Weak convergence of the scaled jump chain and number of mutations of the Kingman coalescent[*]

Martina Favero[†]   Henrik Hult[‡]

Abstract

The Kingman coalescent is a fundamental process in population genetics modelling the ancestry of a sample of individuals backwards in time. In this paper, in a large-sample-size regime, we study asymptotic properties of the coalescent under neutrality and a general finite-alleles mutation scheme, i.e. including both parent independent and parent dependent mutation. In particular, we consider a sequence of Markov chains that is related to the coalescent and consists of block-counting and mutation-counting components. We show that these components, suitably scaled, converge weakly to deterministic components and Poisson processes with varying intensities, respectively. Along the way, we develop a novel approach, based on a change of measure, to generalise the convergence result from the parent independent to the parent dependent mutation setting, in which several crucial quantities are not known explicitly.

1 Introduction

The Kingman coalescent [19] is a classical stochastic process that models the genealogical tree of a sample of individuals. With the aim of including various genetic forces, generalisations of this pivotal model resulted in the establishment of other well-known models, such as the ancestral selection graph [22, 26], the ancestral recombination graph [12, 14], the $\Lambda-$ coalescent [27, 29], and the $\Xi-$ coalescent [25, 31]. Coalescent theory is also closely related to urn models. In fact, a Pólya-like urn structure can be embedded in the coalescent process by matching lineages and balls [34].

In this paper, we consider the Kingman coalescent with a finite-allele mutation scheme evolving under neutrality (i.e. no type has a selective advantage), and study its asymptotic behaviour as the size of the sample grows to infinity. Beside the purely mathematical significance of deriving new properties of a classical model, our analysis is inspired by applied statistical problems which have recently drawn attention to large-sample-size regimes in population genetics and are briefly mentioned in the following.

Coalescent models are often used for inference, combined with Monte Carlo methods to approximate the likelihood of an observed genes configuration. In particular, importance sampling algorithms [15, 32, 6] and sequential Monte Carlo methods [21] have been developed based on simulating backwards the genealogy of the sample (extensions based on generalisations of the coalescent exists [33, 4, 13, 17, 5, 20]). However, it is known empirically that these methods do not scale well with increasing sample size [18]. Another issue is that the coalescent approximation is only accurate when the sample size is sufficiently smaller than the effective population size. Otherwise, as explained in [2], substantial differences between the original models and the coalescent approximation could arise. Studying large-sample-size asymptotic properties of the coalescent provides a tool both for the theoretical asymptotic analysis of inference methods, which could guide their improvement, and for the analysis of the appropriateness of the coalescent as sample sizes in modern studies in genetics grow rapidly.

The contribution of the paper is twofold. The main result (Theorem 2.1) is the convergence of a sequence of Markov chains related to the coalescent, which consists of block-counting and mutation-counting components and is described in Section 2. We show that these components, suitably scaled, converge to deterministic components and Poisson processes with varying intensities, respectively. The convergence result is first proved under the assumption of parent independent mutation (PIM), i.e. the type of the mutated offspring does not depend on the type of the parent. We then develop a novel approach, based on a change of measure and presented in Section 3, to remove the PIM assumption and generalise the convergence result. The change-of-measure approach (Theorem 3.1) constitutes the other main result of the paper. In fact, to the best of our knowledge, it provides the first tool in population genetics to transfer results that are obtained under the widely-spread PIM assumption to the general mutation setting. This setting is notoriously more challenging and often neglected in the literature due to several crucial quantities not being known explicitly. Other technical challenges for the convergence proofs are addressed in Section 4 by constructing a technical framework that allows employing the classical tools developed by Ethier and Kurtz [8]. Section 5 contains the convergence proofs.

2 Outline and main result

The coalescent describes the genealogy of a sample of individuals by describing the evolution of their ancestral lineages which coalesce and mutate. Time evolves backwards from time 0, at which the sample is taken, until only one lineage is left, i.e. the most recent common ancestor of the sample is reached. Here, lineages are *typed*, that is, types are assigned to lineages, rather than superimposed afterwards, as the coalescent evolves backwards conditioned on the initial sample, as in e.g. [15, 7, 10, 9]. We refer to this version of the coalescent as the typed conditional coalescent.

We consider the sample, taken from a population at stationarity, to be of the form $n\mathbf{y}_0^{(n)}$, with $\mathbf{y}_0^{(n)} \in \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\}$, where components represent the numbers of individuals of the various possible types $1, \ldots, d$. While the typed conditional coalescent models the genealogy of the sample, as described below in details, the frequencies of types in the population are modelled by the Wright-Fisher diffusion. The sample thus corresponds

to a multinomial draw from the stationary distribution of the Wright Fisher diffusion. This close connection between the typed conditional coalescent and the Wright-Fisher diffusion finds its explanation in their duality relationship, see e.g. [1, 7, 16, 9, 11].

Besides the number of types d, the key parameters for both the coalescent and the diffusion are the mutation rate θ and the mutation probability matrix $P = (P_{ij})_{i,j=1}^d$, with P_{ij} being the probability that the type changes from i to j when a mutation occurs. Throughout the paper, as it is customary, P is assumed to be irreducible, to ensure the existence of a unique stationary distribution.

Instead of considering the whole tree structure formed by the coalescing lineages, we count its *blocks*, that is, the numbers of lineages of each type over time. Furthermore, time is discretised: the original coalescent evolves in continuous time, as well as its block-counting process, we consider instead the corresponding block-counting jump chain, which jumps to the same states in discrete steps. The kth state of the discrete-time chain corresponds to the state of the continuous-time process right after its kth jump (k^{th} coalescence or mutation event). The block-counting jump chain of the conditional typed coalescent is denoted by $\mathbf{H}^{(n)} = \{\mathbf{H}^{(n)}(k)\}_{k\in\mathbb{N}}$, given the initial sample $\mathbf{H}^{(n)}(0) = n\mathbf{y}_0^{(n)}$ and evolving backwards in time until the most recent common ancestor is reached. Thus, $H_i^{(n)}(k)$ denotes the number of lineages of type i after k jumps in the ancestral history. The transitions of the Markov chain $\mathbf{H}^{(n)}$ are precisely defined in the following, after a scaling. Let $\mathbf{Y}^{(n)} = \{\mathbf{Y}^{(n)}(k)\}_{k\in\mathbb{N}} \subset \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\}$ be defined as

$$\mathbf{Y}^{(n)}(k) = \frac{1}{n}\mathbf{H}^{(n)}(k) = \frac{1}{n}\left(H_1(k), \ldots, H_d(k)\right),$$

see Figure 1. Given a current state $\mathbf{Y}^{(n)}(k) = \mathbf{y} \in \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\}$, the next state $\mathbf{Y}^{(n)}(k+1)$ is of the form $\mathbf{y} - \frac{1}{n}\mathbf{v}$, where, denoting the d-dimensional unit vector along the jth dimension by $\mathbf{e}_j$,

- $\mathbf{v} = \mathbf{e}_i, i = 1, \ldots, d$, with probability $\rho^{(n)}(\mathbf{e}_i \mid \mathbf{y})$, which corresponds to the coalescence of two lineages of type i;

- $\mathbf{v} = \mathbf{e}_i - \mathbf{e}_j, i, j = 1, \ldots, d$, with probability $\rho^{(n)}(\mathbf{e}_i - \mathbf{e}_j \mid \mathbf{y})$, which corresponds to the mutation of a lineage of type j to type i forwards in time (or, equivalently, i to j backwards in time).

The backwards transition probabilities $\rho^{(n)}$ are generally not known in an explicit form, unless mutations are parent independent, see e.g. [15, 32, 6, 13] for a more detailed description. Implicit and explicit expressions for $\rho^{(n)}$, in the general and PIM case respectively, are provided in (A.1) in Appendix A.1 and in (5.1) in Section 5.1.

Furthermore, along with $\mathbf{Y}^{(n)}$, consider the process $\mathbf{M}^{(n)} = \{\mathbf{M}^{(n)}(k)\}_{k\in\mathbb{N}} \subset \mathbb{N}^{d^2}$ which counts the number of mutations of each type occurring in the genealogy. That is, $\mathbf{M}^{(n)}(k) = (M_{ij}^{(n)}(k))_{i,j=1}^d$, where $M_{ij}^{(n)}(k)$ is the cumulative number of mutations from type i to type j (forwards, or j to i backwards) that have occurred in the genealogy during the first k steps, i.e.

$$M_{ij}^{(n)}(k) = \sum_{k'=0}^{k-1} \mathbb{I}_{\{\mathbf{Y}^{(n)}(k') - \mathbf{Y}^{(n)}(k'+1) = \frac{1}{n}(\mathbf{e}_j - \mathbf{e}_i)\}},$$

and $M_{ij}(0) = 0$.

In this paper, we study the asymptotic behaviour of the sequence $\mathbf{Z}^{(n)} = (\mathbf{Y}^{(n)}, \mathbf{M}^{(n)})$, as $n \to \infty$. Note that, the sequence $\mathbf{M}^{(n)}$ is not only interesting for the purpose of studying an additional asymptotic property of the coalescent, but it is also crucial because it naturally appears in our change-of-measure argument, as explained in the next sections. The main convergence result is stated in the following theorem.
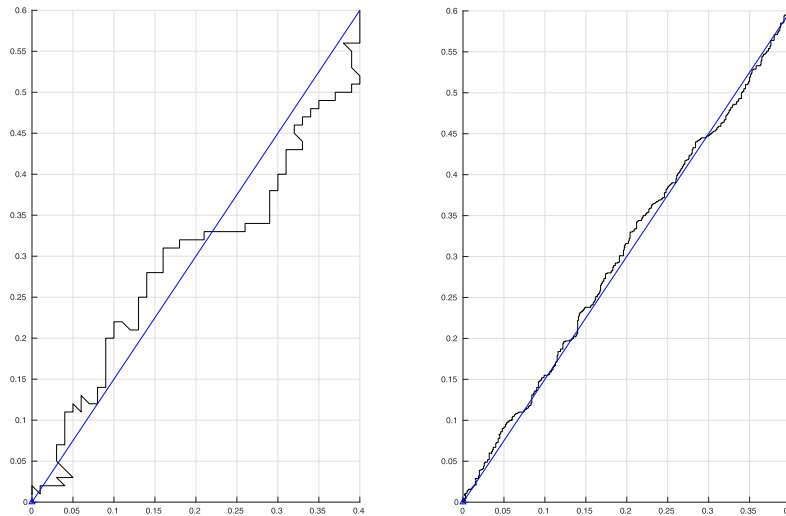
Figure 1: A realisation of the Markov chains $\mathbf{Y}^{(100)}$ (left) and $\mathbf{Y}^{(1000)}$ (right) with starting point $\mathbf{y}_0 = (0.4, 0.6)$. The limiting process $\mathbf{Y}$ is represented by the blue line. Parameters: mutation rate $\theta = 4$, mutation probabilities $Q_1 = Q_2 = 0.5$, number of types $d = 2$.

Theorem 2.1. *Let $\mathbf{y}_0 \in \mathbb{R}_+^d$ and $\mathbf{y}_0^{(n)} \in \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\}, n \in \mathbb{N}$. Assume $\mathbf{Y}^{(n)}(0) = \mathbf{y}_0^{(n)}$, and $\mathbf{y}_0^{(n)} \to \mathbf{y}_0$ as $n \to \infty$. Then, for all $t \in [0, \|\mathbf{y}_0\|_1)$, as $n \to \infty$, the sequence of processes $\tilde{\mathbf{Z}}^{(n)} = \{\mathbf{Z}^{(n)}(\lfloor sn \rfloor)\}_{s \in [0,t]}$ converges weakly to the process $\mathbf{Z} = \{(\mathbf{Y}(s), \mathbf{M}(s))\}_{s \in [0,t]} \subset \mathbb{R}_+^d \times \mathbb{N}^{d^2}$, where $\mathbf{Y}$ is the deterministic process defined by*

$$\mathbf{Y}(s) = \frac{\mathbf{y}_0}{\|\mathbf{y}_0\|_1} \left(\|\mathbf{y}_0\|_1 - s \right),$$

and $\mathbf{M} = (M_{ij})_{i,j=1}^d$, with $M_{ij} = \{M_{ij}(s)\}_{s \in [0,t]}$ being independent time-inhomogeneous Poisson processes with intensities

$$\lambda_{ij}(\mathbf{Y}(s)) = \frac{\theta P_{ij} Y_i(s)}{\|\mathbf{Y}(s)\|_1^2}.$$

Here converging weakly means converging in the Skorokhod space $D_{\mathbb{R}_+^d \times \mathbb{N}^{d^2}}[0, t]$. That is, for any bounded continuous real-valued function g on $D_{\mathbb{R}_+^d \times \mathbb{N}^{d^2}}[0, t]$, it yields

$$\lim_{n \to \infty} \mathbb{E}\left[g\left(\{\tilde{\mathbf{Z}}^{(n)}(s)\}_{s \in [0,t]} \right) \right] = \mathbb{E}\left[g\left(\{\mathbf{Z}(s)\}_{s \in [0,t]} \right) \right].$$

We now provide an intuitive explanation of Theorem 2.1 and illustrate the two main challenges that need to be addressed to obtain a rigorous proof. Define the operator $A^{(n)}$, associated to the Markov chain $\mathbf{Z}^{(n)}$, as

$$A^{(n)} f(\mathbf{y}, \mathbf{m}) = n\mathbb{E}\left[f(\mathbf{Z}^{(n)}(k+1)) - f(\mathbf{Z}^{(n)}(k)) \mid \mathbf{Z}^{(n)}(k) = (\mathbf{y}, \mathbf{m}) \right]$$

$$= \sum_{j=1}^d n \left[f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j, \mathbf{m} \right) - f(\mathbf{y}, \mathbf{m}) \right] \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y})$$

$$+ \sum_{i,j=1}^d \left[f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_i, \mathbf{m} + \mathbf{e}_{ij} \right) - f(\mathbf{y}, \mathbf{m}) \right] n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y}), \quad (2.1)$$

where $\mathbf{e}_{ij}$ is the unit matrix having 1 in position ij and 0 everywhere else. Recent results on the asymptotic behaviour of $\rho^{(n)}$ [10], show that, if $\mathbf{y}^{(n)} \to \mathbf{y} \in \mathbb{R}_+^d$, as $n \to \infty$, then, for $i, j = 1, \ldots, d$,

$$\rho^{(n)}(\mathbf{e}_j \mid \mathbf{y}^{(n)}) \xrightarrow[n \to \infty]{} \frac{y_j}{\|\mathbf{y}\|_1}, \qquad n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y}^{(n)}) \xrightarrow[n \to \infty]{} \lambda_{ij}(\mathbf{y}). \tag{2.2}$$

Therefore, by evaluating (2.1) at $\mathbf{y}^{(n)}$, and letting $\mathbf{y}^{(n)} \to \mathbf{y} \in \mathbb{R}_+^d$ as $n \to \infty$, it is straightforward to see that the sequence of operators $A^{(n)}$ converges to, in some sense to be properly defined,

$$Af(\mathbf{y}, \mathbf{m}) = -\left\langle \nabla_{\mathbf{y}} f(\mathbf{y}, \mathbf{m}), \frac{\mathbf{y}}{\|\mathbf{y}\|_1} \right\rangle + \sum_{i,j=1}^d \left[f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m}) \right] \lambda_{ij}(\mathbf{y}). \tag{2.3}$$

The operator A, defined in (2.3) above, can be proven to be, as the intuition suggests, the infinitesimal generator of the limiting process of Theorem 2.1, see Appendix A.3. In order to make rigorous the convergence of generators sketched above to then prove convergence of the corresponding processes, we need to define the appropriate space of functions, in the domain of the generator A, for which the convergence holds. Two main challenges arise.

The first challenge is given by the lack of an explicit expression for the transition probabilities $\rho^{(n)}$ in the parent dependent mutation case. Even though the pointwise asymptotic behaviour (2.2) of the transition probabilities is known, this is not enough to prove convergence of generators, as evident in Section 5.1, because additional information concerning uniform convergence would be needed. We address this issue by first proving the convergence in the PIM case, for which explicit expressions are available, and by then employing a novel approach based on a change of measure to generalise the result from a PIM setting to a general setting, as explained in details in Section 3. It is important to point out that the counting-mutations sequence $\mathbf{M}^{(n)}$ would naturally appear through the change of measure, even if only the block-counting sequence $\mathbf{Y}^{(n)}$ was considered to begin with (equation (3.1) in Section 3). This strongly motivates the study of the asymptotic behaviour of $\mathbf{M}^{(n)}$, without which the generalisation to the parent dependent setting would not be possible.

The second challenge, which arises already in the simpler PIM setting, is that, if the components of $\mathbf{y}$ are allowed to be arbitrarily close to zero, the scaled transition probabilities of mutation in (2.2) explode, see (5.1) in Section 5.1. As more precisely described in Section 4, this issue can be addressed by constructing an appropriate technical framework which relies on the definition of a suitable metric to avoid components being arbitrarily close to zero and allows employing Ethier-Kurtz classical tools [8].

While sections 3 and 4 are dedicated, respectively, to the change-of-measure argument and the technical framework, Section 5 contains the convergence proofs leading to, and including, the proof of Theorem 2.1.

3 From parent independent to parent dependent mutations: a novel change-of-measure approach

Suppose that the convergence of Theorem 2.1 holds when mutations are parent independent, that is when the mutation probability matrix $Q = (Q_i)_{i=1}^d$ has identical rows. The goal of this section is to illustrate how to generalise the convergence result to a general setting, with mutation probability matrix $P = (P_{ij})_{i,j=1}^d$.

More precisely, using subscripts Q and P to indicate the underlying mutation probability matrices and fixing a bounded continuous function $g : D_{\mathbb{R}_+^d \times \mathbb{N}^{d^2}}[0, t] \to \mathbb{R}$, suppose

that we know

$$\lim_{n\to\infty} \mathbb{E}\left[g\left(\{\tilde{\mathbf{Z}}_Q^{(n)}(s)\}_{s\in[0,t]} \right) \right] = \mathbb{E}\left[g\left(\{\mathbf{Z}_Q(s)\}_{s\in[0,t]} \right) \right],$$

and that we want to prove

$$\lim_{n\to\infty} \mathbb{E}\left[g\left(\{\tilde{\mathbf{Z}}_P^{(n)}(s)\}_{s\in[0,t]} \right) \right] = \mathbb{E}\left[g\left(\{\mathbf{Z}_P(s)\}_{s\in[0,t]} \right) \right].$$

Our approach consists of rewriting the expectations with respect to $\tilde{\mathbf{Z}}_{\mathbf{P}}^{(n)}$ and $\mathbf{Z}_{\mathbf{P}}$ as expectations with respect to $\tilde{\mathbf{Z}}_{\mathbf{Q}}^{(n)}$ and $\mathbf{Z}_{\mathbf{Q}}$, respectively. That is, we apply a change of measure to transform the parent dependent mutation setting in a parent independent mutation setting. We present the core idea in the following.

Consider the probability of a realisation $\boldsymbol{\mathcal{H}}(0), \ldots, \boldsymbol{\mathcal{H}}(K), K \in \mathbb{N}$, of $\mathbf{H}_P(0), \ldots, \mathbf{H}_P(K)$, given $\mathbf{H}_P(0) = \boldsymbol{\mathcal{H}}(0)$,

$$\mathbb{P}\left(\mathbf{H}_P(K) = \boldsymbol{\mathcal{H}}(K), \ldots, \mathbf{H}_P(1) = \boldsymbol{\mathcal{H}}(1) \mid \mathbf{H}_P(0) = \boldsymbol{\mathcal{H}}(0) \right)$$

$$= \frac{1}{p_P(\boldsymbol{\mathcal{H}}(0))} \mathbb{P}\left(\mathbf{H}_P(K) = \boldsymbol{\mathcal{H}}(K), \ldots, \mathbf{H}_P(0) = \boldsymbol{\mathcal{H}}(0) \right)$$

$$= \frac{p_P(\boldsymbol{\mathcal{H}}(K))}{p_P(\boldsymbol{\mathcal{H}}(0))} \prod_{k=0}^{K-1} \mathbb{P}\left(\mathbf{H}_P(k) = \boldsymbol{\mathcal{H}}(k) \mid \mathbf{H}_P(k+1) = \boldsymbol{\mathcal{H}}(k+1) \right),$$

where $p_P(\boldsymbol{\mathcal{H}}(k))$ is the probability of observing a sample $\boldsymbol{\mathcal{H}}(k)$ under the unconditioned coalescent, i.e. the sampling probability, and $\mathbb{P}\left(\mathbf{H}_P(k) = \boldsymbol{\mathcal{H}}(k) \mid \mathbf{H}_P(k+1) = \boldsymbol{\mathcal{H}}(k+1) \right)$ is the explicitly-known forward transition probability. More precisely, p_P, which can be expressed explicitly only when mutations are parent independent, can be defined either through a recursion formula or as a multinomial draw from the stationary distribution of the corresponding Wright-Fisher diffusion, see [15] for details. The stationary distribution, which is Dirichlet in the PIM case, is also not known explicitly in the general case but exists as long as the mutation probability matrix is irreducible [30]. The boundary condition is assumed to be $p_P(e_i) = \pi_{P,i}, i = 1, \ldots, d$, where $\boldsymbol{\pi}_P = (\pi_{P,i})_{i=1}^d$ is the invariant distribution of the mutation probability matrix. Despite the implicit formulation, a large-sample-size asymptotic formula for the sampling probabilities is available in general [10], which makes the rewriting in the last display useful for the asymptotic analysis. The forward transition probabilities instead are always explicitly known [32]. Given a configuration $\mathbf{H}_P(k+1) = \boldsymbol{\mathcal{H}}(k+1)$, a lineage is chosen uniformly at random, i.e. of type i with probability $\frac{\mathcal{H}_i(k+1)}{\|\boldsymbol{\mathcal{H}}(k+1)\|_1}$. The chosen lineage is split with probability $\frac{\|\boldsymbol{\mathcal{H}}(k+1)\|_1 - 1}{\|\boldsymbol{\mathcal{H}}(k+1)\|_1 - 1 + \theta}$, and it is mutated to type j with probability $\frac{\theta}{\|\boldsymbol{\mathcal{H}}(k+1)\|_1 - 1 + \theta} P_{ij}$. That is, the next configuration is $\mathbf{H}_P(k) = \boldsymbol{\mathcal{H}}(k+1) + \mathbf{v}$, where

- $\mathbf{v} = \mathbf{e}_i, i = 1, \ldots, d$, with probability $\frac{\mathcal{H}_i(k+1)}{\|\boldsymbol{\mathcal{H}}(k+1)\|_1} \frac{\|\boldsymbol{\mathcal{H}}(k+1)\|_1 - 1}{\|\boldsymbol{\mathcal{H}}(k+1)\|_1 - 1 + \theta}$;

- $\mathbf{v} = \mathbf{e}_j - \mathbf{e}_i, i, j = 1, \ldots, d$, with probability $\frac{\mathcal{H}_i(k+1)}{\|\boldsymbol{\mathcal{H}}(k+1)\|_1} \frac{\theta}{\|\boldsymbol{\mathcal{H}}(k+1)\|_1 - 1 + \theta} P_{ij}$.

Note that this description clarifies the equivalence between the coalescent and the Pólya-like urn model mentioned in the introduction: according to the probabilities above, typed-balls, which correspond to typed-lineages, are duplicated and replaced, instead of split and mutated respectively [34].

It is now possible to derive the likelihood ratio, the Radon–Nikodym derivative, of the change of measure from parent dependent mutation with mutation probability matrix P to parent independent mutation with mutation probability matrix Q. Assuming, without

loss of generality, that Q is positive, it yields

$$\frac{\mathbb{P}\left(\mathbf{H}_P(K) = \boldsymbol{\mathcal{H}}(K), \ldots, \mathbf{H}_P(1) = \boldsymbol{\mathcal{H}}(1) \mid \mathbf{H}_P(0) = \boldsymbol{\mathcal{H}}(0)\right)}{\mathbb{P}\left(\mathbf{H}_Q(K) = \boldsymbol{\mathcal{H}}(K), \ldots, \mathbf{H}_Q(1) = \boldsymbol{\mathcal{H}}(1) \mid \mathbf{H}_Q(0) = \boldsymbol{\mathcal{H}}(0)\right)}$$

$$= \frac{p_P(\boldsymbol{\mathcal{H}}(K))}{p_Q(\boldsymbol{\mathcal{H}}(K))} \frac{p_Q(\boldsymbol{\mathcal{H}}(0))}{p_P(\boldsymbol{\mathcal{H}}(0))} \prod_{k=0}^{K-1} \frac{\mathbb{P}\left(\mathbf{H}_P(k) = \boldsymbol{\mathcal{H}}(k) \mid \mathbf{H}_P(k+1) = \boldsymbol{\mathcal{H}}(k+1)\right)}{\mathbb{P}\left(\mathbf{H}_Q(k) = \boldsymbol{\mathcal{H}}(k) \mid \mathbf{H}_Q(k+1) = \boldsymbol{\mathcal{H}}(k+1)\right)}$$

$$= \frac{p_P(\boldsymbol{\mathcal{H}}(K))}{p_Q(\boldsymbol{\mathcal{H}}(K))} \frac{p_Q(\boldsymbol{\mathcal{H}}(0))}{p_P(\boldsymbol{\mathcal{H}}(0))} \prod_{i,j=1}^{d} \left(\frac{P_{ij}}{Q_j}\right)^{\sum_{k=0}^{K-1} \mathbb{I}_{\{\boldsymbol{\mathcal{H}}(k) - \boldsymbol{\mathcal{H}}(k+1) = \mathbf{e}_j - \mathbf{e}_i\}}}. \tag{3.1}$$

Note that the likelihood ratio above does not depend on the whole history, but instead only depends on the initial and last configurations, $\boldsymbol{\mathcal{H}}(K), \boldsymbol{\mathcal{H}}(0)$, and on the total number of mutations from type i to type j, $\sum_{k=0}^{K-1} \mathbb{I}_{\{\boldsymbol{\mathcal{H}}(k) - \boldsymbol{\mathcal{H}}(k+1) = \mathbf{e}_j - \mathbf{e}_i\}}$, for $i, j = 1, \ldots, d$. This is the key observation for the change-of-measure approach and it is also the reason why it is crucial for the asymptotic analysis to consider, along with the block-counting sequence, the counting-mutation sequence which naturally appears in the likelihood ratio.

In the next theorem, we tailor the previous calculations to the change of measure from $\tilde{\mathbf{Z}}_P^{(n)}$ to $\tilde{\mathbf{Z}}_Q^{(n)}$ and derive similar calculations for the change of measure from $\mathbf{Z}_P$ to $\mathbf{Z}_Q$. Note that an analogous result can be obtained for any process that is adapted to the block-counting jump-chain of the Kingman coalescent.

Theorem 3.1. *Let* $g : D_{\mathbb{R}_+^d \times \mathbb{N}^{d^2}}[0, t] \to \mathbb{R}$ *be a bounded continuous function. Assume* $\tilde{\mathbf{Y}}_P^{(n)}(0) = \tilde{\mathbf{Y}}_Q^{(n)}(0) = \mathbf{y}_0^{(n)} \in \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\}$. *Then, for all* $t \geq 0$,

$$\mathbb{E}\left[g\left(\{\tilde{\mathbf{Z}}_P^{(n)}(s)\}_{s \in [0,t]}\right)\right] = \mathbb{E}\left[g\left(\{\tilde{\mathbf{Z}}_Q^{(n)}(s)\}_{s \in [0,t]}\right) r_{P,Q}^{(n)}\left(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}\right) c_{P,Q}\left(\tilde{\mathbf{M}}_Q^{(n)}(t)\right)\right], \tag{3.2}$$

where

$$r_{P,Q}^{(n)}(\mathbf{y}^{(n)}, \mathbf{y}_0^{(n)}) = \frac{p_P(n\mathbf{y}^{(n)})}{p_Q(n\mathbf{y}^{(n)})} \frac{p_Q(n\mathbf{y}_0^{(n)})}{p_P(n\mathbf{y}_0^{(n)})}, \quad \mathbf{y}^{(n)} \in \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\},$$

and

$$c_{P,Q}(\boldsymbol{m}) = \prod_{i,j=1}^{d} \left(\frac{P_{ij}}{Q_j}\right)^{m_{ij}}, \quad \boldsymbol{m} \in \mathbb{N}^{d^2}.$$

Furthermore, for all $t \in [0, \|\mathbf{y}_0\|_1)$,

$$\mathbb{E}\left[g\left(\{\mathbf{Z}_P(s)\}_{s \in [0,t]}\right)\right] = \mathbb{E}\left[g\left(\{\mathbf{Z}_Q(s)\}_{s \in [0,t]}\right) c_{P,Q}\left(\mathbf{M}_Q(t)\right)\right]. \tag{3.3}$$

Proof. By definition, $\{\tilde{\mathbf{Z}}_P^{(n)}(s)\}_{s \in [0,t]}$ is a function of n and $\mathbf{H}_P(0), \ldots, \mathbf{H}_P(\lfloor tn \rfloor)$. Thus, we can simply use (3.1) with $K = \lfloor tn \rfloor$ to obtain (3.2).

Since the first component of $\mathbf{Z}_P = (\mathbf{Y}, \mathbf{M}_P)$ and of $\mathbf{Z}_Q = (\mathbf{Y}, \mathbf{M}_Q)$ are deterministic and do not depend on the mutation probability matrix, it is sufficient to focus on the second components. Using that

$$\mathbb{P}\left(\mathbf{M}_P(t) = \mathbf{m} \mid \mathbf{Y}(0) = \mathbf{y}_0\right) = \prod_{i,j=1}^{d} \frac{1}{m_{ij}!} \left(\int_0^t \lambda_{P,ij}(\mathbf{Y}(u)) du\right)^{m_{ij}} e^{-\int_0^t \lambda_{P,ij}(\mathbf{Y}(u)) du}$$

and noting that $\int_0^t \lambda_{P,ij}(\mathbf{Y}(u)) du = \frac{\theta P_{ij} y_{0,i}}{\|\mathbf{y}_0\|_1} \int_0^t \frac{1}{\|\mathbf{y}_0\|_1 - u} du$, we obtain

$$\frac{\mathbb{P}\left(\mathbf{M}_P(t) = \mathbf{m} \mid \mathbf{Y}(0) = \mathbf{y}_0\right)}{\mathbb{P}\left(\mathbf{M}_Q(t) = \mathbf{m} \mid \mathbf{Y}(0) = \mathbf{y}_0\right)} = c_{P,Q}(\mathbf{m}).$$

In fact, $c_{P,Q}(\mathbf{M}_Q(t))$ can be interpreted as the exponential martingale associated to the change of measure that transforms the distribution of the Poisson processes $\{\mathbf{M}_P(s)\}_{s \in [0,t]}$ into the distribution of the Poisson processes $\{\mathbf{M}_Q(s)\}_{s \in [0,t]}$, see e.g. [23]. This proves (3.3) and completes the proof. $\square$

4 Handling the explosion: technical framework

In this section, we construct a technical framework that allows us to address the problem caused by the scaled transition probabilities of mutation in (2.2) exploding near the boundary $\Omega_0 := \{\mathbf{y} = (y_1, \ldots, y_d) : y_i = 0 \text{ for some } i\}$. In such a framework, we can rigorously state the convergence of generators (Proposition 4.1 below) and we can show the weak convergence of the sequence $\mathbf{Z}^{(n)}$ to the process $\mathbf{Z}$ by relying on classical results [8], as explained briefly below and proved in detail in the next section.

A natural state-space of the limiting process $\mathbf{Z}$ is $\mathbb{R}_+^d \times \mathbb{N}^{d^2}$, equipped with the Euclidean metric. Consider instead $E_1 = (0, \infty]^d$, equipped with the metric $\psi_1(\mathbf{y}_1, \mathbf{y}_2) := \left\| \frac{1}{\mathbf{y}_1} - \frac{1}{\mathbf{y}_2} \right\|_2$, where the inversion is component-wise and the inverse of ∞ is by convention 0; and consider $E = E_1 \times \mathbb{N}^{d^2}$, equipped with the product metric $\psi := \psi_1 \oplus \|\cdot\|_2$. For completeness, we naturally define $\mathbf{Z}$ to be constantly equal to its initial state, if one of the components of $\mathbf{Y}(0)$ is equal to ∞, although this type of initial condition does not appear in the relevant setting of Theorem 2.1.

In the new state-space E_1, the roles of 0 and ∞ are reversed for each component, whilst, away from 0 and ∞, the metric ψ_1 is equivalent to the Euclidean metric. In particular, compact sets are bounded away from Ω_0, and thus functions with compact support are equal to zero near Ω_0. See Appendix A.2 for more details on the metric space (E, ψ).

Let $C_c^\infty(E)$ be the space of real-valued continuous functions on (E, ψ) with compact support, equipped with the supremum norm, which is analysed in details in Appendix A.2. This is precisely the space of functions that we need for the convergence of generators, which is stated in the next proposition.

Let $E^{(n)} = \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\} \times \mathbb{N}^{d^2}$ be the state space of $\mathbf{Z}^{(n)}$ and let η_n map any function on E into its restriction on $E^{(n)}$, with value zero on $\Omega_0 \times \mathbb{N}^{d^2}$.

Proposition 4.1. *Assume parent independent mutations. Let $A^{(n)}$ be the operator, defined in (2.1), associated to the Markov chain $\mathbf{Z}^{(n)}$, and let and A be the infinitesimal generator, defined in (2.3) , associated to the Markov process $\mathbf{Z}$. Then, for all $f \in C_c^\infty(E)$,*

$$\lim_{n \to \infty} \sup_{(\mathbf{y}, \boldsymbol{m}) \in E^{(n)}} \left| A^{(n)} \eta_n f(\mathbf{y}, \boldsymbol{m}) - \eta_n A f(\mathbf{y}, \boldsymbol{m}) \right| = 0.$$

Proposition 4.1, which is proved in Subsection 5.1, allows to derive the convergence of the corresponding semigroups, which is proved in Section 5.2. In order to then prove weak convergence (Theorem 2.1), another technical issue needs to be addressed. In fact, as time approaches $\|\mathbf{y}_0\|_1$, $\mathbf{Y}$ approaches the origin, and the intensities of the Poisson processes $\mathbf{M}$ explode, that is, the process $\mathbf{Z}$ exits its state space E in a finite time. This is a classical problem, and the technical framework introduced in this section is constructed precisely to match the framework of Ethier and Kurtz [8] and to employ their classical results which allow us to address the issue and finally prove weak convergence.

Following [8], a new point Δ, called the *cemetery point*, is introduced in the state space. Let $E^\Delta = E \cup \{\Delta\}$ be the one-point compactification of E, which is itself a metric space, see Appendix A.2 and [24] for more details, and let $\mathbf{Z}^\Delta$ be the extension of $\mathbf{Z}$, corresponding to $\mathbf{Z}$ before time $\|\mathbf{y}_0\|_1$, and being equal to Δ from time $\|\mathbf{y}_0\|_1$ on. Furthermore, let ξ_n^Δ be the inclusion map from $E^{(n)}$ into E^Δ, sending $\Omega_0 \times \mathbb{N}^{d^2}$ to Δ.

Instead of proving directly Theorem 2.1, i.e.

$$\{\tilde{\mathbf{Z}}^{(n)}(s)\}_{s\in[0,t]} \xrightarrow[n\to\infty]{} \{\mathbf{Z}(s)\}_{s\in[0,t]} \quad \text{in } D_S[0,t],$$

for a fixed $t \in [0, \|\mathbf{y}_0\|_1)$, we prove an extended convergence

$$\xi_n^{\Delta}\left(\{\tilde{\mathbf{Z}}^{(n)}(s)\}_{s\in[0,\infty)}\right) \xrightarrow[n\to\infty]{} \{\mathbf{Z}^{\Delta}(s)\}_{s\in[0,\infty)} \quad \text{in } D_{E^{\Delta}}[0,\infty) \tag{4.1}$$

which then implies Theorem 2.1, as shown in Subsection 5.3.

5 Convergence proofs

In this section, we prove Theorem 2.1, i.e. the weak convergence of the sequence $\{\tilde{\mathbf{Z}}^{(n)}(s)\}_{s\in[0,t]}$ to the process $\{\mathbf{Z}(s)\}_{s\in[0,t]}$. The plan for the proof of is the following. First, under the assumption of parent independent mutations,

- [Sect. 5.1] the convergence of generators (Proposition 4.1) is proved;

- [Sect. 5.2] the convergence of the corresponding semigroups is proved;

- [Sect. 5.3] the weak convergence (Theorem 2.1) is proved by showing the weak convergence of the extended processes on the extended state space (4.1), which relies on the technical framework of Section 4.

Then, in the general case of possibly parent dependent mutations,

- [Sect. 5.4] we use the change-of-measure approach (Theorem 3.1) to prove the weak convergence (Theorem 2.1).

5.1 Convergence of generators (PIM)

In this subsection, we prove Proposition 4.1. Assume the mutation probability matrix is of the form $Q = (Q_i)_{i=1}^d$. Then, the backwards transition probabilities are explicitly known:

$$\rho^{(n)}(\mathbf{v} \mid \mathbf{y}) = \begin{cases} \frac{y_j(ny_j-1)}{\|\mathbf{y}\|_1(ny_j-1+\theta Q_j)} & \text{if } \mathbf{v} = \mathbf{e}_j, \quad j = 1 \ldots d, \\ \frac{\theta Q_j y_j(ny_i-\delta_{ij}+\theta Q_i)}{\|\mathbf{y}\|_1(n\|\mathbf{y}\|_1-1+\theta)(ny_j-1+\theta Q_j)} & \text{if } \mathbf{v} = \mathbf{e}_j - \mathbf{e}_i, \quad i,j = 1 \ldots d, \\ 0 & \text{otherwise,} \end{cases} \tag{5.1}$$

see e.g. [6] and Appendix A.1. Given $f \in C_c^{\infty}(E)$, let

$$K = \{(\mathbf{y}, \mathbf{m}) \in E : y_j \geq \delta, m_{ij} \leq M, \forall i,j = 1, \ldots, d\}$$

be the compact set that contains the support of f, see Remark A.4 in Appendix A.2. Note that f and all of its derivatives are Lipschitz continuous with respect to ψ and bounded. From now on we consider $n > \frac{1}{\delta}$.

First, we assume that $(\mathbf{y}, \mathbf{m}) \in E^{(n)} \cap K^c$, which implies that $f = Af = 0$ in a neighbourhood of $(\mathbf{y}, \mathbf{m})$, and that the points $(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij})$ and $(\mathbf{y} - \frac{1}{n}\mathbf{e}_j, \mathbf{m})$ belong to $K^c, \forall n \in \mathbb{N}_{>0}, i,j = 1, \ldots, d$. If also $(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_i, \mathbf{m} + \mathbf{e}_{ij}) \in K^c, \forall n \in \mathbb{N}_{>0}, i,j = 1, \ldots, d$, then, $A^{(n)}\eta_n f(\mathbf{y}, \mathbf{m}) = \eta_n Af(\mathbf{y}, \mathbf{m}) = 0$. Otherwise, it must be that for a unique i_0, and for some n, we have that $\delta - \frac{1}{n} \leq y_{i_0} < \delta$; while $y_j \geq \delta$, for all $j \neq i_0$, and $m_{ij} \leq M$, for all $i,j = 1, \ldots, d$. In this case,

$$A^{(n)}\eta_n f(\mathbf{y}, \mathbf{m}) = \left| \sum_{j \neq i_0} \eta_n f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_{i_0}, \mathbf{m} + \mathbf{e}_{i_0 j}\right) n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_{i_0} \mid \mathbf{y}) \right|.$$

Note that, since $y_j \geq \delta$, for $j \neq i_0$, we have $n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_{i_0} \mid \mathbf{y}) < \frac{n\theta Q_j}{n\delta - 1 + \theta Q_j}$, which is bounded by a constant, independent of $\mathbf{y}$. Furthermore, $\left| \eta_n f \left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_{i_0}, \mathbf{m} + \mathbf{e}_{i_0 j} \right) \right|$ is bounded by the supremum of $|f(\mathbf{y}', \mathbf{m}')|$ over those $(\mathbf{y}', \mathbf{m}') \in E^{(n)}$ having y_i' such that $\delta - \frac{1}{n} \leq y_i' < \delta$, for some i. This supremum does not depend on $(\mathbf{y}, \mathbf{m})$ and converges to 0 as $n \to \infty$ since f is continuous with support in K.

Assume next that $(\mathbf{y}, \mathbf{m}) \in E^{(n)} \cap K$, which implies $\mathbf{y} \in \frac{1}{n}\mathbb{N}^d \setminus \Omega_0$. It is to be shown that the difference in the next display is bounded by some function of n, which does not depend on $\mathbf{y}$ and $\mathbf{m}$, and vanishes as $n \to \infty$. The bound is constructed as

$$\left| A^{(n)}\eta_n f(\mathbf{y}, \mathbf{m}) - \eta_n A f(\mathbf{y}, \mathbf{m}) \right|$$

$$= \left| A^{(n)} f(\mathbf{y}, \mathbf{m}) - A f(\mathbf{y}, \mathbf{m}) \right|$$

$$\leq \sum_{j=1}^{d} \left| \frac{f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j, \mathbf{m}\right) - f(\mathbf{y}, \mathbf{m})}{\frac{1}{n}} \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y}) + \frac{\partial f}{\partial y_j}(\mathbf{y}, \mathbf{m}) \frac{y_j}{\|\mathbf{y}\|_1} \right|$$

$$+ \sum_{i,j=1}^{d} \left| \left[f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_i, \mathbf{m} + \mathbf{e}_{ij}\right) - f(\mathbf{y}, \mathbf{m}) \right] n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y}) \right.$$

$$\left. - [f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})] \lambda_{Q,ij}(\mathbf{y}) \right|$$

$$\leq \sum_{j=1}^{d} \left| \frac{f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j, \mathbf{m}\right) - f(\mathbf{y}, \mathbf{m})}{\frac{1}{n}} + \frac{\partial f}{\partial y_j}(\mathbf{y}, \mathbf{m}) \right| \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y})$$

$$+ \sum_{j=1}^{d} \left| \frac{\partial f}{\partial y_j}(\mathbf{y}, \mathbf{m}) \right| \left| \frac{y_j}{\|\mathbf{y}\|_1} - \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y}) \right|$$

$$+ \sum_{i,j=1}^{d} \left| f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_i, \mathbf{m} + \mathbf{e}_{ij}\right) - f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) \right| n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y})$$

$$+ \sum_{i,j=1}^{d} |f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})| \left| n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y}) - \lambda_{ij}(\mathbf{y}) \right|. \tag{5.2}$$

A bound for each term of the sums in the right hand side of (5.2) can be obtained as follows. Applying the mean value theorem yields

$$f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j, \mathbf{m}\right) - f(\mathbf{y}, \mathbf{m}) = -\frac{1}{n}\frac{\partial f}{\partial y_j}\left(a\mathbf{y} + (1-a)\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j\right), \mathbf{m}\right),$$

for some $a \in (0,1)$, therefore, using also that $\rho^{(n)}(\mathbf{e}_j \mid \mathbf{y}) \leq 1$, and that $\frac{\partial f}{\partial y_j}$ is Lipschitz continuous,

$$\left| \frac{f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j, \mathbf{m}\right) - f(\mathbf{y}, \mathbf{m})}{\frac{1}{n}} + \frac{\partial f}{\partial y_j}(\mathbf{y}, \mathbf{m}) \right| \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y})$$

$$= \left| \frac{\partial f}{\partial y_j}\left(a\mathbf{y} + (1-a)\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j\right), \mathbf{m}\right) - \frac{\partial f}{\partial y_j}(\mathbf{y}, \mathbf{m}) \right| \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y})$$

$$\leq \|\nabla_{\mathbf{y}}^2 f\|_\infty \, \psi_1\left(a\mathbf{y} + (1-a)\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j\right), \mathbf{y}\right)$$

$$= \|\nabla_{\mathbf{y}}^2 f\|_\infty \frac{1-a}{y_j(ny_j - 1 + a)}$$

$$\leq \|\nabla_{\mathbf{y}}^2 f\|_\infty \frac{1-a}{\delta(n\delta - 1 + a)}.$$

For the second term in (5.2), by the explicit expression of $\rho^{(n)}(\mathbf{e}_j \mid \mathbf{y})$ in (5.1), it follows that,

$$\left| \frac{\partial f}{\partial y_j}(\mathbf{y}, \mathbf{m}) \right| \left| \frac{y_j}{\|\mathbf{y}\|_1} - \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y}) \right| = \left| \frac{\partial f}{\partial y_j}(\mathbf{y}, \mathbf{m}) \right| \frac{y_j}{\|\mathbf{y}\|_1} \frac{\theta Q_j}{ny_j - 1 + \theta Q_j}$$

$$\leq \left\| \frac{\partial f}{\partial y_j} \right\|_\infty \frac{\theta Q_j}{n\delta - 1 + \theta Q_j}.$$

For the third term in (5.2), since $\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y}) \leq \frac{\theta Q_j}{n\delta - 1 + \theta Q_j} < 1$, and f is Lipschitz continuous,

$$\left| f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_i, \mathbf{m} + \mathbf{e}_{ij}\right) - f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) \right| n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y})$$

$$\leq \|\nabla_{\mathbf{y}} f\|_\infty \psi_1 \left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_i, \mathbf{y}\right)$$

$$= \|\nabla_{\mathbf{y}} f\|_\infty \left\| \frac{\mathbf{e}_j - \mathbf{e}_i}{\mathbf{y}(n\mathbf{y} - \mathbf{e}_j + \mathbf{e}_i))} \right\|_2$$

$$\leq \|\nabla_{\mathbf{y}} f\|_\infty \left(\frac{1}{y_i(ny_i + 1)} + \frac{1}{y_j(ny_j - 1)} \right)$$

$$\leq \|\nabla_{\mathbf{y}} f\|_\infty \left(\frac{1}{\delta(n\delta + 1)} + \frac{1}{\delta(n\delta - 1)} \right).$$

Finally, for the last term in (5.2), recalling that in the PIM case $\lambda_{ij}(\mathbf{y}) = \frac{\theta Q_j y_i}{\|\mathbf{y}\|_1^2}$,

$$|f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})| \left| n\rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y}) - \lambda_{ij}(\mathbf{y}) \right|$$

$$\leq 2\|f\|_\infty \frac{\theta Q_j}{\|\mathbf{y}\|_1} \left| \frac{ny_j}{ny_j - 1 + \theta Q_j} \frac{ny_i - \delta_{ij} + Q_i}{n\|\mathbf{y}\|_1 - 1 + \theta} - \frac{y_i}{\|\mathbf{y}\|_1} \right|$$

$$\leq 2\|f\|_\infty \frac{\mathrm{const}_\theta}{\delta(n\delta - 1)}.$$

Since the bound for (5.2) does not depend on $(\mathbf{y}, \mathbf{m})$ and vanishes as n goes to infinity, the proof is complete.

5.2 Convergence of semigroups (PIM)

In this subsection, we prove the convergence of the semigroups associated to the generators of the previous subsection under the assumption of parent independent mutations. Take $f \in \hat{C}(E)$, the space of continuous real-valued function on (E, ψ) vanishing at infinity, see Remark A.4 for a characterisation. Let $T^{(n)}$ be the semigroup associated to the discrete-time Markov chain $\mathbf{Z}^{(n)}$, that is,

$$T^{(n)} \eta_n f(\mathbf{y}, \mathbf{m}) = \sum_{j=1}^{d} \eta_n f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j, \mathbf{m}\right) \rho^{(n)}(\mathbf{e}_j \mid \mathbf{y})$$

$$+ \sum_{i,j=1}^{d} \eta_n f\left(\mathbf{y} - \frac{1}{n}\mathbf{e}_j + \frac{1}{n}\mathbf{e}_i, \mathbf{m} + \mathbf{e}_{ij}\right) \rho^{(n)}(\mathbf{e}_j - \mathbf{e}_i \mid \mathbf{y}), \tag{5.3}$$

and let T be the semigroup associated to the Markov process $\mathbf{Z}$, described in the following. Let $\Omega_\infty := \{\mathbf{y} = (y_1, \ldots, y_d) : y_i = \infty \text{ for some } i\}$. If $\mathbf{y} \in E_1 \cap \Omega_\infty$, then $T(t)f(\mathbf{y}, \mathbf{m}) = f(\mathbf{y}, \mathbf{m})$, $\forall t \geq 0$. If $\mathbf{y} \in E_1 \setminus \Omega_\infty$, then, for $t \geq \|\mathbf{y}\|_1$, $T(t)f(\mathbf{y}, \mathbf{m}) = 0$, whereas, for $t < \|\mathbf{y}\|_1$,

$$T(t)f(\mathbf{y}, \mathbf{m}) = \sum_{\mathbf{w} \in \mathbb{N}^{d^2}} f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{w}\right) \gamma_{\mathbf{w}}(t, \mathbf{y}), \tag{5.4}$$

where, for $\mathbf{w} = (w_{ij})_{i,j=1,\ldots,d} \in \mathbb{N}^{d^2}$,

$$\gamma_{\mathbf{w}}(t, \mathbf{y}) = \prod_{i,j=1}^{d} \frac{\Lambda_{ij}(t,\mathbf{y})^{w_{ij}}}{w_{ij}!} e^{-\Lambda_{ij}(t,\mathbf{y})}. \tag{5.5}$$

with

$$\Lambda_{ij}(t,\mathbf{y}) = \int_0^t \lambda_{ij}\left(\frac{\mathbf{y}}{\|\mathbf{y}\|_1}\left(\|\mathbf{y}\|_1 - u\right)\right) du = \frac{\theta Q_j y_i}{\|\mathbf{y}\|_1} \log\left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_1 - t}\right)$$

We now prove that, for all $f \in \hat{C}(E)$, for all $t \geq 0$,

$$\lim_{n \to \infty} \sup_{(\mathbf{y},\mathbf{m}) \in E^{(n)}} \left| (T^{(n)})^{\lfloor tn \rfloor} \eta_n f(\mathbf{y}, \mathbf{m}) - \eta_n T(t) f(\mathbf{y}, \mathbf{m}) \right| = 0. \tag{5.6}$$

First, note that $T^{(n)}$ is a linear contraction and that, as shown in Appendix A.3, $\{T(t)\}_{t\geq 0}$ is a strongly continuous contraction semigroup associated to the generator A. Furthermore, $C_c^\infty(E)$ is a core for the generator A. This holds by Proposition 3.3 in [8, Ch.1], since $C_c^\infty(E)$ is dense in $\hat{C}(E)$, $T(t)$ maps $C_c^\infty(E)$ into $C_c^\infty(E)$ and A is the generator of a strongly continuous contraction semigroup. Therefore, by Theorem 6.5 in [8, Ch.1], the convergence of semigroups (5.6) is equivalent to the convergence of generators of Proposition 4.1, which concludes the proof.

5.3 Weak convergence (PIM)

In this subsection, we prove Theorem 2.1 under the assumption of parent independent mutations. We first prove the result for the extended processes (4.1), following [8]. The semigroup T, associated to the Markov process $\mathbf{Z}$, is extended to a conservative semigroup T^Δ as in e.g. [8, Ch. 4], which is associated to the extended Markov process $\mathbf{Z}^\Delta$, by letting, for $f \in \hat{C}(E^\Delta) = C(E^\Delta)$,

$$T^\Delta(t)f = f(\Delta) + T(t)(f - f(\Delta)). \tag{5.7}$$

In the expression above it is implied that $T(t)$ is applied to the restriction to E of the function $f(\cdot) - f(\Delta)$ and that $T^\Delta(t)f$ is a function on E^Δ, equal to $f(\Delta) + T(t)(f - f(\Delta))$ on E and to $f(\Delta)$ on Δ. By Lemma 2.3 in [8, Ch.4], the extended semigroup $\{T^\Delta(t)\}_{t\geq 0}$ inherits all the properties of $\{T(t)\}_{t\geq 0}$ and additionally it is conservative. Hence, it is a Feller semigroup.

By defining the bounded linear operator $\eta_n^\Delta f = f \circ \xi_n^\Delta$, the convergence of semigroups of the previous subsection can be naturally extended to, for all $f \in C(E^\Delta)$ and $t \geq 0$,

$$\lim_{n \to \infty} \sup_{(\mathbf{y},\mathbf{m}) \in E^{(n)}} \left| (T^{(n)})^{\lfloor tn \rfloor} \eta_n^\Delta f(\mathbf{y}, \mathbf{m}) - \eta_n^\Delta T^\Delta(t) f(\mathbf{y}, \mathbf{m}) \right| = 0.$$

Therefore, by Theorem 2.12 in [8, Ch. 4], there exists a Markov process $\mathbf{Z}^\Delta$, associated to the semigroup T^Δ, with sample paths in $D_{E^\Delta}[0,\infty)$, and $\xi_n(\tilde{\mathbf{Z}}^{(n)}) \to \mathbf{Z}^\Delta$ in $D_{E^\Delta}[0,\infty)$. Since T^Δ is the extension of the generator T associated to the process $\mathbf{Z}$, $\mathbf{Z}^\Delta$ coincides with $\mathbf{Z}$ on E.

This easily implies Theorem 2.1, as shown in the following. Since $\mathbf{Y}(0) = \mathbf{y}_0$ and $t < \|\mathbf{y}_0\|_1$, the limiting process at time t is not at the cemetery point, that is, $\mathbf{Z}^\Delta(t) = \mathbf{Z}(t) \in E$. Letting for example $\delta = \frac{1}{2} \min_{i=1,\ldots,d}\{\tilde{Y}_i(t)\}$, which is positive, the convergence of the extended processes (4.1) implies that the probability of all components of $\{\tilde{Y}^{(n)}(s)\}_{0\leq s\leq t}$ being larger than δ converges to 1. Being the paths up to time t bounded away from Ω_0, the convergence of the extended processes (4.1), restricted to the paths up to time t, is equivalent to the convergence of the non-extended processes written in terms of the Euclidean metric.

5.4 Weak convergence (general mutation)

In this subsection, we prove Theorem 2.1 for a general mutation matrix P, knowing from the previous subsection that the theorem holds for a PIM mutation matrix Q, which we assume positive, and using the change-of-measure approach, i.e. Theorem 3.1.

First, we prove the following lemma on the asymptotic properties of the sequences $r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t))$ and $c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))$ which constitute the Radon–Nikodym derivative of the change of measure from P to Q appearing in Theorem 3.1. We use Theorem 2.1 for the sequence $\mathbf{Z}_Q^{(n)}$ and the asymptotic results in [10] for the sampling probabilities.

Remark 5.1. For the following proofs it is useful to recall that, for a sequence of random variables that converges in distribution, convergence of the $\mathcal{L}^1$-norms occurs if and only if the sequence is uniformly integrable.

Lemma 5.2. Let $\mathbf{y}_0 \in \mathbb{R}_+^d$ and $\mathbf{y}_0^{(n)} \in \frac{1}{n}\mathbb{N}^d \setminus \{\mathbf{0}\}, n \in \mathbb{N}$. Assume $\mathbf{Y}_Q^{(n)}(0) = \mathbf{y}_0^{(n)}$, and $\mathbf{y}_0^{(n)} \to \mathbf{y}_0$ as $n \to \infty$. Then, for all $t \in [0, \|\mathbf{y}_0\|_1)$,

(i) $\mathbb{E}\left[c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)})\right] = 1$;

(ii) $\mathbb{E}\left[c_{P,Q}(\mathbf{M}_Q(t))\right] = 1$;

(iii) $r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) \to 1$ in probability, as $n \to \infty$;

(iv) $c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) \to c_{P,Q}(\mathbf{M}_Q(t))$ in distribution, as $n \to \infty$;

(v) $\{c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)})\}_{n \in \mathbb{N}}$ is a uniformly integrable sequence.

Proof. To prove (i) and (ii), simply apply Theorem 3.1 with $g = 1$.

To prove (iii), first recall that, by [10, Thm 4.3], $n^{d-1}p_P(n\mathbf{y}^{(n)}) \to \|\mathbf{y}\|_1^{1-d}\tilde{p}_P(\mathbf{y})$, as $\mathbf{y}^{(n)} \to \mathbf{y}$, where $\tilde{p}_P$ is the stationary density of the Wright-Fisher diffusion that is dual to the Kingman coalescent. Thus

$$\lim_{n\to\infty} r_{P,Q}^{(n)}(\mathbf{y}^{(n)}, \mathbf{y}_0^{(n)}) = \frac{\tilde{p}_P\left(\frac{\mathbf{y}}{\|\mathbf{y}\|_1}\right)\tilde{p}_Q\left(\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|_1}\right)}{\tilde{p}_Q\left(\frac{\mathbf{y}}{\|\mathbf{y}\|_1}\right)\tilde{p}_P\left(\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|_1}\right)}.$$

Furthermore, Theorem 2.1, which is already proven under the PIM assumption, implies that $\tilde{\mathbf{Y}}_Q^{(n)}(t)$ converges in probability to the constant $\mathbf{Y}(t)$, which is equal to $\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|_1}\left(\|\mathbf{y}_0\|_1 - t\right)$. By combining the two convergences above and using the characterization of convergence in probability in terms of almost surely converging subsequences, it is straightforward to conclude that $r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)})$ converges in probability to 1, since

$$\frac{\mathbf{Y}(t)}{\|\mathbf{Y}(t)\|_1} = \frac{\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|_1}\left(\|\mathbf{y}_0\|_1 - t\right)}{\left\|\frac{\mathbf{y}_0}{\|\mathbf{y}_0\|_1}\left(\|\mathbf{y}_0\|_1 - t\right)\right\|_1} = \frac{\mathbf{y}_0}{\|\mathbf{y}_0\|_1}.$$

Proving (iv) is rather straightforward. In fact, since $c_{P,Q}$ is a continuous function, and since, by Theorem 2.1, $\tilde{\mathbf{M}}_Q^{(n)}(t)$ converges in distribution to $\mathbf{M}_Q(t)$, $c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))$ converges in distribution to $c_{P,Q}(\mathbf{M}_Q(t))$, by the continuous mapping theorem. Combining this with (iii) and using Cramér-Slutzky's theorem proves (iv).

Finally, (v) holds since the sequence $c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)})$ converges in distribution to $c_{P,Q}(\mathbf{M}_Q(t))$, as $n \to \infty$, by (iv), and the expectations are constantly equal to 1, by (i), (ii), thus by Remark 5.1 the sequence is uniformly integrable. $\qquad\square$

We now prove Theorem 2.1 for a general mutation matrix P, using properties (iii) and (v) of Lemma 5.2, Theorem 3.1 and the PIM version of Theorem 2.1 (applied to Q).

Fix a bounded continuous function $g : D_{\mathbb{R}_+^d \times \mathbb{N}^{d^2}}[0, t] \to \mathbb{R}$. To shorten the notation, let $G^{(n)}(t) = g(\{\tilde{\mathbf{Z}}_Q^{(n)}(s)\}_{0 \le s \le t})$ and $G(t) = g(\{\mathbf{Z}_Q(s)\}_{0 \le s \le t})$. Because of Theorem 3.1, proving Theorem 2.1 is equivalent to proving that

$$\mathbb{E}\left[G^{(n)}(t) r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))\right] \xrightarrow[n \to \infty]{} \mathbb{E}\left[G(t) c_{P,Q}(\mathbf{M}_Q(t))\right]. \qquad (5.8)$$

By the PIM version of Theorem 2.1 $\{\tilde{\mathbf{Z}}_Q^{(n)}(s)\}_{0 \le s \le t} \to \{\mathbf{Z}_Q(s)\}_{0 \le s \le t}$ in $D_{\mathbb{R}_+^d \times \mathbb{N}^{d^2}}[0, t]$. Thus, since $g \cdot c_{P,Q}$ is a continuous function, the continuous mapping theorem implies that $G^{(n)}(t) c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t)) \to G(t) c_{P,Q}(\mathbf{M}_Q(t))$, in distribution as $n \to \infty$. Therefore, since $r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) \to 1$ in probability, as $n \to \infty$, by Lemma 5.2 (iii), Cramér- Slutzky's theorem implies

$$G^{(n)}(t) c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t)) r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) \xrightarrow[n \to \infty]{} G(t) c_{P,Q}(\mathbf{M}_Q(t)),$$

in distribution. Furthermore, the sequence $\{G^{(n)}(t) r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))\}_{n \in \mathbb{N}}$ is uniformly integrable, because $\{r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))\}_{n \in \mathbb{N}}$ is uniformly integrable, by Lemma 5.2 (v), and $G^{(n)}(t)$ is bounded.

Uniform integrability together with the convergence in distribution implies that $\mathbb{E}\left[|G^{(n)}(t)| r_{P,Q}^{(n)}(\tilde{\mathbf{Y}}_Q^{(n)}(t), \mathbf{y}_0^{(n)}) c_{P,Q}(\tilde{\mathbf{M}}_Q^{(n)}(t))\right]$ converges to $\mathbb{E}\left[|G(t)| c_{P,Q}(\mathbf{M}_Q(t))\right]$ by Remark 5.1. Therefore, if g is nonnegative Theorem 2.1 is proved, otherwise write g as the difference between its positive and negative part and apply the previous argument on both parts.

The proof of Theorem 2.1 is completed. Note also that (4.1) is implied. In fact, with a similar argument to the one in the previous subsection, convergence of $\xi_n^\Delta(\{\tilde{\mathbf{Z}}_P^{(n)}(s)\}_{0 \le s \le t})$ to $\{\mathbf{Z}_P^\Delta(s)\}_{0 \le s \le t}$ in $D_{E^\Delta}[0, t]$, is implied, since t is fixed and $t < \|\mathbf{y}_0\|_1$. Furthermore, the convergence in $D_{E^\Delta}[0, t_1]$ for all $0 < t_1 < |\mathbf{y}_0|$ implies the convergence in $D_{E^\Delta}[0, t_2]$ for all $t_2 > |\mathbf{y}_0|$. This holds by e.g [28, Ch. IV], defining the approximation map $a : D_{E^\Delta}[0, t_2] \to D_{E^\Delta}[0, t_1]$, as $a_{t_1}(z)(s) = z(s \wedge t_1)$ and observing that the previous part of the proof implies that, for all $0 < t_1 < |\mathbf{y}_0|$, $a_{t_1}(\xi_n^\Delta(\tilde{\mathbf{Z}}_P^{(n)})) \to a_{t_1}(\mathbf{Z}_P^\Delta)$ in $D_{E^\Delta}[0, t_2]$. Therefore $\xi_n^\Delta(\tilde{\mathbf{Z}}_P^{(n)})$ converges to $\mathbf{Z}_P^\Delta$ in $D_{E^\Delta}[0, t]$ for all $t > 0$. Convergence in $D_{E^\Delta}[0, t]$ for all $t > 0$ implies convergence in $D_{E^\Delta}[0, \infty)$ by using for example [3, Thm 16.7], with the set of continuity points being $\mathbb{R}_{\ge 0}$ since for all $t > 0$ the probability that the limiting process $\mathbf{Z}_P^\Delta$ is discontinuous at t is zero.

A Appendix

A.1 Backwards transition probabilities

The backward transition probabilities of the block-counting jump chain of the coalescent have the form

$$\rho^{(n)}(\mathbf{v} \mid \mathbf{y}) = \begin{cases} \frac{n y_i (n y_i - 1)}{n \|\mathbf{y}\|_1 (n \|\mathbf{y}\|_1 - 1 + \theta)} \frac{1}{\pi(i \mid n\mathbf{y} - \mathbf{e}_i)} & \text{if } \mathbf{v} = \mathbf{e}_i, \quad i = 1 \dots d, \\ \frac{n y_i \theta P_{ji}}{n \|\mathbf{y}\|_1 (n \|\mathbf{y}\|_1 - 1 + \theta)} \frac{\pi(j \mid n\mathbf{y} - \mathbf{e}_i)}{\pi(i \mid n\mathbf{y} - \mathbf{e}_i)} & \text{if } \mathbf{v} = \mathbf{e}_i - \mathbf{e}_j, \quad i, j = 1 \dots d, \\ 0 & \text{otherwise}, \end{cases}$$

where $\pi(i \mid n\mathbf{y}) = \frac{n y_i + 1}{n \|\mathbf{y}\|_1 + 1} \frac{p(n\mathbf{y} + \mathbf{e}_i)}{p(n\mathbf{y})}$ is the probability of sampling type i after having sampled configuration $n\mathbf{y}$, and p is the sampling probability described in Section 3, see

e.g. [6]. The expression above can be obtained from the expression of the forward transition probabilities by employing Bayes' rule. In the PIM case, it is known that π can be explicitly written as

$$\pi(i \mid n\mathbf{y}) = \frac{ny_i + \theta Q_i}{n \left\|\mathbf{y}\right\|_1 + \theta},$$

which gives an explicit expression also for the backwards transition probabilities, given in (5.1).

A.2 Metric state-spaces and spaces of continuous functions

The goal of this section is to analyse the metric space (E_1, ψ_1), in order to show that compact sets are bounded away from the boundary Ω_0, Remark A.2, and to characterise continuous functions on the metric space (E, ψ), Remark A.3, with the properties of compact support and vanishing at infinity, Remark A.4. Characterisations in terms of the Euclidean metric are provided.

Let $S_1 := \mathbb{R}_+^d$ and $S := S_1 \times \mathbb{N}^{d^2}$ be equipped with the Euclidean metric $\left\|\cdot\right\|_2$. Recall that $\Omega_0 = \{\mathbf{y} = (y_1, \ldots, y_d) : y_i = 0 \text{ for some } i\}$ and $\Omega_\infty := \{\mathbf{y} = (y_1, \ldots, y_d) : y_i = \infty \text{ for some } i\}$. Let $\omega : E_1 \to S_1$ be the homeomorphism defined by $\omega(\mathbf{y}) = \left(\frac{1}{y_1}, \ldots, \frac{1}{y_d}\right)$, with the convention $\frac{1}{\infty} = 0$ and $\frac{1}{0} = \infty$. Recall that for $\mathbf{y}_1, \mathbf{y}_2 \in E_1$,

$$\psi_1(\mathbf{y}_1, \mathbf{y}_2) = \left\|\omega(\mathbf{y}_1) - \omega(\mathbf{y}_2)\right\|_2.$$

By equipping E_1 with the metric ψ_1, the roles of 0 and ∞ are switched, whereas, in the rest of the space, i.e. in $E_1 \setminus \Omega_\infty = S_1 \setminus \Omega_0 = (0, \infty)^d$, the metric ψ_1 is equivalent to the Euclidean metric. In fact, if $\{\mathbf{y}^{(n)}\}_{n=1}^\infty \subset (0, \infty)^d$ and $\mathbf{y} \in (0, \infty)^d$ then $\psi_1(\mathbf{y}^{(n)}, \mathbf{y}) \to 0$ if and only if $\left\|\mathbf{y}^{(n)} - \mathbf{y}\right\|_2 \to 0$. This allows to characterise closed sets and compact sets of (E_1, ψ_1) in terms of the Euclidean metric.

Remark A.1 (Closed sets). A set $C \subset (E_1, \psi_1)$ is closed if and only if $\omega(C) = \{\mathbf{x} \in S_1 : \mathbf{x} = \frac{1}{\mathbf{y}} \text{ for some } \mathbf{y} \in E_1\}$ is closed in $(S_1, \left\|\cdot\right\|_2)$. Equivalently, C is closed if and only if one of the following conditions is satisfied

- $C \cap \Omega_\infty = \emptyset$ and C is bounded and closed in $(E_1 \setminus \Omega_\infty, \left\|\cdot\right\|_2)$;

- $C \setminus \Omega_\infty$ is closed in $(E_1 \setminus \{\infty\}, \left\|\cdot\right\|_2)$ and $\mathbf{y} \in C \cap \Omega_\infty$ for all $\mathbf{y} \in \Omega_\infty$ such that there exists a sequence $\{\mathbf{y}^{(n)}\}_{n \in \mathbb{N}} \subset C$, with $y_i^{(n)} \to y_i = \infty$, for $i \in \mathcal{I} \neq \emptyset$, and $y_j^{(n)} \to y_j \in (0, \infty)$, for $j \in \{1, \ldots, d\} \setminus \mathcal{I}$.

Remark A.2 (Compact sets). A compact set in (E_1, ψ_1) is a closed set which is bounded away from Ω_0. More precisely, a set $K \subset (E_1, \psi_1)$ is compact if and only if K is closed in (E_1, ψ_1) and there exists $\delta > 0$ such that, for all $\mathbf{y} \in K$, $y_j > \delta$, $j = 1, \ldots, d$.

Furthermore, real-valued continuous functions on (E_1, ψ_1) are simply continuous functions on $(E_1 \setminus \Omega_\infty, \left\|\cdot\right\|_2)$ that can be continuously extended on Ω_∞, as specified in the following.

Remark A.3 (Continuous functions). A function $f : (E_1, \psi_1) \to \mathbb{R}$ is continuous if and only if the two following conditions are satisfied

- its restriction $f : (E_1 \setminus \Omega_\infty, \left\|\cdot\right\|_2) \to \mathbb{R}$ is continuous

- $f(\mathbf{y}^{(n)}) \to f(\mathbf{y})$ for any sequence $\{\mathbf{y}^{(n)}\}_{n \in \mathbb{N}} \subset E_1 \setminus \Omega_\infty$, with $y_i^{(n)} \to y_i = \infty$, for $i \in \mathcal{I} \neq \emptyset$, and $y_j^{(n)} \to y_j \in (0, \infty)$, for $j \in \{1, \ldots, d\} \setminus \mathcal{I}$.

Since $E = E_1 \times \mathbb{N}^{d^2}$ is equipped with the product metric $\psi = \psi_1 \oplus \left\|\cdot\right\|_2$, and $\mathbb{N}^{d^2}$ is a discrete space, the above characterisations can be easily extended to (E, ψ). We focus

now on spaces of continuous functions on (E, ψ). In particular, a continuous function on (E, ψ) is simply a function which is continuous with respect to its component in E_1 in the sense of Remark A.3. We say that a function vanishes at infinity if, given any $\epsilon > 0$, there exists a compact set $K_\epsilon \subset E$ such that $|f(\mathbf{z})| < \epsilon$, $\forall \mathbf{z} \in K_\epsilon^c$. Since the component in (E_1, ψ_1) of a compact set is bounded away from Ω_0 and the component in $\mathbb{N}^{d^2}$ is bounded, the following holds.

Remark A.4 (Vanishing at infinity, compact support)**.** A function $f : (E, \psi) \to \mathbb{R}$

- vanishes at infinity if and only if, for any given $\epsilon > 0$, there exist $\delta_\epsilon, M_\epsilon > 0$ such that

$$
\begin{cases}
y_j < \delta_\epsilon \text{ for some } j = 1, \ldots, d \\
\text{or} \\
m_{ij} > M_\epsilon \text{ for some } i, j = 1, \ldots, d
\end{cases}
\Rightarrow |f(\mathbf{y}, \mathbf{m})| < \epsilon
$$

- has compact support if and only if there exist $\delta, M > 0$ such that

$$
\begin{cases}
y_j < \delta \text{ for some } j = 1, \ldots, d \\
\text{or} \\
m_{ij} > M \text{ for some } i, j = 1, \ldots, d
\end{cases}
\Rightarrow f(\mathbf{y}, \mathbf{m}) = 0
$$

The characterisations of this section provide a description of the spaces of continuous functions on (E, ψ) vanishing at infinity, $\hat{C}(E)$, and of smooth functions on (E, ψ) with compact support, $C_c^\infty(E)$. Furthermore, note that $C_c^\infty(E)$ is dense in $\hat{C}(E)$ which is dense in the space of bounded continuous functions on (E, ψ). Moreover, all functions in $C_c^\infty(E)$ and $\hat{C}(E)$ are uniformly continuous, while functions in $C_c^\infty(E)$, and all of their derivatives, are Lipschitz continuous.

Finally, we characterise continuous functions on the extended state-space $E^\Delta = E \cup \{\Delta\}$, which is the one-point, or Alexandroff, compactification of E equipped with the standard topology inherited by the topology in E. Since E is a separable locally compact metric space, its one-point compactification is metrizable, see e.g. [24]. Let (E^Δ, ψ^Δ) be the corresponding metric space.

Remark A.5 (Extended continuous functions)**.** A function $f : (E^\Delta, \psi^\Delta) \to \mathbb{R}$ is continuous if and only if the following conditions are satisfied

- its restriction $f : ((E_1 \setminus \Omega_\infty) \times \mathbb{N}^{d^2}, \|\cdot\|_2) \to \mathbb{R}$ is continuous;

- $\forall \mathbf{m} \in \mathbb{N}^{d^2}, f(\mathbf{y}^{(n)}, \mathbf{m}) \to f(\mathbf{y}, \mathbf{m})$ for any sequence $\{\mathbf{y}^{(n)}\}_{n \in \mathbb{N}} \subset E_1 \setminus \Omega_\infty$, with $y_i^{(n)} \to y_i = \infty$, for $i \in \mathcal{I} \neq \emptyset$, and $y_j^{(n)} \to y_j \in (0, \infty)$, for $j \in \{1, \ldots, d\} \setminus \mathcal{I}$;

- $f(\mathbf{y}^{(n)}, \mathbf{m}^{(n)}) \to f(\Delta)$ for any sequence $\{(\mathbf{y}^{(n)}, \mathbf{m}^{(n)})\}_{n \in \mathbb{N}} \subset (E_1 \setminus \Omega_\infty) \times \mathbb{N}^{d^2}$, with $y_i^{(n)} \to 0$, for some $i \in \{1, \ldots, d\}$ or $m_{ij}^{(n)} \to \infty$ for some $i, j \in \{1, \ldots, d\}$.

Note that, while using the metrics ψ and ψ^Δ allows for a more compact formulation, all the statements concerning continuous functions on E or E^Δ, can be expressed in terms of the Euclidean metric by using the remarks of this section.

A.3 Infinitesimal generator and semigroup of the limiting process (PIM)

In this section, under the assumption of parent independent mutations, we show that the semigroup T, defined in (5.3), is a positive, strongly continuous, contraction semigroup and that the infinitesimal generator A, associated to the semigroup T, is

indeed of the form (2.3). The following, straightforward to derive, expressions for the probabilities in (5.5) are needed

$$\gamma_{\mathbf{0}}(t,\mathbf{y}) = \left(\frac{\|\mathbf{y}\|_1 - t}{\|\mathbf{y}\|_1}\right)^{\theta}, \tag{A.1}$$

$$\gamma_{\mathbf{e}_{ij}}(t,\mathbf{y}) = \left(\frac{\|\mathbf{y}\|_1 - t}{\|\mathbf{y}\|_1}\right)^{\theta} \frac{\theta Q_j y_i}{\|\mathbf{y}\|_1} \log\left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_1 - t}\right), \quad i,j = 1,\ldots,d. \tag{A.2}$$

Furthermore, since the sum of independent Poisson random variables is Poisson distributed, for all $k \in \mathbb{N}$,

$$\sum_{\mathbf{w}\in\mathbb{N}^{d^2}:\|\mathbf{w}\|_1\geq k} \gamma_{\mathbf{w}}(t,\mathbf{y}) = \mathbb{P}\left(\|\mathbf{M}(t+s) - \mathbf{M}(s)\|_1 \geq k\right) = \sum_{h=k}^{\infty} \frac{\Lambda(t,\mathbf{y})^h}{h!} e^{-\Lambda(t,\mathbf{y})}, \tag{A.3}$$

where

$$\Lambda(t,\mathbf{y}) = \sum_{i,j=1}^{d} \Lambda_{ij}(t,\mathbf{y}) = \theta \log\left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_1 - t}\right).$$

Semigroup

T is a positive contraction semigroup, being the semigroup associated to a Markov process. It remains to show that, for all $f \in \hat{C}(E)$,

$$\lim_{t\to 0} \sup_{(\mathbf{y},\mathbf{m})\in E} |T(t)f(\mathbf{y},\mathbf{m}) - f(\mathbf{y},\mathbf{m})| = 0.$$

Take $f \in \hat{C}(E)$, f is continuous and vanishing in the sense of Remark A.4, thus it is uniformly continuous with respect to the metric ψ and bounded. Given $\epsilon > 0$, let $K_\epsilon = \{(\mathbf{y},\mathbf{m}) \in E : y_j \geq \delta_\epsilon, m_{ij} \leq M_\epsilon, \forall i,j = 1,\ldots,d\}$, for some $\delta_\epsilon, M_\epsilon > 0$, be the compact set such that $|f(\mathbf{z})| < \epsilon$, for all $\mathbf{z} \in K_\epsilon^c$. The supremum in the last display is analysed in the three different cases; $\mathbf{y} \in E_1 \cap \Omega_\infty$, and, when $\mathbf{y} \in E_1 \setminus \Omega_\infty$, $\|\mathbf{y}\|_1 \leq t$ and $\|\mathbf{y}\|_1 > t$.

If $\mathbf{y} \in E_1 \cap \Omega_\infty$, then $T(t)f(\mathbf{y},\mathbf{m}) - f(\mathbf{y},\mathbf{m}) = 0, \forall \mathbf{m} \in \mathbb{N}^{d^2}, \forall t \geq 0$. Assume now $\mathbf{y} \in E_1 \setminus \Omega_\infty$. If $\|\mathbf{y}\|_1 \leq t$, then $T(t)f(\mathbf{y},\mathbf{m}) = 0$, and for $t < \delta_\epsilon$, $\|\mathbf{y}\|_1 \leq t < \delta_\epsilon$ implies $(\mathbf{y},\mathbf{m}) \in K_\epsilon^c$ and thus $|f(\mathbf{y},\mathbf{m})| < \epsilon$. Therefore,

$$\sup_{\substack{(\mathbf{y},\mathbf{m})\in E \\ \|\mathbf{y}\|_1 \leq t \\ \mathbf{y}\notin\Omega_\infty}} |T(t)f(\mathbf{y},\mathbf{m}) - f(\mathbf{y},\mathbf{m})| < \epsilon.$$

If $\|\mathbf{y}\|_1 > t$, then, by (5.4),

$$|T(t)f(\mathbf{y},\mathbf{m}) - f(\mathbf{y},\mathbf{m})| = \left|\sum_{\mathbf{w}\in\mathbb{N}^{d^2}} f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1}t, \mathbf{m} + \mathbf{w}\right)\gamma_{\mathbf{w}}(t,\mathbf{y}) - f(\mathbf{y},\mathbf{m})\right|.$$

The remaining argument is divided into the cases $(\mathbf{y},\mathbf{m}) \in K_\epsilon^c$ and $(\mathbf{y},\mathbf{m}) \in K_\epsilon$. If $(\mathbf{y},\mathbf{m}) \in K_\epsilon^c$, then $\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1}t, \mathbf{m} + \mathbf{w}\right) \in K_\epsilon^c$ and $|f(\mathbf{y},\mathbf{m})|, f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1}t, \mathbf{m} + \mathbf{w}\right) < \epsilon$. Since $\sum_{\mathbf{w}\in\mathbb{N}^{d^2}} \gamma_{\mathbf{w}}(t) = 1$, it follows that

$$\sup_{\substack{(\mathbf{y},\mathbf{m})\in K_\epsilon^c \\ \|\mathbf{y}\|_1 > t \\ \mathbf{y}\notin\Omega_\infty}} |T(t)f(\mathbf{y},\mathbf{m}) - f(\mathbf{y},\mathbf{m})| < 2\epsilon.$$

If $(\mathbf{y}, \mathbf{m}) \in K_\epsilon$, consider the following inequality,

$$\left| \sum_{\mathbf{w} \in \mathbb{N}^{d^2}} f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{w}\right) \gamma_{\mathbf{w}}(t, \mathbf{y}) - f(\mathbf{y}, \mathbf{m}) \right|$$

$$\leq \left| f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m}\right) - f(\mathbf{y}, \mathbf{m}) \right| \gamma_{\mathbf{0}}(t, \mathbf{y})$$

$$+ |f(\mathbf{y}, \mathbf{m})| \left(1 - \gamma_{\mathbf{0}}(t, \mathbf{y})\right)$$

$$+ \sum_{\substack{\mathbf{w} \in \mathbb{N}^{d^2} \\ \|\mathbf{w}\|_1 \geq 1}} \left| f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{w}\right) \right| \gamma_{\mathbf{w}}(t, \mathbf{y}).$$

Since f is uniformly continuous, there exists $\eta_\epsilon > 0$, not depending on $\mathbf{y}$, such that $\left| f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m}\right) - f(\mathbf{y}, \mathbf{m}) \right| < \epsilon$, if $\psi\left(\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m}\right), (\mathbf{y}, \mathbf{m})\right) < \eta_\epsilon$. Keeping in mind that $\|\mathbf{y}\|_1, \|\mathbf{y}\|_2 > \delta_\epsilon$, note that

$$\psi\left(\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m}\right), (\mathbf{y}, \mathbf{m})\right) = \psi_1\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{y}\right) = \frac{t}{\|\mathbf{y}\|_1 - t} \sum_{i=1}^{d} \frac{1}{y_i} < \frac{t}{\delta_\epsilon - t} \frac{d}{\delta_\epsilon} \quad \text{(A.4)}$$

is bounded by η_ϵ, if $t < \eta'_\epsilon$, where $\eta'_\epsilon > 0$ only depends on $\delta_\epsilon, \eta_\epsilon$, not on $\mathbf{y}$. Recalling that f is bounded and using (A.1) yields

$$|f(\mathbf{y}, \mathbf{m})| \left(1 - \gamma_{\mathbf{0}}(t, \mathbf{y})\right) \leq \|f\|_\infty \left(1 - \left(1 - \frac{t}{\delta_\epsilon}\right)^\theta\right), \quad \text{(A.5)}$$

which is bounded by ϵ, if $t < \eta''_\epsilon$, for some $\eta''_\epsilon > 0$ which only depends on δ_ϵ, not on $\mathbf{y}$. Furthermore, using (A.3) and again that f is bounded yields

$$\sum_{\substack{\mathbf{w} \in \mathbb{N}^{d^2} \\ \|\mathbf{w}\|_1 \geq 1}} \left| f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{w}\right) \right| \gamma_{\mathbf{w}}(t, \mathbf{y})$$

$$\leq \|f\|_\infty \sum_{\substack{\mathbf{w} \in \mathbb{N}^{d^2} \\ \|\mathbf{w}\|_1 \geq 1}} \gamma_{\mathbf{w}}(t, \mathbf{y})$$

$$= \|f\|_\infty \left(\frac{\|\mathbf{y}\|_1 - t}{\|\mathbf{y}\|_1}\right)^\theta \sum_{h=1}^{\infty} \left[\theta \log\left(\frac{\|\mathbf{y}\|_1}{\|\mathbf{y}\|_1 - t}\right)\right]^h \frac{1}{h!}$$

$$\leq \|f\|_\infty \sum_{h=1}^{\infty} \left(\frac{\theta t}{\delta_\epsilon - t}\right)^h \frac{1}{h!}, \quad \text{(A.6)}$$

where the last quantity is bounded by ϵ if $t < \eta'''_\epsilon$, for some $\eta'''_\epsilon > 0$, which only depends on δ_ϵ, not on $\mathbf{y}$. Therefore, if $t < \min\{\eta'_\epsilon, \eta''_\epsilon, \eta'''_\epsilon\}$,

$$\sup_{\substack{(\mathbf{y}, \mathbf{m}) \in E \\ \|\mathbf{y}\|_1 > t \\ \mathbf{y} \notin \Omega_\infty}} |T(t) f(\mathbf{y}, \mathbf{m}) - f(\mathbf{y}, \mathbf{m})| < 3\epsilon,$$

which completes the proof.

Infinitesimal generator

Let A be the operator defined as in (2.3), for $f \in C_c^\infty(E)$,

$$Af(\mathbf{y}, \mathbf{m}) = -\left\langle \nabla_{\mathbf{y}} f(\mathbf{y}, \mathbf{m}), \frac{\mathbf{y}}{\|\mathbf{y}\|_1} \right\rangle + \sum_{i,j=1}^{d} \left[f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})\right] \lambda_{ij}(\mathbf{y}),$$

We show that A is the infinitesimal generator of the process $\mathbf{Z}$, associated to the semi-group T, by proving that

$$\lim_{t \to 0} \sup_{(\mathbf{y}, \mathbf{m}) \in E} \left| \frac{T(t) f(\mathbf{y}, \mathbf{m}) - f(\mathbf{y}, \mathbf{m})}{t} - A f(\mathbf{y}, \mathbf{m}) \right| = 0.$$

Take $f \in C_c^\infty(E)$, f is smooth with compact support in the sense of Remark A.4. Thus f and all of its derivatives are Lipschitz continuous with respect to ψ and bounded. Let $K = \{(\mathbf{y}, \mathbf{m}) \in E : y_j \geq \delta, m_{ij} \leq M, \forall i, j = 1, \ldots, d\}$ for some $\delta, M > 0$, be the compact set that contains the support of f, so that $f(\mathbf{y}, \mathbf{m}) = 0$, $\forall (\mathbf{y}, \mathbf{m}) \in K^c$.

Since the limit as $t \to 0$ is considered, assume $t < \delta$. Three cases are possible: $\mathbf{y} \in E_1 \cap \Omega_\infty$, and when $\mathbf{y} \in E_1 \setminus \Omega_\infty$, $\|\mathbf{y}\|_1 \leq t$ and $\|\mathbf{y}\|_1 > t$. If $\mathbf{y} \in E_1 \cap \Omega_\infty$, then $T(t) f(\mathbf{y}, \mathbf{m}) = f(\mathbf{y}, \mathbf{m})$ and $A f(\mathbf{y}, \mathbf{m}) = 0$. Assume now $\mathbf{y} \in E_1 \setminus \Omega_\infty$. If $\|\mathbf{y}\|_1 \leq t$, then $T(t) f(\mathbf{y}, \mathbf{m}) = 0$, and $\|\mathbf{y}\|_1 \leq t < \delta$ implies that $(\mathbf{y}, \mathbf{m}) \in K^c$, $(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) \in K^c, i, j = 1, \ldots, d$, and that f is equal to 0 in a neighbourhood of $(\mathbf{y}, \mathbf{m})$, hence $A f(\mathbf{y}, \mathbf{m}) = 0$. If $\|\mathbf{y}\|_1 > t$ and $(\mathbf{y}, \mathbf{m}) \in K^c$, then, as above, $f(\mathbf{y}, \mathbf{m}) = 0$ and $A f(\mathbf{y}, \mathbf{m}) = 0$. Furthermore $\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{w} \right) \in K^c, \forall \mathbf{w} \in \mathbb{N}^{d^2}$, and thus $T(t) f(\mathbf{y}, \mathbf{m}) = 0$

Consider now the supremum over all $(\mathbf{y}, \mathbf{m}) \in K$, with $\|\mathbf{y}\|_1 > t$, $\mathbf{y} \in E_1 \setminus \Omega_\infty$. The aim is to bound the absolute value in the new display (A.7) by some function depending on t and δ, not on $\mathbf{y}$ and $\mathbf{m}$, which vanishes as $t \to 0$. Start by considering the bound,

$$\left| \frac{T(t) f(\mathbf{y}, \mathbf{m}) - f(\mathbf{y}, \mathbf{m})}{t} - A f(\mathbf{y}, \mathbf{m}) \right|$$

$$\leq \left| \frac{f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} \right) - f(\mathbf{y}, \mathbf{m})}{t} \gamma_{\mathbf{0}}(t, \mathbf{y}) + \left\langle \nabla_{\mathbf{y}} f(\mathbf{y}, \mathbf{m}), \frac{\mathbf{y}}{\|\mathbf{y}\|_1} \right\rangle \right|$$

$$+ \sum_{i,j=1}^d \left| \frac{f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{e}_{ij} \right) - f(\mathbf{y}, \mathbf{m})}{t} \gamma_{\mathbf{e}_{ij}}(t, \mathbf{y}) - [f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})] \lambda_{Q,ij}(\mathbf{y}) \right|$$

$$+ \sum_{\mathbf{w} \in \mathbb{N}^{d^2}, \|w\|_1 \geq 2} \left| \frac{f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{w} \right) - f(\mathbf{y}, \mathbf{m})}{t} \right| \gamma_{\mathbf{w}}(t, \mathbf{y}). \tag{A.7}$$

Each of the three terms in the right hand side above are studied separately.

To bound the first term, apply the mean value theorem,

$$f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} \right) - f(\mathbf{y}, \mathbf{m}) = \left\langle \nabla_{\mathbf{y}} f\left(a\mathbf{y} + (1-a)\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t \right), \mathbf{m} \right), -\frac{\mathbf{y}}{\|\mathbf{y}\|_1} t \right\rangle$$

for some $a \in (0, 1)$, to obtain

$$\left| \frac{f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} \right) - f(\mathbf{y}, \mathbf{m})}{t} \gamma_{\mathbf{0}}(t, \mathbf{y}) + \left\langle \nabla_{\mathbf{y}} f(\mathbf{y}, \mathbf{m}), \frac{\mathbf{y}}{\|\mathbf{y}\|_1} \right\rangle \right|$$

$$\leq \left| \left\langle \nabla_{\mathbf{y}} f\left((\|\mathbf{y}\|_1 - (1-a)t) \frac{\mathbf{y}}{\|\mathbf{y}\|_1}, \mathbf{m} \right) - \nabla_{\mathbf{y}} f(\mathbf{y}, \mathbf{m}), \frac{\mathbf{y}}{\|\mathbf{y}\|_1} \right\rangle \right| \gamma_{\mathbf{0}}(t, \mathbf{y})$$

$$+ \left| \left\langle \nabla_{\mathbf{y}} f(\mathbf{y}, \mathbf{m}), \frac{\mathbf{y}}{\|\mathbf{y}\|_1} \right\rangle \right| (1 - \gamma_{\mathbf{0}}(t, \mathbf{y})). \tag{A.8}$$

Note that $\nabla_{\mathbf{y}} f$ is bounded and Lipschitz continuous and

$$\psi_1 \left((\|\mathbf{y}\|_1 - (1-a)t) \frac{\mathbf{y}}{\|\mathbf{y}\|_1}, \mathbf{y} \right) = \frac{(1-a)t}{\|\mathbf{y}\|_1 - (1-a)t} \sum_{i=1}^d \frac{1}{y_i} \leq \frac{t}{\delta - (1-a)t} \frac{d}{\delta}.$$

Using that $\frac{\mathbf{y}}{\|\mathbf{y}\|_1} \le 1$ and $\|\mathbf{y}\|_2, \|\mathbf{y}\|_1 > \delta$, and (A.1), (A.5) for $\gamma_{\mathbf{0}}$ and $1 - \gamma_{\mathbf{0}}$, it follows that (A.8) is bounded by

$$\left\|\nabla_{\mathbf{y}}^2 f\right\|_\infty \frac{dt}{\delta(\delta - (1-a)t)} + \left\|\nabla_{\mathbf{y}} f\right\|_\infty \left(1 - \left(1 - \frac{t}{\delta}\right)^\theta\right).$$

Next, consider the ij^{th} term of the sum in (A.7), which can be bounded by,

$$\left| \frac{f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{e}_{ij}\right) - f(\mathbf{y}, \mathbf{m})}{t} \gamma_{\mathbf{e}_{ij}}(t, \mathbf{y}) - [f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})] \frac{\theta Q_j y_i}{\|\mathbf{y}\|_1^2} \right|$$

$$\le \left| f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{e}_{ij}\right) - f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) \right| \frac{1}{t} \gamma_{\mathbf{e}_{ij}}(t, \mathbf{y})$$

$$+ |f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})| \left| \frac{\gamma_{\mathbf{e}_{ij}}(t, \mathbf{y})}{t} - \frac{\theta Q_j y_i}{\|\mathbf{y}\|_1^2} \right|. \tag{A.9}$$

Recalling that f is Lipschitz continuous, using the bound (A.4) for the distance and expression (A.2) for $\gamma_{\mathbf{e}_{ij}}$, yields,

$$\frac{1}{t} \gamma_{\mathbf{e}_{ij}}(t, \mathbf{y}) \left| f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{e}_{ij}\right) - f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) \right|$$

$$\le -\frac{1}{t} \theta Q_j \log\left(1 - \frac{t}{\delta}\right) \|\nabla_{\mathbf{y}} f\|_\infty \ \psi_1\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{y}\right)$$

$$< -\theta \|\nabla_{\mathbf{y}} f\|_\infty \frac{1}{t} \log\left(1 - \frac{t}{\delta}\right) \frac{td}{\delta(\delta - t)}.$$

Using the explicit expressions for the probabilities (A.1), (A.2), and the bound (A.5) for $1 - \gamma_{\mathbf{0}}$, yields,

$$|f(\mathbf{y}, \mathbf{m} + \mathbf{e}_{ij}) - f(\mathbf{y}, \mathbf{m})| \left| \gamma_{\mathbf{e}_{ij}}(t, \mathbf{y}) - \frac{\theta Q_j y_i}{\|\mathbf{y}\|_1^2} \right|$$

$$\le 2 \|f\|_\infty \frac{\theta Q_j y_i}{\|\mathbf{y}\|_1} \left(\gamma_{\mathbf{0}}(t, \mathbf{y}) \left| \frac{1}{t} \log\left(\frac{\|\mathbf{y}\|_1 - t}{\|\mathbf{y}\|_1}\right) + \frac{1}{\|\mathbf{y}\|_1} \right| + \frac{1}{\|\mathbf{y}\|_1} (1 - \gamma_{\mathbf{0}}(t, \mathbf{y})) \right)$$

$$\le 2\theta \|f\|_\infty \left(-\frac{1}{t} \log\left(\frac{\delta - t}{\delta}\right) - \frac{1}{\delta} + \frac{1}{\delta}\left(1 - \left(1 - \frac{t}{\delta}\right)^\theta\right) \right).$$

Therefore, each term in (A.9) is bounded by a function of t and δ, not depending on $\mathbf{y}$ which vanishes as $t \to 0$.

Similarly to (A.6), the last sum in (A.7) is bounded as follows

$$\sum_{\mathbf{w} \in \mathbb{N}^{d^2}, \|w\|_1 \ge 2} \left| \frac{f\left(\mathbf{y} - \frac{\mathbf{y}}{\|\mathbf{y}\|_1} t, \mathbf{m} + \mathbf{w}\right) - f(\mathbf{y}, \mathbf{m})}{t} \right| \gamma_{\mathbf{w}}(t, \mathbf{y}) \le 2 \|f\|_\infty \frac{1}{t} \sum_{h=2}^\infty \left(\frac{\theta t}{\delta - t}\right)^h \frac{1}{h!},$$

which also vanishes as $t \to 0$. The proof is now complete.

References

[1] A. D. Barbour, S. N. Ethier, and R. C. Griffiths. A transition function expansion for a diffusion model with selection. *Annals of Applied Probability*, 10(1):123–162, 2000. MR1765206

[2] A. Bhaskar, A. G. Clark, and Y. S. Song. Distortion of genealogical properties when the sample is very large. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6):2385–2390, 2014.

[3] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. MR1700749

[4] M. Birkner and J. Blath. Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model. *Journal of mathematical biology*, 57(3):435–465, 2008. MR2411228

[5] M. Birkner, J. Blath, and M. Steinrücken. Importance sampling for lambda-coalescents in the infinitely many sites model. *Theoretical population biology*, 79(4):155–173, 2011.

[6] M. De Iorio and R. C. Griffiths. Importance sampling on coalescent histories. I. *Advances in Applied Probability*, 36(2):417–433, 2004. MR2058143

[7] A. M. Etheridge and R. C. Griffiths. A coalescent dual process in a Moran model with genic selection. *Theoretical Population Biology*, 75:320–330, 2009.

[8] S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 1986. MR0838085

[9] M. Favero, H. Hult, and T. Koski. A dual process for the coupled wright-fisher diffusion. *Journal of Mathematical Biology*, 82(6), 2021. MR4204660

[10] M. Favero and H. Hult. Asymptotic behaviour of sampling and transition probabilities in coalescent models under selection and parent dependent mutations. *Electronic Communications in Probability*, 27:1–13, 2022. MR4440261

[11] M. Favero and P. A. Jenkins. Sampling probabilities, diffusions, ancestral graphs, and duality under strong selection. arXiv:2312.17406, 2023+.

[12] R. C. Griffiths. The two-locus ancestral graph. In I. V. Basawa and R. L. Taylor, editors, *Selected Proceedings of the Sheffield Symposium on Applied Probability*, volume 18 of *Lecture Notes–Monograph Series*, pages 100–117. Institute of Mathematical Statistics, Hayward, CA, 1991. MR1193063

[13] R. C. Griffiths, P. A. Jenkins, and Y. S. Song. Importance sampling and the two-locus model with subdivided population structure. *Advances in applied probability*, 40(2):473–500, 2008. MR2433706

[14] R. C. Griffiths and P. Marjoram. An ancestral recombination graph. In P. Donelly and S. Tavarè, editors, *Progress in Population Genetics and Human Evolution*, page 257–270. Springer-Verlag, Berlin, 1997. MR1493031

[15] R. C. Griffiths and S. Tavaré. Simulating probability distributions in the coalescent. *Theoretical Population Biology*, 46(2):131–159, 1994.

[16] R. C. Griffiths, P. A. Jenkins and S. Lessard. A coalescent dual process for a Wright–Fisher diffusion with recombination and its application to haplotype partitioning. *Theoretical Population Biology*, 112: 126-138, 2016.

[17] A. Hobolth, M. K. Uyenoyama, and C. Wiuf. Importance sampling for the infinite sites model. *Statistical applications in genetics and molecular biology*, 7(1):Article32, 2008. MR2457045

[18] J. Kelleher, A. M. Etheridge, and G. McVean. Efficient coalescent simulation and genealogical analysis for large sample sizes. *PLOS Computational Biology*, 12(5), 2016.

[19] J. F. C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3):235–248, 1982. MR0671034

[20] J. Koskela, P. A. Jenkins, and D. Spanò. Computational inference beyond Kingman's coalescent. *J. Appl. Probab.*, 52(2):519–537, 06 2015. MR3372090

[21] J. Koskela, D. Spanò, and P. A. Jenkins. Inference and rare event simulation for stopped Markov processes via reverse-time sequential Monte Carlo. *Statistics and Computing*, 28(1):131–144, Jan 2018. MR3741642

[22] S. M. Krone and C. Neuhauser. Ancestral processes with selection. *Theoretical Population Biology*, 51:210–237, 1997.

[23] T. G. Kurtz. *Lectures on Stochastic Analysis*. University of Wisconsin – Madison, 2001.

[24] M. Mandelkern. Metrization of the one-point compactification. *Proceedings of the AMS*, 107(4), 1989. MR0991703

[25] M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Annals of Probability*, 29:1547–1562, 2001. MR1880231

[26] C. Neuhauser and S. M. Krone. The genealogy of samples in models with selection. *Genetics*, 154:519–534, 1997.

[27] J. Pitman. Coalescent with multiple collisions. *Annals of Probability*, 27:1870–1902, 1999. MR1742892

[28] D. Pollard. *Convergence of Stochastic Processes*. Springer Series in Statistics. Springer-Verlag New York, 1984. MR0762984

[29] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *Journal of Applied Probability*, 36:1116–1125, 1999. MR1742154

[30] T. Shiga. Diffusion Processes in Population Genetics. *J. Math. Kyoto Univ. (JMKYAZ)*, 21(1):133–151, 1981. MR0606316

[31] J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electronic Journal of Probability*, 5:50pp, 2000. MR1781024

[32] M. Stephens and P. Donnelly. Inference in molecular population genetics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):605–635, 2000. MR1796282

[33] M. Stephens and P. Donnelly. Ancestral inference in population genetics models with selection (with discussion). *Australian & New Zealand Journal of Statistics*, 45(4):395–430, 12 2003. MR2018460

[34] J. Wakeley. Developments in coalescent theory from single loci to chromosomes. *Theoretical population biology*, 133:56–64, 2020.