Stochastic Population Models: Measure-Valued and Partition-Valued Formulations

Diplomarbeit

Humboldt-Universität zu Berlin
Mathematisch-Naturwissenschaftliche Fakultät II
Institut für Mathematik

eingereicht von Nicolas Simon Perkowski
Betreuer: Prof. Dr. Peter Imkeller
Berlin, den 4. November 2009

Acknowledgements

I would like to thank my advisor Peter Imkeller for his support and his mentoring. Also I would like to thank Jochen Blath for giving me great further suggestions and ideas and for fruitful discussions. And last but not least I would like to thank Jean Bertoin for introducing me to the fascinating subjects treated in this work.

Contents

1 Introduction

Consider the easiest model in population genetics: the Wright-Fisher model. That is, we consider a population that develops over time. The population is supposed to be haploid, i.e. each individual has exactly one ancestor. The generations are non-overlapping and of constant size. Further suppose that there is an infinite number of generations both in the future and in the past. Each individual in generation n chooses its ancestor uniformly among the individuals of generation $n-1$, independently of the choices of the other individuals.

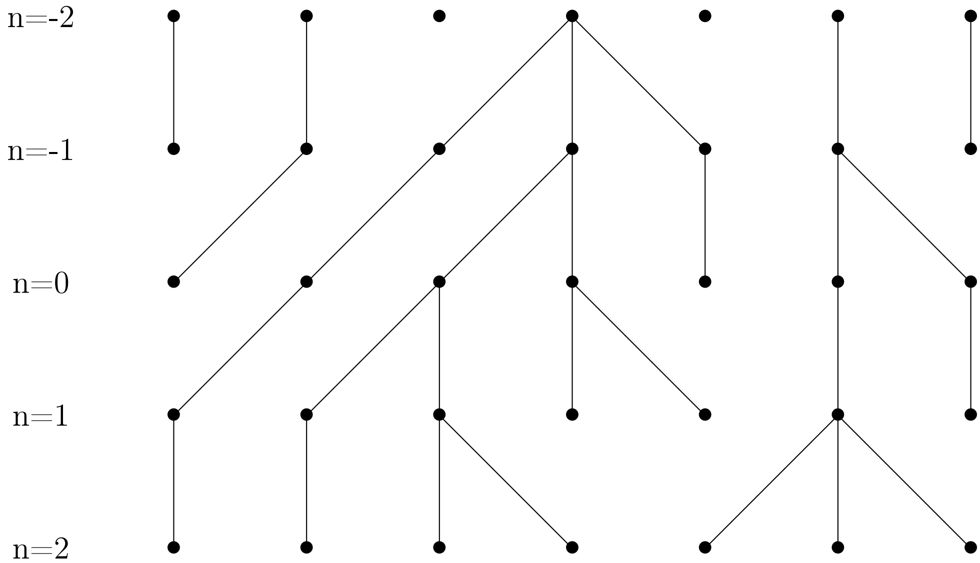

Figure 1: *An example of the genealogical tree for a population of size seven for the generations* -2 *to* 2*.*

If we model the development of the distribution of genetic types forward in time, we obtain a measure-valued process in the limit for large populations: the so called Fleming-Viot process (Kurtz, 1981).

If we model the genealogical tree backward in time, we obtain a partition-valued process in the limit for large populations: Kingman's coalescent (Kingman, 1982b).

Those two processes are dual to each other. This was shown by Dawson and Hochberg (1982). They proved the duality of the Fleming-Viot process to a function-valued process, but their formulation can be easily adapted to prove the duality of Fleming-Viot process and Kingman's coalescent.

The Wright-Fisher model is a special case of a class of population models that was introduced by Cannings (1974, 1975). Möhle and Sagitov (2001) studied the partition-valued formulation of Cannings' model and obtained a general class of coalescents in the limit for large populations, so called *exchangeable coalescents*. Schweinsberg (2000a) classified those exchangeable coalescents and proved that they are in one-to-one correspondance with finite measures Ξ on the infinite simplex

$$\Delta := \left\{ (x_1, x_2, \dots) \in \mathbb{R}^{\mathbb{N}} : x_1 \geq x_2 \geq \dots \geq 0 \text{ and } \sum_{i=1}^{\infty} x_i \leq 1 \right\}$$

This is why exchangeable coalescents are also called Ξ-coalescents. If we consider only measures on Δ that are concentrated on sequences of the form $(x_1, 0, 0, \dots)$ and can thus be interpreted as measures on $[0, 1]$, we also speak of Λ-coalescents.

Bertoin and Le Gall (2003) introduced a generalisation of the Fleming-Viot process, so called Λ-Fleming-Viot processes, for which they gave an explicit Poisson construction. Also they showed that Λ-Fleming-Viot processes and Λ-coalescents are dual to each other.

Ξ-Fleming-Viot processes (that are a generalisation of Λ-Fleming-Viot processes) were introduced explicitly by Birkner et al. (2009) who gave a fundamentally different construction of these processes than Bertoin and Le Gall (2003) gave for their Λ-coalescents.

In this work we want to generalize the result of Bertoin and Le Gall (2003). We will construct Ξ-Fleming-Viot processes and we will show the duality of Ξ-Fleming-Viot processes and Ξ-coalescents. Bertoin and Le Gall (2003) point out the possibility of such a generalisation and they state that "details are left to the interested reader". Having obtained the duality between Ξ-coalescents and Ξ-Fleming-Viot processes, it is not surprising that we will be able to show convergence of the measure-valued formulation of Cannings' model towards Ξ-Fleming-Viot processes.

Finally, we slightly generalize a realistic population model introduced by Schweinsberg (2003). This population model is in the class of Cannings' models and we can use the before obtained convergence results to show the convergence towards coalescents or Fleming-Viot processes, depending on the considered formulation.

In the entire text, we will always consider the Borel σ-algebra, unless it is noted otherwise. We will denote the Borel σ-algebra of a topological space E by $\mathcal{B}(E)$.

2 Preliminaries

Unless it is noted otherwise, everything in this section is a translation of the corresponding sections from Perkowski (2009)

2.1 Exchangeable Random Partitions

In this chapter we introduce the important correspondance between exchangeable random partitions and mass partitions.

2.1.1 Partitions of [n]

Definition 2.1. *1. Let $B \subseteq \mathbb{N}$, $B \neq \varnothing$, be a subset of $\mathbb{N} := \{1, 2, \dots\}$. A **partition** π **of** B is a family of disjoint blocks $(\pi_i : i \in \mathbb{N})$ such that $\bigcup_{i \in \mathbb{N}} \pi_i = B$. We suppose that the π_i are always enumerated by increasing order of their least element.*

2. For a partition π of B, $\boldsymbol{\#\pi} \in \bar{\mathbb{N}} := \mathbb{N} \cup \infty$ is the number of non-empty blocks of π, i.e. $\boldsymbol{\#\pi := \sup\{i : \pi_i \neq \emptyset\}}$.

3. For $i \in B$, $\boldsymbol{\pi(i)}$ is the number of the block of π that contains i.

4. $\boldsymbol{\mathcal{P}_n}$ *is the space of partitions of* $[n] := \{1, \ldots, n\}$*, equipped with the discrete topology.* $\mathbf{0}_n$ *is the partition of* $[n]$ *in singletons.*

5. $\boldsymbol{\mathcal{P}_\infty}$ *is the space of partitions of* $[\infty] := \mathbb{N}$*.* $\mathbf{0}_\infty$ *is the partition of* $\mathbb{N}$ *in singletons.*

6. *For* $n \in \bar{\mathbb{N}}$*,* $m \leq n$ *and* $\pi \in \mathcal{P}_n$ *let* $\boldsymbol{R_m \pi}$ *be the restriction of* π *to* $[m]$*:* $R_m\pi$ *is the unique partition in* $\mathcal{P}_m$ *such that for* $i, j \leq m$*,* i *and* j *are in the same block of* $R_m\pi$ *if and only if they are in the same block of* π*.*

7. *For* $n \in \bar{\mathbb{N}}$ *and* $\pi, \pi' \in \mathbb{P}_n$*, we write* $\boldsymbol{\pi \subseteq \pi'}$ *if* π' *is coarser than* π*, i.e. if* π' *is obtained by coagulating blocks of* π*. We write* $\boldsymbol{\pi \prec \pi'}$ *if* π' *is obtained by coagulating exactly two blocks of* π*.*

We introduce the notation

$$i \overset{\pi}{\sim} j$$

to express that i and j are in the same block of π. We define a distance ρ on $\mathcal{P}_\infty$:

$$\boldsymbol{\rho(\pi, \pi') := 2^{-\inf\{n : R_n\pi \neq R_n\pi'\}}}$$

We would like $\mathcal{P}_\infty$ to be a Polish space. In fact it is even a compact metric space:

Proposition 2.2. *$\mathcal{P}_\infty$ equipped with the distance ρ is a compact metric space.*

Proof. We will show that $(\mathcal{P}_\infty, \rho)$ is complete and that each sequence in $\mathcal{P}_\infty$ admits a Cauchy subsequence.

Let (π_n) be a Cauchy sequence in $\mathcal{P}_\infty$, and let $m \in \mathbb{N}$. So there is $N_m \in \mathbb{N}$ such that for each $n, n' \geq N_m$ we have $\rho(\pi_n, \pi_{n'}) < 2^{-m}$. So the sequence $(R_m\pi_n)_n$ is constant for large enough n. We define a partition $\pi \in \mathcal{P}_\infty$ such that $i \overset{\pi}{\sim} j$ if and only if $i \overset{\pi_n}{\sim} j$ for each n that is large enough. The definition of ρ immediately implies the convergence of π_n towards π.

Let (π_n) be a sequence in $\mathcal{P}_\infty$. We consider $(R_2\pi_n)$. Since $\mathcal{P}_2$ is finite, there is an infinite constant subsequence $(R_2\pi_{n_k})$. Then we consider $(R_3\pi_{n_k})$ and select another infinite constant subsequence $(R_3\pi_{n_{k_l}})$, etc. We obtain a Cauchy subsequence by choosing a diagonal sequence of this collection of subsequences of (π_n). □

2.1.2 Mass Partitions

Definition 2.3. *A* ***mass partition*** *is a real-valued sequence* $(x_1, x_2, \ldots)$ *such that*

$$x_1 \geq x_2 \geq \cdots \geq 0 \text{ and } \sum_{i=1}^{\infty} x_i \leq 1$$

We define

$$x_0 := 1 - \sum_{i=1}^{\infty} x_i$$

We denote by $\boldsymbol{\Delta}$ *the infinite simplex of mass partitions.*

Δ is a compact metric space:

Proposition 2.4. Δ *equipped with the uniform distance*

$$d(x, x') := max\{|x_i - x'_i| : i \in \mathbb{N}\}, \quad x, x' \in \Delta$$

is a compact space.

Uniform convergence is equivalent to simple convergence.

Proof. The equivalence of uniform convergence and simple convergence is a direct consequence of the fact that for each $x = (x_i)$ in Δ we have $x_i \leq 1/i$ for all $i \in \mathbb{N}$.

Let (x^n) be a sequence in Δ. We want to show that (x^n) admits a convergent subsequence. Since $(x_1^n)_n$ is a bounded sequence in $\mathbb{R}$, we can choose a convergent subsequence $x_1^{n_k} \overset{k\to\infty}{\longrightarrow} x_1$. Now we can choose a subsequence $(x_2^{n_{k_l}})$ of $(x_2^{n_k})$ that converges to a $x_2 \in \mathbb{R}$. We repeat this for each $i \in \mathbb{N}$. Then we choose a diagonal subsequence of all those subsequences. Denote that subsequence of (x^n) by (x^m). So for each $i \in \mathbb{N}$, x_i^m converges to x_i when $m \to \infty$. Of course the limit (x_i) is still monotone, i.e. $x_1 \geq x_2 \geq \dots$. Fatou's lemma yields

$$\sum_{i=1}^{\infty} x_i \leq 1$$

Thus $x = (x_i)$ is in Δ. Since uniform convergence is equivalent to simple convergence, (x^m) converges uniformly to x. □

Example 2.5. Let $(\xi_t, 0 \leq t \leq 1)$ be a pure jump subordinator with jumps $a_1 \geq a_2 \geq \dots$ in decreasing order. (In the Appendix B there is an overview of subordinators.) So $(a_1/\xi_1, a_2/\xi_1, \dots)$ is a random point in Δ, and the distribution of (ξ_t) corresponds to a distribution on Δ.

Let $\alpha \in (0, 1)$ and let $(\xi_t, t \in [0, 1])$ be a subordinator with Laplace exponent

$$\Phi(q) = cq^\alpha = \frac{c\alpha}{\Gamma(1-\alpha)} \int_0^\infty (1 - e^{-qx}) x^{-\alpha-1} dx.$$

for some $c > 0$. Here, Γ is the gamma function, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$. Such a (ξ_t) is called **stable subordinator of index $\boldsymbol{\alpha}$**. The Lévy measure of ξ is given by

$$\Lambda_\alpha(dx) = \frac{c\alpha}{\Gamma(1-\alpha)} x^{-\alpha-1} dx$$

It satisfies

$$\Lambda_\alpha(x, \infty) = \frac{c}{\Gamma(1-\alpha)} x^{-\alpha}$$

The corresponding distribution on Δ is called **Poisson-Dirichlet distribution of index ($\boldsymbol{\alpha}$,0), PD($\boldsymbol{\alpha}$,0)**. Note that the parameter c has no influence on the PD$(\alpha, 0)$-distribution, since $k^\alpha c$ corresponds to $(k\xi_t, t \in [0, 1])$ (this can be immediately seen by calculating the Laplace exponent).

2.1.3 Exchangeable Random Partitions

To define exchangeable random partitions, we first need to define permutations: A permutation of $[n]$ for $n \in \mathbb{N}$ is a bijective map from $[n]$ to $[n]$. A permutation of $\mathbb{N}$ is a bijective map σ from $\mathbb{N}$ to $\mathbb{N}$ such that there exists an $N \in \mathbb{N}$ with $\sigma(n) = n$ for each $n \geq N$.

For each permutation σ of $[n], n \in \bar{\mathbb{N}}$ and for each partition $\pi \in \mathcal{P}_n$ we define the partition $\hat{\sigma}\pi$ as follows: for $i, j \in [n]$, $\sigma(i) \overset{\hat{\sigma}\pi}{\sim} \sigma(j)$ if and only if $i \overset{\pi}{\sim} j$.

Definition 2.6. *A random partition π of $[n]$ with $n \in \bar{\mathbb{N}}$ is called **exchangeable** if the law of π is invariant under permutations of $[n]$, i.e. if for each permutation σ of $[n]$, $\hat{\sigma}\pi$ has the same distribution as π.*

Definition 2.7.

*A partition π of $\mathbb{N}$ is said to have **asymptotic frequencies** if for each block B of π:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{i\in B\}} \text{ exists}$$

With the paintbox construction of Kingman (1982b) we can associate an exchangeable random partition to each mass partition:

Definition 2.8. 1. *Let $x \in \Delta$. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed (i.i.d.) random variables, such that*

$$\mathbb{P}(\xi_1 = i) = x_i, i \in \mathbb{N}, \quad \mathbb{P}(\xi_1 = 0) = 1 - \sum_{i=1}^{\infty} x_i$$

Given the values of the ξ_n, we define a partition $\pi \in \mathcal{P}_\infty$ such that $i \neq j$ are in the same block of π if and only if

$$\xi_i = \xi_j > 0$$

So all i with $\xi_i = 0$ are singletons of π. We denote the distribution of π by $\boldsymbol{P^x}$. P^x is called a **paint box distribution***. To motivate this name, imagine that each number i corresponds to a color. 0 corresponds to a magic paint that has a different color each time it is used. Each element $j \in \mathbb{N}$ is painted with the colour ξ_j. Then all the elements with the same color are put in the same block of π.*

2. *For a distribution ν on Δ we define a **mixture of paint boxes**:*

$$P^\nu(d\pi) := \int_\Delta P^x(d\pi)\nu(dx)$$

It is easily verified that those paint boxes correspond to exchangeable partitions that almost surely (a.s.) possess asymptotic frequencies. The second statement is obtained with the law of large numbers. Indeed, every exchangeable random partition is given by a mixture of paint boxes. To prove this, we will need de Finetti's theorem. The following version is Theorem (3.1) of Aldous (1985):

Theorem 2.9 (de Finetti)**.** *Let $(Z_i)_{i\in\mathbb{N}}$ be an exchangeable sequence of real-valued random variables. That is, for each permutation σ of $\mathbb{N}$, $(Z_{\sigma(i)})_{i\in\mathbb{N}}$ has the same distribution as $(Z_i)_{i\in\mathbb{N}}$. Then there exists a random probability measure μ on $\mathbb{R}$ (cf. Definition A.1), such that*

$$(Z_i) \text{ is i.i.d. conditionally on the } \sigma\text{-algebra created by } \mu$$
$$\mathbb{P}(Z_i \in A|\mu)(\omega) = \mu(\omega, A)$$

Now we are ready to state the main result of this section. This theorem was established by Kingman (1978). The following proof is taken from Aldous (1985), Proposition (11.9), and we use details from the more elaborate version of Bertoin (2006), Theorem 2.1.

Theorem 2.10 (Kingman). *Let π be an exchangeable random partition of $\mathbb{N}$. Then π a.s. possesses asymptotic frequencies. Let $X_1 \geq X_2 \geq \dots$ be the ordered sequence of the asymptotic frequencies of the different blocks of π where $X_n := 0$ if π has less than n non-empty blocks. Then $X = (X_1, X_2, \dots)$ is a.s. in Δ, and conditionally on X, π has the distribution P^X. In particular*

$$\mathbb{P}(\pi \in A) = \int_\Delta P^x(A) G(dx)$$

where G is the distribution of X.

Proof.

1. $b : \mathbb{N} \to \mathbb{N}$ is called selection map for the partition η if for all i, j in the same block of η we have $b(i) = b(j) = k$ where k is an element of the same block of η. So let b be a selection map for π.

   Let $(\xi_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence that is uniformly distributed on $[0, 1]$ (notation: $\xi_i \simeq U([0, 1])$), independent of π and of b. We define $Z_i := \xi_{b(i)}$. Since b and π are independent of (ξ_i), the distribution of $(Z_i)_{i\in\mathbb{N}}$ does not depend on the selection map b.

2. The sequence (Z_i) is exchangeable: Let σ be a permutation of $\mathbb{N}$. We have

   $$Z_{\sigma(i)} = \xi_{b(\sigma(i))} = \xi'_{b'(i)}$$

   where

   $$\xi'_i := \xi_{\sigma(i)} \text{ and } b'(i) := (\sigma^{-1} \circ b \circ \sigma)(i)$$

   b' is a selection map for $\widehat{\sigma^{-1}}\pi$: Let i and j be in the same block of $\widehat{\sigma^{-1}}\pi$: Then $\sigma(i)$ and $\sigma(j)$ are in the same block of π, and thus

   $$b(\sigma(i)) = b(\sigma(j)) = \sigma(k)$$

   for a certain k such that $\sigma(k)$ and $\sigma(i)$ are in the same block of π. But that means that k and i are in the same block of $\widehat{\sigma^{-1}}\pi$. Further we have

   $$b'(i) = b'(j) = \sigma^{-1} \circ \sigma(k) = k$$

   and therefore b' is a selection map for $\widehat{\sigma^{-1}}\pi$. (ξ'_i) is an i.i.d. sequence that is uniformly distributed on $[0, 1]$ and that is independent of $\widehat{\sigma^{-1}}\pi$ and of b'. Since π is exchangeable, $\widehat{\sigma^{-1}}\pi$ has the same distribution as π, and thus $Z_{\sigma(i)}$ has the same distribution as (Z_i).

3. We use de Finetti's theorem (Theorem 2.9). Let μ be a random probability measure for (Z_i) as in the theorem. We can choose it such that for each ω, the mass of $\mu(\omega)$ is concentrated on $[0, 1]$. Let $f(\mu)(\omega)$ be the ordered sequence $\mu_1(\omega) \geq \mu(\omega)_2 \geq \dots$ of atoms of $\mu(\omega)$. That is, $\mu_1(\omega)$ is the size of the largest atom of $\mu(\omega)$, etc. We define $\mu_n(\omega) := 0$ if $\mu(\omega)$ has less than n atoms. Conditionally on μ, the distribution of π is given by $P^{f(\mu)}$:

   Let

   $$q(x) := \inf\{y : \mu([0, y]) \geq x\}$$

be the (random) quantile function of μ. We define

$$\theta := \{x \in (0,1) : \exists\, \epsilon > 0 \text{ such that } q(x) = q(y) \text{ if } |y - x| < \epsilon\}$$

The intervall lengths of θ correspond to the atom sizes of μ. Let $(V_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence, $V_i \simeq U([0,1])$, independent of π, of (Z_i), and of μ. Then

$$\mathbb{P}(q(V_1) \leq x|\mu) = \mathbb{P}(\mu([0,x]) \geq V_1|\mu) = \mu([0,x])$$

so conditionally on μ, $(q(V_i))$ has the same distribution as (Z_i). We define a partition π' such that i and j are in the same block of π' if and only if $q(V_i) = q(V_j)$. Conditionally on μ, π' has the same distribution as π. But i and j are in the same block of π' if and only if V_i and V_j are in the same intervall of θ. So conditionally on μ, π' (and therefore also π) has the paint box distribution $P^{f(\mu)}$. (We could define $W_i := k$ if V_i is in the k-th largest intervall of θ and $W_i := 0$ if V_i is in no intervall of θ to see that we are really in the paint box setting.)

4. Conclusion: We have

$$\mathbb{P}(\pi \in A|\mu) = P^{f(\mu)}(A)$$

and conditionally on μ, π has asymptotic frequencies $f(\mu)$. In particular, π a.s. possesses asymptotic frequencies. By taking expectations on both sides we get

$$\mathbb{P}(\pi \in A) = \int_{\Delta} P^x(A)G(dx)$$

where G is the distribution of $f(\mu)$, i.e. the distribution of the asymptotic frequencies of π.

□

2.2 Exchangeable Coalescents

2.2.1 Definition and Classification

We introduce coalescents with simultaneous multiple collisions and we show a correspondance between such coalescents and finite measures on the infinite simplex Δ.

Definition 2.11. *A **coalescent** is a stochastic process $(\Pi(t))_{t\geq 0}$ with values in $\mathcal{P}_n$ for $n \in \bar{\mathbb{N}}$ that is a.s. right-continuous and possesses left limits (càdlàg) and such that for all $s > t \geq 0$: $\Pi(t)$ is a refinement of $\Pi(s)$, i.e. $\Pi(t) \subseteq \Pi(s)$.*

Definition 2.12. *Let $B \subseteq \mathbb{N}$ be a subset of $\mathbb{N}$. Let π be a partition of B. Let $m \geq \#\pi$ and let $\pi' \in \mathcal{P}_m$. We define the partition $\mathbf{Coag(\pi, \pi')}$ as follows:*

$$\mathrm{Coag}(\pi,\pi')_j := \bigcup_{i\in\pi'_j} \pi_i, \quad j \leq \#\pi'$$

where $\mathrm{Coag}(\pi,\pi')_j$ is the jth block of $\mathrm{Coag}(\pi,\pi')$.

The coagulation operator has two elementary properties that will be very useful:

1. For $\pi, \pi' \in \mathcal{P}_\infty, n \in \mathbb{N}$, we have

$$R_n \operatorname{Coag}(\pi, \pi') = \operatorname{Coag}(R_n\pi, R_n\pi')$$

2. If all the terms in the following equation are well-defined we have

$$\operatorname{Coag}(\pi, \operatorname{Coag}(\pi', \pi'')) = \operatorname{Coag}(\operatorname{Coag}(\pi, \pi'), \pi'')$$

Definition 2.13. *Let $b, r \in \mathbb{N}$, $k_1, \dots, k_r \geq 2$, $s \in \mathbb{N}_0 := \{0, 1, \dots\}$, and $b = k_1 + \dots + k_r + s$. $\pi \in \mathcal{P}_b$ is called a* $\boldsymbol{(b; k_1, \dots, k_r; s)}$***-partition*** *if π has (non-ordered) blocks $B'_1, \dots, B'_r$ of respective sizes $k_1, \dots, k_r$, and s singletons.*

Definition 2.14. *Let $\pi \in \mathcal{P}_n$, $n \in \bar{\mathbb{N}}$ and $\infty > b = \#\pi$. π' is a* $\boldsymbol{(b; k_1, \dots, k_r; s)}$***-collision*** *of π if $\pi' = \operatorname{Coag}(\pi, \pi'')$ where π'' is any $(b; k_1, \dots, k_r; s)$-partition.*

Here we will only consider coalescents that are Markov processes and for which the rate of each $(b; k_1, \dots, k_r; s)$-collision is the same.

Definition 2.15. *Let $m \in \bar{\mathbb{N}}$. A coalescent $(\Pi(t))_{t\geq 0}$ with values in $\mathcal{P}_m$ is called* ***coalescent with simultaneous multiple collisions (c.s.m.c.)*** *or* ***exchangeable coalescent*** *if for all $n, m \in \mathbb{N}$, $n \leq m$:*

$$(R_n\Pi(t))_{t\geq 0} \text{ is a Markov chain with values in } \mathcal{P}_n$$

and

when $R_n\Pi(t)$ has b blocks, each $(b; k_1, \dots, k_r; s)$-collision
happens with rate $\lambda_{b;k_1,\dots,k_r;s}$

If $\Pi(0) = \mathbf{0}_m$, then Π is called ***standard****.*

An important example of such coalescents is given by Kingman's coalescent. For this coalescent, the collision rates are $\lambda_{b;2;b-2} = 1$ for each b, and every other rate is 0. This means that the jump rate from π to π' is 1 if π' is formed from π by coagulating exactly 2 of its blocks, and otherwise the rate is 0. This process was introduced by Kingman (1982b) to study the genealogy of large populations. The new idea that proved to be very successful was to consider a process with values in $\mathcal{P}_n$. Kingman proved that this coalescent arises in the limit for large populations in a number of models: The Wright-Fisher model, the Moran model (which we will not study here), but also the general Cannings' model if we assume the family sizes to be sufficiently bounded (this will be expressed more precisely later in this text). The mathematical properties of Kingman's coalescent are described in Kingman (1982a).

In 1998, Bolthausen and Sznitman (1998) introduced another exchangeable coalescent. This paved the way for the general classification of those processes:

In 1999, Pitman (1999) and Sagitov (1999) introduced independently of each other coalescents with multiple collisions. Those are exchangeable coalescents with $\lambda_{b;k_1,\dots,k_r;s} = 0$ for $r > 1$, i.e. each λ that is not of the form $\lambda_{b;k;b-k}$ is 0. This evidently means that for such coalescents we can have a collision of several blocks (not just of two blocks as for Kingman's coalescent), but a.s. no two such collisions happen at the same time.

Coalescents with simultaneous multiple collisions were obtained the first time by Möhle and Sagitov (2001) as limits of Cannings' population models. A classification of c.s.m.c.'s was given by Schweinsberg (2000a). In this article Schweinsberg proved that c.s.m.c.'s are in one-to-one correspondance with finite measures on the space of mass partitions Δ:

Theorem 2.16. *Let $\{\lambda_{b;k_1,\dots,k_r;s} : r, b \in \mathbb{N}, k_1, \dots, k_r \geq 2, s \in \mathbb{N}_0, b = \sum_{j=1}^r k_j + s\}$ be a family of positive (i.e. ≥ 0) numbers. Then there exists a standard coalescent with simultaneous multiple collisions with values in $\mathcal{P}_\infty$ with collision rates $\lambda_{b;k_1,\dots,k_r;s}$, if and only if there is a finite measure Ξ on Δ,*

$$\Xi = \Xi_0 + c\delta_{\mathbf{0}}$$

where Ξ_0 has no atom in $\mathbf{0} := (0, 0, \dots)$, $\delta_{\mathbf{0}}$ is the Dirac mass in $\mathbf{0}$ and $c \geq 0$, such that

$$\lambda_{b;k_1,\dots,k_r;s} = \int_\Delta \frac{Q_{k_1,\dots,k_r;s}(x)}{\sum_{j=1}^\infty x_j^2} \Xi_0(dx) + c\mathbb{1}_{\{r=1,k=2\}} \text{ with} \tag{1}$$

$$Q_{k_1,\dots,k_r;s}(x) := \sum_{l=0}^{s} \sum_{i_1 \neq \dots \neq i_{r+l}} \binom{s}{l} x_{i_1}^{k_1} \dots x_{i_r}^{k_r} x_{i_{r+1}} \dots x_{i_{r+l}} \left(1 - \sum_{j=1}^\infty x_j\right)^{s-l} \tag{2}$$

For each c.s.m.c., the associated measure Ξ is uniquely determined.

Remark. 1. Note that the integral in (1) is well-defined, as Ξ_0 has no atom in $\mathbf{0}$.

2. The formula (1) is the formula that was originally established by Schweinsberg (2000a). There is another formula given by Bertoin (2006). Bertoin considers the measure

$$\nu(dx) := \left(1 \middle/ \sum_{j=1}^\infty x_j^2\right) \Xi_0(dx) + c\delta_{\mathbf{0}}$$

that is not necessarily finite on Δ.

Definition 2.17. *A c.s.m.c. $(\Pi(t))_{t\geq 0}$ with rates $\lambda_{b;k_1,\dots,k_r;s}$ given by (1) is called **Ξ-coalescent**.*

Poissonian Construction To show that condition (1) is sufficient, we construct a Ξ-coalescent with a Poisson point process construction (cf. Appendix A for an overview of Poisson point processes). This construction was originally given by Schweinsberg (2000a), but we present the slightly adapted version of Bertoin (2006), Chapter 4.2.3. Nonetheless some details in the proof are taken from Schweinsberg (2000a).

Let ν be a σ-finite measure on Δ such that

$$\nu(\{\mathbf{0}\}) = 0 \text{ and } \int_\Delta \sum_{j=1}^\infty x_j^2 \nu(dx) < \infty \tag{3}$$

Let $c \geq 0$. We associate a σ-finite measure μ on $\mathcal{P}_\infty$ to ν and c: For $i, j \in \mathbb{N}$ let $\kappa(i, j)$ be the unique partition of $\mathbb{N}$ that consists of one block of size two, $\{i, j\}$, and otherwise only of singletons. We define

$$\mu(d\pi) := \int_\Delta P^x(d\pi)\nu(dx) + c \sum_{i=1}^\infty \sum_{j=i+1}^\infty \mathbb{1}_{\kappa(i,j)}(d\pi) \tag{4}$$

Since ν is σ-finite, μ is σ-finite as well.

Let $(e(t))_{t\geq 0}$ be a Poisson point process of intensity μ. We will use $(e(t))$ to construct processes $(\Pi_n(t))_{t\geq 0}$ with values in $\mathcal{P}_n$. Then we will see that all the Π_n are compatible: a.s.

$R_m\Pi_n(t) = \Pi_m(t)$ for each t. Therefore we can define a process $(\Pi(t))_{t\geq 0}$ with values in $\mathcal{P}_\infty$ such that $R_n\Pi(t) = \Pi_n(t)$ for $t \geq 0$, $n \in \mathbb{N}$.

For $n \in \mathbb{N}$ we define

$$A_n := \{\pi \in \mathcal{P}_\infty : R_n\pi \neq \mathbf{0}_n\}$$

and for $k, l \in \mathbb{N}$:

$$A_{k,l} := \{\pi \in \mathcal{P}_\infty : k \text{ and } l \text{ are in the same block of } \pi\} \tag{5}$$

We have

$$\begin{aligned}\mu(A_n) &\leq \sum_{k=1}^{n}\sum_{l=k+1}^{n} \mu(A_{k,l}) = \sum_{k=1}^{n}\sum_{l=k+1}^{n}\left(\int_\Delta P^x(A_{k,l})\nu(dx) + 1\right) \\ &= \binom{n}{2}\left(\int_\Delta \sum_{j=1}^{\infty} x_j^2 \nu(dx) + 1\right) < \infty\end{aligned}$$

The last inequality comes from (3).

We define $T_{0,n} := 0$ and for $k \geq 1$: $T_{k,n} := \inf\{t > T_{k-1,n} : e(t) \in A_n\}$. Since $\mu(A_n) < \infty$, the $T_{k,n}$ correspond to jump times of a Poisson process. Thus they are without cluster point and we have $e(T_{k,n}) \in A_n$ for $k \geq 1$. Given a partition $\pi \in \mathcal{P}_\infty$, we define $\Pi_n^\pi(0) := R_n\pi$ and

$$\Pi_n^\pi(T_{k,n}) := \mathrm{Coag}(\Pi_n^\pi(T_{k-1,n}), e(T_{k,n}))$$

Now let $m < n \in \mathbb{N}$. Since $A_m \subseteq A_n$, Π_m^π and Π_n^π are constant on the intervall $[T_{k,n}, T_{k+1,n})$ for each $k \geq 0$. Thus it suffices to verify the equality $\Pi_m^\pi(t) = R_m\Pi_n^\pi(t)$ a.s. for $t \in \{T_{k,n} : k \geq 0\}$. For $k = 0$ this is trivial. Let $k \geq 1$. Recall that for a partition η, $\eta(i)$ is the number of the block containing i. Let $i, j \in [m]$. Then i and j are in the same block of $\Pi_m^\pi(T_{k,n})$ if and only if $\Pi_m^\pi(T_{k-1,n})(i)$ and $\Pi_m^\pi(T_{k-1,n})(j)$ are in the same block of $e(T_{k,n})$. On the other side i and j are in the same block of $\Pi_n^\pi(T_{k,n})$ (and thus of $R_m\Pi_n^\pi(T_{k,n})$) if and only if $\Pi_n^\pi(T_{k-1,n})(i)$ and $\Pi_n^\pi(T_{k-1,n})(j)$ are in the same block of $e(T_{k,n})$. But since the blocks of partitions are enumerated by increasing order of their least element, and since by induction hypothesis $\Pi_m^\pi(T_{k-1,n}) = R_m\Pi_n^\pi(T_{k-1,n})$, we have $\Pi_n^\pi(T_{k-1,n})(i) = \Pi_m^\pi(T_{k-1,n})(i)$ for each $i \in [m]$. We obtain $\Pi_m^\pi(T_{k,n}) = R_m\Pi_n^\pi(T_{k,n})$.

The construction of Π^π is now evident: Let $i, j \in \mathbb{N}$, then i and j are in the same block of $\Pi^\pi(t)$ if they are in the same block of $\Pi_{\max\{i,j\}}^\pi(t)$. Using the definition of the topology on $\mathcal{P}_\infty$ it is evident that Π^π is càdlàg and that for each $t < s$, $\Pi^\pi(t)$ is a refinement of $\Pi^\pi(s)$. Therefore we constructed a coalescent.

Given a finite measure $\Xi = \Xi_0 + c\delta_{\mathbf{0}}$ on Δ, we define $\nu(dx) := \frac{\Xi_0(dx)}{\sum_{j=1}^\infty x_j^2}$ and we construct Π^π exactly like we just did. It remains to show that Π^π is a Ξ-coalescent.

Proposition 2.18 (Sufficient Condition of Theorem 2.16)**.** *The process $(\Pi^\pi(t))_{t\geq 0}$ constructed as above is a Ξ-coalescent.*

Proof. 1. $R_n\Pi^\pi$ is a Markov chain:

$R_n\Pi^\pi = \Pi_n^\pi$ where Π_n^π is the process of the construction. By using the construction and the "independent increments" (55) of Poisson point processes, it is easily verified that Π_n^π is a Markov chain.

2. Each $(b;k_1,\dots,k_r;s)$-collision has the rate $\lambda_{b;k_1,\dots,k_r;s}$:
   Let $n\in\mathbb{N}$. Let $\pi\in\mathcal{P}_\infty$ such that $R_n\pi$ has b blocks. Let π' be a $(b;k_1,\dots,k_r;s)$-partition. We denote its (non-ordered) blocks of size ≥ 2 by $B'_1,\dots,B'_r$. The jump rate of $R_n\Pi^\pi(0)=\Pi^\pi_n(0)$ to $\mathrm{Coag}(R_n\pi,\pi')$ is given by $\mu(A_{\infty,\pi'})$ with $A_{\infty,\pi'}:=\{\eta\in\mathcal{P}_\infty : R_b\eta=\pi'\}$. We calculate $P^x(A_{\infty,\pi'})$: Recall that P^x was constructed by i.i.d. variables (ξ_i). If $\mathbb{R}_b\eta=\pi'$, there exist necessarily

$$
\begin{aligned}
&0\leq l\leq s,\ i_1\neq\dots\neq i_{r+l} \text{ all} \neq 0 \text{ and } 1\leq m_1<\dots<m_l\leq b \text{ such that}\\
&\xi_m=i_j \text{ for } m\in B'_j, 1\leq j\leq r\\
&\xi_{m_j}=i_{r+j},\ 1\leq j\leq l\\
&\xi_m=0 \text{ for } m\leq b, m\notin \left(\cup_{j=1,\dots,r}B'_j\right)\cup\{m_1,\dots,m_l\}
\end{aligned}
$$

   By summing up all the possible combinations we obtain

$$
\begin{aligned}
P^x(A_{\infty,\pi'}) &= \sum_{l=0}^{s}\binom{s}{l}\sum_{i_1\neq\dots\neq i_{r+l}} x_{i_1}^{k_1}\dots x_{i_r}^{k_r}x_{i_{r+1}}\dots x_{i_{r+l}}\left(1-\sum_{j=1}^{\infty}x_j\right)^{s-l}\\
&= Q_{k_1,\dots,k_r;s}(x) \qquad (6)
\end{aligned}
$$

   This implies

$$
\begin{aligned}
\lambda_{b;k_1,\dots,k_r;s} = \mu(A_{\infty,\pi'}) &= \int_\Delta Q_{k_1,\dots,k_r;s}(x)\nu(dx)+c\mathbb{1}_{\{r=1,k=2\}}\\
&= \int_\Delta\left(Q_{k_1,\dots,k_r;s}(x)/\sum_{j=1}^{\infty}x_j^2\right)\Xi_0(dx)+c\mathbb{1}_{\{r=1,k=2\}}
\end{aligned}
$$

   and this is the desired formula (1).

□

Necessary Condition of Theorem 2.16 Given the $\lambda_{b;k_1,\dots,k_r;s}$, we will construct a σ-finite measure μ on $\mathcal{P}_\infty$. Then we will associate a σ-finite measure ν on Δ and a $c\geq 0$ to μ. We will see that ν satisfies (3), and we will be able to define a finite measure on Δ by setting $\Xi(dx):=\sum_{j=1}^\infty x_j^2\nu(dx)+c\delta_{\mathbf{0}}$. Then we will see that the rates $\lambda_{b;k_1,\dots,k_r;s}$ are given by (1).

We choose this complicated way to obtain the results of Schweinsberg (2000a) (that we want to use) with the methods of Bertoin (2006) (that reveal more about the structure of coalescents with simultaneous multiple collisions).

Definition 2.19. *Given $\pi\in\mathcal{P}_n$, $n\in\mathbb{N}$, we define for $m\in\bar{\mathbb{N}}$, $m>n$:*

$$\boldsymbol{A_{m,\pi}}:=\{\pi'\in\mathcal{P}_m : R_n\pi'=\pi\}$$

Proposition 2.20. *There is an unique measure μ on $\mathcal{P}_\infty$ such that $\mu(A_{\infty,\pi})=\lambda_{b;k_1,\dots,k_r;s}$ for each $(b;k_1,\dots,k_r;s)$-partition π. This measure satisfies*

1. *μ is invariant under permutations of $\mathbb{N}$ (then μ is called* ***exchangeable****),*
2. *$\mu(\{\mathbf{0}_\infty\})=0$,*
3. *$\mu(\{\pi\in\mathcal{P}_\infty : R_n\pi\neq\mathbf{0}_n\})<\infty$ for each $n\in\mathbb{N}$*

Proof. For each $(b; k_1, \ldots, k_r; s)$-partition π with $r > 0$ let

$$q_\pi := \lambda_{b;k_1,\ldots,k_r;s}$$

We define

$$\mathcal{A}_n := \sigma(\{A_{\infty,\pi} : \pi \in \mathcal{P}_n \backslash \{\mathbf{0}_n\}\}) \text{ and } \mathcal{A} := \bigcup_{n\in\mathbb{N}} \mathcal{A}_n$$

It is easily verified that $\mathcal{A}$ is an algebra. We define a measure μ_0 on $\mathcal{A}$ by

$$\mu_0(A_{\infty,\pi}) := q_\pi$$

To verify that μ_0 is σ-additive, we consider $\pi \in \mathcal{P}_n$ and $m > n$. Since $R_n R_m \Pi^{\mathbf{0}_\infty} = R_n \Pi^{\mathbf{0}_\infty}$, we have

$$q_\pi = \sum_{\pi' \in A_{m,\pi}} q_{\pi'} \tag{7}$$

which is the same as

$$\mu_0(A_{\infty,\pi}) = \mu_0\left(\dot{\bigcup}_{\pi' \in A_{m,\pi}} A_{\infty,\pi'}\right) = \sum_{\pi' \in A_{m,\pi}} \mu_0(A_{\infty,\pi'})$$

μ_0 is evidently additive on $\mathcal{A}_n$, thus we have a σ-additive measure on an algebra $\mathcal{A}$. We can use Caratheodory's extension theorem to extend μ_0 to an unique measure μ on $\mathcal{B}(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\}) = \sigma(\mathcal{A})$ if we consider each $\mathcal{A}_n$ as sub-set of $\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\}$ rather than $\mathcal{P}_\infty$. To obtain a measure on $\sigma(\mathcal{P}_\infty)$, we define $\mu(\{\mathbf{0}_\infty\}) := 0$. μ satisfies condition 2 by definition. Condition 3 is satisfied since

$$\mu(\{\pi \in \mathcal{P}_\infty : R_n\pi \neq \mathbf{0}_n\}) = \sum_{\pi \in \mathcal{P}_n \backslash \{\mathbf{0}_n\}} q_\pi$$

Condition 1 is satisfied since $q_\pi = q_{\hat{\sigma}\pi}$ for each permutation σ of $[n]$. $\square$

Proposition 2.21. *Let μ be the measure of Proposition 2.20. There are a unique measure ν on Δ and a unique $c \geq 0$ such that*

$$\mu(d\pi) = \int_\Delta P^x(d\pi)\nu(dx) + c\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} \delta_{\kappa(i,j)}(d\pi).$$

ν satisfies

$$\nu(\mathbf{0}) = 0 \text{ and } \int_\Delta \sum_{j=1}^{\infty} x_j^2 \nu(dx) < \infty.$$

We even have a stronger result:

1. *μ-almost every (a.e.) π has asymptotic frequencies*

2. *ν is given by*

$$\nu(dx) = \mathbb{1}_{\{x\neq\mathbf{0}\}}\mu(|\pi|^\downarrow \in dx)$$

   where $|\pi|^\downarrow$ denotes the asymptotic frequency of π, and

$$\mathbb{1}_{\{|\pi|^\downarrow\neq 0\}}\mu(d\pi) = \int_\Delta P^x(d\pi)\nu(dx)$$

3. $\mathbb{1}_{\{|\pi|^\downarrow=0\}}\mu(d\pi) = c\sum_{i=1}^\infty \sum_{j=i+1}^\infty \delta_{\kappa(i,j)}(d\pi)$

Proof. 1. For $n \in \mathbb{N}$ we introduce

$$\mu_n(d\pi) := \mathbb{1}_{\{R_n\pi\neq\mathbf{0}_n\}}\mu(d\pi).$$

Since $\mu(\{\pi : R_n\pi \neq \mathbf{0}_n\}) < \infty$ (cf. Proposition 2.20), μ_n is a finite measure on $\mathcal{P}_\infty$. Let $\vec{\mu}_n$ be the image measure of μ_n under

$$\pi \mapsto \vec{\pi} \quad \text{where} \quad i \overset{\vec{\pi}}{\sim} j \Longleftrightarrow n+i \overset{\pi}{\sim} n+j$$

Since μ is exchangeable, $\vec{\mu}_n$ is a finite exchangeable measure on $\mathcal{P}_\infty$. From Kingman's theorem (applied to $\vec{\mu}_n(.)/\vec{\mu}_n(\mathcal{P}_\infty)$) we obtain that $\vec{\mu}_n$-a.e. π possesses asymptotic frequencies and that $\vec{\mu}_n$ is given by

$$\vec{\mu}_n(d\pi) = \int_\Delta P^x(d\pi)\vec{\mu}_n(|\pi|^\downarrow \in dx) \tag{8}$$

Let $A := \{\pi : \pi \text{ possesses asymptotic frequencies}\}$. We have $\mu(\{\mathbf{0}_\infty\}) = 0$ and the asymptotic frequencies of a partition π do not depend on $R_n\pi$ for $n < \infty$. Thus

$$\begin{aligned}\mu(A) &= \lim_{n\to\infty}\mu_n(A) = \lim_{n\to\infty}\mu_n(\{\pi : \vec{\pi} \text{ possesses asymptotic frequencies}\})\\ &= \lim_{n\to\infty}\vec{\mu}_n(A) = 1\end{aligned}$$

which yields the first statement of the theorem.

2. By using the same measure extension argument as in the proof of Proposition 2.20, we see that it suffices to show

$$\mu(R_k\pi = \pi_k, |\pi|^\downarrow \neq \mathbf{0}) = \int_\Delta P^x(R_k\pi = \pi_k)\mathbb{1}_{\{x\neq\mathbf{0}\}}\mu(|\pi|^\downarrow \in dx) \tag{9}$$

for $k \in \mathbb{N}$ and $\pi_k \in \mathcal{P}_k$. So let k and π_k be given. By monotone convergence we obtain

$$\begin{aligned}&\mu(R_k\pi = \pi_k, |\pi|^\downarrow \neq \mathbf{0})\\ &\quad= \lim_{n\to\infty}\mu(R_k\pi = \pi_k, |\pi|^\downarrow \neq \mathbf{0}, \pi|_{\{k+1,\dots,k+n\}} \neq \mathbf{0}_{\{k+1,\dots,k+n\}})\end{aligned}$$

where $\mathbf{0}_{\{k+1,\dots,k+n\}}$ is the partition of $\{k+1,\dots,k+n\}$ into singletons. Since μ is exchangeable, this expression equals

$$\begin{aligned}&= \lim_{n\to\infty}\vec{\mu}_n(R_k\pi = \pi_k, |\pi|^\downarrow \neq \mathbf{0})\\ &\overset{(8)}{=} \lim_{n\to\infty}\int_\Delta P^x(R_k\pi = \pi_k)\mathbb{1}_{\{x\neq\mathbf{0}\}}\vec{\mu}_n(|\pi|^\downarrow \in dx)\end{aligned}$$

With the same argument that we used in the proof of 1., we see that $|\pi|^\downarrow$ does not change under $\pi \mapsto \vec{\pi}$; hence we obtain

$$\begin{aligned}&= \lim_{n\to\infty}\int_\Delta P^x(R_k\pi = \pi_k)\mathbb{1}_{\{x\neq\mathbf{0}\}}\mu_n(|\pi|^\downarrow \in dx)\\ &= \lim_{n\to\infty}\int_\Delta P^x(R_k\pi = \pi_k)\mathbb{1}_{\{x\neq\mathbf{0}\}}\mu(|\pi|^\downarrow \in dx, R_n\pi \neq \mathbf{0}_n)\end{aligned}$$

By monotone convergence and using $\mu(\{\mathbf{0}_\infty\}) = 0$, we obtain the desired equation:

$$\mu(R_k\pi = \pi_k, |\pi|^\downarrow \neq \mathbf{0}) = \int_\Delta P^x(R_k\pi = \pi_k)\mathbb{1}_{\{x\neq\mathbf{0}\}}\mu(|\pi|^\downarrow \in dx)$$

Hence it suffices to define $\nu(dx) := \mathbb{1}_{\{x\neq\mathbf{0}\}}\mu(|\pi|^\downarrow \in dx)$. It remains to show that

$$\int_\Delta \sum_{j=1}^\infty x_j^2 \nu(dx) < \infty$$

But this is easy now:

$$\begin{aligned}\int_\Delta \sum_{j=1}^\infty x_j^2 \nu(dx) &= \int_\Delta P^x(1 \stackrel{\pi}{\sim} 2)\nu(dx) \\ &\stackrel{(9)}{=} \mu(R_2\pi = \{1,2\}, |\pi|^\downarrow \neq \mathbf{0}) \leq \mu(R_2\pi \neq \mathbf{0}_2) < \infty\end{aligned}$$

The last inequality is condition 3 of Proposition 2.20.

3. Let $\bar{\mu}(d\pi) := \mathbb{1}_{\{1\stackrel{\pi}{\sim}2, |\pi|^\downarrow=\mathbf{0}\}}\mu(d\pi)$ and let $\vec{\mu}$ be the image measure of $\bar{\mu}$ under

$$\pi \mapsto \vec{\pi} \quad \text{where} \quad i \stackrel{\vec{\pi}}{\sim} j \iff 2+i \stackrel{\pi}{\sim} 2+j$$

$\vec{\mu}$ is a finite exchangeable measure on $\mathcal{P}_\infty$ and under $\vec{\mu}$, a.e. π has the asymptotic frequency $\mathbf{0}$. Hence $\vec{\mu}$ is a Dirac mass in $\mathbf{0}_\infty$. Since μ is exchangeable and $\bar{\mu}(\mathcal{P}_\infty) < \infty$,

$$\bar{\mu}(\exists\, j \geq 3 : 1 \stackrel{\pi}{\sim} j) = \sum_{j=1}^\infty \bar{\mu}(1 \stackrel{\pi}{\sim} j) = 0.$$

Therefore $\bar{\mu} = c\delta_{\kappa(1,2)}$ for some $c \geq 0$. Since μ is exchangeable, we deduce

$$\mathbb{1}_{\{|\pi|^\downarrow=\mathbf{0}\}}\mu(d\pi) = c\sum_{i=1}^\infty \sum_{j=i+1}^\infty \delta_{\kappa(i,j)}$$

□

To obtain the rates $\lambda_{b;k_1,\dots,k_r;s}$, we first calculate $\lambda_{b;k_1,\dots,k_r;s}$ as a function of ν: Let π be a $(b;k_1,\dots,k_r;s)$-partition. Then

$$\begin{aligned}\lambda_{b;k_1,\dots,k_r;s} &= \mu(A_{\infty,\pi}) = \int_\Delta P^x(A_{\infty,\pi})\nu(dx) + c\sum_{i=1}^\infty\sum_{j=i+1}^\infty \delta_{\kappa(i,j)}(A_{\infty,\pi}) \\ &\stackrel{(6)}{=} \int_\Delta Q_{k_1,\dots,k_r;s}(x)\nu(dx) + c\mathbb{1}_{\{r=1,k_1=2\}}\end{aligned}$$

By defining $\Xi(dx) := \sum_{j=1}^\infty x_j^2\nu(dx)$, we obtain a finite measure on Δ such that the $\lambda_{b;k_1,\dots,k_r;s}$ are given by (1).

2.2.2 Examples

Without doubt the most prominent example of an exchangeable coalescent is Kingman's coalescent. It corresponds to $\Xi = \delta_{\mathbf{0}}$. This coalescent has some interesting properties:

Proposition 2.22. *Let $(\Pi(t), t \geq 0)$ be a standard Kingman coalescent with values in $\mathcal{P}_\infty$.*

1. *Π **comes down from infinity**. This means that for each $t > 0$, a.s. $\#\Pi(t) < \infty$. Further, a.e. block of $\Pi(t)$ is of infinite size.*
2. *$(D_t := \#\Pi(t), t > 0)$ is a pure death process with death rate $(\binom{k}{2}, k \in \mathbb{N})$. More precisely,$(D_t)$ is a Markov process with values in $\mathbb{N}$ and with jump rates*
$$\lim_{h \downarrow \infty} \frac{1}{h} \mathbb{P}(D_{t+h} = l | D_t = k) = \begin{cases} \binom{k}{2}, & l = k-1 \\ 0, & \textit{otherwise} \end{cases}$$
for all k.
3. *Each trajectory of $(\Pi(t))$ passes by a sequence*
$$\ldots \mathcal{R}_k \prec \mathcal{R}_{k-1} \prec \cdots \prec \mathcal{R}_2 \prec \mathcal{R}_1$$
where $\mathcal{R}_k$ is the state of Π when $\#\Pi = k$. The sequence $(\mathcal{R}_k)$ is independent of (D_t), it is Markovian, and for each k, conditionally on $\mathcal{R}_{k+1} = \pi$, $\mathcal{R}_k$ is distributed uniformly on the $\binom{k+1}{2}$ partitions that are obtained by coagulating exactly two blocks of π.
4. *As a consequence of 2. and 3. we obtain: For all $S \in \mathcal{B}(\mathcal{P}_\infty)$*
$$\mathbb{P}(R_t \in S) = \sum_{k=1}^{\infty} \mathbb{P}(D_t = k) \mathbb{P}(\mathcal{R}_k \in S)$$

The proof of this proposition can be found in Kingman (1982a), Theorem 4.

An entire class of Ξ-coalescents that are particularly easy to describe are coalescents with multiple (asynchronous) collisions that were introduced independently by Pitman (1999) and Sagitov (1999).

Definition 2.23. *A **coalescent with multiple asynchronous collisions** or **simple coalescent** is an exchangeable coalescent that corresponds to a finite measure Λ on Δ which satisfies*
$$\Lambda(\{x = (x_1, x_2, \ldots) : x_2 > 0\}) = 0$$

In this case we could rather consider the image measure of Λ under the projection $(x_1, x_2, \ldots) \mapsto x_1$. Thus we can view Λ as a finite measure on $[0,1]$. In this setting the rates $\lambda_{b;k_1,\ldots,k_r;s}$ are given by
$$\lambda_{b;k} := \lambda_{b;k;b-k} = \int_{[0,1]} x^{k-2}(1-x)^{b-k} \Lambda(dx)$$
and all other rates are 0. In words, a simple coalescent is an exchangeable coalescent without simultaneous collisions. At each collision time, several blocks are selected and united to form a single new block.

Pitman showed in Proposition 23 of Pitman (1999) that each simple coalescent comes down from infinity or **stays infinite**, which means that the coalescent a.s. has an infinite number of blocks at each time t.

Example 2.24. For $r, s > 0$, we can consider $\Lambda = \text{Beta}(r, s)$. $\text{Beta}(r, s)$ is the distribution on $[0, 1]$ with density

$$\frac{x^{r-1}(1-x)^{s-1}}{B(r,s)}$$

where B is the beta-function,

$$B(r,s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)} = \int_0^1 x^{r-1}(1-x)^{s-1}dx$$

In this case the jump rates are given by

$$\lambda_{b;k} = \frac{B(k+r-2, b+s-k)}{B(r,s)}.$$

Schweinsberg showed in Schweinsberg (2000b), Example 15, that a standard $\text{Beta}(r, s)$-coalescent comes down from infinity if and only if $r < 1$.

In the case $r = s = 1$, $\text{Beta}(1, 1)$ is the uniform distribution on $[0, 1]$. We denote it by U. The U-coalescent has jump rates

$$\lambda_{b;k} = \frac{(k-2)!(b-k)!}{(b-1)!}$$

and was introduced by Bolthausen and Sznitman (1998). The standard U-coalescent does not come down from infinity.

2.2.3 Some Properties of Coalescents

Elementary Properties

Remark. 1. With the Poisson-construction one can easily see that a Ξ-coalescent $(\Pi^\pi(t))_{t\geq 0}$ with $\Pi^\pi(0) = \pi$ is obtained by defining

$$\Pi^\pi(t) := \text{Coag}(\pi, \Pi(t)), \quad t \geq 0$$

where $(\Pi(t))_{t\geq 0}$ is a standard Ξ-coalescent. We even have a stronger result: Conditionally on $\Pi^\pi(t)$, $(\Pi^\pi(t+s))_{s\geq 0}$ has the same distribution as $\text{Coag}(\Pi^\pi(t), \Pi(s))_{s\geq 0}$.

2. If $(\Pi(t))_{t\geq 0}$ is a standard exchangeable coalescent, then for each $t \geq 0$, $\Pi(t)$ is a random exchangeable partition. This is equally verified with the Poissonian construction since the measure μ that we had constructed on $\mathcal{P}_\infty$ was exchangeable and the coagulation of two independent exchangeable partitions is still exchangeable (cf. Bertoin (2006), Lemma 4.3).

3. Let Ξ be a finite measure on Δ with $\Xi(\Delta) \neq 0$. The case $\Xi(\Delta) = 0$ is trivial, since in that case all jump rates are 0. We define $G := \Xi/\Xi(\Delta)$. Then G is a probability on Δ, and with the definition of the jump rates (1), we see that the rates of the G-coalescent are given by dividing the rates of the Ξ-coalescent by $\Xi(\Delta)$. Modulo a change of the time scale we can therefore suppose $\Xi(\Delta) = 1$.

Lemma 2.25. *The jump rates*

$$\{\lambda_{b;k_1,\dots,k_r;s} : r,b \in \mathbb{N}, k_1,\dots,k_r \geq 2, s \in \mathbb{N}_0, b = k_1 + \cdots + k_r + s\}$$

of an exchangeable coalescent satisfy the following consistency relation:

$$\begin{aligned}\lambda_{b;k_1,\dots,k_r;s} = &\sum_{m=1}^{r} \lambda_{b+1;k_1,\dots,k_{m-1},k_m+1,k_{m+1},\dots,k_r;s} + s\lambda_{b+1;k_1,\dots,k_r,2;s-1} \\ &+ \lambda_{b+1;k_1,\dots,k_r;s+1}\end{aligned} \tag{10}$$

where we define $\lambda_{b;k_1,\dots,k_r;-1} := 0$. *This equation can be rewritten as*

$$\begin{aligned}\lambda_{b;k_1,\dots,k_r;s+1} =&\lambda_{b;k_1,\dots,k_r;s} - \sum_{m=1}^{r} \lambda_{b+1;k_1,\dots,k_{m-1},k_m+1,k_{m+1},\dots,k_r;s} \\ &- s\lambda_{b+1;k_1,\dots,k_r,2;s-1}\end{aligned} \tag{11}$$

This is a recurrence equation that allows us the calculate all the rates when we are only given the

$$\lambda_{b;k_1,\dots,k_r;0}, \quad b,r \in \mathbb{N}, k_1,\dots,k_r \geq 2, b = k_1 + \cdots + k_r$$

We do not give the proof here. This is Lemma 18 in Schweinsberg (2000a). The proof is elementary and it is based on the fact that $R_n R_{n+1}\Pi = R_n \Pi$ for an exchangeable coalescent Π. Noting this, one distinguishes the different possibilities for the behaviour of $n+1$ in $R_{n+1}\Pi$, and one gets the desired equation.

Behaviour at Collision Times

Lemma 2.26. *Let* Ξ *be a probability on* Δ *and let* $(\Pi(t))_{t\geq 0}$ *be a standard* Ξ*-coalescent. Let for* $i \neq j$ $\tau_{i,j} := \inf\{t \geq 0 : i \overset{\Pi(t)}{\sim} j\}$. *Let* $B_1, B_2, \dots$ *be the blocks of* $\Pi(\tau_{i,j}-)$ *(that are possibly empty for large enough* k*). Let* $\pi \in \mathcal{P}_{\#\Pi(\tau_{i,j}-)}$ *be the unique partition with* $k \overset{\pi}{\sim} l$ *if and only if* B_k *and* B_l *are in the same block of* $\Pi(\tau_{i,j})$. *Then* π *is the restriction of a partition* $\pi' \in \mathcal{P}_\infty$ *to* $\{1,\dots,\#\Pi(\tau_{i,j}-)\}$. π' *is invariant under permutations of* $\mathbb{N}$ *that do not change* $\Pi(\tau_{i,j}-)(i)$ *and* $\Pi(\tau_{i,j}-)(j)$*, and* π' *a.s. possesses asymptotic frequencies that have distribution* Ξ.

Sketch of the proof.

1. Without loss of generality we suppose that Π is given by the Poissonian construction. Since $\Pi(t)$ is exchangeable for each t, it suffices to show the statement for $i,j = 1,2$.

2. We have

$$\tau_{1,2} = \inf\{t \geq 0 : e(t) \in A_{1,2}\}$$

where $A_{1,2}$ is defined as in (5). It suffices to show $|e(\tau_{1,2})|^{\downarrow} \simeq \Xi$. Let $S \in \mathcal{B}(\Delta)$. We define

$$A_{1,2}^S := \{\varepsilon \in A_{1,2} : |\varepsilon|^{\downarrow} \in S\}.$$

The formula (57) of the Appendix A yields:

$$\begin{aligned}\mathbb{P}(|e(\tau_{1,2})|^{\downarrow} \in S) &= \mathbb{P}(e(\tau_{1,2}) \in A_{1,2}^S) \overset{(57)}{=} \frac{\mu(A_{1,2}^S)}{\mu(A_{1,2})} \\ &= \frac{1}{\mu(A_{1,2})}\left[\int_\Delta \sum_{j=1}^{\infty} x_j^2 \mathbb{1}_{\{x\in S\}} \Big/ \sum_{j=1}^{\infty} x_j^2 \Xi_0(dx) + c\sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} \delta_{\{\kappa(i,j)\in A_{1,2}^S\}}\right] \\ &= \frac{1}{\mu(A_{1,2})}\left[\Xi_0(S) + c\mathbb{1}_{\{\mathbf{0}\in S\}}\right] = \frac{1}{\mu(A_{1,2})}\Xi(S)\end{aligned}$$

Since $\mu(A_{1,2}) = 1$, the proof is complete.

$\square$

Coming from Infinity and Proper Frequencies Let $(\Pi_t : t \geq 0)$ be a simple standard coalescent with rates $\lambda_{b;k}$. We denote by γ_b the rate with which the number of blocks of $R_b\Pi$ decreases, i.e.

$$\boldsymbol{\gamma_b := \sum_{k=2}^{b} (k-1)\binom{b}{k}\lambda_{b;k}}$$

Schweinsberg (2000b) showed that Π comes down from infinity if and only if

$$\sum_{b=2}^{\infty} \gamma_b^{-1} < \infty$$

For general exchangeable coalescents we do not know an equally simple condition that is equivalent to the coming down from infinity. But there is a nice result on the asymptotic frequencies:

Definition 2.27. *Let $\pi \in \mathcal{P}_\infty$ be a partition that possesses asymptotic frequencies and let $(x_1, x_2, \dots) \in \Delta$ be the ordered sequence of its frequencies. We say that π has **proper frequencies** if*

$$\sum_{j=1}^{\infty} x_j = 1$$

*Otherwise we say that π has **dust**.*

Proposition 2.28. *Let $\Xi = \Xi_0 + c\delta_{\mathbf{0}}$ be a finite measure on Δ with $\Xi(\mathbf{0}) = 0$ and $c \geq 0$. Let $(\Pi_t : t \geq 0)$ be a standard Ξ-coalescent and let $t > 0$. Then Π_t a.s. has proper frequencies if and only if $c > 0$ or if*

$$\int_\Delta \sum_{j=1}^{\infty} x_j \Big/ \sum_{j=1}^{\infty} x_j^2 \Xi_0(dx) = \infty$$

Proof. Let ν be the distribution of the asymptotic frequencies of Π_t. Then the distribution of Π_t is given by

$$\int_\Delta P^x(d\pi)\nu(dx)$$

With the definition of the paint box P^x we see that Π_t a.s. has proper frequencies if and only if $\{1\}$ a.s. is not a block of Π_t.

Without loss of generality we suppose that Π is given by the Poisson construction with Poisson point process $(e(t))_{t\geq 0}$ of intensity μ. We define

$$A := \{\pi \in \mathcal{P}_\infty : \{1\} \text{ is no block of } \pi\}$$

and $T_A := \inf\{t \geq 0 : e(t) \in A\}$. Then $\{1\}$ is a block of Π_t if and only if $T_A > t$. We have $\mathbb{P}(T_A > t) = 0$ if and only if $\mu(A) = \infty$. But

$$\begin{aligned}
\mu(A) &= \int_\Delta P^x(A)\Xi_0(dx) \Big/ \sum_{j=1}^{\infty} x_j^2 + c \sum_{i=1}^{\infty} \sum_{j=i+1}^{\infty} \mathbb{1}_{\{\kappa(i,j)\}}(A) \\
&= \int_\Delta \sum_{j=1}^{\infty} x_j \Big/ \sum_{j=1}^{\infty} x_j^2 \Xi_0(dx) + c \sum_{j=2}^{\infty} 1
\end{aligned}$$

and this is infinite if and only if $c > 0$ or $\int_\Delta \sum_{j=1}^{\infty} x_j \Big/ \sum_{j=1}^{\infty} x_j^2 \Xi_0(dx) = \infty$. $\square$

Feller Property We recall the definition of a Feller process (cf. Revuz and Yor (1999), Definition (2.1) and Definition (2.5) of Chapter III.). Since $\mathcal{P}_\infty$ is compact, $C_0(\mathcal{P}_\infty) = C(\mathcal{P}_\infty)$.

Definition 2.29. *A* ***Feller semi-group*** *on $C(\mathcal{P}_\infty)$ is a family of linear positive (i.e. $f \geq 0 \Rightarrow P_t f \geq 0$) operators $(T_t)_{t\geq 0}$ on $C(\mathcal{P}_\infty)$ such that*

1. *$T_0 = \mathrm{Id}$ and $||T_t|| \leq 1$ for $t \geq 0$,*
2. *$T_{t+s} = T_t \circ T_s$ for $t, s \geq 0$ and*
3. *$\lim_{t\to 0} T_t f = f$ in $C(\mathcal{P}_\infty)$ for each $f \in C(\mathcal{P}_\infty)$.*

A ***Feller process*** *is a Markov process with a Feller semi-group.*

Proposition 2.30 (Feller Property)**.** *Let Ξ be a finite measure on Δ. Each Ξ-coalescent is a Feller process in its canonic filtration. Its semi-group is given by*

$$P_t f(\pi) = \mathbb{E}(f(\mathrm{Coag}(\pi, \Pi_t)))$$

where Π is a standard Ξ-coalescent.

Proof. Let $(\Pi_t : t \geq 0)$ be a Ξ-coalescent. Without loss of generality we can suppose that Π is given by the Poissonian construction with Poisson point process e. We already remarked that conditionally on Π_t, $(\Pi_{t+s} : s \geq 0)$ has the same distribution as $(\mathrm{Coag}(\Pi(t), \bar{\Pi}(s)) : s \geq 0)$ where $\bar{\Pi}$ is a standard Ξ-coalescent that is independent of Π. Indeed this remains true if we condition on $(\Pi_r : 0 \leq r \leq t)$ since $(e(t+s))_{s\geq 0}$ is independent of $(e(r))_{0\leq r<t}$. So Π_t is a Markov process. It remains to show that its semi-group $(P_t : t \geq 0)$ is Feller. It suffices to show that $P_t C(\mathcal{P}_\infty) \subseteq C(\mathcal{P}_\infty)$ and that for each $f \in C(\mathcal{P}_\infty)$ and for each $\pi \in \mathcal{P}_\infty$, we have $\lim_{t\to 0} P_t f(\pi) = f(\pi)$ (cf. Proposition (2.4) of Chapter III of Revuz and Yor (1999)).

Let $\pi \in \mathcal{P}_\infty$, let $(\Pi^\pi_t : t \geq 0)$ be a Ξ-coalescent with $\Pi^\pi(0) = \pi$ and let $(\Pi_t : t \geq 0)$ be a standard Ξ-coalescent that is independent of Π^π. Let $f \in C(\mathcal{P}_\infty)$. We have

$$P_t f(\pi) = \mathbb{E}(f(\Pi^\pi_t)) = \mathbb{E}(f(\mathrm{Coag}(\pi, \Pi_t)))$$

But it is easily verified that the Coag operator is continuous from $\mathcal{P}_\infty \times \mathcal{P}_\infty$ to $\mathcal{P}_\infty$. With dominated convergence we obtain the continuity of $P_t f$. It remains to show that $\lim_{t\to 0} P_t f(\pi) = f(\pi)$. But this follows immediately since Π is right-continuous, $\Pi_0 = \mathbf{0}_\infty$ and $\mathrm{Coag}(\pi, \mathbf{0}_\infty) = \pi$. Then we use once again dominated convergence and we obtain the desired result. □

We remark that as a consequence each Ξ-coalescent admits the strong Markov property (cf. Theorem (3.1) in Chapter III. of Revuz and Yor (1999)).

2.2.4 Exchangeable Coalescents and Martingale Problems

We want to show that the Ξ-coalescent is the unique solution to an easily described martingale problem. Let $\lambda_{b;k_1,\ldots,k_r;s}$ be the rates of a Ξ-coalescent. We write $\lambda_\pi := \lambda_{b;k_1,\ldots,k_r;s}$ for every $(b; k_1, \ldots, k_r; s)$-partition π. Let $\mathcal{D} := \{F \in C(\mathcal{P}_\infty) : \exists n \in \mathbb{N}, \tilde{F} \in C(\mathcal{P}_n), F(\pi) = \tilde{F}(R_n \pi) \forall \pi\}$. We define an operator

$$Q : \mathcal{D} \to C(\mathcal{P}_\infty), F(\cdot) \mapsto \sum_{\eta \in \mathcal{P}_n} \lambda_\eta(\tilde{F}(\mathrm{Coag}(R_n \cdot, \eta)) - \tilde{F}(R_n \cdot))$$

Because of the consistency relation (10) this operator is well-defined, and of course it is just the restriction of the infinitesimal generator of the Ξ-coalescent to $\mathcal{D}$. So we know that the Ξ-coalescent with starting distribution ν is a solution to the (Q,ν)-martingale problem (cf. Appendix C for an overview of martingale problems).

Proposition 2.31. *For Q and ν as above, every solution to the (Q,ν)-martingale problem has the same finite-dimensional distributions as the Ξ-coalescents starting with distribution ν. Any solution with càdlàg paths is a Ξ-coalescent.*

Proof. Let Π be a solution. Then for any $n \in \mathbb{N}$, $R_n\Pi$ is a solution to the (Q_n,ν_n)-martingale problem with

$$Q_n : B(\mathcal{P}_n) \to B(\mathcal{P}_n), \quad Q_nF(\cdot) = \sum_{\eta\in\mathcal{P}_n} \lambda_\eta(F(\mathrm{Coag}(\cdot,\eta)) - F(\cdot))$$

where $B(\mathcal{P}_n)$ is the space of bounded measurable functions on $\mathcal{P}_n$ and $\nu_n := \nu \circ R_n^{-1}$. But for a finite state space there is uniqueness for any martingale problem (cf. example in Appendix C). That means that for every solution Π of the (Q,ν)-martingale problem the finite-dimensional distributions of $R_n\Pi$ are uniquely determined. The functions depending only on $R_n\pi$ form an algebra in $C(\mathcal{P}_\infty)$ that separates points and contains constants. So it is dense in the uniform topology by the Stone-Weierstrass theorem. Thus we obtain the uniqueness of the finite-dimensional distributions for solutions to the martingale problem. Since the Ξ-coalescents is a solution, this means that any solution has the same finite-dimensional distributions as the Ξ-coalescent.

We immediately obtain from Proposition C.3 in Appendix C that a solution with càdlàg paths is a Ξ-coalescent. □

2.2.5 Exchangeable Coalescents in Discrete Time

In this section we introduce a discrete time version of the Ξ-coalescent. Under certain assumptions we will obtain such processes as limits of Cannings' population models. For a Ξ-coalescent to exist it is necessary that Ξ satisfies an additional condition.

Proposition 2.32. *Let $\{p_{b;k_1,\dots,k_r;s} : b,r \in \mathbb{N}, k_1,\dots,k_r \geq 2, s \in \mathbb{N}_0, b = \sum_{j=1}^r k_j + s\}$ be a family of non-negative numbers. Then there exists a discrete time process $(Y(m) : m \in \mathbb{N}_0)$ with values in $\mathcal{P}_\infty$ with $Y(0) = \mathbf{0}_\infty$ and such that for $n \in \mathbb{N}$, $(R_nY(m))$ is a Markov chain that satisfies for all π with $\#\pi = b$, for each $(b;k_1,\dots,k_r;s)$-collision ε of π and for all $m \in \mathbb{N}_0$:*

$$\mathbb{P}(R_nY(m+1) = \varepsilon | R_nY(m) = \pi) = p_{b;k_1,\dots,k_r;s}$$

if and only if

$$p_{b;k_1,\dots,k_r;s} = \int_\Delta \frac{Q_{k_1,\dots,k_r;s}(x)}{\sum_{j=1}^\infty x_j^2} \Xi(dx) \tag{12}$$

for a finite measure Ξ on Δ, without atom in $\mathbf{0}$, which satisfies

$$\int_\Delta 1 \Big/ \sum_{j=1}^\infty x_j^2 \, \Xi(dx) \leq 1 \tag{13}$$

In this case, the measure Ξ is uniquely determined.

In the demonstration we simply reduce the discrete time case to the continuous time case.

Proof. 1. Necessary condition and uniqueness:

Let Y be such a discrete time process. Let $(N_t, t \geq 0)$ be a Poisson process with parameter 1, independent of Y. Define $\Pi(t) := Y(N_t)$, $t \geq 0$. Then Π is a standard exchangeable coalescent with jump rates $\{p_{b;k_1,\ldots,k_r;s}\}$. Hence there exists a unique finite measure Ξ on Δ such that the $p_{b;k_1,\ldots,k_r;s}$ are given by (12).

Let λ_b be the total collision rate of a Ξ-coalescent with b blocks, i.e.

$$\begin{aligned}\lambda_b &= \sum_{r=1}^{\lfloor b/2 \rfloor} \sum_{\{k_1,\ldots,k_r\}} N(b;k_1,\ldots,k_r;s)\lambda_{b;k_1,\ldots,k_r;s} \\ &= \sum_{r=1}^{\lfloor b/2 \rfloor} \sum_{\{k_1,\ldots,k_r\}} N(b;k_1,\ldots,k_r;s)p_{b;k_1,\ldots,k_r;s} \leq 1\end{aligned}$$

where $N(b;k_1,\ldots,k_r;s)$ is the number of $(b;k_1,\ldots,k_r;s)$-partitions in $\mathcal{P}_b$, and $\lfloor x \rfloor$ is the largest integer that is smaller than x. We necessarily have $\lambda_b \leq 1$ for all b. Let μ and $c \geq 0$ be associated to Ξ like in the Poissonian construction. We have

$$\begin{aligned}\lambda_b = \mu(\{\eta : R_b\eta \neq \mathbf{0}_b\}) &\xrightarrow{b\to\infty} \int_\Delta \frac{P^x(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\})}{\sum_{j=1}^\infty x_j^2}\Xi_0(dx) + c\sum_{i=1}^\infty \sum_{j=i+1}^\infty \delta_{\kappa(i,j)}(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\}) \\ &= \int_\Delta \frac{1}{\sum_{j=1}^\infty x_j^2}\Xi_0(dx) + c\sum_{i=1}^\infty \sum_{j=i+1}^\infty \delta_{\kappa(i,j)}(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\})\end{aligned}$$

For this expression to be ≤ 1, it is necessary that Ξ has no atom in $\mathbf{0}$ and satisfies (13).

2. Sufficient condition:

Let Ξ be a finite measure on Δ that has no atom in $\mathbf{0}$ and that satisfies (13). Let $(\Pi(t))$ be a standard Ξ-coalescent, given by the Poissonian construction. Let $(e(t))$ and μ be as in the Poissonian construction. We have

$$\mu(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\}) = \int_\Delta \frac{P^x(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\})}{\sum_{j=1}^\infty x_j^2}\Xi(dx) \leq 1$$

If we define $T_0 := 0$, $T_k := \inf\{t > T_{k-1} : e(t) \in \mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\}\}$, $k \geq 1$, we obtain a sequence $0 = T_0 < T_1 < \ldots$. Let $(I_m : m \in \mathbb{N}_0)$ be an i.i.d. sequence of Bernoulli variables, independent of e, such that $\mathbb{P}(I_m = 1) = \mu(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\})$. Let $S_m := I_1 + \cdots + I_m$. We define a discrete time Markov process Y by setting $Y(m) := \Pi(T_{S_m})$.

Let $n \in \mathbb{N}$, let $\pi \in \mathcal{P}_n$ with b blocks, and let ε be a $(b;k_1,\ldots,k_r;s)$-collision of π, $\varepsilon = \mathrm{Coag}(\pi,\eta)$ with $\eta \in \mathcal{P}_b$. Using the strong Markov property of the Poisson point process e and the property (57) from Appendix A, we obtain

$$\begin{aligned}&\mathbb{P}(R_nY(m+1) = \varepsilon | R_nY(m) = \pi) = \mathbb{P}(I_{m+1} = 1)\mathbb{P}(R_n\Pi(T_{S_{m+1}}) = \varepsilon | R_n\Pi(T_{S_m}) = \pi) \\ &= \mu(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\})\mathbb{P}(R_be(T_{S_{m+1}}) = \eta) = \mu(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\})\frac{p_{b;k_1,\ldots,k_r;s}}{\mu(\mathcal{P}_\infty \backslash \{\mathbf{0}_\infty\})} \\ &= p_{b;k_1,\ldots,k_r;s}\end{aligned}$$

$\square$

Since we saw that for every discrete time exchangeable coalescent with transition probabilities $p_{b;k_1,\dots,k_r;s}$ there exists a continuous time exchangeable coalescent with jump rates $p_{b;k_1,\dots,k_r;s}$, we know that the $p_{b;k_1,\dots,k_r;s}$ must also satisfy the recursion (11).

2.3 Exchangeable Coalescents and Flows of Bridges

2.3.1 Bridges and Exchangeable Partitions

In this chapter we present an interesting correspondance between exchangeable coalescents and flows of bridges that was established by Bertoin and Le Gall (2003).

Definition 2.33. *A **bridge** is a stochastic process $(B(r) : r \in [0,1])$ such that*

1. *$B(0) = 0, B(1) = 1$, B has increasing càdlàg paths.*
2. *For all $n \in \mathbb{N}$: $(B(1/n) - B(0), B(2/n) - B(1/n), \dots, B(1) - B(1-1/n))$ is an exchangeable vector.*

The general classification of processes with exchangeable increments was given by Kallenberg (1973), Theorem 2.1. In our setting this result can be expressed as follows:

Proposition 2.34 (Kallenberg)**.** *$(B(r) : r \in [0,1])$ is a bridge if and only if there is a random variable X with values in Δ and an i.i.d. sequence $(U_i)_{i\in\mathbb{N}}$ of uniform variables on $[0,1]$, independent of X, such that $(B(r) : r \in [0,1])$ has the same distribution as*

$$\left(\left(1 - \sum_{j=1}^{\infty} X_j\right) r + \sum_{j=1}^{\infty} X_j \mathbb{1}_{\{U_j \le r\}} : r \in [0,1]\right)$$

In the following we will always assume that B is of this form.

We can associate an exchangeable partition to each flow of bridges. We define the càdlàg inverse of B:

$$B^{-1}(s) := \inf\{r \in [0,1] : B(r) > s\}, s \in [0,1) \text{ et } B^{-1}(1) := 1.$$

The lengths of the constant intervalls of B^{-1} correspond exactly to the jump sizes of B. Let $(V_i)_{i\in\mathbb{N}}$ be an i.i.d. sequence of uniform random variables on $[0,1]$. We define a partition $\pi(B)$ such that

$$i \overset{\pi(B)}{\sim} j \text{ if and only if } B^{-1}(V_i) = B^{-1}(V_j)$$

In what follows we suppose that the sequence (V_i) to define $\pi(B)$ is always the same, for each choice of B. By combining Theorem 36 of Pitman (1999) with Theorem 2.3 of Kallenberg (1973) we obtain:

Proposition 2.35. *Let (B^n) be a sequence of bridges with respective jump sizes $(X_i^n)_{i\in\mathbb{N}} \in \Delta$, and let B be a bridge with jump sizes $X \in \Delta$. Then the following conditions are equivalent:*

1. *$\pi(B^n) \overset{n\to\infty}{\longrightarrow} \pi(B)$ in distribution on $\mathcal{P}_\infty$*
2. *$X^n \overset{n\to\infty}{\longrightarrow} X$ in distribution on Δ*
3. *$B^n \overset{n\to\infty}{\longrightarrow} B$ in distribution on the space $D([0,1],[0,1])$ of càdlàg functions on $[0,1]$ with values in $[0,1]$, equipped with the Skorohod topology.*

Remark. 1. If B and B' are independent bridges, then $B \circ B'$ is a bridge as well:

The only property that is not obvious is the exchangeability of the increments. Let $n \in \mathbb{N}$ and let $f : \mathbb{R}^n \to \mathbb{R}$ be a bounded measurable function. By conditioning on B' and by using the independence of B and B' we obtain

$$\begin{aligned}&\mathbb{E}(f(B \circ B'(1/n) - B \circ B'(0/n), \ldots, B \circ B'(n/n) - B \circ B'((n-1)/n))) \\ &\quad = \mathbb{E}(\phi(B'(0), B'(1/n), \ldots, B'(n)))\end{aligned}$$

with $\phi(t_0, \ldots, t_n) = \mathbb{E}(f(B(t_1) - B(t_0), \ldots, B(t_n) - B(t_1)))$. But B has exchangeable increments, so ϕ only depends on $(t_1 - t_0, \ldots, t_n - t_{n-1})$. Let ψ be such that $\psi(t_1 - t_0, \ldots, t_n - t_{n-1}) = \phi(t_0, \ldots, t_n)$. Then we have

$$\begin{aligned}&\mathbb{E}(f(B \circ B'(1/n) - B \circ B'(0/n), \ldots, B \circ B'(n/n) - B \circ B'((n-1)/n))) \\ &\quad = \mathbb{E}(\psi(B'(1/n) - B'(0), \ldots, B'(1) - B'(1 - 1/n)))\end{aligned}$$

Since B' has exchangeable increments and since ψ is a bounded measurable function, we obtain the exchangeability of the increments of $B \circ B'$.

2. $(B \circ B')^{-1} = B'^{-1} \circ B^{-1}$

The following result is Corollary 1 of Bertoin and Le Gall (2003). We do not give the proof here, but it is not at all trivial.

Proposition 2.36. *Let $k \geq 2$, and let $B^1, \ldots, B^k$ be independent bridges. We define*

$$C^l := B^1 \circ \cdots \circ B^l, l = 1, \ldots, k$$

Then conditionally on $(\pi(C^1), \ldots, \pi(C^{l-1}))$, $\pi(C^l)$ has the same distribution as the coagulation of $\pi(C^{l-1})$ by an independent partition that is distributed like $\pi(B^l)$.

2.3.2 Flows of Bridges

Definition 2.37. *A family $(B_{s,t} : -\infty < s \leq t < \infty)$ of bridges is a **flow of bridges** if*

1. *For each $s \leq t \leq u$: $B_{s,u} = B_{s,t} \circ B_{t,u}$.*
2. *The distribution of $B_{s,t}$ does not depend on $t - s$.*
3. *For $-\infty < t_1 < \cdots < t_n < \infty$, the bridges $B_{t_1,t_2}, \ldots, B_{t_{n-1},t_n}$ are independent.*
4. *$B_{s,s} = \mathrm{Id}$ for all s and $B_{0,t} \xrightarrow{t \to 0} \mathrm{Id}$ in probability in the Skorohod topology.*

We can associate an exchangeable coalescent to each flow of bridges:

Proposition 2.38. *Let B be a flow of bridges. We define for each $t \geq 0$ $\Pi_t := \pi(B_{0,t})$. Then $(\Pi_t : t \geq 0)$ has a càdlàg modification that is a standard exchangeable coalescent.*

Proof. Let $0 \leq t_0 < \cdots < t_n$. By Proposition 2.36, conditionally on $(\Pi_{t_0}, \ldots, \Pi_{t_{n-1}})$, Π_{t_n} has the same distribution as the coagulation of $\Pi_{t_{n-1}}$ by an independent partition that is distributed like $\pi(B_{0,t_n - t_{n-1}}) = \Pi_{t_n - t_{n-1}}$. So Π is a Markov process with semi-group

$$P_t f(\eta) = \mathbb{E}(f(\mathrm{Coag}(\eta, \pi(B_{0,t}))))$$

This is a Feller semi-group. This is shown exactly as in the proof of Proposition 2.30: We use the continuity of Coag from $\mathcal{P}_\infty \times \mathcal{P}_\infty$ to $\mathcal{P}_\infty$ and the fact that $B_{0,t}$ converges in probability to the identity when t tends to 0. Then we obtain the convergence of $P_t f(\eta)$ to $f(\eta)$ when $t \to 0$ from Proposition 2.35. Since Π is a Feller process, it has a càdlàg modification (cf. Theorem (2.7) in Chapter III of Revuz and Yor (1999)).

It remains to show that for each $n \in \mathbb{N}$ $R_n\Pi$ is a Markov process such that each $(b; k_1, \dots, k_r; s)$-collision has the same rate $\lambda_{b;k_1,\dots,k_r;s}$. The Markov property is easily obtained with the property $R_n \operatorname{Coag}(\eta, \varepsilon) = \operatorname{Coag}(R_n\eta, R_n\varepsilon)$ of the coagulation operator. Like this we see that $R_n\Pi$ has the semi-group

$$P_t^n f(\eta) = \mathbb{E}(f(\operatorname{Coag}(\eta, R_n\Pi_t)))$$

Since Π_t is an exchangeable partition for each t, each $(b; k_1, \dots, k_r; s)$-collision has the same rate. Therefore the càdlàg modification of Π is an exchangeable coalescent. $\square$

We would like to establish a correspondance between flows of bridges and exchangeable coalescents. It remains to show the injectivity and the surjectivity of the map $(B_{s,t}) \mapsto (\pi(B_{0,t}))$. More precisely we would like to show:

1. Let B and B' be two flows of bridges with the same finite-dimensional distributions. Then $(\pi(B_{0,t}))_{t\geq 0}$ has the same finite-dimensional distributions as $(\pi(B'_{0,t}))_{t\geq 0}$.

2. Let Π be a standard exchangeable coalescent. Then there exists a flow of bridges B such that Π and $(\pi(B_{0,t}))_{t\geq 0}$ have the same finite-dimensional distributions.

The first statement is more or less obvious: This is just Proposition 2.35 and an application of the stationarity and independence properties of flows of bridges.

We will show the second statement with a Poissonian construction. Let $(u_i)_{i\in\mathbb{N}} \in [0,1]^{\mathbb{N}}$ and let $(x_i)_{i\in\mathbb{N}} \in \Delta$. We define

$$b_{(u_i),(x_i)}(r) := \left(1 - \sum_{i=1}^{\infty} x_i\right) r + \sum_{i=1}^{\infty} x_i \mathbb{1}_{\{r \geq u_i\}}$$

Note that if (u_i) is an i.i.d. sequence of uniform variables on $[0,1]$, then $b_{(u_i),(x_i)}$ is a bridge.

Let ν be a finite measure on Δ with $\nu(\{\mathbf{0}\}) = 0$. Let $U^{\otimes\mathbb{N}} := U \otimes U \otimes \dots$ on $[0,1]^{\mathbb{N}}$ (where U is the uniform distribution on $[0,1]$). Let $(e(t) : t \in \mathbb{R})$ be a Poisson point process of intensity $U^{\otimes\mathbb{N}} \otimes \nu$ on $[0,1]^{\mathbb{N}} \times \Delta$. A Poisson point process with real-valued index t (instead of positive t) is defined exactly as an usual Poisson point process, just that in this case we consider a Poisson random measure on $\mathbb{R} \times E$ rather than $\mathbb{R}_+ \times E$. Since ν and $U^{\otimes\mathbb{N}}$ are finite measures, e a.s. only has a finite number of points on $(s,t]$ for all finite $s \leq t$. Let

$$(t_1, (u_i^1), (x_i^1)), \dots, (t_k, (u_i^k), (x_i^k))$$

be those points with $s < t_1 < \dots < t_k \leq t$. We define

$$B_{s,t} := b_{(u_i^1),(x_i^1)} \circ \dots \circ b_{(u_i^k),(x_i^k)} \tag{14}$$

If e has no points on $(s,t]$, we define $B_{s,t} := Id$.

Proposition 2.39. *$(B_{s,t} : -\infty < s \leq t < \infty)$ is a flow of bridges*

Proof. All the properties of flows of bridges are trivially satisfied. The only thing we need to show is that $B_{s,t}$ actually is a bridge for all $s \leq t$. We argue conditionally on the number K of points of e on $(s,t]$. Conditionally on $K = k$, the variables (u_i^1), ..., (u_i^k), (x_i^1), ..., (x_i^k) are independent, and $(u_i^j)_i$ has the distribution $U^{\otimes \mathbb{N}}$ for all $j \leq k$. Thus the processes $b_{(u_i^j)_i,(x_i^j)_i}$ are independent bridges, and conditionally on $K = k$,

$$B_{s,t} = b_{(u_i^1),(x_i^1)} \circ \cdots \circ b_{(u_i^k),(x_i^k)}$$

is a bridge. Since a mixture of bridge laws preserves the bridge properties, $B_{s,t}$ is a bridge. $\square$

We consider $\Pi_t = \pi(B_{0,t})$. Let $t > 0$ be a jump time of Π, corresponding to the point $(t, (u_i), (x_i))$ of e. Then Π_t is the coagulation of Π_{t-} by $\pi(b_{(u_i),(x_i)})$, and $\pi(b_{(u_i),(x_i)})$ is an exchangeable partition, independent of Π_{t-}, with distribution

$$\int_\Delta P^x(d\pi) \frac{\nu(dx)}{\nu(\Delta)}$$

If we compare this formula with the formula (4) of the Poisson construction of exchangeable coalescents, we see that Π is a standard $\sum_{i=1}^\infty x_i^2 \nu(dx)$-coalescent (since Π is càdlàg by construction).

Let $\Xi = \Xi_0 + c\delta_{\mathbf{0}}$ be a finite measure on Δ such that $\Xi_0(\{\mathbf{0}\}) = 0$ and $c \geq 0$. Then we can find a sequence (Ξ_n) of finite measures on $\Delta \backslash \{\mathbf{0}\}$ such that $(\sum_{i=1}^\infty x_i^2)\Xi_n(dx)$ converges weakly to Ξ. We can take for example a sequence $x^n = (x_i^n)_{i \in \mathbb{N}} \in \Delta$ converging to $\mathbf{0}$ in Δ, and then define for $n \in \mathbb{N}$:

$$\Xi_n(dx) := \frac{c}{\sum_{i=1}^\infty (x_i^n)^2} \delta_{(x_i^n)_i} + \mathbb{1}_{\{\sum_{i=1}^\infty x_i^2 \geq 1/n\}}(x) \frac{\Xi(dx)}{\sum_{i=1}^\infty x_i^2}$$

Proposition 2.40. *Let Ξ be a finite measure on Δ. Let Ξ_n be a sequence of finite measures on Δ such that $\sum_{i=1}^\infty x_i^2 \Xi_n(dx)$ converges weakly to Ξ. Let for $n \in \mathbb{N}$ B^n be the flow of bridges associated to $\sum_{i=1}^\infty x_i^2 \Xi_n(dx)$. Then the finite-dimensional distributions of B^n converge weakly to the finite-dimensional distributions of a flow of bridges $(B_{s,t} : s \leq t)$ such that the associated exchangeable coalescent is a standard Ξ-coalescent. In particular, for each standard exchangeable coalescent Π we can find a flow of bridges B such that $(\pi(B_{0,t}))_{t \geq 0}$ and $(\Pi_t)_{t \geq 0}$ have the same finite-dimensional distributions.*

Proof. For each finite measure Ξ on Δ let Q^Ξ be the distribution on $D([0,\infty), \mathcal{P}_\infty)$ of a standard Ξ-coalescent. We will show later that $\Xi \mapsto Q^\Xi$ is a continuous map (cf. Proposition 3.5). So we obtain that the standard $\sum_{i=1}^\infty x_i^2 \Xi_n(dx)$-coalescent converges in distribution to the standard Ξ-coalescent. Without loss of generality we suppose that all the B^n are given by the Poisson construction. Then $(\pi(B_{0,t}^n))$ is a standard $\sum_{i=1}^\infty x_i^2 \Xi_n(dx)$-coalescent and thus $(\pi(B_{0,t}^n))$ converges in distribution to a standard Ξ-coalescent $(\Pi_t)_{t \geq 0}$. So for each t, $\pi(B_{0,t}^n)$ converges in distribution. We obtain the convergence in distribution of $B_{0,t}^n$ from Proposition 2.35. Denote the limit by B_t. Then for all $t \geq 0$: $\pi(B_t)$ has the same distribution as Π_t. Let $t, s > 0$ and let B_s' be a copy of B_s, independent of B_t. Then

$$\pi(B_t \circ B_s') \simeq \mathrm{Coag}(\pi(B_t), \pi(B_s'))$$

where $\simeq$ denotes equality in law. But $\pi(B_s') \simeq \Pi_s$ and therefore $\pi(B_t \circ B_s')$ has the same distribution as $\Pi_{t+s} \simeq \pi(B_{t+s})$. Another application of Proposition 2.35 yields $B_t \circ B_s' \simeq B_{t+s}$.

Note that $D([0,1],[0,1])$ equipped with the Skorohod topology is a Polish space. So we can construct a family of bridges $(B_{s,t}, -\infty < s \leq t < \infty)$ with the Daniell-Kolmogorov extension theorem such that for all $s \leq t$ $B_{s,t} \simeq B_{t-s}$ and such that for $-\infty < t_1 < \cdots < t_n < \infty$, $B_{t_1,t_2}, \ldots, B_{t_{n-1},t_n}$ are independent. This family is a flow of bridges: $B_{0,0} = \mathrm{Id}$ is evident since $\pi(B_0) = \Pi_0 = \mathbf{0}_\infty$. The convergence in probability of $B_{0,t}$ to the identity when $t \to \infty$ is obtained from the continuity in probability of Π. Π is continuous in probability since $R_n\Pi$ is a jump-hold process without fixed jump times and because of the definition of the topology on $\mathcal{P}_\infty$.

So we have the convergence of $B^n_{s,t}$ to $B_{s,t}$ for all fixed s, t and this implies the distribution of finite-dimensional distributions: For $s_1, t_1, \ldots, s_m, t_m$ we cut the intervalls $(s_i, t_i]$ into disjoint or equal intervalls. So we obtain the convergence in distribution of $(B^n_{s_1,t_1}, \ldots, B^n_{s_m,t_m})$ to $(B_{s_1,t_1}, \ldots, B_{s_m,t_m})$ by using these intervall decompositions and the independence properties of flows of bridges. □

2.4 Fleming-Viot Process

This section was not included in Perkowski (2009). We present a measure-valued process that was introduced by Fleming and Viot (1979). Let E be a compact metric space, and let $\mathcal{M}_1(E)$ be the space of probability measures on E, equipped with the topology of weak convergence.

For $f : E^p \to \mathbb{R}$ bounded and measurable we define

$$< f, \mu^{\otimes p} > := \int_{E^p} f(x_1, \ldots, x_p) \mu^{\otimes p}(dx_1, \ldots, dx_p)$$

Let $\mathcal{D} := \{\Phi_f : \mathcal{M}_1(E) \to \mathbb{R}, \Phi_f(\mu) =< f, \mu^{\otimes p} > \text{ for some } p \in \mathbb{N}, f \in C(E^p)\}$ We define a linear operator A:

$$A : C(\mathcal{M}_1(E)) \supset \mathcal{D} \to C(\mathcal{M}_1(E))$$

such that for $\Phi_f(\mu) =< f, \nu^{\otimes p} >$:

$$A\Phi_f(\mu) = \sum_{1 \leq i < j \leq p} \int [f(x_1, \ldots, x_i, \ldots, x_i, \ldots, x_p) - f(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_p)] \mu^{\otimes p}(dx)$$

Definition 2.41. *Let ν be a probability on $\mathcal{M}_1(E)$. A* **Fleming-Viot process** *starting with distribution ν is an $\mathcal{M}_1(E)$-valued process $(\rho_t : t \geq 0)$ that is a solution to the (A, ν)-martingale problem.*

Existence and uniqueness of the solution to that martingale problem were shown in Fleming and Viot (1979). We do not give the proof here because we will show existence and uniqueness of solutions to a more general class of martingale problems later. It it shown in Kurtz (1981), Theorem 10.1, that the Fleming-Viot process arises as the limit for large populations in the Wright-Fisher model if the time is suitably rescaled. Essentially the same proof can also be found in Ethier and Kurtz (1986), Theorem 4.1 of Chapter 10.

3 Weak Convergence Results

Before we continue, we need to establish some convergence results on which we will rely heavily in what follows.

3.1 Convergence of Rescaled Markov Chains

Assume we are in the following setting:

Let (E, d) be a compact metric space, equipped with its Borel σ-algebra $\mathcal{E}$. Let

$$A : C(E) \supseteq \mathcal{D}(A) \to C(E)$$

be an operator on $C(E)$. Let ν be a probability measure on $(E, \mathcal{E})$. We want to approximate a solution to the (A, ν)-martingale problem. That is, we want to find a sequence $(X_N)_{N\in\mathbb{N}}$ of processes in $D([0, \infty), E)$, such that X_N converges in distribution in the Skorohod-topology to some $X \in D([0, \infty), E)$, and X is a solution to the (A, ν) martingale problem.

We want to show convergence of processes of the following type: Let for every $n \in \mathbb{N}$ $(Y_N(m) : m \in \mathbb{N}_0)$ be a discrete time homogenous Markov process with values in some compact metric space E_N. Let P_N denote its transition probability, i.e. for all $x \in E_N$ and for all Borel sets B of E_N we have

$$\mathbb{P}(Y_N(1) \in B | Y_N(0) = x) = P_N(x, B)$$

Define the operator T_N on $B(E_N)$, the space of bounded measurable functions on E_N, equipped with the topology of uniform convergence, as follows:

$$T_N f(x) := \int_{E_N} f(y) P_N(x, dy)$$

Let (c_N) be a sequence of strictly positive numbers. Define

$$A_N : B(E_N) \to B(E_N), \quad A_N F := \frac{1}{c_N} (T_N - I) F$$

Let for all N $\pi_N : E_N \to E$ be a measurable map. We want to show convergence of $(X_N(t) := \pi_N Y_N(\lfloor t/c_N \rfloor) : t \geq 0)$ to X.

We are now able to formulate our convergence theorem. This result is shown in Ethier and Kurtz (1986) in a more general setting. But the proof there is scattered over many chapters, and in our setting we can give a simpler and more direct proof. Nonetheless this proof is using some of the techniques from Ethier and Kurtz (1986)

Theorem 3.1. *Let (Y_N) be as above. We make the following assumptions:*

- *There is uniqueness for the (A, ν)-martingale problem,*
- *$\mathcal{D}(A)$ contains an algebra $\mathcal{A}$ that contains a constant function $\neq 0$ and that separates points,*
- *$c_N \to 0$ when $N \to \infty$,*
- *The distribution of $\pi_N Y_N(0)$ converges weakly to ν when $N \to \infty$,*
- *For every $f \in \mathcal{D}(A)$ there exists a sequence (f_N) with $f_N \in B(E_N)$ such that*

$$\sup_{y\in E_N} |f(\pi_N y) - f_N(y)| \to 0, \quad N \to \infty \tag{15}$$

and

$$\sup_{y\in E_N} |Af(\pi_N y) - A_N f_N(y)| \to 0, \quad N \to \infty \tag{16}$$

Then $(X_N(t) := \pi_N Y_N(\lfloor t/c_N \rfloor) : t \geq 0)$ converges in distribution on $D([0,\infty),E)$ to the unique solution X of the (A,ν)-martingale problem.

Proof. It suffices to show that the sequence (X_N) is tight in $D([0,\infty),E)$, and that every cluster point of the sequence is a solution to the (A,ν)-martingale problem.

1. First we show the tightness of $(f(X_N))$ in $D([0,\infty),\mathbb{R})$ for all $f \in \mathcal{A}$. Let $f \in \mathcal{A}$ and let (f_N) be a sequence for f satisfying (15) and (16). Further let (g_N) be a sequence for f^2 satisfying (15) and (16). f^2 is in $\mathcal{D}(A)$ because $\mathcal{A}$ is an algebra. Let $\mathcal{G}_k^N := \sigma(Y_N(m) : m \leq k)$ be the canonic filtration for Y_N. We set $\mathcal{F}_t^N := \mathcal{G}_{\lfloor t/c_N \rfloor}^N$. Since Y_N is a discrete time Markov process, for every bounded measurable $\varphi : E_N \to \mathbb{R}$ we know that

$$\varphi(Y_N(m)) - \sum_{i=0}^{m-1}(T_N - I)\varphi(Y_N(i)), \quad m \in \mathbb{N}_0$$

is a martingale with respect to the filtration $(\mathcal{G}_\cdot^N)$. We define the following sequences of processes:

$$\begin{aligned}
\varphi_N(t) &:= f_N(Y_N(\lfloor t/c_N \rfloor)) + c_N\,(t/c_N - \lfloor t/c_N \rfloor)\,A_N f_N(Y_N(\lfloor t/c_N \rfloor)) \\
&\qquad - \int_0^t A_N f_N(Y_N(\lfloor s/c_N \rfloor))ds \\
&= f_N(Y_N(\lfloor t/c_N \rfloor)) + c_N\,(t/c_N + \lfloor t/c_N \rfloor)\,A_N f_N(Y_N(\lfloor t/c_N \rfloor)) \\
&\qquad - c_N \sum_{i=0}^{\lfloor t/c_N \rfloor - 1} A_N f_N(Y_N(i)) - c_N\,(t/c_N - \lfloor t/c_N \rfloor)\,A_N f_N(Y_N(\lfloor t/c_N \rfloor)) \\
&= f_N(Y_N(\lfloor t/c_N \rfloor)) - \sum_{i=0}^{\lfloor t/c_N \rfloor - 1} (T_N - I) f_N(Y_N(i))
\end{aligned}$$

and

$$\begin{aligned}
\psi_N(t) &:= g_N(Y_N(\lfloor t/c_N \rfloor)) + c_N\,(t/c_N - \lfloor t/c_N \rfloor)\,A_N g_N(Y_N(\lfloor t/c_N \rfloor)) \\
&\qquad - \int_0^t A_N g_N(Y_N(\lfloor s/c_N \rfloor))ds \\
&= g_N(Y_N(\lfloor t/c_N \rfloor)) - \sum_{i=0}^{\lfloor t/c_N \rfloor - 1} (T_N - I) g_N(Y_N(i))
\end{aligned}$$

φ_N and ψ_N are thus both martingales with respect to the filtration $(\mathcal{F}_\cdot^N)$. We have

$$\begin{aligned}
\mathbb{E}[(f(X_N(t+s)) - f(X_N(t)))^2 | \mathcal{F}_t^N] &= \mathbb{E}[f^2(X_N(t+s)) - f^2(X_N(t)) | \mathcal{F}_t^N] \\
&\quad - 2f(X_N(t))\mathbb{E}[f(X_N(t+s)) - f(X_N(t)) | \mathcal{F}_t^N] \\
&= \mathbb{E}[f^2(X_N(t+s)) - \psi_N(t+s) - (f^2(X_N(t)) - \psi_N(t)) | \mathcal{F}_t^N] \\
&\quad - 2f(X_N(t))\mathbb{E}[f(X_N(t+s)) - \varphi_N(t+s) - (f(X_N(t)) - \varphi_N(t)) | \mathcal{F}_t^N]
\end{aligned}$$

We examine the term $f(X_N(t)) - \varphi_N(t)$:

$$\begin{aligned}
f(X_N(t)) - \varphi_N(t) &= [f(\pi_N Y_N(\lfloor t/c_N \rfloor)) - f_N(Y_N(\lfloor t/c_N \rfloor))] \\
&\quad - c_N\,(t/c_N - \lfloor t/c_N \rfloor)\,A_N f_N(Y_N(\lfloor t/c_N \rfloor)) + \int_0^t A_N f_N(Y_N(\lfloor u/c_N \rfloor))du \qquad (17)
\end{aligned}$$

The first term on the right hand side will converge to 0 by (15). For large enough N, the second term is bounded by $c_N(1+\epsilon)\sup_{y\in E} Af(y)$ by (16), which also tends to 0 since c_N converges to 0 and since Af is bounded. Only the last term might pose a problem. But that one we can combine with the corresponding term from $f(X_N(t+s)) - \varphi_N(t+s)$. Therefore we obtain

$$
\begin{aligned}
&\mathbb{E}[(f(X_N(t+s)) - f(X_N(t)))^2|\mathcal{F}_t^N] \leq 2\sup_{y\in E_N}\left|f^2(\pi_N y) - g_N(y)\right| \\
&+ 2c_N\left(\sup_{y\in E_N}\left|Af^2(\pi_N y) - A_N g_N(y)\right| + ||Af^2||\right) \\
&+ \int_t^{t+s}\left|\sup_{y\in E_N}\left|Af^2(\pi_N y) - A_N g_N(y)\right| + ||Af^2||\right| du \\
&+ 4||f||\sup_{y\in E_N}|f(\pi_N y) - f_N(y)| \\
&+ 4||f||c_N\left(\sup_{y\in E_N}|Af(\pi_N y) - A_N f_N(y)| + ||Af||\right) \\
&+ 2||f||\int_t^{t+s}\left|\sup_{y\in E_N}|Af(\pi_N y) - A_N f_N(y)| + ||Af||\right| du
\end{aligned}
$$

$||\cdot||$ denotes the supremum norm on $C(E)$. For $s \leq \delta$ we obtain an inequality where the right hand side does not depend on s or t any more:

$$
\begin{aligned}
&\mathbb{E}[(f(X_N(t+s)) - f(X_N(t)))^2|\mathcal{F}_t^N] \leq 2\sup_{y\in E_N}\left|f^2(\pi_N y) - g_N(y)\right| + (2c_N+\delta)||Af^2|| \\
&+ (2cN+\delta)\sup_{y\in E_N}\left|Af^2(\pi_N y) - A_N g_N(y)\right| + 4||f||\sup_{y\in E_N}|f(\pi_N y) - f_N(y)| \\
&+ 2||f||(2c_N+\delta)\sup_{y\in E_N}|Af(\pi_N y) - A_N f_N(y)| + 2||f||(2c_N+\delta)||Af||
\end{aligned}
$$

Since f is bounded we can apply Lemma 3.2 and we obtain that for any $\mathcal{F}^N$-stopping times $T \leq S \ \leq T+\delta$

$$
\begin{aligned}
&\mathbb{E}((f(X_N(S)) - f(X_N(T)))^2) \leq 6\{2\sup_{y\in E_N}\left|f^2(\pi_N y) - g_N(y)\right| + (2c_N+2\delta)||Af^2|| \\
&+ (2cN+2\delta)\sup_{y\in E_N}\left|Af^2(\pi_N y) - A_N g_N(y)\right| + 4||f||\sup_{y\in E_N}|f(\pi_N y) - f_N(y)| \\
&+ 2||f||(2c_N+2\delta)\sup_{y\in E_N}|Af(\pi_N y) - A_N f_N(y)| + 2||f||(2c_N+2\delta)||Af||\}
\end{aligned}
$$

Denote by $\mathcal{S}^N$ the set of all $\mathcal{F}^N$-stopping times. We apply Markov's inequality and obtain for any $\lambda > 0$:

$$
\limsup_{N\to\infty}\sup_{S,T\in\mathcal{S}^N, T\leq S\leq T+\delta}\mathbb{P}(|f(X_N(S)) - f(X_N(T))| > \lambda) \leq \frac{12\delta||Af^2|| + 24\delta||f||\times||Af||}{\lambda}
$$

and therefore

$$
\lim_{\delta\to 0}\limsup_{N\to\infty}\sup_{S,T\in\mathcal{S}^N, T\leq S\leq T+\delta}\mathbb{P}(|f(X_N(S)) - f(X_N(T))| > \lambda) = 0
$$

Since f is bounded, $\sup_{t\geq 0} f(X_N(t))$ is obviously tight in $\mathbb{R}$. Therefore we can apply Aldous' criterion (cf. Theorem 4.5 in Chapter VI of Jacod and Shiryaev (2002)) to obtain the tightness of $(f(X_N))$ in $D([0,\infty),\mathbb{R})$.

2. Since $\mathcal{A}$ is an algebra that separates points, it is dense in the uniform topology on $C(E)$ by the Stone-Weierstrass theorem. E is compact, so X_N automatically satisfies the compact containment condition:

$$\inf_{N\in\mathbb{N}} \mathbb{P}(X_N(t) \in E, t \geq 0) = 1$$

So (X_N) is a sequence of processes that satisfies the compact containment condition and such that $(f(X_N))$ is tight in $D([0,\infty),\mathbb{R})$ for all f in a dense subset of $C(E)$. We can apply Theorem 9.1 in Chapter 3 of Ethier and Kurtz (1986) to obtain the tightness of (X_N) in $D([0,\infty),E)$.

3. It remains to show that every cluster point of the sequence (X_N) is a solution to the (A,ν)-martingale problem. Since the distribution of $X_N(0)$ converges weakly to ν, every cluster point X must satisfy $X(0) \simeq \nu$. Therefore it suffices to show that for every cluster point X of (X_N) and for every $f \in \mathcal{D}(A)$

$$f(X(t)) - \int_0^t Af(X_s)ds, \quad t \geq 0$$

is a martingale. By a version of the monotone class theorem (cf. Corollary 4.4 of the appendix of Ethier and Kurtz (1986)) it suffices to show that for every $0 \leq t_1 < \cdots < t_n < t_{n+1} < \infty$ and for all bounded continuous functions $h_1,\ldots,h_n$ on E we have

$$\mathbb{E}\left[\left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} Af(X(s))ds\right)\prod_{k=1}^n h_k(X(t_k))\right] = 0$$

First we consider only times t_i with $\mathbb{P}(X_{t_i-} = X_{t_i}) = 1$. For such t_i we have

$$\begin{aligned}
&\mathbb{E}\left[\left(f(X(t_{n+1})) - f(X(t_n)) - \int_{t_n}^{t_{n+1}} Af(X_s)ds\right)\prod_{k=1}^n h_k(X(t_k))\right]\\
&= \lim_{N\to\infty}\mathbb{E}\left[\left(f(X_N(t_{n+1})) - f(X_N(t_n)) - \int_{t_n}^{t_{n+1}} Af(X_N(s))ds\right)\prod_{k=1}^n h_k(X_N(t_k))\right]
\end{aligned}$$

Since we know that for all N φ_N defined as above is a martingale with respect to the filtration $\mathcal{F}^N_\cdot$, we can insert $-\varphi_N(t_{n+1}) + \varphi_N(t_n)$ in the brackets:

$$\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}\left[\left(f(X_N(t_{n+1})) - f(X_N(t_n)) - \int_{t_n}^{t_{n+1}} Af(X_N(s))ds\right)\prod_{k=1}^n h_k(X_N(t_k))\right]\\
&\quad= \lim_{N\to\infty}\mathbb{E}[\{(f(X_N(t_{n+1})) - \varphi_N(t_{n+1})) - (f(X_N(t_n)) - \varphi_N(t_n))\\
&\qquad - \int_{t_n}^{t_{n+1}} Af(X_N(s))ds\}\prod_{k=1}^n h_k(X_N(t_k))]
\end{aligned}$$

By (17) we know

$$\lim_{N\to\infty} f(X_N(t_{n+1})) - \varphi_N(t_{n+1}) = \lim_{N\to\infty}\int_0^{t_{n+1}} A_N f_N(Y_N(\lfloor u/c_N\rfloor))du$$

and therefore we obtain by applying dominated convergence several times

$$\mathbb{E}\left[\left(f(X(t_{n+1}))-f(X(t_n))-\int_{t_n}^{t_{n+1}}Af(X_s)ds\right)\prod_{k=1}^{n}h_k(X(t_k))\right]$$
$$=\lim_{N\to\infty}\mathbb{E}\left[\left(\int_{t_n}^{t_{n+1}}\{A_Nf_N(Y_N(\lfloor s/c_N\rfloor)-Af(\pi_NY_N(\lfloor s/c_N\rfloor))\}ds\right)\prod_{k=1}^{n}h_k(X_N(t_k))\right]$$
$$\overset{(16)}{=}0$$

For general t_i we remark that

$$\{t\geq 0:\mathbb{P}(X_t\neq X_{t-})<1\}$$

is at most countable by Lemma 7.7 of Chapter 3 of Ethier and Kurtz (1986). Thus we can use the right-continuity of X and bounded convergence to obtain the equality for all $0\leq t_1<\cdots<t_{n+1}$.

□

Remark. We can replace the assumption that E is compact by assuming that E is a Polish space and that the X_N satisfy the compact containment condition: For every $\epsilon>0$ and every $T>0$ there is a compact set $K_{T,\epsilon}\subseteq E$ such that

$$\liminf_{N\to\infty}\mathbb{P}\left(X_N(t)\in K_{T,\epsilon},0\leq t\leq T\right)\geq 1-\epsilon$$

At one point we assumed that $\mathcal{A}$ is dense in $C(E)$, which only follows from the Stone-Weierstrass theorem if E is compact. But we only needed this to apply Theorem 9.1 of Chapter 3 of Ethier and Kurtz (1986), and for this theorem we only need the density of $\mathcal{A}$ with respect to the topology of uniform convergence on compact subsets. In the non-compact case case we need to work on $C_b(E)$ rather than on $C(E)$.

We used nowhere in the proof that the E_N are compact. It suffices to assume that they are Polish and to work with $C_b(E_N)$ instead of $C(E_N)$.

In the proof of the following lemma we take some ideas from the proof of Theorem 8.6 in Chapter 3 of Ethier and Kurtz (1986).

Lemma 3.2. *Let $(X_t:t\geq 0)$ be a real-valued stochastic process with globally bounded càdlàg paths, adapted to some filtration $(\mathcal{F}_t)_{t\geq 0}$. Assume X satisfies*

$$\mathbb{E}((X_s-X_t)^2|\mathcal{F}_t)\leq C(\delta)$$

for some function C of δ and for all s and t with $t\leq s\leq t+\delta$. Then

$$\mathbb{E}((X_S-X_T)^2)\leq 6C(2\delta)$$

for all finite $\mathcal{F}_\cdot$-stopping times S and T such that a.s. $T\leq S\leq T+\delta$.

Proof. First we proof that under the assumption we have

$$\mathbb{E}((X_{T+s}-X_T)^2|\mathcal{F}_T)\leq C(\delta)$$

for any $\mathcal{F}_\cdot$-stopping time T and any $s \leq \delta$. Let T be a stopping time that takes only finitely many values, $t_1, \dots, t_n$. Then

$$\mathbb{E}((X_{T+s} - X_T)^2 | \mathcal{F}_T) = \sum_{k=1}^{n} \mathbb{E}(\mathbb{1}_{\{T=t_k\}}(X_{t_k+s} - X_{t_k})^2 | \mathcal{F}_T)$$

But $\mathbb{E}(\mathbb{1}_{\{T=t_k\}}(X_{t_k+s} - X_{t_k})^2 | \mathcal{F}_T) = \mathbb{1}_{\{T=t_k\}}\mathbb{E}((X_{t_k+s} - X_{t_k})^2 | \mathcal{F}_{t_k})$: Let $A \in \mathcal{F}_T$. Then

$$\begin{aligned}\mathbb{E}(\mathbb{1}_A \mathbb{1}_{\{T=t_k\}}(X_{t_k+s} - X_{t_k})^2) &= \mathbb{E}[\mathbb{E}(\mathbb{1}_{\{T=t_k\}\cap A}\mathbb{E}((X_{t_k+s} - X_{t_k})^2 | \mathcal{F}_{t_k})] \\ &= \mathbb{E}[\mathbb{E}(\mathbb{1}_A \mathbb{1}_{\{T=t_k\}}\mathbb{E}((X_{t_k+s} - X_{t_k})^2 | \mathcal{F}_{t_k})]\end{aligned}$$

Therefore

$$\mathbb{E}((X_{T+s} - X_T)^2 | \mathcal{F}_T) = \sum_{k=1}^{n} \mathbb{1}_{\{T=t_k\}}\mathbb{E}((X_{t_k+s} - X_{t_k})^2 | \mathcal{F}_{t_k}) \leq \sum_{k=1}^{n} \mathbb{1}_{\{T=t_k\}} C(\delta) = C(\delta)$$

Now let T be any finite stopping time. Then T can be approached by a sequence of stopping times (T_N) taking only finitely many values and such that $T_N \geq T$ for all N. We use the right-continuity of X, the fact that X is globally bounded, and that we have $\mathcal{F}_{T_N} \supseteq \mathcal{F}_T$ since $T_N \geq T$. Like this we obtain

$$\begin{aligned}\mathbb{E}((X_{T+s} - X_T)^2 | \mathcal{F}_T) &= \lim_{N\to\infty} \mathbb{E}((X_{T_N+s} - X_{T_N})^2 | \mathcal{F}_T) \\ &= \lim_{N\to\infty} \mathbb{E}[\mathbb{E}((X_{T_N+s} - X_{T_N})^2 | \mathcal{F}_{T_N}) | \mathcal{F}_T] \\ &\leq \lim_{N\to\infty} \mathbb{E}(C(\delta) | \mathcal{F}_T) = C(\delta)\end{aligned}$$

Therefore for any stopping time T and any $s \leq \delta$:

$$\mathbb{E}((X_{T+s} - X_T)^2) \leq C(\delta)$$

Now let S and T be suitable stopping times. We have

$$\begin{aligned}(X_S - X_T)^2 &\leq \frac{1}{\delta}\int_{\delta}^{2\delta} 2((X_{T+x} - X_T)^2 + (X_{T+x} - X_S)^2)dx \\ &\leq \frac{2}{\delta}\left(\int_{\delta}^{2\delta} (X_{T+x} - X_T)^2 dx + \int_{0}^{2\delta} (X_{S+x} - X_S)^2 dx\right)\end{aligned}$$

and therefore

$$\mathbb{E}((X_S - X_T)^2) \leq \frac{2}{\delta}\left(\int_{\delta}^{2\delta} C(2\delta)dx + \int_{0}^{2\delta} C(2\delta)dx\right) = 6C(2\delta)$$

□

3.2 Convergence of Markov Processes

We only need to change the proof of Theorem 3.1 a little bit to obtain a convergence result for continuous time Markov processes:

Let (E, d), ν, A, E_N and π_N be as above. Let for every N A_N be a linear operator with domain $\mathcal{D}(A_N) \subseteq B(E_N)$. Let $Y_N \in D([0,\infty), E_N)$ be a solution of the (A_N, ν_N)-martingale problem for some distribution ν_N on E_N.

Proposition 3.3. *We make the following assumptions:*

- *There is uniqueness for the (A, ν)-martingale problem,*
- *$\mathcal{D}(A)$ contains an algebra $\mathcal{A}$ that contains a constant function $\neq 0$ and that separates points,*
- *The distribution of $\pi_N Y_N(0)$ converges weakly to ν when $N \to \infty$,*
- *For every $f \in \mathcal{D}(A)$ there is a sequence (f_N) with $f_N \in \mathcal{D}(A_N)$ such that*

$$\sup_{y \in E_N} |f(\pi_N y) - f_N(y)| \to 0, \quad N \to \infty \tag{18}$$

and

$$\sup_{y \in E_N} |Af(\pi_N y) - A_N f_N(y)| \to 0, \quad N \to \infty \tag{19}$$

Then $(X_N(t) := \pi_N Y_N(t) : t \geq 0)$ converges in distribution on $D([0, \infty), E)$ to the unique solution X of the (A, ν)-martingale problem.

Proof. The proof is exactly the same as the proof of Theorem 3.1, only that we need to take different φ_N and ψ_N:

$$\varphi_N(t) := f_N(Y_N(t)) - \int_0^t A_N f_N(Y_N(s)) ds$$

$$\psi_N(t) := g_N(Y_N(t)) - \int_0^t A_N g_N(Y_N(s)) ds$$

Since Y_N is a solution to the (A_N, ν_N)-martingale problem, φ_N and ψ_N are $\mathcal{F}_\cdot^N$-martingales.

The rest of the proof is identical. □

3.3 An Application

As a first application of the obtained convergence results we can show that if $Q^{\Xi,\nu}$ denotes the law on $D([0, \infty), \mathcal{P}_\infty)$ of a Ξ-coalescent starting with distribution ν, then the map

$$(\Xi, \nu) \mapsto Q^{\Xi,\nu}$$

is continuous. Here we equip the space of probabilities on $\mathcal{P}_\infty$, $\mathcal{M}_1(\mathcal{P}_\infty)$, and the space of probabilities on $D([0, \infty), \mathcal{P}_\infty)$, $\mathcal{M}_1(D([0, \infty), \mathcal{P}_\infty))$, with the topology of weak convergence. First we need to establish the following lemma which is taken from Schweinsberg (2000a).

Lemma 3.4. *Let $r \geq 1$ and $k_1, \ldots, k_r \geq 2$ let*

$$g_{k_1,\ldots,k_r} : \Delta \to \mathbb{R}, \quad x \mapsto \begin{cases} \displaystyle\sum_{i_1 \neq \cdots \neq i_r} x_{i_1}^{k_1} \cdots x_{i_r}^{k_r} \Big/ \sum_{i=1}^{\infty} x_i^2 \,, & x \neq \mathbf{0} \\ \mathbb{1}_{r=1,k_1=2}, & x = \mathbf{0} \end{cases}$$

Then $g_{k_1,\ldots,k_r}$ is a continuous and bounded map.

Proof. $g := g_{k_1,\dots,k_r}$ is obviously bounded since for $x \neq 0$ $g(x) \leq \sum_{i=1}^{\infty} x_i^2 / \sum_{i=1}^{\infty} x_i^2 = 1$. To see that g is continuous we define for $n \in \mathbb{N}$

$$f^{(n)} : \Delta \to \mathbb{R}, \quad (x_1, x_2, \dots) \mapsto \sum_{\substack{i_1,\dots,i_r=1 \\ \text{all distinct}}}^{n} x_{i_1}^{k_1} \dots x_{i_r}^{k_r}$$

all the $f^{(n)}$ are continuous and we will show the uniform convergence of $f^{(n)}$ to $f(x) := \sum_{i_1 \neq \dots \neq i_r} x_{i_1}^{k_1} \dots x_{i_r}^{k_r}$, which implies the continuity of f: Let $x = (x_1, x_2, \dots) \in \Delta$. Then

$$\begin{aligned} |f^{(n)}(x) - f(x)| &\leq \sum_{j=1}^{r} \sum_{i_j=n+1}^{\infty} \sum_{\substack{i_1,\dots,i_{j-1},i_{j+1},\dots,i_r=1 \\ \text{all distinct}}}^{\infty} x_{i_1}^{k_1} \dots x_{i_r}^{k_r} \leq r \sum_{i_1=n+1}^{\infty} \sum_{i_2,\dots,i_r=1}^{\infty} x_{i_1}^2 \dots x_{i_r}^2 \\ &\leq r \sum_{i_1=n+1}^{\infty} x_{i_1}^2 \leq r \sum_{i_1=n+1}^{\infty} \frac{1}{i_1^2} \end{aligned}$$

since for $x \in \Delta$ and for all $i \in \mathbb{N}$ $x_i \leq 1/i$. This bound tends to 0 when $n \to \infty$, uniformly in x. Thus g is the ratio of two continuous functions and therefore continuous whenever $x \neq \mathbf{0}$. To see the continuity in $\mathbf{0}$, we first consider the case $r = 1, k_1 = 2$. In this case we have $g(x) = 1$ for all x, which is of course continuous. Otherwise let $x \in \Delta \backslash \{\mathbf{0}\}$ with $d(x, \mathbf{0}) < 1/n$ where d denotes the distance on Δ. Then for all i we have $x_i < 1/n$, and since $\sum_{i=1}^{\infty} x_i \leq 1$ we obtain

$$\sum_{i=1}^{\infty} x_i^k \leq \sum_{i=1}^{\infty} x_i^2 \frac{1}{n} \leq \frac{1}{n} n \frac{1}{n^2} = \frac{1}{n^2}$$

for all $k > 2$. Therefore

$$g(x) = \sum_{i_1 \neq \dots \neq i_r} x_{i_1}^{k_1} \dots x_{i_r}^{k_r} \Bigg/ \sum_{i=1}^{\infty} x_i^2 \leq \sum_{i_1=1}^{\infty} x_{i_1}^2 \frac{1}{n} \Bigg/ \sum_{i=1}^{\infty} x_i^2 = \frac{1}{n}$$

so g is continuous in 0. $\square$

The following proposition was proven in Schweinsberg (2000a) for Dirac masses ν. The proof here is different from Schweinsberg's proof since we use our weak convergence results.

Proposition 3.5. *Let Ξ be a finite measure on Δ and let $\nu \in \mathcal{M}_1(\mathcal{P}_\infty)$. Let $Q^{\Xi,\nu}$ be the distribution of a Ξ-coalescent Π with $\Pi_0 \simeq \nu$. Then the map*

$$(\Xi, \nu) \mapsto Q^{\Xi,\nu}$$

is continuous.

Proof. We know that there is uniqueness for the martingale problem for the Ξ-coalescent. We want to apply Proposition 3.3. Let Ξ_N be a sequence of finite measures on Δ that converges weakly to Ξ and denote by $\lambda^N_{b;k_1,\dots,k_r;s}$ respectively λ^N_π the rates of the Ξ_N-coalescent. Denote by $\lambda_{b;k_1,\dots,k_r;s}$ respectively λ_π the rates of the Ξ-coalescent. We introduce the operators A_N and A which are defined as in section 2.2.4:

$$\mathcal{D} := \{F \in C(\mathcal{P}_\infty) : \exists n \in \mathbb{N}, \tilde{F} \in C(\mathcal{P}_n), F(\pi) = \tilde{F}(R_n\pi) \forall \pi\}$$

$$A_N : \mathcal{D} \to C(\mathcal{P}_\infty), F(\cdot) \mapsto \sum_{\eta \in \mathcal{P}_n} \lambda_\eta^N (\tilde{F}(\mathrm{Coag}(R_n \cdot, \eta)) - \tilde{F}(R_n \cdot))$$

and

$$A : \mathcal{D} \to C(\mathcal{P}_\infty), F(\cdot) \mapsto \sum_{\eta \in \mathcal{P}_n} \lambda_\eta (\tilde{F}(\mathrm{Coag}(R_n \cdot, \eta)) - \tilde{F}(R_n \cdot))$$

The Ξ-coalescent Π with $\Pi_0 \simeq \nu$ is the unique càdlàg solution to the (A, ν)-martingale problem. So by Proposition 3.3 it suffices to show that $\mathcal{D}$ contains an algebra that separates points and contains constants (which is obvious), and that

$$\lambda^N_{b;k_1,\ldots,k_r;s} \to \lambda_{b;k_1,\ldots,k_r;s},\ N \to \infty$$

for all $b = k_1 + \cdots + k_r + s$. Since by the consistency relation (10) every other rate can be expressed as a finite linear combination of rates with $s = 0$, it suffices to show the convergence for $s = 0$. But for $s = 0$ we have

$$\lambda^N_{b;k_1,\ldots,k_r;0} = \int_\Delta g_{k_1,\ldots,k_r}(x) \Xi_N(dx)$$

which converges to

$$\int_\Delta g_{k_1,\ldots,k_r}(x) \Xi(dx) = \lambda_{b;k_1,\ldots,k_r;0}$$

by Lemma 3.4. □

4 Ξ-Fleming-Viot Processes

We will present generalisations of the Fleming-Viot process, so called Ξ-Fleming-Viot processes. We will prove that Ξ-Fleming-Viot processes and Ξ-coalescents are dual to each other, which will yield a characterization of the Ξ-Fleming-Viot process as the unique solution to a certain martingale problem.

The Λ-Fleming-Viot process was introduced by Bertoin and Le Gall (2003). The Ξ-Fleming-Viot process was introduced by Birkner et al. (2009). Here we work in the setting of Bertoin and Le Gall (2003) and extend their results to the Ξ-case.

4.1 Definition and Construction of the Ξ-Fleming-Viot Process

We want to generalize the martingale problem that characterized the Fleming-Viot process. Let E be a compact metric space. We introduce the following notation: For a partition $\pi \in \mathcal{P}_n$ for some $n \in \bar{\mathbb{N}}$ and for $i \in [n]$ let $\pi_{[i]} := \min\{j \in [n] : i \overset{\pi}{\sim} j\}$. This notation is a little unfortunate since we already introduced $\pi(i)$ and π_i, so we have to be careful. With this notation we can rewrite the generator of the Fleming-Viot process:

$$A\Phi_f(\mu) = \sum_{\substack{\pi \in \mathcal{P}_p: \\ \#\pi = p-1}} \int [f(x_{\pi_{[1]}}, \ldots, x_{\pi_{[p]}}) - f(x_1, \ldots, x_p)] \mu^{\otimes p}(dx)$$

With this notation it is quite obvious how to generalize the generator: Let $\Xi = \Xi_0 + c\delta_{\mathbf{0}}$ be a finite measure on Δ with $\Xi_0(\{\mathbf{0}\}) = 0$. Then for every $(b; k_1, \dots, k_r; s)$-partition π we define

$$\lambda_\pi := \lambda_{b;k_1,\dots,k_r;s} = \int_\Delta \frac{Q_{k_1,\dots,k_r;s}(x)}{\sum_{j=1}^\infty x_j^2} \Xi_0(dx) + c\mathbb{1}_{\{r=1,k=2\}}$$

as in (1). Let $\mathcal{D} := \{\Phi_f : \mathcal{M}_1(E) \to \mathbb{R}, \Phi(\mu) =< f, \mu^{\otimes p} > \text{ for some } p \in \mathbb{N}, f \in C(E^p)\}$ be the domain of the generator of the Fleming-Viot process. We generalize A by defining an operator

$$G : \mathcal{D} \to C(\mathcal{M}_1(E))$$

such that

$$G\Phi_f(\mu) = \sum_{\substack{\pi \in \mathcal{P}_p: \\ \pi \neq \mathbf{0}_p}} \lambda_\pi \int [f(x_{\pi_{[1]}}, \dots, x_{\pi_{[p]}}) - f(x_1, \dots, x_p)] \mu^{\otimes p}(dx)$$

Definition 4.1. *Let ν be a probability on $\mathcal{M}_1(E)$. A* **Ξ-Fleming-Viot process** *starting with distribution ν is an $\mathcal{M}_1(E)$-valued process $(\rho_t : t \geq 0)$ that is a solution to the (Ξ, ν)-martingale problem. If we just have a family of rates λ_π and we do not want to refer explicitly to the measure Ξ, then ρ is also called a* **generalized Fleming-Viot process***.*

Thus the Fleming-Viot process is the special case of the Ξ-Fleming-Viot process corresponding to $\Xi = \delta_{\mathbf{0}}$. Every function $f : E^n \to \mathbb{R}$ can be interpreted as a function $\tilde{f} : E^{n+1} \to \mathbb{R}$ with $\tilde{f}(x_1, \dots, x_{n+1}) = f(x_1, \dots, x_n)$. So we must have the consistency relation (7) for the rates λ_π, and we can repeat the proof for the classification of Ξ-coalescents to see that every generalized Fleming-Viot process is indeed a Ξ-Fleming-Viot process for some finite measure Ξ on Δ. A priori it is not obvious that a Ξ-Fleming-Viot process exists, nor that the solution to the considered martingale problem is unique.

Remark. Consider $\bar{\mathcal{D}} := \{\Phi_f : \mathcal{M}_1(E) \to \mathbb{R}, \Phi(\mu) =< f, \mu^{\otimes p} > \text{ for some } p \in \mathbb{N}, f = \prod_{i=1}^p \varphi_i,\ \varphi_i \in C(E)\}$, and let $\bar{G}$ be the restriction of G to $\bar{\mathcal{D}}$. Then for any ν, the (G, ν)- and the $(\bar{G}, \nu)$-martingale problem are equivalent, i.e. any solution of the (G, ν)-martingale problem is a solution of the $(\bar{G}, \nu)$-martingale problem and vice versa. Of course any solution of the (G, ν)-martingale problem is a solution of the $(\bar{G}, \nu)$-martingale problem. To see the opposite inclusion, note that the functions of the type $\prod_{i=1}^p \varphi_i(x_i)$ are dense in the uniform topology of $C(E^p)$ by the Stone-Weierstrass theorem. This means that for any $\Phi_f \in \mathcal{D}$ there is a sequence $(\Phi_{f_n}) \subseteq \bar{\mathcal{D}}$ such that Φ_{f_n} tends uniformly to Φ_f. By the definition of G we see that then also $\bar{G}\Phi_{f_n} = G\Phi_{f_n}$ tends uniformly to $G\Phi_f$. So for any bounded $\mathcal{F}_t$-measurable random variable Z and for any solution ρ of the $(\bar{G}, \nu)$-martingale problem we have by uniform convergence

$$\begin{aligned}
&\mathbb{E}\left[\left(\Phi_f(\rho_{t+s}) - \Phi_f(\rho_t) - \int_t^{t+s} G\Phi_f(\rho_u)du\right) Z\right] \\
&= \lim_{n\to\infty} \mathbb{E}\left[\left(\Phi_{f_n}(\rho_{t+s}) - \Phi_{f_n}(\rho_t) - \int_t^{t+s} \bar{G}\Phi_{f_n}(\rho_u)du\right) Z\right] = 0
\end{aligned}$$

Proposition 4.2. *Let E be any compact metric space and let ν be a distribution on $\mathcal{M} := \mathcal{M}_1(E)$. Then a càdlàg version of the Ξ-Fleming-Viot process with values in $\mathcal{M}$ and with starting distribution ν exists. If $(\rho_t : t \geq 0)$ and $(\eta_t : t \geq 0)$ are two Ξ-Fleming-Viot processes such that ρ and η have the same starting distribution, then they have the same finite-dimensional distributions. In particular any two càdlàg Ξ-Fleming-Viot processes with the same starting distribution have the same distribution on $D([0, \infty), \mathcal{M})$.*

Remark. The existence of càdlàg Ξ-Fleming-Viot processes was also shown by Birkner et al. (2009) with a particle system construction. Here we take a different approach.

Proof. We prove the proposition in several steps: First we prove uniqueness of the solutions. Then we show that for measures Ξ satisfying $\Xi(\{0\}) = 0$ and $\int_\Delta (1/\sum_{i=1}^\infty x_i^2)\,\Xi(dx) < \infty$ the Ξ-Fleming-Viot process exists as a jump-hold process. Finally we obtain general Ξ-Fleming-Viot processes as limits of those jump-hold processes.

1. To prove uniqueness we will show the duality of Ξ-Fleming-Viot processes and Ξ-coalescents. Let $p \in \mathbb{N}$ and let $f \in C(E^p)$. Bertoin and Le Gall (2003) introduced a cleverly chosen function on $\mathcal{M}_1(E) \times \mathcal{P}_p$ that gives us the duality: Let for $\pi \in \mathcal{P}_p$ and for $(x_1, \dots, x_p) \in E^p$ $Y(\pi; x_1, \dots, x_p) := (y_1, \dots, y_p)$ where $y_i = x_j$ if and only if i is in π_j. We define

$$\Theta_f : \mathcal{M}_1(E) \times \mathcal{P}_p \to \mathbb{R}, \quad \Theta_f(\mu, \pi) := \int_{E^p} f(Y(\pi; x_1, \dots, x_p))\mu^{\otimes p}(dx_1, \dots, dx_p)$$

When we fix a partition $\pi \in \mathcal{P}_p$, $\Theta_f(\cdot, \pi)$ is of the form $\int_{E^p} g(x_1, \dots, x_p)\mu^{\otimes p}(dx_1, \dots, dx_p)$ for some $g \in C(E^p)$. Therefore we can define $G\Theta_f(\cdot, \pi)$. Let $(\Pi_p(t) : t \geq 0)$ be the standard Ξ-coalescent with values in $\mathcal{P}_p$. We assume that Π_p is independent of ρ. Denote Q the generator of Π_p. We have

$$QF(\pi) = \sum_{\substack{\eta \in \mathcal{P}_p: \\ \eta \neq \mathbf{0}_p}} \lambda_\eta (F(\mathrm{Coag}(\pi, \eta)) - F(\pi))$$

for any function F on $\mathcal{P}_p$. Since for fixed μ, $\Theta_f(\mu, \cdot)$ is a function on $\mathcal{P}_p$, we can define $Q\Theta_f(\mu, \cdot)$. We readily see that

$$G\Theta_f(\mu, \pi) = Q\Theta_f(\mu, \pi)$$

for any $\mu \in \mathcal{M}_1(E)$ and $\pi \in \mathcal{P}_p$:

$$G\Theta_f(\mu, \pi) = \sum_{\substack{\eta \in \mathcal{P}_p: \\ \eta \neq \mathbf{0}_p}} \lambda_\pi \int [f(Y(\pi; x_{\eta_{[1]}}, \dots, x_{\eta_{[p]}})) - f(Y(\pi; x_1, \dots, x_p))]\mu^{\otimes p}(dx)$$

and

$$Q\Theta_f(\mu, \pi) = \sum_{\substack{\eta \in \mathcal{P}_p: \\ \eta \neq \mathbf{0}_p}} \lambda_\pi \int [f(Y(\mathrm{Coag}(\pi, \eta); x_1, \dots, x_p)) - f(Y(\pi; x_1, \dots, x_p))]\mu^{\otimes p}(dx)$$

Let $x_1 \neq \cdots \neq x_p \in E$. Let $Y(\pi; x_{\eta_{[1]}}, \dots, x_{\eta_{[p]}}) = (y_1, \dots, y_p)$ and $Y(\mathrm{Coag}(\pi, \eta); x_1, \dots, x_p) = (\tilde{y}_1, \dots, \tilde{y}_p)$. Let $i, j \leq p$, $i \neq j$. Then $y_i = y_j$ if and only if $i \in \pi_k$, $j \in \pi_l$ and $\eta_{[k]} = \eta_{[l]}$. But this is the case if and only if k and l are in the same block of η and thus if and only if i and j are in the same block of $\mathrm{Coag}(\pi, \eta)$. Thus for all $x_1 \neq \cdots \neq x_p$ we have $Y(\pi; x_{\eta_{[1]}}, \dots, x_{\eta_{[p]}}) = Y(\mathrm{Coag}(\pi, \eta); x_1, \dots, x_p)$. Of course this also holds for general choices of $x_1, \dots, x_p$. Hence

$$G\Theta_f(\mu, \pi) = Q\Theta_f(\mu, \pi)$$

for all μ and π. This implies that for every Ξ-Fleming-Viot process $(\rho_t : t \geq 0)$ and for all $f \in C(E^p)$ we have

$$\mathbb{E}(\Theta_f(\rho_t, \Pi_p(0))) = \mathbb{E}(\Theta_f(\rho_0, \Pi_p(t)))$$

and thus that ρ and Π_p are dual with respect to Θ_f in the sense of Liggett (1985). The following arguments are taken from Etheridge (2000): Let $\pi \in \mathcal{P}_p$. We have

$$\mathbb{E}(\Theta_f(\rho_t, \pi)) = \mathbb{E}\left(\int_0^t G\Theta_f(\rho_s, \Pi_p(0))ds\right) + \mathbb{E}(\Theta_f(\rho_0, \pi))$$

and therefore

$$\frac{d}{ds}\mathbb{E}(\Theta_f(\rho_s, \pi)) = \mathbb{E}(G\Theta_f(\rho_s, \pi))$$

Analogously we obtain for $\mu \in \mathcal{M}_1(E)$:

$$\frac{d}{ds}\mathbb{E}(\Theta_f(\mu, \Pi_p(s))) = \mathbb{E}(Q\Theta_f(\mu, \Pi_p(s)))$$

Therefore for fixed t and for $0 \leq s \leq t$:

$$\frac{d}{ds}\mathbb{E}(\Theta_f(\rho_s, \Pi_p(t-s))) = \mathbb{E}(G\Theta_f(\rho_s, \Pi_p(t-s))) - \mathbb{E}(Q\Theta_f(\rho_s, \Pi_p(t-s))) = 0$$

and thus

$$0 = \int_0^t \frac{d}{ds}\mathbb{E}(\Theta_f(\rho_s, \Pi_p(t-s)))ds = \mathbb{E}(\Theta_f(\rho_t, \Pi_p(0))) - \mathbb{E}(\Theta_f(\rho_0, \Pi_p(t)))$$

But $\Pi_p(0) = \mathbf{0}_p$ and thus $\Theta_f(\cdot, \Pi_p(0)) = \Phi_f(\cdot)$. Since $\mathcal{D}$ is an algebra that separates points on $\mathcal{M}_1(E)$ and that contains constant functions, it is dense in the uniform topology on $C(\mathcal{M}_1(E))$. Therefore the one-dimensional marginals of the Ξ-Fleming-Viot process are uniquely determined by its starting distribution. But for the solution of a martingale problem it is sufficient to have uniqueness of one-dimensional distributions to obtain uniqueness of finite-dimensional distributions (cf. Theorem 4.2 of Chapter 4 of Ethier and Kurtz (1986)). This proves our uniqueness statement. It remains to show existence of Ξ-Fleming-Viot processes.

2. Let Ξ be a finite measure on Δ satisfying $\Xi(\{\mathbf{0}\}) = 0$ and $\int_\Delta (1/\sum_{i=1}^\infty x_i^2)\,\Xi(dx) < \infty$. Define $\bar{\Xi}(dx) := 1/\sum_{i=1}^\infty x_i^2 \Xi(dx)$. Consider the following transition function P on $\mathcal{M} \times \mathcal{B}(\mathcal{M})$ ($\mathcal{B}(\mathcal{M})$ being the Borel σ-algebra of $\mathcal{M}$):

$$P(\mu, B) := \int_\Delta \int_{E^\mathbb{N}} \mathbb{1}_B \left(\left(1 - \sum_{i=1}^\infty x_i\right)\mu + \sum_{i=1}^\infty x_i \delta_{y_i}\right) \mu^{\otimes\mathbb{N}}(dy) \frac{\bar{\Xi}(dx)}{\bar{\Xi}(\Delta)}$$

Consider the operator $A : B(\mathcal{M}) \to B(\mathcal{M})$,

$$Af(\mu) := \bar{\Xi}(\Delta) \int_\mathcal{M} (f(\eta) - f(\mu))P(\mu, d\eta)$$

Since the jump rate $\bar{\Xi}(\Delta)$ is bounded, there exists a jump-hold Markov process $(\rho_t : t \geq 0)$ with generator A (cf. Chapter 4.2 of Ethier and Kurtz (1986)), starting with distribution ν. We can even construct it explicitly: Let $(Y_m : m \in \mathbb{N}_0)$ be a discrete time Markov process with transition function P and with starting distribution ν. Let $(N_t : t \geq 0)$ be a Poisson process with parameter $\bar{\Xi}(\Delta)$ that is independent of Y. Define

$$\rho_t := Y(N_t),\ t \geq 0$$

Then ρ is a Markov process with the desired generator. Now let $f(x_1, \ldots, x_p) = \varphi_1(x_1) \ldots \varphi_p(x_p)$ with $\varphi_i \in C(E)$ for all i. We want to evaluate $A\Phi_f(\mu)$. We have

$$\begin{aligned}
&\Phi_f\left(\left(1-\sum_{i=1}^{\infty} x_i\right)\mu + \sum_{i=1}^{\infty} x_i \delta_{y_i}\right) = \prod_{j=1}^{p}\left[\left(1-\sum_{i=1}^{\infty} x_i\right) < \varphi_j, \mu > + \sum_{i=1}^{\infty} x_i \varphi_j(y_i)\right] \\
&\quad = \sum_{J \subseteq [p]} \left(1-\sum_{i=1}^{\infty} x_i\right)^{p-|J|} \prod_{j \in [p]\setminus J} < \varphi_j, \mu > \prod_{j \in J}\left(\sum_{i=1}^{\infty} x_i \varphi_j(y_i)\right) \\
&\quad = \sum_{J \subseteq [p]} \left(1-\sum_{i=1}^{\infty} x_i\right)^{p-|J|} \prod_{j \in [p]\setminus J} < \varphi_j, \mu > \times \\
&\qquad \left[\sum_{\pi \in \mathcal{P}_J} \sum_{i_1 \neq \cdots \neq i_{\#\pi}} \left(\prod_{j \in \pi_1} x_{i_1} \varphi_j(y_{i_1})\right) \ldots \left(\prod_{j \in \pi_{\#\pi}} x_{i_{\#\pi}} \varphi_j(y_{i_{\#\pi}})\right)\right]
\end{aligned}$$

and therefore

$$\begin{aligned}
&\int_{[0,1]^{\mathbb{N}}} \Phi_f\left(\left(1-\sum_{i=1}^{\infty} x_i\right)\mu + \sum_{i=1}^{\infty} x_i \delta_{y_i}\right) \mu^{\otimes \mathbb{N}}(dy) \\
&\quad = \sum_{J \subseteq [p]} \left(1-\sum_{i=1}^{\infty} x_i\right)^{p-|J|} \prod_{j \in [p]\setminus J} < \varphi_j, \mu > \times \\
&\qquad \left[\sum_{\pi \in \mathcal{P}_J} \sum_{i_1 \neq \cdots \neq i_{\#\pi}} x_{i_1}^{|\pi_1|} < \prod_{j \in \pi_1} \varphi_j, \mu > \ldots x_{i_{\#\pi}}^{|\pi_{\#\pi}|} < \prod_{j \in \pi_{\#\pi}} \varphi_j, \mu >\right]
\end{aligned}$$

Note that

$$\begin{aligned}
&\sum_{J \subseteq [p]} \left(1-\sum_{i=1}^{\infty} x_i\right)^{p-|J|} \sum_{\pi \in \mathcal{P}_J} \sum_{i_1 \neq \cdots \neq i_{\#\pi}} x_{i_1}^{|\pi_1|} \ldots x_{i_{\#\pi}}^{|\pi_{\#\pi}|} \\
&\quad = \sum_{J \subseteq [p]} \left(1-\sum_{i=1}^{\infty} x_i\right)^{p-|J|} \left(\sum_{i=1}^{\infty} x_i\right)^{|J|} = \left(1-\sum_{i=1}^{\infty} x_i + \sum_{i=1}^{\infty} x_i\right)^{p} = 1
\end{aligned}$$

and therefore

$$\int_{[0,1]^{\mathbb{N}}} \Phi_f\left(\left(1-\sum_{i=1}^\infty x_i\right)\mu + \sum_{i=1}^\infty x_i\delta_{y_i}\right)\mu^{\otimes\mathbb{N}}(dy) - \Phi_f(\mu)$$

$$= \sum_{J\subseteq[p]}\left(1-\sum_{i=1}^\infty x_i\right)^{p-|J|} \sum_{\pi\in\mathcal{P}_J}\sum_{i_1\neq\cdots\neq i_{\#\pi}} x_{i_1}^{|\pi_1|}\ldots x_{i_{\#\pi}}^{|\pi_{\#\pi}|}\times$$

$$\left[\prod_{j\in[p]\setminus J} <\varphi_j,\mu><\prod_{j\in\pi_1}\varphi_j,\mu>\cdots<\prod_{j\in\pi_{\#\pi}}\varphi_j,\mu> - \prod_{i=1}^p<\varphi_i,\mu>\right]$$

$$= \sum_{\substack{\pi\in\mathcal{P}_p\setminus\mathbf{0}_p \\ (p;k_1,\ldots,k_r;s)-\text{partition}}} \sum_{l=0}^s\binom{s}{l}\left(1-\sum_{i=1}^\infty x_i\right)^{s-l}\sum_{i_1\neq\cdots\neq i_{r+l}} x_{i_1}^{k_1}\ldots x_{i_r}^{k_r}x_{i_{r+1}}\ldots x_{i_{r+l}}\times$$

$$\left[<\prod_{j\in\pi_1}\varphi_j,\mu>\cdots<\prod_{j\in\pi_{\#\pi}}\varphi_j,\mu> - \prod_{i=1}^p<\varphi_i,\mu>\right]$$

$$= \sum_{\substack{\pi\in\mathcal{P}_p\setminus\mathbf{0}_p: \\ (p;k_1,\ldots,k_r;s)-\text{partition}}} Q_{k_1,\ldots,k_r;s}(x)\times$$

$$[<f(y_{\pi_{[1]}},\ldots,y_{\pi_{[p]}}),\mu^{\otimes p}> - <f(y_1,\ldots,y_p),\mu^{\otimes p}>]$$

where the sum over the partitions means that we sum over all partitions π and just distinguish the different types that π can have. So finally we obtain

$$A\Phi_f(\mu) = \int_{E^{\mathbb{N}}}\mu^{\otimes\mathbb{N}}(dy)\int_\Delta\bar{\Xi}(dx)\left[\Phi_f\left(\left(1-\sum_{i=1}^\infty x_i\right)\mu+\sum_{i=1}^\infty x_i\delta_{y_i}\right)-\Phi_f(\mu)\right]$$

$$= \sum_{\substack{\pi\in\mathcal{P}_p\setminus\mathbf{0}_p: \\ (p;k_1,\ldots,k_r;s)-\text{partition}}} \int_\Delta \frac{Q_{k_1,\ldots,k_r;s}(x)}{\sum_{i=1}^\infty x_i^2}\Xi(dx)\times$$

$$[<f(y_{\pi_{[1]}},\ldots,y_{\pi_{[p]}}),\mu^{\otimes p}> - <f(y_1,\ldots,y_p),\mu^{\otimes p}>]$$

$$= \sum_{\substack{\pi\in\mathcal{P}_p\setminus\mathbf{0}_p: \\ (p;k_1,\ldots,k_r;s)-\text{partition}}} \lambda_{b;k_1,\ldots,k_r;s}\times$$

$$[<f(y_{\pi_{[1]}},\ldots,y_{\pi_{[p]}}),\mu^{\otimes p}> - <f(y_1,\ldots,y_p),\mu^{\otimes p}>]$$

$$= G\Phi_f(\mu) = \bar{G}\Phi_f(\mu)$$

So ρ is a solution to the $(\bar{G},\mu)$-martingale problem, which by our previous remark implies that ρ is a solution to the (G,μ)-martingale problem, i.e. it is a Ξ-Fleming-Viot process starting with distribution ν.

3. It remains to show the statement for general Ξ. We choose a sequence of finite measures Ξ_N on Δ with $\Xi_N(\{\mathbf{0}\}) = 0$ for all N, such that $\int_\Delta 1/\sum_{i=1}^\infty x_i^2\,\Xi_N(dx) < \infty$ for all N, and such that Ξ_N converges weakly to Ξ. For example we can choose for $\Xi = \Xi_0 + c\delta_{\mathbf{0}}$ a sequence $x^N \neq \mathbf{0}$ that converges to $\mathbf{0}$ in Δ, and then define

$$\Xi_N(dx) := \mathbb{1}_{\{\sum_{i=1}^\infty x_i^2\geq 1/N\}}(x)\Xi(dx) + c\delta_{x^N}$$

For every N we construct a Ξ_N-Fleming-Viot process with values in $\mathcal{M}$, ρ^N, with starting distribution ν like in the previous step. We showed in the proof of Proposition 3.5 that the rates λ_π^N of the Ξ_N-coalescent converge to the rates λ_π of the Ξ-coalescent. Since the ρ^N have càdlàg paths by construction and since the domain $\mathcal{D}$ contains an algebra that separates points and contains constants, we can apply Proposition 3.3 to get the convergence of ρ^N to the Ξ-Fleming-Viot process, which therefore has to exist.

$\square$

Remark. Bertoin and Le Gall (2003) gave a Poisson point process construction for the Λ-Fleming-Viot process. Since this construction only seems to work for the case $E = [0,1]$, we rather constructed the process with a combination of a discrete time Markov process and a Poisson process. Most of the proof consists just of generalisation of notation for the Λ-case that was proven by Bertoin and Le Gall (2003). However to obtain Ξ-Fleming-Viot processes for general Ξ as limits of Ξ_N-Fleming-Viot ρ^N processes with $1/\sum_{i=1}^\infty x_i^2 \Xi_N(dx) < \infty$, we needed to change the argumentation a little:

In the Λ-setting we obtain that for every bounded measurable function φ on $[0,1]$,

$$\left(< \varphi, \rho_t^N >: t \geq 0\right)$$

is a martingale with quadratic variation

$$\int_{[0,1]} x^2 \bar{\Lambda}_N(dx) \int_0^t \left(< \varphi^2, \rho_s^N > - < \varphi, \rho_s^N >^2\right) ds$$

where $\bar{\Lambda}_N(dx) = x^{-2}\Lambda_N(dx)$ and the Λ_N correspond to the Ξ_N of our proof. So the sequence $\int_{[0,1]} x^2 \bar{\Lambda}_N(dx)$ is bounded and therefore the quadratic variation of the martingale is C-tight, which implies the tightness of the sequence $< \varphi, \rho^N >$ by Theorem 4.13 in Chapter VI. of Jacod and Shiryaev (2002). This in turn yields the tightness of the sequence ρ^N by Theorem 9.1 in Chapter 3 of Ethier and Kurtz (1986).

However in the Ξ-case we obtain the same sequence of martingales, but now their quadratic variation is given by

$$\int_{[0,1]} \left(\sum_{i=1}^\infty x_i\right)^2 \bar{\Xi}_N(dx) \int_0^t \left(< \varphi^2, \rho_s^N > - < \varphi, \rho_s^N >^2\right) ds$$

and in general the sequence

$$\int_\Delta \left(\sum_{i=1}^\infty x_i\right)^2 \bar{\Xi}_N(dx) = \int_\Delta \left(\sum_{i=1}^\infty x_i\right)^2 \Big/ \sum_{i=1}^\infty x_i^2 \, \Xi_N(dx)$$

is not bounded. To show tightness of $< \varphi, \rho^N >$ directly does not seem to be very easy either since the jump-rate $\bar{\Xi}_N(\Delta)$ of the Poisson processes used to construct ρ_N tends to infinity when $N \to \infty$.

4.2 Some Properties of the Ξ-Fleming-Viot Process

Proposition 4.3. *Let E be a compact metric space. Let Ξ be a finite measure on Δ Then any Ξ-Fleming-Viot process ρ with values in $\mathcal{M}_1(E)$ has the strong Markov property with respect to* *the filtration $(\mathcal{F}_t)$, i.e. for any bounded measurable function f and for any finite $(\mathcal{F}_t)$-stopping time τ we have*

$$\mathbb{E}_\mu(f(\rho_{\tau+t})|\mathcal{F}_\tau) = \mathbb{E}_{\rho_\tau}(f(\rho_t))$$

Proof. This is just Theorem 4.2 c) of Chapter 4 in Ethier and Kurtz (1986), since we already established the uniqueness of the martingale problem for the Ξ-Fleming-Viot process. The only thing we still need to show is that if $\mathbb{P}_\mu$ denotes the law on $D([0,\infty),\mathcal{M}_1(E))$ of the Ξ-Fleming-Viot process starting in μ, then for any Borel set B in $D([0,\infty),\mathcal{M}_1(E))$ the map

$$\mu \mapsto \mathbb{P}_\mu(B)$$

is Borel measurable. But by Theorem 4.6 of Chapter 4 in Ethier and Kurtz (1986) this follows if $\mathcal{M}_1(E)$ is complete and separable and if $C_b(\mathcal{M}_1(E))$ is separable. Since E is compact, $\mathcal{M}_1(E)$ is compact by Prohorov's theorem (Theorem 2.2 of Chapter 3 of Ethier and Kurtz (1986)). The topology of weak convergence is generated by the Prohorov distance (cf. Theorem 3.1 in Chapter 3 of Ethier and Kurtz (1986)), so $\mathcal{M}_1(E)$ is a compact metric space, so it is complete and separable. Also, $C_b(\mathcal{M}_1(E)) = C(\mathcal{M}_1(E))$ is separable by a Stone-Weierstrass argument. □

Remark. In the case $E = [0,1]$ it is easy to see that any Ξ-Fleming-Viot process is in fact a Feller process. This can be shown by using a connection between flows of bridges and Ξ-Fleming-Viot processes, and it is explained in Bertoin and Le Gall (2003). In the general case this result is more complicated and it was shown by Birkner et al. (2009):

Proposition 4.4. *Let E be a compact metric space and let Ξ be a finite measure on Δ. The Ξ-Fleming-Viot process with values in $\mathcal{M}_1(E)$ is a Feller process.*

Proof. This is Proposition 4.3 (respectively Remark 4.4 a)) of Birkner et al. (2009). There it is shown that the operator that we used to introduce the Ξ-Fleming-Viot process satisfies a necessary and sufficient condition for its closure to generate a Feller semi-group. □

Proposition 4.5. *For a distribution ν on $\mathcal{M}_1(E)$ and for a finite measure Ξ on Δ denote by $P^{\nu,\Xi}$ the law on $D([0,\infty),\mathcal{M}_1(E))$ of a càdlàg Ξ-Fleming-Viot Process ρ with $\rho_0 \simeq \nu$. Then the map*

$$\mathcal{M}_1(\mathcal{M}_1(E)) \times \mathcal{M}_f(\Delta) \ni (\nu,\Xi) \mapsto P^{\nu,\Xi} \in \mathcal{M}_1(D([0,\infty),\mathcal{M}_1(E)))$$

is continuous. Here $\mathcal{M}_f(\Delta)$ is the space of finite measures on Δ and of course all the spaces of measures are equipped with the topology of weak convergence.

Proof. We already proved everything that we need to get the continuity of this map: If A_Ξ is the operator that we used to define the Ξ-Fleming-Viot process, then the (A_Ξ,ν)-martingale problem has a unique solution by Proposition 4.2. If Ξ_N converges to Ξ and λ_π^N are the rates of the Ξ_N-Fleming-Viot process and λ_π are the rates of the Ξ-Fleming-Viot process, then λ_π^N converges to λ_π for every π (which was shown in the proof of Proposition 3.5). Therefore $A_{\Xi_N}\Phi_f$ converges uniformly to $A_\Xi\Phi_f$ for all $\Phi_f \in \mathcal{D}$. Since the domain $\mathcal{D}$ contains an algebra that separates points, we can apply Proposition 3.3 to obtain the continuity. □

4.3 Discrete Time Ξ-Fleming-Viot Processes

We introduce a discrete time Ξ-Fleming-Viot process for measures Ξ satisfying $\Xi(\{\mathbf{0}\}) = 0$ and $\int_\Delta 1/\sum_{i=1}^\infty x_i^2 \Xi(dx) \leq 1$. Then we show that this process is the unique solution to a discrete time martingale problem.

Define $\bar{\Xi}(dx) := 1/\sum_{i=1}^\infty x_i^2 \Xi(dx)$ and consider the transition function from the proof of Proposition 4.2:

$$P : \mathcal{M}_1(E) \times \mathcal{B}(\mathcal{M}_1(E)) \to [0,1],$$

$$P(\mu, B) := \int_\Delta \int_{E^\mathbb{N}} \mathbb{1}_B \left(\left(1 - \sum_{i=1}^\infty x_i \right) \mu + \sum_{i=1}^\infty x_i \delta_{y_i} \right) \mu^{\otimes \mathbb{N}}(dy) \bar{\Xi}(dx) + (1 - \bar{\Xi}(\Delta)) \mathbb{1}_B(\mu)$$

A **discrete time Ξ-Fleming-Viot process** is a discrete time Markov process with transition function P. Define the operator

$$T : B(\mathcal{M}_1(E)) \to B(\mathcal{M}_1(E)),\ Tf(\cdot) := \int_{\mathcal{M}_1(E)} f(\mu) P(\cdot, d\mu)$$

We know that for any discrete time Markov process $(Y(m) : m \in \mathbb{N}_0)$ with transition function P and for any bounded measurable function f, the process

$$M_f(m) := f(Y(m)) - \sum_{i=0}^{m-1} (T - I) f(Y_k),\ m \in \mathbb{N}_0$$

is a martingale with respect to the filtration $\mathcal{F}_k := \sigma(Y_0, \ldots, Y_k)$. ($I$ is the identity map). Conversely we know that if for every bounded measurable f M_f is a martingale with respect to some filtration $(\mathcal{F}_k)$, then Y is a Markov process with respect to $\mathcal{F}$, and its transition function is given by $P(\mu, B) := T\mathbb{1}_B(\mu)$. We want to examine T on a certain set of functions $\mathcal{D}$, and then show that the discrete time Ξ-Fleming-Viot process is the unique process for which M_f is a martingale for all $f \in \mathcal{D}$.

Proposition 4.6. *Let $\mathcal{D} := \{\Phi_f \in C(\mathcal{M}_1(E)) : \exists f \in C(E^p)$ s.t. $\Phi_f(\mu) =< f, \mu^{\otimes p} >\}$. Let Ξ be a finite measure on Δ with $\Xi(\{\mathbf{0}\}) = 0$ and such that $\int_\Delta 1/\sum_{i=1}^\infty x_i^2 \Xi(dx) < \infty$. Let λ_π be the rates of the Ξ-Fleming-Viot process. Define*

$$G : \mathcal{D} \to C(\mathcal{M}_1(E)),\ G\Phi_f(\mu) = \sum_{\substack{\pi \in \mathcal{P}_p: \\ \pi \neq \mathbf{0}_p}} \lambda_\pi \int [f(x_{\pi_{[1]}}, \ldots, x_{\pi_{[p]}}) - f(x_1, \ldots, x_p)] \mu^{\otimes p}(dx)$$

If for a discrete time process $(Y(m) : m \in \mathbb{N}_0)$ for every $\Phi_f \in \mathcal{D}$

$$M_f(m) := \Phi_f(Y(m)) - \sum_{i=0}^{m-1} G\Phi_f(Y(i)),\ m \in \mathbb{N}_0$$

is a martingale with respect to the filtration $\mathcal{F}_k = \sigma(Y(0), \ldots, Y(k))$, then Y is a discrete time Ξ-Fleming-Viot process

Proof. Let $\Phi_f \in \mathcal{D}$. Then

$$\mathbb{E}(\Phi_f(Y(m)) | \mathcal{F}_{m-1}) = M_f(m-1) + \sum_{i=0}^{m-1} G\Phi_f(Y(i)) = \Phi_f(Y(m-1)) + G\Phi_f(Y(m-1))$$

In the proof of Proposition 4.2 we saw that for $\Phi_f \in \mathcal{D}$

$$\begin{aligned} G\Phi_f(\mu) &= \int_\Delta \int_{E^{\mathbb{N}}} \left[\Phi_f \left(\left(1 - \sum_{i=1}^{\infty} x_i \right) \mu + \sum_{i=1}^{\infty} x_i \delta_{y_i} \right) - \Phi_f(\mu) \right] \mu^{\otimes \mathbb{N}}(dy) \Xi(dx) \\ &= T\Phi_f(\mu) - \Phi_f(\mu) \end{aligned}$$

where T is the transition operator of the discrete time Ξ-Fleming-Viot process. So for such Φ_f

$$\mathbb{E}(\Phi_f(Y(m))|\mathcal{F}_{m-1}) = T\Phi_f(Y_{m-1})$$

Since the domain $\mathcal{D}$ is an algebra that separates points and contains constants and since $\mathcal{M}_1(E)$ is compact, $\mathcal{D}$ is dense in the uniform topology of $C(\mathcal{M}_1(E))$. Thus

$$\mathbb{E}(F(Y(m))|\mathcal{F}_{m-1}) = TF(Y_{m-1})$$

for every $F \in C(\mathcal{M}_1(E))$. We can apply the monotone class theorem to obtain

$$\mathbb{E}(F(Y(m))|\mathcal{F}_{m-1}) = TF(Y_{m-1})$$

for every bounded measurable F: The set of functions satisfying this equation is closed under uniform convergence and under bounded point-wise convergence (F_N converges bounded point-wise to F if for all $\mu \in \mathcal{M}_1(E)$ $F_N(\mu)$ converges to $F(\mu)$ and if $\sup_N ||F_N|| < \infty$), and it contains the continuous functions which are closed under multiplication. So Corollary 4.4 in the Appendix of Ethier and Kurtz (1986) yields that the set contains all bounded measurable functions. So Y is a discrete time Markov process with transition operator T. Therefore it is a discrete time Ξ-Fleming-Viot process. □

5 Cannings' Population Model

5.1 The Model

We consider a population model introduced by Cannings (1974, 1975). In this model, we assume we are given a haploid population with non-overlapping generations, and that in every generation the population has the constant size N. We suppose there is an infinite number of generations both in the past and in the future, i.e. for every $m \in \mathbb{Z}$ we are given a generation. The model is described by a family of random variables $\{(\nu_{1,N}^m, \ldots, \nu_{N,N}^m) : m \in \mathbb{Z}\}$, where $\nu_{i,N}^m$ is the number of descendants of the ith individual in generation m of a population of size N. Since the size of the population stays constant in all generations, we necessarily have

$$\nu_{1,N}^m + \cdots + \nu_{N,N}^m = N \quad \forall m \in \mathbb{Z} \tag{20}$$

We suppose that the reproduction in different generations is independent and of the same law, i.e.

$$(\nu_{1,N}^m, \ldots, \nu_{N,N}^m), m \in \mathbb{Z}, \text{ are i.i.d.} \tag{21}$$

So if we are only interested in the distribution of $(\nu_{1,N}^m, \ldots, \nu_{N,N}^m)$, we can omit the index m. Finally, we suppose that the reproduction of an individual i does not depend on the index i, i.e. that

$$(\nu_{1,N}, \ldots, \nu_{N,N}) \text{ is an exchangeable random vector.} \tag{22}$$

We suppose that the individuals in generation m are distributed randomly on the families in generation $m-1$. For example the i-th individual in generation m is a descendant of the j-th individual in generation $m-1$ with probability $\nu_{j,N}^{m-1}/N$.

Clearly the Wright-Fisher model is a special case of this class of models. In that case, $(\nu_{1,N}, \ldots, \nu_{N,N})$ has the multinomial distribution with parameters $(N; 1/N, \ldots, 1/N)$.

Assume we are interested in the genealogy of a sample of the population. Say we sample $n \leq N$ individuals in generation 0, and we introduce a process $(\Pi_{n,N}(m) : m \in \mathbb{N}_0)$ with values in $\mathcal{P}_n$: i and j are in the same block of $\Pi_{n,N}(m)$ if and only if the i-th and the j-th individual have the same acestor in the $-m$-th generation. Of course if two individuals have the same ancestor in the $-m$-th generation, then this is also the case in the $-m-1$-th generation, and therefore $\Pi_{n,N}(m+1)$ is always coarser then $\Pi_{n,N}(m)$, i.e. $\Pi_{n,N}(m+1)$ is obtained by coagulating blocks of $\Pi_{n,N}(m)$.

On the other side, if we wish to model the distribution of genetic types, we can introduce a measure-valued process $(\rho_N(m) : m \in \mathbb{N}_0)$: Assume that in generation 0 every individual has some genetic type, which we will represent by an element x of some metric compact space E. Assume that every individual in generation m inherits its genetic type without mutation from its ancestor in generation $m-1$. We introduce a process $(Y^N(m) : m \in \mathbb{N}_0)$ with values in E^N, such that $Y_i^N(m)$ is the genetic type of individual i in generation m. Then we define

$$\rho_N(m) := \sum_{i=1}^{N} \frac{1}{N} \delta_{Y_i^N(m)}$$

which therefore is a process with values in $\mathcal{M}^N(E) := \{\sum_{i=1}^N \frac{1}{N}\delta_{x_i} : (x_1, \ldots, x_N) \in E^N\}$.

5.2 Convergence Results

We want to let the size N of the population tend to infinity to obtain diffusion approximations for our processes $\Pi_{n,N}$ and ρ^N. To obtain a diffusion approximation, obviously we will need to rescale the time. The right factor with which to rescale the time is the probability of two individuals in generation m to have the same ancestor in generation $m-1$, which is

$$c_N = \sum_{i=1}^{N} \frac{\mathbb{E}((\nu_{i,N})_2)}{(N)_2} = \frac{\mathbb{E}((\nu_{1,N})_2)}{N-1} = \frac{\sigma_N^2}{N-1}$$

where $(N)_k := N(N-1)\ldots(N-k+1)$ and where σ_N^2 is the variance of $\nu_{1,N}$ (and the last equality is true because $\mathbb{E}(\nu_{1,N}) = 1$).

Let $\pi \in \mathcal{P}_b$ be a $(b; k_1, \ldots, k_r; s)$-partition - where we could have $r = 0$. If we take b individuals in generation m and label them from 1 to b, then the probability that exactly the individuals whose numbers are in the same block of π have the same ancestor in generation $m-1$ is given by

$$\begin{aligned} &\sum_{\substack{i_1,\ldots,i_{r+s}=1 \\ \text{all distinct}}}^{N} \frac{\mathbb{E}((\nu_{i_1,N})_{k_1} \ldots (\nu_{i_r,N})_{k_r} \nu_{i_{r+1},N} \ldots \nu_{i_{r+s},N})}{(N)_b} \\ &\qquad = \frac{(N)_{r+s}}{(N)_b} \mathbb{E}((\nu_{1,N})_{k_1} \ldots (\nu_{r,N})_{k_r} \nu_{r+1,N} \ldots \nu_{r+s,N}) \end{aligned} \tag{23}$$

We will first state the most general convergence theorem, and later we will present some criteria to check whether the assumptions of the theorem hold. First we give the partition-valued formulation which was proven by Möhle and Sagitov (2001). Here we use their ideas, only in the end of the proof we use the general weak convergence results that we established, rather than using the arguments from Möhle (1999), where convergence in distribution in the Skorohod-topology is proven with coupling techniques.

Theorem 5.1. *Suppose that for every $r \in \mathbb{N}, k_1, \ldots, k_r \geq 2$, the limits*

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{k_1} \ldots (\nu_{r,N})_{k_r})}{N^{k_1+\cdots+k_r-r} c_N} =: \Phi_r(k_1, \ldots, k_r) \tag{24}$$

exist.

1. *Suppose $\lim_{N\to\infty} c_N = 0$. Then for all $n \in \mathbb{N}$, $(\Pi_{n,N}(\lfloor t/c_N \rfloor) : t \geq 0)$ converges in distribution in the Skorohod topology to an exchangeable coalescent $(\Pi_{n,\infty}(t) : t \geq 0)$ with values in $\mathcal{P}_n$. The transition rates $\lambda_{b;k_1,\ldots,k_r;0}$ of $\Pi_{n,\infty}$ are given by $\Phi_r(k_1, \ldots, k_r)$, and these rates determine all the $\lambda_{b;k_1,\ldots,k_r;s}$.*

2. *If $\lim_{N\to\infty} c_N = c > 0$, then for all $n \in \mathbb{N}$, $(\Pi_{n,N}(m) : m \in \mathbb{N}_0)$ converges in distribution to a discrete time exchangeable coalescent $(\Pi_{n,\infty}(m) : m \in \mathbb{N}_0)$ with values in $\mathcal{P}_n$. The transition probabilities $p_{b;k_1,\ldots,k_r;0}$ of $\Pi_{n,\infty}$ are given by $c \times \Phi_r(k_1, \ldots, k_r)$, and these transition probabilities determine all the $p_{b;k_1,\ldots,k_r;s}$.*

Before we begin with the proof, we show two lemmas that we will need for this proof as well as for the proof of convergence for the measure-valued formulation. They are both shown in Möhle and Sagitov (2001).

Lemma 5.2. *Define*

$$\Psi_{r,s}(k_1, \ldots, k_r) := \lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{k_1} \ldots (\nu_{r,N})_{k_r} \nu_{r+1,N} \ldots \nu_{r+s,N})}{N^{k_1+\cdots+k_r-r} c_N}$$

if the limit exists. Then $\Psi_{r,0} = \Phi_r$, and the existence of the Φ_r implies the existence of all $\Psi_{r,s}$, since the $\Psi_{r,s}$ satisfy the following recursion:

$$\begin{aligned} \Psi_{r,s+1}(k_1, \ldots, k_r) =& \Psi_{r,s}(k_1, \ldots, k_r) - \sum_{j=1}^{r} \Psi_{r,s}(k_1, \ldots, k_{j-1}, k_j + 1, k_{j+1}, \ldots, k_r) \\ & - s\Psi_{r+1,s-1}(k_1, \ldots, k_r, 2) \end{aligned} \tag{25}$$

for all $s \in \mathbb{N}_0, r \in \mathbb{N}, k_1, \ldots, k_r \geq 2$ (where $\Psi_{r,-1} := 0$).

Proof. We have

$$
\begin{aligned}
&(N-r-s)\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s+1,N})\\
&\quad\overset{(22)}{=}\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N}(\nu_{r+s+1,N}+\dots+\nu_{N,N}))\\
&\quad\overset{(20)}{=}\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N}(N-\nu_{1,N}-\dots-\nu_{r+s,N}))\\
&\quad=\mathbb{E}[(\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N}(N-k_1-\dots-k_r-s\\
&\qquad\qquad\qquad\qquad\qquad\qquad-\sum_{i=1}^{r}(\nu_{i,N}-k_i)-\sum_{i=r+1}^{r+s}(\nu_{i,N}-1))]\\
&\quad=(N-k_1-\dots-k_r-s)\mathbb{E}[(\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N}]\\
&\qquad-\sum_{i=1}^{r}\mathbb{E}[(\nu_{1,N})_{k_1}\dots(\nu_{i,N})_{k_i+1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N}]\\
&\qquad-s\mathbb{E}[(\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}(\nu_{r+1,N})_2\nu_{r+2,N}\dots\nu_{r+s,N}]
\end{aligned}
$$

Then we divide by $N^{k_1+\dots+k_r+1-r}c_N$ and let N tend to infinity to obtain the recursion. □

Lemma 5.3. *Define*

$$\gamma_b := \lim_{N\to\infty}\frac{1-\mathbb{E}(\nu_{1,N}\dots\nu_{b,N})}{c_N}$$

if the limit exists. Then the existence of the Φ_r implies the existence of all γ_b, since the γ_b satisfy the following recursion:

$$\gamma_{b+1}=\gamma_b+b\Psi_{1,b-1}(2)$$

Proof. This is shown exactly like the previous lemma. □

Proof of Theorem 5.1. 1. Let $\pi\in\mathcal{P}_n$ with b blocks and let $\eta\in\mathcal{P}_n$ be a $(b;k_1,\dots,k_r;s)$-collision of π. Then the transition probability of $\Pi_{n,N}$ from π to η is given by

$$
\begin{aligned}
p^N_{\pi\eta}&:=\mathbb{P}(\Pi_{n,N}(m+1)=\eta|\Pi_{n,N}(m)=\pi)\\
&\overset{(23)}{=}\frac{(N)_{r+s}}{(N)_b}\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N})
\end{aligned}
\tag{26}
$$

In particular, the transition probability from π to π is given by

$$\mathbb{E}(\nu_{1,N}\dots\nu_{b,N})$$

2. Suppose $\lim_{N\to\infty}c_N=0$. In this case we can apply Theorem 3.1 to obtain the weak convergence: Define for every $(n;k_1,\dots,k_r;s)$-partition η $\Psi_\eta:=\Psi_{r,s}(k_1,\dots,k_r)$. Let $\mathcal{D}(A):=\{F:\mathcal{P}_n\to\mathbb{R}\}$ and

$$AF(\pi):=\sum_{\eta\in\mathcal{P}_n}\Psi_\eta(F(\mathrm{Coag}(\pi,\eta))-F(\pi))$$

This martingale problem has at most one solution since $\mathcal{P}_n$ is finite (cf. example in Appendix C). We can rewrite the recursion (25) to see that the $\Psi_{r,s}$ satisfy the same consistency relation (10) as the $\lambda_{b;k_1,\dots,k_r;s}$. This allows us to rewrite A. We define for every $\pi \in \mathcal{P}_n$ with $\#\pi = b$ and for every $(b;k_1,\dots,k_r;s)$-collision η of π $\Psi_{\pi\eta} := \Psi_{r,s}(k_1,\dots,k_r)$. Then we have

$$AF(\pi) := \sum_{\substack{\eta\in\mathcal{P}_n:\\ \eta\supseteq\pi}} \Psi_{\pi\eta}(F(\eta)-F(\pi))$$

On the other side, A_N is given by

$$A_N F(\pi) = \sum_{\substack{\eta\in\mathcal{P}_n:\\ \eta\supseteq\pi}} \frac{p^N_{\pi\eta}}{c_N}(F(\eta)-F(\pi))$$

To apply Theorem 3.1, all we still need to check is whether for every $\pi \subsetneq \eta \in \mathcal{P}_n$ we have

$$\lim_{N\to\infty} \frac{p^N_{\pi\eta}}{c_N} = \Psi_{\pi\eta}$$

But if η is a $(b;k_1,\dots,k_r;s)$-collision of π, then by (26) we have

$$\begin{aligned}
\lim_{N\to\infty} \frac{p^N_{\pi\eta}}{c_N} &= \lim_{N\to\infty} \frac{(N)_{r+s}}{(N)_b c_N}\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N}) \\
&= \lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N})}{N^{b-r-s}c_N} \\
&= \lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r}\nu_{r+1,N}\dots\nu_{r+s,N})}{N^{k_1-\dots-k_r-r}c_N} = \Psi_{r,s}(k_1,\dots,k_r) = \Psi_{\pi\eta} \qquad (27)
\end{aligned}$$

3. Now suppose $\lim_{N\to\infty} c_N = c > 0$. Set for $\pi \nsubseteq \eta$ $\Psi_{\pi\eta} := 0$ and set for every π with $\#\pi = b$ $\Psi_{\pi\pi} := \lim_{N\to\infty}\mathbb{E}(\nu_{1,N}\dots\nu_{b,N})/c_N$ (which exists under the assumption $\lim_{N\to\infty} c_N > 0$ because of Lemma 5.3). We have for every $\pi \in \mathcal{P}_n$

$$\sum_{\substack{\eta\in\mathcal{P}_n:\\ \eta\supseteq\pi}} c\Psi_{\pi\eta} = \sum_{\substack{\eta\in\mathcal{P}_n:\\ \eta\supseteq\pi}} \lim_{N\to\infty} c\frac{p^N_{\pi\eta}}{c_N} = \lim_{N\to\infty} \sum_{\substack{\eta\in\mathcal{P}_n:\\ \eta\supseteq\pi}} p^N_{\pi\eta} = 1$$

So let $(\Pi_{n,\infty}(m) : m \in \mathbb{N}_0)$ be the Markov chain with transition matrix $P = (c\Psi_{\pi\eta})_{\pi,\eta\in\mathcal{P}_n}$. Then the finite-dimensional distributions of $\Pi_{n,N}$ converge to the finite-dimensional distributions of $\Pi_{n,\infty}$. But of course for discrete time processes, convergence of finite-dimensional distributions is equivalent to convergence in distribution of the processes (cf. e.g. Proposition 4.6 in Chapter 3 of Ethier and Kurtz (1986) or p. 19 of Billingsley (1968)).

□

Now we present the most general convergence result for the measure-valued formulation: We recall that we assumed that every individual has a genetic type that can be described by an element x in some metric compact space E. Further we assumed that every individual inherits the genetic type of its ancestor. The distribution of genetic types in the model with N elements is at each time given by a measure $\mu \in M^N(E) = \{\sum_{i=1}^N \delta_{x_i}/N : (x_1,\dots,x_N) \in E^N\}$. We start in generation 0 with a random distribution of genetic types, and then we follow the development of the distribution of genetic types forward in time. Like that we will obtain a measure-valued stochastic process $(\rho_N(m) : m \in \mathbb{N}_0)$.

Theorem 5.4. *Suppose that for every $r \in \mathbb{N}, k_1, \ldots, k_r \geq 2$, the limits*

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{k_1} \ldots (\nu_{r,N})_{k_r})}{N^{k_1+\cdots+k_r-r} c_N} =: \Phi_r(k_1, \ldots, k_r) \tag{28}$$

exist and suppose that the distribution of $\rho_N(0)$ converges weakly to some distribution ν on $\mathcal{M}_1(E)$.

1. *Suppose $\lim_{N\to\infty} c_N = 0$. Then $(\rho_N(\lfloor t/c_N \rfloor) : t \geq 0)$ converges in distribution in the Skorohod topology to a generalized Fleming-Viot process $(\rho(t) : t \geq 0)$. The rates $\lambda_{b;k_1,\ldots,k_r;0}$ of ρ are given by $\Phi_r(k_1, \ldots, k_r)$, and these rates determine all the $\lambda_{b;k_1,\ldots,k_r;s}$.*

2. *If $\lim_{N\to\infty} c_N = c > 0$, then $(\rho_N(m) : m \in \mathbb{N}_0)$ converges in distribution to a discrete time generalized Fleming-Viot process $(\rho(m) : m \in \mathbb{N}_0)$. The transition probabilities $p_{b;k_1,\ldots,k_r;0}$ of ρ are given by $c \times \Phi_r(k_1, \ldots, k_r)$, and these transition probabilities determine all the $p_{b;k_1,\ldots,k_r;s}$.*

Proof. 1. Let for $i \leq N$ $Y_i^N(m) \in E$ be the genetic type of individual i in generation m of the population with N individuals. Let $f(x_1, \ldots, x_p) = \prod_{i=1}^p \varphi_i(x_i)$ with $\varphi_i \in C(E)$ for all i. Let $(x_1, \ldots, x_N) \in E^N$ and let $\mu_N := \sum_{i=1}^N \frac{1}{N}\delta_{x_i}$. We want to evaluate

$$\frac{1}{c_N}\{\mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0) = \mu_N\right] - \Phi_f(\mu_N)\}$$

to apply Theorem 3.1. We have

$$\begin{aligned}\mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0) = \mu_N\right] &= \mathbb{E}\left[\prod_{i=1}^p \sum_{j=1}^N \frac{1}{N}\varphi_i(Y_j^N(1)) \middle| Y^N(0) = (x_1, \ldots, x_N)\right] \\ &= N^{-p} \sum_{j_1,\ldots,j_p=1}^N \mathbb{E}\left[\prod_{i=1}^p \varphi_i(Y_{j_i}^N(1)) \middle| Y^N(0) = (x_1, \ldots, x_N)\right]\end{aligned}$$

Let for $i, j \leq N$ $A_{i,j}^N$ be the event that the j-th individual in generation 1 is a descendant of the i-th individual in generation 0. Then we have

$$\begin{aligned}&\mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0) = \mu_N\right] \\ &= N^{-p} \sum_{\pi\in\mathcal{P}_p} \sum_{\substack{j_1,\ldots,j_{\#\pi}=1 \\ \text{all distinct}}}^N \sum_{l_1,\ldots,l_{\#\pi}=1}^N \\ &\qquad \mathbb{E}\left[\prod_{i\in\pi_1} \varphi_i(Y_{j_1}^N(1)) \cdots \prod_{i\in\pi_{\#\pi}} \varphi_i(Y_{j_{\#\pi}}^N(1)) \mathbb{1}_{\cap_{i=1}^{\#\pi} A_{l_i,j_i}^N} \middle| Y^N(0) = (x_1, \ldots, x_N)\right]\end{aligned}$$

Of course A_{l_i,j_i}^N is independent of $Y^N(0)$ by our assumptions, and therefore

$$\begin{aligned}&\mathbb{E}\left[\prod_{i\in\pi_1} \varphi_i(Y_{j_1}^N(1)) \cdots \prod_{i\in\pi_{\#\pi}} \varphi_i(Y_{j_{\#\pi}}^N(1)) \mathbb{1}_{\cap_{i=1}^{\#\pi} A_{l_i,j_i}^N} \middle| Y^N(0) = (x_1, \ldots, x_N)\right] \\ &= \mathbb{E}\left[\prod_{i\in\pi_1} \varphi_i(Y_{j_1}^N(1)) \cdots \prod_{i\in\pi_{\#\pi}} \varphi_i(Y_{j_{\#\pi}}^N(1)) \middle| \{Y^N(0) = (x_1, \ldots, x_N)\} \cap_{i=1}^{\#\pi} A_{l_i,j_i}^N\right] \mathbb{P}\left[\cap_{i=1}^{\#\pi} A_{l_i,j_i}^N\right] \\ &= \prod_{i\in\pi_1} \varphi_i(x_{l_1}) \cdots \prod_{i\in\pi_{\#\pi}} \varphi_i(x_{l_{\#\pi}}) \mathbb{P}\left[\cap_{i=1}^{\#\pi} A_{l_i,j_i}^N\right]\end{aligned}$$

which yields

$$
\begin{aligned}
&\mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0)=\mu_N\right] \\
&\quad = N^{-p}\sum_{\pi\in\mathcal{P}_p}\sum_{\eta\in\mathcal{P}_{\#\pi}}\sum_{\substack{j_1,\dots,j_{\#\pi}=1\\ \text{all distinct}}}^{N}\sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \text{all distinct}}}^{N}\prod_{m=1}^{\#\eta}\left(\prod_{i\in\mathrm{Coag}(\pi,\eta)_m}\varphi_i(x_{l_m})\right)\mathbb{P}\left[\bigcap_{k=1}^{\#\eta}\left(\cap_{i\in\eta_k}A^N_{l_k,j_i}\right)\right] \\
&\quad = N^{-p}\sum_{\pi\in\mathcal{P}_p}\sum_{\eta\in\mathcal{P}_{\#\pi}}(N)_{\#\pi}\sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \text{all distinct}}}^{N}\prod_{m=1}^{\#\eta}\left(\prod_{i\in\mathrm{Coag}(\pi,\eta)_m}\varphi_i(x_{l_m})\right)\frac{\mathbb{E}[(\nu_{l_1,N})_{|\eta_1|}\dots(\nu_{l_{\#\eta},N})_{|\eta_{\#\eta}|}]}{(N)_{\#\pi}}
\end{aligned}
\tag{29}
$$

2. Let $\pi\neq\mathbf{0}_p$ and $\eta\neq\mathbf{0}_{\#\pi}$. Set $C:=\prod_{i=1}^p||\varphi_i||$. Then we have

$$
\begin{aligned}
&\left|N^{-p}(N)_{\#\pi}\sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \text{all distinct}}}^{N}\prod_{m=1}^{\#\eta}\left(\prod_{i\in\mathrm{Coag}(\pi,\eta)_m}\varphi_i(x_{l_m})\right)\frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{(N)_{\#\pi}}\right| \\
&\qquad\leq N^{-p}(N)_{\#\eta}C\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}] \\
&\qquad\leq\frac{C}{N^{p-\#\pi}}\frac{\mathbb{E}[(\nu_{l_1,N})_{|\eta_1|}\dots(\nu_{l_{\#\eta},N})_{|\eta_{\#\eta}|}]}{N^{\#\pi-\#\eta}}=o(c_N)
\end{aligned}
\tag{30}
$$

$o(c_N)$ means that this term tends to 0 when it is divided by c_N and when N tends to infinity. This is true because by Lemma 5.2 $\mathbb{E}[(\nu_{l_1,N})_{|\eta_1|}\dots(\nu_{l_{\#\eta},N})_{|\eta_{\#\eta}|}]/(N^{\#\pi-\#\eta}c_N)$ converges and because $\#\pi<p$.

3. Let $\pi=\mathbf{0}_p$ and $\eta\in\mathcal{P}_p$. Then

$$
\begin{aligned}
&N^{-p}(N)_p\sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \text{all distinct}}}^{N}\prod_{m=1}^{\#\eta}\left(\prod_{i\in\mathrm{Coag}(\pi,\eta)_m}\varphi_i(x_{l_m})\right)\frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{(N)_p} \\
&\quad=\sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \text{all distinct}}}^{N}\prod_{m=1}^{\#\eta}\left(\frac{\prod_{i\in\eta_m}\varphi_i(x_{l_m})}{N}\right)\frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^{p-\#\eta}} \\
&\quad=<\prod_{i\in\eta_1}\varphi_i,\mu_N>\cdots<\prod_{i\in\eta_{\#\eta}}\varphi_i,\mu_N>\frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^{p-\#\eta}} \\
&\qquad-\sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \exists i\neq j:l_i=l_j}}^{N}\prod_{m=1}^{\#\eta}\left(\prod_{i\in\eta_m}\varphi_i(x_{l_m})\right)\frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^p}
\end{aligned}
\tag{31}
$$

But for $\eta \neq \mathbf{0}_p$ the "minus-term" is of order $o(c_N)$ by Lemma 5.2:

$$\begin{aligned}
&\left| \sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \exists i\neq j: l_i=l_j}}^{N} \prod_{m=1}^{\#\eta} \left(\prod_{i\in\eta_m} \varphi_i(x_{l_m}) \right) \frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^p} \right| \\
&\leq \binom{\#\eta}{2} N^{\#\eta-1} C \frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^p} \\
&= \frac{C}{N}\binom{\#\eta}{2} \frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^{p-\#\eta}} = o(c_N)
\end{aligned} \tag{32}$$

For $\eta = \mathbf{0}_p$ we can rewrite the "minus-term" as follows:

$$\begin{aligned}
&\sum_{\substack{l_1,\dots,l_p=1\\ \exists i\neq j: l_i=l_j}}^{N} \prod_{m=1}^{p} \varphi_i(x_{l_m}) \frac{\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]}{N^p} \\
&\quad = \sum_{\sigma\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}} \sum_{\substack{l_1,\dots,l_{\#\sigma}=1\\ \text{all distinct}}}^{N} \prod_{m=1}^{\#\sigma} \left(\prod_{i\in\sigma_m} \varphi_i(x_{l_m}) \right) \frac{\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]}{N^p}
\end{aligned} \tag{33}$$

4. Finally let $\pi \neq \mathbf{0}_p$ but $\eta = \mathbf{0}_{\#\pi}$. Then

$$\begin{aligned}
&N^{-p}(N)_{\#\pi} \sum_{\substack{l_1,\dots,l_{\#\eta}=1\\ \text{all distinct}}}^{N} \prod_{m=1}^{\#\eta} \left(\prod_{i\in\mathrm{Coag}(\pi,\eta)_m} \varphi_i(x_{l_m}) \right) \frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{(N)_{\#\pi}} \\
&= \sum_{\substack{l_1,\dots,l_{\#\pi}=1\\ \text{all distinct}}}^{N} \prod_{m=1}^{\#\pi} \left(\prod_{i\in\pi_m} \varphi_i(x_{l_m}) \right) \frac{\mathbb{E}[\nu_{1,N}\dots\nu_{\#\pi,N}]}{N^p}
\end{aligned} \tag{34}$$

5. We combine (29) - (34) to obtain

$$\begin{aligned}
&\mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0)=\mu_N\right] \\
&= o(c_N) + \sum_{\eta\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}} < \prod_{i\in\eta_1} \varphi_i, \mu_N > \dots < \prod_{i\in\eta_{\#\eta}} \varphi_i, \mu_N > \frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^{p-\#\eta}} \\
&\quad - o(c_N) + \prod_{i=1}^{p} < \varphi_i, \mu_N > \mathbb{E}[\nu_{1,N}\dots\nu_{p,N}] \\
&\quad - \sum_{\sigma\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}} \sum_{\substack{l_1,\dots,l_{\#\sigma}=1\\ \text{all distinct}}}^{N} \prod_{m=1}^{\#\sigma} \left(\prod_{i\in\sigma_m} \varphi_i(x_{l_m}) \right) \frac{\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]}{N^p} \\
&\quad + \sum_{\pi\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}} \sum_{\substack{l_1,\dots,l_{\#\pi}=1\\ \text{all distinct}}}^{N} \prod_{m=1}^{\#\pi} \left(\prod_{i\in\pi_m} \varphi_i(x_{l_m}) \right) \frac{\mathbb{E}[\nu_{1,N}\dots\nu_{\#\pi,N}]}{N^p}
\end{aligned}$$

But

$$\Bigg| - \sum_{\sigma\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}} \sum_{\substack{l_1,\dots,l_{\#\sigma}=1\\ \text{all distinct}}}^{N} \prod_{m=1}^{\#\sigma}\left(\prod_{i\in\sigma_m}\varphi_i(x_{l_m})\right)\frac{\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]}{N^p}$$

$$+ \sum_{\pi\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}} \sum_{\substack{l_1,\dots,l_{\#\pi}=1\\ \text{all distinct}}}^{N} \prod_{m=1}^{\#\pi}\left(\prod_{i\in\pi_m}\varphi_i(x_{l_m})\right)\frac{\mathbb{E}[\nu_{1,N}\dots\nu_{\#\pi,N}]}{N^p}\Bigg|$$

$$\leq \sum_{\pi\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}} (N)_{\#\pi} C\left|\frac{\mathbb{E}[\nu_{1,N}\dots\nu_{\#\pi,N}]-1-(\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]-1)}{N^p}\right|$$

$$\leq \sum_{\pi\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}} N^{\#\pi-p} C\left|(\mathbb{E}[\nu_{1,N}\dots\nu_{\#\pi,N}]-1)-(\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]-1)\right| = o(c_N)$$

by Lemma 5.3 and since for $\pi\neq\mathbf{0}_p$ $\#\pi<p$.
So if we define for $\eta\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}$

$$\Psi_\eta := \lim_{N\to\infty}\frac{\mathbb{E}[(\nu_1,N)_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{N^{p-\#\eta}c_N}$$

(which exists by Lemma 5.2), then

$$\frac{1}{c_N}\left(\mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0)=\mu_N\right]-\prod_{i=1}^{p}<\varphi_i,\mu_N>\right)$$

$$= \sum_{\eta\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}} <\prod_{i\in\eta_1}\varphi_i,\mu_N>\cdots<\prod_{i\in\eta_{\#\eta}}\varphi_i,\mu_N>\frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{c_N N^{p-\#\eta}}$$

$$-\prod_{i=1}^{p}<\varphi_i,\mu_N>\frac{1-\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]}{c_N}+\frac{o(c_N)}{c_N}$$

$$= \sum_{\eta\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}} \Psi_\eta <\prod_{i\in\eta_1}\varphi_i,\mu_N>\cdots<\prod_{i\in\eta_{\#\eta}}\varphi_i,\mu_N> -\gamma_p\prod_{i=1}^{p}<\varphi_i,\mu_N>$$

$$+\frac{o(c_N)}{c_N}+\epsilon(N)$$

with

$$\epsilon(N) = \sum_{\eta\in\mathcal{P}_p\backslash\{\mathbf{0}_p\}}\left(\frac{\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]}{c_N N^{p-\#\eta}}-\Psi_\eta\right)<\prod_{i\in\eta_1}\varphi_i,\mu_N>\cdots<\prod_{i\in\eta_{\#\eta}}\varphi_i,\mu_N>$$

$$-\left(\frac{1-\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]}{c_N}-\gamma_p\right)\prod_{i=1}^{p}<\varphi_i,\mu_N>$$

which tends to 0 when $N\to\infty$, uniformly in μ_N. Note that also the $o(c_N)/c_N$-term tends to 0 uniformly in μ_N. We have

$$\Psi_\eta = \lim_{N\to\infty}\frac{1}{c_N}\frac{(N)_{\#\eta}}{(N)_p}\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]$$

If we consider the partition-valued formulation of the model for a sample of size p, then $(N)_{\#\eta}/(N)_p\mathbb{E}[(\nu_1, N)_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]$ is the transition probability $p^N_{\mathbf{0}_p\eta}$ of $\Pi_{p,N}$ from $\mathbf{0}_p$ to η (cf. (26)). $\mathbb{E}[\nu_{1,N}\dots\nu_{p,N}]$ is the probability $p^N_{\pi\pi}$ of $\Pi_{p,N}$ to stay in π. Therefore

$$\begin{aligned}\sum_{\eta\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}}\Psi_\eta &= \lim_{N\to\infty}\frac{1}{c_N}\sum_{\eta\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}}\frac{(N)_{\#\pi}}{(N)_p}\mathbb{E}[(\nu_{1,N})_{|\eta_1|}\dots(\nu_{\#\eta,N})_{|\eta_{\#\eta}|}]\\ &= \lim_{N\to\infty}\frac{1}{c_N}\sum_{\eta\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}}p^N_{\mathbf{0}_p\eta} = \lim_{N\to\infty}\frac{1}{c_N}(1-p^N_{\mathbf{0}_p\mathbf{0}_p}) = \gamma_p\end{aligned}$$

So finally we obtain

$$\begin{aligned}&\frac{1}{c_N}\left(\mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0)=\mu_N\right]-\Phi_f(\mu_N)\right)\\ &= \sum_{\eta\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}}\Psi_\eta\left(<\prod_{i\in\eta_1}\varphi_i,\mu_N>\dots<\prod_{i\in\eta_{\#\eta}}\varphi_i,\mu_N> - \prod_{i=1}^p<\varphi_i,\mu_N>\right)+\frac{o(c_N)}{c_N}+\epsilon(N)\\ &= \sum_{\eta\in\mathcal{P}_p\setminus\{\mathbf{0}_p\}}\Psi_\eta\int_{E^p}\left(\prod_{i=1}^p\varphi_i(x_{\eta_{[i]}})-\prod_{i=1}^p\varphi_i(x_i)\right)\mu_N^{\otimes p}(dx_1,\dots,dx_p)+\frac{o(c_N)}{c_N}+\epsilon(N)\\ &= G\Phi_f(\mu_N)+\frac{o(c_N)}{c_N}+\epsilon(N)\end{aligned}$$

where G is the operator that we used to introduce the generalized Fleming-Viot process corresponding to the rates Ψ_η. Recall that $\mathcal{M}^N(E)$ was defined as $\{\sum_{i=1}^N\delta_{x_i}/N : (x_1,\dots,x_N)\in E^N\}$. Let T_N be the transition operator of ρ_N (i.e. $T_N\Phi_f(\mu_N) = \mathbb{E}\left[\Phi_f(\rho_N(1))\,|\rho_N(0)=\mu_N\right]$). Then we have

$$\sup_{\mu_N\in\mathcal{M}^N(E)}\left|\frac{1}{c_N}\left(T_N-I\right)\Phi_f(\mu_N)-G\Phi_f(\mu_N)\right|\to 0,\ N\to\infty \tag{35}$$

6. Suppose $c_N\to 0$. We can apply Theorem 3.1 with $E_N=\mathcal{M}^N(E)$ and with π_N being the inclusion map from $\mathcal{M}^N(E)$ to $\mathcal{M}_1(E)$. We obtain that $(\rho_N(\lfloor t/c_N\rfloor) : t\geq 0)$ converges in distribution in the Skorohod-topology to the unique solution to the $(\bar{G},\nu)$-martingale problem, i.e. the generalized Fleming-Viot process with rates Ψ_π, starting with distribution ν.

7. Suppose $\lim_{N\to\infty}c_N = c > 0$. Note that $\mathcal{M}_1(E)^{\mathbb{N}}$ is a compact space as a product of compact spaces. This is easy to see with a diagonal sequence argument, since $\mathcal{M}_1(E)$ is a metric space when equipped with the Prohorov distance. Also, the statement is true by Tychonoff's theorem (which uses the axiom of choice, cf. Munkres (2000), Theorem 37.3). So by Prohorov's theorem (Theorem 2.2 in Chapter 3 of Ethier and Kurtz (1986)), any sequence of discrete time processes with values in $\mathcal{M}_1(E)$ is tight. By Proposition 4.6 it therefore suffices to show that any cluster point of the sequence ρ_N is a solution to the discrete time $(\bar{G},\nu)$-martingale problem (since the discrete time (G,ν)- and $(\bar{G},\nu)$-martingale problems are equivalent, just as in the continuous time case). Let $m\in\mathbb{N}_0$, let h be a bounded and measurable function on $\mathcal{M}_1(E)^{m+1}$, and let $\Phi_f\in\bar{\mathcal{D}}$. Using (35)

and bounded convergence, we obtain

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\left[\left(\Phi_f(\rho_N(m+1)) - \Phi_f(\rho_N(m)) - cG\Phi_f(\rho_N(m))\right) h(\rho_N(0),\dots,\rho_N(m))\right] \\
&= \lim_{N\to\infty} \mathbb{E}\left[\left(\Phi_f(\rho_N(m+1)) - \Phi_f(\rho_N(m)) - (T_N - I)\Phi_f(\rho_N(m))\right) h(\rho_N(0),\dots,\rho_N(m))\right] \\
&= 0
\end{aligned}$$

which completes the proof.

□

5.3 Convergence Criteria

Sometimes it is not easy to check the conditions of Theorem 5.1 respectively Theorem 5.4. So we present two criteria for that. We will not give the proofs here.

The first result is Theorem 4. b) of Möhle (2000).

Proposition 5.5. *Suppose*

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_1, N)_3)}{N^2 c_N} = 0$$

Then c_N tends to 0 when N tends to infinity, and for any b and for any $\pi \in \mathcal{P}_b$ with $\#\pi < b-1$, we have

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{|\pi_1|}\dots(\nu_{\#\pi,N})_{|\#\pi|})}{N^{b-\#\pi} c_N} = 0$$

That means, that the limit in the partition-valued formulation will be Kingman's coalescent, and the limit in the measure-valued formulation will be the classical Fleming-Viot process.

The second result is Theorem 2.1 respectively Remark 1 from Möhle and Sagitov (1998).

Proposition 5.6. *Suppose* $\lim_{N\to\infty} c_N = 0$ *and*

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_2(\nu_{2,N})_2)}{N^2 c_N} = 0$$

Also, suppose that there exists a probability Λ on $[0,1]$ such that

$$\lim_{N\to\infty} \frac{N}{c_N}\mathbb{P}(\nu_{1,N} > Nx) = \int_{[x,1]} y^{-2}\Lambda(dy)$$

for all $x \in (0,1)$ where the limiting function is continuous. Then for any $(b; k_1,\dots,k_r; s)$-partition $\pi \in \mathcal{P}_b$ with $r > 1$ we have

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{|\pi_1|}\dots(\nu_{\#\pi,N})_{|\#\pi|})}{N^{b-\#\pi} c_N} = 0$$

and for a $(b; k_1,\dots,k_r; s)$-partition π with $r = 1$ and $k \geq 2$ we have

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{|\pi_1|}\dots(\nu_{\#\pi,N})_{|\#\pi|})}{N^{b-\#\pi} c_N} = \int_{[0,1]} x^{k-2}(1-x)^{b-k}\Lambda(dx)$$

That means, that the limit will be a Λ-coalescent or a Λ-Fleming-Viot process.

6 Convergence Results for Schweinsberg's Model

We present a realistic population model that was introduced by Schweinsberg (2003). The model is a special case of Cannings' model, and we show convergence results.

Suppose we have a haploid population with non-overlapping discrete generations, an infinite number of generations both in the future and in the past. Suppose every individual in every generation has the same reproduction law, which is independent of the reproduction of all the other individuals. Further we suppose that the population size is restricted due to some external influence. So only a fixed number of the descendants in each generation can survive. This model can be described mathematically in the following way:

Let $(X_i^m : i \in \mathbb{N}, m \in \mathbb{Z})$ be a family of i.i.d. variables with values in $\mathbb{N}_0$. If we are only interested in the distribution of the X_i^m we can therefore omit the index m. We suppose

$$\mathbb{E}(X_1) > 1 \tag{36}$$

We interpret the X_i^m as reproduction laws of a supercritical Galton-Watson process. The Galton-Watson process describes the size of a population. It is given by $Y_0 := N$ and $Y_{k+1} := X_1^k + \cdots + X_{Y_k}^k$. The restriction of the population size can be modelled as follows:

If $X_1^k + \cdots + X_N^k > N$, we choose randomly N individuals which will be the descendants from generation k that actually survive. The size of family i in generation k, $\nu_{i,N}^k$, is thus given by the number of chosen descendants of X_i^k.

If $X_1^k + \cdots + X_N^k < N$, we define $(\nu_{1,N}^k, \dots, \nu_{N,N}^k) := (1, \dots, 1)$. The probability of this event will tend to 0 when N tends to infinity because of (36).

So we have a haploid population with non-overlapping generations, infinitely many both in the past and in the future. The population has a fixed size, the family size vectors $(\nu_{1,N}^k, \dots, \nu_{N,N}^k)$ are i.i.d. and exchangeable. So we are in the setting of Cannings' model, and we can use our previous results.

For most of the results we will suppose that there exists $a > 0$ such that the tail of the distribution of X_1 is of **regular variation with index $-a$**, which means that for any $C > 0$ we have

$$\lim_{k\to\infty} \frac{\mathbb{P}(X_1 \geq Ck)}{\mathbb{P}(X_1 \geq k)} = C^{-a} \tag{37}$$

Cf. Appendix D for an overview of functions of regular variation.

In the original article Schweinsberg (2003), the assumption was slightly stronger. There it was supposed that there would be some constant $C > 0$ such that

$$\mathbb{P}(X_1 \geq k) \sim Ck^{-a} \tag{38}$$

where $\sim$ means that the ratio of the two sides tends to 1 when k tends to infinity. This special case was also presented in Perkowski (2009).

Let for $n \in \mathbb{N}$ $(\Pi_{n,N}(m) : m \in \mathbb{N}_0)$ be the partition-valued formulation of the population model, and let $(\rho_N(m) : m \in \mathbb{N}_0)$ be the measure-valued formulation with values in $\mathcal{M}_1(E)$.

Theorem 6.1. *Suppose (36) and that the distribution of $\rho_N(0)$ converges weakly to some distribution ν on $\mathcal{M}_1(E)$.*

1. *If $\mathbb{E}(X_1^2) < \infty$, then $(\Pi_{n,N}(\lfloor t/c_N \rfloor))_{t\geq 0}$ converges in distribution in the Skorohod topology to Kingman's n-coalescent when $N \to \infty$, and $(\rho_N(\lfloor t/c_N \rfloor) : t \geq 0)$ converges in distribution in the Skorohod topology to the classical Fleming-Viot process starting with distribution ν. If X_1 satisfies the assumption (37) with $a > 2$, then $\mathbb{E}(X_1^2) < \infty$.*

2. *Under assumption (37) with $a = 2$, $(\Pi_{n,N}(\lfloor t/c_N \rfloor))_{t\geq 0}$ converges in distribution in the Skorohod topology to Kingman's n-coalescent when $N \to \infty$, and $(\rho_N(\lfloor t/c_N \rfloor) : t \geq 0)$ converges in distribution in the Skorohod topology to the classical Fleming-Viot process starting with distribution ν.*

3. *(37) with $1 < a < 2$ implies the convergence of $(\Pi_{n,N}(\lfloor t/c_N \rfloor))_{t\geq 0}$ in the Skorohod topology towards a* Beta$(2-a,a)$*-coalescent with values in $\mathcal{P}_n$ when $N \to \infty$. Recall that* Beta*-coalescents are coalescent with multiple asynchronous collisions. Under assumption (37) with $a \in (1,2)$, $(\rho_N(\lfloor t/c_N \rfloor) : t \geq 0)$ converges in distribution in the Skorohod topology to the* Beta$(2-a,a)$*-Fleming-Viot process starting with distribution ν.*

   The transition rates of the limit processes are given by

$$\lambda_{b;k} = \frac{B(k-a, b-k+a)}{B(2-a,a)}$$

4. *(38) (attention: not (37)!) with $a = 1$ implies the convergence of $(\Pi_{n,N}(\lfloor t/c_N \rfloor))_{t\geq 0}$ in the Skorohod topology towards a* Beta$(1,1) = U$*-coalescent with values in $\mathcal{P}_n$ when $N \to \infty$. Under assumption (38) with $a = 1$, $(\rho_N(\lfloor t/c_N \rfloor) : t \geq 0)$ converges in distribution in the Skorohod topology to the U-Fleming-Viot process starting with distribution ν.*

   The transition rates of the limit processes are given by

$$\lambda_{b;k} = B(k-1, b-k+1)$$

5. *Let $0 < a < 1$ and let $\Theta_a(dx)$ be the probability measure on Δ that corresponds to the* PD$(a,0)$*-distribution. Define*

$$\Xi_a(dx) := \sum_{j=1}^{\infty} x_j^2 \Theta_a(dx)$$

   Under assumption (37) with $0 < a < 1$, $(\Pi_{n,N}(m))_{m\in\mathbb{N}_0}$ converges in distribution to a discrete time Ξ_a-coalescent with values in $\mathcal{P}_n$ when $N \to \infty$, and $(\rho_N(m) : m \in \mathbb{N}_0)$ converges in distribution to a discrete time Ξ_a-Fleming-Viot process starting with distribution ν.

   The transition probabilities of the limit processes are given by

$$p_{b;k_1,\ldots,k_r;s} = \frac{a^{r+s-1}(r+s-1)!}{(b-1)!} \prod_{i=1}^{r} (k_i - 1 - a)_{k_i - 1}$$

Remark. Note that for $a = 1$ we do not show the generalisation, but we just quote Schweinsberg's result. This is not because this case is fundamentally different and because in this case the generalisation will not be true. Probably the generalisation is also true in that case, and in fact this can be easily shown if $\mathbb{E}(X_1)$ is finite, or if $\mathbb{P}(X_1 \geq k) = k^{-1}l(k)$ for a function of slow variation l that is bounded and bounded away from 0, or if l is given by a suitable function of the logarithm, e.g. $l = \log$ or $l = 1/\log^n$ or $l(x) = \log\log x^2$.

The reason why we are not able to show the generalisation for general l is that $a = 1$ is a special case in Karamata's theorem (Theorem D.3), in which we can not control the behaviour of $\int_0^x \mathbb{P}(X_1 \geq y)dy$ for $x \to \infty$ as precisely as in the other cases. So this seems to be only a technical difficulty.

The importance of Schweinsberg's work lays in the fact that he introduced a natural population model in which other coalescents than Kingman's coalescent are obtained as limits. This result motivated the deeper study of the Beta-coalescents.

In the proof we will always argue for the convergence of coalescents. But of course those arguments stay valid for Fleming-Viot processes since in both cases we prove convergence with some variation of Theorem 5.1 respectively Theorem 5.4.

6.1 Preliminary Results

We remark that assumption (37) yields the existence of a function of slow variation l such that

$$\mathbb{P}(X_1 \geq k) = k^{-a} l(k) \quad \forall k \geq 1 \tag{39}$$

Of course in this case l has to satisfy $l(k) \leq k^a$ for all $k \geq 1$ and therefore it is locally bounded.

Lemma 6.2. *Let $g : \mathbb{N}_0 \to \mathbb{R}$ and let X be a random variable with values in $\mathbb{N}_0$. Then*

$$\sum_{k=0}^{N} g(k)\mathbb{P}(X = k) = g(0) - g(N)\mathbb{P}(X \geq N + 1) + \sum_{k=1}^{N} [g(k) - g(k-1)]\mathbb{P}(X \geq k)$$

If $\lim_{N\to\infty} g(N)\mathbb{P}(X \geq N + 1) = 0$, we obtain

$$\mathbb{E}(g(X)) = g(0) + \sum_{k=1}^{\infty} [g(k) - g(k-1)]\mathbb{P}(X \geq k)$$

This lemma is proven by summation by parts. Before continuing, we introduce a new notation: We define $\mu := \mathbb{E}(X_1)$ and $S_N := X_1 + \cdots + X_N$.

Lemma 6.3. *If $\mu > 1$, then there is an $A < 1$ such that $\mathbb{P}(S_N \leq N) \leq A^N$ for all $N \in \mathbb{N}$.*

Proof. Let $\rho(r) := \mathbb{E}(r^{X_1})$, $r \in [0, 1]$, be the generating function of X_1. ρ is continuously differentiable on $(0, 1)$, and we have $\rho'(1) = \mu$ if $\mu < \infty$ but also if $\mu = \infty$ (cf. for example Klenke (2008), Theorem 3.2). So $\rho(1) = 1$ and $\rho'(1) > 1$. Therefore there exists $r < 1$ such that $\rho(r) < r$. We define $A := \frac{\rho(r)}{r}$. With Markov's inequality and because the X_i are i.i.d. we obtain

$$\mathbb{P}(S_N \leq N) \leq \frac{\mathbb{E}(r^{S_N})}{r^N} = A^N$$

□

The following lemma is essential for the proofs of all parts of Theorem 6.1 since it expresses the important limits

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{k_1} \dots (\nu_{r,N})_{k_r})}{N^{k_1+\cdots+k_r-r} c_N}$$

in terms of the X_i.

Lemma 6.4. *For $r \geq 1$, $k_1, \dots, k_r \geq 2$, we have*

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_{k_1} \dots (\nu_{r,N})_{k_r})}{N^{k_1+\cdots+k_r-r} c_N} = \lim_{N\to\infty} \frac{N^r}{c_N} \mathbb{E}\left(\frac{(X_1)_{k_1} \dots (X_r)_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \right) \tag{40}$$

This means that the existence of one of those limits implies the existence of the other one, and in this case the two sides are equal. Further we have

$$c_N \sim N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) \tag{41}$$

and there exists $A_1 > 0$ such that

$$c_N \geq \frac{A_1}{N} \quad \forall N \geq 1 \tag{42}$$

Proof. 1. We order the individuals of a fixed generation randomly. Independently of that we order the individuals of the preceding generation randomly. Let $B_{k_1,\dots,k_r}$ be the event that the k_1 first individuals of the present generation descend from the first individual of the preceding generation, the next k_2 individuals descend from the second individual of the preceding generation, etc. We have

$$\mathbb{P}(B_{k_1,\dots,k_r}) = \frac{\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r})}{(N)_{k_1+\dots+k_r}} \tag{43}$$

Further we have

$$\begin{aligned}\mathbb{P}(B_{k_1,\dots,k_r}) &= \mathbb{E}(\mathbb{P}(B_{k_1,\dots,k_r}\cap\{\mathcal{S}_N\geq N\}|X_1,\dots,X_N)) + \mathbb{P}(B_{k_1,\dots,k_r}\cap\{S_N<N\})\\ &= \mathbb{E}\left(\frac{(X_1)_{k_1}\dots(X_r)_{k_r}}{(S_N)_{k_1+\dots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right) + \mathbb{P}(B_{k_1,\dots,k_r}\cap\{S_N<N\}) \end{aligned}\tag{44}$$

2. For c_N we have the following inequality:

$$\begin{aligned}c_N &= \frac{N\mathbb{E}((\nu_{1,N})_2)}{(N)_2} = N\mathbb{P}(B_2) \overset{(44)}{\geq} N\mathbb{E}\left(\frac{(X_1)_2}{(S_N)_2}\mathbb{1}_{\{S_N\geq N\}}\right) \geq N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right)\\ &\geq \frac{N}{2}\mathbb{E}\left(\left(\frac{X_1}{S_N}\right)^2\mathbb{1}_{\{X_1\geq 2,S_N\geq N\}}\right)\end{aligned}$$

Jensen's inequality yields

$$\begin{aligned}&\geq \frac{N}{2}\left[\mathbb{E}\left(\frac{X_1}{S_N}\mathbb{1}_{\{X_1\geq 2,S_N\geq N\}}\right)\right]^2\\ &= \frac{N}{2}\left[\mathbb{E}\left(\frac{X_1}{S_N}\middle| X_1\geq 2, S_N\geq N\right)\mathbb{P}(X_1\geq 2, S_N\geq N)\right]^2\\ &\geq \frac{N}{2}\left[\mathbb{E}\left(\frac{X_1}{S_N}\middle| X_1\geq 2, S_N\geq N\right)(\mathbb{P}(X_1\geq 2) - A^N)\right]^2\end{aligned}$$

since the X_i are i.i.d. this is

$$\geq \frac{N}{2}\left[\frac{\mathbb{P}(X_1\geq 2) - A^N}{N}\right]^2 = \frac{(\mathbb{P}(X_1\geq 2)-A^N)^2}{2N}$$

Let N_0 be such that $\mathbb{P}(X_1\geq 2) - A^N > 0$ for $N \geq N_0$. If we define

$$A_1 := \min\{(\mathbb{P}(X_1\geq 2) - A^{N_0})^2/2, 1c_1, \dots, N_0 c_{N_0}\}$$

then $A_1 > 0$ and

$$c_N \geq \frac{A_1}{N} \quad \forall N \geq 1$$

3. Equation (43) yields

$$\frac{N^r}{c_N}\mathbb{P}(B_{k_1,\ldots,k_r}) \sim \frac{\mathbb{E}((\nu_{1,N})_{k_1}\ldots(\nu_{r,N})_{k_r})}{N^{k_1+\cdots+k_r-r}c_N}$$

To prove (40) it therefore suffices to show

$$\lim_{N\to\infty}\frac{N^r}{c_N}\mathbb{P}(B_{k_1,\ldots,k_r}) = \lim_{N\to\infty}\frac{N^r}{c_N}\mathbb{E}\left(\frac{(X_1)_{k_1}\ldots(X_r)_{k_r}}{S_N^{k_1+\cdots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right)$$

Since $\mathbb{P}(S_N \leq N) \leq A^N$ and $c_N \geq A_1/N$, we have $\lim_{N\to\infty}\frac{N^r}{c_N}\mathbb{P}(B_{k_1,\ldots,k_r}\cap\{S_N < N\}) = 0$. With (44) we obtain

$$\begin{aligned}\lim_{N\to\infty}\frac{N^r}{c_N}\mathbb{P}(B_{k_1,\ldots,k_r}) &\overset{(44)}{=} \lim_{N\to\infty}\frac{N^r}{c_N}E\left(\frac{(X_1)_{k_1}\ldots(X_r)_{k_r}}{(S_N)_{k_1+\cdots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right)\\ &= \lim_{N\to\infty}\frac{N^r}{c_N}\mathbb{E}\left(\frac{(X_1)_{k_1}\ldots(X_r)_{k_r}}{S_N^{k_1+\cdots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right)\end{aligned}$$

4. (41) is a special case of (40) with $r = 1$ and $k_1 = 2$ since in this case the left side of (40) is equal to 1.

$\square$

We will need another estimation of c_N for which the proof is a little technical.

Lemma 6.5. *If $\mu < \infty$, then there exists $A_2 > 0$ such that*

$$c_N \geq A_2 N\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right)$$

for all large enough N.

Proof. We have

$$\begin{aligned}N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) &\geq N\mathbb{E}\left(\frac{(X_1)_2}{(X_1+2(N-1)\mu)^2}\mathbb{1}_{\{S_N\geq N\}}\mathbb{1}_{\{X_2+\cdots+X_N\leq 2(N-1)\mu\}}\right)\\ &\geq N\mathbb{E}\left(\frac{(X_1)_2}{(X_1+2(N-1)\mu)^2}\mathbb{1}_{\{X_2+\cdots+X_N\leq 2(N-1)\mu\}}\right) - N\mathbb{P}(S_N < N)\\ &\geq N\mathbb{E}\left(\frac{(X_1)_2}{(X_1+2(N-1)\mu)^2}\right)\mathbb{P}(X_2+\cdots+X_N \leq 2(N-1)\mu) - NA^N\\ &\text{with Markov's inequality we get } \mathbb{P}(X_2+\cdots+X_N \leq 2(N-1)\mu\}) \geq 1/2\\ &\geq \frac{N}{8\mu^2}\mathbb{E}\left(\frac{(X_1)_2}{(X_1+N)^2}\right) - NA^N \geq N\left(\frac{1}{32\mu^2}\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2,N^2\}}\right) - A^N\right)\end{aligned}$$

But there exists a $c > 0$ such that

$$\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right) \geq \frac{c}{N^2}$$

for all N large enough:

$$\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right) = \sum_{k=0}^{N} \mathbb{P}(X_1 = k)\frac{k(k-1)}{N^2} + \sum_{k=N+1}^{\infty} \mathbb{P}(X_1 = k)\frac{k^2 - k}{k^2}$$
$$\geq \frac{1}{N^2}\sum_{k=0}^{N} \mathbb{P}(X_1 = k) \geq \frac{\mathbb{P}(X_1 \leq L)}{N^2}$$

for $L \leq N$. We choose L such that $\mathbb{P}(X_1 \leq L) > 0$ and we obtain the desired inequality for all $N \geq L$ with $c := \mathbb{P}(X_1 \leq L)$. This inequality yields

$$A^N = o\left(\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right)\right)$$

when $N \to \infty$. This means that for all N that are large enough we have

$$A^N \leq \frac{1}{64\mu^2}\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right)$$

With (41) we can find a $c' > 0$ such that

$$c_N \geq c' N \mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N \geq N\}}\right) \geq N\frac{c'}{64\mu^2}\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right)$$

for large enough N. We define $A_1 := \frac{c'}{64\mu^2}$. $\square$

Lemma 6.6. *If Y is a positive random variable such that $P(Y \geq k)$ is of regular variation with index $-a < -k$ for some $k \in \mathbb{N}$, then $\mathbb{E}(Y^k) < \infty$.*

Proof. We have

$$\mathbb{E}(Y^k) = \int_0^\infty kx^{k-1}\mathbb{P}(Y \geq x)dx = \int_0^\infty kx^{k-1}x^{-a}l(x)dx$$

for some function of slow variation l by (39). By Karamata's theorem (Theorem D.3), we have

$$\int_y^\infty x^{k-a-1}l(x)dx \sim \frac{y^{k-a}l(y)}{a-k}$$

which tends to 0 when y tends to infinity by Proposition D.1. At the same time

$$\int_0^y kx^{k-1}\mathbb{P}(Y \geq x)dx$$

is finite for every finite y, which yields the finiteness of $\int_0^\infty kx^{k-1}\mathbb{P}(Y \geq x)dx$. $\square$

6.2 Proof of Theorem 6.1, 1.

By Lemma 6.6, if X_1 satisfies (37) with $a > 2$, then $\mathbb{E}(X_1^2) < \infty$.

We would like to use Proposition 5.5 to show convergence to the n-coalescent. So we need to show $\lim_{N\to\infty} \mathbb{E}((\nu_{1,N})_3)/(N^2 c_N) = 0$. With Lemma 6.4 this is equivalent to

$$\lim_{N\to\infty} \frac{N}{c_N}\mathbb{E}\left(\frac{(X_1)_3}{S_N^3}\mathbb{1}_{\{S_N\geq N\}}\right) = 0$$

And since $c_N \geq A_1/N$ by (42), it suffices to show that

$$\lim_{N\to\infty} N^2\mathbb{E}\left(\frac{(X_1)_3}{S_N^3}\mathbb{1}_{\{S_N\geq N\}}\right) = 0$$

We have

$$\begin{aligned} N^2\mathbb{E}\left(\frac{(X_1)_3}{S_N^3}\mathbb{1}_{\{S_N\geq N\}}\right) &\leq N^2\mathbb{E}\left(\frac{X_1^3}{\max\{X_1^3, N^3\}}\right) \\ &= N^2\left(\sum_{k=0}^{N-1}\frac{k^3}{N^3}\mathbb{P}(X_1 = k) + \sum_{k=N}^{\infty}\mathbb{P}(X_1 = k)\right) \\ &= \frac{1}{N}\sum_{k=0}^{N-1}k^3\mathbb{P}(X_1 = k) + N^2\mathbb{P}(X_1 \geq N) \end{aligned} \tag{45}$$

The second term tends to 0 when $N \to \infty$ since $\mathbb{E}(X_1^2) < \infty$. The first term also tends to 0: Let $L \leq N$.

$$\begin{aligned} \frac{1}{N}\sum_{k=0}^{N-1}k^3\mathbb{P}(X_1 = k) &= \frac{1}{N}\sum_{k=0}^{L-1}k^3\mathbb{P}(X_1 = k) + \frac{1}{N}\sum_{k=L}^{N-1}k^3\mathbb{P}(X_1 = k) \\ &\leq \frac{L}{N}\sum_{k=0}^{L-1}k^2\mathbb{P}(X_1 = k) + \sum_{k=L}^{N-1}k^2\mathbb{P}(X_1 = k) \leq \frac{L\mathbb{E}(X_1^2)}{N} + \mathbb{E}(X_1^2\mathbb{1}_{\{X_1\geq L\}}) \end{aligned}$$

Since we can choose L arbitrarily large, this expression tends to 0 for $N \to \infty$. Therefore we proved the first part of Theorem 6.1.

6.3 Proof of Theorem 6.1, 2.

1. Under assumption (37) with $a = 2$ we have $\mu < \infty$ (cf. Lemma 6.6). Therefore we can apply Lemma 6.5. We will need a preliminary result: Under (37) with $a = 2$, we have

$$\lim_{N\to\infty}\frac{l(N)}{Nc_N} = 0 \tag{46}$$

where l is the function of slow variation satisfying $P(X_1 \geq k) = k^{-2}l(k)$. This is equivalent to $\lim_{N\to\infty} Nc_N/l(N) = \infty$.

$$
\begin{aligned}
\frac{Nc_N}{l(N)} &\overset{\text{Lemma 6.5}}{\geq} \frac{NA_2N\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2,N^2\}}\right)}{l(N)} \geq \frac{N^2A_2\sum_{k=0}^{N}(k)_2\mathbb{P}(X_1=k)/N^2}{l(N)} \\
&\overset{\text{Lemma 6.2}}{=} \frac{A_2}{l(N)}\left(-N(N-1)\mathbb{P}(X_1\geq N+1)+\sum_{k=1}^{N}\underbrace{((k)_2-(k-1)_2)}_{=2(k-1)}\mathbb{P}(X_1\geq k)\right) \\
&\geq \frac{A_2}{l(N)}\left(-N(N-1)(N+1)^{-2}l(N+1)+\int_2^N 2(x-1)\mathbb{P}(X_1\geq x)dx\right) \\
&\geq A_2\left(-N(N-1)(N+1)^{-2}\frac{l(N+1)}{l(N)}+\frac{2}{l(N)}\int_2^N 2\left(x-\frac{1}{2}x\right)x^{-2}l(x)dx\right) \\
&= A_2\left(-N(N-1)(N+1)^{-2}\frac{l(N+1)}{l(N)}+\frac{2}{l(N)}\int_2^N \frac{l(x)}{x}dx\right)
\end{aligned}
$$

The first term between the brackets tends to -1, and by Theorem D.3, the second term tends to $+\infty$ when N tends to infinity. Since $A_2 > 0$, this yields (46).

2. To obtain the convergence of $(\Pi_{n,N}(\lfloor t/c_N\rfloor))$ to the n-coalescent, it suffices to show that

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{E}\left(\frac{(X_1)_3}{S_N^3}\mathbb{1}_{\{S_N\geq N\}}\right)=0$$

(cf. the proof of Theorem 6.1, 1.). We use (45) to obtain

$$\frac{N}{c_N}\mathbb{E}\left(\frac{(X_1)_3}{S_N^3}\mathbb{1}_{\{S_N\geq N\}}\right)\leq\frac{1}{N^2c_N}\sum_{k=0}^{N-1}k^3\mathbb{P}(X_1=k)+\frac{N}{c_N}\mathbb{P}(X_1\geq N)$$

For the first term we get

$$
\begin{aligned}
\lim_{N\to\infty}\frac{1}{N^2c_N}\sum_{k=0}^{N-1}k^3\mathbb{P}(X_1=k) &\overset{\text{Lemma 6.2}}{\leq} \lim_{N\to\infty}\frac{1}{N^2c_N}\sum_{k=1}^{N-1}(k^3-(k-1)^3)\mathbb{P}(X_1\geq k) \\
&\leq \lim_{N\to\infty}\frac{1}{N^2c_N}\int_0^{N-1}3(x+1)^2\mathbb{P}(X_1\geq x)dx \\
&\leq \lim_{N\to\infty}\frac{\text{const}}{N^2c_N}+\frac{1}{N^2c_N}\int_1^N 3(x+x)^2\mathbb{P}(X_1\geq x)dx \\
&\overset{(42)}{\leq} \lim_{N\to\infty}\frac{\text{const}}{NA_1}+\frac{12Nl(N)}{N^2c_N}\int_1^N l(x)dx\Big/(Nl(N)) \\
&\overset{\text{Theorem D.3}}{=} \lim_{N\to\infty}12\frac{l(N)}{Nc_N}\overset{(46)}{=}0
\end{aligned}
$$

and the second term tends to 0 as well:

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}(X_1\geq N)=\lim_{N\to\infty}\frac{l(N)}{Nc_N}\overset{(46)}{=}0$$

So the proof of Theorem 6.1, 2. is complete.

6.4 Proof of Theorem 6.1, 3.

The following result is true under assumption (37) with $a \leq 1 < 2$:

Lemma 6.7. *If (39) is satisfied with $1 \leq a < 2$, then*

$$\lim_{M\to\infty} \frac{1}{\mathbb{P}(X_1 \geq M)} \mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right) = \lim_{M\to\infty} \frac{M^a}{l(M)} \mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right) = aB(2-a, a)$$

Proof. 1. We can express

$$\lim_{M\to\infty} \frac{M^a}{l(M)} \mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right)$$

in terms of

$$\lim_{M\to\infty} 2M^{1+a} \int_L^\infty \frac{x^{1-a} l(x)}{(x+M)^3} dx$$

This means that the existence of the second limit yields the existence of the first limit, and in this case the two are equal: We use Lemma 6.2 to obtain

$$\mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right) = \sum_{k=1}^{\infty} \left(\frac{(k)_2}{(k+M)^2} - \frac{(k-1)_2}{(k-1+M)^2}\right) \mathbb{P}(X_1 \geq k)$$

Let $\epsilon > 0$. We choose L large enough such that for $k, M \geq L$

$$\begin{aligned}(1-\epsilon) \int_{k-1}^k \frac{2Mx^{1-a} l(x)}{(x+M)^3} dx &\leq \left(\frac{(k)_2}{(k+M)^2} - \frac{(k-1)_2}{(k-1+M)^2}\right) k^{-a} l(k) \\ &\leq (1+\epsilon) \int_{k-1}^k \frac{2Mx^{1-a} l(x)}{(x+M)^3} dx\end{aligned}$$

This is possible because the derivative of $x(x-1)/(x+M)^2$ is asymptotically equal to $2Mx/(x+M)^3$ for $M \to \infty$. Hence

$$\left(\frac{(k)_2}{(k+M)^2} - \frac{(k-1)_2}{(k-1+M)^2}\right) k^{-a} \sim k^{-a} \int_{k-1}^k \frac{2Mx}{(x+M)^3} dx$$

for large M, and for large values of k we get

$$\left(\frac{(k)_2}{(k+M)^2} - \frac{(k-1)_2}{(k-1+M)^2}\right) k^{-a} \sim \int_{k-1}^k \frac{2Mx^{1-a}}{(x+M)^3} dx$$

So finally with Theorem D.2:

$$\left(\frac{(k)_2}{(k+M)^2} - \frac{(k-1)_2}{(k-1+M)^2}\right) k^{-a} l(k) \sim \int_{k-1}^k \frac{2Mx^{1-a} l(x)}{(x+M)^3} dx$$

For all $L \in \mathbb{N}$ we have

$$\begin{aligned}0 &\leq \lim_{M\to\infty} M^a \sum_{k=1}^{L} \left(\frac{(k)_2}{(k+M)^2} - \frac{(k-1)_2}{(k-1+M)^2}\right) \mathbb{P}(X_1 \geq k) \\ &\leq \lim_{M\to\infty} M^a \sum_{k=1}^{L} \frac{(k)_2}{(k+M)^2} = 0\end{aligned}$$

since $a < 2$. So we obtain the two inequalities that confirm the statement that we want to prove:

$$\limsup_{M\to\infty} M^a \mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right) \leq \limsup_{M\to\infty}(1+\epsilon)2M^{1+a}\int_L^\infty \frac{x^{1-a}l(x)}{(x+M)^3}dx$$

and

$$\liminf_{M\to\infty} M^a \mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right) \geq \liminf_{M\to\infty}(1-\epsilon)2M^{1+a}\int_L^\infty \frac{x^{1-a}l(x)}{(x+M)^3}dx$$

2. With the substitution $y = M/(M+x)$ we obtain

$$\int_L^\infty \frac{x^{1-a}l(x)}{(x+M)^3}dx = \int_0^{M/(M+L)} \left(\frac{M(1-y)}{y}\right)^{1-a} l\left(\frac{M(1-y)}{y}\right)\left(\frac{y}{M}\right)^3 (My^{-2})dy$$

Hence

$$\begin{aligned}
&\lim_{M\to\infty}\frac{M^a}{l(M)}\mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right)\\
&\quad= \lim_{M\to\infty} 2M^{1+a}M^{-1-a}\int_0^{M/(M+L)} y^a(1-y)^{1-a}\frac{l(M(1-y)/y)}{l(M)}dy
\end{aligned}$$

Now $l(M(1-y)/y)/l(M)$ tends pointwise to 1. We want to exchange the limit with the integral. We have

$$y^a(1-y)^{1-a}\frac{l(M(1-y)/y)}{l(M)} = (1-y)\frac{\mathbb{P}(X_1 \geq M(1-y)/y)}{\mathbb{P}(X_1 \geq M)}$$

On $(0,1/2]$, $(1-y)/y$ is larger than 1, so $\mathbb{P}(X_1 \geq M(1-y)/y)/\mathbb{P}(X_1 \geq M) \leq 1$. On every compact subset of $[1/2,\infty)$, l is bounded away from 0 and ∞ (if $\lim_{x\to x_0} l(x) = 0$, then $\lim_{x\to x_0}\mathbb{P}(X_1 \geq x) = 0$ and therefore $\mathbb{P}(X_1 \geq x_0) = 0$). So we can apply Potter's bound (cf. Proposition D.5) on $[1/2,\infty)$ with some $\delta > 0$ such that $a+\delta < 2$:

$$y^a(1-y)^{1-a}\frac{l(M(1-y)/y)}{l(M)} \leq y^a(1-y)^{1-a}C_\delta\left(\frac{y}{1-y}\right)^\delta = C_\delta y^{(1+a+\delta)-1}(1-y)^{(2-a-\delta)-1}$$

and this is integrable on $[0,1]$ (since its integral is $C_\delta B(a+\delta+1, 2-a-\delta)$). So finally we obtain with dominated convergence

$$\begin{aligned}
\lim_{M\to\infty}\frac{M^a}{l(M)}\mathbb{E}\left(\frac{(X_1)_2}{(X_1+M)^2}\right) &= 2\int_0^1 y^a(1-y)^{1-a}dy = 2B(a+1,2-a)\\
&= 2\frac{\Gamma(a+1)\Gamma(2-a)}{\Gamma(3)} = 2\frac{a\Gamma(a)\Gamma(2-a)}{2\Gamma(2)} = aB(2-a,a)
\end{aligned}$$

□

In both cases, $a = 1$ and $2 > a > 1$, we would like to use Proposition 5.6 to show convergence to the Beta$(2-a,a)$-coalescent. Thus we need to show:

$$\lim_{N\to\infty} c_N = 0,$$

$$\lim_{N\to\infty}\frac{\mathbb{E}((\nu_{1,N})_2(\nu_{2,N})_2)}{N^2c_N}=0$$

and for all $x\in(0,1)$

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}(\nu_{1,N}>Nx)=\int_{[x,1]}y^{-2}\frac{y^{2-a-1}(1-y)^{a-1}}{B(2-a,a)}dy$$

We will show in a series of lemmas that under assumption (37) with $a\in(1,2)$ we have

$$\lim_{N\to\infty}\frac{c_N}{N\mathbb{P}(X_1\geq N)}=\lim_{N\to\infty}\frac{N^{a-1}c_N}{l(N)}=a\mu^{-a}B(2-a,a)$$

(since $a\in(1,2)$, this yields $\lim_{N\to\infty}c_N=0$ by Proposition D.1),

$$\lim_{N\to\infty}\frac{\mathbb{E}((\nu_{1,N})_2(\nu_{2,N})_2)}{N^2c_N}=0 \tag{47}$$

and for all $x\in(0,1)$

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right)=\frac{1}{B(2-a,a)}\int_x^1 y^{-1-a}(1-y)^{a-1}dy \tag{48}$$

With these lemmas it suffices to show

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}(\nu_{1,N}>Nx)=\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right)$$

Let $x\in(0,1)$ and let $\epsilon>0$, $\epsilon<x$. On $\{S_N\geq N\}$, conditionally on $X_1,\dots,X_N$, $\nu_{1,N}$ has the hypergeometric distribution with parameters (X_1,S_N,N). In Chvátal (1979) we find the following bound for the tails of the hypergeometric distribution:

Let Z be hypergeometrically distributed with parameters (N,M,n). Then for all $\epsilon>0$ we have

$$\mathbb{P}\left(Z\geq\left(\frac{M}{N}+\epsilon\right)n\right)\leq e^{-2\epsilon^2n^2}$$

With the symmetry of the hypergeometric distribution and a small calculation this also yields

$$\mathbb{P}\left(Z\leq\left(\frac{M}{N}-\epsilon\right)n\right)\leq e^{-2\epsilon^2n^2}$$

We apply these bounds and the fact that $\lim_{N\to\infty}N/c_N\mathbb{P}(S_N<N)=0$ since $c_N\geq A_1/N$ and $\mathbb{P}(S_N\leq N)\leq A^N$. Like this we obtain

$$\begin{aligned}\limsup_{N\to\infty}\frac{N}{c_N}\mathbb{P}(\nu_{1,N}>Nx)&=\limsup_{N\to\infty}\frac{N}{c_N}\mathbb{E}(\mathbb{P}(\nu_{1,N}>Nx|X_1,\dots,X_N)\mathbb{1}_{\{S_N\geq N\}}\mathbb{1}_{\{X_1/S_N\geq x-\epsilon\}})\\&\leq\limsup_{N\to\infty}\frac{N}{c_N}\mathbb{P}(\{S_N\geq N\}\cap\{X_1/S_N\geq x-\epsilon\})\\&=\limsup_{N\to\infty}\frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x-\epsilon\right)\end{aligned}$$

and

$$\liminf_{N\to\infty} \frac{N}{c_N}\mathbb{P}(\nu_{1,N} > Nx) \ge \liminf_{N\to\infty} \frac{N}{c_N}\mathbb{E}(\mathbb{P}(\nu_{1,N} > Nx|X_1,\dots,X_N)\mathbb{1}_{\{S_N\ge N\}}\mathbb{1}_{\{X_1/S_N\ge x+\epsilon\}})$$
$$= \liminf_{N\to\infty} \frac{N}{c_N}\mathbb{P}(\{S_N \ge N\}\cap\{X_1/S_N \ge x+\epsilon\})$$
$$= \liminf_{N\to\infty} \frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\ge N\}} \ge x+\epsilon\right)$$

Now we let ϵ tend to 0 and obtain the equality of the limits.

Note that the same calculation works in the case $a = 1$ if we know $c_N \to 0$ and if we know (47) and (48) for $a = 1$.

It remains to prove the mentioned lemmas.

Lemma 6.8. *Under (37) with $a \in (1, 2)$ we have*

$$\lim_{N\to\infty} \frac{c_N}{N\mathbb{P}(X_1 \ge N)} = \lim_{N\to\infty} \frac{N^{a-1}c_N}{l(N)} = \mu^{-a}aB(2-a, a)$$

Proof. We showed in Lemma 6.4 that

$$c_N \sim N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\ge N\}}\right)$$

So is suffices to show that

$$\lim_{N\to\infty} \frac{N^a}{l(N)}\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\ge N\}}\right) = \mu^{-a}aB(2-a, a)$$

We want to use Lemma 6.7. Let $\epsilon > 0$ and $\delta > 0$ such that $(1-\delta)\mu > 1$. By the law of large numbers

$$\mathbb{P}((1-\delta)N\mu \le X_2 + \cdots + X_N \le (1+\delta)N\mu) > 1-\epsilon \tag{49}$$

for large enough N. For such N we have

$$\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\ge N\}}\right) = \mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\ge N\}}\mathbb{1}_{\{X_2+\cdots+X_N<(1-\delta)N\mu\}}\right)$$
$$+\,\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{X_2+\cdots+X_N\ge(1-\delta)N\mu\}}\right)$$
$$\le \epsilon\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right) + \mathbb{E}\left(\frac{(X_1)_2}{(X_1+(1-\delta)N\mu)^2}\right)$$
$$\le 4\epsilon\mathbb{E}\left(\frac{(X_1)_2}{(X_1+N)^2}\right) + \mathbb{E}\left(\frac{(X_1)_2}{(X_1+(1-\delta)N\mu)^2}\right)$$

Because l is of slow variation, this yields

$$\limsup_{N\to\infty} \frac{N^a}{l(N)}\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\ge N\}}\right)$$
$$\le \limsup_{N\to\infty} \frac{N^a}{l(N)}4\epsilon\mathbb{E}\left(\frac{(X_1)_2}{(X_1+N)^2}\right) + ((1-\delta)\mu)^{-a}\frac{((1-\delta)N\mu)^a}{l((1-\delta)N\mu)}\mathbb{E}\left(\frac{(X_1)_2}{(X_1+(1-\delta)N\mu)^2}\right)$$
$$\overset{\text{Lemma 6.7}}{=} 4\epsilon aB(2-a, a) + ((1-\delta)\mu)^{-a}aB(2-a, a)$$

For N large enough that (49) holds, we also have

$$\begin{aligned}\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) &\geq \mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{X_2+\cdots+X_N\leq(1+\delta)N\mu\}}\right)\\ &\geq (1-\epsilon)\mathbb{E}\left(\frac{(X_1)_2}{(X_1+(1+\delta)N\mu)^2}\right)\end{aligned}$$

which implies

$$\begin{aligned}&\liminf_{N\to\infty}\frac{N^a}{l(N)}\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right)\\ &\geq (1-\epsilon)((1+\delta)\mu)^{-a}\frac{((1+\delta)N\mu)^a}{l((1+\delta)N\mu)}\mathbb{E}\left(\frac{(X_1)_2}{(X_1+(1+\delta)N\mu)^2}\right)\\ &\overset{\text{Lemma 6.7}}{=} (1-\epsilon)((1+\delta)\mu)^{-a}aB(2-a,a)\end{aligned}$$

so by letting $\epsilon,\delta\to 0$ we get the desired limit. $\square$

Lemma 6.9. *Under (37) with $a\in(1,2)$:*

$$\lim_{N\to\infty}\frac{\mathbb{E}((\nu_{1,N})_2(\nu_{2,N})_2)}{N^2c_N}=0$$

Proof. With Lemma 6.4 it suffices to show:

$$\lim_{N\to\infty}\frac{N^2}{c_N}\mathbb{E}\left(\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbb{1}_{\{S_N\geq N\}}\right)=0$$

We have

$$\mathbb{E}\left(\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbb{1}_{\{S_N\geq N\}}\right)\leq\mathbb{E}\left(\frac{(X_1)_2(X_2)_2}{\max\{X_1^2,N^2\}\max\{X_2^2,N^2\}}\right)=\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2,N^2\}}\right)^2$$

By Lemma 6.5 we know

$$\mathbb{E}\left(\frac{(X_1)^2}{\max\{X_1^2,N^2\}}\right)\leq\frac{c_N}{A_2N}$$

Since by Lemma 6.8 c_N tends to 0 when N tends to infinity, we obtain

$$\limsup_{N\to\infty}\frac{N^2}{c_N}\mathbb{E}\left(\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbb{1}_{\{S_N\geq N\}}\right)\leq\limsup_{N\to\infty}\frac{N^2}{c_N}\left(\frac{c_N}{A_2N}\right)^2=\limsup_{N\to\infty}\frac{c_N}{A_2^2}=0$$

$\square$

Lemma 6.10. *Under (37) with $a\in(1,2)$ we have for all $x\in(0,1)$:*

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right)=\frac{1}{B(2-a,a)}\int_x^1 y^{-1-a}(1-y)^{a-1}dy$$

Proof. This proof is based on Lemma 6.8. Let $x\in(0,1)$. Let $\epsilon>0$, $\delta>0$ such that $(1-\delta)\mu>1$. For N large enough we have

$$\mathbb{P}\left[(1-\delta)N\mu\leq X_2+\cdots+X_N\leq(1+\delta)N\mu\right]>1-\epsilon$$

So for such N:

$$\begin{aligned}
\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right) &= \mathbb{P}\left(\left\{\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right\}\cap\{X_2+\cdots+X_N<(1-\delta)N\mu\}\right)\\
&\quad+\mathbb{P}\left(\left\{\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right\}\cap\{X_2+\cdots+X_N\geq(1-\delta)N\mu\}\right)\\
&\leq \epsilon\mathbb{P}\left(\frac{X_1}{N}\geq x\right)+\mathbb{P}\left(\frac{X_1}{X_1+(1-\delta)N\mu}\geq x\right)
\end{aligned}$$

By taking the lim sup we get

$$\begin{aligned}
&\limsup_{N\to\infty}\frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right)\\
&\quad\leq \limsup_{N\to\infty}\frac{N}{c_N}\left(\epsilon\mathbb{P}\left(\frac{X_1}{N}\geq x\right)+\mathbb{P}\left(\frac{X_1}{X_1+(1-\delta)N\mu}\geq x\right)\right)\\
&\quad= \limsup_{N\to\infty}\frac{N}{c_N}\left(\epsilon l(Nx)(Nx)^{-a}+l\left(\frac{x}{1-x}(1-\delta)N\mu\right)\left(\frac{x}{1-x}(1-\delta)N\mu\right)^{-a}\right)\\
&\quad= \limsup_{N\to\infty}\frac{l(N)}{N^{a-1}c_N}\left(\epsilon x^{-a}+\left(\frac{x}{1-x}(1-\delta)\mu\right)^{-a}\right)\\
&\quad\overset{\text{Lemma 6.8}}{=}\frac{1}{B(2-a,a)}\left(\frac{\epsilon x^{-a}\mu^a}{a}+\frac{1}{a}\left(\frac{1-x}{x}\right)^a(1-\delta)^{-a}\right)
\end{aligned}$$

We need a similar estimate for the lim inf: For large enough N we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right)\\
&\overset{(1-\delta)\mu>1}{\geq}\mathbb{P}\left(\left\{\frac{X_1}{S_N}\geq x\right\}\cap\{(1-\delta)N\mu\leq X_2+\cdots+X_N\leq(1+\delta)N\mu\}\right)\\
&\quad\geq(1-\epsilon)\mathbb{P}\left(\frac{X_1}{X_1+(1+\delta)N\mu}\geq x\right)=(1-\epsilon)\mathbb{P}\left(X_1\geq\frac{x}{1-x}(1+\delta)N\mu\right)
\end{aligned}$$

Thus

$$\begin{aligned}
&\liminf_{N\to\infty}\frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right)\\
&\quad\geq\liminf_{N\to\infty}\frac{N}{c_N}(1-\epsilon)l\left(\frac{x}{1-x}(1+\delta)N\mu\right)\left(\frac{x}{1-x}(1+\delta)N\mu\right)^{-a}\\
&\quad=\liminf_{N\to\infty}\frac{l(N)}{N^{a-1}c_N}\mu^{-a}(1-\epsilon)\left(\frac{1-x}{x}\right)^a(1+\delta)^{-a}\\
&\quad\overset{\text{Lemma 6.8}}{=}\frac{1}{aB(2-a,a)}\left(\frac{1-x}{x}\right)^a(1-\epsilon)(1+\delta)^{-a}
\end{aligned}$$

By letting ϵ and δ tend to 0 we get

$$\lim_{N\to\infty}\frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}}\geq x\right)=\frac{1}{B(2-a,a)}\frac{1}{a}\left(\frac{1-x}{x}\right)^a$$

But with the substitution $z = (1-y)/y$, $dz = -y^{-2}dy$ we easily see that

$$\frac{1}{a}\left(\frac{1-x}{x}\right)^a = \int_x^1 y^{-1-a}(1-y)^{a-1}dy$$

□

6.5 Proof of Theorem 6.1, 4.

We will show in a series of lemmas that under the assumption (38) with $a = 1$ we have

$$\lim_{N\to\infty} c_N \log N = 1$$

(in particular this yields $\lim_{N\to\infty} c_N = 0$),

$$\lim_{N\to\infty} \frac{\mathbb{E}((\nu_{1,N})_2(\nu_{2,N})_2)}{N^2 c_N} = 0$$

and for all $x \in (0,1)$

$$\lim_{N\to\infty} \frac{N}{c_N}\mathbb{P}\left(\frac{X_1}{S_N}\mathbb{1}_{\{S_N\geq N\}} \geq x\right) = \int_x^1 y^{-2}dy$$

With these lemmas the case $a = 1$ is proven exactly as the case $1 < a < 2$.

Note that under the assumption (38) with $a = 1$, there exist $C', C'' > 0$, such that for all $k \geq 1$

$$C'k^{-1} \leq \mathbb{P}(X_1 \geq k) \leq C''k^{-1} \tag{50}$$

Lemma 6.11. *Under (38) with $a = 1$ we have*

$$\lim_{N\to\infty} c_N \log N = 1$$

Proof. With Lemma 6.4 it suffices to show

$$\lim_{N\to\infty} \log N\left(N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right)\right) = 1.$$

1. Let $B > 0$. We define $Y_i := \mathbb{1}_{\{X_i\leq BN\}}X_i$. We will show that

$$\lim_{N\to\infty}\frac{\mathbb{E}(Y_1)}{\log N} = C$$

where C is the constant from $\mathbb{P}(X_1 \geq k) \sim Ck^{-1}$. Let $1 \leq L \leq BN$. Then

$$\begin{aligned}
\mathbb{E}(Y_1) &= \int_0^\infty \mathbb{P}(Y_1 \geq x)dx = \int_0^{BN} \mathbb{P}(BN \geq X_1 \geq x)dx \\
&= \int_0^{BN} (\mathbb{P}(X_1 \geq x) - \mathbb{P}(X_1 > BN))dx \\
&= \int_0^L \mathbb{P}(X_1 \geq x)dx + \int_L^{BN} \mathbb{P}(X_1 \geq x)dx - BN\mathbb{P}(X_1 > BN)
\end{aligned}$$

Since (50) implies $\mathbb{P}(X_1 > BN) \leq C''(BN)^{-1}$, we obtain

$$\lim_{N\to\infty} \frac{1}{\log N} \left(\int_0^L \mathbb{P}(X_1 \geq x)dx - BN\mathbb{P}(X_1 > BN) \right) = 0$$

Let $\eta > 0$. If L is large enough, then we have for all $k \geq L$:

$$C(1-\eta)\frac{1}{k} \leq \mathbb{P}(X_1 \geq x) \leq C(1+\eta)\frac{1}{x}$$

Since the logarithm is a function of slow variation, we have

$$\lim_{N\to\infty} \frac{1}{\log N} C \int_L^{BN} \frac{1}{x} dx = \lim_{N\to\infty} \frac{C}{\log N}(\log(BN) - \log L) = C$$

By letting $\eta \to 0$ we get

$$\lim_{N\to\infty} \frac{\mathbb{E}(Y_1)}{\log N} = C$$

2. We will need a number of auxiliary inequalities: With Lemma 6.2 and (50) we see that

$$\operatorname{var} Y_1 \leq \mathbb{E}(Y_1^2) = \sum_{k=1}^{\infty} (k^2 - (k-1)^2)\mathbb{P}(Y_1 \geq k) \leq \sum_{k=1}^{\lfloor BN \rfloor} 2k\mathbb{P}(X_1 \geq k) \leq 2C''BN \tag{51}$$

We will also need the following inequality

$$\begin{aligned} \lim_{N\to\infty} \mathbb{P}\left(\max_{1\leq i\leq N} X_i > BN \right) &= \lim_{N\to\infty} 1 - (1 - \mathbb{P}(X_1 > BN))^N \\ &\overset{(38)}{=} \lim_{N\to\infty} 1 - \left(1 - \frac{C}{BN}\right)^N = 1 - \exp(-C/B) \end{aligned} \tag{52}$$

Let $\epsilon > 0$ and $0 < \delta < 1/2$. We choose B large enough that the following condition (2a) is satisfied and then we choose N large enough that the other conditions are satisfied:

(a) $1 - e^{-C/B} < \epsilon/4$

(b) $C(1-\delta)\log N > 1$

(c) $\left|1 - \mathbb{E}\left(\frac{Y_2+\cdots+Y_N}{CN\log N}\right)\right| < \frac{\delta}{2}$

(d) $|\mathbb{P}(\max_{1\leq i\leq N} X_i > BN) - (1 - e^{-C/B})| < \epsilon/4$ and finally

(e) $(8C''B)/(C^2\delta^2(\log N)^2) < \epsilon/2$

3. We evaluate the probabilities of two events that we will need in the proof: With (2c) we obtain

$$\begin{aligned} \mathbb{P}\left(\left| \frac{X_2+\cdots+X_N}{CN\log N} - 1 \right| \geq \delta \right) \leq & \mathbb{P}\left(\left| \frac{Y_2+\cdots+Y_N}{CN\log N} - \mathbb{E}\left(\frac{Y_2+\cdots+Y_N}{CN\log N}\right) \right| \geq \frac{\delta}{2} \right) \\ & + \mathbb{P}\left(\max_{1\leq i\leq N} X_i > BN \right) \end{aligned}$$

We use Chebyshev's inequality, (2a), and (2d) to see that this is

$$\leq \operatorname{var}\left(\frac{Y_2+\cdots+Y_N}{CN\log N}\right)\left(\frac{\delta}{2}\right)^{-2}+\frac{\epsilon}{2}$$

with (51) and then (2e) we finally obtain

$$\mathbb{P}\left(\left|\frac{X_2+\cdots+X_N}{CN\log N}-1\right|\geq\delta\right)\leq\frac{8C''BN(N-1)}{\delta^2C^2N^2(\log N)^2}+\frac{\epsilon}{2}\leq\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon \tag{53}$$

The other event is $\{X_2+\cdots+X_N\leq\frac{C}{2}N\log N\}$:

$$\mathbb{P}\left(X_2+\cdots+X_N\leq\frac{C}{2}N\log N\right)\leq\mathbb{P}\left(\frac{Y_2+\cdots+Y_N}{CN\log N}\leq\frac{1}{2}\right)$$

with (2c) and since $\delta<1/2$:

$$\leq\mathbb{P}\left(\left|\frac{Y_2+\cdots+Y_N}{CN\log N}-\mathbb{E}\left(\frac{Y_2+\cdots+Y_N}{CN\log N}\right)\right|\geq\frac{1}{4}\right)$$

with Chebyshev's inequality this is

$$\leq 16\operatorname{var}(Y_1)\frac{N-1}{C^2N^2(\log N)^2}$$

and with (51) we get

$$\mathbb{P}\left(X_2+\cdots+X_N\leq\frac{C}{2}N\log N\right)\leq 32C''BN\frac{N-1}{C^2N^2(\log N)^2}\leq\frac{32C''B}{C^2(\log N)^2} \tag{54}$$

4. After this technical preparation we are now able to calculate the limit

$$\lim_{N\to\infty}\log N\left(N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right)\right)$$

We define the events

$$D_1:=\left\{X_2+\cdots+X_N\leq\frac{C}{2}N\log N\right\}$$

(we showed in (54) that $\mathbb{P}(D_1)\leq\frac{32C''B}{C^2(\log N)^2}$ for B and N large enough)

$$D_2:=\left\{\frac{C}{2}N\log N<X_2+\cdots+X_N\leq C(1-\delta)N\log N\right\}$$

($\mathbb{P}(D_2)\leq\mathbb{P}\left(\left|\frac{X_2+\cdots+X_N}{CN\log N}-1\right|\geq\delta\right)\overset{(53)}{\leq}\epsilon$ for large enough N)

$$D_3:=\{X_2+\cdots+X_N>C(1-\delta)N\log N\}$$

So

$$\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) \leq \mathbb{P}(D_1)\mathbb{E}\left(\frac{(X_1)_2}{\max\{X_1^2, N^2\}}\right) + \mathbb{P}(D_2)\mathbb{E}\left(\frac{(X_1)_2}{(X_1+\frac{C}{2}N\log N)^2}\right)$$
$$+\mathbb{E}\left(\frac{(X_1)_2}{(X_1+C(1-\delta)N\log N)^2}\right)$$
$$\leq \frac{128C''B}{C^2(\log N)^2}\mathbb{E}\left(\frac{(X_1)_2}{(X_1+N)^2}\right) + \epsilon\mathbb{E}\left(\frac{(X_1)_2}{(X_1+\frac{C}{2}N\log N)^2}\right)$$
$$+\mathbb{E}\left(\frac{(X_1)_2}{(X_1+C(1-\delta)N\log N)^2}\right)$$

We use Lemma 6.7 with $M = N$, $M = \frac{C}{2}N\log N$ and $M = C(1-\delta)N\log N$ to obtain

$$\limsup_{N\to\infty} N\log N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) \leq \limsup_{N\to\infty}\frac{128C''B}{C^2\log N}C + \epsilon\frac{2}{C}C + \frac{1}{C(1-\delta)}C$$
$$= 2\epsilon + \frac{1}{1-\delta}$$

To evaluate the $\liminf$ we introduce

$$D_4 := \{C(1-\delta)N\log N \leq X_2+\cdots+X_N \leq C(1+\delta)N\log N\}$$

We calculated in (53) that $\mathbb{P}(D_4^c)\leq\epsilon$ for large enough N. With (2b) we get

$$\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) \geq \mathbb{P}(D_4)\mathbb{E}\left(\frac{(X_1)_2}{(X_1+C(1+\delta)N\log N)^2}\right)$$

We use Lemma 6.7 with $M = C(1+\delta)N\log N$ to obtain

$$\liminf_{N\to\infty} N\log N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) \geq (1-\epsilon)\frac{1}{C(1+\delta)}C = \frac{1-\epsilon}{1+\delta}$$

The proof is completed by letting $\epsilon,\delta\to 0$. □

Lemma 6.12. *Under (38) with $a=1$ we have*

$$\lim_{N\to\infty}\frac{\mathbb{E}((\nu_{1,N})_2(\nu_{2,N})_2)}{N^2c_N} = 0$$

Sketch of the proof. With Lemma 40 it suffices to show

$$\lim_{N\to\infty}\frac{N^2}{c_N}\mathbb{E}\left(\frac{(X_1)_2(X_2)_2}{S_N^4}\mathbb{1}_{\{S_N\geq N\}}\right) = 0$$

We distinguish the events

$$D := \{X_3+\cdots+X_N \leq \frac{C}{2}N\log N\}$$

and D^c. We showed in the proof of 6.11, (54) that there is a $K>0$ such that $\mathbb{P}(D)\leq\frac{K}{(\log N)^2}$ for N large enough. We use Lemma 6.7 and Lemma 6.11, the rest is a more or less elementary calculation. □

Lemma 6.13. *Under (38) with $a = 1$, we have for all $x \in (0, 1)$*

$$\lim_{N\to\infty} \frac{N}{c_N} \mathbb{P}\left(\frac{X_1}{S_N} \mathbb{1}_{\{S_N \geq N\}} \geq x\right) = \int_x^1 y^{-2} dy$$

Proof. The proof is similar to the last part of the proof of Lemma 6.11. Let $\epsilon > 0$ and $0 < \delta < 1/2$. Let D_1, D_2, D_3, D_4 be as in the proof of Lemma 6.11. For large enough N we have $P(D_1) \leq K/(\log N)^2$ for some $K > 0$ and $\mathbb{P}(D_2) < \epsilon$, exactly as in the proof of Lemma 6.11. So

$$\mathbb{P}\left(\frac{X_1}{S_N} \mathbb{1}_{\{S_N \geq N\}} \geq x\right) \leq \frac{K}{(\log N)^2} \mathbb{P}\left(\frac{X_1}{N} > x\right) + \epsilon \mathbb{P}\left(\frac{X_1}{X_1 + \frac{C}{2} N \log N} \geq x\right)$$
$$+ \mathbb{P}\left(\frac{X_1}{X_1 + C(1-\delta) N \log N} \geq x\right)$$

With Lemma 6.11 and (38) we get

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{N}{c_N} \mathbb{P}\left(\frac{X_1}{S_N} \mathbb{1}_{\{S_N \geq N\}} \geq x\right) \\
&\quad \leq \limsup_{N\to\infty} N \log N \\
&\quad \left(\frac{K}{(\log N)^2} C (Nx)^{-1} + \epsilon C \left(\frac{xCN \log N}{2(1-x)}\right)^{-1} + C \left(\frac{xC(1-\delta) N \log N}{1-x}\right)^{-1}\right) \\
&\quad = \epsilon \frac{2(1-x)}{x} + \frac{1}{1-\delta} \frac{1-x}{x}
\end{aligned}$$

For large enough N we have $\mathbb{P}(D_4) \geq (1 - \epsilon)$. So

$$\begin{aligned}
\liminf_{N\to\infty} \frac{N}{c_N} \mathbb{P}\left(\frac{X_1}{S_N} \mathbb{1}_{\{S_N \geq N\}} \geq x\right) &\geq \liminf_{N\to\infty} N \log N (1-\epsilon) \mathbb{P}\left(\frac{X_1}{X_1 + C(1+\delta) N \log N} \geq x\right) \\
&\overset{(38)}{=} \liminf_{N\to\infty} N \log N (1-\epsilon) C \left(\frac{xC(1+\delta) N \log N}{1-x}\right)^{-1} = \frac{1-\epsilon}{1+\delta} \frac{1-x}{x}
\end{aligned}$$

By letting $\epsilon, \delta \to 0$, we obtain

$$\lim_{N\to\infty} \frac{N}{c_N} \mathbb{P}\left(\frac{X_1}{S_N} \mathbb{1}_{\{S_N \geq N\}} \geq x\right) = \frac{1-x}{x} = \int_x^1 y^{-2} dy$$

□

6.6 Proof of Theorem 6.1, 5.

Let $(\xi(t) : t \in [0, 1])$ be a stable subordinator of index a with Lévy measure

$$\Lambda_a(dx) = a x^{-a-1} dx$$

Let g be an asymptotic inverse of $f(x) := \mathbb{P}(X_1 \geq x)$. This is a positive function which diverges to ∞ when N tends to infinity and which satisfies $f(g(x)) \sim 1/x$ for $x \to \infty$ (cf. Proposition D.4). Let $Z_1 \geq Z_2 \geq \dots$ be the ordered jumps of ξ. For all N, let $Y_{1,N} \geq \dots \geq Y_{N,N}$ be the decreasing sequence of the values of $X_1/g(N), \dots, X_N/g(N)$.

Lemma 6.14. *For all $j \in \mathbb{N}$: $(Y_{1,N}, \dots, Y_{j,N})$ converges in distribution on $\mathbb{R}^j$ to $(Z_1, \dots, Z_j)$ when $N \to \infty$.*

Proof. Let $x_1 \geq \dots \geq x_j > 0$ be given. We define $x_0 := \infty$ and by convention $x_0^{-a} := 0$. We define

$$L_i^N := \#\{k : Y_{k,N} \in [x_i, x_{i-1})\} \text{ and}$$
$$K_i := \#\{k : Z_k \in [x_i, x_{i-1})\}$$

K_i corresponds to the number of atoms on $[0,1] \otimes [x_i, x_{i-1})$ of a Poisson random measure with intensity $\lambda \otimes \Lambda_a$ where λ is the Lebesgue measure on $[0,1]$. Therefore K_i has a Poisson distribution with parameter $\Lambda_a([x_i, x_{i-1})) = (x_i^{-a} - x_{i-1}^{-a})$. Also, all the K_i are independent.

$(L_1^N, \dots, L_j^N, N - L_1^N - \dots - L_j^N)$ has a multinomial distribution with parameters $(N; p_{1,N}, \dots, p_{j,N}, p_N)$ where $p_{i,N} := \mathbb{P}(X_1/g(N) \in [x_i, x_{i-1}))$ and $p_N := 1 - p_{1,N} - \dots - p_{j,N}$. We have

$$\begin{aligned} p_{i,N} &= \mathbb{P}(X_1/g(N) \geq x_i) - \mathbb{P}(X_1/g(N) \geq x_{i-1}) = f(g(N)x_i) - f(g(N)x_{i-1}) \\ &\sim f(g(N))x_i^{-a} - f(g(N))x_{i-1}^{-a} \sim N^{-1}(x_i^{-a} - x_{i-1}^{-a}) \text{ and} \\ p_N^{N-n_1-\dots-n_j} &= (1 - \mathbb{P}(X_1/g(N) \geq x_j))^{N-n_1-\dots-n_j} \sim (1 - N^{-1}x_j^{-a})^N \sim e^{-x_j^{-a}} \end{aligned}$$

So for all $(n_1, \dots, n_j) \in \mathbb{N}^j$:

$$\begin{aligned} \mathbb{P}(L_1^N = n_1, \dots, L_j^N = n_j) &= \frac{N!}{n_1! \dots n_j!(N - \sum_{i=1}^j n_i)!} p_{1,N}^{n_1} \dots p_{j,N}^{n_j} p_N^{N-n_1-\dots-n_j} \\ &\sim \frac{N^{n_1+\dots+n_j}}{n_1! \dots n_j!} \left(\prod_{i=1}^j N^{-ni}(x_i^{-a} - x_{i-1}^{-a})^{n_i} \right) e^{-x_j^{-a}} \\ &= \prod_{i=1}^j \frac{e^{-(x_i^{-a} - x_{i-1}^{-a})}((x_i^{-a} - x_{i-1}^{-a}))^{n_i}}{n_i!} = \prod_{i=1}^j \mathbb{P}(K_i = n_i) \\ &= \mathbb{P}(K_1 = n_1, \dots, K_j = n_j) \end{aligned}$$

We have $Y_{i,N} \geq x_i$ if and only if $L_1^N + \dots + L_i^N \geq i$ and $Z_i \geq x_i$ if and only if $K_1 + \dots + K_i \geq i$. So

$$\begin{aligned} \lim_{N\to\infty} (Y_{1,N} \geq x_1, \dots, Y_{j,N} \geq x_j) &= \lim_{N\to\infty} \mathbb{P}(L_1^N + \dots + L_i^N \geq i,\, 1 \leq i \leq j) \\ &= \mathbb{P}(K_1 + \dots + K_i \geq i,\, 1 \leq i \leq j) \\ &= \mathbb{P}(Z_1 \geq x_1, \dots, Z_j \geq x_j) \end{aligned}$$

With the remarks from the section "Euclidean Space" of Chapter 1, 3. of Billingsley (1968) (p. 17) we get the convergence in distribution. □

Lemma 6.15. *For all $j \in \mathbb{N}$: When $N \to \infty$, $(Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^N Y_{i,N})$ converges in distribution on $\mathbb{R}^{j+1}$ to $(Z_1, \dots, Z_j, \sum_{i=j+1}^\infty Z_j)$.*

Proof. Let d be the Prohorov distance on the space of probabilities on $\mathbb{R}^{j+1}$. d is defined as

$$d(P,Q) := \inf\{r > 0 : P(A) \leq Q(A^r) + r \text{ and } Q(A) \leq P(A^r) + r \text{ for all } A \in \mathcal{B}(\mathbb{R}^{j+1})\}$$

where $A^r := \{x \in \mathbb{R}^{j+1} : |y - x| < r \text{ for some } y \in A\}$. Convergence in distribution on $\mathbb{R}^{j+1}$ is equivalent to convergence of the distributions in the Prokhorov distance (cf. Theorem 3.3.1 of Ethier and Kurtz (1986).)

Let for $M \leq N$ $Q_{M,N}$ be the distribution of $(Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^{M} Y_{i,N})$ and let Q_N be the distribution of $(Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^{N} Y_{i,N})$. Let $\epsilon > 0$. We choose a $B(\epsilon) > \delta > 0$ where $B(\epsilon)$ is a certain bound depending on ϵ, which will be found later. For large enough M we have $\mathbb{P}(Z_M \geq \delta) < \epsilon/4$. We showed in the preceding lemma that $Y_{M,N}$ converges in distribution to Z_M. Portmanteau's theorem (cf. Theorem 3.3.1 of Ethier and Kurtz (1986)) yields

$$\limsup_{N\to\infty} \mathbb{P}(Y_{M,N} \in [\delta, \infty)) \leq \mathbb{P}(Z_M \geq \delta)$$

Therefore we have for large enough M and N $\mathbb{P}(Y_{M,N} \geq \delta) \leq \epsilon/2$. Hence

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{N} Y_{i,N} \mathbb{1}_{\{Y_{i,N} \leq \delta\}}\right) &= \frac{N}{g(N)} \mathbb{E}\left(X_1 \mathbb{1}_{\{X_1 \leq g(N)\delta\}}\right) \\
&= \frac{N}{g(N)} \int_0^\infty \mathbb{P}\left(X_1 \mathbb{1}_{\{X_1 \leq g(N)\delta\}} \geq x\right) dx \\
&\leq \frac{N}{g(N)} \int_0^{g(N)\delta} x^{-a} l(x) dx
\end{aligned}$$

By Karamata's theorem (Theorem D.3),

$$\begin{aligned}
\int_0^{g(N)\delta} x^{-a} l(x) dx &\sim \frac{(g(N)\delta)^{1-a} l(g(N)\delta)}{1-a} = \frac{g(N)\delta f(g(N)\delta)}{1-a} \\
&\sim \frac{g(N)\delta f(g(N))\delta^{-a}}{1-a} \sim \frac{g(N)}{N} \frac{\delta^{1-a}}{1-a}
\end{aligned}$$

So for large enough N

$$\mathbb{E}\left(\sum_{i=1}^{N} Y_{i,N} \mathbb{1}_{\{Y_{i,N} \leq \delta\}}\right) \leq \frac{N}{g(N)} (1+\epsilon) \frac{g(N)}{N} \frac{\delta^{1-a}}{1-a} = (1+\epsilon) \frac{\delta^{1-a}}{1-a}$$

With Markov's inequality we obtain

$$\begin{aligned}
\mathbb{P}\left(\sum_{i=M+1}^{N} Y_{i,N} \geq \epsilon\right) &\leq \mathbb{P}(Y_{M,N} \geq \delta) + \mathbb{P}\left(\sum_{i=1}^{N} Y_{i,N} \mathbb{1}_{\{Y_{i,N} \leq \delta\}} \geq \epsilon\right) \\
&\leq \frac{\epsilon}{2} + \frac{1+\epsilon}{\epsilon} \frac{\delta^{1-a}}{1-a}
\end{aligned}$$

for large enough M and N. For the right $B(\epsilon)$ and for $\delta < B(\epsilon)$ we therefore obtain

$$\mathbb{P}\left(\sum_{i=M+1}^{N} Y_{i,N} \geq \epsilon\right) \leq \epsilon$$

Hence for all $A \in \mathcal{B}(\mathbb{R}^{j+1})$:

$$Q_{M,N}(A) = \mathbb{P}\left((Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^{M} Y_{i,N}) \in A\right)$$
$$\leq \mathbb{P}\left((Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^{N} Y_{i,N}) \in A^{\epsilon}\right) + \mathbb{P}\left(\sum_{i=M+1}^{N} Y_i > \epsilon\right) \leq Q_N(A^{\epsilon}) + \epsilon$$

and analogously $Q_N(A) \leq Q_{M,N}(A^{\epsilon}) + \epsilon$, hence $d(Q_N, Q_{M,N}) \leq \epsilon$.

Let P be the distribution of $(Z_1, \dots, Z_j, \sum_{i=j+1}^{\infty} Z_i)$ and let P_M be the distribution of $(Z_1, \dots, Z_j, \sum_{i=j+1}^{M} Z_i)$.

Since $\sum_{i=j+1}^{M} Z_i$ converges a.s. (and therefore also in distribution) to $\sum_{i=j+1}^{\infty} Z_i$ when $M \to \infty$, we have for large enough M $d(P_M, P) < \epsilon$.

For all M, $(Y_{1,N}, \dots, Y_{M,N})$ converges in distribution to $(Z_1, \dots, Z_M)$ according to the last lemma. Hence

$$(Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^{M} Y_{i,N}) \text{ converges in distribution to } (Z_1, \dots, Z_j, \sum_{i=j+1}^{M} Z_i)$$

and for N (depending on M) large enough we have $d(Q_{M,N}, P_M) < \epsilon$.

We thus choose M large enough such that $\mathbb{P}(Z_M \geq \delta) < \epsilon/4$ and such that $d(P_M, P) < \epsilon$. Then we choose N_0 large enough such that every $N \geq N_0$ satisfies all the other conditions that we needed. So for all $N \geq N_0$:

$$d(Q_N, P) \leq d(Q_N, Q_{M,N}) + d(Q_{M,N}, P_M) + d(P_M, P) \leq \epsilon + \epsilon + \epsilon = 3\epsilon$$

□

We define $W_i := Z_i / \sum_{j=1}^{\infty} Z_j$ for all $i \geq 1$. So $(W_1, W_2, \dots)$ has the PD$(a, 0)$ distribution.

Lemma 6.16. *When $N \to \infty$,*

$$\left(\frac{Y_{1,N}}{\sum_{i=1}^{N} Y_{i,N}}, \dots, \frac{Y_{N,N}}{\sum_{i=1}^{N} Y_{i,N}}, 0, \dots\right) \textit{ converges in distribution on } \Delta \textit{ to } (W_1, W_2, \dots)$$

Proof. We just showed that for all j,

$$(Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^{N} Y_{i,N}) \text{ converges in distribution to } (Z_1, \dots, Z_j, \sum_{i=j+1}^{\infty} Z_j)$$

We define

$$h : \mathbb{R}^{j+1} \to \mathbb{R}^j, (x_1, \dots, x_{j+1}) \mapsto \left(\frac{x_1}{x_1 + \dots, x_{j+1}}, \dots, \frac{x_j}{x_1 + \dots, x_{j+1}}\right)$$

h is continuous on $\mathbb{R}^{j+1} \backslash \{0\}$. But $\mathbb{P}(Z_1 + \dots + Z_j + \sum_{i=j+1}^{\infty} Z_i = 0) = 0$. Therefore the continuous mapping theorem (Corollary 3.1.9 in Ethier and Kurtz (1986)) implies that

$$h((Y_{1,N}, \dots, Y_{j,N}, \sum_{i=j+1}^{N} Y_{i,N})) \text{ converges in distribution to } h((Z_1, \dots, Z_j, \sum_{i=j+1}^{\infty} Z_j))$$

Hence we have the convergence in distribution of

$$\left(\frac{Y_{1,N}}{\sum_{i=1}^N Y_{i,N}}, \ldots, \frac{Y_{j,N}}{\sum_{i=1}^N Y_{i,N}} \right) \text{ to } (W_1, \ldots, W_j)$$

for all j. But the functions on Δ that only depend on a finite number of coordinates are convergence determining (cf. Theorem 3.4.5 in Ethier and Kurtz (1986), take e.g. the coordinate projections as strongly separating subset). Therefore we have the convergence in distribution on Δ. □

Lemma 6.17. *For all $r \in \mathbb{N}$ and $k_1, \ldots, k_r \geq 2$ we have*

$$\lim_{N\to\infty} N^r \mathbb{E} \left(\frac{(X_1)_{k_1} \ldots (X_r)_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \right) = \sum_{\substack{i_1,\ldots,i_r=1 \\ \text{all distinct}}}^{\infty} \mathbb{E}(W_{i_1}^{k_1} \ldots W_{i_r}^{k_r})$$

Proof. We have

$$\begin{aligned} N^r \mathbb{E} \left(\frac{(X_1)_{k_1} \ldots (X_r)_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \right) &\sim (N)_r \mathbb{E} \left(\frac{(X_1)_{k_1} \ldots (X_r)_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \right) \\ &= \sum_{\substack{i_1,\ldots,i_r=1 \\ \text{all distinct}}}^{N} \mathbb{E} \left(\frac{(X_{i_1})_{k_1} \ldots (X_{i_r})_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \right) \end{aligned}$$

and

$$\begin{aligned} &\lim_{N\to\infty} \sum_{\substack{i_1,\ldots,i_r=1 \\ \text{all distinct}}}^{N} \mathbb{E} \left(\frac{(X_{i_1})_{k_1} \ldots (X_{i_r})_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \right) \\ &= \lim_{N\to\infty} \sum_{\substack{i_1,\ldots,i_r=1 \\ \text{all distinct}}}^{N} \mathbb{E} \left(\frac{X_{i_1}^{k_1} \ldots X_{i_r}^{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \right) \end{aligned}$$

since

$$\begin{aligned} &\sum_{\substack{i_1,\ldots,i_r=1 \\ \text{all distinct}}}^{N} \mathbb{E} \left(\frac{(X_{i_1})_{k_1} \ldots (X_{i_r})_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \mathbb{1}_{\{X_{i_j} \geq N^{1/4}, 1 \leq j \leq r\}} \right) \\ &\sim \sum_{\substack{i_1,\ldots,i_r=1 \\ \text{all distinct}}}^{N} \mathbb{E} \left(\frac{X_{i_1}^{k_1} \ldots X_{i_r}^{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \mathbb{1}_{\{X_{i_j} \geq N^{1/4}, 1 \leq j \leq r\}} \right) \end{aligned}$$

and since

$$\begin{aligned} &\sum_{\substack{i_1,\ldots,i_r=1 \\ \text{all distinct}}}^{N} \mathbb{E} \left(\frac{(X_{i_1})_{k_1} \ldots (X_{i_r})_{k_r}}{S_N^{k_1+\cdots+k_r}} \mathbb{1}_{\{S_N \geq N\}} \mathbb{1}_{\{X_{i_1} < N^{1/4}\}} \right) \\ &\leq \sum_{i_1=1}^{N} \left(\frac{N^{1/4}}{N} \right)^{k_1} \sum_{i_2,\ldots,i_r=1}^{N} \mathbb{E} \left(\left(\frac{X_{i_2}}{S_N} \right)^{k_2} \ldots \left(\frac{X_{i_r}}{S_N} \right)^{k_r} \right) \\ &\leq N \left(\frac{N^{1/4}}{N} \right)^2 \sum_{i_2,\ldots,i_r=1}^{N} \mathbb{E} \left(\left(\frac{X_{i_2}}{S_N} \right) \ldots \left(\frac{X_{i_r}}{S_N} \right) \right) \leq \frac{1}{\sqrt{N}} \end{aligned}$$

and this inequality remains valid if we replace $\mathbb{1}_{\{X_{i_1}<N^{1/4}\}}$ by $\mathbb{1}_{\{X_{i_j}<N^{1/4}\}}$.

Further

$$\lim_{N\to\infty} \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{N} \mathbb{E}\left(\frac{X_{i_1}^{k_1}\dots X_{i_r}^{k_r}}{S_N^{k_1+\dots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right) = \lim_{N\to\infty} \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{N} \mathbb{E}\left(\frac{X_{i_1}^{k_1}\dots X_{i_r}^{k_r}}{S_N^{k_1+\dots+k_r}}\right)$$

since by Lemma 6.3 $\mathbb{P}(S_N < N) \leq A^N$ and hence

$$\lim_{N\to\infty} \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{N} \mathbb{E}\left(\frac{X_{i_1}^{k_1}\dots X_{i_r}^{k_r}}{S_N^{k_1+\dots+k_r}}\mathbb{1}_{\{S_N<N\}}\right) \leq \lim_{N\to\infty} N^r A^N = 0$$

So

$$\lim_{N\to\infty} N^r\mathbb{E}\left(\frac{(X_1)_{k_1}\dots(X_r)_{k_r}}{S_N^{k_1+\dots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right) = \lim_{N\to\infty} \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{N} \mathbb{E}\left(\frac{X_{i_1}^{k_1}\dots X_{i_r}^{k_r}}{S_N^{k_1+\dots+k_r}}\right)$$

$$= \lim_{N\to\infty} \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{N} \mathbb{E}\left(\left(\frac{Y_{i_1,N}}{Y_{1,N}+Y_{N,N}}\right)^{k_1}\dots\left(\frac{Y_{i_r,N}}{Y_{1,N}+Y_{N,N}}\right)^{k_r}\right)$$

We introduce the function

$$f:\Delta\to\mathbb{R},\ (x_1,x_2,\dots)\mapsto \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{\infty} x_{i_1}^{k_1}\dots x_{i_r}^{k_r}$$

By Lemma 6.16 it suffices to show that f is continuous and bounded to obtain the desired convergence. Of course every continuous function on Δ is bounded, since Δ is compact. And the continuity we already showed in the proof of Lemma 3.4. $\square$

Now we can combine these lemmas to complete the proof of Theorem 6.1, 4.

We define the measures Θ_a and Ξ_a as in Theorem 6.1: Θ_a is the probability on Δ that corresponds to the PD$(a,0)$ distribution and $\Xi_a(dx) := \sum_{j=1}^{\infty} x_j^2\Theta_a(dx)$. To obtain the convergence of $(\Pi_{n,N}(m) : m \in \mathbb{N}_0)$ to a discrete time Ξ_a-coalescent with values in $\mathcal{P}_n$ we use Theorem 5.1. Thus we need to show:

1. $\lim_{N\to\infty} c_N = c > 0$
2. For all $r\in\mathbb{N}$, $k_1,\dots,k_r\geq 2$:

$$\lim_{N\to\infty}\frac{\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r})}{N^{k_1+\dots+k_r-r}} = \int_\Delta \sum_{i_1\neq\dots\neq i_r} x_{i_1}^{k_1}\dots x_{i_r}^{k_r}\Big/\sum_{j=1}^{\infty}x_j^2\Xi_a(dx)$$

Under assumption 1 we already proved condition 2: In this case we obtain from Lemma 6.4

$$\lim_{N\to\infty}\frac{\mathbb{E}((\nu_{1,N})_{k_1}\dots(\nu_{r,N})_{k_r})}{N^{k_1+\dots+k_r-r}} = \lim_{N\to\infty} N^r\mathbb{E}\left(\frac{(X_1)_{k_1}\dots(X_r)_{k_r}}{S_N^{k_1+\dots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right)$$

with Lemma 6.17 this is

$$\lim_{N\to\infty} N^r\mathbb{E}\left(\frac{(X_1)_{k_1}\dots(X_r)_{k_r}}{S_N^{k_1+\dots+k_r}}\mathbb{1}_{\{S_N\geq N\}}\right) = \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{\infty} \mathbb{E}(W_{i_1}^{k_1}\dots W_{i_r}^{k_r})$$
$$= \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{\infty} \int_\Delta x_{i_1}^{k_1}\dots x_{i_r}^{k_r}\Theta_a(dx) = \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{\infty} \int_\Delta x_{i_1}^{k_1}\dots x_{i_r}^{k_r}\Big/\sum_{j=1}^{\infty} x_j^2\;\Xi_a(dx)$$

To prove the condition 1 and to calculate the transition probabilities, we will need a result of Pitman. This is Proposition 9 in Pitman (1995): Let Π be a random partition in $\mathcal{P}_\infty$ whose distribution is given by the paint box mixture corresponding to Θ_a. Let $k_1,\dots,k_r\geq 2$ such that $k_1+\dots+k_r=n$ and let $\pi\in\mathcal{P}_n$ be the unique partition with blocks $\{1,\dots,k_1\},\{k_1+1,\dots,k_1+k_2\},\dots,\{k_1+\dots+k_{r-1}+1,\dots,k_1+\dots+k_r\}$. Then

$$\mathbb{P}(R_n\Pi=\pi) = \frac{a^{r-1}(r-1)!}{(n-1)!}\prod_{i=1}^{r}[1-a]_{k_i-1}$$

where $[x]_0:=1$ and $[x]_k:=x(x+1)\dots(x+k-1)$ for $k\geq 1$. On the other side we obtain from the paint box construction:

$$\mathbb{P}(R_n\Pi=\pi) = \sum_{\substack{i_1,\dots,i_r=1\\ \text{all distinct}}}^{\infty} \mathbb{E}(W_{i_1}^{k_1}\dots W_{i_r}^{k_r})$$

Now it is easy to see that $\lim_{N\to\infty} c_N>0$: We use Lemma 6.4 and Lemma 6.17 to obtain

$$\lim_{N\to\infty} c_N = \lim_{N\to\infty} N\mathbb{E}\left(\frac{(X_1)_2}{S_N^2}\mathbb{1}_{\{S_N\geq N\}}\right) = \sum_{i=1}^{\infty}\mathbb{E}(W_i^2) = \mathbb{P}(R_2\Pi=\{1,2\}) = 1-a>0$$

Thus $(\Pi_{n,N}(m): m\in\mathbb{N}_0)$ converges to a discrete time Ξ_a-coalescent with values in $\mathcal{P}_n$. The transition probabilities are given by

$$p_{b;k_1,\dots,k_r;0} = \int_\Delta \sum_{i_1\neq\dots\neq i_r} x_{i_1}^{k_1}\dots x_{i_r}^{k_r}\Big/\sum_{j=1}^{\infty}x_j^2\;\Xi_a(dx) = \int_\Delta \sum_{i_1\neq\dots\neq i_r} x_{i_1}^{k_1}\dots x_{i_r}^{k_r}\Theta_a(dx)$$
$$= \sum_{i_1\neq\dots\neq i_r}\mathbb{E}(W_{i_1}^{k_1}\dots W_{i_r}^{k_r}) = \frac{a^{r-1}(r-1)!}{(b-1)!}\prod_{i=1}^{r}[1-a]_{k_i-1}$$

To calculate the transition probabilities for $s>0$ we will need the **exchangeable probability function** of Π. This is a function on the space of finite sequences of positive integers. For $k_1,\dots,k_r\geq 1$ let π be a partition of $k_1+\dots+k_r$ with blocks of respective sizes $k_1,\dots,k_r$. Then

$$p(k_1,\dots,k_r) = \mathbb{P}(R_{k_1+\dots+k_r}\Pi=\pi) = \frac{a^{r-1}(r-1)!}{(k_1+\dots+k_r-1)!}\prod_{i=1}^{r}[1-a]_{k_i-1}$$

In Proposition 10 of Pitman (1995) it is shown that

$$p(k_1,\dots,k_r) = \sum_{j=1}^{r} p(k_1,\dots,k_{j-1},k_j+1,k_{j+1},\dots,k_r) + p(k_1,\dots,k_r,1)$$

For $k_1, \dots, k_r \geq 2$ we define $p_s(k_1, \dots, k_r) := p(k_1, \dots, k_r, \underbrace{1, \dots, 1}_{s})$. So

$$\begin{aligned} p_{s+1}(k_1, \dots, k_r) = p_s(k_1, \dots, k_r) - &\sum_{j=1}^{r} p_s(k_1, \dots, k_{j-1}, k_j + 1, k_{j+1}, \dots, k_r) \\ &- s p_{s-1}(k_1, \dots, k_r, 2) \end{aligned}$$

and this is the same recursion that we have for the $p_{b;k_1,\dots,k_r;s}$ (cf. the remark in the proof of Theorem 5.1). Since $p_{b;k_1,\dots,k_r;0} = p_0(k_1, \dots, k_r)$, we therefore have

$$\begin{aligned} p_{b;k_1,\dots,k_r;s} = p_s(k_1, \dots, k_r) &= \frac{a^{r+s-1}(r+s-1)!}{(b-1)!} \prod_{j=1}^{r} [1-a]_{k_j-1} \\ &= \frac{a^{r+s-1}(r+s-1)!}{(b-1)!} \prod_{j=1}^{r} (k_j - a - 1)_{k_j-1} \end{aligned}$$

for all $b, r \in \mathbb{N}$, $s \in \mathbb{N}_0$, $k_1, \dots, k_r \geq 2$ such that $b = k_1 + \dots + k_r + s$.

Appendix

A Poisson Point Processes

Let $(E, \mathcal{E})$ be a Polish space, equipped with its Borel σ-algebra.

Definition A.1. *A* ***random measure*** *on E is a map $\nu : \Omega \times \mathcal{E} \mapsto \mathbb{R}_+$ such that*

1. *For all $\omega \in \Omega$, $\nu(\omega, .)$ is a measure on $(E, \mathcal{E})$.*
2. *For all $A \in \mathcal{E}$, $\nu(., A)$ is a random variable.*

Let μ be a σ-finite measure on E.

Definition A.2. *A* ***Poisson random measure*** *of intensity μ is a random measure M on E such that for all A with $\mu(A) < \infty$ we have*

$$\begin{aligned} &\mathbb{P}(M(., A) = k) = e^{-\mu(A)} \frac{\mu(A)^k}{k!} \text{ for all } k \in \mathbb{N}_0 \text{ and} \\ &\text{if } A \cap B = \emptyset, \quad M(., A) \text{ and } M(., B) \text{ are independent.} \end{aligned}$$

Now let μ be a σ-finite measure on E and let λ be the Lebesgue measure on $\mathbb{R}_+$. Let M be a Poisson random measure on $\mathbb{R}_+ \times E$ of intensity $\lambda \otimes \mu$. With the definition of a Poisson process that is given in Revuz and Yor (1999), Chapter XII. Definition (1.3), it is easy to see that for all $A \in \mathcal{E}$ with $\mu(A) < \infty$,

$$M_t^A(\omega) := M(\omega, [0, t] \times A)$$

defines a Poisson process with intensity $\mu(A)$. Since μ is σ-finite, we therefore have $M(\{t\} \times E) \in \{0, 1\}$ a.s. Since E is Polish and $\mathcal{E}$ is its Borel σ-algebra: If $M(\{t\} \times E) = 1$, then there exists $x \in E$ such that $M(\{(t, x)\}) = 1$.

Now we are able to define a **Poisson point process** of intensity μ: Let $\delta \notin E$ be a point that does not belong to E. We define

$$e(t) := \begin{cases} \delta, & \text{if } M(\{t\} \times E) = 0 \\ x, & \text{if } M(\{(t,x)\}) = 1 \end{cases}$$

It is easy to see that a Poisson point process $(e(t))_{t\geq 0}$ satisfies for all $s \geq 0$:

$$(e(t+s))_{t\geq 0} \perp\!\!\!\perp (e(t))_{0\leq t<s} \text{ and} \tag{55}$$

$$(e(t+s))_{t\geq 0} \simeq (e(t))_{t\geq 0} \text{ where } \simeq \text{ denotes equality in law.} \tag{56}$$

Let $A \in \mathcal{E}$ such that $\mu(A) < \infty$ and let $B \subseteq A$ be a Borel subset of A. Let $T_A := \inf\{t \geq 0 : e(t) \in A\}$. With the elementary properties of exponential random variables we obtain

$$\mathbb{P}(e(T_A) \in B) = \mu(B)/\mu(A) \tag{57}$$

B Subordinators

We consider a measure Λ on $(0,\infty)$ that satisfies

$$\int_{(0,\infty)} (1 \wedge x)\Lambda(dx) < \infty \tag{58}$$

Let $(e(t), 0 \leq t \leq 1)$ be a Poisson point process on $[0,1]$ with intensity Λ. We define

$$\xi_t := \sum_{0\leq s\leq t: e(s)\neq\delta} e(s), \; t \in [0,1]$$

$(\xi_t, 0 \leq t \leq 1)$ is called a **subordinator** and Λ is its **Lévy measure**. We remark that this is not the most general form of a subordinator. This special case is also called **pure jump subordinator**.

The condition (58) assures that ξ_1 is finite. It is easily verified that (ξ_t) is an increasing process with independent and stationary increments, hence it is a Lévy process. With the Lévy-Khintchine formula we obtain the Laplace exponent of (ξ_t) (cf. Bertoin (1996), p. 72):

$$\mathbb{E}(e^{-q\xi_t}) = \exp(-t\Phi(q)) \text{ where}$$

$$\Phi(q) = \int_{(0,\infty)} (1 - e^{-qx})\Lambda(dx)$$

On the other side, this exponent determines the law of the subordinator.

Note that because of condition (58), for $x > 0$ there is only a finite number of jumps of (ξ_t) of size $> x$. Therefore we can order the jumps of (ξ_t) in decreasing order: $a_1 \geq a_2 \geq \ldots$

C Martingale Problems

Definition C.1. *Assume we are given a Polish space E, a distribution ν on E, and an operator*

$$A : B(E) \supseteq \mathcal{D}(A) \to B(E)$$

where $B(E)$ denotes the space of bounded measurable functions on E. We call a process $(X_t : t \geq 0)$ with values in E a **solution to the (A, ν)-martingale problem** *if and only if*

$$X_0 \sim \nu$$

and for all $f \in \mathcal{D}(A)$ the process

$$f(X_t) - \int_0^t Af(X_s)ds, \quad t \geq 0$$

is a martingale with respect to the filtration

$$\mathcal{F}_t := \sigma(X_s : s \leq t) \vee \sigma\left(\int_0^s h(X_u)du : s \leq t, h \in B(E)\right)$$

We say that there is **uniqueness for the (A, ν)-martingale problem** *if any two solutions have the same finite-dimensional distributions.*

Example C.2. Let X be a Markov process with starting distribution ν and with infinitesimal generator G. Then X is a solution to the (G, ν)-martingale problem. Cf. Ethier and Kurtz (1986), Proposition 1.7 of Chapter 4.

Proposition C.3. *Let E be a Polish space and let A be an operator on $B(E)$. Suppose that for every distribution ν on E the one-dimensional distributions of the solution of the (A, ν)-martingale problem are uniquely determined. That is, for every two solutions X and Y of the (A, ν)-martingale problem and for every $t \geq 0$ we have*

$$X_t \simeq Y_t$$

where $\simeq$ denotes equality in law. Then any two solutions X and Y have the same finite-dimensional distributions, and any solution X is a Markov process with respect to the filtration $(\mathcal{F}_t)$. If X and Y are two solutions with càdlàg paths, then they have the same distribution on $D([0, \infty), E)$ since from Proposition 7.1 in Chapter 3 of Ethier and Kurtz (1986) we obtain easily that the distribution of a process on $D([0, \infty), E)$ is determined by its finite-dimensional distributions.

Proof. Cf. Theorem 4.2 of Chapter 4 in Ethier and Kurtz (1986). □

Example C.4. Let $E := \{1, \ldots, N\}$ and let A be any operator whose domain includes all functions on E. Let ν be any distribution on E. Then the (A, ν)-martingale problem has at most one solution (i.e. any two solutions have the same finite-dimensional distributions): Let f be a function on E and let X be a solution to the (A, ν)-martingale problem. Then

$$\begin{aligned}\mathbb{E}(f(X_{t+s})|\mathcal{F}_t) &= f(X_t) - \int_0^t Af(X_u)du + \mathbb{E}\left(\int_0^{t+s} Af(X_u)du|\mathcal{F}_t\right) \\ &= f(X_t) + \int_t^{t+s} A\mathbb{E}(f(X_u)|\mathcal{F}_t)du\end{aligned}$$

And this integral equation has a unique solution

$$\mathbb{E}(f(X_{t+s})|\mathcal{F}_t) = e^{sA}f(X_t)$$

In particular we have

$$\mathbb{E}(f(X_t)) = \int_E e^{tA}f(y)\nu(dy)$$

which shows that the one-dimensional distributions of any solution are uniquely determined which by Proposition C.3 implies the uniqueness of the solutions.

D Regular Variation

A function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is said to be of **regular variation with index a** $\in \mathbb{R}$, if for any $C > 0$

$$\lim_{x\to\infty} \frac{f(Cx)}{f(x)} = C^a \tag{59}$$

If $a = 0$, then f is said to be of **slow variation**. In particular, every function f of regular variation with index a can be written as

$$f(x) = x^a x^{-a} f(x) = x^a l(x)$$

for $x > 0$. $l(x) := x^{-a} f(x)$ is a function of slow variation:

$$\lim_{x\to\infty} \frac{l(Cx)}{l(x)} = \lim_{x\to\infty} \frac{C^{-a}x^{-a}f(Cx)}{x^{-a}f(x)} = 1$$

Proposition D.1. *If l is a function of slow variation, then $\lim_{x\to\infty} x^a l(x) = \infty$ for all $a > 0$, and $\lim_{x\to\infty} x^a l(x) = 0$ for all $a < 0$.*

Cf. Proposition 1.3.6 of Bingham et al. (1989).
In fact, for functions of slow variation the convergence (59) is uniformly in C:

Theorem D.2. *Let l be a function of slow variation. Then for any compact set $K \subset (0, \infty)$ we have*

$$\lim_{x\to\infty} \sup_{C\in K} \frac{l(Cx)}{l(x)} = 1$$

One of the most useful results for functions of regular variation is Karamata's theorem:

Theorem D.3 (Karamata)**.** *Let l be a function of regular variation that is bounded on each compact interval. Then we have for all $K > 0$*

1. *For $a > -1$*

$$\int_K^x y^a l(y)dy \sim \frac{x^{a+1}l(x)}{a+1},\ x \to \infty$$

2. *For $a = -1$, $\int_K^x l(y)y^{-1}dy$ is of regular variation and*

$$\frac{1}{l(x)} \int_K^x \frac{l(y)}{y} dy \to \infty,\ x \to \infty$$

3. *For $a < -1$, $\int_x^\infty y^a l(y)dy$ converges when x tends to infinity, and*

$$\int_x^\infty y^a l(y)dy \sim \frac{x^{a+1}l(x)}{-a-1},\ x \to \infty$$

This is shown in Bingham et al. (1989), Proposition 1.5.8 to 1.5.10.

Proposition D.4. *If f is a function of regular variation with index $-a$ for some $a > 0$, then there exists an* **asymptotic inverse** *g of f. g is of regular variation with index $1/a$ and satisfies*

$$f(g(x)) \sim \frac{1}{x}$$

This is Theorem 1.5.12 in Bingham et al. (1989).

Proposition D.5 (Potter's bound)**.** *Let l be a function of slow variation that is bounded away from 0 and from ∞ on every compact subset of $[K, \infty)$ for some $K \geq 0$. Then for every $\delta > 0$ there exists a constant C_δ such that*

$$\frac{l(x)}{l(y)} \leq C_\delta \max \left\{ \left(\frac{x}{y} \right)^{\delta}, \left(\frac{x}{y} \right)^{-\delta} \right\}$$

for all $x, y > K$.

References

D.J. Aldous. Exchangeability and related topics. In P.L. Hennequin, editor, *École d'Été de Probabilités de Saint-Flour XIII - 1983*, volume 1117 of *Lecture Notes in Mathematics*. Springer-Verlag, 1985.

J. Bertoin. *Lévy Processes*, volume 121 of *Cambridge Tracts in Mathematics*. Cambridge University Press, 1996.

J. Bertoin. *Random Fragmentation and Coagulation Processes*, volume 102 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2006.

J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields*, 126:261–288, 2003.

P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., 1968.

N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1989.

M. Birkner, J. Blath, M. Möhle, M. Steinrücken, and J. Tams. A modified lookdown construction for the xi-fleming-viot process with mutation and populations with recurrent bottlenecks. *ALEA Lat. Am. J. Probab. Math. Stat.*, 6:25–61, 2009.

E. Bolthausen and A.-S. Sznitman. On ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.

C. Cannings. The latent roots of certain markov chains arising in genetics: a new approach, i. haploid models. *Adv. Appl. Probab.*, 6:260–290, 1974.

C. Cannings. The latent roots of certain markov chains arising in genetics: a new approach, ii. further haploid models. *Adv. Appl. Probab.*, 7:264–282, 1975.

V. Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25:285–287, 1979.

D. Dawson and K. Hochberg. Wandering random measures in the fleming-viot model. *Ann. Probab.*, 10(3): 554–580, 1982.

A.M. Etheridge. *An Introduction to Superprocesses*, volume 20 of *University Lecture Series*. American Mathematical Society, 2000.

S. Ethier and T. Kurtz. *Markov Processes, Characterization and Convergence.* Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons, Inc., 1986.

W.H. Fleming and M. Viot. Some measure-valued markov processes in population genetics theory. *Indiana University Mathematics Journal*, 28(5):817–843, 1979.

J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes*, volume 288 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, 2002.

O. Kallenberg. Canonical representations and convergence criteria for processes with interchangeable increments. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 27:23–36, 1973.

J.F.C. Kingman. The representation of partition structures. *J. London Math. Soc.*, 18:374–380, 1978.

J.F.C. Kingman. The coalescent. *Stochast. Process. Appl.*, 13:235–248, 1982a.

J.F.C. Kingman. On the genealogy of large populations. *J. Appl. Probab.*, 19A:27–43, 1982b.

A. Klenke. *Probability Theory. A Comprehensive Course.* Universitext. Springer-Verlag, 2008.

T. Kurtz. *Approximation of Population Processes*, volume 36 of *CBMS-NSF Regional Conf. Series in Appl. Math.* SIAM, 1981.

T.M. Liggett. *Interacting Particle Systems.* Springer-Verlag, 1985.

M. Möhle. Weak convergence to the coalescent in neutral population models. *J. Appl. Probab.*, 36:446–460, 1999.

M. Möhle. Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.*, 32:983–993, 2000.

M. Möhle and S. Sagitov. A characterization of ancestral limit processes arising in haploid population genetics models. preprint, ftp://ftp.mathematik.uni-mainz.de/pub/publications/misc/Moehle/general.ps.Z, 1998.

M. Möhle and S. Sagitov. A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.*, 29(4):1547–1562, 2001.

J. Munkres. *Topology.* Prentice Hall, 2nd edition, 2000.

N. Perkowski. Coalescents échangeables obtenus des processus de galton-watson. Master's thesis, Université Pierre et Marie Curie, Paris, 2009.

J. Pitman. Exchangeable and partially exchangeable random partitions. *Probab. Theory Relat. Fields*, 102: 145–158, 1995.

J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27:1870–1902, 1999.

D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*, volume 293 of *Grundlehren der mathematischen Wissenschaften.* Springer-Verlag, 3rd edition, 1999.

S. Sagitov. The general coalescent with asynchroneous mergers of ancestral lines. *J. Appl. Probab.*, 36: 1116–1125, 1999.

J. Schweinsberg. Coalescents with simultaneous multiple collisions. *Electron. J. Probab.*, 5(12):1–50, 2000a.

J. Schweinsberg. A necessary and sufficient condition for the lambda-coalescent to come down from infinity. *Electron. Comm. Probab.*, 5:1–11, 2000b.

J. Schweinsberg. Coalescent processes obtained from supercritical galton-watson processes. *Stochast. Process. Appl.*, 106:107–139, 2003.

Selbstständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Datum Unterschrift

Einverständniserklärung

Hiermit erkläre ich mich einverstanden, dass ein Exemplar meiner Diplomarbeit für die Bibliothek des Institutes für Mathematik verbleibt.

Datum Unterschrift