WARWICK
THE UNIVERSITY OF WARWICK

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

http://wrap.warwick.ac.uk/169248

Copyright and reuse:
This thesis is made available online and is protected by original copyright.
Please scroll down to view the document itself.
Please refer to the repository record for this item for information to help you to cite it.
Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Small–Time Asymptotics of Coalescent and Fleming–Viot Processes

Philip A. Hanson

Thesis submitted for the degree of

Doctor of Philosophy of Mathematics

Supervisors: Dr. Dario Spanò, Dr. Paul A. Jenkins

Dr. Jere Koskela

Department of Mathematics and Statistics

University of Warwick

February 2022

Contents

Acknowledgements

I wish to express gratitude to my supervisors Dario Spanò, Paul Jenkins and Jere Koskela who have been a great help throughout this project. I would also like to thank the University of Warwick Mathematics Department and the EPSRC for their financial support. Finally, I wish to extend my appreciation to my partner Bree for her help and support through the writing of this thesis.

Declaration

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree at any university. Except where stated otherwise by reference or acknowledgment, the work presented is entirely my own.

Abstract

Measure-valued and coalescent processes have a long and rich history in the field of population genetics. Through duality, these processes are intimately linked and the study of one often leads to insights in the other. In this thesis we investigate the small-time behaviour of backwards-in-time genealogical processes to establish a diffusion limit at $t = 0$. Forwards in time we consider linear combinations of an i.i.d. collection of Fleming-Viot processes, describing LLN and CLT style limits. In order to establish convergence in the second chapter, we require tightness. The duality of these processes and results from our first chapter are essential in proving this and much of the second chapter.

Notation

$C(A, B)$	The set of continuous functions $f : A \to B$
$\mathcal{B}([0, 1])$	The Borel σ-algebra on $[0, 1]$
$\mathcal{P}([0, 1])$	The set of probability measures on $[0, 1]$
$\mathcal{M}([0, 1])$	The set of signed measures on $[0, 1]$
$\mathcal{D}([0, 1])$	The set of Schwartz distributions on $[0, 1]$
$\langle f, \mu \rangle$	$\int_0^1 f(x)\mu(dx)$
$n_{(k)}$	Rising factorial, $n(n+1)\ldots(n+k-1)$
$n_{[k]}$	Falling factorial, $n(n-1)\ldots(n-k+1)$
$f(n) \sim g(n)$	$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$
$f(n) = O(g(n))$	$\limsup_{n\to\infty} \frac{\lvert f(n)\rvert}{g(n)} < \infty$
$f(n) = o(g(n))$	$\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$
$X \sim \mu$	The RV X is distributed according to μ

Introduction

Nearly every characteristic or behaviour exhibited in the natural world is influenced by genetics. Understanding the interplay between our biology and environment - as well as the influence of random mutations - is the goal of the study of population genetics. In this field there are a number of models that intend to describe a number of phenomena, several of which we study in this thesis.

The models in this field can be broadly divided into two categories. The first are backwards-in-time models, like the Kingman Coalescent, that describe the structure of the ancestry of a population. The second are forward in time models, like the Fleming–Viot (FV) process, that intend to capture the effects of genetic drift, random mutations and natural selection on the prevalence of different genetic types in a population. One of the beautiful aspects of this theory is that the forward and backwards in time models are often dual to each other. We define this notion more concretely in the following section but for now this can be thought of as taking a question about models in one temporal direction and answering it by studying the models in the other. Indeed the two halves of the thesis concern models in these two temporal directions and results from the first are vital in proving results from the second.

Backwards in Time

As mentioned, the first half of this thesis concerns models that move backwards in time that represent the genealogy of a population. In this thesis we are concerned with two of these; namely the Kingman coalescent and the Ancestral Selection Graph (ASG).

Firstly, the Kingman coalescent is a Markov process taking values in the space of partitions of the natural numbers. First proposed by Kingman [1, 2, 3], the dynamics are as follows: when the process is at a partition with a finite number of blocks, any two blocks merge at rate 1. This means that when the process is at n blocks, it reduces to $n-1$ blocks at a rate of $\binom{n}{2}$. Genetically, two numbers in $\mathbb{N}$ represent ancestral lineages and these numbers being in the same block at time $t > 0$ indicate that those two lineages have a common ancestor t units of time in the past. One can show that in fact we can start this process from the finest possible partition of $\mathbb{N}$, $\{\{1\}, \{2\}, ...\}$, after which the process instantly transitions to a partition with a finite number of blocks. This phenomenon is called "coming down from infinity" and is of particular interest in the first chapter. Throughout this thesis we are mainly concerned with the associated block-counting process $(N_t^{0,0})_{t\geq 0}$, with $N_0^{0,0} = +\infty$.

To incorporate mutation into this model one allows any single block to be deleted at a rate of $\theta/2 > 0$, representing that lineage mutating to a different type. The block counting process of this model is denoted $N_t^{0,\theta}$. This extra linear push downwards does not affect how the speed at which the process comes down from infinity. In fact, as we will see later in Chapter 1, mutation does not affect the asymptotic distribution of $N_t^{0,\theta}$ as $t \to 0$ at all, as long as the mutation rate is fixed.

The other important model we consider in Chapter 1 is the Ancestral Selection Graph (ASG). First proposed by Neuhauser and Krone [4, 5], it aims to include the notion of natural selection into the coalescent framework. The way selection is modelled forwards in time is by having individuals with a selective advantage leave more offspring. Going backwards in time then, some reproductive events may or may not occur depending on the types of the individuals involved. Since we do not know the types a priori then we must consider both possibilities. This is reflected in the ASG block-counting process - denoted $N_t^{\sigma,\theta}$ - by introducing a birth rate of $\sigma/2$ per block, representing that lineage splitting into two. This provides a linear push upwards that turns the block-counting process into a birth-death process. We will characterise $N_t^{\sigma,\theta}$ close to $t = 0$ and see that it in fact has the same limiting distribution and second order asymptotics as the Kingman coalescent.

Forwards in Time

The general way that forwards-in-time models have been developed is as follows: start from a finite population, define mechanisms by which individuals in subsequent generations inherit genetic information and then, with a suitable scaling of time and space, let the number of individuals grow to infinity. In this limiting process variables like mutation and selection rates also have to be scaled appropriately to appear in the limit.

A simple starting point for this limiting procedure is the Wright–Fisher model. In this model one starts with a finite population of N individuals, each having one of a finite number of genetic types, say $A_1, \ldots, A_k$. In the subsequent generation, new individuals pick their type uniformly at random from the previous generation. Mutation can be added into this framework by changing (or not changing) the type inherited from the original according to some underlying probability distribution. If one scales time and the mutation parameters appropriately, what results in the infinite population limit is the Wright–Fisher diffusion. This process models the prevalence of $k \in \mathbb{N}$ genetic types in a population over time and lives on the $k-1$-simplex

$$\Delta_{k-1} = \left\{ (p_1, \ldots, p_k) \,\middle|\, \sum_{i=1}^{k} p_i = 1 \right\},$$

and has the following generator:

$$L = \frac{1}{2}\sum_{i,j=1}^{k} p_i(\delta_{i,j} - p_j)\frac{\partial^2}{\partial p_i \partial p_j} + \sum_{j=1}^{k}\left(\sum_{i=1}^{k} q_{i,j}p_i\right)\frac{\partial}{\partial p_j},$$

where $q_{i,j}$ is the rate at which individuals of type i mutate to type j and the domain is simply twice continuously-differentiable functions on $\mathbb{R}^k$, restricted to Δ_k. Generally, mutation is taken to be parent-independent and $q_{i,j}$ is set to $\theta_j/2 > 0$.

What one would like to do next is let the number of types k grow to infinity. However, this is non-trivial. If we simply let $k \to \infty$ without any re-ordering of types then the mass for any individual type would become zero. One way around this is to construct the limiting process as a measure-valued diffusion. This is what W. Fleming and M. Viot did in 1979 [6]. This probability measure-valued process gives, at each time t, the distribution of genetic types at a single locus in the population. The types themselves are normally contained in a compact metric space S and the process, denoted $(F_t)_{t\geq 0}$, has the following generator:

$$\begin{aligned}(\mathcal{L}\phi)(\mu) =& \frac{1}{2}\int_S\int_S \frac{\partial^2\phi(\mu)}{\partial\mu(x)\partial\mu(y)}(\delta_x(\mathrm{d}y) - \mu(\mathrm{d}y))\mu(\mathrm{d}x) \\ &+ \int_S A\left(\frac{\partial\phi(\mu)}{\partial\mu(\cdot)}\right)(x)\mu(\mathrm{d}x), \qquad (0.1)\end{aligned}$$

where A is the generator of a Feller semigroup on $C(S, \mathbb{R})$ and

$$\partial\phi(\mu)/\partial\mu(x) = \lim_{\epsilon \to 0^+} \epsilon^{-1} \left\{ \phi(\mu + \epsilon\delta_x) - \phi(\mu) \right\},$$

where δ_x is the Dirac measure at x. The domain of $\mathcal{L}$ is

$$\begin{aligned} \mathcal{D}(\mathcal{L}) = \{\phi : \phi(\mu) = F(\langle f_1, \mu \rangle, \ldots, \langle f_k, \mu \rangle), \\ F \in C^2(\mathbb{R}^k, \mathbb{R}), f_1, \ldots, f_k \in \mathcal{D}(A), k \geq 1\}. \end{aligned}$$

The type space is generally taken to be any compact metric space, though in this thesis we take it to be $[0, 1]$.

In this thesis we wish to take this generalisation further, hoping to describe a large number of loci at the same time. Our approach to achieving this is to take a FV process for each locus and consider scalings of linear combinations giving rise to limit processes described respectively by a law of large numbers (LLN) and central limit theorem (CLT). In doing this we will attain processes in the spaces $\mathcal{P}([0, 1])$ and $\mathcal{D}([0, 1])$, probability measures and Schwartz distributions on $[0, 1]$ respectively. Furthermore we also establish the continuity of the limiting Gaussian Schwartz distribution-valued process that characterises the CLT fluctuations. Finally, we drop the assumption of identical starting measures and consider a Lyapunov-style CLT, establishing tightness and convergence of finite-dimensional distributions in that setting also.

1 Diffusion Limits at Small Times for the ASG

The Kingman coalescent [1, 2, 3] is one of the fundamental models in the field of population genetics, though it has many mathematical qualities that are of independent interest. If one takes a large population of individuals under random mating with no natural selection, recombination or population structure and considers their ancestry into the past, a process that models the structure of this ancestry is the Kingman coalescent.

This partition-valued Markov process can be started from the partition of singletons $\{\{1\},\{2\},\dots\}$, in which case the associated block-counting process – $(N_t^{0,0})_{t\geq 0}$ – starts at $+\infty$ and instantly becomes finite almost surely [7]. This phenomenon is called "coming down from infinity" (CDI) and motivates investigation into the behaviour of $N_t^{0,0}$ close to $t = 0$.

When considering the CDI behaviour of the coalescent, a natural question to ask is how quickly this happens. In answering this, one looks for a function ν_t such that

$$\lim_{t\downarrow 0} \frac{N_t^{0,0}}{\nu_t} = 1, \quad a.s. \tag{1.1}$$

In [8] it is proved that for those general birth/death processes which come down from infinity, one can define ν as

$$\nu_t := \inf\{n \geq 0;\ \mathbb{E}_\infty[T_n] \leq t\}, \tag{1.2}$$

where T_n is the first hitting time of n and the subscript on $\mathbb{E}$ denotes that the process started from infinity. In [9], Griffiths derives asymptotic expressions for $N_t^{0,0}$ when t is small. There it is shown that, as $t \to 0$, $N_t^{0,0}$ is approximately Gaussian with mean $2/t$ and variance $2/(3t)$. These approximations are often used when simulating the Wright–Fisher diffusion [10]. For the coalescent then, the function $2/t$ satisfies (1.1); see also [7, Theorem 4.9]. It should be noted that any function which satisfies $\nu_t \sim 2/t$ as $t \to 0$ will also satisfy (1.1).

In [11] the authors fully characterised the fluctuations of $N_t^{0,0}$ as $t \to 0$ by proving the convergence of

$$X_\epsilon^{0,0}(t) := \epsilon^{-1/2}\left(\frac{\epsilon t}{2} N_{\epsilon t}^{0,0} - 1\right), \quad t > 0, \quad X_\epsilon^{0,0}(0) = 0,$$

to

$$Z_t := \frac{1}{\sqrt{2}t}\int_0^t u \,\mathrm{d}W_u, \quad t > 0, \quad Z_0 = 0, \tag{1.3}$$

as $\epsilon \to 0$, where W is a Brownian Motion. This convergence is in law in the Skorokhod space $D([0,\infty),\mathbb{R})$ of real-valued càdlàg functions equipped with the J_1 topology that makes it a completely separable metric space.

Two key genetic mechanisms *not* present in Kingman's original construction of the coalescent are mutation and selection. Parent-independent mutation is incorporated by having mutations occur at a constant rate $\theta/2 \geq 0$ independently on each line of ancestry. Moving forwards in time, a mutation event causes a new individual to enter the ancestry and start leaving descendants. Therefore, if we look backwards in time a mutation event on a line of ancestry causes it to end. This addition now means that the number of lineages moves from n to $n-1$ at rate

$$\frac{n(n-1+\theta)}{2}.$$

Here we see that the addition of mutation only slightly affects the rate at which lineages are lost. Asymptotically this expression is still $O(n^2)$ as $n \to \infty$; as such, one expects the same speed of coming down from infinity and indeed we prove this in Section 1.2.

The basic mechanic of selection is that one genetic type may have an evolutionary advantage over another. Forward in time we model this by having fit types reproduce at a higher rate. When looking backwards in time—without knowing the types of all individuals—there is uncertainty as to whether an offspring arose as a result of a normal reproduction event or one involving a fit type. In the ancestral process we model this by having lineages split into two independently at rate $\sigma/2 \geq 0$, tracking both possibilities. Rather than a tree, the resulting genealogical structure is a graph known as the Ancestral Selection Graph (ASG), introduced to the coalescent framework by Neuhauser and Krone in [4] and [5]. These additions result in the number of lineages — denoted $N_t^{\sigma,\theta}$ — becoming a birth/death process with the following transition rates:

$$\begin{aligned} n &\mapsto n-1 \text{ at rate } \frac{n(n-1+\theta)}{2}, \\ n &\mapsto n+1 \text{ at rate } \frac{\sigma n}{2}. \end{aligned}$$

In this chapter we show that, close to $t = 0$, the quadratic rate of coalescence dominates both the linear rate of upward jumps and the linear perturbation in the rate of downward jumps, and in fact the ASG has the same limiting behaviour as the coalescent in the following sense:

Theorem 1.1. *Let $(N_t^{\sigma,\theta})_{t\geq 0}$ be the number of lineages in the ASG at time t. Then the process*

$$X_\epsilon^{\sigma,\theta}(t) := \epsilon^{-1/2}\left(\frac{\epsilon t}{2}N_{\epsilon t}^{\sigma,\theta} - 1\right), \quad t > 0, \quad X_\epsilon^{\sigma,\theta}(0) = 0, \tag{1.4}$$

converges in law in $D([0,\infty),\mathbb{R})$ as $\epsilon \to 0$ to the Gaussian process Z given by (1.3)*.*

Note that if one sets $\sigma = 0$ then the above result implies that the diffusion limit also holds for the Kingman coalescent with mutation. Furthermore, since the number of lineages in the Ancestral Recombination Graph (ARG) has a similar birth/death structure — with a recombination rate $\rho \geq 0$ giving linear births — the above theorem also applies to that model [12].

The duality between these coalescent processes and corresponding diffusions is a well established and often used tool in the study of population genetics. Indeed in [13] the duality is exploited to express the transition function of the Wright–Fisher (WF) diffusion in terms of

$$d_n^{\theta}(t) := \mathbb{P}(N_t^{0,\theta} = n),$$

and this offers a direct way to actually simulate the diffusion [10, 14].

In the next section we outline a Poisson random measure construction that allows us to consider all of these processes together on the same probability space. Section 1.2 contains an analysis of hitting times of these processes with Section 1.3 containing the necessary lemmas that will be used to prove Theorem 1.1 in Section 1.4.

1.1 A Poisson Random Measure Construction

In order to analyse the number of lineages in the ASG and Kingman coalescent, a construction of them via a Poisson random measure (PRM) is very useful. Here we extend the construction of [11] to account for mutation and selection.

First, we define the following spaces:

$$\Delta := \{(i,j) : i,j \in \mathbb{N}, i < j\},$$

and

$$\bar{\Delta} := \Delta \cup \{(i,\infty); i \in \mathbb{N}\} \cup \{(i,0); i \in \mathbb{N}\},$$

where a standard element of $\bar{\Delta}$ will be denoted $\boldsymbol{k} = (i,j)$. On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ let π be a PRM on $\mathbb{R}_+ \times \bar{\Delta}$ with intensity measure

$$\nu := \ell \otimes \left(\sum_{(i,j)\in\Delta} \delta_{(i,j)} + \sum_{i\in\mathbb{N}} \frac{\theta}{2}\delta_{(i,\infty)} + \sum_{i\in\mathbb{N}} \frac{\sigma}{2}\delta_{(i,0)} \right),$$

where ℓ is the Lebesgue measure.

The block-counting process of the ASG can be obtained from π as follows: an arrival of type $(t,\boldsymbol{k}) = (t,(i,j))$ represents a potential coalescence, mutation or selection event. If $0 < j < \infty$ then the lineages i and j coalesce if the process has at least j lineages at time t. If $j = \infty$ then the ith lineage is lost — again if the process has at least i lineages at time t. If $j = 0$ then the ith lineage is split into two new lineages.

We also let $\hat{\pi} := \pi - \nu$ be the compensated PRM associated with π and

$$\begin{aligned}
\Delta_n &:= \{(i,j) \in \Delta : 1 \leq i < j \leq n\}, \\
\bar{\Delta}_n &:= \Delta_n \cup \{(i,\infty) \in \bar{\Delta} : 1 \leq i \leq n\} \\
&\qquad \cup \{(i,0) \in \bar{\Delta} : 1 \leq i \leq n\}.
\end{aligned}$$

If we consider starting the ASG with a finite number of lineages n, we can express the number of lineages at time t in the following way:

$$N_{t,n}^{\sigma,\theta} := n - \int_0^t \int_{\bar{\Delta}} \mathbb{1}_{\bar{\Delta}_{N_{s-,n}^{\sigma,\theta}}}(\boldsymbol{k}) \left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k})\right] \pi(\mathrm{d}s, \mathrm{d}\boldsymbol{k}). \tag{1.5}$$

We can then define the $\mathbb{P}$-almost sure limit

$$N_t^{\sigma,\theta} := \lim_{n\to\infty} N_{t,n}^{\sigma,\theta}$$

whose existence is assured by the lookdown construction of Donnelly and Kurtz [15, Section 6]; see also Section 1.4 later in this chapter, namely the proof of Lemma 1.13.

We can also obtain the Kingman coalescent with and without mutation - denoted $N^{0,\theta}$ and $N^{0,0}$ - from this PRM by thinning the different arrival types. We can do this by considering a similar construction to (1.5) and letting $n \to \infty$. However, simply doing this is not sufficient for our analysis. Our main tool in investigating the small time behaviour of the ASG is an analysis and comparison of neutral and non-neutral hitting times. In the proof of Proposition 1.11 we need the hitting times of our processes to satisfy the following:

- $T^{0,\theta}_{n,n-1} \leq T^{\sigma,\theta}_{n,n-1}$ almost surely, for each $n \in \mathbb{N}$,
- $T^{0,\theta}_{n,n-1}$ must be independent of $T^{\sigma,\theta}_{n}$ for each $n \in \mathbb{N}$.

Here $T^{\sigma,\theta}_{n,n-1}$ and $T^{0,\theta}_{n,n-1}$ are the time taken for $N^{\sigma,\theta}$ and $N^{0,\theta}$ respectively to move from n to $n-1$ and $T^{\sigma,\theta}_{n} := T^{\sigma,\theta}_{\infty,n}$ is the time taken to get from infinity to n for the ASG.

With just the thinning applied to the PRM the first inequality would only be true in distribution and the second would not necessarily be true; for example if the neutral process reaches $n-1$ before the non-neutral process has reached n, the two hitting times would rely on the same Poisson arrivals and hence not be independent.

To rectify this we construct the pair of processes $N^{\sigma,\theta}$ and $N^{0,\theta}$ from a series of separate couplings for each level n. To achieve this we first construct $N^{\sigma,\theta}$ and a new death process $\hat{N}$; this process has similar behaviour to the coalescent but will need a time-change in order to become $N^{0,\theta}$. Considering then these two processes started from $n+1$, we wait until arrivals in the PRM cause $\hat{N}$ to hit n; any arrival of type (i,j) with $i < j \leq n+1$ or $i \leq n+1, j = \infty$ will do this. It could be the case that the non-neutral process will hit n at the same time; at which point we stop and the coupling for level $n+1$ is complete. However if there is a birth before a death then $\hat{N}$ may hit n strictly before $N^{\sigma,\theta}$ does. In this case we let $\hat{T}^{-}_{n}$ be the time that this happens and hold $\hat{N}$ at level n, ignoring further Poisson arrivals until the non-neutral process has also hit n. The time that this then happens is labelled $\hat{T}_n^+$. We can do this for each level and represent $\hat{N}$ started from n blocks in a similar way to (1.5) as an integral with respect to the PRM:

$$\hat{N}_{t,n} := n - \int_0^t \int_{\bar{\Delta}} \mathbb{1}_{\bar{\Delta}_{\hat{N}_{s-,n}}}(\boldsymbol{k}) \mathbb{1}_{j>0}(\boldsymbol{k}) \mathbb{1}_A(s) \pi(\mathrm{d}s, \mathrm{d}\boldsymbol{k}),$$

where

$$A := \bigcap_{k \geq 1} \{s : \hat{T}_k^- < s \leq \hat{T}_k^+\}^c.$$

Similarly to the ASG construction we can then define the $\mathbb{P}$-almost sure limit

$$\hat{N}_t := \lim_{n \to \infty} \hat{N}_{t,n},$$

which again is assured by [15]. This coupling then produces the block counting process of the ASG and a version of the Coalescent that is occasionally held still whilst waiting for the ASG to "catch up".

In order to obtain $N^{0,\theta}$ we need to skip over these periods of time that $\hat{N}$ is held still. To do this we let

$$\phi(s) := s - \int_0^s \sum_{k \in \mathbb{N}} \mathbb{1}(\hat{T}_k^- < r \leq \hat{T}_k^+)\, dr,$$

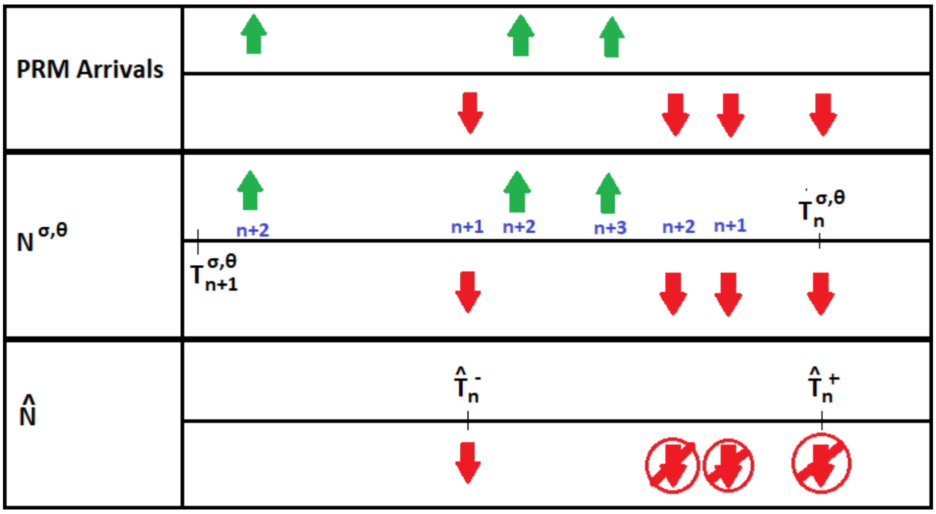
and define

$$\phi^{-1}(t) = \inf\{s \geq 0 : \phi(s) \geq t\}.$$

Running $\hat{N}$ on the timescale of ϕ^{-1} now skips over the periods of time when it is held still, producing the Coalescent; i.e. we can define

$$\left(N_t^{0,\theta}\right)_{t\geq 0}(\omega) := \left(\hat{N}_{\phi^{-1}(t)}\right)_{t\geq 0}(\omega).$$

for each $\omega \in \Omega$. The following diagram illustrates one step in the construction:



In this realisation $\hat{N}$ hits n as soon as there is a death arrival (red down arrow) in the PRM. When this occurs (at time $\hat{T}_n^-$) the next two death arrivals are ignored since the non-neutral process $(N^{\sigma,\theta})$ hasn't hit n yet. The final death arrival that causes $N^{\sigma,\theta}$ to hit n is also ignored and after this both processes are at n, completing the coupling for this level. Since the PRM arrivals that govern $T_n^{\sigma,\theta}$ are to the left of $\hat{T}_n^+$ and those governing $T_{n,n-1}^{0,\theta}$ are to the right, these two random variables are independent.

With this construction our hitting times have the two desired qualities above. The first follows simply from the fact that, for each n, any arrival that causes $N^{\sigma,\theta}$ to decrease also causes $N^{0,\theta}$ to decrease. The second follows because the behaviour of $N^{0,\theta}$ after it has hit level n does not depend on Poisson arrivals that affect $T_n^{\sigma,\theta}$. Furthermore, we also have the following inequalities

$$N_t^{0,\theta}(\omega) \le N_t^{0,0}(\omega), \quad t \ge 0, \tag{1.6}$$

$$N_t^{0,\theta}(\omega) \le N_t^{\sigma,\theta}(\omega), \quad t \ge 0, \tag{1.7}$$

for almost every $\omega \in \Omega$. These inequalities follow immediately from the fact that any arrival of type (i, j), $0 < j < \infty$ that causes $N_t^{0,\theta}$ to decrease will also cause $N_t^{0,0}$ and $N_t^{\sigma,\theta}$ to decrease.

From here on, when used with any three of these processes, $\mathbb{E}$ is the expectation operator of the probability space outlined in this section.

1.2 An Analysis of Hitting Times

In the study of general birth/death processes one can learn a lot about the behaviour of a process through analysis of its hitting times. Of course in the case of the Kingman coalescent with mutation, the time taken to get from n to $n-1$ is simply an exponential clock:

$$T_{n,n-1}^{0,\theta} \sim \text{Exp}\left(\frac{n(n-1+\theta)}{2}\right),$$

which has a finite kth moment for all $k \in \mathbb{N}$. The ASG however experiences both births and deaths and so the situation is more complex. A first-step analysis leads to the following recurrence relation for the corresponding hitting time for the ASG:

$$T^{\sigma,\theta}_{n,n-1} \stackrel{d}{=} \xi_n + \mathbb{1}_{E_n}\left(\widehat{T}^{\sigma,\theta}_{n+1,n} + \widehat{T}^{\sigma,\theta}_{n,n-1}\right), \tag{1.8}$$

where $\xi_n \sim \operatorname{Exp}\left(n(n-1+\theta+\sigma)/2\right)$ is the holding time at level n, E_n is the event that the ASG jumps up after reaching level n for the first time and $\widehat{T}^{\sigma,\theta}_{n+1,n}$ and $\widehat{T}^{\sigma,\theta}_{n,n-1}$ are independent copies of $T^{\sigma,\theta}_{n+1,n}$ and $T^{\sigma,\theta}_{n,n-1}$ respectively. This recurrence relation will be used to prove the asymptotic behaviour of the moments of $T^{\sigma,\theta}_{n,n-1}$, but first we state a lemma on conditions needed for the hitting times of a general birth/death process to have finite kth moments for all $k \in \mathbb{N}$:

Lemma 1.2. *Let $T^{\sigma,\theta}_{n,n-1}$ be the hitting time of $n-1$ from n for $N^{\sigma,\theta}$. Then for all $k \in \mathbb{N}$*

$$\lim_{n\to\infty} \mathbb{E}\left[\left(T^{\sigma,\theta}_{n,n-1}\right)^k\right] = 0. \tag{1.9}$$

Remark 1.3. *It should be noted that this lemma is primarily used as a starting point for the recursion in Proposition 1.4 where it is proved that in fact these moments are $O(n^{-2k})$ as $n \to \infty$. There the starting point for the recursion is that these moments are at least $O(1)$ as $n \to \infty$, a weaker statement than the above.*

Proof. As a reminder, the birth and death rates at for $N^{\sigma,\theta}$ at level n (conventionally denoted λ_n and μ_n) are

$$\begin{aligned}\lambda_n =& \frac{\sigma n}{2} \\ \mu_n =& \frac{n(n-1+\theta)}{2}.\end{aligned}$$

We look to verify conditions (2.9) of Lemma 2.2 (ii) in [8]. These conditions are

$$\sup_{n,i\geq 1} \frac{\mu_n}{\mu_{n+i}} < \infty, \tag{1.10}$$

$$\frac{\lambda_n}{\mu_n} \to 0 \text{ as } n \to \infty, \tag{1.11}$$

$$\sum_{n\geq 1} \frac{1}{\mu_n} < \infty. \tag{1.12}$$

(1.11) and (1.12) are immediate from the form of λ_n and μ_n. The fraction in (1.10) is maximised when $i = 1$. Since $\mu_n/\mu_{n+1} \to 1$ as $n \to \infty$, all three conditions are now satisfied.

These conditions imply both (1.1) and (1.3) in [8]. (1.1) is necessary and sufficient for the process $N^{\sigma,\theta}$ almost surely to be absorbed at zero i.e. the random variable

$$T^{\sigma,\theta}_{\infty,0} := \sum_{n\geq 1} T^{\sigma,\theta}_{n,n-1}$$

is almost surely finite. This in turn implies that the random variables $T^{\sigma,\theta}_{n,n-1}$ tend to zero almost surely and by the continuous mapping theorem so too the random variables

$\left(T^{\sigma,\theta}_{n,n-1}\right)^k$.

Combining (1.1) and (1.3) with Proposition 2.2 in [8] ensures that all moments of $T^{\sigma,\theta}_{n,n-1}$ and $T^{\sigma,\theta}_{\infty,0}$ are finite. Finally, since the kth power of $T^{\sigma,\theta}_{n,n-1}$ is clearly smaller than the corresponding power of $T^{\sigma,\theta}_{\infty,0}$ for all n, the dominated convergence theorem is sufficient to ensure (1.9). $\square$

With this established we proceed by more accurately determining the rate at which these moments tend to zero, via a comparison with the neutral case. A lower bound, $\mathbb{E}[(T^{0,\theta}_{n,n-1})^k] \leq \mathbb{E}[(T^{\sigma,\theta}_{n,n-1})^k]$, is immediate. The following proposition provides a corresponding upper bound:

Proposition 1.4. *For sufficiently large n,*

$$\mathbb{E}\left[\left(T^{\sigma,\theta}_{n,n-1}\right)^k\right] \leq \mathbb{E}\left[\left(T^{0,\theta}_{n,n-1}\right)^k\right] + \frac{B^{\sigma,\theta}_k}{n^{2k+1}}. \tag{1.13}$$

where $B^{\sigma,\theta}_k \in \mathbb{R}^+$.

Proof. Recalling the definition of ξ_n from (1.8) we note that

$$\mathbb{E}\left[\xi_n^k\right] \leq \mathbb{E}\left[\left(T^{0,\theta}_{n,n-1}\right)^k\right]. \tag{1.14}$$

This substitution will be used in the recursion (1.8) since a comparison of the ASG to the Kingman coalescent is the goal. Now, we take both sides of (1.8) to the power k and apply expectations, remembering the independence of the event E_n. Using the inequalities (1.14) and $(x+y)^j \leq 2^j(x^j + y^j)$ for

$j \in \mathbb{N}$, $x, y \geq 0$ this becomes

$$
\begin{aligned}
&\mathbb{E}\left[\left(T^{\sigma,\theta}_{n,n-1}\right)^k\right] \\
&\quad \leq \mathbb{E}\left[\left(T^{0,\theta}_{n,n-1}\right)^k\right] \\
&\qquad + \sum_{j=1}^{k} \binom{k}{j} \frac{2^j \sigma}{n-1+\theta+\sigma} \\
&\qquad\quad \times \mathbb{E}\left[\xi_n^{k-j}\left(\left(\widehat{T}^{\sigma,\theta}_{n+1,n}\right)^j + \left(\widehat{T}^{\sigma,\theta}_{n,n-1}\right)^j\right)\right].
\end{aligned}
$$

Applying the independence of the ξ_n random variable to each term in the sum, using (1.14) again and letting $a_{n,k} := \mathbb{E}[(T^{\sigma,\theta}_{n,n-1})^k]$, $x_{n,k} := \mathbb{E}[(T^{0,\theta}_{n,n-1})^k]$, we obtain the following recursion formula:

$$
a_{n,k} \leq x_{n,k} + \sum_{j=1}^{k} \binom{k}{j} \frac{2^j \sigma}{n-1+\theta+\sigma} x_{n,k-j}\left(a_{n+1,j} + a_{n,j}\right). \tag{1.15}
$$

Next we need to show that $a_{n,k} = O(n^{-2k})$ as $n \to \infty$. We start by noting that Lemma 1.2 implies that $a_{n,k} = O(1)$ as $n \to \infty$, for all $k \in \mathbb{N}$. We can now recursively use (1.15) to improve these asymptotics. To this end, suppose that $a_{n,k} = O(n^{-i})$ for some $0 \leq i \leq 2k-1$ so that there exists a constant C_i such that $a_{n,k} \leq C_i n^{-i}$ for n large enough. Furthermore assume that $a_{n,j} = O(n^{-2j})$ for all $1 \leq j < k$ and let the constants C_j be such that $a_{n,j} \leq C_j n^{-2j}$ for n large enough. Since $x_{n,j} = j! 2^j n^{-j}(n-1+\theta)^{-j}$ there exists

$B_j^{0,\theta}$ such that

$$x_{n,j} \leq \frac{B_j^{0,\theta}}{n^{2j}}, \qquad \forall j, n \in \mathbb{N}. \tag{1.16}$$

Substituting these upper bounds into (1.15) we obtain the following:

$$\begin{aligned} a_{n,k} \leq \sum_{j=0}^{k-1} \binom{k}{j} \frac{2^j \sigma B_{k-j}^{0,\theta}}{n^{2(k-j)}(n-1+\theta+\sigma)} \left(\frac{C_j}{(n+1)^{2j}} + \frac{C_j}{n^{2j}} \right) \\ + \frac{2^k \sigma}{n-1+\theta+\sigma} \left(\frac{C_i}{(n+1)^i} + \frac{C_i}{n^i} \right) \end{aligned} \tag{1.17}$$

Multiplying this inequality by n^{i+1} we get

$$\begin{aligned} &n^{i+1} a_{n,k} \\ &\quad \leq \sum_{j=0}^{k-1} \binom{k}{j} \frac{2^j \sigma B_{k-j}^{0,\theta} n^{i+1}}{n^{2(k-j)}(n-1+\theta+\sigma)} \left(\frac{C_j}{(n+1)^{2j}} + \frac{C_j}{n^{2j}} \right) \\ &\qquad + \frac{2^k \sigma n^{i+1}}{n-1+\theta+\sigma} \left(\frac{C_i}{(n+1)^i} + \frac{C_i}{n^i} \right) \end{aligned}$$

Tallying up the indices of n we see that the largest term is when $j = k$, which is $O(1)$ as $n \to \infty$. The summands with $j = 0$ to $k-1$ are all of order $O(n^{i-2k})$ where $i - 2k \leq -1$ thanks to the restriction on i. This means that in fact $a_{n,k} = O(n^{-i-1})$ and so we can recursively apply this argument until we obtain $a_{n,k} = O(n^{-2k})$. Performing this iteration another time, where we would have $i = 2k$ in (1.17) and multiply by n^{2k+1}, results in terms that diverge as $n \to \infty$ and so this iterative process terminates here.

To conclude the proof we multiply (1.15) by n^{2k}:

$$
\begin{aligned}
n^{2k}a_{n,k} \leq n^{2k}x_{n,k} & \\
&+ \sum_{j=1}^{k} \binom{k}{j} \frac{2^j\sigma}{n-1+\theta+\sigma} n^{2k}x_{n,k-j}\left(a_{n+1,j}+a_{n,j}\right)
\end{aligned}
$$

and note that the summands are all $O(n^{-1})$ as $n \to \infty$. Thus we have

$$n^{2k}a_{n,k} \leq n^{2k}x_{n,k} + O(n^{-1}),$$

yielding (1.13). $\square$

We now consider the hitting times $T_n^{\sigma,\theta} := T_{\infty,n}^{\sigma,\theta}$. Before looking at this expression for the ASG we state a brief lemma on the large n asymptotics of the hitting time $T_n^{0,\theta}$ for the Kingman coalescent with mutation:

Lemma 1.5. *For all $k \in \mathbb{N}$ there exists $C_k^{0,\theta} \in \mathbb{R}^+$ such that*

$$\mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] \leq \frac{C_k^{0,\theta}}{n^k} \qquad \forall n \in \mathbb{N}. \tag{1.18}$$

Proof. For this proof we will take advantage of the fact that the hitting time $T_n^{0,\theta}$ can be written as a sum of the hitting times $T_{i,i-1}^{0,\theta}$:

$$T_n^{0,\theta} = \sum_{i \geq n+1} T_{i,i-1}^{0,\theta}.$$

Taking the kth power of this expression we get

$$\begin{aligned}\left(T_n^{0,\theta}\right)^k &\overset{d}{=} \left(\sum_{i=n+1}^{\infty} T_{i,i-1}^{0,\theta}\right)^k \\ &\overset{d}{=} \sum_{j=1}^{k} \sum_{n+1\le i_1<\cdots<i_j} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \prod_{l=1}^{j} \left(T_{i_l,i_l-1}^{0,\theta}\right)^{m_l}.\end{aligned}$$

Since the i_l are distinct we can apply expectations to the above and separate out each of the terms in the product by independence. Doing so, along with applying (1.16), yields the following:

$$
\begin{aligned}
\mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] &= \sum_{j=1}^{k} \sum_{n+1\le i_1<\cdots<i_j} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j\\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \prod_{l=1}^{j} \mathbb{E}\left[\left(T_{i_l,i_l-1}^{0,\theta}\right)^{m_l}\right] \\
&\le \sum_{j=1}^{k} \sum_{n+1\le i_1<\cdots<i_j} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j\\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \prod_{l=1}^{j} \frac{B_{m_l}^{0,\theta}}{i_l^{2m_l}} \\
&\le \sum_{j=1}^{k} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j\\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \prod_{l=1}^{j} \sum_{i_l\ge n+1} \frac{B_{m_l}^{0,\theta}}{i_l^{2m_l}} \\
&\le \sum_{j=1}^{k} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j\\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \prod_{l=1}^{j} \int_n^\infty \frac{B_{m_l}^{0,\theta}}{x^{2m_l}}\,\mathrm{d}x \\
&= \sum_{j=1}^{k} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j\\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \prod_{l=1}^{j} \frac{B_{m_l}^{0,\theta}}{(2m_l-1)n^{2m_l-1}} \\
&= \sum_{j=1}^{k} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j\\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \frac{1}{n^{2k-j}} \prod_{l=1}^{j} \frac{B_{m_l}^{0,\theta}}{(2m_l-1)} \\
&\le \frac{C_k^{0,\theta}}{n^k},
\end{aligned}
$$

for some $C_k^{0,\theta} \in \mathbb{R}^+$,where $|\boldsymbol{m}| = m_1 + \cdots + m_j$. □

Using Proposition 1.4 and a similar approach as the previous lemma we can analyse the large n asymptotics of $T_n^{\sigma,\theta}$. Again we find that the hitting times for the coalescent and the ASG are "close" in the sense that the immediate bound $\mathbb{E}[(T_n^{0,\theta})^k] \le \mathbb{E}[(T_n^{\sigma,\theta})^k]$ is complemented by the following result:

Proposition 1.6. *For n sufficiently large, we have*

$$\begin{aligned}\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] &\leq \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] + \frac{C_k^{\sigma,\theta}}{n^{k+1}} \\ &\leq \frac{C_k^{0,\theta}}{n^k} + \frac{C_k^{\sigma,\theta}}{n^{k+1}} \end{aligned} \tag{1.19}$$

where $C_k^{\sigma,\theta} \in \mathbb{R}^+$.

Proof. As in the previous lemma, we can write the kth moment of $T_n^{\sigma,\theta}$ in the following way:

$$\begin{aligned}\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] = &\sum_{j=1}^{k} \sum_{n+1 \leq i_1 < \cdots < i_j} \sum_{\substack{\boldsymbol{m} \in \mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \\ &\times \prod_{l=1}^{j} \mathbb{E}\left[\left(T_{i_l, i_l - 1}^{\sigma,\theta}\right)^{m_l}\right]. \end{aligned} \tag{1.20}$$

Substituting the inequality from (1.13) into (1.20) we obtain the following:

$$\begin{aligned}&\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] \\ &\quad \leq \sum_{j=1}^{k} \sum_{n+1 \leq i_1 < \cdots < i_j} \sum_{\substack{\boldsymbol{m} \in \mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \\ &\qquad \times \prod_{l=1}^{j} \left(\mathbb{E}\left[\left(T_{i_l, i_l - 1}^{0,\theta}\right)^{m_l}\right] + \frac{B_{m_l}^{\sigma,\theta}}{i_l^{2m_l+1}} \right). \end{aligned}$$

From this point on we will use the shorthand $x_{i,m} = \mathbb{E}\left[\left(T_{i,i-1}^{0,\theta}\right)^m\right]$ and $y_{i,m} := B_m^{\sigma,\theta}/i^{2m+1}$. We also let $B_m^{0,\theta}$ be the constant from (1.16) such that $x_{i,m} \leq B_m^{0,\theta}/i^{2m}$ and consider n large enough so that $x_{i,m} > y_{i,m}$ for all $i \geq n$

and $1 \le m \le k$.

Thinking about expanding the brackets inside the product we see that the leading term recovers exactly $\mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right]$. Meanwhile, we can encode the remaining terms as follows:

$$\begin{aligned}\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] \le \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] \\ + \sum_{j=1}^{k} \sum_{n+1 \le i_1 < \cdots < i_j} \sum_{\substack{\boldsymbol{m} \in \mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} \\ \sum_{\substack{\boldsymbol{\alpha} \in \{0,1\}^j \\ |\boldsymbol{\alpha}| \neq j}} \prod_{l=1}^{j} x_{i_l,m_l}^{\alpha_l} y_{i_l,m_l}^{1-\alpha_l}.\end{aligned} \tag{1.21}$$

Since we are considering n large enough so that, for $i_l \ge n$, $x_{i_l,m_l} > y_{i_l,m_l}$, the largest term in the sum is the one of the form $x_{i_1,m_1} \dots x_{i_{l-1},m_{l-1}} y_{i_l,m_l} x_{i_{l+1},m_{l+1}} \dots x_{i_j,m_j}$ for some $l = 1, \dots, j$. We do not know which one of these terms will be largest so we use the following upper bound:

$$\begin{aligned}\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] \le \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] \\ + \sum_{j=1}^{k} \sum_{n+1 \le i_1 < \cdots < i_j} \sum_{\substack{\boldsymbol{m} \in \mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} (2^j - 1) \\ \sum_{l=1}^{j} y_{i_l,m_l} \prod_{p \neq l} x_{i_p,m_p}.\end{aligned}$$

Note that the factor of $2^j - 1$ appears as this is the total number of terms in the sum over $\boldsymbol{\alpha}$ in (1.21). We can now bound this above by separating out the sums over the different indices, dropping the restriction of $i_1 < \ldots < i_j$, and using the bound $x_{i_l,m_l} \leq B^{0,\theta}_{m_l}/i_l^{2m_l}$. This gives us

$$\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] \leq \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] + \sum_{j=1}^{k} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} (2^j - 1) \sum_{l=1}^{j} \sum_{i_l \geq n+1} \frac{B^{\sigma,\theta}_{m_l}}{i_l^{2m_l+1}} \prod_{p\neq l} \sum_{i_p \geq n+1} \frac{B^{0,\theta}_{m_p}}{i_p^{2m_p}}.$$

Using integral bounds on each of these sums we obtain the following:

$$\begin{aligned}
\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] &\leq \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] + \sum_{j=1}^{k} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} (2^j - 1) \sum_{l=1}^{j} \int_n^\infty \frac{B^{\sigma,\theta}_{m_l}}{x^{2m_l+1}}\,\mathrm{d}x \prod_{p\neq l} \int_n^\infty \frac{B^{0,\theta}_{m_p}}{x^{2m_p}}\,\mathrm{d}x \\
&= \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] + \sum_{j=1}^{k} \sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j \\ |\boldsymbol{m}|=k}} \binom{k}{\boldsymbol{m}} (2^j - 1) \sum_{l=1}^{j} \frac{B^{\sigma,\theta}_{m_l}}{2m_l n^{2m_l}} \prod_{p\neq l} \frac{B^{0,\theta}_{m_p}}{(2m_p - 1)n^{2m_p-1}}.
\end{aligned}$$

Considering that all sums are now finite and the m_i sum to

k, the right hand side becomes

$$\begin{aligned}\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] \leq \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right]& \\ &+\sum_{j=1}^{k}\sum_{\substack{\boldsymbol{m}\in\mathbb{N}^j\\ |\boldsymbol{m}|=k}}\binom{k}{\boldsymbol{m}}(2^j-1) \\ &\sum_{l=1}^{j}\left(\frac{B_{m_j}^{\sigma,\theta}}{2m_j}\prod_{i\neq l}\frac{B_{m_i}^{0,\theta}}{(2m_i-1)}\right)\frac{1}{n^{2k-j+1}}.\end{aligned}$$

Since all the sums above are finite then there exist $D_j^{\sigma,\theta}\in\mathbb{R}^+$, $j=1,\ldots,k$ such that

$$\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^k\right] \leq \mathbb{E}\left[\left(T_n^{0,\theta}\right)^k\right] + \sum_{j=1}^{k}\frac{D_j^{\sigma,\theta}}{n^{2k-j+1}},$$

and so we have (1.19). □

1.3 Controlling the ASG at Small Times

In this section we discuss the speed of coming down from infinity for the ASG and establish the mode of its convergence to that speed.

Before talking about the ASG we formally establish that the Kingman coalescent with mutation has the same speed of coming down from infinity as that without mutation. It should be noted that we consider this an unsurprising result but have not seen a formal proof, and so we present one here:

Proposition 1.7.

$$\lim_{t\to 0}\frac{tN_t^{0,\theta}}{2} = 1 \quad a.s. \tag{1.22}$$

Proof. In order to ensure that there exists a function ν_t such that $N_t^{0,\theta}/\nu_t$ converges to 1 almost surely, we turn to [8, Theorem 5.1] where this limit is considered for general birth/death processes with birth and death rates λ_n and μ_n respectively. There, two conditions are sufficient to ensure the existence of ν:

- $\lim_{n\to\infty}\lambda_n/\mu_n = 0$
- $(\mu_n)_{n\geq 1}$ varies regularly with index $\rho > 1$.

In our case $\lambda_n = 0$ so there is only the second condition to check; that the death rate μ_n varies regularly with an index $\rho > 1$ as $n \to \infty$. Recall that a sequence of real nonzero numbers $(a_n)_{n\geq 1}$ varies regularly with index $\rho \neq 0$ if, for all $b > 0$,

$$\lim_{n\to\infty}\frac{a_{[bn]}}{a_n} = b^\rho.$$

Since our death rates are a quadratic polynomial they satisfy the above condition with $\rho = 2$. This gives us our a.s. convergence.

We proceed to verify the form of ν by recalling (1.2) and considering the expected hitting time of n by $N^{0,\theta}$:

$$\mathbb{E}\left[T_n^{0,\theta}\right] = \sum_{k=n+1}^{\infty}\frac{2}{k(k-1+\theta)}.$$

We can then sandwich this moment between the following sums:

$$\sum_{k>n+\theta+1} \frac{2}{k(k-1)} \leq \mathbb{E}\left[T_n^{0,\theta}\right] \leq \sum_{k=n+1}^{\infty} \frac{2}{k(k-1)}$$

which are both $2/n + O(n^{-2})$ as $n \to \infty$ and so $\mathbb{E}[T_n^{0,\theta}]$ has these asymptotics as well. To see that this implies that ν_t is asymptotically equivalent to $2/t$ as $t \to 0$ we use the definition of ν_t to get the following set of inequalities:

$$\begin{aligned} \mathbb{E}\left[T_{\nu_t}^{0,\theta}\right] \leq t < \mathbb{E}\left[T_{\nu_t-1}^{0,\theta}\right] \\ \frac{2}{\nu_t} + O((\nu_t)^{-2}) \leq t < \frac{2}{\nu_t - 1} + O((\nu_t)^{-2}) \\ \frac{2}{t} + O((\nu_t t)^{-1}) \leq \nu_t < \frac{2}{t}\left(\frac{\nu_t}{\nu_t - 1}\right) + O((\nu_t t)^{-1}). \end{aligned} \tag{1.23}$$

Now, when we divide the error term on either side by $2/t$ we get a function that grows at most like $(\nu_t)^{-1}$ and so tends to zero as $t \to 0$. This tells us that ν_t is asymptotically equivalent to $2/t$ as $t \to 0$. $\square$

Thanks to the asymptotics established in the previous section, the equivalent result for the ASG is now simple to prove:

Proposition 1.8.

$$\lim_{t \to 0} \frac{t N_t^{\sigma,\theta}}{2} = 1 \quad a.s. \tag{1.24}$$

Proof. Again we appeal to [8, Theorem 5.1] in order to verify the above limit. For the ASG, the death rates are the same as Kingman's coalescent with mutation so μ_n still varies regularly with index 2. Furthermore, since our birth rates are linear in n we still have that $\lambda_n/\mu_n \to 0$ as $n \to \infty$. Thus [8, Theorem 5.1] gives us our almost sure convergence.

Looking at the equivalent of equation (1.23) for the ASG we can use (1.19) to state the following series of inequalities for sufficiently small t:

$$\begin{aligned}
\mathbb{E}\left[T^{0,\theta}_{\nu_t}\right] \leq \mathbb{E}\left[T^{\sigma,\theta}_{\nu_t}\right] \leq t < \mathbb{E}\left[T^{\sigma,\theta}_{\nu_t-1}\right] \leq \mathbb{E}\left[T^{0,\theta}_{\nu_t-1}\right] + \frac{C_1^{\sigma,\theta}}{(\nu_t-1)^2} \\
\frac{2}{\nu_t} + O((\nu_t)^{-2}) \leq t < \frac{2}{\nu_t - 1} + O((\nu_t)^{-2}) \\
\frac{2}{t} + O((\nu_t t)^{-1}) \leq \nu_t < \frac{2}{t}\left(\frac{\nu_t}{\nu_t - 1}\right) + O((\nu_t t)^{-1}).
\end{aligned}$$

In the same way as in the previous proof we have that $\nu_t \sim 2/t$ as $t \to 0$ and so (1.24) is verified. $\square$

In order to establish Theorem 1.1 the above almost sure convergence is not enough. In fact, in order to verify the convergence of (1.4) to (1.3) we need more control over $N_t^{\sigma,\theta}$ near zero. Namely we need a result analogous to [16, Theorem 2] for the ASG. We first establish separately the result under neutrality as it will be used in the proof of the equivalent statement when selection is present:

Proposition 1.9.

$$\lim_{t\to 0} \mathbb{E}\left[\sup_{s\leq t}\left(\frac{sN_s^{0,\theta}}{2} - 1\right)^k\right] = 0, \qquad \forall k \in \mathbb{N}. \tag{1.25}$$

Proof. Using [16, Theorem 2] with $d = k$ on $N_s^{0,0}$, there exists a time $t_1 > 0$ such that

$$\mathbb{E}\left[\sup_{s\le t_1}\left(\frac{sN_s^{0,0}}{2} - 1\right)^k\right] \le 1.$$

This combined with the fact that $(sN_s^{0,0}/2)^k$ can grow at most like s^k for $s > t_1$ ensures that we have

$$\mathbb{E}\left[\sup_{s\le t}\left(\frac{s}{2}N_s^{0,0}\right)^k\right] < \infty,$$

for all $t \ge 0$, $k \in \mathbb{N}$. This now immediately gives us

$$\mathbb{E}\left[\sup_{s\le t}\left(\frac{s}{2}N_s^{0,\theta}\right)^k\right] < \infty, \tag{1.26}$$

for all $k \in \mathbb{N}$ thanks to (1.6). In order to establish (1.25) we first show the almost sure convergence of

$$\sup_{s\le t}\left(\frac{sN_s^{0,\theta}}{2} - 1\right)^k \tag{1.27}$$

to zero as $t \to 0$. This is easily done with an application of the continuous mapping theorem and (1.22); in fact since we are considering a right limit we rely only on the right continuity of the function and the supremum from the right. With this, we then bound (1.27) above by $\sup_{s\le t}(sN_s^{0,\theta}/2)^k \vee 1$ and use the dominated convergence theorem together with (1.26) to obtain (1.25). $\square$

In order to obtain the same result for the ASG we need to take a closer look at the behaviour of the process $(sN_s^{\sigma,\theta}/2)^k$.

We consider what happens to this process between successive hitting times $T_n^{\sigma,\theta}$ and $T_{n-1}^{\sigma,\theta}$. Between these times the process will sit on one of the curves in Figure 1.

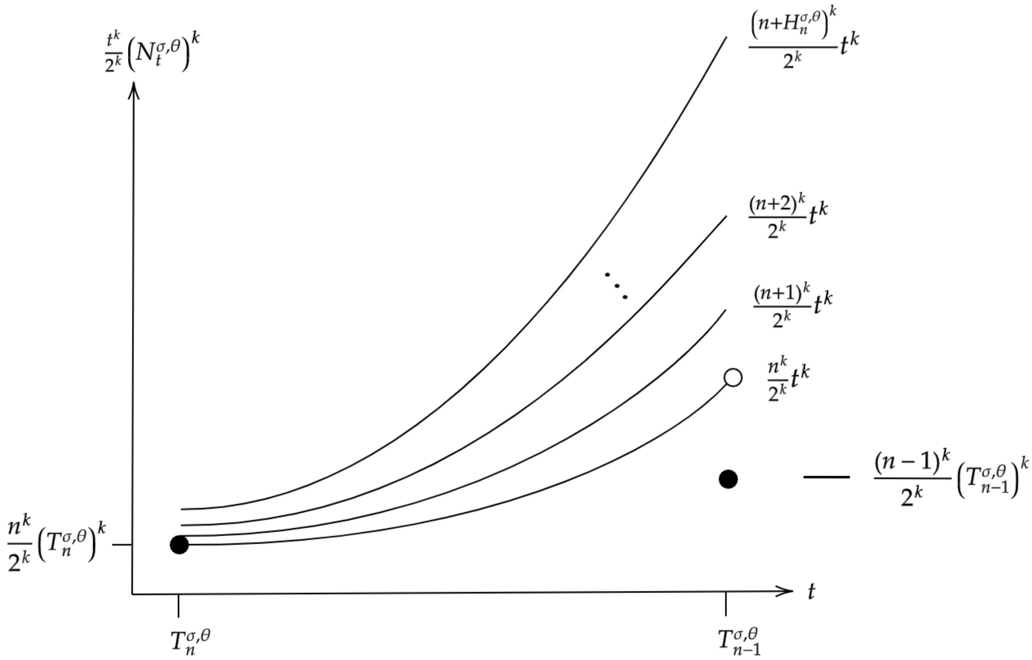


Figure 1: Possible trajectories of the function $(sN_s^{\sigma,\theta}/2)^k$ between the first hitting times of level n and $n-1$.

Here the random variable $H_n^{\sigma,\theta}$ is defined as

$$H_n^{\sigma,\theta} := \#\{s \in [T_n^{\sigma,\theta}, T_{n-1}^{\sigma,\theta}) : N_s^{\sigma,\theta} - N_{s-}^{\sigma,\theta} > 0\}$$

or simply the number of times the ASG jumps up in this time window. From the diagram we see that the highest point that the process can reach in this time window is

$$\frac{\left(n+H_n^{\sigma,\theta}\right)^k}{2^k}\left(T_{n-1}^{\sigma,\theta}\right)^k=\sum_{j=0}^{k}\binom{k}{j}\frac{n^{k-j}}{2^k}\left(H_n^{\sigma,\theta}\right)^j\left(T_{n-1}^{\sigma,\theta}\right)^k. \tag{1.28}$$

We can now proceed by looking at the supremum of the above random variables over all n since

$$\sup_{s\le t}\left(\frac{sN_s^{\sigma,\theta}}{2}\right)^k\le\sup_{n\in\mathbb{N}}\frac{\left(n+H_n^{\sigma,\theta}\right)^k}{2^k}\left(T_{n-1}^{\sigma,\theta}\right)^k, \tag{1.29}$$

for any $t\ge 0$. Before stating our result however, we need to know more about the moments of $H_n^{\sigma,\theta}$. In the proof of [8, Lemma 4.1] it is shown that, for a general birth/death process N_t with birth/death rates λ_n and μ_n, if

$$\limsup_{n\to\infty}\frac{\lambda_n}{\mu_n}<1 \tag{1.30}$$

holds then the associated H_n random variable

$$H_n:=\#\{s\in[T_n,T_{n-1}):N_s-N_{s-}>0\}$$

satisfies

$$\mathbb{E}_\infty\left[H_n^2\right]\le C\frac{\lambda_n}{\mu_n},$$

for some constant $C\ge 0$, for all $n\in\mathbb{N}$, where T_n are the first hitting times of level n for the process N_t. Since we need to consider higher moments of $H_n^{\sigma,\theta}$, this inequality is not enough. As such, we extend the above inequality to all moments of H_n for general birth/death processes:

Lemma 1.10. *Let* $(N_t)_{t\ge0}$ *be a birth/death process with* *birth/death rates* λ_n *and* μ_n *respectively. If* (1.30) *holds then for each* $k \in \mathbb{N}$ *there exists* $c_k > 0$ *such that*

$$\mathbb{E}_\infty \left[H_n^k\right] \leq c_k \frac{\lambda_n}{\mu_n}, \quad n \geq 1. \tag{1.31}$$

Proof. We follow a similar method of proof as in [8, Lemma 4.1] by first noting the random variable H_n is simply the number of positive jumps before T_{n-1} of a discrete-time random walk started at n with transition probabilities $p_{i,i+1} = \lambda_i/(\lambda_i + \mu_i)$ and $p_{i,i-1} = \mu_i/(\lambda_i + \mu_i)$.

Next, we note that (1.30) gives us the existence of $n_0 \in \mathbb{N}$ such that

$$p = \sup_{n \geq n_0} \frac{\lambda_n}{\lambda_n + \mu_n} < \frac{1}{2}.$$

Thus H_n, for $n \geq n_0$, is stochastically dominated by T; the hitting time of $n-1$ by a discrete-time simple random walk started at n with up/down transition probabilities p, $1-p$ respectively. Since $p < 1/2$, all moments of T—and thus H_n for $n \geq n_0$—are finite. Finiteness of moments of H_n for $n < n_0$ can be obtained through the following recursive relationship:

$$H_n \stackrel{d}{=} \mathbb{1}_{E_n}(1 + \hat{H}_{n+1} + \hat{H}_n),$$

where E_n is the event that the first jump after T_n was an upward one. We now consider the Laplace transform of H_n, denoted $\hat{G}_n(a) := \mathbb{E}_\infty \left[\exp(-aH_n)\right]$, and the recursion formula (4.3) from [8]:

$$\hat{G}_n(a) = \frac{\mu_n}{\lambda_n + \mu_n} + \frac{\lambda_n}{\lambda_n + \mu_n} e^{-a} \hat{G}_n(a) \hat{G}_{n+1}(a),$$

with $a \geq 0$, $n \geq 1$. Differentiating k times with respect to t gives us

$$\hat{G}_n^{(k)}(a) = \frac{\lambda_n}{\lambda_n + \mu_n} \sum_{\substack{|\boldsymbol{m}|=k \\ \boldsymbol{m} \in \mathbb{N}_0^3}} \binom{k}{\boldsymbol{m}} (-1)^{m_1} e^{-a} \hat{G}_n^{(m_2)}(a) \hat{G}_{n+1}^{(m_3)}(a),$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. If we now evaluate the above at $a = 0$ we get

$$\mathbb{E}_\infty \left[H_n^k \right] = \frac{\lambda_n}{\lambda_n + \mu_n} \sum_{\substack{|\boldsymbol{m}|=k \\ \boldsymbol{m} \in \mathbb{N}_0^3}} \binom{k}{\boldsymbol{m}} \mathbb{E}_\infty \left[H_n^{m_2} \right] \mathbb{E}_\infty \left[H_{n+1}^{m_3} \right],$$

since a factor of $(-1)^k$ can be cancelled from both sides. Rearranging this we see that

$$\frac{\mu_n}{\lambda_n} \mathbb{E}_\infty \left[H_n^k \right] = \sum_{\substack{|\boldsymbol{m}|=k \\ \boldsymbol{m} \neq (0,k,0)}} \binom{k}{\boldsymbol{m}} \mathbb{E}_\infty \left[H_n^{m_2} \right] \mathbb{E}_\infty \left[H_{n+1}^{m_3} \right].$$

Now, the right hand side of this equation is a finite combination of moments; each of which are bounded by the respective moment of T. Thus the right hand side can be bounded by some constant $c_k \geq 0$. This then gives us (1.31). $\square$

With this we now know enough about the asymptotic behaviour of the random variables in (1.28) to state the following:

Proposition 1.11.

$$\lim_{t\to 0}\mathbb{E}\left[\sup_{s\le t}\left(\frac{sN_s^{\sigma,\theta}}{2}-1\right)^k\right]=0, \qquad \forall k\in\mathbb{N}. \tag{1.32}$$

Proof. The first thing to note is that as soon as the right hand side of (1.29) is bounded in expectation the result follows thanks to an application of (1.24) and the dominated convergence theorem in a similar way to the proof of Proposition 1.9.

We start by looking at the terms on the right hand side of (1.28) for $j \neq 0$. The aim here is to show that the expectations of these random variables are summable in n and hence their supremum over n has finite mean. Applying (1.19), (1.31), and the Cauchy-Schwarz inequality to these terms we obtain the following:

$$\begin{aligned}
&\mathbb{E}\left[\binom{k}{j}\frac{n^{k-j}}{2^k}\left(H_n^{\sigma,\theta}\right)^j\left(T_{n-1}^{\sigma,\theta}\right)^k\right]\\
&\le \binom{k}{j}\frac{n^{k-j}}{2^k}\mathbb{E}\left[\left(H_n^{\sigma,\theta}\right)^{2j}\right]^{1/2}\mathbb{E}\left[\left(T_{n-1}^{\sigma,\theta}\right)^{2k}\right]^{1/2}\\
&\le \binom{k}{j}\frac{n^{k-j}}{2^k}\frac{c_{2j}^{1/2}}{n^{1/2}}\left(\frac{C_{2k}^{0,\theta}}{n^{2k}}+\frac{C_{2k}^{\sigma,\theta}}{n^{2k+1}}\right)^{1/2}\\
&= O\left(n^{-j-1/2}\right).
\end{aligned}$$

This then gives us that the expectation of the terms on the right hand side of (1.28) are summable in n as long as

$j \neq 0$. Thus we have that their supremum over n is finite in expectation.

Next we need to deal with the first term on the right hand side of (1.28); i.e. the term corresponding to $j = 0$. Since

$$\sup_{n\in\mathbb{N}} \frac{n^k}{2^k}\left(T_{n-1}^{\sigma,\theta}\right)^k \leq \sup_{n\in\mathbb{N}}\left[\frac{n^k}{2^k}\left(T_{n-1}^{\sigma,\theta}\right)^k - \frac{n^k}{2^k}\left(T_{n-1}^{0,\theta}\right)^k\right] + \sup_{n\in\mathbb{N}} \frac{n^k}{2^k}\left(T_{n-1}^{0,\theta}\right)^k \tag{1.33}$$

it suffices to show that the right hand side of the above is finite in expectation. For the second term this follows immediately from (1.26) since $n^k(T_{n-1}^{0,\theta})^k/2^k$ are the right end points of the continuous pieces of the process $t^k(N_t^{0,\theta})^k/2^k$. For the first term we consider the increments in n of these random variables and show they are absolutely summable. This gives the desired result since, for any real sequence x_n,

$$\sup_{n\in\mathbb{N}} x_n \leq x_1 + \sum_{n\geq 1} |x_{n+1} - x_n|. \tag{1.34}$$

For us, the corresponding increment is

$$\begin{aligned}
&\left| \frac{(n+1)^k}{2^k} \left(T_n^{\sigma,\theta}\right)^k - \frac{(n+1)^k}{2^k} \left(T_n^{0,\theta}\right)^k \right. \\
&\qquad \left. - \frac{n^k}{2^k} \left(T_{n-1}^{\sigma,\theta}\right)^k + \frac{n^k}{2^k} \left(T_{n-1}^{0,\theta}\right)^k \right| \\
&= \left| \left(\frac{(n+1)^k}{2^k} \left(T_n^{\sigma,\theta}\right)^k - \frac{n^k}{2^k} \left(T_{n-1}^{\sigma,\theta}\right)^k \right) \right. \\
&\qquad \left. - \left(\frac{(n+1)^k}{2^k} \left(T_n^{0,\theta}\right)^k - \frac{n^k}{2^k} \left(T_{n-1}^{0,\theta}\right)^k \right) \right| . \qquad (1.35)
\end{aligned}$$

Now, the terms inside the brackets are just the difference between successive right endpoints of continuous pieces of the processes $\left(sN_s^{\sigma,\theta}/2\right)^k$ and $\left(sN_s^{0,\theta}/2\right)^k$ respectively. These can be expressed in the following way:

$$\begin{aligned}
&\frac{(n+1)^k}{2^k} \left(T_n^{\sigma,\theta}\right)^k - \frac{n^k}{2^k} \left(T_{n-1}^{\sigma,\theta}\right)^k \\
&\quad = \left[(n+1)^k - n^k\right] \frac{\left(T_n^{\sigma,\theta}\right)^k}{2^k} - \frac{n^k}{2^k} \left(\left(T_{n-1}^{\sigma,\theta}\right)^k - \left(T_n^{\sigma,\theta}\right)^k \right) \\
&\quad = \frac{\left(T_n^{\sigma,\theta}\right)^k}{2^k} \left(\sum_{j=1}^{k} \binom{k}{j} \frac{n^{k-j}}{2^k} \right) \\
&\qquad\quad - \frac{n^k}{2^k} \left(\sum_{j=1}^{k} \binom{k}{j} \left(T_n^{\sigma,\theta}\right)^{k-j} \left(T_{n,n-1}^{\sigma,\theta}\right)^j \right) ,
\end{aligned} \qquad (1.36)$$

since $(T_{n-1}^{\sigma,\theta})^k = (T_n^{\sigma,\theta} + T_{n,n-1}^{\sigma,\theta})^k$. Note that one can set $\sigma = 0$ in the above meaning these equalities apply to both the Kingman coalescent and the ASG. Using the substitution

(1.36), (1.35) is then equal to

$$\left|\sum_{j=1}^{k}\binom{k}{j}\frac{n^{k-j}}{2^k}\left[\left(T_n^{\sigma,\theta}\right)^k-\left(T_n^{0,\theta}\right)^k\right]\right.$$
$$-\frac{n^k}{2^k}\sum_{j=1}^{k}\binom{k}{j}\left(\left(T_n^{\sigma,\theta}\right)^{k-j}\left(T_{n,n-1}^{\sigma,\theta}\right)^j\right.$$
$$\left.\left.-\left(T_n^{0,\theta}\right)^{k-j}\left(T_{n,n-1}^{0,\theta}\right)^j\right)\right|,$$

which we bound above by

$$\leq\sum_{j=1}^{k}\binom{k}{j}\frac{n^{k-j}}{2^k}\left|\left(T_n^{\sigma,\theta}\right)^k-\left(T_n^{0,\theta}\right)^k\right| \tag{1.37}$$
$$+\frac{n^k}{2^k}\sum_{j=1}^{k}\binom{k}{j}\left|\left(T_n^{\sigma,\theta}\right)^{k-j}\left(T_{n,n-1}^{\sigma,\theta}\right)^j\right.$$
$$\left.-\left(T_n^{0,\theta}\right)^{k-j}\left(T_{n,n-1}^{0,\theta}\right)^j\right|.$$

If we now apply expectations to (1.37), then the expectation of (1.35) is bounded above by

$$
\begin{aligned}
&\sum_{j=1}^{k}\binom{k}{j}\frac{n^{k-j}}{2^k}\mathbb{E}\left[\left|\left(T_n^{\sigma,\theta}\right)^k-\left(T_n^{0,\theta}\right)^k\right|\right] \\
&\qquad+\frac{n^k}{2^k}\sum_{j=1}^{k}\binom{k}{j}\mathbb{E}\left[\left|\left(T_n^{\sigma,\theta}\right)^{k-j}\left(T_{n,n-1}^{\sigma,\theta}\right)^j\right.\right. \\
&\qquad\qquad\qquad\qquad\qquad\left.\left.-\left(T_n^{0,\theta}\right)^{k-j}\left(T_{n,n-1}^{0,\theta}\right)^j\right|\right] \\
&\quad\leq\sum_{j=1}^{k}\binom{k}{j}\frac{n^{k-j}}{2^k}\mathbb{E}\left[\left|\left(T_n^{\sigma,\theta}\right)^k-\left(T_n^{0,\theta}\right)^k\right|\right] \\
&\qquad+\frac{n^k}{2^k}\sum_{j=1}^{k}\binom{k}{j}\mathbb{E}\left[\left(T_n^{\sigma,\theta}\right)^{k-j}\left|\left(T_{n,n-1}^{\sigma,\theta}\right)^j-\left(T_{n,n-1}^{0,\theta}\right)^j\right|\right] \\
&\qquad+\frac{n^k}{2^k}\sum_{j=1}^{k}\binom{k}{j}\mathbb{E}\left[\left(T_{n,n-1}^{0,\theta}\right)^j\left|\left(T_n^{\sigma,\theta}\right)^{k-j}-\left(T_n^{0,\theta}\right)^{k-j}\right|\right].
\end{aligned}
$$

Recalling equations (1.6) and (1.7) from Section 1.1, we see that on the joint probability space we have constructed the hitting times for the ASG are almost surely longer than the corresponding hitting times for the Kingman coalescent with mutation. This allows us to drop modulus signs from inside the above expectations. Utilising the independence of $T_{n,n-1}^{0,\theta}$ and $T_n^{\sigma,\theta}$ established in Section 1.1 along with (1.13), (1.18), and (1.19) results in the following set of inequalities:

$$\begin{aligned}
&\mathbb{E}\left[\left|\frac{(n+1)^k}{2^k}\left(T_n^{\sigma,\theta}\right)^k - \frac{(n+1)^k}{2^k}\left(T_n^{0,\theta}\right)^k\right.\right.\\
&\qquad\qquad\qquad\left.\left. - \frac{n^k}{2^k}\left(T_{n-1}^{\sigma,\theta}\right)^k + \frac{n^k}{2^k}\left(T_{n-1}^{0,\theta}\right)^k\right|\right]\\
&\le \sum_{j=1}^{k}\binom{k}{j}\frac{n^{k-j}}{2^k}\frac{C_k^{\sigma,\theta}}{n^{k+1}}\\
&\quad + \frac{n^k}{2^k}\sum_{j=1}^{k}\binom{k}{j}\left[\left(\frac{C_{k-j}^{0,\theta}}{n^{k-j}} + \frac{C_{k-j}^{\sigma,\theta}}{n^{k-j+1}}\right)\frac{B_j^{\sigma,\theta}}{n^{2j+1}}\right.\\
&\qquad\qquad\qquad\qquad\left. + \frac{j!2^j}{n^j(n-1+\theta)^j}\frac{C_j^{\sigma,\theta}}{n^{k-j+1}}\right]\\
&= \frac{1}{2^k}\sum_{j=1}^{k}\binom{k}{j}\frac{C_k^{\sigma,\theta}}{n^{j+1}}\\
&\quad + \frac{1}{2^k}\sum_{j=1}^{k}\binom{k}{j}\left[\left(C_{k-j}^{0,\theta} + \frac{C_{k-j}^{\sigma,\theta}}{n}\right)\frac{B_j^{\sigma,\theta}}{n^{j+1}}\right.\\
&\qquad\qquad\qquad\qquad\left. + \frac{j!2^jC_j^{\sigma,\theta}}{n(n-1+\theta)^j}\right]\\
&= O(n^{-2}).
\end{aligned}$$

From this we can conclude that the expectation of (1.35) is summable in n. This then means that, thanks to (1.34), the left hand side of (1.33) is finite in expectation; when combined with (1.29) this gives (1.32). $\square$

1.4 Proof of Theorem 1.1

In order to prove Theorem 1.1 we extend the arguments of Limic and Talarczyk [11] with modifications where needed. Lemmas 1.13 and 1.15 are analogues of Lemmas 2.2 and 2.3 from [11] for the ASG. Though the methods of proof are similar, they rely heavily on our PRM construction of the ASG and the results from Section 1.3; most importantly Proposition 1.11. Regardless of the similarities we present the results here in full in order to keep the proof self-contained.

Before establishing the key lemmas needed to prove Theorem 1.1 we state a technical lemma from [16] that is used repeatedly in what follows.

Lemma 1.12 ([16, Lemma 10])**.** *Suppose $f, g : [a,b] \to \mathbb{R}$ are càdlàg functions such that*

$$\sup_{x\in[a,b]} \left| f(x) + \int_a^x g(u)\,\mathrm{d}u \right| \leq K,$$

for some $K < \infty$. If, in addition $f(x)g(x) > 0$, $x \in [a,b]$, whenever $f(x) \neq 0$, then

$$\sup_{x\in[a,b]} \left| \int_a^x g(u)\,\mathrm{d}u \right| \leq K \text{ and } \sup_{x\in[a,b]} |f(x)| \leq 2K. \tag{1.38}$$

The first lemma we establish gives a formula for $tN_t^{\sigma,\theta}/2$ which helps us in analysing its small time behaviour.

Lemma 1.13. *Under the assumptions of Theorem 1.1 we* *have*

$$\frac{tN_t^{\sigma,\theta}}{2} = 1 - \int_0^t \left(\frac{sN_s^{\sigma,\theta}}{2} - 1\right)\frac{1}{s}\,\mathrm{d}s - M_t^{\sigma,\theta} + R_t^{\sigma,\theta}, \quad t \geq 0, \tag{1.39}$$

where

$$M_t^{\sigma,\theta} = \frac{1}{2}\int_0^t \int_{\bar{\Delta}} s\mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k})\left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k})\right]\hat{\pi}(\mathrm{d}s, \mathrm{d}\boldsymbol{k}), \tag{1.40}$$

$t \geq 0$*, and* $R^{\sigma,\theta}$ *is a continuous process such that for any* $T > 0$ *there exists* $C_1 > 0$ *such that*

$$\mathbb{E}\left[\sup_{s\leq t}|R_s^{\sigma,\theta}|\right] \leq C_1 t, \quad t \leq T. \tag{1.41}$$

Remark 1.14. *It should be noted that* C_1 *and several of the constants in this section have an implicit dependence on* T*. Fortunately, we are only concerned with behaviour close to* $t = 0$ *and so this dependence is unimportant.*

Proof. Letting $0 < r \leq t$, the PRM construction in Section

1.1 allows us to write $N_t^{\sigma,\theta}$ as:

$$\begin{aligned}
N_t^{\sigma,\theta}& \\
=&N_r^{\sigma,\theta} - \int_{(r,t]}\int_{\bar{\Delta}} \mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k})\left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k})\right]\pi(\mathrm{d}s,\mathrm{d}\boldsymbol{k}),\\
=&N_r^{\sigma,\theta} - \int_{(r,t]}\int_{\bar{\Delta}} \mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k})\left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k})\right]\nu(\mathrm{d}s,\mathrm{d}\boldsymbol{k})\\
&-\int_{(r,t]}\int_{\bar{\Delta}} \mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k})\left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k})\right]\hat{\pi}(\mathrm{d}s,\mathrm{d}\boldsymbol{k}),\\
=&N_r^{\sigma,\theta} - \int_{(r,t]} \frac{N_s^{\sigma,\theta}(N_s^{\sigma,\theta}-1+\theta-\sigma)}{2}\,\mathrm{d}s\\
&-\int_{(r,t]}\int_{\bar{\Delta}} \mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k})\left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k})\right]\hat{\pi}(\mathrm{d}s,\mathrm{d}\boldsymbol{k}).
\end{aligned}$$

Since all jump times are isolated and countable then this representation is permissible. From integration by parts we also get that

$$tN_t^{\sigma,\theta} = rN_r^{\sigma,\theta} + \int_{(r,t]} N_s^{\sigma,\theta}\,\mathrm{d}s + \int_{(r,t]} s\,\mathrm{d}N_s^{\sigma,\theta}.$$

Combining these we obtain

$$\begin{aligned}
\frac{tN_t^{\sigma,\theta}}{2} =&\frac{rN_r^{\sigma,\theta}}{2} + \int_{(r,t]}\left(\frac{N_s^{\sigma,\theta}}{2} - s\frac{N_s^{\sigma,\theta}(N_s^{\sigma,\theta}-1+\theta-\sigma)}{4}\right)\mathrm{d}s\\
&-\int_{(r,t]}\int_{\bar{\Delta}} \frac{s}{2}\mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k})\left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k})\right]\hat{\pi}(\mathrm{d}s,\mathrm{d}\boldsymbol{k}).
\end{aligned}\tag{1.42}$$

Next we need to check that one can formally let $r = 0$ in the above expression, recognise that the final term is then equal to $M_t^{\sigma,\theta}$, and rearrange the drift term to reflect the formula in (1.39).

Starting with the final term in (1.42), we first need to fix $T \geq 0$ and use (1.32) to find that, for $t \leq T$

$$\mathbb{E}\left[\sup_{s\leq t}\left(\frac{sN_s^{\sigma,\theta}}{2}\right)^k\right] < \infty. \tag{1.43}$$

We can then use properties of the compensated Poisson integral to bound the second moment of the martingale (1.40):

$$\begin{aligned}\mathbb{E}[(M_t^{\sigma,\theta})^2] &= \mathbb{E}\left[\int_0^t \int_{\bar{\Delta}} \frac{s^2}{4}\mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k})\nu(\mathrm{d}s, \mathrm{d}\boldsymbol{k})\right] \\ &= \mathbb{E}\left[\int_0^t s^2\frac{N_s^{\sigma,\theta}(N_s^{\sigma,\theta} - 1 + \theta + \sigma)}{8}\,\mathrm{d}s\right] \\ &\leq \mathbb{E}\left[\int_0^t s^2\frac{(N_s^{\sigma,\theta})^2}{8} + Ts\frac{(\theta+\sigma)N_s^{\sigma,\theta}}{8}\,\mathrm{d}s\right] \\ &\leq C_2 t, \qquad t \leq T,\end{aligned} \tag{1.44}$$

where $C_2 \geq 0$ comes from (1.43) with $k = 1, 2$. Thanks to [17, Theorem 8.23] and (1.44) we get that $M^{\sigma,\theta}$ given by (1.40) is a well defined square integrable martingale. Moreover, Doob's L^2 maximal inequality gives us the following bound:

$$\mathbb{E}\left[\sup_{s\leq t}(M_s^{\sigma,\theta})^2\right] \leq 4C_2 t, \qquad \forall t \leq T. \tag{1.45}$$

We note that the final term in (1.42) is equal to $M_t^{\sigma,\theta} - M_r^{\sigma,\theta}$ and by (1.45) we have $M_r^{\sigma,\theta} \to 0$ in L^2 as $r \to 0$.

Next, we rewrite the integral with respect to s in (1.42) as

$$
\begin{aligned}
A_r^{\sigma,\theta}(t) &:= \frac{1}{2}\int_{(r,t]} N_s^{\sigma,\theta}\,\mathrm{d}s - \frac{1}{2}\int_{(r,t]} s\frac{N_s^{\sigma,\theta}(N_s^{\sigma,\theta}-1+\theta-\sigma)}{2}\,\mathrm{d}s \\
&= -\int_{(r,t]} \frac{N_s^{\sigma,\theta}}{2}\left(\frac{s}{2}N_s^{\sigma,\theta}-1\right)\mathrm{d}s \\
&\qquad + \int_{(r,t]} \frac{s(1-\theta+\sigma)}{4}N_s^{\sigma,\theta}\,\mathrm{d}s.
\end{aligned}
$$

This allows us to rearrange (1.42) into the following:

$$
\begin{aligned}
&\frac{t}{2}N_t^{\sigma,\theta} - 1 + \int_{(r,t]} \frac{N_s^{\sigma,\theta}}{2}\left(\frac{s}{2}N_s^{\sigma,\theta}-1\right)\mathrm{d}s \\
&\quad = \frac{r}{2}N_r^{\sigma,\theta} - 1 + \int_{(r,t]} \frac{(1-\theta+\sigma)s}{4}N_s^{\sigma,\theta}\,\mathrm{d}s \\
&\qquad\qquad - (M_t^{\sigma,\theta} - M_r^{\sigma,\theta}).
\end{aligned}
$$

Applying (1.38) with $f(s) = sN_s^{\sigma,\theta}/2 - 1$, $g(s) = N_s^{\sigma,\theta}(sN_s^{\sigma,\theta}/2-1)/2$, $a = r$, and $b = t$ we find that

$$
\begin{aligned}
&\sup_{r\le s\le t}\left|\frac{s}{2}N_s^{\sigma,\theta}-1\right| \\
&\quad \le 2\Bigg(\left|\frac{r}{2}N_r^{\sigma,\theta}-1\right| + |M_r^{\sigma,\theta}| \\
&\qquad\qquad + \sup_{r\le s\le t}|M_s^{\sigma,\theta}| + \int_{(r,t]}\left|\frac{(1-\theta+\sigma)s}{4}N_s^{\theta}\right|\mathrm{d}s\Bigg).
\end{aligned}
$$

Letting $r \to 0$ in the above and using (1.24) gives us

$$\sup_{s\le t}\left|\frac{s}{2}N_s^{\sigma,\theta}-1\right| \\ \le 2\left(\sup_{s\le t}|M_s^{\sigma,\theta}|+\int_0^t\left|\frac{(1-\theta+\sigma)s}{4}N_s^{\sigma,\theta}\right|\,\mathrm{d}s\right), \quad (1.46)$$

where letting $r=0$ in the integral on the right hand side is permissible thanks to (1.43) with $k=1$. Squaring both sides of (1.46), applying expectations and using (1.45) and (1.43) gives us that there exists $C_3>0$ such that

$$\mathbb{E}\left[\sup_{s\le t}\left(\frac{sN_s^{\sigma,\theta}}{2}-1\right)^2\right]\le C_3 t, \qquad \forall t\le T. \quad (1.47)$$

We can now use this bound, which improves on (1.32), to ensure the integral term in (1.39) is well-defined. Letting $X_t^{\sigma,\theta}:=X_1^{\sigma,\theta}(t)=tN_t^{\sigma,\theta}/2-1$, (1.47) (along with Cauchy-Schwarz) allows us to control $|X_t^{\sigma,\theta}|$ as follows:

$$\mathbb{E}\left[|X_t^{\sigma,\theta}|\right]\le\sqrt{C_3 t},\ t\le T. \quad (1.48)$$

If we now consider the integral term in (1.39) we see that by (1.48)

$$\begin{aligned}\mathbb{E}\left[\int_0^t|X_s^{\sigma,\theta}|\frac{1}{s}\,\mathrm{d}s\right]&=\int_0^t\mathbb{E}\left[|X_s^{\sigma,\theta}|\right]\frac{1}{s}\,\mathrm{d}s\\&\le\int_0^t\frac{\sqrt{C_3}}{\sqrt{s}}\,\mathrm{d}s\\&=2\sqrt{C_3 t}.\end{aligned}$$

This shows that the integral with respect to s in (1.39) has finite expectation and thus is finite a.s. and so is well-defined.

Next, we can express $A_r^{\sigma,\theta}(t)$ as

$$A_r^{\sigma,\theta}(t) = -\int_{(r,t]} \left(\frac{s}{2}N_s^{\sigma,\theta} - 1\right)^2 \frac{1}{s}\,\mathrm{d}s - \int_{(r,t]} \left(\frac{s}{2}N_s^{\sigma,\theta} - 1\right) \frac{1}{s}\,\mathrm{d}s + \frac{1}{2}\int_{(r,t]} \frac{(1-\theta+\sigma)s}{2} N_s^{\sigma,\theta}\,\mathrm{d}s.$$

Using (1.47) along with (1.43) we find that

$$\mathbb{E}\left[\left|A_r^{\sigma,\theta}(t) + \int_{(r,t]} \left(\frac{s}{2}N_s^{\sigma,\theta} - 1\right)\frac{1}{s}\,\mathrm{d}s\right|\right] \le C_4 t, \qquad \forall t \le T,$$

where $C_4 > 0$ does not depend on r. This shows that, as $r \to 0$, $A_r^{\sigma,\theta}(t)$ converges in L^1 to

$$-\int_0^t \left(\frac{sN_s^{\sigma,\theta}}{2} - 1\right)\frac{1}{s}\,\mathrm{d}s + R_t^{\sigma,\theta},$$

where

$$R_t^{\sigma,\theta} := -\int_0^t \left(\frac{s}{2}N_s^{\sigma,\theta} - 1\right)^2 \frac{1}{s}\,\mathrm{d}s + \frac{1}{2}\int_0^t \frac{(1-\theta+\sigma)s}{2} N_s^{\sigma,\theta}\,\mathrm{d}s. \tag{1.49}$$

To conclude, (1.47) applied to the first term on the right hand side of (1.49) and (1.43) applied to the second term yield the bound in (1.41). □

Next, we will reduce the problem of convergence of (1.4) to convergence of the following process defined via the martingale (1.40):

$$
\begin{aligned}
Y_t^{\sigma,\theta} &:= -\frac{1}{t}\int_0^t u \, \mathrm{d}M_u^{\sigma,\theta}, \quad t > 0, \quad Y_0^{\sigma,\theta} = 0, \\
Y_\epsilon^{\sigma,\theta}(t) &:= \epsilon^{-1/2} Y_{\epsilon t}^{\sigma,\theta}.
\end{aligned}
$$

Lemma 1.15. *The process $(Y_t^{\sigma,\theta})_{t\geq 0}$ satisfies the equation*

$$
Y_t^{\sigma,\theta} = -\int_0^t Y_s^{\sigma,\theta} \frac{1}{s} \, \mathrm{d}s - M_t^{\sigma,\theta}. \tag{1.50}
$$

Moreover, there exists $C_5 > 0$ such that for any $t \leq T$

$$
\mathbb{E}\left[\sup_{s\leq t} \left(Y_s^{\sigma,\theta}\right)^2\right] \leq C_5 t. \tag{1.51}
$$

Finally, we have that

$$
\lim_{\epsilon\to 0} \mathbb{E}\left[\sup_{s\leq t} \left| X_\epsilon^{\sigma,\theta}(s) - Y_\epsilon^{\sigma,\theta}(s) \right|\right] = 0. \tag{1.52}
$$

Proof. First we consider $L_t^{\sigma,\theta} := tY_t^{\sigma,\theta}$ so that $L^{\sigma,\theta}$ is a square integrable martingale with quadratic variation

$$
\left[L^{\sigma,\theta}\right]_t = \frac{1}{4}\int_0^t \int_{\bar{\Delta}} s^4 \mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k}) \pi(\mathrm{d}s, \mathrm{d}\boldsymbol{k}),
$$

which has expectation

$$\begin{aligned}\mathbb{E}\left[[L^{\sigma,\theta}]_t\right] &= \frac{1}{4}\mathbb{E}\left[\int_0^t \int_{\bar{\Delta}} s^4 \mathbb{1}_{\bar{\Delta}_{N^{\sigma,\theta}_{s-}}}(\boldsymbol{k})\nu(\mathrm{d}s, \mathrm{d}\boldsymbol{k})\right] \\ &= \frac{1}{4}\mathbb{E}\left[\int_0^t s^2 \left(s^2 \frac{N_s^{\sigma,\theta}(N_s^{\sigma,\theta} - 1 + \theta + \sigma)}{2}\right) \mathrm{d}s\right] \\ &\leq C_2 \int_0^t s^2 \,\mathrm{d}s = \frac{C_2}{3}t^3,\end{aligned} \tag{1.53}$$

where we use the same bound from (1.44) on the expectation of the bracketed term in (1.53).

This now gives us that

$$\mathbb{E}\left[\left(Y_t^{\sigma,\theta}\right)^2\right] = \frac{1}{t^2}\mathbb{E}\left[\left(L_t^{\sigma,\theta}\right)^2\right] = \frac{1}{t^2}\mathbb{E}\left[[L^{\sigma,\theta}]_t\right] \leq \frac{C_2}{3}t. \tag{1.54}$$

In order to obtain (1.50) we first use integration by parts to write

$$\begin{aligned}Y_t^{\sigma,\theta} - Y_r^{\sigma,\theta} &= \int_{(r,t]} \frac{1}{s^2}\int_0^s u \,\mathrm{d}M_u^{\sigma,\theta}\,\mathrm{d}s - \int_{(r,t]} \mathrm{d}M_s^{\sigma,\theta} \\ &= -\int_{(r,t]} \frac{1}{s} Y_s^{\sigma,\theta}\,\mathrm{d}s - M_t^{\sigma,\theta} + M_r^{\sigma,\theta},\end{aligned}$$

and let $r \to 0$; using (1.54) and Jensen's inequality to bound $\mathbb{E}[\int_{(r,t]} |Y_s^{\sigma,\theta}|/s\,\mathrm{d}s]$ uniformly in $r > 0$ by $2\sqrt{Ct}/\sqrt{3}$, ensuring that $\int_0^t Y_s^{\sigma,\theta}/s\,\mathrm{d}s$ exists almost surely in the same way as $\int_0^t X_s^{\sigma,\theta}/s\,\mathrm{d}s$.

Now, we apply (1.38) with $f(s) = Y_s^{\sigma,\theta}$, $g(s) = Y_s^{\sigma,\theta}/s$,

$a = 0$ and $b = t$ to get that

$$\sup_{s \le t} |Y_s^{\sigma,\theta}| \le 2 \sup_{s \le t} M_s^{\sigma,\theta}.$$

We can then square both sides and apply expectations along with (1.45) to obtain (1.51).

Finally, we prove (1.52) by recalling that $X_t^{\sigma,\theta} = tN_t^\theta/2 - 1$ and so by Lemma 1.13 and (1.50),

$$X_t^{\sigma,\theta} - Y_t^{\sigma,\theta} = -\int_0^t (X_s^{\sigma,\theta} - Y_s^{\sigma,\theta}) \frac{1}{s} \, \mathrm{d}s + R_t^\theta.$$

Applying (1.38) to the above we find that

$$\sup_{s \le t} |X_s^{\sigma,\theta} - Y_s^{\sigma,\theta}| \le 2 \sup_{s \le t} |R_s^{\sigma,\theta}|.$$

This leads to

$$\begin{aligned}
&\mathbb{E}\left[\sup_{s \le t} \left|X_\epsilon^{\sigma,\theta}(s) - Y_\epsilon^{\sigma,\theta}(s)\right|\right] \\
&= \epsilon^{-1/2} \mathbb{E}\left[\sup_{s \le t} \left|X^{\sigma,\theta}(\epsilon s) - Y^{\sigma,\theta}(\epsilon s)\right|\right] \\
&\le 2\epsilon^{-1/2} \mathbb{E}\left[\sup_{s \le t} |R_{\epsilon s}^{\sigma,\theta}|\right] \\
&\le 2C_1 \sqrt{\epsilon} t,
\end{aligned}$$

using (1.41) for the final inequality. Letting $\epsilon \to 0$ in the above gives us (1.52). □

We now have everything in place to prove Theorem 1.1. The proof consists of first checking the conditions of [18, Chapter

7, Theorem 1.4] are satisfied for the process

$$L_\epsilon^{\sigma,\theta}(t) := -tY_\epsilon^{\sigma,\theta}(t) = \frac{1}{\epsilon^{3/2}} \int_0^{\epsilon t} u \, \mathrm{d}M_u^{\sigma,\theta}, \qquad t \geq 0,$$

and proving its convergence to $(tZ_t)_{t\geq 0}$ as $\epsilon \to 0$. Then, as in [11], we take advantage of Steps 2–4 in the proof of [19, Lemma 4.8] to extend this to the convergence of $Y_\epsilon^{\sigma,\theta}$ to Z as $\epsilon \to 0$. Essentially, the continuity of $t \mapsto 1/t$ away from zero along with the control we have over these processes near zero is what ensures this convergence. Finally, (1.52) then gives us the convergence of (1.4) to (1.3).

Proof of Theorem 1.1. Starting with the process $L_\epsilon^{\sigma,\theta}$, we first note that, since $L_\epsilon^{\sigma,\theta}(t) = -\epsilon^{-3/2} L^{\sigma,\theta}(\epsilon t)$, it is an L^2-martingale of the form

$$\begin{aligned} &L_\epsilon^{\sigma,\theta}(t) \\ &= \epsilon^{-3/2} \frac{1}{2} \int_0^{\epsilon t} \int_{\bar{\Delta}} s^2 \mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k}) \left[\mathbb{1}_{j>0}(\boldsymbol{k}) - \mathbb{1}_{j=0}(\boldsymbol{k}) \right] \hat{\pi}(\mathrm{d}s, \mathrm{d}\boldsymbol{k}). \end{aligned} \tag{1.55}$$

The compensator of the square of this process is given by

$$\begin{aligned} \langle L_\epsilon^{\sigma,\theta} \rangle_t &= \frac{1}{4\epsilon^3} \int_0^{\epsilon t} \int_{\bar{\Delta}} s^4 \mathbb{1}_{\bar{\Delta}_{N_{s-}^{\sigma,\theta}}}(\boldsymbol{k}) \nu(\mathrm{d}s, \mathrm{d}\boldsymbol{k}), \\ &= \frac{1}{4\epsilon^3} \int_0^{\epsilon t} s^4 \frac{N_s^{\sigma,\theta}(N_s^{\sigma,\theta} - 1 + \theta + \sigma)}{2} \mathrm{d}s \\ &= \frac{1}{2} \int_0^t s^2 (\epsilon s)^2 \frac{N_{\epsilon s}^{\sigma,\theta}(N_{\epsilon s}^{\sigma,\theta} - 1 + \theta + \sigma)}{4} \mathrm{d}s. \end{aligned} \tag{1.56}$$

We now wish to verify that the assumptions (b) in [18, Theorem 1.4, Chapter 7] are satisfied with M_n

corresponding to $L_\epsilon^{\sigma,\theta}$, A_n corresponding to $\langle L_\epsilon^{\sigma,\theta}\rangle$ and $C_t = \int_0^t u^2\,\mathrm{d}u/2$. A simplified version of this theorem can be found in Appendix A.

First, we note that since $\langle L_\epsilon^{\sigma,\theta}\rangle$ is continuous we only need to show (i) that $\langle L_\epsilon^{\sigma,\theta}\rangle_t$ converges as $\epsilon \to 0$ to $\int_0^t u^2/2\,\mathrm{d}u$ in probability for each $t > 0$, and (ii) that for any $T > 0$,

$$\lim_{\epsilon\to 0} \mathbb{E}\left[\sup_{t\leq T} |L_\epsilon^{\sigma,\theta}(t) - L_\epsilon^{\sigma,\theta}(t-)|^2\right] = 0. \tag{1.57}$$

The first claim (i) follows from the representation (1.56) along with Proposition 1.11. To address (ii): the limit in (1.57) holds since, thanks to the representation (1.55), the jumps of $L_\epsilon^{\sigma,\theta}$ on $[0,T]$ are isolated and uniformly bounded by $\epsilon^{-3/2}(\epsilon T)^2/2$ which converges to zero as $\epsilon \to 0$. Thus we have the convergence of $L_\epsilon^{\sigma,\theta} \to (tZ_t)_{t\geq 0}$ in law in $D([0,\infty),\mathbb{R})$.

Before concluding the proof one needs the following bound on the limiting process Z:

$$\mathbb{E}\left[\sup_{s\leq t} Z_s^2\right] \leq C_6 t, \qquad C_6 \geq 0. \tag{1.58}$$

This bound follows from the formula

$$Z_t = -\int_0^t Z_s \frac{1}{s}\,\mathrm{d}s + \frac{1}{\sqrt{2}} W_t,$$

in combination with (1.38), similarly to the proof of (1.51).

Now, thanks to continuity away from zero of $t \mapsto 1/t$ we have the desired convergence of $Y_\epsilon^{\sigma,\theta} \to Z$, but not at zero. To get the full convergence we consider the processes

$$Y_\epsilon^{\sigma,\theta,(b)}(t) := -\left(\frac{1}{b}\mathbb{1}_{[0,b]}(t) + \frac{1}{t}\mathbb{1}_{(b,\infty)}(t)\right) L_\epsilon^{\sigma,\theta}(t),$$

and

$$Z^{(b)}(t) := \left(\frac{t}{b}\mathbb{1}_{[0,b]}(t) + \mathbb{1}_{(b,\infty)}(t)\right) Z(t)$$

and look at what happens as $b \to 0$. Again, thanks to the continuity of $t \mapsto 1/t$ away from zero we have that $Y_\epsilon^{\sigma,\theta,(b)}$ converges in law in $D([0,\infty),\mathbb{R})$ to $Z^{(b)}$ as $\epsilon \to 0$. Applying the triangle inequality twice yields

$$\begin{aligned} \mathbb{E}\left\|Y_\epsilon^{\sigma,\theta} - Z\right\|_\infty \leq \mathbb{E}\left\|Y_\epsilon^{\sigma,\theta} - Y_\epsilon^{\sigma,\theta,(b)}\right\|_\infty + \mathbb{E}\left\|Y_\epsilon^{\sigma,\theta,(b)} - Z^{(b)}\right\|_\infty \\ + \mathbb{E}\left\|Z^{(b)} - Z\right\|_\infty, \qquad (1.59) \end{aligned}$$

where $\|\cdot\|_\infty$ denotes the sup norm in $D([0,\infty),\mathbb{R})$. Considering the first term on the right hand side of (1.59) we see that

$$\begin{aligned} \left\|Y_\epsilon^{\sigma,\theta} - Y_\epsilon^{\sigma,\theta,(b)}\right\|_\infty &= \sup_{t\leq b}\left(\frac{1}{t} - \frac{1}{b}\right)\left|L_\epsilon^{\sigma,\theta}(t)\right| \\ &= \sup_{t\leq b}\left(1 - \frac{t}{b}\right)|Y_\epsilon^{\sigma,\theta}(t)| \\ &\leq \sup_{t\leq b}|Y_\epsilon^{\sigma,\theta}(t)|. \end{aligned}$$

Applying expectations and (1.51) this goes to zero as $b \to 0$, independently of ϵ. The same calculation can be performed with the third term in (1.59) thanks to (1.58). As discussed, the middle term in (1.59) goes to zero thanks to continuity away from zero and so we have the desired convergence. $\square$

2 A Central Limit Theorem for the Fleming–Viot Process

Moving now onto processes that model populations forward in time, we have objects like the Wright–Fisher (WF) diffusion and the Fleming–Viot (FV) process that represent the distribution of genetic types in the population at a single locus over time; with the former modelling a finite number of types and the latter an infinite number of types. There have even been extensions to two and n locus models. These have been defined through Markov generators and can be characterised as couplings of one locus models with a degree of interaction between loci (either through recombination or gene conversion, see [20] for details).

Here our motivation comes from the field of Quantative Genetics. Understanding how genetic differences determine complex traits is one of the fundamental questions in this field. Though there are examples of areas of the human genome that contribute heavily to a single phenotype - for example the HLA (human leukocyte antigen) locus for several inflammatory diseases [21] - it is largely believed that most phenotypes for complex traits are influenced by small contributions from a large number of loci [22, 23]. With this viewpoint in mind, we look to use the existing single locus processes to model long genomes with a large number of loci. To do this we consider a growing number of loci each with negligible effect on the overall genotype and let the number of loci tend to infinity, considering LLN and CLT limits.

Mathematically we start with a collection of i.i.d. neutral FV processes $((F_{i,t})_{t\geq 0})_{i\in\mathbb{N}}$ with parent-independent mutation defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. With our space of genetic types being $[0,1]$, these processes take values in $\mathcal{P}([0,1])$ where $F_{i,t}$ represents the genotypic distribution at the i^{th} locus in the population. The mutation mechanism has two parameters, $(\theta, \nu_0) \in [0,\infty) \times \mathcal{P}([0,1])$, that represent the total mutation rate and the distribution from which new types are picked, respectively. We also enforce that all of the starting points $F_{i,0}$ are independent draws from the same distribution $Q_0 \in \mathcal{P}(\mathcal{P}([0,1]))$, with shared mean μ_0.

Along with independence of these loci, assuming that all of the FV processes start from the same probability distribution is less than desirable; however establishing this simple case is useful in order to expand these ideas further.

Alongside this discussion of genotypes is an attempt to describe phenotypes - i.e. the expressed characteristics of organisms, like eye and hair colour for example. One can think of a phenotype as a function of genotypes, plus some environmental noise. In our case, we track genotypes via probability distributions so it makes sense to think of phenotypes as integrals:

$$\text{Phenotype} = \int f(x)\text{Genotype}(\mathrm{d}x) + \text{Noise}.$$

The function f here assigns a value to each genetic type that represents its contribution to the phenotype. If we replace "Genotype(dx)" with $F_{1,t}(\mathrm{dx})$ then we will only see the contributions to a phenotype from a single locus. With our model though we hope to capture the aggregate behaviour of all loci.

We start this chapter by considering the LLN and CLT limits

$$S_t^{(N)} := \frac{1}{N}\sum_{i=1}^{N} F_{i,t} \tag{2.1}$$

and

$$\eta_t^{(N)} := \frac{1}{\sqrt{N}}\sum_{i=1}^{N} F_{i,t} - \sqrt{N}\mathbb{E}_{Q_0}\left[F_{1,t}\right] \tag{2.2}$$

for a finite number of time points, ensuring that the finite dimensional distributions of our processes converge as the number of loci grows to infinity. Since we are concerned with weak convergence we will mainly be considering integrals of continuous functions against these measures. From here on we will use the following notation:

$$\langle f, \mu\rangle := \int_{[0,1]} f(x)\mu(dx).$$

With this established we then look to prove tightness of both the LLN and CLT collections. The CLT object will at first be considered as a signed measure in $\mathcal{M}([0,1])$, though later will be considered as a Schwartz distribution in the space $\mathcal{D}([0,1])$ as there it is easier to obtain tightness. This lifting argument is outlined in Section 2.3.

The transition function of the FV process is essential in proving tightness and will be utilised to analyse the process's small-time behaviour. This transition function exploits the duality between these processes and the Kingman coalescent with mutation - denoted $(N_t)_{t\geq 0}$ in this chapter as it is the only coalescent used - to control how much the process can change over a small time window. The first three sections of this chapter will constitute the proof of the following theorem:

Theorem 2.1. *Let $(F_i)_{i=1}^{\infty} \subset C([0,\infty),\mathcal{P}([0,1]))$ be an i.i.d. collection of FV processes, all started at independent draws from $Q_0 \in \mathcal{P}(\mathcal{P}([0,1]))$, with parent-independent mutation, parameters $(\theta, \nu_0) \in [0,\infty) \times \mathcal{P}([0,1])$. Then the processes $S^{(N)} \in C([0,\infty),\mathcal{P}(0,1))$ and $\eta^{(N)} \in C([0,\infty),\mathcal{D}([0,1]))$ defined by* (2.1) *and* (2.2) *respectively converge in distribution in their respective spaces to μ and η which are characterised via Propositions 2.3 and 2.5.*

After this, we investigate the continuity of the Gaussian Schwartz distribution-valued process η further; considering its continuity in the strong dual topology. We will then conclude by considering the situation where our collection of FV processes, $(F_{i,t})_{i\in\mathbb{N}}$, are started from different random probability measures and prove the non-identical version of the above theorem in Section 2.6. First though, we start with some background on the Dirichlet process and parent-independent mutation.

2.1 The Dirichlet Process and Parent-Independent Mutation

One of the fundamental building blocks of the Fleming-Viot process is the Dirichlet Process. We will see in Section 2.4 the precise way in which these interact but first we start with the definition of a Dirichlet process from [24]:

Definition 2.2. *A random probability measure F is distributed as a Dirichlet Process on a Polish Space $(S, \mathcal{F})$ with parameter μ (a positive, finite measure on S) if, for every $A_1, ..., A_k$ a measurable partition of S, we have that*

$$(F(A_1), ..., F(A_k)) \sim Dir_{\mu(A_1),...,\mu(A_k)}$$

where $Dir(\alpha_1, ..., \alpha_k)$ is the Dirichlet distribution on the simplex Δ_{k-1}

$$\Delta_{k-1} = \left\{ (p_1, \ldots, p_k) \middle| \sum_{i=1}^{k} p_i = 1 \right\},$$

and

$$Dir_{\alpha_1,...,\alpha_k}(\mathrm{d}y_1, ..., \mathrm{d}y_k) \propto \prod_{i=1}^{k} y_i^{\alpha_i - 1}\mathrm{d}y_i.$$

We denote this as $F \sim \Pi_\mu$.

An equivalent definition of the Dirichlet distribution on the real line found in [24] uses the gamma distribution - denoted $\mathcal{G}(\alpha, \beta)$ with shape parameter $\alpha \geq 0$ and scale parameter $\beta > 0$. When $\alpha = 0$ the distribution is degenerate at zero, with $\alpha > 0$ the distribution has a density with respect to Lebesgue measure on $\mathbb{R}$ as follows:

$$f(z|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} e^{-z/\beta} z^{\alpha-1} \mathbb{1}_{(0,\infty)}(z).$$

If we let $Z_1, ..., Z_k$ be independent with $Z_j \sim \mathcal{G}(\alpha_j, 1)$ (where at least one of the $\alpha_j > 0$) then the Dirichlet distribution with parameter $(\alpha_1, ..., \alpha_k)$ can be defined as the distribution of $(Y_1, ..., Y_k)$, where

$$Y_j = \frac{Z_j}{\sum_{i=1}^{k} Z_i}, \quad j = 1, ..., k.$$

If $(Y_1, ..., Y_k)$ has a Dirichlet distribution with parameter $(\alpha_1, ..., \alpha_k)$ then, defining $\alpha := \sum_{i=1}^{k} \alpha_i$ we have the following moments of the Y_i:

- $\mathbb{E}[Y_i] = \frac{\alpha_i}{\alpha}$,
- $\mathbb{E}[Y_i^2] = \frac{\alpha_i(\alpha_i+1)}{\alpha(\alpha+1)}$,
- $\mathbb{E}[Y_i Y_j] = \frac{\alpha_i \alpha_j}{\alpha(\alpha+1)}$.

Oftentimes we decompose the measure μ down to $\mu = \theta\nu_0$ where $\theta = \mu(S) > 0$ and so ν_0 is a probability measure on S. In this case we use the notation $F \sim \Pi_{\theta,\nu_0}$. The above definition is quite abstract so we will highlight an alternative characterisation of the Dirichlet Process (DP). These characterisations also highlight an important feature of DPs in that any realisation of a DP is almost surely a discrete measure.

With $\theta > 0$ and ν_0 as above, consider a Poisson point process (PPP) on $(0, \infty)$ with intensity $\theta x^{-1} e^{-x} \mathrm{d}x$. Almost surely, we can order the points of this process $\Gamma_1 > \Gamma_2 > \ldots$. Letting $\Sigma = \sum_i \Gamma_i$, $P_i = \Gamma_i / \Sigma$ and

$$F = \sum_{i=1}^{\infty} P_i \delta(X_i), \tag{2.3}$$

where the X_i are i.i.d(ν_0) (and are independent of the PPP), F is then distributed as a Dirichlet Process with parameters (θ, ν_0) [18, 25]. In this case the (P_i) are said to have a Poisson-Dirichlet (PD) distribution with parameter θ. We also have that $\Sigma \sim \mathcal{G}(\theta, 1)$ and is independent of F [24]. Alternatively, one can construct the PD(θ) distribution via stick-breaking. Here one needs a collection $(W_i)_{i \geq 1}$ of i.i.d. Beta(1,θ) random variables and then defining $\tilde{P}_j$ as

$$\tilde{P}_j = \left[\prod_{i=1}^{j-1} (1 - W_i) \right] W_j$$

we obtain a collection such that $\sum_{i=1}^{\infty} \tilde{P}_i = 1$ and we say that the $(\tilde{P}_j)$ have the GEM(θ) distribution. If we then rank a GEM(θ) distribution and re-label the sequence $P_1 \geq P_2 \geq \ldots$ we now have that (P_i) has the PD(θ) distribution [26].

Parent-independent mutation is often the preferred model used in population genetics due to its mathematical tractability. In the infinitely-many alleles model, every time a mutation occurs it is to a new allelic type never before seen in the population. The way this is incorporated into the generator of the FV process (0.1) is by setting

$$(Af)(x) := \frac{\theta}{2} \int_S (f(\xi) - f(x))\nu_0(d\xi),$$

henceforth this is referred to as parent-indpendent mutation with parameters $(\theta, \nu_0) \in \mathbb{R}^+ \times \mathcal{P}(S)$. Here mutations occur with intensity $\theta/2$ and the new type is picked according to the probability measure ν_0. ν_0 is assumed to be non-atomic so that, when a new type is picked, it is almost surely a new one; in-line with the infinitely-many alleles model.

We will see in Section 2.4 that, at any time $t > 0$, the FV process is distributed according to a Dirichlet Process. The first parameter (from $\mathbb{R}^+$) will be of the form $N_t + \theta$ where $N_t \in \mathbb{N}_0$ is the number of blocks in the Kingman coalescent at time t. The probability measure parameter is the measure $\theta\nu_0$ plus an N_t number of atoms from the starting distribution F_0 divided by $(N_t + \theta)$ (to ensure it has total mass 1). From the representation (2.3) we see that, since the X_i will be picked from a mixture of these atomic and non-atomic measures, there will be an influence from the past dependent on the strength of mutation. If mutation is strong (corresponding to a large value of θ) then N_t will be smaller (since mutations cause this process to decrease) and more weighting will be given to the new types picked by ν_0. If mutation is weak then it will take longer to get away from the influence of the ancestors in the original population.

With this in place we move now to set up our LLN and CLT constructions and ensure their convergence.

2.2 A Law of Large Numbers for the Fleming–Viot Process

In this section we consider the sequence of probability measure-valued stochastic processes

$$\left(S_t^{(N)}\right)_{t\geq 0} := \left(\frac{1}{N}\sum_{i=1}^{N} F_{i,t}\right)_{t\geq 0} \tag{2.4}$$

and wish to show that, as $N \to \infty$, these converge weakly in $C([0,\infty), \mathcal{P}([0,1]))$ to the deterministic limit $(\mu_t)_{t\geq 0}$ where

$$\mu_t := \sum_{n=0}^{\infty} \frac{d_n^\theta(t)}{n+\theta}\left(n\mu_0 + \theta\nu_0\right) = \mathbb{E}_{Q_0}\left[F_{1,t}\right], \tag{2.5}$$

where $\mu_0 = \mathbb{E}_{Q_0}[F_{1,0}]$, with the subscript denoting the distribution of the starting measure.

It should be noted that, behind the scenes, there is an intermediate step in obtaining equation (2.5). First, we condition on knowing the starting measure:

$$\mathbb{E}\left[F_{1,t}|F_{1,0}\right] = \sum_{n=0}^{\infty} \frac{d_n^\theta(t)}{n+\theta}\left(nF_{1,0} + \theta\nu_0\right).$$

Then we apply expectations to both sides of the above expression to obtain (2.5) via the tower law; where the expectation operator can be taken through the sum since the collection $(d_n^\theta(t))_{n\in\mathbb{N}}$ forms a probability distribution. This mechanism is implicitly working behind the scenes whenever we state expressions for higher and mixed moments of a Fleming-Viot process at time $t > 0$.

The presence of $d_n^\theta(t) = \mathbb{P}_\infty(N_t = n)$ allows us to write the above as

$$\mu_t = \mathbb{E}_\infty\left[\frac{N_t}{N_t+\theta}\right]\mu_0 + \mathbb{E}_\infty\left[\frac{\theta}{N_t+\theta}\right]\nu_0. \tag{2.6}$$

Tavaré [27] in fact provides a way to exactly solve these moments. There at the bottom of page 157 they define a process

$$Z_m(t) = \frac{\rho_m^{-1}(t)(N_t)_{(m)}}{(N_t+\theta)_{(m)}}$$

for $N_0 = i$, where $\rho_m(t) = \exp(-m(m+\theta-1)t/2)$. The top of page 158 tells us that

$$\mathbb{E}\left[Z_m(t)\right] = \frac{i_{[m]}}{(i+\theta)_{(m)}}.$$

Letting $i \to \infty$ and setting $m = 1$ then results in

$$\mathbb{E}_\infty \left[\frac{N_t}{N_t + \theta} \right] = \rho_1(t) = e^{-\theta t/2}.$$

This in turn gives us

$$\mu_t = e^{-\theta t/2}\mu_0 + (1 - e^{-\theta t/2})\nu_0.$$

As in Chapter 1 the subscripts here indicate that the coalescent in question starts from infinity. The above relationship is due to the moment duality between the FV process and the Kingman coalescent; this is best exemplified in the transition function of the FV process which is discussed in Section 2.4.

In order to obtain weak convergence of probability distributions over $C([0,\infty),\mathcal{P}([0,1]))$ we needs two things: weak convergence of finite dimensional distributions and relative compactness of the collection over N [18, Chapter 3, Theorem 7.8]. We start by considering the finite-dimensional distributions:

Proposition 2.3. *Let $k \in \mathbb{N}$ and $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$. Then*

$$(S_{t_1}^{(N)}, \ldots, S_{t_k}^{(N)}) \xrightarrow{d} (\mu_{t_1}, ..., \mu_{t_k}) \tag{2.7}$$

as $N \to \infty$.

Proof. In order to show the above weak convergence of finite dimensional distributions we let $f_1, \ldots, f_k \in C([0,1],\mathbb{R})$ and consider the object

$$\left(\langle f_1, S_{t_1}^{(N)}\rangle, \ldots, \langle f_k, S_{t_k}^{(N)}\rangle\right). \tag{2.8}$$

The Portmanteau Theorem now means that in order to determine (2.7) we just need to ensure that (2.8) converges to the corresponding object for $(\mu_t)_{t\geq 0}$. Using (2.4) and the linearity of integrals we see that (2.8) is equal to

$$\frac{1}{N}\sum_{i=1}^{N}\left(\langle f_1, F_{i,t_1}\rangle, \ldots, \langle f_k, F_{i,t_k}\rangle\right),$$

which, thanks to the i.i.d. nature of the F_i, converges to

$$\left(\langle f_1, \mu_{t_1}\rangle, \ldots, \langle f_k, \mu_{t_k}\rangle\right)$$

by the Classic Law of Large Numbers. $\square$

Now onto relative compactness. Using Theorem II.4.1 in [28] we can reduce the problem of showing relative compactness of

$$\left((S_t^{(N)})_{t\geq 0}\right)_{N\in\mathbb{N}}$$

in $C([0,\infty), \mathcal{P}([0,1]))$ to showing relative compactness of

$$\left((\langle f, S_t^{(N)}\rangle)_{t\geq 0}\right)_{N\in\mathbb{N}}$$

in $C([0,\infty), \mathbb{R})$ for all $f \in C([0,1], \mathbb{R})$.

To establish relative compactness of these real-valued processes we will appeal to Theorem I.4.3 of [29] and show that, for each $f \in C([0,1], \mathbb{R})$, $\langle f, S_t^{(N)}\rangle$ satisfies the following:

$$\mathbb{E}\left[\left|\langle f, S_t^{(N)}\rangle - \langle f, S_s^{(N)}\rangle\right|^4\right] \leq C_T|t-s|^2, \tag{2.9}$$

for all $s, t \in [0, T]$, $T = 1, 2, 3, ...$, where $C_T > 0$ is a constant that depends only on T. Before ensuring this condition is satisfied, we fix $f \in C([0,1], \mathbb{R})$, let $X_{t-s} := \langle f, F_{1,t}\rangle - \langle f, F_{i,s}\rangle$ and bound (2.9) above by the following:

$$\begin{aligned}
&\frac{1}{N^4}\mathbb{E}\left[\left(\sum_{i=1}^{N}\langle f, F_{i,t}\rangle - \langle f, F_{i,s}\rangle\right)^4\right] \\
&= \frac{1}{N^4}\left\{N\mathbb{E}_{Q_0}\left[X_{t-s}^4\right] + 4N_{[2]}\mathbb{E}_{Q_0}\left[X_{t-s}^3\right]\mathbb{E}_{Q_0}\left[X_{t-s}\right]\right. \\
&\quad + 3N_{[2]}\mathbb{E}_{Q_0}\left[X_{t-s}^2\right]^2 + 6N_{[3]}\mathbb{E}_{Q_0}\left[X_{t-s}^2\right]\mathbb{E}_{Q_0}\left[X_{t-s}\right]^2 \\
&\quad \left. + N_{[4]}\mathbb{E}_{Q_0}\left[X_{t-s}\right]^4\right\} \\
&\leq \mathbb{E}_{Q_0}\left[X_{t-s}^4\right] + 4\left|\mathbb{E}_{Q_0}\left[X_{t-s}^3\right]\mathbb{E}_{Q_0}\left[X_{t-s}\right]\right| \qquad (2.10) \\
&\quad + 3\mathbb{E}_{Q_0}\left[X_{t-s}^2\right]^2 + 6\mathbb{E}_{Q_0}\left[X_{t-s}^2\right]\mathbb{E}_{Q_0}\left[X_{t-s}\right]^2 \\
&\quad + \mathbb{E}_{Q_0}\left[X_{t-s}\right]^4.
\end{aligned}$$

Here we are using the i.i.d. nature of the summands along with bounding the powers of N outside the expectations above by 1. This simplification now allows us to instead consider the regularity of a continuous function integrated against a single FV process.

We move now to consider the convergence of the finite dimensional distributions of the corresponding CLT object before concluding tightness of the LLN object. Once this is completed we will return to the notion of tightness and show relative compactness for both the LLN and CLT collections together in Section 2.4.

2.3 A Central Limit Theorem for the Fleming–Viot Process

In this section we define and investigate the limiting behaviour of the CLT object

$$\eta_t^{(N)} := \frac{1}{\sqrt{N}} \sum_{i=1}^{N} F_{i,t} - \sqrt{N}\mu_t. \tag{2.11}$$

This is a signed measure and the corresponding process lives in the space $C([0,\infty),\mathcal{M}([0,1]))$, with $\mathcal{M}([0,1])$ endowed with the weak* topology; the coarsest topology that ensures that the map

$$\mu \mapsto \langle f, \mu \rangle$$

is continuous for all $f \in C([0,1],\mathbb{R})$. The continuity in t of these objects for $N \in \mathbb{N}$ is assured by the continuity of the summands $F_{i,t}$ and their shared mean μ_t.

Unfortunately for us, there seem to be no equivalents to Theorem 9.1.VI in [30] or Theorem II.4.1 in [28] that allow us to consider

$$\langle f, \eta_t^{(N)} \rangle, \quad f \in C([0,1],\mathbb{R}),$$

instead of the signed measures themselves when proving convergence of finite dimensional distributions and tightness. Furthermore, due to the nature of the FV process, at any time $t > 0$, F_t is almost surely an atomic measure. Its mean however is diffuse and so $\eta_t^{(N)}$ is already in its Hahn decomposition and its total variation is $2\sqrt{N}$; which of course is unbounded as N tends to infinity.

Since compact sets of $\mathcal{M}([0,1])$ are difficult to characterise, and the total variation of our object blows up as $N \to \infty$, it seems we need to consider a less restrictive topology so that the distributions of $((\eta^{(N)})_{t\geq 0})_{N\in\mathbb{N}}$ are uniformly tight.

Before doing this we briefly note that the space of finitely additive signed measures with bounded variation can be viewed as the topological dual to the space of continuous functions, $C([0,1],\mathbb{R})$ [31]. The space $\mathcal{M}([0,1])$ is then of course a subspace of this dual space. $C([0,1],\mathbb{R})$ is a very large collection of functions (including almost every possible path of a Brownian motion of length 1) and induces a relatively restrictive topology when we take its dual. If however we consider a much smaller space of functions and take its dual we end up with a more general class of operators known as Distributions.

This smaller space of functions we will consider is $C^\infty([0,1],\mathbb{R})$, the space of infinitely differentiable real-valued functions on $[0,1]$. Distributions are then continuous linear functionals on this space when it is endowed with the *Canonical LF-Topology.* This topology is difficult to describe directly, though Trèves [32] provides the following general proposition to check if a linear functional on $C_c^\infty(U, \mathbb{R})$ (continuous functions on U with compact support, U an open subset of $\mathbb{R}^n$) is a distribution:

Proposition ([32], Proposition 21.1)**.** *A linear functional T on $C_c^\infty(U, \mathbb{R})$ is a distribution if and only if it possesses the following equivalent properties:*
1) To every compact subset K of U there is an integer $m \geq 0$ and a constant $C > 0$ such that, for all f with support in the set K,

$$|T(f)| \leq C \sup_{|p| \leq m} \left(\sup_{x \in U} \{\partial^p f(x)\} \right),$$

where p is a multi-index.
2) If a sequence of functions $\{f_n\}_{n \in \mathbb{N}}$ converge uniformly to zero, as well as all their derivatives, and if the functions f_n have support contained in a compact subset K of U, independent of the index n, then $T(f_n) \to 0$.

Using these conditions one can see that any signed measure can be considered as a distribution. In our case, since our underlying space is compact, our space of distributions are in fact Schwartz Distributions. We refer to this space as $\mathcal{D}([0, 1])$. For more details on these objects, see [33].

With this in place we can now consider $(\eta^{(N)})_{N \in \mathbb{N}}$ as distribution-valued processes and utilise Theorem 5.3 of [34]. This theorem allows us to invoke the existence of a process in $C([0,1],\mathcal{D}([0,1]))$ which is the weak limit of $(\eta_t^{(N)})_{0\leq t\leq 1}$ as $N\to\infty$. We state part 1 of this theorem here with our notation as it is an integral part of our CLT convergence. It is appealed to several times; not only to prove convergence of our processes but also to invoke the existence of the limiting process η:

Theorem ([34],Theorem 5.3)**.** *Suppose that the sample paths of X^n are elements of $C([0,1],\mathcal{D}([0,1]))$ for every $n\in\mathbb{N}$. Further suppose that for each f in $C([0,1],\mathbb{R})$ the sequence of distributions of $(\langle f,X_t^n\rangle)_{t\geq 0}$ is tight in $C([0,1],\mathbb{R})$ and for any finite elements $f_1,\ldots,f_k\in C([0,1],\mathbb{R})$ and points $t_1<\cdots<t_k$ in $[0,1]$, the distribution of*

$$(\langle f_1,X_{t_1}^n\rangle,\ldots,\langle f_k,X_{t_k}^n\rangle)$$

converges in law to some k-dimensional probability distribution. Then there exists the limit process X whose sample paths are elements in $C([0,1],\mathcal{D}([0,1]))$ such that $X^n\xrightarrow{d}X$.

Remark 2.4. *Note here that this theorem only relates to continuous paths of length 1. For our purposes, we invoke this theorem repeatedly for the time periods $[m,m+1]$ for $m=0,1,2,\ldots$. The continuity of our process then ensures that the concatenation of these time periods results in a well defined process on $C([0,\infty),\mathcal{D}([0,1]))$.*

The convergence of finite dimensional distributions is one requirement of this theorem. Before stating and proving this convergence we note that, from here on, $(F_t)_{t\geq 0}$ denotes a generic Fleming–Viot process with parent-independent mutation, parameters (θ, ν_0), with starting measure F_0. When seen inside an expectation or covariance operator, the subscript on the operator indicates the distribution of F_0. With that we move onto the proposition:

Proposition 2.5. *Let $(\eta_t^{(N)})_{t\geq 0}$ be as in (2.11), fix $k \in \mathbb{N}$, $f_1, \ldots, f_k \in C^\infty([0,1], \mathbb{R})$ and $0 \leq t_1 < \cdots < t_k \leq 1$. Then, as $N \to \infty$,*

$$\left(\langle f_1, \eta_{t_1}^{(N)}\rangle, \ldots, \langle f_k, \eta_{t_k}^{(N)}\rangle\right) \xrightarrow{d} \left(\langle f_1, \eta_{t_1}\rangle, \ldots, \langle f_k, \eta_{t_k}\rangle\right), \tag{2.12}$$

where the object on the right is a k-dimensional Gaussian random variable with covariance matrix

$$\Sigma_{\underline{t}}^2 := (\mathrm{Cov}_{Q_0}(\langle f_j, F_{t_j}\rangle, \langle f_l, F_{t_l}\rangle))_{j,l=1}^k.$$

Proof. Since the pre-limiting object on the left hand side of (2.12) is equal to

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\langle f_1, F_{i,t_1}\rangle, \ldots, \langle f_k, F_{i,t_k}\rangle\right) - \sqrt{N}\left(\langle f_1, \mu_{t_1}\rangle, \ldots, \langle f_k, \mu_{t_k}\rangle\right)$$

then an application of the classic CLT is all we need; provided the moments,

$$\mathbb{E}_{Q_0}\left[\langle f_j, F_{t_j}\rangle\right]$$

and

$$\mathbb{E}_{Q_0}\left[\langle f_j, F_{t_j}\rangle\langle f_l, F_{t_l}\rangle\right]$$

are finite for $1 \leq j, l \leq k$. Thankfully, the fact that these are integrals of continuous functions against probability measures over a compact space ensures this fact. $\square$

The other requirement of [34, Theorem 5.3] is tightness of the collection of $\mathbb{R}$-valued processes

$$\left(\langle f, \eta_t^{(N)} \rangle\right)_{t \geq 0}, \quad f \in C^\infty([0,1], \mathbb{R}).$$

Similarly to Section 2.2, we show this via Kolmogorov continuity; i.e. we show that there exists constants $C_T > 0$ such that

$$\mathbb{E}\left[\left|\langle f, \eta_t^{(N)} \rangle - \langle f, \eta_s^{(N)} \rangle\right|^4\right] \leq C_T |t-s|^2, \tag{2.13}$$

for all $N \in \mathbb{N}$ with $s, t \in [0, T]$, $T = 1, 2, 3, \ldots$. Before proceeding to the analysis of conditions (2.9) and (2.13) we let $X_{i,t-s} := \langle f, F_{i,t} \rangle - \langle f, F_{i,s} \rangle$ and bound the left hand side of (2.13) above by the following:

$$
\begin{aligned}
&\frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^{N} X_{i,t-s} - \mathbb{E}_{Q_0}\left[X_{i,t-s}\right]\right)^4\right] \\
&= \frac{1}{N}\Big(\mathbb{E}_{Q_0}\left[X_{1,t-s}^4\right] - 4\mathbb{E}_{Q_0}\left[X_{1,t-s}^3\right]\mathbb{E}_{Q_0}\left[X_{1,t-s}\right] \\
&\quad +6\mathbb{E}_{Q_0}\left[X_{1,t-s}^2\right]\mathbb{E}_{Q_0}\left[X_{1,t-s}\right]^2 - 3\mathbb{E}_{Q_0}\left[X_{1,t-s}\right]^4\Big) \\
&\quad + 3\frac{N_{[2]}}{N^2}\Big(\mathbb{E}_{Q_0}\left[X_{1,t-s}^2\right]^2 - 2\mathbb{E}_{Q_0}\left[X_{1,t-s}^2\right]\mathbb{E}_{Q_0}\left[X_{1,t-s}\right]^2 \\
&\qquad\qquad +\mathbb{E}_{Q_0}\left[X_{1,t-s}\right]^4\Big) \\
&\leq \mathbb{E}_{Q_0}\left[X_{1,t-s}^4\right] + 4\left|\mathbb{E}_{Q_0}\left[X_{1,t-s}^3\right]\mathbb{E}_{Q_0}\left[X_{1,t-s}\right]\right| \\
&\quad + 6\mathbb{E}_{Q_0}\left[X_{1,t-s}^2\right]\mathbb{E}_{Q_0}\left[X_{1,t-s}\right]^2 + 3\mathbb{E}_{Q_0}\left[X_{1,t-s}^2\right]^2 \\
&\quad + 3\mathbb{E}_{Q_0}\left[X_{1,t-s}\right]^4. \qquad (2.14)
\end{aligned}
$$

As with (2.10), we use the i.i.d nature of the summands to simplify the expression along with bounding fractions involving N above by 1. More details on the cancellations that occur above can be found in Appendix B. As with (2.10), this upper bound allows us to analyse the small time behaviour of the CLT object by considering what happens to a single FV process between times s and t.

2.4 The Transition Function and Tightness

With these simplifications in place we are ready to tackle the regularity in time of a continuous function integrated against a single FV process. In order to analyse the moments in (2.10) and (2.14) we will utilise the process's transition function. First discovered in 1993 by Ethier and Griffiths

[35], it has the following form:

$$P(t,\mu,d\nu) = d_0^\theta(t)\Pi_{\theta,\nu_0}(d\nu) + \sum_{n=1}^{\infty} d_n^\theta(t)\int_{[0,1]^n}\mu^n(dx_1\times\cdots\times dx_n) \Pi_{n+\theta,(n+\theta)^{-1}\{n\zeta_n(x_1,\ldots,x_n)+\theta\nu_0\}}(d\nu), \tag{2.15}$$

where μ is the starting measure, $\Pi_{\lambda,\nu}$ is the distribution of a Dirichlet Process with parameters $\lambda > 0$ and $\nu \in \mathcal{P}([0,1])$,

$$\zeta_n(x_1,\ldots,x_n) := \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \in \mathcal{P}([0,1])$$

is the empirical distribution of the points $x_1,\ldots,x_n$ and $d_n^\theta(t)$, as in Section 2.2, are the transition probabilities of the Kingman coalescent with mutation.

With the form of this transition function, the duality with the coalescent is clear. Heuristically what happens is that if one wishes to determine the state of a FV process at some time $t > 0$ you effectively follow this simple procedure:

- Run a Kingman coalescent, starting from infinity, for $t > 0$ units of time.
- The number of ancestors that survive, N_t, all have types drawn from the initial distribution $\mu \in \mathcal{P}([0,1])$.
- These ancestors then influence the present day population by having their types incorporated into the parameters of the Dirichlet Process from which we

draw the FV process.

- Types in the FV process at time t can now be one of the N_t original types from μ or are new types, drawn from the distribution ν_0.
- The size of the θ partly determines how prevalent the old types are; a higher total mutation parameter means the types in the new population are more likely to be drawn from ν_0 than $\zeta_n(x_1, \ldots, x_n)$.

This Kingman coalescent-induced mixture of Dirichlet Processes is precisely the reason for the form of (2.6) where the expectation of F_t is a mixture of the expected starting and base measures μ_0 and ν_0 respectively. For a more precise description of higher moments of the FV process at a single time we have the following formula from [35] which concerns moments of the form

$$\mathbb{E}_{Q_0}\left[\langle f_1, F_t\rangle \ldots \langle f_m, F_t\rangle\right]$$

for some $f_1, \dots, f_m \in C([0,1], \mathbb{R})$:

$$\mathbb{E}_{Q_0}\left[\langle f_1, F_t\rangle \dots \langle f_m, F_t\rangle\right] = \sum_{n=0}^{\infty} d_n^{\theta}(t) \sum_{M \subset \{1,\dots,m\}} \frac{1}{(n+\theta)_{(m)}}$$
$$\times \left\{ \sum_{k=1}^{|M|} n_{[k]} \sum_{\beta \in \pi(M,k)} |\beta_1|! \cdots |\beta_k|! \prod_{j=1}^{k} \left\langle \prod_{i \in \beta_j} f_i, \mu_0 \right\rangle \right\}$$
$$\times \left\{ \sum_{l=1}^{|M^c|} \sum_{\gamma \in \pi(M^c,l)} (|\gamma_1|-1)! \cdots (|\gamma_l|-1)! \theta^l \prod_{j=1}^{l} \left\langle \prod_{i \in \gamma_j} f_i, \nu_0 \right\rangle \right\}, \tag{2.16}$$

where $\pi(A, i)$ is the set of all possible partitions of A that have i elements and an empty sum is considered to be equal to 1.

This formula, with $m = 1$ and f ranging over all continuous functions, is how we derive equation (2.5). If we consider m up to 4 and let all f_i be the same function f we can analyse higher moments of the FV process and show the following inequalities:

Proposition 2.6. *Let $0 \le s < t$, then there exist constants* *depending on f and θ such that*

$$\mathbb{E}_{Q_0}\left[\langle f, F_t\rangle - \langle f, F_s\rangle\right] \leq C_{f,\theta}^{(1,1)} \mathbb{E}_\infty\left[\frac{1}{N_{t-s}+\theta}\right] \tag{2.17}$$

$$\mathbb{E}_{Q_0}\left[\left(\langle f, F_t\rangle - \langle f, F_s\rangle\right)^2\right] \leq C_{f,\theta}^{(2,1)} \mathbb{E}_\infty\left[\frac{1}{N_{t-s}+\theta}\right] + C_{f,\theta}^{(2,2)} \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)^2}\right] \tag{2.18}$$

$$\mathbb{E}_{Q_0}\left[\left(\langle f, F_t\rangle - \langle f, F_s\rangle\right)^3\right] \leq C_{f,\theta}^{(3,1)} \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)^2}\right] + C_{f,\theta}^{(3,2)} \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)^3}\right] \tag{2.19}$$

$$\mathbb{E}_{Q_0}\left[\left(\langle f, F_t\rangle - \langle f, F_s\rangle\right)^4\right] \leq C_{f,\theta}^{(4,1)} \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)^2}\right] + C_{f,\theta}^{(4,2)} \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)^3}\right] + C_{f,\theta}^{(4,3)} \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)^4}\right] \tag{2.20}$$

Proof. Considering the first moment, we start by conditioning on the event $\{F_s = \mu\}$, $\mu \in \mathcal{P}([0,1])$, and use (2.15) with $m = 1$ and $f_1 = f$, along with the Markov property to obtain the following:

$$\begin{aligned}
&\mathbb{E}_{Q_0}\left[\langle f, F_t\rangle - \langle f, F_s\rangle \middle| F_s = \mu\right] \\
&= \mathbb{E}\left[\langle f, F_{t-s}\rangle - \langle f, \mu\rangle \middle| F_0 = \mu\right] \\
&= \sum_{n=0}^{\infty} \frac{d_n^\theta(t-s)}{n+\theta}\left[n\langle f, \mu\rangle + \theta\langle f, \nu_0\rangle - (n+\theta)\langle f, \mu\rangle\right] \\
&= \sum_{n=0}^{\infty} \frac{\theta d_n^\theta(t-s)}{n+\theta}\left[\langle f, \nu_0\rangle - \langle f, \mu\rangle\right] \\
&= \theta(\langle f, \nu_0\rangle - \langle f, \mu\rangle)\mathbb{E}_\infty\left[\frac{1}{N_{t-s}+\theta}\right]. \qquad (2.21)
\end{aligned}$$

Since f is a continuous function on a compact space, its integral against any probability measure is bounded by $\sup\{f(x), x \in [0,1]\} < \infty$. Thus we can bound the constant in front of the expectation on the right hand side of (2.21) by

$$C_{f,\theta}^{(1,1)} := 2\theta \sup(|f|(x)).$$

Note that this bound is independent of μ and so if we take expectations on both sides and apply the tower law on the left we obtain (2.17).

Applying the same tactic to higher moments, again conditioned on the event $\{F_s = \mu\}$, we get similar (though more complicated) expressions with different coalescent moments on the right hand side. The constants in front of these moments are mixtures of integrals of powers of f against μ and ν_0. These can again be bounded above independently of the measures, leading to the constants

$C^{(i,j)}_{f,\theta}$ for $i = 2,3,4$, $j = 1,2$, and $(i,j) = (4,3)$. The coalescent moments there are not yet in the correct form though. For the second moment we have

$$\begin{aligned}\mathbb{E}_{\mu_0}&\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^2 \,|F_s = \mu\right] \\ &\leq C^{(2,1)}_{f,\theta}\mathbb{E}_\infty\left[\frac{N_{t-s}}{(N_{t-s}+\theta)_{(2)}}\right] \\ &\quad + C^{(2,2)}_{f,\theta}\mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)_{(2)}}\right],\end{aligned}$$

but these are trivially bounded above by (2.18). The third and fourth moments have similar expressions, and the exact same tactics are used to produce the formulas (2.19) and (2.20). For the complete calculations, see Appendix C. □

We now have upper bounds on the Kolmogorov continuity conditions (2.9) and (2.13) via the simplifications (2.10) and (2.14) in terms of inverse moments of the Kingman coalescent. The next step is to show that these expressions behave asymptotically as we expect them to. As we know from Chapter 1, $N_t \sim 2/t$ as $t \to 0$ and indeed

$$\mathbb{E}_\infty\left[N_t^k\right] \sim 2^k/t^k$$

as $t \to 0$ for all $k \in \mathbb{N}$ (Proposition 1.9). We would like this to also be true for $k = -1,-2,-3,-4$; more specifically we need

$$\mathbb{E}_\infty\left[(N_t+\theta)^k\right] \sim 2^k/t^k \tag{2.22}$$

as $t \to 0$ for $k = -1,-2,-3,-4$. Once this is in place we can then substitute the asymptotic expressions for these coalescent moments into the right hand sides of (2.17), (2.18), (2.19) and (2.20) in combination with (2.10) and (2.14) to obtain the required Kolmogorov continuity. The following proposition proves a more general form of (2.22), taking advantage of an expression for $\frac{d}{dt}d_n^\theta(t)$ from [35] and L'Hôpital's rule:

Proposition 2.7. *Let $(N_t)_{t\geq 0}$ be the block-counting process of the Kingman coalescent with parent-independent mutation, rate $\theta > 0$. Let $k \in \mathbb{N}$ and define a_n as*

$$a_n = n^{-k} + \sum_{j=1}^{\infty} A_{j,k} n^{-k-j}, \tag{2.23}$$

where $A_{j,k}$ is a polynomial in k of order at most j. Then,

$$\lim_{t\to 0} \frac{\mathbb{E}_\infty\left[a_{N_t}\right]}{t^k/2^k} = 1. \tag{2.24}$$

Proof. We first express the moment in (2.24) as follows:

$$\mathbb{E}_\infty\left[a_{N_t}\right] = \sum_{n=0}^{\infty} a_n d_n^\theta(t). \tag{2.25}$$

In order to analyse this moment close to $t = 0$ we will utilise the above representation along with several parts of [35, Lemma 2.5]. In the following we fix a $T > 0$ and only consider $t \leq T$, since we are considering a limit as $t \to 0$ this T is arbitrary.

Firstly, the fact that $a_n \leq 1$ for sufficiently large n and $(d_n^\theta(t))_{n\in\mathbb{N}}$ is a probability mass function tells us that the series in (2.25) is convergent for all $t \geq 0$ and so the moment is well-defined. Part (ii) of [35, Lemma 2.5] then tells us that

$$\lim_{t \to 0} \mathbb{E}_\infty \left[a_{N_t} \right] = \lim_{n \to \infty} a_n = 0.$$

Thus, in order to analyse the limit in (2.24) we will look to apply L'Hôpital's rule several times. In doing this we would like to differentiate (2.25) and exchange this with the infinite sum, differentiating term by term. In order for this to be permissable we must (along with the convergence of the original series) have uniform - in t - convergence of the partial sums of the derivatives

$$\sum_{n=0}^{N} a_n \frac{d}{dt} d_n^\theta(t).$$

Part (iv) of [35, Lemma 2.5] tells us that this is equal to

$$\begin{aligned} \sum_{n=0}^{\infty} a_n \frac{d}{dt} d_n^\theta(t) &= \sum_{n=0}^{\infty} a_n (\lambda_{n+1} d_{n+1}^\theta(t) - \lambda_n d_n^\theta(t)) \\ &= \sum_{n=1}^{\infty} \lambda_n (a_{n-1} - a_n) d_n^\theta(t), \end{aligned}$$

where $\lambda_n = n(n-1+\theta)/2$ is the death rate of N_t when it is equal to n. If we now consider the following upper bound

$$\begin{aligned} &\sup_{t \geq 0} \left| \sum_{n=N+1}^{\infty} \lambda_n (a_{n-1} - a_n) d_n^\theta(t) \right| \\ &\leq \sup_{t \geq 0} \left(\sum_{n=N+1}^{\infty} (\lambda_n (a_{n-1} - a_n))^2 \right)^{1/2} \left(\sum_{n=N+1}^{\infty} \left(d_n^\theta(t) \right)^2 \right)^{1/2}, \end{aligned}$$

we see that on the right hand side, if $k \geq 2$, the second term is bounded by 1 and the first term tends to zero as N tends to infinity, since it is the tail of a convergent series (we will see later that $\lambda_n(a_{n-1} - a_n) \sim n^{-k+1}$ as $n \to \infty$). If $k = 1$ then $\lambda_n(a_{n-1} - a_n) = O(1)$ as $n \to \infty$ and so there exists a constant $C > 0$ such that

$$\sum_{n=1}^{\infty} \lambda_n(a_{n-1} - a_n) d_n^{\theta}(t) \leq \sum_{n=1}^{\infty} C d_n^{\theta}(t) = C,$$

and thus we have uniform convergence, in t, of the sequence of partial sums of the derivatives.

If we want to characterise the above as $t \to 0$ we need to consider what happens to $\lambda_n(a_{n-1} - a_n)$ as $n \to \infty$. First considering $a_{n-1} - a_n$ we see that

$$\begin{aligned}
&a_{n-1} - a_n \\
&\quad =(n-1)^{-k} - n^{-k} + \sum_{j=1}^{\infty} A_{j,k} \left[(n-1)^{-k-j} - n^{-k-j} \right] \\
&\quad =kn^{-k-1} + \sum_{j=2}^{\infty} \binom{k}{j} n^{-k-j} + \sum_{j=1}^{\infty} A_{j,k} \sum_{i=1}^{\infty} \binom{k+j}{i} n^{-k-j-i} \\
&\quad =kn^{k-1} + \left[\binom{k}{2} A_{1,k}(k+1) \right] n^{-k-2} \\
&\quad\quad + \left[\binom{k}{3} + A_{2,k}(k+2) + A_{1,k} \binom{k+1}{2} \right] n^{-k-3} + \dots \\
&\quad =kn^{-k-1} + \sum_{j=2}^{\infty} B_{j,k} n^{-k-j},
\end{aligned}$$

where we define

$$\binom{k}{j} := \frac{k(k+1)\dots(k+j-1)}{j!} = \frac{k_{(j)}}{j!}$$

for all $k, j \in \mathbb{N}$ and

$$B_{j,k} := \binom{k}{j} + \sum_{i=1}^{j-1} A_{i,k} \binom{k+i}{j-i}.$$

Note that $B_{j,k}$ is a polynomial in k of order at most j.

Now, $\lambda_n = n^2/2 + n(\theta-1)/2$ and so

$$\begin{aligned}
\lambda_n(a_{n-1} - a_n) =& \frac{k}{2} n^{-k+1} + \frac{k(\theta-1)}{2} n^{-k} \\
&+ \sum_{j=2}^{\infty} \frac{B_{j,k}}{2} \left[n^{-k-j+2} + (\theta-1) n^{-k-j+1} \right] \\
=& \frac{k}{2} n^{-k+1} + \sum_{j=2}^{\infty} C_{j,k} n^{-k-j+2} =: b_n,
\end{aligned}$$

where

$$C_{2,k} := \frac{k(\theta-1)}{2} + \frac{B_{2,k}}{2}$$

and

$$C_{j,k} := \frac{B_{j,k}}{2} + \frac{B_{j-1,k}(\theta-1)}{2}$$

for $j \geq 3$.

Having now applied L'Hôpital's rule once we have that

$$\lim_{t \to 0} \frac{\sum_{n=0}^{\infty} a_n d_n^\theta(t)}{t^k / 2^k} = \lim_{t \to 0} \frac{\sum_{n=0}^{\infty} b_n d_n^\theta(t)}{k t^{k-1} / 2^k}.$$

where

$$b_n = \frac{k}{2} n^{-k+1} + \sum_{j=2}^{\infty} C_{j,k} n^{-k-j+2},$$

and again $C_{j,k}$ is a polynomial in k of order at most j.

We look to repeat this procedure $k-1$ more times in order to analyse our limit. Though the power of n has shifted by one our justification for moving the differential inside the infinite sum is the same as before. Due to the nature of the a_n, repeating this procedure $k-1$ more times results in

$$\frac{d^k}{dt^k} \sum_{n=0}^{\infty} a_n d_n^{\theta}(t) = \sum_{n=0}^{\infty} c_n d_n^{\theta}(t)$$

where

$$c_n = \frac{k!}{2^k} + \sum_{j=k+1}^{\infty} D_{j,k} n^{-j+k}$$

and $D_{j,k}$ is a polynomial in k of order at most j.

Finally, if we consider the left hand side of (2.24) and apply L'Hôpital's rule k times we have that

$$\begin{aligned}
\lim_{t \to 0} \frac{\mathbb{E}_{\infty}\left[a_{N_t}\right]}{t^k/2^k} &= \lim_{t \to 0} \frac{\sum_{n=0}^{\infty} c_n d_n^{\theta}(t)}{k!/2^k}, \\
&= \frac{2^k}{k!} \lim_{t \to 0} \sum_{n=0}^{\infty} c_n d_n^{\theta}(t), \\
&= \frac{2^k}{k!} \lim_{n \to \infty} c_n = 1.
\end{aligned}$$

□

Since
$$(n+\theta)^{-k} = n^{-k} + \sum_{j=1}^{\infty} (-1)^j \binom{k}{j} \theta^j n^{-k-j}$$
the moments in (2.17) to (2.20) satisfy Proposition 2.7 and become:

$$\mathbb{E}_{\mu_0}\left[\langle f, F_t\rangle - \langle f, F_s\rangle\right] \leq C_{f,\theta}^{(1,1)}(t-s) + o(t-s) \tag{2.26}$$
$$\mathbb{E}_{\mu_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^2\right] \leq C_{f,\theta}^{(2,1)}(t-s) + C_{f,\theta}^{(2,2)}(t-s)^2 + o(t-s) \tag{2.27}$$
$$\mathbb{E}_{\mu_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^3\right] \leq C_{f,\theta}^{(3,1)}(t-s)^2 + C_{f,\theta}^{(3,2)}(t-s)^3 + o((t-s)^2) \tag{2.28}$$
$$\mathbb{E}_{\mu_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^4\right] \leq C_{f,\theta}^{(4,1)}(t-s)^2 + C_{f,\theta}^{(4,2)}(t-s)^3 + C_{f,\theta}^{(4,3)}(t-s)^4 + o((t-s)^2) \tag{2.29}$$

as $(t-s) \to 0$. Letting $\epsilon > 0$, we can choose $t-s$ small enough so that the $o(t-s)$ and $o((t-s)^2)$ terms above are smaller than $\epsilon(t-s)$ and $\epsilon(t-s)^2$ respectively. Reminding ourselves of (2.10) we have:

$$\begin{aligned}
&\frac{1}{N^4}\mathbb{E}\left[\left(\sum_{i=1}^{N}\langle f, F_{i,t}\rangle - \langle f, F_{i,s}\rangle\right)^4\right] \\
&\quad \le \mathbb{E}_{Q_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^4\right] \\
&\qquad + 4\mathbb{E}_{Q_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^3\right]\mathbb{E}_{Q_0}\left[\langle f, F_t\rangle - \langle f, F_s\rangle\right] \\
&\qquad + 3\mathbb{E}_{Q_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^2\right]^2 \\
&\qquad + 6\mathbb{E}_{Q_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^2\right]\mathbb{E}_{Q_0}\left[\langle f, F_t\rangle - \langle f, F_s\rangle\right]^2 \\
&\qquad + \mathbb{E}_{Q_0}\left[\langle f, F_t\rangle - \langle f, F_s\rangle\right]^4 .
\end{aligned}$$

Now, substituting (2.26) to (2.29) into the above yields (2.9), with C_T equal to a polynomial in ϵ and $C_{f,\theta}^{i,j}$, $i = 1, 2, 3, 4$, $j = 1, 2, 3$. A similar procedure ensures that substituting (2.26) to (2.29) into (2.14) yields (2.13).

For $t - s$ large, say $t - s \ge \epsilon > 0$, we can bound N_{t-s} below by zero and attain the bound

$$\mathbb{E}_\infty\left[\frac{1}{(N_{t-s} + \theta)^k}\right] \le \frac{1}{\theta^k}$$

for $k = 1, 2, 3, 4$. This can then be used in (2.17) to (2.20) to bound these expressions independently of N_{t-s}. Then, using these in (2.10) and (2.14) bound them above by constants, say, B_1 and B_2. Choosing C_T to be larger than both B_1/ϵ^2 and B_2/ϵ^2 then gives us (2.9) and (2.13).

All of this work to obtain tightness and convergence of finite dimensional distributions of these processes constitutes the proof of Theorem 2.1, which we restate here for completeness:

Theorem 2.1. *Let $(F_i)_{i\in\mathbb{N}} \subset C([0,\infty),\mathcal{P}([0,1]))$ be an i.i.d. collection of FV processes, all started at independent draws from $Q_0 \in \mathcal{P}(\mathcal{P}([0,1]))$, with parent-independent mutation, parameters $(\theta,\nu_0) \in [0,\infty) \times \mathcal{P}([0,1])$. Then the processes $S^{(N)} \in C([0,\infty),\mathcal{P}(0,1))$ and $\eta^{(N)} \in C([0,\infty),\mathcal{D}([0,1]))$ defined by* (2.1) *and* (2.2) *respectively converge in distribution in their respective spaces to μ and η which are characterised via Propositions 2.3 and 2.5.*

2.5 Continuity

What we have shown up to this point is the existence of and convergence to two stochastic processes that are meant to characterise the "mean" genotype of a population and the fluctuations of that genotype around its mean via a Gaussian Schwartz distribution-valued process. In this section we investigate the continuity of the latter process.

To investigate the continuity of η we need to decide upon a topology on $\mathcal{D}([0,1])$. The most natural choice is the strong dual topology induced by the space's duality with $C^\infty([0,1])$. Fortunately, for a sequence to converge in the strong dual topology it is necessary and sufficient for it to converge in the weak* topology [32, p351-359]. As far as Schwartz distributions are concerned this is simply pointwise convergence. Using the Kolmogorov continuity theorem (also known as the Kolmogorov-Chentsnov theorem) and the implications above, we can establish the continuity of η:

Proposition 2.8. *$(\eta_t)_{t\geq 0}$ has a modification that is sample continuous, taking values in $\mathcal{D}([0,1])$ endowed with the strong dual topology.*

Proof. As established above, a Schwartz distribution μ_n converges to μ in the strong dual topology if

$$\langle f, \mu_n \rangle \to \langle f, \mu \rangle$$

as $n \to \infty$, for all $f \in C^\infty([0,1],\mathbb{R})$. A $\mathcal{D}([0,1])$-valued process $(\mu_t)_{t\geq 0}$ then has continuous sample paths if, for almost all $\omega \in \Omega$, $(\mu_t(\omega))_{t\geq 0}$ is a continuous process; i.e.

$$(\langle f, \mu_t \rangle(\omega))_{t\geq 0}$$

is continuous in t for all $f \in C^\infty([0,1],\mathbb{R})$. Thus we wish to show the above for a $\mathcal{D}([0,1])$-valued process $\tilde{\eta}$ that is a modification of η; i.e.

$$\forall t \geq 0, \mathbb{P}(\tilde{\eta}_t = \eta_t) = 1.$$

We proceed then by defining, for $f \in C^\infty([0,1],\mathbb{R})$,

$$Y_t^{(N),f} := \langle f, \eta_t^{(N)} \rangle,$$

and

$$Y_t^f := \langle f, \eta_t \rangle.$$

As established in Section 2.4, for any $T > 0$ there exists a constant $C > 0$ such that

$$\mathbb{E}\left[\left(Y_t^{(N),f} - Y_s^{(N),f}\right)^4\right] \leq C(t-s)^2, \tag{2.30}$$

for $0 < t, s \leq T$. Since the bound (2.30) is independent of N, we let $N \to \infty$ and obtain

$$\mathbb{E}\left[\left(Y_t^f - Y_s^f\right)^4\right] \leq C(t-s)^2.$$

By the Kolmogorov continuity theorem we then have that there exists a continuous modification of Y^f, denoted $\tilde{Y}^f$, where the finite-dimensional distributions of the two processes are equal.

The collection $(\tilde{Y}_t^f)_{f \in C^\infty([0,1],\mathbb{R})}$ for a single time $t \geq 0$ defines a random linear operator $\tilde{\eta}_t$ on the space $C([0,1],\mathbb{R})$ by

$$\tilde{\eta}_t(f) := \tilde{Y}_t^f,$$

provided that $\tilde{Y}_t^{af+bg} = a\tilde{Y}_t^f + b\tilde{Y}_t^g$ for constants $a, b \in \mathbb{R}$. This additivity is assured by the fact that $Y_t^{af+bg} = aY_t^f + bY_t^g$ and that $\tilde{Y}$ is a modification of Y. If we consider Proposition 21.1 [32] (a statement of which can be found in Section 2.3) then we see that $\tilde{\eta}_t$ satisfies 2, since if the collection $\{f_n\}_{n \in \mathbb{N}}$ converge uniformly to zero as $n \to \infty$, along with all of their derivatives, then $\tilde{\eta}_t(f_n) \to 0$ since

$$\tilde{\eta}_t(f_n) \sim \mathcal{N}(0, \mathrm{Var}_{Q_0}(\langle f_n, F_t \rangle))$$

converges to zero almost surely as $n \to \infty$.

If we now let $\{f_n\}_{n\in\mathbb{N}} \subset C^\infty([0,1],\mathbb{R})$ be the collection of polynomials with rational coefficients, this class is rich enough so as to be separating since, if $T, \tilde{T} \in \mathcal{D}([0,1])$ and

$$T(f_n) = \tilde{T}(f_n)$$

for all $n \in \mathbb{N}$ then for a general $f \in C^\infty([0,1],\mathbb{R})$, there exists a subsequence $(f_{n_k})_{k\in\mathbb{N}}$ that converges to f uniformly as $k \to \infty$ and so

$$T(f) = \lim_{k\to\infty} T(f_{n_k}) = \lim_{k\to\infty} \tilde{T}(f_{n_k}) = \tilde{T}(f),$$

and $T = \tilde{T}$.

In order to establish sample-continuity of $\tilde{\eta}$ it is sufficient to find a set $A \subset \Omega$ such that, for all $f \in C^\infty([0,1],\mathbb{R})$, $(\langle f, \tilde{\eta}_t\rangle(\omega))_{t\geq 0}$ is continuous in t for all $\omega \in \Omega \setminus A$ with $\mathbb{P}(A) = 0$.

For each f_n we let A_n be a set such that $(\tilde{Y}_t^{f_n}(\omega))_{t\geq 0} = (\langle f_n, \tilde{\eta}_t\rangle(\omega))_{t\geq 0}$ is continuous in t for every $\omega \in \Omega \setminus A_n$, where $\mathbb{P}(A_n) = 0$. The existence of this set comes from the sample-continuity of $\tilde{Y}^{f_n}$. Our set A will then be the union of all of these, $\cup_{n\in\mathbb{N}} A_n$.

To establish continuity of $\langle f, \tilde{\eta}_t\rangle$ for a general $f \in C^\infty([0,1],\mathbb{R})$, let $\epsilon > 0$ and f_n be such that

$||f - f_n||_\infty < \epsilon/3$. Furthermore, let $\delta > 0$ be such that

$$|\langle f_n, \tilde{\eta}_t \rangle - \langle f_n, \tilde{\eta}_s \rangle| < \frac{\epsilon}{3}$$

for all $|t - s| < \delta$. Then we have that, for $|t - s| < \delta$

$$\begin{aligned}
|\langle f, \tilde{\eta}_t \rangle - \langle f, \tilde{\eta}_s \rangle| &\leq |\langle f, \tilde{\eta}_t \rangle - \langle f_n, \tilde{\eta}_t \rangle| + |\langle f_n, \tilde{\eta}_t \rangle - \langle f_n, \tilde{\eta}_s \rangle| \\
&\quad + |\langle f_n, \tilde{\eta}_s \rangle - \langle f, \tilde{\eta}_s \rangle| \\
&\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

Thus, $(\langle f, \tilde{\eta}_t \rangle(\omega))_{t \geq 0}$ is continuous in t for every $\omega \in \Omega \setminus A$. This then gives us continuity in t of $(\tilde{\eta}_t(\omega))_{t \geq 0}$ for every $\omega \in \Omega \setminus A$.

Finally, we fix $t \geq 0$ and let the set B_n be such that $\mathbb{P}(B_n) = 0$ and

$$\langle f_n, \eta_t \rangle(\omega) = Y_t^{f_n}(\omega) = \tilde{Y}_t^{f_n}(\omega) = \langle f_n, \tilde{\eta}_t \rangle(\omega)$$

for all $\omega \in \Omega \setminus B_n$. Taking the union of these sets to be $B = \cup_{n \in \mathbb{N}} B_n$, we have that

$$\mathbb{P}(\langle f_n, \tilde{\eta}_t \rangle = \langle f_n, \eta_t \rangle, \forall n \in \mathbb{N}) = 1 - \mathbb{P}(B) = 1$$

and so since the collection $\{f_n\}_{n \in \mathbb{N}}$ is separating we have that $\tilde{\eta}_t$ is equal to η_t almost surely. $\square$

2.6 Different Starting Measures

The last section of this chapter will focus on the setting in which each of the independent FV processes start from different starting measures.

The set up is similar to before; we have an independent collection of FV processes $((F_{i,t})_{t\geq 0})_{i\in\mathbb{N}}$ which now each start at probability measures $F_{i,0}$ each distributed according to $Q_{i,0} \in \mathcal{P}(\mathcal{P}([0,1]))$, with $\mathbb{E}_{Q_{i,0}}[F_{i,0}] = \mu_{i,0}$, and consider the collections $(\langle f, F_{i,t}\rangle)_{t\geq 0}$ for $f \in C([0,1],\mathbb{R})$.

The LLN case is relatively simple: we define

$$\left(\bar{S}_t^{(N)}\right)_{t\geq 0} := \left(\frac{1}{N}\sum_{i=1}^{N} F_{i,t}\right)_{t\geq 0}, \tag{2.31}$$

and require that

$$\frac{1}{N}\sum_{i=1}^{N} F_{i,0} \to \overline{F}_0, \tag{2.32}$$

weakly as $N \to \infty$ for some $\overline{F}_0 \sim \overline{Q}_0 \in \mathcal{P}(\mathcal{P}([0,1]))$. Note that this also implies that the means converge

$$\frac{1}{N}\sum_{i=1}^{N} \mu_{i,0} \to \bar{\mu}_0,$$

where $\bar{\mu}_0 = \mathbb{E}_{\overline{Q}_0}[F_0]$.

With this we state the analogue of Proposition 2.3 in this case:

Proposition 2.9. *Let $k \in \mathbb{N}$ and $0 \leq t_1 \leq t_2 \leq \cdots \leq t_k$. Then*

$$(\bar{S}_{t_1}^{(N)}, \ldots, \bar{S}_{t_k}^{(N)}) \xrightarrow{d} (\bar{\mu}_{t_1}, ..., \bar{\mu}_{t_k})$$

as $N \to \infty$, where

$$\bar{\mu}_t := \mathbb{E}_{\overline{Q}_0}\left[F_t\right] = \sum_{n=0}^{\infty} \frac{d_n^{\theta}(t)}{n+\theta}\left\{n\bar{\mu}_0 + \theta\nu_0\right\}.$$

Proof. As with the proof of Proposition 2.3, we start by letting $f_1, \dots, f_k \in C([0,1], \mathbb{R})$ and consider the object

$$\langle \underline{f}, \bar{S}_{\underline{t}}^{(N)} \rangle := \left(\langle f_1, \bar{S}_{t_1}^{(N)} \rangle, \dots, \langle f_k, \bar{S}_{t_k}^{(N)} \rangle \right). \tag{2.33}$$

as $N \to \infty$, checking that this converges to the corresponding object for $(\bar{\mu}_t)_{t\geq 0}$. For this we appeal to Kolmogorov's strong law [36]; a slightly modified LLN in which the sequence of independent random variables are not identically distributed. The result states that the sequence of differences between the sample mean and the expected sample mean tend to zero. In our case this means that

$$\langle \underline{f}, \bar{S}_{\underline{t}}^{(N)} \rangle - \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{Q_{i,0}}\left[(\langle f_1, F_{i,t_1} \rangle, \dots, \langle f_k, F_{i,t_k} \rangle) \right] \to 0,$$

as $N \to \infty$. This convergence holds if each $\langle f_j, F_{i,t_j} \rangle$ has finite second moment and

$$\sum_{i=1}^{\infty} \frac{1}{i^2} \mathrm{Cov}_{Q_{i,0}}(\langle f_j, F_{i,t_j} \rangle, \langle f_l, F_{i,t_l} \rangle) < \infty,$$

for each $1 \leq j, l \leq k$. The above holds since, as used several times in this chapter, these moments can be bounded uniformly in i using the suprema of the functions $f_1^2, \dots, f_k^2$, hence leaving the above sum finite.

Let us now consider the expected sample mean for a single function f and time $t \geq 0$:

$$\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{Q_{i,0}}\left[\langle f, F_{i,t}\rangle\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{\infty} \frac{d_n^\theta(t)}{n+\theta}\left\{n\langle f, \mu_{i,0}\rangle + \theta\langle f, \nu_0\rangle\right\} \\
&= \sum_{n=0}^{\infty} \frac{d_n^\theta(t)}{n+\theta}\left\{n\frac{1}{N}\sum_{i=1}^{N}\langle f, \mu_{i,0}\rangle + \theta\langle f, \nu_0\rangle\right\} \\
&\overset{N\to\infty}{\to} \sum_{n=0}^{\infty} \frac{d_n^\theta(t)}{n+\theta}\left\{n\langle f, \bar{\mu}_0\rangle + \theta\langle f, \nu_0\rangle\right\}, \\
&= \mathbb{E}_{\overline{Q}_0}\left[\langle f, F_t\rangle\right].
\end{aligned}$$

Considering a similar limit for the components of (2.33), ranging over all collections $f_1, \ldots, f_k \in C([0,1], \mathbb{R})$ we obtain our result. □

In order to investigate the equivalent of $\eta^{(N)}$ when we have different starting measures we again have to lift the CLT object into distribution space and appeal to Theorem 5.3 of [34] to obtain our convergence. However, since the starting measures are different we have to consider a Lindeberg-Feller CLT. The formulation of this that we will appeal to can be found in [37, Proposition 2.27]. Our pre-limiting object is the following:

$$\bar{\eta}_t^{(N)} := \frac{1}{\sqrt{N}}\sum_{i=1}^{N} F_{i,t} - \mathbb{E}_{Q_{i,0}}\left[F_{i,t}\right].$$

In order to prove the convergence of finite dimensional distributions we let $k \in \mathbb{N}$, $0 \leq t_1 < \cdots < t_k \leq 1$, $f_1, \ldots, f_k \in C([0,1], \mathbb{R})$ and define

$$\begin{aligned}
\langle \underline{f}, F_{i,\underline{t}} \rangle &:= \left(\langle f_1, F_{i,t_1} \rangle, \ldots, \langle f_k, F_{i,t_k} \rangle \right), \\
\langle \underline{f}, \mu_{i,\underline{t}} \rangle &:= \mathbb{E}_{Q_{i,0}} \left[\langle \underline{f}, F_{i,\underline{t}} \rangle \right], \\
\Sigma^2_{i,\underline{t}} &:= \left(\mathrm{Cov}_{Q_{i,0}}(\langle f_j, F_{i,t_j} \rangle, \langle f_l, F_{i,t_l} \rangle) \right)_{j,l=1}^{k}, \\
\langle \underline{f}, \bar{\eta}^{(N)}_{\underline{t}} \rangle &:= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \langle \underline{f}, F_{i,\underline{t}} \rangle - \langle \underline{f}, \mu_{i,\underline{t}} \rangle. \qquad (2.34)
\end{aligned}$$

We can now state and prove the following proposition:

Proposition 2.10. *As $N \to \infty$ (2.34) converges to a k-dimensional Gaussian distribution with mean zero and Covariance Matrix*

$$\overline{\Sigma}^2_{\underline{t}} := \left(\mathrm{Cov}_{\overline{Q}_0}(\langle f_j, F_{t_j} \rangle, \langle f_l, F_{t_l} \rangle) \right)_{j,l=1}^{k}, \qquad (2.35)$$

if (2.32) *and the following condition are met:*

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \langle f, \mu_{i,0} \rangle \langle g, \mu_{i,0} \rangle = \langle f, \bar{\mu}_0 \rangle \langle g, \bar{\mu}_0 \rangle, \qquad (2.36)$$

for all $f, g \in C([0,1], \mathbb{R})$.

Remark 2.11. *Condition* (2.36) *may seem unintuitive at first so we offer an alternative interpretation. If we let $(X_i)_{i \in \mathbb{N}}$ be a collection of independent random variables, each distributed according to $\mu_{i,0}$ and X be distributed as $\bar{\mu}_0$ then* (2.36) *is equivalent to:*

$$\frac{1}{N} \sum_{i=1}^{N} \mathrm{Cov}_{\mu_{i,0}}(f(X_i), g(X_i)) \to \mathrm{Cov}_{\bar{\mu}_0}(f(X), g(X))$$

Proof. Two conditions are needed in order to ensure the convergence of (2.34) to a Gaussian random variable. The first is that the sum of covariance matrices converges to (2.35) as the sum tends to infinity. The $(j,l)^{\text{th}}$ entry in this sum is the following:

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\mathrm{Cov}_{Q_{i,0}}(\langle f_j, F_{i,t_j}\rangle, \langle f_l, F_{i,t_l}\rangle) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{n,m=0}^{\infty}\frac{d_n^{\theta}(t_l-t_j)d_m^{\theta}(t_j)}{(n+\theta)(m+\theta)_{(2)}} \\
&\qquad \times \{2mn\langle f_j f_l, \mu_{i,0}\rangle + nm(m-1)\langle f_l, \mu_{i,0}\rangle\langle f_j, \mu_{i,0}\rangle \\
&\qquad\quad + nm\theta\langle f_l, \mu_{i,0}\rangle\langle f_j, \nu_0\rangle + n\theta\langle f_j f_l, \nu_0\rangle \\
&\qquad\quad + m\theta(n+m+\theta+1)\langle f_j, \mu_{i,0}\rangle\langle f_l, \nu_0\rangle \\
&\qquad\quad + \theta^2(n+m+\theta+1)\langle f_l, \nu_0\rangle\langle f_j, \nu_0\rangle\} \\
&\quad - \frac{1}{N}\sum_{i=1}^{N}\sum_{n,m=0}^{\infty}\frac{d_n^{\theta}(t_j)d_m^{\theta}(t_l)}{(n+\theta)(m+\theta)} \\
&\qquad \times \{nm\langle f_j, \mu_{i,0}\rangle\langle f_l, \mu_{i,0}\rangle + n\theta\langle f_l, \mu_{i,0}\rangle\langle f_j, \nu_0\rangle \\
&\qquad\quad + m\theta\langle f_j, \mu_{i,0}\rangle\langle f_l, \nu_0\rangle + \theta^2\langle f_j, \nu_0\rangle\langle f_l, \nu_0\rangle\},
\end{aligned}
\tag{2.37}
$$

with $j < l$. The derivation of this formula via Chapman-Kolmogorov can be found in Appendix D. Switching the order of summation, we see that condition (2.36), along with (2.32), ensures that (2.37) converges to (2.35).

The second condition needed is that the collection of random variables

$$\left(\frac{1}{\sqrt{N}}\langle \underline{f}, F_{i,\underline{t}}\rangle\right)_{i\in\mathbb{N}}$$

satisfy Lindeberg's condition. This is implied by Lyapunov's condition which we opt to show instead (see Appendix E for a brief proof of this implication). Thus we wish to show that there exists a $\delta > 0$ such that

$$\frac{1}{(\sqrt{N})^{2+\delta}}\sum_{i=1}^{N}\mathbb{E}_{Q_{i,0}}\left[||\langle \underline{f}, F_{i,\underline{t}}\rangle||^{2+\delta}\right] \overset{N\to\infty}{\to} 0, \tag{2.38}$$

where $||\cdot||$ denotes the Euclidean norm. If we let $\delta = 2$ then the left hand side of (2.38) is bounded above by

$$\frac{1}{N^2}\sum_{i=1}^{N}\left(\sum_{j=1}^{k}(\sup_{x\in[0,1]} f_j(x))^2\right)^2,$$

which, since the summand in i is uniformly bounded, converges to zero as $N \to \infty$.

With these two conditions met, Proposition 2.27 in [37] gives us our result. □

To conclude the convergence of our LLN and CLT objects in the non-identical case we only have to deal with tightness. Fortunately this follows very quickly from the bounds we have already established. As in the identical case we only need tightness of our processes integrated against (certain) continuous functions.

In a similar way to (2.10) and (2.14) we can simplify the expressions

$$\mathbb{E}\left[\left(\langle f,\bar{S}_t^{(N)}\rangle-\langle f,\bar{S}_s^{(N)}\rangle\right)^4\middle|F_{i,0}\sim Q_{i,0},1\leq i\leq N\right]$$

and

$$\mathbb{E}\left[\left(\langle f,\bar{\eta}_t^{(N)}\rangle-\langle f,\bar{\eta}_s^{(N)}\rangle\right)^4\middle|F_{i,0}\sim Q_{i,0},1\leq i\leq N\right]$$

using the independence of the summands. If we let $0\leq s<t$ and $X_{i,t-s}:=\langle f,F_{i,t}\rangle-\langle f,F_{i,s}\rangle$ then the above expressions simplify to

$$\begin{aligned}
&\frac{1}{N^4}\mathbb{E}\left[\left(\sum_{i=1}^{N}X_{i,t-s}\right)^4\right] \qquad (2.39)\\
=&\frac{1}{N^4}\Bigg[\sum_{i=1}^{N}\mathbb{E}[X_{i,t-s}^4]+4\sum_{i\neq j}\mathbb{E}[X_{i,t-s}^3]\mathbb{E}[X_{j,t-s}]\\
&\quad+6\sum_{i<j}\mathbb{E}[X_{i,t-s}^2]\mathbb{E}[X_{j,t-s}^2]\\
&\quad+12\sum_{\substack{i\neq j<k\\ i\neq k}}\mathbb{E}[X_{i,t-s}^2]\mathbb{E}[X_{j,t-s}]\mathbb{E}[X_{k,t-s}]\\
&\quad+24\sum_{i<j<k<l}\mathbb{E}[X_{i,t-s}]\mathbb{E}[X_{j,t-s}]\mathbb{E}[X_{k,t-s}]\mathbb{E}[X_{l,t-s}]\Bigg]
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{1}{N^2}\mathbb{E}\left[\left(\sum_{i=1}^{N} X_{i,t-s} - \mathbb{E}[X_{i,t-s}]\right)^4\right] \qquad (2.40)\\
=&\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^4] - 4\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^3]\mathbb{E}_{Q_{i,0}}[X_{i,t-s}]\\
&\qquad + 6\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^2]\mathbb{E}_{Q_{i,0}}[X_{i,t-s}]^2 - 3\mathbb{E}_{Q_{i,0}}[X_{i,t-s}]^4\\
&+ \frac{6}{N^2}\sum_{i<j}\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^2]\mathbb{E}_{Q_{i,0}}[X_{j,t-s}^2]\\
&\qquad - \mathbb{E}_{Q_{i,0}}[X_{i,t-s}^2]\mathbb{E}_{Q_{i,0}}[X_{j,t-s}]^2\\
&\qquad - \mathbb{E}_{Q_{i,0}}[X_{i,t-s}]^2\mathbb{E}_{Q_{i,0}}[X_{j,t-s}^2]\\
&\qquad + \mathbb{E}_{Q_{i,0}}[X_{i,t-s}]^2\mathbb{E}_{Q_{i,0}}[X_{j,t-s}]^2\\
\le&\frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^4] + 4\left|\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^3]\mathbb{E}_{Q_{i,0}}[X_{i,t-s}]\right|\\
&\qquad + 6\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^2]\mathbb{E}_{Q_{i,0}}[X_{i,t-s}]^2\\
&+ \frac{6}{N^2}\sum_{i<j}\mathbb{E}_{Q_{i,0}}[X_{i,t-s}^2]\mathbb{E}_{Q_{i,0}}[X_{j,t-s}^2]\\
&\qquad + \mathbb{E}_{Q_{i,0}}[X_{i,t-s}]^2\mathbb{E}_{Q_{i,0}}[X_{j,t-s}]^2
\end{aligned}
$$

where the conditioning on $\{F_{i,0} \sim Q_{i,0}, 1 \le i \le N\}$ is implicit and dropped here for readability. For more details on the cancellations and simplifications that occur in (2.40), see Appendix B.

In the same way as in Section 2.4 we can apply the bounds (2.26), (2.27), (2.28) and (2.29) to (2.39) and (2.40). We can apply these bounds equally to all of the summands since they are independent of the starting measure. In the LLN case the number of summands is at most $N_{[4]}$ (the last sum over i, j, k and l) and so the factor of $1/N^4$ ensures that any terms involving N can be bounded above by 1. Similarly in the CLT case, the number of terms in each sum is less than $N_{[2]}$ so the fractions $1/N^2$ and $6/N^2$ ensure that we can again bound the terms in N above by 1. This ensures that the collections

$$(((\langle f, \bar{S}_t^{(N)} \rangle))_{t\geq 0})_{N\in\mathbb{N}}$$

and

$$(((\langle f, \bar{\eta}_t^{(N)} \rangle))_{t\geq 0})_{N\in\mathbb{N}}$$

are tight in $C([0,\infty),\mathbb{R})$ thanks to Theorem I.4.3 of [29]. With this complete we can now state the final theorem in this thesis:

Theorem 2.12. *Let $(F_i)_{i=1}^{\infty} \subset C([0,\infty),\mathcal{P}([0,1]))$ be an independent collection of FV processes, each started at $F_{i,0} \sim Q_{i,0}$ with means $\mu_{i,0}$, with parent-independent mutation, parameters $(\theta,\nu_0) \in [0,\infty) \times \mathcal{P}([0,1])$.*

If there exists a $\overline{F}_0 \in \mathcal{P}([0,1])$ and $\overline{Q}_0 \in \mathcal{P}(\in \mathcal{P}([0,1]))$ such that $\overline{F}_0 \sim \overline{Q}_0$ and the following are satisfied:

$$\frac{1}{N}\sum_{i=1}^{N}\langle f, F_{i,0}\rangle \overset{N\to\infty}{\longrightarrow} \langle f, \overline{F}_0\rangle$$

$$\frac{1}{N}\sum_{i=1}^{N}\langle f, \mu_{i,0}\rangle\langle g, \mu_{i,0}\rangle \overset{N\to\infty}{\longrightarrow} \langle f, \bar{\mu}_0\rangle\langle g, \bar{\mu}_0\rangle$$

for all $f, g \in C([0,1], \mathbb{R})$ *then the processes* $\bar{S}^{(N)} \in C([0,\infty), \mathcal{P}(0,1))$ *and* $\bar{\eta}^{(N)} \in C([0,\infty), \mathcal{D}([0,1]))$ *defined by* (2.31) *and* (2.34) *respectively converge in distribution in to* $\bar{\mu}$ *and* $\bar{\eta}$ *which are characterised via Propositions 2.9 and 2.10.*

A Simplified Martingale Convergence Theorem

Here we state a simplified version of Theorem 1.4 from Chapter 7 of [18] (namely we only consider one dimensional processes and condition (b)) which is used to verify the convergence of $Y_\epsilon^{\sigma,\theta}$ to Z_t:

Theorem. *For $n = 1, 2, ...$ let $\{\mathcal{F}_t^n\}$ be a filtration and M_n be an $\mathcal{F}_t^n$ local martingale with sample paths in $D([0,\infty),\mathbb{R})$ and $M_n(0) = 0$. Let A_n be a process with sample paths in $D([0,\infty),\mathbb{R})$ such that $A_n(t) - A_n(s) \geq 0$ for $t > s \geq 0$. Assume the following conditions hold:*

- *For each $T > 0$, $\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T} |A_n(t) - A_n(t-)|\right] = 0$,*
- $\lim_{n\to\infty} \mathbb{E}\left[\sup_{t\leq T} |M_n(t) - M_n(t-)|^2\right] = 0$,
- *$M_n^2(t) - A_n(t)$ is an $\mathcal{F}_t^n$-local martingale*
- *$A_n(t) \to C_t$ in probability,*

where C is a deterministic continuous process such that $C_0 = 0$ and $C_t - C_s \geq 0$ for $t > s \geq 0$.

Then $M_n \Rightarrow L = (W(C_t))_{t\geq 0}$ in law in $D([0,\infty),\mathbb{R})$, where W is a standard Brownian motion.

B CLT Simplifications

Let $(X_i)_{i\in\mathbb{N}}$ be an independent collection of random variables with means $\mu_i \in \mathbb{R}$. Letting $S_n := \sum_{i=1}^n X_i - \mu_i$ we look to find the fourth moment of S_n. First, we calculate the fourth power of S_n:

$$
\begin{aligned}
S_n^4 &= \left(\sum_{i=1}^n X_i - \mu_i\right)^4 \qquad \text{(B.1)}\\
&= \sum_{i=1}^n (X_i - \mu_i)^4 + 4\sum_{i\neq j}(X_i-\mu_i)^3(X_j-\mu_j)\\
&\quad + 6\sum_{i<j}(X_i-\mu_i)^2(X_j-\mu_j)^2\\
&\quad + 12\sum_{\substack{i\neq j<k\\ i\neq k}}(X_i-\mu_i)^2(X_j-\mu_j)(X_k-\mu_k)\\
&\quad + 24\sum_{i<j<k<l}(X_i-\mu_i)(X_j-\mu_j)(X_k-\mu_k)(X_l-\mu_l).
\end{aligned}
$$

Pairwise independence within each sum means that, when applying expectations to both sides of (B.1), any linear term will be equal to zero since

$$\mathbb{E}[X_i - \mu_i] = \mu_i - \mu_i = 0.$$

Thus, the fourth moment of S_n can be simplified to the following:

$$\begin{aligned}
\mathbb{E}\left[S_n^4\right] =& \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i - \mu_i\right)^4\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu_i)^4\right] \\
&\quad + 6\sum_{i<j} \mathbb{E}[(X_i - \mu_i)^2 (X_j - \mu_j)^2] \\
&= \sum_{i=1}^{n} \mathbb{E}[X_i^4] - 4\mathbb{E}[X_i^3]\mu_i + 6\mathbb{E}[X_i^2]\mu_i^2 - 3\mu_i^4 \\
&\quad + 6\sum_{i<j} \mathbb{E}[X_i^2]\mathbb{E}[X_j^2] - \mathbb{E}[X_i^2]\mu_j^2 - \mu_i^2\mathbb{E}[X_j^2] + \mu_i^2\mu_j^2
\end{aligned}$$

C Fleming-Viot Regularity Calculations

For completeness we re-state the general formula for moments of a Fleming-Viot process (2.16) as it is used several times here:

$$\begin{aligned}
\mathbb{E}_{Q_0}\left[\langle f_1, F_t\rangle \ldots \langle f_m, F_t\rangle\right] = \sum_{n=0}^{\infty} d_n^{\theta}(t) \sum_{M\subset\{1,\ldots,m\}} \frac{1}{(n+\theta)_{(m)}} \\
\times \left\{ \sum_{k=1}^{|M|} n_{[k]} \sum_{\beta\in\pi(M,k)} |\beta_1|! \cdots |\beta_k|! \prod_{j=1}^{k} \left\langle \prod_{i\in\beta_j} f_i, \mu_0 \right\rangle \right\} \\
\times \left\{ \sum_{l=1}^{|M^c|} \sum_{\gamma\in\pi(M^c,l)} (|\gamma_1|-1)! \cdots (|\gamma_l|-1)! \theta^l \prod_{j=1}^{l} \left\langle \prod_{i\in\gamma_j} f_i, \nu_0 \right\rangle \right\}.
\end{aligned}$$

Considering the second moment regularity we have

$$\begin{aligned}
&\mathbb{E}_{Q_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^2 \middle| F_s = \mu\right] \\
&= \mathbb{E}\left[(\langle f, F_{t-s}\rangle - \langle f, \mu\rangle)^2 \middle| F_0 = \mu\right] \\
&= \mathbb{E}\left[\langle f, F_{t-s}\rangle^2 \middle| F_0 = \mu\right] - 2\langle f, \mu\rangle \mathbb{E}\left[\langle f, F_{t-s}\rangle | F_0 = \mu\right] \qquad \text{(C.1)} \\
&\qquad + \langle f, \mu\rangle^2 \\
&= \sum_{n=0}^{\infty} \frac{d_n^\theta(t-s)}{(n+\theta)_{(2)}} G_2(n, \theta, f, \mu, \nu_0) \\
&= \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)_{(2)}} G_2(N_{t-s}, \theta, f, \mu, \nu_0)\right], \qquad \text{(C.2)}
\end{aligned}$$

where

$$\begin{aligned}
G_2(n, \theta, f, \mu, \nu_0) = {} & 2n[\langle f^2, \mu\rangle - \langle f, \mu\rangle^2] + \theta\langle f^2, \nu_0\rangle \\
& + \theta^2\langle f, \nu_0\rangle^2 - 2\theta(\theta+1)\langle f, \nu_0\rangle\langle f, \mu\rangle \\
& + \theta(\theta+1)\langle f, \mu\rangle^2.
\end{aligned}$$

For the third and fourth moments we have

$$\begin{aligned}
&\mathbb{E}_{Q_0}\left[(\langle f, F_t\rangle - \langle f, F_s\rangle)^3 \middle| F_s = \mu\right] \\
&= \mathbb{E}\left[(\langle f, F_{t-s}\rangle - \langle f, \mu\rangle)^3 | F_0 = \mu\right] \\
&= \mathbb{E}\left[\langle f, F_{t-s}\rangle^3 | F_0 = \mu\right] - 3\langle f, \mu\rangle \mathbb{E}\left[\langle f, F_{t-s}\rangle^2 | F_0 = \mu\right] \\
&\quad + 3\langle f, \mu\rangle^2 \mathbb{E}\left[\langle f, F_{t-s}\rangle | F_0 = \mu\right] - \langle f, \mu\rangle^3 \qquad \text{(C.3)} \\
&= \sum_{n=0}^{\infty} \frac{d_n^\theta(t-s)}{(n+\theta)_{(3)}} G_3(n, \theta, f, \mu, \nu_0) \\
&= \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)_{(3)}} G_3(N_{t-s}, \theta, f, \mu, \nu_0)\right], \qquad \text{(C.4)}
\end{aligned}$$

and

$$
\begin{aligned}
&\mathbb{E}_{Q_0}\left[\left(\langle f, F_t\rangle - \langle f, F_s\rangle\right)^4 \middle| F_s = \mu\right] \\
&= \mathbb{E}\left[\left(\langle f, F_{t-s}\rangle - \langle f, \mu\rangle\right)^4 |F_0 = \mu\right] \\
&= \mathbb{E}\left[\langle f, F_{t-s}\rangle^4 |F_0 = \mu\right] - 4\langle f, \mu\rangle \mathbb{E}\left[\langle f, F_{t-s}\rangle^3 |F_0 = \mu\right] \\
&\quad + 6\langle f, \mu\rangle^2 \mathbb{E}\left[\langle f, F_{t-s}\rangle^2 |F_0 = \mu\right] \qquad\qquad (C.5) \\
&\quad - 4\langle f, \mu\rangle^3 \mathbb{E}\left[\langle f, F_{t-s}\rangle |F_0 = \mu\right] + \langle f, \mu\rangle^4 \\
&= \sum_{n=0}^{\infty} \frac{d_n^\theta(t-s)}{(n+\theta)_{(4)}} G_4(n, \theta, f, \mu, \nu_0) \\
&= \mathbb{E}_\infty\left[\frac{1}{(N_{t-s}+\theta)_{(4)}} G_4(N_{t-s}, \theta, f, \mu, \nu_0)\right], \qquad\qquad (C.6)
\end{aligned}
$$

where

$$
\begin{aligned}
&G_3(n, \theta, f, \mu, \nu_0) \\
&\quad = n[6\langle f^3, \mu\rangle - 6(\theta+3)\langle f^2, \mu\rangle\langle f, \mu\rangle + 6(\theta+2)\langle f, \mu\rangle^3 \\
&\qquad\qquad + 6\theta\langle f^2, \mu\rangle\langle f, \nu_0\rangle - 6\theta\langle f, \mu\rangle^2\langle f, \nu_0\rangle] \\
&\qquad + 2\theta\langle f^3, \nu_0\rangle + 3\theta^2\langle f^2, \nu_0\rangle\langle f, \nu_0\rangle + 3\theta^3\langle f, \nu_0\rangle^3 \\
&\qquad - 3\theta(\theta+2)\langle f, \mu\rangle\langle f^2, \nu_0\rangle - 3\theta^2(\theta+2)\langle f, \mu\rangle\langle f, \nu_0\rangle^2 \\
&\qquad + 3\theta(\theta+1)(\theta+2)\langle f, \mu\rangle^2\langle f, \nu_0\rangle - \theta(\theta+1)(\theta+2)\langle f, \mu\rangle^3,
\end{aligned}
$$

and

$$
\begin{aligned}
&G_4(n,\theta,f,\mu,\nu_0)\\
&\quad=12n^2[\langle f^2,\mu\rangle^2-2\langle f^2,\mu\rangle\langle f,\mu\rangle^2+\langle f,\mu\rangle^4+\theta\langle f^3,\mu\rangle\langle f,\nu_0\rangle]\\
&\qquad+2n[12\langle f^4,\mu\rangle-12(\theta+4)\langle f,\mu\rangle\langle f^3,\mu\rangle-6\langle f^2,\mu\rangle^2\\
&\qquad\qquad+(6\theta^2+42\theta+73)\langle f^2,\mu\rangle\langle f,\mu\rangle^2\\
&\qquad\qquad-6(\theta^2+5\theta+7)\langle f,\mu\rangle^4+12\theta\langle f^3,\mu\rangle\langle f,\nu_0\rangle\\
&\qquad\qquad-12\theta(\theta+4)\langle f,\mu\rangle\langle f^2,\mu\rangle\langle f,\nu_0\rangle\\
&\qquad\qquad+12\theta(\theta+3)\langle f,\mu\rangle^3\langle f,\nu_0\rangle-6\theta^2\langle f,\mu\rangle^2\langle f,\nu_0\rangle^2\\
&\qquad\qquad-6\theta\langle f,\mu\rangle^2\langle f^2,\nu_0\rangle-4\theta^3\langle f,\mu\rangle\langle f,\nu_0\rangle^3\\
&\qquad\qquad+6\theta\langle f^2,\mu\rangle\langle f^2,\nu_0\rangle+6\theta^2\langle f^2,\mu\rangle\langle f,\nu_0\rangle^2]\\
&\qquad+6\theta\langle f^4,\nu_0\rangle+8\theta^2\langle f,\nu_0\rangle\langle f^3,\nu_0\rangle3\theta^2\langle f^2,\nu_0\rangle^2\\
&\qquad+6\theta^3\langle f,\nu_0\rangle^2\langle f^2,\nu_0\rangle+\theta^4\langle f,\nu_0\rangle^4\\
&\qquad+\theta(\theta+1)(\theta+2)(\theta+3)\langle f,\mu\rangle^4\\
&\qquad-4\theta(\theta+1)(\theta+2)(\theta+3)\langle f,\mu\rangle^3\langle f,\nu_0\rangle\\
&\qquad-8\theta(\theta+3)\langle f,\mu\rangle\langle f^3,\nu_0\rangle\\
&\qquad-12\theta^2(\theta+3)\langle f,\mu\rangle\langle f,\nu_0\rangle\langle f^2,\nu_0\rangle\\
&\qquad-12\theta^3(\theta+3)\langle f,\mu\rangle\langle f,\nu_0\rangle^3\\
&\qquad+6\theta(\theta+2)(\theta+3)\langle f,\mu\rangle^2\langle f^2,\nu_0\rangle\\
&\qquad+6\theta^2(\theta+2)(\theta+3)\langle f,\mu\rangle^2\langle f,\nu_0\rangle^2.
\end{aligned}
$$

The important thing about these calculations is not the exact values of the functions G_2, G_3 and G_4. The crucial fact that allows us to obtain tightness is the absence of n^2 or higher terms in G_2 and G_3 and n^3 or higher terms in G_4. These higher order terms cancel when applying the formula for moments of a FV process (see the start of this appendix or (2.16)) to (C.1), (C.3) and (C.5). This cancellation is essential since it leaves the terms inside the expectation of (C.2) as approximately N_{t-s}^{-1} as $N_{t-s} \to \infty$, $t-s \to 0$. The terms inside the expectations of (C.4) and (C.6) are of order N_{t-s}^{-2} as $N_{t-s} \to, \infty$, $t-s \to 0$. It is these asymptotics combined with Proposition 2.7 that lead us to the bounds (2.26) - (2.29).

D Autocovariance Calculations

Utilising the Chapman-Kolmogorov equation we can re-write

$$\mathbb{E}_{Q_0}\left[\langle f, F_t\rangle\langle g, F_s\rangle\right]$$

where $t > s$, as

$$\int_{\mathcal{P}([0,1])}\int_{\mathcal{P}([0,1])}\langle f, \mu\rangle\langle g, \nu\rangle P(t-s, \nu, d\mu)P(s, \mu_0, d\nu),$$

where P is the transition function of the FV process, as in (2.15). From this we can perform the following string of equalities

$$\begin{aligned}
&\int_{\mathcal{P}([0,1])}\int_{\mathcal{P}([0,1])}\langle f,\mu\rangle\langle g,\nu\rangle P(t-s,\nu,d\mu)P(s,\mu_0,d\nu)\\
&=\int_{\mathcal{P}([0,1])}\left[\int_{\mathcal{P}([0,1])}\langle f,\mu\rangle P(t-s,\nu,d\mu)\right]\langle g,\nu\rangle P(s,\mu_0,d\nu)\\
&=\int_{\mathcal{P}([0,1])}\sum_{n=0}^{\infty}\frac{d_n^\theta(t-s)}{n+\theta}\left(n\langle f,\nu\rangle+\theta\langle f,\nu_0\rangle\right)\langle g,\nu\rangle P(s,\mu_0,d\nu)\\
&=\sum_{n=0}^{\infty}\frac{d_n^\theta(t-s)}{n+\theta}\left(n\int_{\mathcal{P}([0,1])}\langle f,\nu\rangle\langle g,\nu\rangle P(s,\mu_0,d\nu)\right.\\
&\qquad\qquad\left.+\theta\langle f,\nu_0\rangle\int_{\mathcal{P}([0,1])}\langle g,\nu\rangle P(s,\mu_0,d\nu)\right).
\end{aligned}$$

If we now apply (2.16) (see also the formula at the start of Appendix C) then we obtain the following:

$$\begin{aligned}
&\mathbb{E}_{Q_0}\left[\langle f,F_t\rangle\langle g,F_s\rangle\right]\\
&\qquad=\sum_{n,m=0}^{\infty}\frac{d_n^\theta(t-s)d_m^\theta(s)}{(n+\theta)(m+\theta)_{(2)}}G(m,n,f,g,\mu_0,\nu_0),
\end{aligned}\tag{D.1}$$

where

$$\begin{aligned}
G(m,n,f,g,\mu_0,\nu_0)=&2mn\langle fg,\mu_0\rangle+nm(m-1)\langle f,\mu_0\rangle\langle g,\mu_0\rangle\\
&+nm\theta\langle f,\mu_0\rangle\langle g,\nu_0\rangle+n\theta\langle fg,\nu_0\rangle\\
&+m\theta(n+m+\theta+1)\langle g,\mu_0\rangle\langle f,\nu_0\rangle\\
&+\theta^2(n+m+\theta+1)\langle f,\nu_0\rangle\langle g,\nu_0\rangle.
\end{aligned}$$

(D.1) minus the product

$$\mathbb{E}_{Q_0}\left[\langle f,F_t\rangle\right]\mathbb{E}_{Q_0}\left[\langle g,F_s\rangle\right]$$

then gives the formula for $\mathrm{Cov}_{Q_0}(\langle f, F_t\rangle, \langle g, F_s\rangle)$.

E Lyapunov and Lindeberg

Let $(Y_{n,i})_{n,i\in\mathbb{N}}$ be random variables in $\mathbb{R}^k$. We wish to show that the Lyapunov condition

$$\exists\delta > 0 \text{ such that } \lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[||Y_{n,i}||^{2+\delta}\right] = 0 \tag{E.1}$$

implies the Lindeberg condition

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[||Y_{n,i}||^{2}\mathbb{1}\{||Y_{n,i}|| > \epsilon\}\right] = 0, \quad \forall\epsilon > 0, \tag{E.2}$$

where $||\cdot||$ denotes the Euclidean norm. To see this first note that, on the event $\{||Y_{n,i}|| > \epsilon\}$,

$$\frac{||Y_{n,i}||^{\delta}}{\epsilon^{\delta}} > 1.$$

Then, if we consider the Lindeberg condition we can bound it above by the following:

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}\left[||Y_{n,i}||^{2}\mathbb{1}\{||Y_{n,i}|| > \epsilon\}\right] \\
&\qquad\leq \frac{1}{\epsilon^{\delta}}\sum_{i=1}^{n}\mathbb{E}\left[||Y_{n,i}||^{2+\delta}\mathbb{1}\{||Y_{n,i}|| > \epsilon\}\right] \\
&\qquad\leq \frac{1}{\epsilon^{\delta}}\sum_{i=1}^{n}\mathbb{E}\left[||Y_{n,i}||^{2+\delta}\right] \to 0,
\end{aligned}$$

and so (E.1) implies (E.2).

References

[1] J. F. C. Kingman, “The coalescent,” *Stochastic Processes and their Applications*, vol. 13, no. 3, pp. 235 – 248, 1982.

[2] J. F. C. Kingman, “On the genealogy of large populations,” *Journal of Applied Probability*, vol. 19, pp. 27–43, 7 1982.

[3] J. F. C. Kingman, “Exchangeability and the evolution of large populations,” in *Proceedings of the International Conference on Exchangeability in Probability and Statistics* (G. Koch and F. Spizzichino, eds.), pp. 97–112, North-Holland Elsevier, Amsterdam, 6th - 9th April 1982.

[4] C. Neuhauser and S. M. Krone, “The genealogy of samples in models with selection,” *Genetics*, vol. 145, no. 2, pp. 519–534, 1997.

[5] S. M. Krone and C. Neuhauser, “Ancestral processes with selection,” *Theoretical Population Biology*, vol. 51, no. 3, pp. 210 – 237, 1997.

[6] W. H. Fleming and M. Viot, “Some measure-valued Markov processes in population genetics theory,” *Indiana University Mathematics Journal*, vol. 28, no. 5, pp. 817–843, 1979.

[7] N. Berestycki, “Recent progress in coalescent theory,” *Ensaios Matemáticos*, vol. 16, pp. 1–193, 2009.

[8] V. Bansaye, S. Méléard, and M. Richard, "Speed of coming down from infinity for birth-and-death processes," *Advances in Applied Probability*, vol. 48, no. 4, p. 1183–1210, 2016.

[9] R. C. Griffiths, "Asymptotic line-of-descent distributions," *Journal of Mathematical Biology*, vol. 21, pp. 67–75, 1984.

[10] P. A. Jenkins and D. Spanò, "Exact simulation of the Wright–Fisher diffusion," *The Annals of Applied Probability*, vol. 27, p. 1478–1509, Jun 2017.

[11] V. Limic and A. Talarczyk, "Diffusion limits at small times for Λ-coalescents with a Kingman component," *Electron. J. Probab.*, vol. 20, p. 20 pp., 2015.

[12] R. Griffiths and P. Marjoram, "An ancestral recombination graph," in *Progress in population genetics and human evolution*, pp. 257 – 270, Springer, 1997. Progress in population genetics and human evolution ; Conference date: 01-01-1997.

[13] R. C. Griffiths, "A transition density expansion for a multi-allele diffusion model," *Advances in Applied Probability*, vol. 11, no. 2, pp. 310–325, 1979.

[14] R. C. Griffiths and W.-H. Li, "Simulating allele frequencies in a population and the genetic differentiation of populations under mutation pressure," *Theoretical Population Biology*, vol. 23, pp. 19–33, 1983.

[15] P. Donnelly and T. G. Kurtz, "Genealogical processes for Fleming-Viot models with selection and recombination," *Annals of Applied Probability*, vol. 9, pp. 1091–1148, 11 1999.

[16] J. Berestycki, N. Berestycki, and V. Limic, "The Λ-coalescent speed of coming down from infinity," *The Annals of Probability*, vol. 38, pp. 207–233, 01 2010.

[17] S. Peszat and J. Zabczyk, *Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach.* Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2007.

[18] S. N. Ethier and T. G. Kurtz, *Markov processes – characterization and convergence.* Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, New York: John Wiley & Sons Inc., 1986.

[19] V. Limic and A. Talarczyk, "Second-order asymptotics for the block counting process in a class of regularly varying Λ-coalescents," *Annals of Probability*, vol. 43, pp. 1419–1455, 05 2015.

[20] S. N. Ethier and T. G. Kurtz, "Fleming-Viot processes in population genetics," *SIAM Journal of Control and Optimization*, vol. 31, no. 2, pp. 345–386, 1993.

[21] H. Shi, G. Kichaev, and B. Pasaniuc, "Contrasting the genetic architecture of 30 complex traits from summary

association data," *The American Journal of Human Genetics*, vol. 99, pp. 139–153, 7 2016.

[22] E. A. Boyle, Y. I. Li, and J. K. Pritchard, "An expanded view of complex traits: From polygenic to omnigenic," *Cell*, vol. 169, pp. 1177–1186, 6 2017.

[23] M. Lynch, B. Walsh, *et al.*, *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA, 1998.

[24] T. S. Ferguson, "A bayesian analysis of some nonparametric problems," *The Annals of Statistics*, vol. 1, pp. 209–230, 03 1973.

[25] S. Ethier and T. G. Kurtz, "Convergence to fleming-viot processes in the weak atomic topology," *Stochastic Processes and their Applications*, vol. 54, no. 1, pp. 1–27, 1994.

[26] J. Pitman, *Some developments of the Blackwell-MacQueen urn scheme*, vol. Volume 30 of *Lecture Notes–Monograph Series*, pp. 245–267. Hayward, CA: Institute of Mathematical Statistics, 1996.

[27] S. Tavaré, "Line-of-descent and genealogical processes, and their applications in population genetics models," *Theoretical Population Biology*, vol. 26, no. 2, pp. 119 – 164, 1984.

[28] D. Dawson and E. Perkins, *Superprocesses at Saint-Flour*. Probability at Saint-Flour, Springer-Verlag Berlin Heidelberg, 2012.

[29] N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes.* Kodansha scientific books, North-Holland, 1989.

[30] P. Billingsley, ed., *Convergence of Probability Measures.* John Wiley & Sons, Inc., 7 1999.

[31] M. Munroe, *Introduction to Measure and Integration.* Addison-Wesley, 1953.

[32] F. Trèves, *Topological Vector Spaces, Distributions and Kernels*, vol. 25 of *Pure and Applied Mathematics.* Elsevier, 1967.

[33] L. Schwartz, *Théorie des distributions.* Hermann, Paris, 1966.

[34] I. Mitoma, "Tightness of Probabilities on C([0, 1]; Y') and D([0, 1]; Y')," *The Annals of Probability*, vol. 11, pp. 989–999, 1983.

[35] S. N. Ethier and R. C. Griffiths, "The transition function of a Fleming-Viot process," *The Annals of Probability*, vol. 21, pp. 1571–1590, 07 1993.

[36] P. K. Sen and J. M. Singer, *Large Sample Methods in Statistics (1994).* CRC Press, 11 2017.

[37] A. W. van der Vaart, *Asymptotic Statistics.* Cambridge University Press, 10 1998.