \mathbf{1}_{i=2}(f(1) - f(2)) + \theta_2 \mathbf{1}_{i=1}(f(2) - f(1))$ (i.e. type 1 mutates into type 2 at rate θ_2, the other direction a rate θ_1). It suffices to consider, say the evolution of $\tilde{Z}_t$, the fraction of type 2 currently in the population. We see from Remark 8 that $\tilde{Z}$ is a diffusion with generator

$$\frac{1}{2} z(1-z)\varphi'' + \big(-\theta_1 z + \theta_2(1-z) - sz(1-z) \big)\varphi',$$

a Wright-Fisher diffusion with mutation and selection.

The ancestral influence graph

Imagine we sample n individuals from the (infinite) population at time T. Recall that a very convenient property of the coalescent with neutral mutations is that one can first generate the genealogy and then simply superimpose the mutation process on it. While this is no longer the case in a selective scenario, the modified look down construction allows one to construct rigorously a 'supragenealogy' which does not only keep track of the actual ancestors of the sampled individuals, but also of those who could potentially be ancestors if they carry the right type in a selective reampling event. Once we have constructed the supragenealogy, we can superimpose the types at time 0 and the mutation process and decide for each selective resampling event who among the two potential ancestors actually is a real ancestor. Thus the supragenealogy will be a super-set of the real genealogical tree(s) connecting the sample.

We restrict ourselves to an intuitive discussion and refer to [DK99], Section 8 for details (and the notational overhead): Start with n lineages. Each pair of lineages coalesces at rate 1. Additionally, each lineage branches at rate $\bar{\sigma}$. In such a 'branch event', two new lineages are created, and we think of the then three lineages as labelled with o ('original'), 1 and 2. Additionally, we label each such branch event with a uniform($[0,1]$)-distributed random value. See the example below.
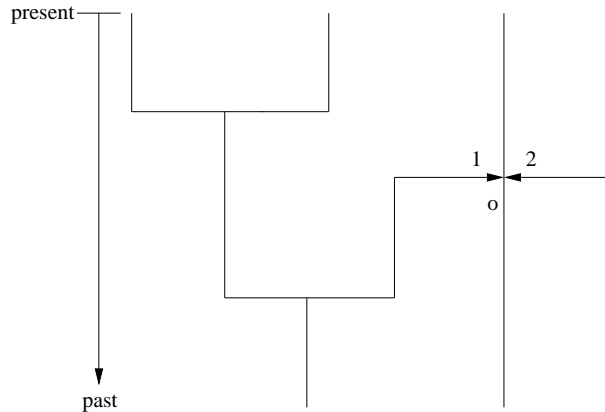
Once we arrive at time T backwards (which corresponds to time 0 in the population model, i.e. the founding population), each of the lineages is coloured according to an independent pick from $Z(0)$. Then we run forwards in time independent copies of the mutation process on each line. Assume that we come to a branch point which is labelled with $u \in [0,1]$ and that the types at the three incoming edges labelled o, 1, 2 are e_o, e_1, e_2 respectively. Then the 'outcoming' type (in the forwards time direction) will be e_1 if $\bar{\sigma}u \leq \sigma(e_1, e_2)$, corresponding to a successful selective resampling event. Otherwise it will be (remain) type e_o.

Note that if we go back far enough into the past, the number of lineages will eventually become 1 because the coalescence rate is quadratic and the branching rate is only linear in the number of active lineages. Then we have found the *ultimate ancestor* of the sample. Note that it need not necessarily be the actual most recent common ancestor of the sampled individual because the latter is only specified once we have assigned the types and sorted out the selective resamplings.

Krone & Neuhauser's ancestral selection graph

The ancestral influence graph (which in its fully-fledged form also includes recombination, an issue that we have not discussed at all) generalises the so-called 'ancestral selection graph' from [KN97]. Let us briefly consider this. Krone & Neuhauser work in the two-type scenario as described in Remark 9. As the selection function $\sigma(i, j) = s\mathbf{1}_{i=1}$ does not depend on the second input, it suffices to consider double branching instead of triple branching as above (i.e. we can ignore the branches labelled with "2" in the above construction. We will also need no uniform

36

marks on the branching events.) Furthermore, as a two-type mutation can always be realised in a parent-independent way, we can construct the supragenealogy in this special case as follows: Start with n lineages. Each pair of lineages coalesces at rate 1, each lineage branches in two at rate s. Once we arrive at 'backwards' time T, assign founding types (independent coin tosses, type 2 with probability $\tilde{Z}_0$). Proceed forwards in time in the following way: Throw down independent Poisson processes $P^{(1)}$ with rate θ_1 and $P^{(2)}$ with rate θ_2 on each line. We prescribe that the type carried by a line changes to i (irrespective of its present type) whenever it 'hits' a point of $P^{(i)}$. If we come to a branching event, the 'outcoming' type will be 1 if either one of the incoming branches carries type 1, otherwise it will be type 2.

We can make use of this e.g. in order to compute moments of $\tilde{Z}_t$: Note that $\mathbb{E}_{z_0}\,(\tilde{Z}_T)^n$ is the probability to observe only type 2 in a sample of size n from the population at time T. On the other hand, we can compute this probability differently. Imagine we are constructing the ancestral selection graph backwards in time and 'raining down' the two types of Poisson points simultaneously. Note that once a lineage has hit a mutation point, we know its type, and we can safely discontinue it. The event in question requires that no 'active' lineage ever hits a type 1-mutation point, and also all lineages arriving at time 0 must be assigned type 2. Thus

$$\mathbb{E}_{z_0}\,(\tilde{Z}_t)^n = \mathbb{E}_n\Big[\exp\big(-\theta_1\int_0^t L_u\,du\big)z_0^{L_t}\Big], \qquad (1.73)$$

where (L_t) is a birth and death process with birth rate $\lambda(k) = sk$ and death rate $\mu(k) = \binom{k}{2}+\theta_2 k$. Note that by taking $t \to \infty$, this allows to express moments of the equilibrium distribution of $\tilde{Z}$ via the Laplace transform of the time to extinction for (L_t).

As a check, we can derive (1.73) in a completely different way. Put $f_{n,z}(t) := \mathbb{E}_{z_0}\,(\tilde{Z}_t)^n$, $g_{n,z}(t) := E_n\Big[\exp\big(-\theta_1\int_0^t L_u\,du\big)z_0^{L_t}\Big]$. Obviously $f_{n,z}(0) = g_{n,z}(0)$. Itô's Formula yields

$$d\tilde{Z}^n = n\tilde{Z}^{n-1}\sqrt{\tilde{Z}(1-\tilde{Z})}dW + \Big(-n\theta_1\tilde{Z}^n + n\theta_2\big(\tilde{Z}^{n-1}-\tilde{Z}^n\big) + ns\big(\tilde{Z}^{n+1}-\tilde{Z}^n\big) + \frac{n(n-1)}{2}\big(\tilde{Z}^{n-1}-\tilde{Z}^n\big)\Big)dt,$$

so

$$\frac{\partial}{\partial t}f_{n,z}(t) = \Big(\theta_2 n + \binom{n}{2}\Big)\big(f_{n-1,z}(t) - f_{n,z}(t)\big) + sn\big(f_{n+1,z}(t) - f_{n,z}(t)\big) - \theta_1 n f_{n,z}(t).$$

On the other hand, Kolmogorov's backward equation for the continuous-time Markov chain (L_t) shows that $g_{n,z}$ solves the same equations, hence $f_{n,z}(t) = g_{n,z}(t)$.

Approximating with finite systems

Finally, let us come back to the finite Moran systems we started from. With the machinery developed above, we can obtain the n-particle system $X^{(n)}$ as the (unique) solution of ($1 \leq j \leq n$)

$$
\begin{aligned}
X_j^{(n)}(t) \;=\; & X_j(0) + \int_{[0,1]\times(0,t]} h(X_j^{(n)}(s-),u) - X^{(n)}(s-)\,M_j(du,ds) & (1.74)\\
& + \sum_{i<j}\int_{(0,t]} X_i^{(n)}(s-) - X_j^{(n)}(s-)\,L_{ij}(ds)\\
& + \sum_{1\leq i<k<j}\int_{(0,t]} X_{j-1}^{(n)}(s-) - X_j^{(n)}(s-)\,L_{ij}(ds)\\
& + \int_{[0,1]^3\times(0,t]} \big(X_{k_n(u_1,s)}^{(n)}(s-) - X_j^{(n)}(s-)\big)\\
& \qquad \times \mathbf{1}\big(\bar{\sigma}u_3 \leq \sigma(X_{k_n(u_1,s)}^{(n)}(s-),X_{k_n(u_2,s)}^{(n)}(s-))\big)K_j(du_1,du_2,du_3,ds),
\end{aligned}
$$

constructed from the same driving Poisson processes and using the same neutral marker processes as the infinite system.

37

Proposition 3. *For each $T > 0$, $j = 1, 2, \ldots$, we have*

$$\mathbb{P}\big(X_j^{(n)}(t) = X_j(t) \text{ for } 0 \le t \le T\big) \longrightarrow \quad \text{as } n \to \infty. \tag{1.75}$$

Let $Z^{(n)}(t) := \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j^{(n)}(t)}$ be the empirical distribution process of the n-particle system. For any $m \in \mathbb{N}$ and $f : E^m \to \mathbb{R}$ bounded measurable we have

$$\mathbb{E}\left| \langle f, Z(t)^m \rangle - \langle f, \big(Z^{(n)}(t)\big)^m \rangle \right| \longrightarrow \quad \text{as } n \to \infty. \tag{1.76}$$

We will not give a proof (which can be found in Section 7 of [DK99]), but note that at least the author finds (1.75) quite plausible in view of the above considerations: $X_j(t)$ and $X_j^{(n)}(t)$ will agree until the first time when $X_{k_n(u_1,s)}^{(n)}(s-) \ne \rho(X, s, u_1)$ for some atom (u_1, u_2, u_3, s) of K_j, of which there are only finitely many in each bounded time interval. As the type of $\rho(X, s, u_1)$ is decided 'on some finite level', it is very likely for large n that it is equal to $X_{k_n(u_1,s)}^{(n)}(s-)$.

Note that once we have (1.75), we obtain (1.76) easily because $(X_1(t), \ldots, X_m(t))$ is distributed like m independent picks from $Z(t)$.

Chapter 2

Branching models

2.1 A crash course on Galton-Watson processes

Galton-Watson processes are a simple model of a stochastically reproducing 'population': Let $\xi_{n,i}$, $n, i \in \mathbb{N}$ be i.i.d. copies of an $\mathbb{N}_0$-valued random variable ξ. For $n \in \mathbb{N}$ and given Z_{n-1}, the size of the n-th generation is given by

$$Z_n = \sum_{i=1}^{Z_{n-1}} \xi_{n,i}.$$

Note that $(Z_n)_n$ is a discrete Markov chain and has the *branching property:* For $a, b \in \mathbb{N}$,

$$\mathscr{L}(Z_n | Z_0 = a + b) = \mathscr{L}(Z_n | Z_0 = a) * \mathscr{L}(Z_n | Z_0 = b).$$

We assume that $\mu := \mathbb{E}\,\xi < \infty$. We denote $p_k := \mathbb{P}(\xi = k)$. To avoid trivial cases, we also assume that $p_0, p_1 < 1$. Let $\phi(s) := \mathbb{E}\,s^\xi$ be the generating function of ξ, $f_n(s) := \mathbb{E}\big[s^{Z_n} \big| Z_0 = 1\big]$. We have

$$\mathbb{E}_1\big[s^{Z_n}\big] = \mathbb{E}_1\big[\mathbb{E}[s^{Z_n}|Z_1]\big] = \mathbb{E}_1\big[f_{n-1}(s)^{Z_1}\big] = \phi(f_{n-1}(s)),$$

hence

$$f_n(s) = \big(\underbrace{\phi \circ \cdots \circ \phi}_{n \text{ times}}\big)(s).$$

A Galton-Watson process is called *subcritical* if $\mu < 1$, *critical* if $\mu = 1$, and *supercritical* if $\mu > 1$.

Proposition 4. *Let $q_n := \mathbb{P}_1(Z_n = 0)$. Then $q_n \nearrow q = \mathbb{P}_1(\lim_{n \to \infty} Z_n = 0)$, the extinction probability. q is the smallest fixed point of ϕ. We have $q = 1$ if $\mu \leq 1$, and $q < 1$ if $\mu > 1$.*

Proof. Obviously $\{Z_n = 0\} \subset \{Z_{n+1} = 0\}$, hence $q_n \leq q_{n+1}$. Decomposing according to the size of the first generation shows that

$$q_{n+1} = \mathbb{P}_1(Z_{n+1} = 0) = \mathbb{E}\big[\mathbb{P}_1(Z_n = 0)^{Z_1}\big] = \phi(q_n),$$

so taking $n \to \infty$ yields $q = \phi(q)$ by continuity of ϕ. Let $[0, 1] \ni q' = \phi(q')$ be a fixed point of ϕ. By monotonicity of ϕ,

$$q_1 = \phi(0) \leq \phi(q') = q'.$$

Apply ϕ n-times to this inequality to obtain $q_n \leq q'$, hence also $q \leq q'$.

Finally recall that ϕ is a strictly convex, increasing function on $[0, 1]$ with $\phi(1) = 1$, and that $\mu = \sum_{k=1}^{\infty} k p_k = \phi'(1-)$. Hence $\mu \leq 1$ implies $\phi(s) > s$ for all $s \in [0, 1)$, so that $q = 1$ is the only fixed point of ϕ in this case. On the other hand, if $\mu > 1$, there must be a second fixed point $q < 1$. $\qquad\square$

We denote by $\sigma^2 := \mathrm{Var}(\xi) = \sum_k (k^2 - \mu^2) p_k$ the variance of the individual offspring number.

Lemma 6.
$$\mathbb{E}\, Z_n = \mu^n \mathbb{E}\, Z_0,$$

$$\mathbb{E}\, Z_n^2 = \begin{cases} \frac{\sigma^2}{(1/\mu)-1}\left(\mu^n - \mu^{2n}\right)\mathbb{E}\, Z_0 + \mu^{2n}\mathbb{E}\, Z_0^2 & \mu \neq 1, \\ n\sigma^2 \mathbb{E}\, Z_0 + \mathbb{E}\, Z_0^2 & \mu = 1. \end{cases}$$

Proof. We have

$$\mathbb{E}\, Z_n = \mathbb{E}\Big[\mathbb{E}\big[\sum_{i=1}^{Z_{n-1}} \xi_{n,i}\big|Z_{n-1}\big]\Big] = \mu\mathbb{E}\, Z_{n-1} = \cdots = \mu^n \mathbb{E}\, Z_0$$

and

$$\begin{aligned} \mathbb{E}\,(Z_n)^2 &= \mathbb{E}\Big[\mathbb{E}\big[\big(\sum_{i=1}^{Z_{n-1}} \xi_{n,i}\big)^2\big|Z_{n-1}\big]\Big] = \mu^2\mathbb{E}\big[Z_{n-1}(Z_{n-1}-1)\big] + \mathbb{E}[\xi^2]\mathbb{E}[Z_{n-1}] \\ &= \mu^2\mathbb{E}\,(Z_{n-1})^2 + \sigma^2\mathbb{E}\, Z_{n-1} = \mu^2\mathbb{E}\,(Z_{n-1})^2 + \sigma^2\mu^{n-1}\mathbb{E}\, Z_0, \end{aligned}$$

so $a_n := (\mathbb{E}\,(Z_n)^2)/\mu^{2n}$ satisfies

$$a_n = a_{n-1} + \sigma^2\mu^{-n-1}\mathbb{E}\, Z_0 = a_0 + \sigma^2\mathbb{E}\, Z_0 \sum_{i=1}^{n} \mu^{-i-1} = \begin{cases} \mathbb{E}\, Z_0^2 + \sigma^2\mathbb{E}\, Z_0 \frac{\mu^{-n}-1}{(1/\mu)-1} & \mu \neq 1, \\ \mathbb{E}\, Z_0^2 + n\sigma^2\mathbb{E}\, Z_0 & \mu = 1. \end{cases}$$

$\square$

Note that

$$W_n := \frac{Z_n}{\mu^n}$$

is a nonnegative martingale (with respect to the filtration generated by Z).

In the subcritical case, we obtain $\mathbb{P}_1(Z_n > 0) \leq \mathbb{E}_1 Z_n = \mu^n$, and it is known that in fact $\mathbb{P}_1(Z_n > 0) \sim \mu^n$ if ξ has slightly more than a first moment, namely if $\mathbb{E}\,\xi \log^+ \xi < \infty$ (a result of Heathcote, Seneta and Vere-Jones, cf. Thm. 11.4 in [LP05] and the references given there).

In the supercritical case, if additionally $\mathbb{E}\,\xi^2 < \infty$, we see from Lemma 6 that

$$\sup_n \mathbb{E}_1 W_n^2 = 1 + \frac{\sigma^2}{1 - (1/\mu)} < \infty.$$

In particular, (W_n) is then a uniformly integrable martingale, so $\mathbb{E}_1\big[\lim_{n\to\infty} W_n\big] = 1$. Thus in the supercritical case, if the population does not die out, it tends to grow geometrically with rate μ up to a 'random constant'. The assumption $\sigma^2 < \infty$ is too strong: A result of Kesten and Stigum (cf. Chapter 11 in [LP05] and the references given there) tells that a necessary and sufficient condition for $\mathbb{E}_1\big[\lim_{n\to\infty} W_n\big] = 1$ is again $\mathbb{E}\,\xi \log^+ \xi < \infty$.

Theorem 6. *Assume $\mu = 1$ and $\sigma^2 < \infty$. Then*

$$\mathbb{P}_1(Z_n > 0) \sim \frac{2}{\sigma^2 n}. \tag{2.1}$$

Remark 10. This result is known as Kolmogorov's estimate, who proved it under an additional third moment assumption. Kesten, Ney and Spitzer removed the third moment assumption. We follow the proof of Jochen Geiger, in the version presented in Chapter 11.4 of [LP05].
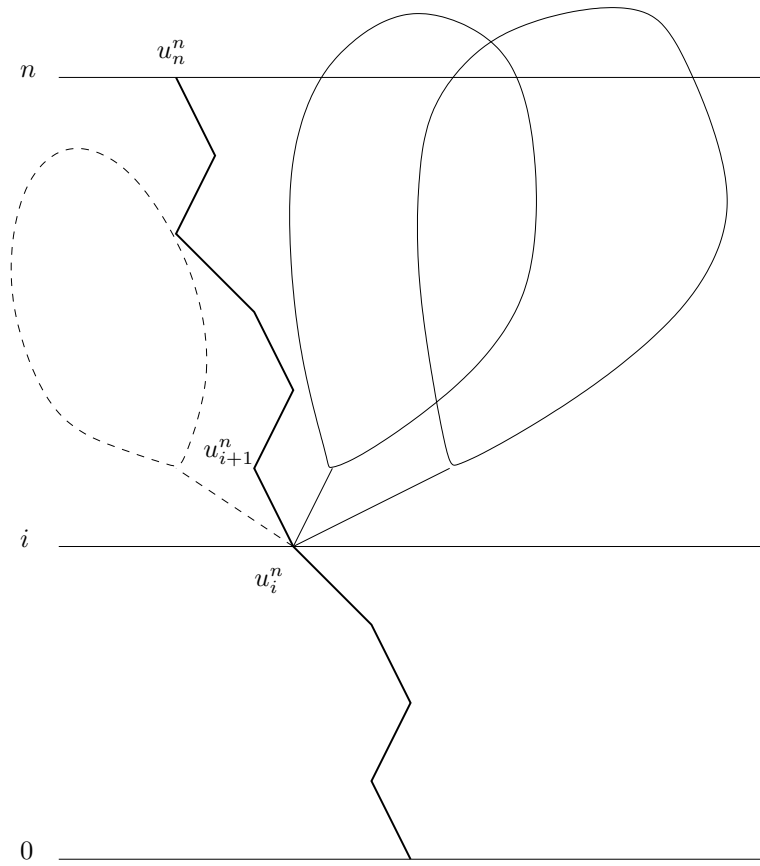
Figure 2.1: Here, u_{i+1}^N is the second child of u_i^n, and $X_i^{(n)} = 2$.

Proof. Observe that the Galton-Watson process (Z_n), started from $Z_0 = 1$, gives rise to an ordered or planar tree if we think of the offspring of each individual as being ordered and draw them in order, say from left to right, into the plane.

On the event $\{Z_n > 0\}$, there must be a left-most living individual in generation n. Let us denote this individual by u_n^n, and let u_i^n, $i = 0, 1, \ldots, n-1$ be its ancestor in generation i (u_0^n=root of the tree). Denote by $X_i^{(n)}$ the number of children of u_i^n 'to the right of u_{i+1}^n' (i.e. direct descendants which the chosen ordering puts behind u_{i+1}^n), and let

$$\tilde{X}_i^{(n)} = \#\{\text{individuals in generation } n \text{ who are descendants of } u_i^n, \text{ but not of } u_{i+1}^n\}, \quad i = 0, 1\ldots, n-1$$

(see Figure 2.1). Note that

$$Z_n = 1 + \sum_{i=0}^{n-1} \tilde{X}_i^{(n)}.$$

Let $a_n = \mathbb{P}_1(Z_n > 0)$, and $D_{n-1} \sim \text{geom}(a_{n-1})$, i.e. $\mathbb{P}(D_{n-1} = j) = a_{n-1}(1 - a_{n-1})^{j-1}$, $j = 1, 2, \ldots$, independent of Z_1. Then

$$\mathscr{L}(Z_1 | Z_n > 0) = \mathscr{L}(Z_1 | Z_1 \geq D_n). \tag{2.2}$$

In order to see this imagine that we produce the Galton-Watson process as follows: First we take infinitely many 'potential individuals' in the first generation, numbered (from 'left to right') with $1, 2, \ldots$. Each of them founds an independent Galton-Watson process, which survives for (the remaining) $n - 1$ generations with probability a_{n-1}. Let $D_{n-1} = $ (index of the) leftmost individual whose descendancy survives for at least $n-1$ generations. In the next step, we choose Z_1, an independent copy of ξ, and declare the first Z_1 individuals of the potential first generation to be the 'real' first generation. Thus in this coupling, $\{Z_n > 0\} = \{Z_1 \geq D_{n-1}\}$.

Lemma 7 applied to $A = Z_1$, $B = D_{n-1}$, $C = D_n$ then gives

$$\mathscr{L}(Z_1 | Z_n > 0) \quad \text{is stochastically increasing in } n. \tag{2.3}$$

(Note that if $B \sim \text{geom}(\beta)$, $C \sim \text{geom}(\gamma)$ with $\gamma \leq \beta$, then

$$\frac{\mathbb{P}(C \leq x)}{\mathbb{P}(B \leq x)} = \frac{1 - (1 - \gamma)^x}{1 - (1 - \beta)^x} \leq \frac{1 - (1 - \gamma)^{x+1}}{1 - (1 - \beta)^{x+1}} = \frac{\mathbb{P}(C \leq x + 1)}{\mathbb{P}(B \leq x + 1)}.$$

A way to see this is to observe that for $x \in \mathbb{N}$, $(0, 1) \ni a \mapsto (1 - a^x)/(1 - a^{x+1})$ is decreasing.)

Note that because $a_n \to_{n \to \infty} 0$, we have

$$\mathbb{P}(Z_1 = k | Z_1 \geq D_{n-1}) = \frac{p_k\big(1 - (1 - a_{n-1})^k\big)}{\sum_{j=1}^{\infty} p_j\big(1 - (1 - a_{n-1})^j\big)} \underset{n \to \infty}{\sim} \frac{kp_k}{\sum_{j=1}^{\infty} jp_j} =: \hat{p}_k,$$

and for $\ell \leq k$

$$\mathbb{P}(Z_1 = k, D_{n-1} = \ell | Z_1 \geq D_{n-1}) = \frac{p_k a_{n-1}(1 - a_{n-1})^{\ell-1}}{\sum_{j=1}^{\infty} p_j\big(1 - (1 - a_{n-1})^j\big)} \underset{n \to \infty}{\sim} \frac{p_k}{\sum_{j=1}^{\infty} jp_j} = \hat{p}_k \frac{1}{k}.$$

Thus, $\mathscr{L}(Z_1 | Z_1 \geq D_{n-1})$ converges to that of $\hat{\xi}$, the *size-biasing* of ξ. Furthermore, on the event $\{Z_1 \geq D_{n-1}\}$, given $Z_1 = k$, D_{n-1} is asymptotically uniformly distributed on $\{1, 2, \ldots, k\}$. The above and (2.3) imply that

$$\mathbb{E}\big[Z_1 \big| \{Z_n > 0\}\big] \to \mathbb{E}\hat{\xi} = \sum_j j^2 p_j = \sigma^2 + 1 \quad \text{as } n \to \infty$$

(note that $\mathbb{E}\big[Z_1 \big| \{Z_n > 0\}\big] = \sum_{k=1}^{\infty} \mathbb{P}(Z_1 \geq k | \{Z_n > 0\})$), and each summand is increasing in n, so that in this case convergence in distribution implies convergence of the means).

$$\mathscr{L}(Z_1 - D_{n-1} | \{Z_1 \geq D_{n-1}\}) \to \mathscr{L}([\hat{\xi}U]),$$

42

where $U \sim \text{uniform}([0,1])$, independent of $\hat{\xi}$, and $[\hat{\xi}U]$ denotes the biggest integer $\leq \hat{\xi}U$. Hence

$$\mathbb{E}\big[X_0^{(n)}\big|\{Z_n > 0\}\big] = \mathbb{E}\big[Z_1 - D_{n-1}\big|\{Z_1 \geq D_{n-1}\}\big] \to \mathbb{E}\big[[\hat{\xi}U]\big] = \frac{1}{2}\mathbb{E}\big[\hat{\xi} - 1\big] = \frac{\sigma^2}{2} \quad \text{as } n \to \infty.$$

(In order to verify convergence of the means in this case write

$$\mathbb{E}\big[Z_1 - D_{n-1}\big|\{Z_1 \geq D_{n-1}\}\big] = \mathbb{E}\big[\mathbf{1}_{\{Z_1 \leq K\}}(Z_1 - D_{n-1})\big|\{Z_1 \geq D_{n-1}\}\big] + \mathbb{E}\big[\mathbf{1}_{\{Z_1 > K\}}(Z_1 - D_{n-1})\big|\{Z_1 \geq D_{n-1}\}\big].$$

The first term converges to $\mathbb{E}\big[\mathbf{1}_{\{\hat{\xi} \leq K\}}[\hat{\xi}U]\big]$ as $n \to \infty$, while by (2.3), the second is bounded uniformly in n by $\mathbb{E}[\mathbf{1}_{\{\hat{\xi} > K\}}\hat{\xi}]$. Take $K \to \infty$ to conclude.)

The same arguments can be applied to any generation $i < n$, and thus we obtain

$$\mathbb{E}\big[X_i^{(n)}\big|\{Z_n > 0\}\big] \to \frac{\sigma^2}{2} \quad \text{whenever } n - i \to \infty. \tag{2.4}$$

Furthermore $\mathbb{E}\big[\tilde{X}_i^{(n)}\big|\{Z_n > 0\}\big] = \mathbb{E}\big[X_i^{(n)}\big|\{Z_n > 0\}\big]$, because by definition none of the children of u_i^n to the left of u_{i+1}^n have descendants in generation n, and each of the $X_i^{(n)}$ siblings to the right of u_{i+1}^n founds an independent, critical (unconditioned) Galton-Watson tree. Thus we find

$$\frac{1}{n\mathbb{P}_1(Z_n > 0)} = \frac{1}{n}\frac{\mathbb{E}\big[Z_n\mathbf{1}_{\{Z_n > 0\}}\big]}{\mathbb{P}_1(Z_n > 0)} = \frac{1}{n}\mathbb{E}\big[Z_n\big|\{Z_n > 0\}\big] = \frac{1}{n} + \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}\big[\tilde{X}_i^{(n)}\big|\{Z_n > 0\}\big] \to \frac{\sigma^2}{2},$$

which yields (2.1) $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Lemma 7. *Let A, B, C be $\mathbb{N}_0$-valued random variables, A independent from (B, C), satisfying $\mathbb{P}(A \geq B), \mathbb{P}(A \geq C) > 0$ and*

$$x \mapsto \mathbb{P}(C \leq x)/\mathbb{P}(B \leq x) \quad \text{is increasing in } x. \tag{2.5}$$

Then

$$\mathscr{L}(A|A \geq B) \quad \text{is stochastically smaller than} \quad \mathscr{L}(A|A \geq C).$$

Note that (2.5) implies (but is stronger than) $\mathscr{L}(B) \ll \mathscr{L}(C)$.

Proof. Abbreviate $\tilde{a}_x := \mathbb{P}(A = x)$, $b_x := \mathbb{P}(B \leq x)$, $c_x := \mathbb{P}(B \leq x)$. Let $y \in \mathbb{N}$ be such that $\mathbb{P}(B \leq A \leq y) > 0$ (and then necessarily also $\mathbb{P}(C \leq A \leq y) > 0$). (2.5) implies

$$\frac{c_{y+1}}{b_{y+1}} \geq \frac{\sum_{x=0}^{y}\tilde{a}_x b_x \frac{c_x}{b_x}}{\sum_{x=0}^{y}\tilde{a}_x b_x} = \frac{\sum_{x=0}^{y}\tilde{a}_x c_x}{\sum_{x=0}^{y}\tilde{a}_x b_x},$$

hence

$$\tilde{a}_{y+1}c_{y+1}\sum_{x=0}^{y}\tilde{a}_x b_x \geq \tilde{a}_{y+1}b_{y+1}\sum_{x=0}^{y}\tilde{a}_x c_x.$$

Adding $\mathbb{P}(B \leq A \leq y)\mathbb{P}(C \leq A \leq y)$ to both sides we obtain

$$\Big(\sum_{x=0}^{y}\tilde{a}_x b_x\Big)\Big(\sum_{x=0}^{y+1}\tilde{a}_x c_x\Big) \geq \Big(\sum_{x=0}^{y+1}\tilde{a}_x b_x\Big)\Big(\sum_{x=0}^{y}\tilde{a}_x c_x\Big),$$

which proves that

$$y \mapsto \frac{\mathbb{P}(B \leq A \leq y)}{\mathbb{P}(C \leq A \leq y)}$$

is decreasing in y. In particular,

$$\frac{\mathbb{P}(B \leq A \leq y)}{\mathbb{P}(C \leq A \leq y)} \geq \frac{\mathbb{P}(B \leq A)}{\mathbb{P}(C \leq A)},$$

which is the claim. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Bibliography

[Ber96]   Jean Bertoin. *Lévy processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1996.

[DK99]   Peter Donnelly and Thomas G. Kurtz. Genealogical processes for fleming-viot models with selection and recombination. *Annals of Applied Probability*, 9(4):1091–1148, 1999.

[Dur02]   Rick Durrett. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer-Verlag, New York, 2002.

[EG87]   S. N. Ethier and R. C. Griffiths. The infinitely-many-sites model as a measure-valued diffusion. *Ann. Probab.*, 15(2):515–545, 1987.

[EK86]   Stewart N. Ethier and Thomas G. Kurtz. *Markov processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. Characterization and convergence.

[EK93]   S. N. Ethier and Thomas G. Kurtz. Fleming-Viot processes in population genetics. *SIAM J. Control Optim.*, 31(2):345–386, 1993.

[Ewe04]   Warren J. Ewens. *Mathematical population genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2004. Theoretical introduction.

[GL04]   Robert Griffiths and Sabin Lessard. Ewens' sampling formula and related formulae: combinatorial proofs, extensions to variable population size and applications to ages of alleles. *Preprint*, 2004.

[Hop84]   Fred M. Hoppe. Pólya-like urns and the Ewens' sampling formula. *J. Math. Biol.*, 20(1):91–94, 1984.

[JK72]   Norman L. Johnson and Samuel Kotz. *Distributions in statistics: continuous multivariate distributions*. John Wiley & Sons Inc., New York, 1972. Wiley Series in Probability and Mathematical Statistics.

[JT87]   Paul Joyce and Simon Tavaré. Cycles, permutations and the structure of the Yule process with immigration. *Stochastic Process. Appl.*, 25(2):309–314, 1987.

[Kal02]   Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[KN97]   Stephen M. Krone and Claudia Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51:210–231, 1997.

[LP05]   Russell Lyons and Yuval Peres. *Probability on trees and networks,* a book in progress. http://mypage.iu.edu/∼rdlyons/prbtree/prbtree.html, 2005.

[RW00] L. C. G. Rogers and David Williams. *Diffusions, Markov processes, and martingales. Vol. 2.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2000. Itô calculus, Reprint of the second (1994) edition.

[SD00] Matthew Stephens and Peter Donnelly. Inference in molecular population genetics. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 62(4), 2000. With discussion and a reply by the authors.

[Tav87] Simon Tavaré. The birth process with immigration, and the genealogical structure of large populations. *J. Math. Biol.*, 25(2):161–168, 1987.