
Convergence of the Genealogy of the Spatial Cannings Model

Dissertation zur Erlangung
des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen
im Promotionsstudiengang „Mathematical Sciences“
der Georg-August University School of Science (GAUSS)

vorgelegt von
Benjamin Heuer
aus Bassum

Göttingen, 2016

Betreuungsausschuss

Prof. Dr. Anja Sturm, IMS

Prof. Dr. Axel Munk, IMS

Mitglieder der Prüfungskommission

Referentin: Prof. Dr. Anja Sturm, IMS

Korreferent: Prof. Dr. Dominic Schuhmacher, IMS

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Axel Munk, IMS

PD. Dr. Ulf-Rainer Fiebig, IMS

Prof. Dr. D. Russell Luke, NAM

Prof. Dr. Thomas Schick, MI

Tag der mündlichen Prüfung: 23.09.2016

Abstract

In this thesis we consider the genealogy of a spatial Cannings model. This is a population model in which individuals are distributed over a countable set of sites G . The reproduction of individuals at each site is panmictic (exchangeable) and preserves the local population size. The offspring then migrate to other sites in G , also in an exchangeable manner.

We consider the spatial coalescent introduced by sampling n individuals at present time and tracking their ancestral lines back in time. The resulting process is the spatial Cannings coalescent.

Our main result shows, that an appropriately time-rescaled spatial Cannings coalescent converges to a spatial Ξ -coalescent in the large population limit. The key feature of our result is that the spatial structure is preserved into the limit as opposed to a fast migration limit. The influence of the migration on the local population size can yield a time-inhomogeneous limit and, in case of sites with a small population size, our limiting process may not have a strongly continuous semigroup.

Table of Contents

1	The Forward and the Backward Model	1
1.1	Introduction	1
1.2	The Forward Model	6
1.3	The Backward Model	9
2	Basic Properties of the Ξ-Coalescent	15
3	Main Result: Convergence to the Limiting Coalescent	21
3.1	Assumptions	21
3.2	Examples of Models Meeting the Assumptions	30
3.3	Definition of the Limiting Coalescent	37
3.4	The Limiting result	40
3.5	Discussion of the Result	43
3.6	Overview Over the Rest of the Thesis	47
4	The Vague Topology	49
4.1	Definition and Basic Theory	49
4.2	Convergence of Integer Valued Random Measures	62
5	The Graphical Representation and the Tracking Map	75
5.1	The Space of Graphical Representations	75
5.2	Tracking a Graphical Representation	76
6	Time-Inhomogeneous Regular Jump Processes	81
6.1	Definitions and Basic Properties	81
6.2	Some Set Theory	89
6.3	Identifying the Distribution of a Time-Inhomogeneous Regular Jump Process	96
7	System of Flows	103
7.1	Flows and Particle Tracking	103
7.2	Uniform Convergence of Flows	110

8	Proof of the Limiting Result	115
8.1	Verification of the Properties of \mathcal{F}, \mathcal{R}	115
8.2	Non-Explosiveness of the Limiting Process	117
8.3	Modeling Π^N with Graphical Representations	118
8.4	Convergence of the Graphical Representations	121
8.5	Concluding the Proof	126
9	Outlook	129

Acknowledgments

I would like to thank Anja Sturm, not only for all the encouragement, patience and help she gave me as my doctoral advisor, but also for sparking my interest in stochastics with her great lectures in the first place. Also, I thank my second advisor Axel Munk for being always there in case I had questions. Moreover, I thank Dominic Schuhmacher for sharing his insights into Bernoulli and Poisson processes and for agreeing to be my second referee on relatively short notice. Additionally, I thank Thomas Schick, Ulf Fiebig and Russell Luke for taking some of their time as part of my oral exam committee. I also thank my colleagues at the Institute for Mathematical Stochastics in Göttingen for providing me with a pleasant work environment. In particular, I would like to thank my friends and colleagues Alexander Hartmann and Fabian Telschow, who made for great targets to bounce mathematical ideas and questions off and also for their input regarding some formal as well as formatting questions concerning this thesis. Last, but not least, I would like to thank my parents Jürgen and Gabriela Heuer for their great support over the years.

Chapter 1

The Forward and the Backward Model

1.1 Introduction

Population genetics is the part of evolutionary biology which studies the way in which alleles of genes are passed from one generation of a population to the next over time. For instance, take a population of drosophilae and select a particular gene in their DNA. We can consider questions about the mutation rate of that gene or about the evolutionary advantage, the so called *fitness*, of different alleles of the gene. If we had perfect information about how the population behaved at all times it would be fairly easy to give estimators for these things, but in practice the amount of information available is severely limited. But especially for large populations a lot of the “microscopic” structure may be negligible. There are two approaches to reduce the information considered.

The forward-in-time approach tracks the frequencies of different types of alleles in the population forward in time. Mathematically, this gives rise to a stochastic process taking values in the space of type distributions, a so called *Flemming-Viot-process*.

For the backward-in-time approach we consider a sample of n individuals taken at present time. Each individual gives rise to an ancestral line backwards in time by considering its (genetic) ancestor in each generation. These ancestral lines can merge. Such a merger (also called *collision*) happens whenever the affected ancestral lines find their most recent common ancestor. Mathematically this is modelled as a stochastic process taking values in the space of partitions of $\{1, \dots, n\}$. Such a partition consists of disjoint subsets of $\{1, \dots, n\}$ which we call *blocks*. We name the resulting process the

genealogical process or also the *coalescent of the population*. We now explain the connection of ancestral lines and a partition of $\{1, \dots, n\}$ in greater detail. We start by enumerating the sampled individuals using the numbers $1, \dots, n$. Each (non-empty) block of the partition corresponds to a common ancestor of exactly the sampled individuals which are contained in the block. Thus the block is also represented by a particular individual in the population at any point in time, the common ancestor.

The two approaches are often equivalent in the sense that there is a distribution determining duality between the forward Fleming-Viot-process and the backward genealogical-process for $n \rightarrow \infty$. In this thesis we focus on the backwards-in-time approach and thus on the analysis of the genealogical process. In order to answer biological questions the model has to be chosen simple enough, such that a theoretical treatment is possible, but rich enough such that important features of the population are still observable. To simplify the model usually two steps are taken:

In the first step restrictions are stipulated for the reproduction in the population. In the second step a large population limit is considered together with a suitable time rescaling.

If the restrictions of the first step are sensible, then the large population limit in the second step exists and yields a well understood coalescent. In fact, it turns out that under suitable restrictions whole families of population models may give the same type of coalescent in the large population limit. This is usually referred to as the *robustness of the coalescent*.

One example for such a robustness result is when we consider the so called *Wright-Fisher-model*. This is a particular reproduction mechanism in which each individual chooses its parent uniformly at random from the previous generation. Since multiple individuals can choose the same parent the genealogy of this population would feature multiple mergers (mergers in which more than 2 ancestral lines merge into one line at the same time) and even simultaneous multiple mergers (mergers in which multiple collections of ancestral lines merge into single lines each). But if the coalescent of the sample is sped up linearly with the population size we see a so called *Kingman-coalescent* emerge in the limit. That is a coalescent process, introduced first by Kingman [17], in which only pairs of ancestral lines merge independently and with a fixed rate. In particular neither simultaneous nor multiple mergers occur in the large population limit.

This result can be generalized to a robustness result for coalescents of so called *Cannings-models*. The defining property of these population models, which were considered by Cannings [3] and [4], is that the reproduction in each generation leaves the population size fixed and the vector of offspring distributions is exchangeable (i.e., invariant under permutation) in each

generation. Biologically this means that we consider a population of stable size and we do not have additional information about the fertility of sampled individuals. It is commonly known that the Wright-Fisher-model is a special case of a Cannings-model.

It was shown by Möhle and Sagitov [20] that the coalescent of these Cannings-Models, under certain conditions on the moments of the offspring distributions, converges in the large population limit to a so called Λ - n -coalescent or (under weaker assumptions) to the more general Ξ - n -coalescent. In order to define the Λ - n -coalescent we first choose a finite measure Λ on $[0, 1]$. We set for $2 \leq k \leq m$:

$$\lambda_k^m := \int_0^1 x^{k-2}(1-x)^{m-k} d\Lambda(x).$$

The Λ - n -coalescent is a Markov chain in continuous time (MCCT) on the space of partitions of $\{1, \dots, n\}$ first independently considered by Pitman [22] and Sagitov [23]. In this process multiple mergers may be possible. Indeed, if m ancestral lines are present in the coalescent, then any selection of k of these lines merges into one line independently with the rate λ_k^m . It should be noted that the Kingman-coalescent is given by the special case $\Lambda = \delta_0$. Simultaneous mergers, however, are still impossible. In the more general situation considered by Möhle and Sagitov [20] the limiting process is a Ξ - n -coalescent, which is again a MCCT with coalescent rates given by a finite measure Ξ on the infinite simplex. In this coalescent simultaneous multiple mergers are possible. The connection to a unique measure Ξ was made explicit by Schweinsberg [27]. Since the difference between Λ - and Ξ - n -coalescents is mostly just additional notation we omit the detailed definition of the rates in case of the Ξ - n -coalescent for the purpose of this introduction. Λ - or even Ξ - n -coalescents arise in particular if the variation of the offspring distribution is large. This is, for example, the case in the reproduction mechanisms analysed by Eldon and Wakeley [7] or can occur due to the presence of recurring selective sweeps as considered by Durrett and Schweinsberg [5] and [6].

Let the aforementioned population of drosophilae be separated in colonies that live on piles of rotting fruit on a wine orchard (example given by Wakeley [29]). This situation introduces an additional piece of information for each individual – its location. Since a single drosophila is unlikely to visit all of the fruit piles in its live time, we can expect to see an impact of the spatial structure on the genealogies of our sample. Therefore it is necessary to extend the theory to accommodate the spatial information.

The theoretical treatment of spatial settings like this has become an important new part of population genetics and is also the main topic of

this thesis. We now extend our model to accommodate spatial information. We start by generalizing the population model considered by Cannings to a spatial model which we call the *spatial Cannings-model*. We consider a population of fixed size N living on an at most countable (discrete) set of sites G . In each generation $k \in \mathbb{N}$ there is a reproduction and a migration step. In the reproduction step the individuals at each site reproduce according to a nonspatial Cannings-model. This means that the offspring distribution is exchangeable and the population size at each site (also called local population size) stays constant. In the migration step the individuals can migrate freely to another site, but they have to do so in an exchangeable manner. To be more precise we count for each two sites $x, y \in G$ and for any generation the number of migrants from x to y in that generation. Exchangeability in this context means, that conditioned on all of these numbers the individuals of the population located at x migrate in a way, such that the resulting distribution is invariant under permutation of the individuals at x and independent of the migration out of all the other sites.

In order to define a spatial version of the coalescent, we sample $n \ll N$ individuals at present time and look at their genealogy. The ancestral line of individual $i \in \{1, \dots, n\}$ of the sample is, at any given time, represented by the ancestor of i . This ancestor has a location in G and thus we can add the spatial information to our genealogical process by labeling every ancestral line with that location. This gives rise to the spatial coalescent of our population. Lines merge whenever they find a most recent common ancestor and lines migrate due to the migration of the ancestor representing the line.

The main result of this thesis is a robustness result similar to the nonspatial case. We give conditions on the distributions which govern the reproduction and migration mechanisms in order to ensure convergence of this spatial coalescent, after application of a properly chosen time rescaling, to a time-continuous limiting process (see Theorem 3.14). In this limiting process lines merge at each site according to a time-rescaled Ξ - n -coalescent (Ξ may depend on the location) and their labels migrate independently on G . This migration happens according to a possibly time-inhomogeneous process, which we call the *particle tracking process*. The name is chosen since the process is reminiscent of the tracking a water molecule through a discrete system of water tanks.

To give a justification and more details for the migration, we assign a mass of $1/N$ to each individual and consider the migratory behavior of the mass in the population backwards in time. For each site $x \in G$ and generation $k \in \mathbb{N}_0$ we have a *total mass* $\mathcal{R}_{k;x}^N \in [0, 1]$ at that site. Furthermore, for each two sites $x \neq y \in G$ and each generation $k \in \mathbb{N}_0$ we can consider the total amount of mass that has moved from x to y up to generation k . We call this the

cumulative mass flow $\mathcal{F}_{k;x,y}^N$. We assume that these discrete-time processes converge in the large population limit, after a proper time-rescaling, to time-continuous processes $(\mathcal{F}_{t;x,y})_{t \in \mathbb{R}_+}$, $\mathcal{R}_{t;x}$ such that $t \mapsto \mathcal{F}_{t;x,y}$ is absolutely continuous and thus weakly differentiable. In our interpretation with water molecules, the $\mathcal{R}_{t;x}$ describes the amount of water in tank x at time t and $\mathcal{F}_{t;x,y}$ describes the amount of water that has flown from tank x to tank y up to time t . The infinitesimal rate with which the tracked molecule migrates from x to y at time t is given by $\mathcal{F}'_{t;x,y}/\mathcal{R}_{t;x}$. Indeed this is exactly the infinitesimal rate we use for our reproduction mechanism in the limiting process.

Previous work on such spatial settings include results for the structured coalescent by Herbots [11] and a generalization to spatial Λ -coalescents by Heuer and Sturm [12]. However, in those cases the migration mechanism was assumed to be balanced and the total number of individuals at each site as well as the total number of migrants from x to y were assumed to be deterministic. This is no longer required in our setting. It should be noted that in such a balanced situation the processes \mathcal{R}_x and $\mathcal{F}'_{x,y}$ are deterministic and constant which leads to a time-homogeneous limit.

Usually such a limiting result would be proven using the theory of Feller-processes and martingale problems but the spatial component of our limiting process can make simple application of this strategy problematic. It is noteworthy, that we allow for large variations in the local population size due to migration, in particular the paths $t \mapsto \mathcal{R}_{t;x}$ may not be constant. This may make the setup time-inhomogeneous, even in the limit. Even worse, the local population may become very small (of order $o(N)$), which would lead to the existence of times $s \leq t$ with $\mathcal{R}_{s;x} > \mathcal{R}_{t;x} = 0$. As a consequence, the migration mechanism of the limiting coalescent is not a Feller-process and indeed does not even feature a strongly continuous semigroup.

For that reason we have to take a different approach in order to show convergence. We use the following steps: First, we use *graphical representations* to construct our processes. Second, we show that these graphical representations converge with respect to the *vague topology*. Third, we show that the mapping which maps a graphical representation to a process is continuous if the limiting process is *non-explosive*. Forth, we show that our limiting process is non-explosive by analysing distributions of *time-inhomogeneous Markov chains*. There is a more detailed sketch of the proof after the limiting result (Theorem 3.14). The complete proof is done in Chapter 8.

1.2 The Forward Model

In this section we describe our population model in detail. Note that a tilde over a process denotes, that the process runs forward in time. The notation for time-reversed versions of the processes later on drops the tilde. Furthermore, we use the notation $[m] := \{1, \dots, m\}$ for $m \in \mathbb{N}_+$. First, we define processes to model the migration of the population.

Definition 1.1 (The Migration Mechanism). Let $N \in \mathbb{N}_+$ be the total number of individuals and G a countable set of all available colonies to the population. Let $\tilde{\mathcal{N}}_0^N := (\tilde{\mathcal{N}}_{0;x}^N)_{x \in G}$ be a random variable on \mathbb{N}_0^G with $\sum_{x \in G} \tilde{\mathcal{N}}_{0;x}^N = N$. The number $\tilde{\mathcal{N}}_{0;x}^N$ counts the individuals at site x in generation 0. For each generation $k \in \mathbb{N}_0$ let $\tilde{\mathcal{M}}_k^N := (\tilde{\mathcal{M}}_{k;x,y}^N)_{(x,y) \in G^2}$ be a random variable on $\mathbb{N}_0^{G \times G}$. We consider $\tilde{\mathcal{M}}_{k;x,y}^N$ to be the total number of migrants who moved from x in generation k to y in generation $k+1$. In order for $\tilde{\mathcal{M}}_k^N$ to describe migration in this way, it needs to fulfill an additional property which ensures, that all individuals are accounted for: Define the number of individuals at site x in generation k by

$$\tilde{\mathcal{N}}_{k;x}^N := \tilde{\mathcal{N}}_{0;x}^N + \sum_{l=1}^k \sum_{y \in G} (\tilde{\mathcal{M}}_{l;y,x}^N - \tilde{\mathcal{M}}_{l;x,y}^N), \quad \tilde{\mathcal{N}}_k^N := (\tilde{\mathcal{N}}_{k;x}^N)_{x \in G}.$$

We then only allow $\tilde{\mathcal{M}}_k^N$ with the property $\sum_{y \in G} \tilde{\mathcal{M}}_{k;x,y}^N = \tilde{\mathcal{N}}_{k-1;x}^N$ for $k \in \mathbb{N}_+$. Define the discrete time processes $\tilde{\mathcal{N}}^N = (\tilde{\mathcal{N}}_k^N)_{k \in \mathbb{N}_0}$ (the process of the total numbers) and $\tilde{\mathcal{M}}^N = (\tilde{\mathcal{M}}_k^N)_{k \in \mathbb{N}_0}$ (the process of the total migration). Given $\tilde{\mathcal{N}}^N$ and $\tilde{\mathcal{M}}^N$ we now define the migration process $\tilde{X}^N = (\tilde{X}_k^N)_{k \in \mathbb{N}_0}$ which tracks the migration of each individual in the population. Let $\tilde{X}_k^N = (\tilde{X}_{k;i}^N)_{i \in [N]}$ be a random variable in $G^{[N]}$. We define the process \tilde{X}^N by the following procedure: define an urn with N balls which are colored, using the elements of G . Let $\tilde{\mathcal{N}}_{0;x}^N$ be the number of balls with color $x \in G$. For each $i \in [N]$ we draw a ball from the urn without replacement. Let $\tilde{X}_{0;i}^N \in G$ be the result of the i -th draw. Now assume that we have already defined \tilde{X}_{k-1}^N for some $k \in \mathbb{N}_+$. For $x \in G$ let $I_{k;x} \subseteq [N]$ be the set of indices i with $\tilde{X}_{k-1;i}^N = x$. For each $x \in G$ we define an urn with $|I_{k;x}| = \tilde{\mathcal{N}}_{k-1;x}^N$ balls with colors in G . Moreover, let $\tilde{\mathcal{M}}_{k;x,y}^N$ be the number of balls of color $y \in G$. For each $i \in I_{k;x}$ we draw a ball out of the urn without replacement. We then set $\tilde{X}_{k;i}^N \in G$ to be the result of the i -th draw. Since the procedure yields $\sum_{x \in G} I_{k;x} = N$ this defines \tilde{X}_k^N completely. Note that by construction each path of the migration process \tilde{X}^N represents a migration of the population which is consistent with the total numbers $\tilde{\mathcal{N}}^N$ and the total migrants $\tilde{\mathcal{M}}^N$.

Remark 1.2. By symmetry, \tilde{X}^N run up to a time k yields the uniform distribution on all migrations that are consistent with the processes $\tilde{\mathcal{N}}^N, \tilde{\mathcal{M}}^N$. This could be used as an alternative, shorter definition for \tilde{X}^N . But since the step by step definition using urns will be useful later on we chose the above definition instead.

We now define random variables in order to describe the reproduction.

Definition 1.3 (The Reproduction Mechanism). Let $N \in \mathbb{N}_+$. For each site $x \in G$, generation $k \in \mathbb{N}$ and each potential choice of individuals $I \subseteq [N]$ let $\tilde{\nu}_{k;x}^I = (\tilde{\nu}_{k;x,i}^I)_{i \in I}$ be an exchangeable random vector with $\tilde{\nu}_{k;x,1}^I \in \mathbb{N}_0$ and

$$\sum_{i \in I} \tilde{\nu}_{k;x,i}^I = |I|.$$

Here $\tilde{\nu}_{k;x,i}^I$ is the number of offspring of the i -th individual at site x in generation k if there is a collection of individuals I present at x . Let these random vectors be mutually independent from each other and $\tilde{\mathcal{N}}^N, \tilde{\mathcal{M}}^N$ as well as \tilde{X}^N . Furthermore, let $\tilde{\nu}_{k;x}^I$ have identical distributions for $k \in \mathbb{N}_0$ and only dependent on the total number of individuals $|I|$ (as opposed to the specific choice of $I \subseteq [N]$).

We can now define the population model by combining the migration and the reproduction mechanisms.

Definition 1.4 (The Population Model). Let $N \in \mathbb{N}_+$. Again, we label the individuals with numbers $i \in [N]$. At generation $k = 0$ individual i of the population is at site $\tilde{X}_{0;i}^N$. Now whenever we reach a new generation k our population goes through the following steps:

1. At site $x \in G$ let $I_{k;x}$ be the set of labels of individuals in x in generation $k - 1$ (see Definition 1.1). Each individual $i \in I_{k;x}$ has a number of offspring given by $\tilde{\nu}_{k;x,i}^{I_{k;x}}$ (see Definition 1.3). Due to

$$\sum_{i \in I_{k;x}} \tilde{\nu}_{k;x,i}^{I_{k;x}} = |I_{k;x}|$$

the total number of offspring at site x is equal to the number of available parents.

2. We now label the offspring, using the collection of labels $I_{k;x}$ of their parents. We may do so in an arbitrary manner.
3. For each $i \in I_{k;x}$ the offspring with label i migrates to the site $\tilde{X}_{k;i}^N$.

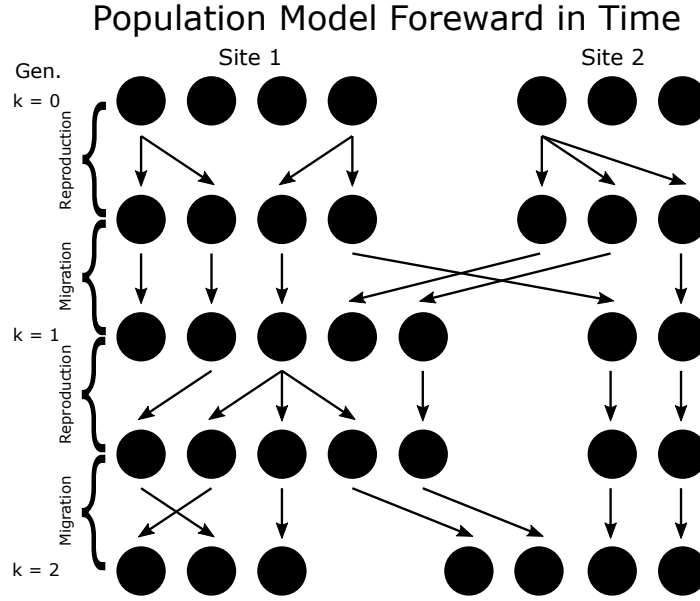


Figure 1.1: Visualization of the reproduction and migration steps of our population model for $N = 7$, $G = \{1, 2\}$ over two generations.

These steps are illustrated for a special case in Figure 1.1. Here we consider the case $N = 7$, $G = \{1, 2\}$ and the starting conditions $\tilde{\mathcal{N}}_{0;1}^N = 4$, $\tilde{\mathcal{N}}_{0;2}^N = 3$. Due to exchangeability it is not necessary to enumerate the individuals in the figure and it suffices to represent them using black balls. Arrows in the reproduction step point from parent to offspring, arrows in the migration step indicate movement of individuals.

Remark 1.5. It is noteworthy, that by Remark 1.2 and since $\tilde{\nu}_{k;x}^I$ does not depend on the specific choice of I other than $|I|$ any choice of the initial labeling of the population as well as any choice for the labeling of the offspring yields the same distribution. This, of course, is the reason why we did not specify these choices further.

Examples 1.6. Common examples for the reproduction mechanism include:

1. The *Moran model* in which we choose $(\tilde{\nu}_{k;x,i}^{I_{k;x}})_{i \in I_{k;x}}$ to be uniformly distributed on

$$\{x \in \{0, 1, 2\}^{|I_{k;x}|} \mid \exists i, k \in [I_{k;x}] : x_i = 0, x_k = 2, x_l = 1 \forall l \in [I_{k;x}] \setminus \{i, k\}\}$$

or in words: the vector $(\tilde{\nu}_{k;x,i}^{I_{k;x}})_{i \in I_{k;x}}$ is uniformly chosen from all vectors which have exactly one component equal to 0, one component equal to 2 and the remaining components equal to 1.

2. The *Wright-Fisher model* in which we choose $(\tilde{V}_{k;x,i}^{I_{k;x}})_{i \in I_{k;x}}$ to be a vector of i.i.d. Poisson distributed random variables conditioned on their sum being equal $|I_{k;x}|$. An alternative way of describing this distribution would be that the vector is multinomially distributed. More precisely we consider an urn containing one ball for each color $i \in I_{k;x}$. Now we draw $|I_{k;x}|$ times with replacement from the urn and set $\tilde{V}_{k;x,i}^{I_{k;x}}$ to be the total number of draws of color i .

Examples for the migration mechanism include:

1. *Balanced, time homogeneous migration* where $\tilde{\mathcal{M}}_k^N$ are identically distributed over $k \in \mathbb{N}_0$ and we have for all $x \in G$:

$$\sum_{y \in G} (\tilde{\mathcal{M}}_{1;y,x}^N - \tilde{\mathcal{M}}_{1;x,y}^N) = 0$$

thus implying that $k \mapsto \tilde{\mathcal{N}}_k^N$ is constant. In particular there is the special case in which the processes $\tilde{\mathcal{M}}^N$ and $\tilde{\mathcal{N}}^N$ are also deterministic.

2. *Migration via exchangeable random walkers*: We consider an independent collection of time-homogeneous Markov chains in discrete time $(X^{i,N})_{i \in [N]}$ on G with the same transition matrix P^N such that the starting vector $(X_0^{1,N}, \dots, X_0^{N,N})$ is exchangeable and set for $x \in G$

$$\tilde{\mathcal{N}}_{0;x}^N := \sum_{i \in [N]} \mathbb{1}_{\{X_0^{i,N} = x\}}$$

and for $x, y \in G$ with $x \neq y$ and $k \in \mathbb{N}_0$:

$$\tilde{\mathcal{M}}_{k;x,y}^M := \sum_{i \in [N]} \mathbb{1}_{\{X_k^{i,N} = x, X_{k+1}^{i,N} = y\}}.$$

Since the vectors $(X_k^{1,N}, \dots, X_k^{N,N})$ stay exchangeable for all $k \in \mathbb{N}_0$ we can construct the migration mechanism now defined by the matrix $\tilde{\mathcal{M}}$ and the vector $\tilde{\mathcal{N}}_0$ by assigning the random walk $X^{i,N}$ to individual i .

1.3 The Backward Model

The goal of this section is to define the coalescent of the population model defined in Section 1.4. First, we reverse time.

Definition 1.7. Let $l \in \mathbb{N}$ and define for $k \in [l]$

$$\mathcal{M}_k^N := (\tilde{\mathcal{M}}_{l-k+1}^N)^T \quad (\text{the transposed matrix})$$

as well as $\mathcal{N}_0^N := \tilde{\mathcal{N}}_l^N$. For $x \in G$ and $I \subseteq [N]$ define $\nu_{k;x}^I := \tilde{\nu}_{l-k+1;x}^I$. Furthermore we now define the processes \mathcal{N}^N and X^N in the same manner in which we defined the respective tilde-processes but using \mathcal{M}^N and \mathcal{N}_0^N instead of $\tilde{\mathcal{M}}^N$ and $\tilde{\mathcal{N}}_0^N$.

The next proposition shows that the construction of the migration using the reversed processes \mathcal{N} and \mathcal{M} is equivalent to reversing the processes $\tilde{\mathcal{N}}^N$ and \tilde{X}^N .

Proposition 1.8. We have $\mathcal{N}^N = (\mathcal{N}_{l-k}^N)_{k \in \{0, \dots, l\}}$ and $X^N \stackrel{\mathcal{Q}}{=} (\tilde{X}_{l-k}^N)_{k \in \{0, \dots, l\}}$.

Proof. We prove the first equality by induction over k . By definition the processes are identical for $k = 0$. Assuming $\mathcal{N}_{k-1}^N = \tilde{\mathcal{N}}_{l-k+1}^N$ we have for $x \in G$

$$\begin{aligned} \mathcal{N}_{k;x}^N &= \mathcal{N}_{k-1;x}^N + \sum_{y \in G} (\mathcal{M}_{k;y,x}^N - \mathcal{M}_{k;x,y}^N) \\ &= \tilde{\mathcal{N}}_{l-k+1;x}^N + \sum_{y \in G} (\tilde{\mathcal{M}}_{l-k+1;x,y}^N - \tilde{\mathcal{M}}_{l-k+1;y,x}^N) \\ &= \tilde{\mathcal{N}}_{l-k;x}^N + \sum_{y \in G} (\tilde{\mathcal{M}}_{l-k+1;y,x}^N - \tilde{\mathcal{M}}_{l-k+1;x,y}^N) + \sum_{y \in G} (\tilde{\mathcal{M}}_{l-k+1;x,y}^N - \tilde{\mathcal{M}}_{l-k+1;y,x}^N) \\ &= \tilde{\mathcal{N}}_{l-k;x}^N \end{aligned}$$

where the first equality is due to the construction of \mathcal{N}^N and the third equality is due to the construction of $\tilde{\mathcal{N}}^N$.

The second claim is a simple consequence from Remark 1.2: Since \tilde{X}^N up to time l is given by the uniform distribution on all paths belonging to migrations that are consistent with $\tilde{\mathcal{M}}^N$ as well as $\tilde{\mathcal{N}}^N$ and each such path backwards in time corresponds to a path consistent to \mathcal{M}^N as well as \mathcal{N}^N (since migration from x to y forward in time will be migration from y to x backward in time). The claim follows since X^N up to time k also yields the uniform distribution on all paths belonging to migrations that are consistent with \mathcal{M}^N as well as \mathcal{N}^N . \square

Thus, the migration backwards in time is exchangeable if the migration forward in time is exchangeable. Since we are interested in the genealogy of the population we will, from now on, only look at the dynamics of the backwards-in-time processes \mathcal{N}^N , \mathcal{M}^N , X^N . Moreover, we assume that they are given as processes on \mathbb{N}_0 instead of $[l]$. We also redefine our $k = 0$ time

point as the present time at which we can sample from the population. This is more natural for analysis of the backwards processes, than fixing a generation 0 far in the past.

Remark 1.9. Justifying the possibility to extend the time reversal infinitely far into the past may be questionable without additional properties of the process $(\tilde{\mathcal{N}}^N, \tilde{\mathcal{M}}^N)$. If $(\tilde{\mathcal{N}}^N, \tilde{\mathcal{M}}^N)$ is a time-homogeneous, irreducible Markov chain with transition matrix P and equilibrium distribution $\mu = (\mu_i)$, then the backwards process can be extended to \mathbb{N}_0 as the time-reversal of the Markov chain. More precisely we define $(\mathcal{N}^N, \mathcal{M}^N)$ as the Markov chain which has the transition matrix \hat{P} defined by the equations

$$\mu_j \hat{P}_{j,i} = \mu_i P_{i,j} \quad \text{for all states } i, j.$$

It should be noted though, that we do not require this Markov property in our results. Any model for which we can define $(\mathcal{N}^N, \mathcal{M}^N)$ for all times in a sensible manner can be considered.

Before we define the coalescent of the population we first want to specify the respective state space and some notation.

Definition 1.10 (The Spaces of Partitions and of Labeled Partitions). Let $n \in \mathbb{N}_+ \cup \{\infty\}$. We define \mathcal{P}_n as the set of all *partitions* of $[n]$. In the case $n = \infty$ we omit the subscript n . We may represent a partition $\pi \in \mathcal{P}_n$ either by the equivalence relation \sim_π it defines on $[n]$ or by its non-empty equivalence classes $(B_k)_{k \in [l]}$, also called *blocks* (l denotes the number of non-empty equivalence classes in π). We order the blocks B_k by their smallest elements, writing $\pi = (B_1, \dots, B_l)$. We call π trivial if it only has blocks of size 1, called *singletons* (i.e., $\pi = (\{1\}, \dots, \{n\})$). For a spatial setting we have to extend this definition. Given a set G let $\mathcal{P}_{G,n}$ be the set of *labeled partitions* of $[n]$, meaning that we have a partition in the above sense but each block B_k also carries a label $L_k \in G$. Again, we drop the subscript n in case of $n = \infty$. To be precise we write $\pi = (B_k, L_k)_{k \in [l]} \in \mathcal{P}_{G,n}$ exactly, if we have $(B_k)_{k \in [l]} \in \mathcal{P}_n$ and $L_k \in G$ for all $k \in [l]$.

For $m \leq n$ we can define a restriction map

$$\tau_m^n: \mathcal{P}_n \rightarrow \mathcal{P}_m.$$

For $\pi \in \mathcal{P}_n$ we define $\tau_m^n(\pi)$ to be the restriction of the equivalence relation \sim_π on $[n]$ to $[m]$. In terms of blocks this means that if we have $\pi = (B_k)_{k \in [l]}$ we get $\tau_m^n(\pi) = (B_k \cap [m])_{k \in [l']}$ where $l' \in [l]$ is the largest natural number with $B_{l'} \cap [m] \neq \emptyset$. As before we omit the superscript in the case $n = \infty$. In the same manner we can define a restriction $\tau_{G,m}^n$ for labeled partitions by defining

the restriction of $\pi = (B_k, L_k)_{k \in [l]} \in \mathcal{P}_{G,n}$ to be $\tau_{G,m}^n(\pi) = (B_k \cap [m], L_k)_{k \in [l']}$ where $l' \in [l]$ is defined as in the nonspatial case. For the case $n = \infty$ we omit the superscript. With these restrictions we can now define metrics on \mathcal{P}_n and $\mathcal{P}_{G,n}$ by setting

$$d_n(\pi, \pi') = \sup_{k \in [n]} 2^{-k} \mathbf{1}_{\{\tau_k^n(\pi) \neq \tau_k^n(\pi')\}} \quad \text{for } \pi, \pi' \in \mathcal{P}_n$$

and

$$d_{G,n}(\pi, \pi') = \sup_{k \in [n]} 2^{-k} \mathbf{1}_{\{\tau_{G,k}^n(\pi) \neq \tau_{G,k}^n(\pi')\}} \quad \text{for } \pi, \pi' \in \mathcal{P}_{G,n}.$$

Again, we omit the subscript n in the case that $n = \infty$.

Remark 1.11. One can see that in case $n = \infty$ the spaces (\mathcal{P}, d) and (\mathcal{P}_G, d_G) are Polish spaces. In the case $n \in \mathbb{N}_+$ this is trivially true since the spaces then are even finite, discrete spaces. Since we will only consider the case $n < \infty$ in this thesis, we omit the proof for $n = \infty$.

The following definition specifies what we mean by a coalescent mathematically.

Definition 1.12 (Collisions and Coalescent Processes). Let $n, m \in \mathbb{N}_+ \cup \{\infty\}$ with $n \geq m$. Let $\mu \in \mathcal{P}_m$ and $\pi \in \mathcal{P}_n$. Then we define the π -collision of μ as the unique partition in \mathcal{P}_m given by merging exactly the collections of blocks in μ which have their index in a mutual block of π . More precisely let $\mu = (A_1, A_2, \dots)$ and $\pi = (B_1, B_2, \dots)$ then we define the π -collision of μ as the partition given by ordering the blocks $C_1, C_2 \dots$ given by

$$C_j := \bigcup_{i \in B_j} A_i$$

with respect to their smallest elements.

Now let $n \in \mathbb{N}_+$. We call a stochastic process $\Pi = (\Pi_t)_{t \in \mathbb{R}_+}$ on \mathcal{P}_n an (n) -coalescent if it is a càdlàg process and if the jumps of the paths of Π are given by π -collisions for some suitable partition π . We call a stochastic process $\Pi = (\Pi_t)_{t \in \mathbb{R}_+}$ on \mathcal{P} a coalescent if $\tau_n(\Pi)$ is an n -coalescent for all $n \in \mathbb{N}_+$.

Now let G be a topological space. We call a stochastic process $\Pi = (\Pi_t)_{t \in \mathbb{R}_+}$ on $\mathcal{P}_{G,n}$ a spatial (n) -coalescent if it is a càdlàg process and if the process in \mathcal{P}_n given by forgetting the labels of blocks in Π is an n -coalescent. We call a stochastic process $\Pi = (\Pi_t)_{t \in \mathbb{R}_+}$ on \mathcal{P}_G a spatial coalescent if $\tau_{G,n}(\Pi)$ is a spatial n -coalescent for all $n \in \mathbb{N}_+$.

We use analogous definitions in the discrete-time case by identifying a process in discrete time with its right-continuous, constant extension to continuous time.

of this block in the population. Applying the migration and coalescence steps allows us to derive Π_{k+1}^N from Π_k^N and thus by successive application of the steps we can define Π^N completely.

In order to visualize this procedure we used the situation of Figure 1.1 and sampled $n = 4$ individuals at present time. We marked these individuals red in our diagram and tracked them backwards through the arrows of the diagram. This yields Figure 1.2. At present time $k = 0$ we have $\Pi_0^N = ((\{1\}, 1), (\{2\}, 2), (\{3\}, 2), (\{4\}, 2))$. Going through the diagram one generation backwards in time ($k = 1$) the ancestral lines in our sample with index 1 and 2 have found a common ancestor at site 1 and the line with index 3 migrated from site 2 to site 1, we get $\Pi_1^N = ((\{1, 2\}, 1), (\{3\}, 1), (\{4\}, 2))$. Going back another generation we get $\Pi_2^N = ((\{1, 2, 4\}, 1), (\{3\}, 2))$.

Note that, since the construction only entails drawing without replacement and since the offspring distributions are exchangeable, we can define the coalescent for a smaller sample size $m \leq n$ by taking the coalescent for the sample size n and then “forgetting” the individuals $m + 1, \dots, n$. This property is called the *consistency relation*. More precisely, in terms of the notation in Definition 1.10 we may get the coalescent for sample size m by applying $\tau_{G,m}^n$ to the coalescent for sample size n .

Chapter 2

Basic Properties of the Ξ -Coalescent

Before we continue with our spatial setting we want to introduce the Ξ -*coalescent* which arises in the nonspatial case as the large population limit. The theory in this chapter is an excerpt of Schweinsberg [27] though our notation will differ slightly.

Definition 2.1 (The Ξ -(n -)Coalescent). Consider a family of rates

$$\{\lambda_\pi \in \mathbb{R}_+ \mid \pi \in \mathcal{P}_m \text{ nontrivial for some } m \in \mathbb{N}_+\}$$

such that for all $m \in \mathbb{N}_+$ and all $\pi \in \mathcal{P}_m$ nontrivial the following consistency property holds:

$$\lambda_\pi = \sum_{\tau_m^{m+1}(\mu)=\pi} \lambda_\mu \quad (\text{consistency}). \quad (2.1)$$

Furthermore let λ_π only depend on the ordered sequence of the blocksizes of π , this is usually referred to as the exchangeability of the coalescent. In particular, if π has i nonempty blocks with sizes $l_1 \geq l_2 \geq \dots \geq l_i$ we write

$$\lambda_{l_1, \dots, l_i} := \lambda_\pi \quad (\text{exchangeability}). \quad (2.2)$$

Now let Π be an n -coalescent which is also a time-homogeneous Markov chain such that λ_π is the rate with which a π collision happens if the chain is currently in a state with i nonempty blocks. We call Π a Ξ - n -coalescent and if Π is started in the trivial partition we call it a standard Ξ - n -coalescent.

Let Π be a coalescent taking values in \mathcal{P} such that for all $n \in \mathbb{N}_+$ the restriction $\tau_n(\Pi)$ is a (standard) Ξ - n -coalescent then we call Π a (standard) Ξ -coalescent.

Remark 2.2. We can use the exchangeability (2.2) to rewrite the consistency (2.1) in the following manner:

$$\lambda_{l_1, \dots, l_i} = \lambda_{l_1, \dots, l_{i,1}} + \sum_{j=1}^i \lambda_{l_{\sigma(1)}, \dots, l_{\sigma(j)+1}, \dots, l_{\sigma(i)}}. \quad (2.3)$$

Where $\sigma: [i] \rightarrow [i]$ is a permutation which reorders the parameters if necessary:

$$\sigma(l_{\sigma(1)}) \geq \dots \geq \sigma(l_{\sigma(j)} + 1) \geq \dots \geq l_{\sigma(i)}.$$

Definition 2.1 does not explain the meaning of the Ξ in the name of the Ξ -coalescent. It turns out that there are 3 major equivalent ways to represent the rates of a Ξ -coalescent.

Theorem 2.3. *Let $\{\lambda_\pi \in \mathbb{R}_+ \mid \pi \in \mathcal{P}_m \text{ nontrivial for some } m \in \mathbb{N}_+\}$ be a family of rates. Then the following statements are equivalent:*

1. *The properties (2.1) and (2.2) hold.*
2. *There exists a unique finite measure Ξ on the infinite simplex*

$$\Delta := \left\{ (x_1, x_2, \dots) \in [0, 1]^{\mathbb{N}} \mid x_1 \geq x_2 \geq \dots \geq 0, \sum_{j=1}^{\infty} x_j \leq 1 \right\}$$

such that for all $r \in \mathbb{N}_+$, $s \in \mathbb{N}_0$ and all $l_1 \geq \dots \geq l_{r+s}$ with $l_r \geq 2$ and $l_{r+1} = 1$ the rate $\lambda_{l_1, \dots, l_{r+s}}$ is given by:

$$\int_{\Delta} \frac{1}{\|x\|_2^2} \sum_{k=0}^s \sum_{i_1 \neq \dots \neq i_{r+k}} \binom{s}{k} x_{i_1}^{l_1} \cdot \dots \cdot x_{i_{r+k}}^{l_{r+k}} \cdot (1 - \|x\|_1)^{s-k} d\Xi(x). \quad (2.4)$$

Note that we used the following notations for $x = (x_1, x_2, \dots) \in \Delta$:

$$\|x\|_2^2 := \sum_{j=1}^{\infty} x_j^2, \quad \|x\|_1 := \sum_{j=1}^{\infty} x_j.$$

Furthermore, the sum over $i_1 \neq \dots \neq i_{r+k}$ is taken over all $(r+k)$ -tuples of mutually different indices in \mathbb{N}_+ .

3. *There exists a unique sequence $(F_r)_{r \in \mathbb{N}_+}$ such that F_r is a symmetric, finite measure on the r -simplex:*

$$\Delta_r := \left\{ (x_1, \dots, x_r) \in [0, 1]^r \mid \sum_{j=1}^r x_j \leq 1 \right\},$$

we have $F_1(\Delta_1) \geq F_2(\Delta_2) \geq \dots$ and for all $r \in \mathbb{N}_+$, $s \in \mathbb{N}_0$ and all $l_1 \geq \dots \geq l_{r+s}$ with $l_r \geq 2$ and $l_{r+1} = 1$ the rate $\lambda_{l_1, \dots, l_{r+s}}$ is given by:

$$\sum_{k=r}^{\lfloor r+s/2 \rfloor} \int_{\Delta_k} x_1^{l_1-2} \cdot \dots \cdot x_r^{l_r-2} T_{r,s}^{(k)}(x_1, \dots, x_k) dF_k(x_1, \dots, x_k) \geq 0, \quad (2.5)$$

where $T_{r,s}^{(k)}(x_1, \dots, x_k)$ are polynomials given by the following formulas:

$$T_{k,s}^{(k)}(x_1, \dots, x_k) := \left(1 - \sum_{j=1}^k x_j \right)^s$$

and for $i \in [k]$:

$$T_{k-i,s}^{(k)}(x_1, \dots, x_k) := (-1)^{i+1} \sum_{k_i=2i-1}^{k_{i+1}-2} \dots \sum_{k_1=1}^{k_2-2} \prod_{p=0}^i k_p \left(1 - \sum_{j=1}^{k-p} x_j \right)^{k_{p+1}-k_p-2}$$

where we set $k_0 := -1$ and $k_{i+1} := s + 1$.

Proof. The equivalence of representations follows from Lemma 18 (for consistency), Theorem 2 (for the representation with Ξ) and Proposition 8 (for the representation with $(F_r)_{r \in \mathbb{N}_+}$) in [27]. Uniqueness of Ξ follows from Proposition 4 in [27]. Uniqueness of $(F_r)_{r \in \mathbb{N}_+}$ follows from Proposition 8 in [27]. \square

Remark 2.4. Note that we defined the infinite simplex Δ to only contain decreasing sequences. This choice is required to ensure the uniqueness of Ξ in Theorem 2.3.

Furthermore, in order for the sequence of symmetric measures $(F_r)_{r \in \mathbb{N}_+}$ to define a Ξ -coalescent, we require (2.5) to only yield nonnegative numbers. This is noteworthy since it can be hard to check. This issue does not arise when working with Ξ since (2.4) by definition is always nonnegative.

It should be noted that a Ξ - n -coalescent only requires makes use of rates λ_π with $\pi \in \mathcal{P}_m$, $m \leq n$ (see Definition 2.1). But in order to identify a Markov chain as a Ξ - n -coalescent it does not suffice to just check consistency and exchangeability of the rates for $m \leq n$ since the system may not be extendable to larger n and thus not be representable by a measure Ξ in the sense of Theorem 2.3.

Example 2.5. This is an example for a consistent collection of rates $(\lambda_{l_1, \dots, l_i})$ for sample size $n \leq 4$ which can not be extended to $n = 5$. We define:

$$\lambda_2 = 2, \quad \lambda_{2,1} = \lambda_3 = 1, \quad \lambda_4 = \lambda_{2,2} = \lambda_{2,1,1} = 0, \quad \lambda_{3,1} = 1.$$

It is easy to check that these rates are consistent but if we try to extend the system to $n = 5$ in a consistent manner the following equations have to hold:

$$\begin{aligned} 0 &= \lambda_4 = \lambda_5 + \lambda_{4,1}, \\ 0 &= \lambda_{2,2} = 2\lambda_{3,2} + \lambda_{2,2,1}, \\ 0 &= \lambda_{2,1,1} = \lambda_{3,1,1} + 2\lambda_{2,2,1} + \lambda_{2,1,1,1}. \end{aligned}$$

Since all summands are nonnegative all rates appearing on the right hand side of these equations would have to be zero. But if the system would be consistent we would also have

$$1 = \lambda_{3,1} = \lambda_{4,1} + \lambda_{3,2} + \lambda_{3,1,1} = 0$$

which yields a contradiction. We will see later in Proposition 3.4 that for our purposes the system always allows for the choice of a Ξ , even if we restrict the sample size of our setting.

Definition 2.6 (The Λ -Coalescent and Kingman-Coalescents). Let $n \in \mathbb{N}_+$.

1. A Ξ -(n -)coalescent in which no simultaneous collisions are possible ($\lambda_\pi = 0$ whenever π has at least two blocks of size greater equal 2) is called a Λ -(n -)coalescent.
2. A Ξ -(n -)coalescent in which only pairs of blocks can merge and do so with rate 1 is called a *Kingman*-(n -)coalescent.

The Kingman-coalescent as well as Λ -coalescents correspond to special choices for the finite measure Ξ .

Examples 2.7.

1. Let Λ be a finite measure on $[0, 1]$. Define Ξ a the measure on Δ induced by the inclusion

$$\iota: [0, 1] \rightarrow \Delta, \quad x_1 \mapsto (x_1, 0, 0, \dots).$$

Then (2.4) is only non-zero if $r = 1$ and simplifies to

$$\lambda_{l_1, 1, 1, \dots, 1} = \int_0^1 x^{-2} x^{l_1} (1-x)^s d\Lambda(x).$$

Thus, given n blocks any collection of k blocks merges independently with rate

$$\lambda_k^n := \int_0^1 x^{k-2} (1-x)^{n-k} d\Lambda(x).$$

Alternatively we can define $F_1 := \Lambda$ and $F_r := 0$ for $r \geq 2$. In this case (2.5) is only non-zero if $r = 1$ and simplifies to

$$\lambda_{l_1, 1, 1, \dots, 1} = \int_0^1 x^{l_1-2} T_{1,s}^{(1)} d\Lambda(x)$$

which also gives the rates of a Λ -coalescent.

Now assume that Ξ is not supported on $\iota([0, 1]) \subseteq \Delta$. Let $r = 2$, $l_1 = 2$, $l_2 = 2$ and $s = 0$. Using (2.4) and since the function $x \mapsto \sum_{i_1 \neq i_2} x_{i_1}^2 x_{i_2}^2$ is positive on $\Delta \setminus \iota([0, 1])$, we get

$$\lambda_{2,2} = \int_{\Delta} \frac{1}{\|x\|_2^2} \sum_{i_1 \neq i_2} x_{i_1}^2 x_{i_2}^2 d\Xi(x) > 0.$$

Thus Ξ allows for simultaneous mergers and therefore can not define a Λ -coalescent.

2. Consider the finite measure $\Xi = a\delta_0$ where $0 = (0, 0, \dots) \in \Delta$ and $a > 0$. Then (2.4) is only non-zero if $l_1 = 2$ and $r = 1$ and we get $\lambda_{2,1,\dots,1} = a$ for all $s \in \mathbb{N}_+$. Thus we only see pairwise mergers and any pair of blocks in the coalescent merges independently with rate a .

Alternatively we can choose $F_1 = a\delta_0$ and $F_r = 0$ for all $r \geq 1$. In this case (2.5) is only non-zero if $l_1 = 2$ and $r = 1$ and yields $\lambda_{2,1,\dots,1} = a$ for all $s \in \mathbb{N}_+$. Thus this choice of Ξ yields a Kingman-coalescent sped up by a .

Now assume that Ξ is not supported on 0 . Let $r = 1$, $l_1 = 3$ and $s = 0$. Using (2.4) and since the function $x \mapsto \sum_{i=1}^{\infty} x_i^3$ is positive on $\Delta \setminus \{0\}$, we get

$$\lambda_3 = \int_{\Delta} \frac{1}{\|x\|_2^2} \sum_{i=1}^{\infty} x_i^3 d\Xi(x) > 0.$$

Thus Ξ allows for multiple mergers and therefore can not define a (sped up) Kingman-coalescent.

The representation of the Ξ -coalescent using the measure Ξ allows for a specific construction of the process Π using Poisson point processes. This construction also gives an interpretation for (2.4). We present this construction for the case $\Xi(\{0\}) = 0$ and in an informal way. For a rigorous construction in the general case see Schweinsberg [27] Section 3. We start by considering a Poisson point process η on $\mathbb{R}_+ \times \Delta$ with intensity measure given by $\|x\|_2^{-2} dt d\Xi(x)$. For each atom (x, t) of η we choose an i.i.d. sequence $(Y_k)_{k \in \mathbb{N}_+}$, independent from η , of \mathbb{N}_0 -valued random variables with $\mathbb{P}(Y_1 = m) = x_m$ where we set

$x_0 := 1 - \sum_{j=1}^{\infty} x_j$. We now construct a Ξ - n -coalescent Π using the following recursion. Let (x, t) be an atom of η and Π be already defined up to (but not including) $t \in \mathbb{R}_+$. Consider the event $\Pi_{t-} = \pi \in \mathcal{P}_n$. Then we color the k -th block of π with the color Y_k if $Y_k \geq 1$ or not at all if $Y_k = 0$. Afterwards we merge all blocks with the same color in order to define Π_t . To see that this procedure indeed yields the rates given in (2.4) let $\pi \in \mathcal{P}_n$ with block sizes $l_1 \geq l_2 \geq \dots \geq l_{r+s}$. We consider the event that the block in π belonging to l_1 gets colored with i_1 , the block belonging to l_2 gets colored with i_2 and so on up to the block belonging to l_r corresponding to the color i_r . Furthermore we assume that k additional colors i_{r+1}, \dots, i_{r+k} only show up exactly once and $s - k$ many blocks were not colored at all. There are $\binom{s}{k}$ possibilities to color or not color the remaining s blocks in π in this fashion. The probability of the event that the i.i.d. sequence $(Y_k)_{k \in \mathbb{N}_+}$ gives us a π -collision with these colors is

$$\binom{s}{k} x_{i_1}^{l_1} \cdot \dots \cdot x_{i_r}^{l_r} \cdot x_{i_{r+1}} \cdot \dots \cdot x_{i_{r+k}} (1 - \|x\|_1)^{s-k}.$$

Now note that the choice of k and of the actual colors does not matter for the merging which yields the sums in the integrand of (2.4). By the Coloring Theorem for Poisson point processes it follows that the rate with which we see a corresponding merger in Π is given by (2.4).

Chapter 3

Main Result: Convergence to the Limiting Coalescent

3.1 Assumptions

Our goal in this chapter is to show that the coalescent Π^N of our population model converges given the proper time rescaling and certain assumptions to a (potentially time-inhomogeneous) spatial Ξ -coalescent in the large population limit. Before we list our assumptions we have to make some definitions.

Definition 3.1 (The Mass Process and the Flow Process). We assign a mass of $1/N$ to each individual in the population. Define the process $\mathcal{R}^N := \mathcal{N}^N/N$. We call \mathcal{R}^N the *mass process*. Note that for $x \in G$ and $k \in \mathbb{N}_0$ the number $\mathcal{R}_{k;x}^N \in [0, 1]$ is the total mass at site x in time k . For $x, y \in G$, $x \neq y$, $k \in \mathbb{N}$ define

$$\mathcal{F}_{k;x,y}^N := \frac{1}{N} \sum_{l=1}^k \mathcal{M}_{l;x,y}^N$$

and set $\mathcal{F}_k^N := (\mathcal{F}_{k;x,y}^N)_{x,y \in G}$ as well as $\mathcal{F}^N := (\mathcal{F}_k^N)_{k \in \mathbb{N}}$. We call \mathcal{F}^N the (*cumulative*) *flow process*. Note that $\mathcal{F}_{k;x,y}^N$ is the total amount of mass that has flown from x to y up to time k .

Definition 3.2.

1. For $m \in [N]$, $x \in G$ and $k \in \mathbb{N}$ we define

$$c_x^m := \frac{\text{Var}(\nu_{k;x,i}^I)}{m-1} \tag{3.1}$$

for $i \in I \subseteq [N]$ with $|I| = m$. Note that by definition of the offspring laws the right-hand side indeed only depends on x and m .

2. For $p, q \in \mathbb{N}_0$ we use the notation: $(p)_q := p!/(p-q)!$. Let $\pi \in \mathcal{P}_n$ be a nontrivial partition with j blocks of sizes $l_1, l_2, \dots, l_j > 0$. Note that π nontrivial implies that there is a $i \in [j]$ with $l_i \geq 2$. We define (whenever the limit exists)

$$\phi_x(\pi) := \phi_{x,j}(l_1, \dots, l_j) := \lim_{m \rightarrow \infty} \frac{\mathbb{E}((\nu_{k;x,1}^{[m]})_{l_1} \cdots (\nu_{k;x,j}^{[m]})_{l_j})}{m^{l_1 + \dots + l_j - j} c_x^m}. \quad (3.2)$$

As in the nonspatial case c_x^m has an important meaning for the coalescence.

Proposition 3.3. *The constant c_x^m is the probability for two given ancestral lines at x at time $k-1$, after the migration step, to coalesce at time k if there are m individuals present at x .*

Proof. Assume that $I \subseteq \mathbb{N}$ with $|I| = m$ is the set of indices of individuals present at site x and time $k-1$ after the migration step. Fix two ancestral lines $r, s \in [n]$ at site x after the migration step. We first note that by exchangeability:

$$m = \mathbb{E} \left(\sum_{i \in I} \nu_{k;x,i}^I \right) = \sum_{i \in I} \mathbb{E} (\nu_{k;x,i}^I) = m \mathbb{E} (\nu_{k;x,1}^I)$$

Thus we have $\mathbb{E}(\nu_{k;x,1}^I) = 1$ and therefore

$$\begin{aligned} & \mathbb{P}(\text{Lines } r \text{ and } s \text{ merge in the previous generation}) \\ &= \sum_{i \in I} \mathbb{P}(\text{Lines } r \text{ and } s \text{ have the same parent } i \in I) \\ &= \sum_{i \in I} \mathbb{E} \left(\frac{\nu_{k;x,i}^I}{m} \cdot \frac{\nu_{k;x,i}^I - 1}{m-1} \right) = \frac{\mathbb{E}((\nu_{k;x,1}^I)^2) - 1}{m-1} = \frac{\text{Var}(\nu_{k;x,1}^I)}{m-1} \end{aligned}$$

□

The expressions $\phi_{x,i}(l_1, \dots, l_i)$ will later be connected to the event that l_j ancestral lines for each $j \in [i]$ in a sample of size $n = l_1 + \dots + l_i$ merge simultaneously. Therefore we expect a consistency property to hold. The first two claims of following Proposition are results already provided by Möhle and Sagitov [20].

Proposition 3.4. *Consider the situation of 2. of Definition 3.2 and let $x \in G$.*

1. *We have for all $j \leq i \in \mathbb{N}_+$ and all $m_1 \geq \dots \geq m_j \in \mathbb{N}_+$ as well as $l_1 \geq \dots \geq l_i \in \mathbb{N}_+$ with $l_1 \geq m_1, \dots, l_j \geq m_j$ and $m_1 \geq 2$:*

$$\phi_{x,i}(l_1, \dots, l_i) \leq \phi_{x,j}(m_1, \dots, m_j). \quad (3.3)$$

The inequality even holds if we use \limsup instead of \lim in (3.2). We have in particular

$$\phi_{x,i}(l_1, \dots, l_i) \leq \phi_{x,1}(2) = 1$$

and thus the sequences appearing on the right-hand side of (3.2) are always bounded.

2. *Let $i \in \mathbb{N}_+$, $l_1 \geq \dots \geq l_i \in \mathbb{N}_+$ with $l_1 \geq 2$. If existence of the limit in (3.2) is known for all but one term in the following equation then the limit for the remaining term also exists and the equation holds:*

$$\phi_{x,i}(l_1, \dots, l_i) = \phi_{x,i+1}(l_1, \dots, l_i, 1) + \sum_{j=1}^i \phi_{x,i}(l_1, \dots, l_j + 1, \dots, l_i). \quad (3.4)$$

3. *There exists a finite measure Ξ_x on the infinite simplex Δ such that for all $i \in \mathbb{N}_+$ and all $l_1 \geq \dots \geq l_i \in \mathbb{N}_+$, $l_1 \geq 2$ the limit $\phi_{x,i}(l_1, \dots, l_i)$ is given by (2.4) whenever it exists. If all the limits exist, then Ξ_x is unique.*
4. *Let $\phi_{x,2}(2, 2) = 0$ and assume that the limits $\phi_{x,1}(k)$ exist for all $k \geq 2$. Then all limits ϕ exist and Ξ_x corresponds to a Λ -coalescent.*
5. *Let $\phi_{x,1}(3) = 0$. Then all limits ϕ exist and Ξ_x corresponds to a Kingman-coalescent.*

Proof. We first note that (3.3) corresponds to (18) in Möhle and Sagitov [20] and (3.4) to Lemma 3.3 in Möhle and Sagitov [20]. Note that with the $\psi_{j,s}$ in the notation of [20] we have with our notation for ϕ (we omit the spatial index x):

$$\psi_{i,s}(l_1, \dots, l_i) = \phi_{i+s}(l_1, \dots, l_i, 1, \dots, 1)$$

with $l_1, \dots, l_i \geq 2$. We still provide a proof for completeness' sake and due to these notational differences.

We start with some preliminary calculations. Let $x \in G$, $m \in \mathbb{N}_+$, $I = [m]$, $k \in \mathbb{N}_0$, $i \in \mathbb{N}_+$ and $l_1 \geq \dots \geq l_i$ with $l_1 \geq 2$. Moreover, let $n := l_1 + \dots + l_i \leq m - 1$. We have:

$$\begin{aligned}
& \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1 + \dots + l_i + 1 - (i+1)} c_x^m} \\
& + \sum_{j=1}^i \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,j}^I)_{l_j+1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i})}{m^{l_1 + \dots + l_j + 1 + \dots + l_i - i} c_x^m} \\
& = \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1 + \dots + l_i - i} c_x^m} \\
& + \sum_{j=1}^i \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot (\nu_{k;x,j}^I - l_j))}{m^{l_1 + \dots + l_i + 1 - i} c_x^m} \\
& = \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1 + \dots + l_i - i} c_x^m} \\
& + \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot (\sum_{j=1}^i (\nu_{k;x,j}^I - l_j)))}{m^{l_1 + \dots + l_i + 1 - i} c_x^m} \\
& \stackrel{\star}{=} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1 + \dots + l_i - i} c_x^m} \\
& + \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot (m - n - \sum_{j=i+1}^m (\nu_{k;x,j}^I)))}{m^{l_1 + \dots + l_i + 1 - i} c_x^m} \\
& \stackrel{\star\star}{=} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1 + \dots + l_i - i} c_x^m} \\
& + \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot (m - n - (m - i) \nu_{k;x,i+1}^I))}{m^{l_1 + \dots + l_i + 1 - i} c_x^m} \\
& = \frac{m - n}{m} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i})}{m^{l_1 + \dots + l_i - i} c_x^m} \\
& + \frac{i}{m} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdot \dots \cdot (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1 + \dots + l_i - i} c_x^m}.
\end{aligned}$$

Here \star follows from the the fact that $\sum_{i \in I} \nu_{k;x,i}^I = |I| = m$ and $\star\star$ from the exchangeability of $(\nu_{k;x,i}^I)_{i \in I}$. Rearranging terms and multiplying both sides

with $m/(m-n)$ yields

$$\begin{aligned}
& \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdots (\nu_{k;x,i}^I)_{l_i})}{m^{l_1+\cdots+l_i-i} c_x^m} \\
&= \frac{m-i}{m-n} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdots (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1+\cdots+l_i+1-(i+1)} c_x^m} \\
&+ \frac{m}{m-n} \sum_{j=1}^i \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdots (\nu_{k;x,j}^I)_{l_j+1} \cdots (\nu_{k;x,i}^I)_{l_i})}{m^{l_1+\cdots+l_j+1+\cdots+l_i-i} c_x^m}. \tag{3.5}
\end{aligned}$$

Now we can start showing the claims.

1. Since all terms in (3.5) are nonnegative and applying the $\limsup_{m \rightarrow \infty}$ we get the inequalities

$$\begin{aligned}
& \limsup_{m \rightarrow \infty} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdots (\nu_{k;x,j}^I)_{l_j+1} \cdots (\nu_{k;x,i}^I)_{l_i})}{m^{l_1+\cdots+l_j+1+\cdots+l_i-i} c_x^m} \\
&\leq \limsup_{m \rightarrow \infty} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdots (\nu_{k;x,i}^I)_{l_i})}{m^{l_1+\cdots+l_i-i} c_x^m}
\end{aligned}$$

and

$$\begin{aligned}
& \limsup_{m \rightarrow \infty} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdots (\nu_{k;x,i}^I)_{l_i} \cdot \nu_{k;x,i+1}^I)}{m^{l_1+\cdots+l_i+1-(i+1)} c_x^m} \\
&\leq \limsup_{m \rightarrow \infty} \frac{\mathbb{E}((\nu_{k;x,1}^I)_{l_1} \cdots (\nu_{k;x,i}^I)_{l_i})}{m^{l_1+\cdots+l_i-i} c_x^m}.
\end{aligned}$$

This corresponds to the inequalities (where we use \limsup instead of \lim in the definition of ϕ):

$$\phi_{x,i}(l_1, \dots, l_j+1, \dots, l_i) \leq \phi_{x,i}(l_1, \dots, l_i)$$

and

$$\phi_{x,i}(l_1, \dots, l_i, 1) \leq \phi_{x,i}(l_1, \dots, l_i).$$

The inequality (3.3) now follows from successive application of these inequalities. In particular we get due to $l_1 \geq 2$, $\mathbb{E}(\nu_{k,x,i}^I) = 1$ and (3.1):

$$\begin{aligned}
\phi_{x,i}(l_1, \dots, l_i) &\leq \phi_{x,1}(2) = \lim_{m \rightarrow \infty} \frac{\mathbb{E}(\nu_{k,x,i}^I (\nu_{k,x,i}^I - 1))}{m c_x^m} \\
&= \lim_{m \rightarrow \infty} \frac{\text{Var}(\nu_{k,x,i}^I)}{m c_x^m} \\
&= \lim_{m \rightarrow \infty} \frac{m-1}{m} = 1.
\end{aligned}$$

2. Assume that convergence for all but one summand in equation (3.5) is known. Solving for the remaining summand and letting $m \rightarrow \infty$ shows convergence of the remaining summand and we get (3.4) since all the limits exist.
3. If all the limits $\lambda_{l_1, \dots, l_i} := \phi_{x,i}(l_1, \dots, l_i)$ exist they yield a family of rates fulfilling (2.1) and (2.2). By Theorem 2.3 there exists a unique measure Ξ_x on Δ such that $\phi_{x,i}(l_1, \dots, l_i)$ is given by (2.4). Now we omit the assumption on the convergence of the $\phi_{x,i}(l_1, \dots, l_i)$. By the first part of this proposition we know that the sequences defining $\phi_{x,i}(l_1, \dots, l_i)$ have a finite limit superior. Thus by a diagonal scheme we can find a subsequence such that all limits $\phi_{x,i}(l_1, \dots, l_i)$ exist for that subsequence. Thus we can apply the previous result and again get a finite measure Ξ_x on Δ such that $\phi_{x,i}(l_1, \dots, l_i)$ is given by (2.4). In particular all the limits $\phi_{x,i}(l_1, \dots, l_i)$ which were already existing before the choice of the subsequence are represented by Ξ_x .
4. The inequality (3.3) yields $\phi_{x,i}(l_1, \dots, l_i) = 0$ if there are $j_1 \neq j_2 \in [i]$ with $l_{j_1} \geq 2$ and $l_{j_2} \geq 2$. The consistency (3.4) simplifies to

$$\phi_{x,i+1}(k, 1, \dots, 1) = \phi_{x,i}(k, 1, \dots, 1) - \phi_{x,i}(k+1, 1, \dots, 1)$$

for all $i \in \mathbb{N}_+$, $k \geq 2$. By induction over the number of ones in the argument of $\phi_{x,i+1}(k, 1, \dots, 1)$ we can show that since the limit $\phi_{x,1}(k)$ exists for all $k \geq 2$ all limits $\phi_{x,i+1}(k, 1, \dots, 1)$ have to exist as well. Moreover, by 1. of Examples 2.7 $\phi_{x,2}(2, 2) = 0$ is equivalent to Ξ_x defining a Λ_x -coalescent.

5. We note that by 2. of Examples 2.7 the only possible measure Ξ_x which would yield $\phi_{x,1}(3) = 0$ is a sped up Kingman-coalescent. We have shown in the first claim of this proposition that $\phi_{x,1}(2) = 1$, thus $\Xi_x = \delta_0$ is uniquely determined. This implies that all convergent subsequences considered in the proof of 3 of this proposition have the same limit. Since the sequences are also bounded they have to converge even without the choice of a subsequence. As already noted, the limit has to correspond to a Kingman-coalescent.

□

Remark 3.5.

1. Möhle and Sagitov [20] and Schweinsberg [27] considered a situation in which all the limits defining ϕ exist. Our approach will generalize this a little bit by only requiring the limits to exist up to a certain sample size n . However, Example 2.5 shows that it would theoretically be possible to observe a consistent system for a small sample size, that can not be extended to larger sample sizes and therefore does not come from a finite measure Ξ on Δ . 3. of Proposition 3.4 ensures that this can not happen for our limiting coalescent. The rates given by ϕ always come from a measure Ξ even if the limits only exist for small sample sizes.
2. Note that (3.4) yields the recursion

$$\phi_{x,i+1}(l_1, \dots, l_i, 1) = \phi_{x,i}(l_1, \dots, l_i) - \sum_{j=1}^i \phi_{x,i}(l_1, \dots, l_j + 1, \dots, l_i). \quad (3.6)$$

By successive application of this recursion we can get rid of all the parameters $l_i = 1$ without increasing the sample size considered. In particular it suffices to show existence of the limits $\phi_{x,i}(l_1, \dots, l_i)$ for all $i \in \mathbb{N}_+$, $l_1, \dots, l_i \geq 2$ with $\sum_{j=1}^i l_j \leq n$ in order to conclude existence of $\phi_{x,p}(l_1, \dots, l_p)$ for all $p \in \mathbb{N}$ and $l_1, \dots, l_p \geq 1$ with $\sum_{j=1}^p l_j \leq n$.

3. It is worthwhile to point out that 5. of Proposition 3.4 did not require $\phi_{x,2}(2, 2) = 0$. If we had $\phi_{x,2}(2, 2) = 0$ then the consistency (3.4) would suffice to show that $\phi_{x,i}(2, 1, \dots, 1) = 1$ for all $i \in \mathbb{N}_+$ and thus that the rates belong to a Kingman-coalescent. But without $\phi_{x,2}(2, 2) = 0$ consistency is insufficient. In fact consider for $n = 4$:

$$\lambda_2 = 1, \quad \lambda_3 = 0, \lambda_{2,1} = 1, \quad \lambda_4 = 0, \lambda_{3,1} = 0, \lambda_{2,2} = 1/2, \lambda_{2,1,1} = 1/2.$$

It is a simple calculation to check that this system is consistent but allows for $\lambda_3 = 0$ and $\lambda_{2,2} > 0$. This shows that the existence of a Ξ -measure extending a given system of rates ϕ as shown in 3. of Proposition 3.4 can provide us with some additional information which is not already implied by 1. and 2. of Proposition.

We can now state our assumptions.

Assumptions 3.6. All the weak convergence assumptions are actually for the joint distribution of all appearing sequences.

1. We assume that there exists an $\alpha > 0$ such that we have for $x \in G$ and for $m \rightarrow \infty$:

$$c_x^m \cdot m^\alpha \rightarrow \beta_x \quad \text{with } \beta_x \in \mathbb{R}_+.$$

2. We assume that the limit defining $\phi_{x,i}(l_1, \dots, l_i) < \infty$ exists for all $i \in \mathbb{N}_+$, $l_1, \dots, l_i \in \mathbb{N}_+$ with $l_j \geq 2$ for all $j \in [i]$ and $l_1 + \dots + l_i \leq n$, $x \in G$.

3. For $t \in \mathbb{R}_+$ and $N \in \mathbb{N}_+$ set $k_t^N := \lfloor t \cdot N^\alpha \rfloor$. We have weak convergence of the finite dimensional distributions of the flow processes

$$(\mathcal{F}_{k_t^N}^N)_{t \in \mathbb{R}_+} \rightarrow (\mathcal{F}_t)_{t \in \mathbb{R}_+} \quad \text{for } N \rightarrow \infty.$$

We call the process \mathcal{F} the *limiting, (cumulative) flow process*.

Let $x \neq y \in G$. Furthermore we assume that the processes $\mathcal{F}_{x,y}$ have almost surely absolutely continuous paths. In particular weak differentials $\mathcal{F}'_{\bullet;x,y}$ exist almost surely.

4. The paths

$$t \mapsto \sum_{x \in G} \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;x,y}$$

are almost surely integrable on compact subsets of \mathbb{R}_+ . In words: the total amount of mass moved in the system up to some finite time is finite.

5. We have weak convergence of the starting masses: $\mathcal{R}_0^N \rightarrow \mathcal{R}_0$. Furthermore the limit has to still be normalized:

$$\sum_{x \in G} \mathcal{R}_{0;x} = 1 \quad \text{almost surely.}$$

6. Let $x \in G$, then:

$$\sum_{x \in G \setminus \{y\}} \mathcal{F}_{k_t^N;x,y}^N \rightarrow \sum_{x \in G \setminus \{y\}} \mathcal{F}_{\bullet;x,y}, \quad \sum_{x \in G \setminus \{y\}} \mathcal{F}_{k_t^N;y,x}^N \rightarrow \sum_{x \in G \setminus \{y\}} \mathcal{F}_{\bullet;y,x}$$

as weak convergence of the finite dimensional distributions of the processes. In words: The total flow in and out of a given site needs to converge weakly.

7. We define for $x \in G$, $t \in \mathbb{R}_+$:

$$\mathcal{R}_{t;x} := \mathcal{R}_{0;x} + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} - \mathcal{F}_{t;x,y}.$$

We call the process \mathcal{R} the *(limiting) mass process*.

Since $k \mapsto \mathcal{F}_{k;x,y}^N$ was nondecreasing we also have that $t \mapsto \mathcal{F}_{t;x,y}$ is nondecreasing and thus measure generating. We assume for $x, y \in G$, $x \neq y$:

$$\int_0^\infty \mathbb{1}_{\{\mathcal{R}_{t;x}=0\}} \mathcal{F}_{dt;x,y} = \int_0^\infty \mathbb{1}_{\{\mathcal{R}_{t;x}=0\}} \mathcal{F}_{dt;y,x} = 0 \quad (3.7)$$

almost surely. In words: There is no flow into or out of a site which carries no mass.

Some remarks regarding the Assumptions 3.6.

Remarks 3.7.

1. Let G be finite. Then Assumptions 4 and 6 as well as the normalization of \mathcal{R}_0 in Assumption 5 are implied by the remaining assumptions.
2. The definition of $k_t^N = [t \cdot N^\alpha]$ corresponds to a polynomial global rescaling $1/N^\alpha$ which is a consequence of the limiting behavior of \mathcal{C}_x^m (see 1. of Assumptions 3.6). The polynomial behavior is not necessary in order to attain a limit. However, this choice has the advantage that it behaves particularly well when the limiting mass of sites differs (see Remark 3.15 for more details).
3. It will later turn out, that the convergence of finite dimensional distributions of the rescaled flows \mathcal{F}^N extendeds to uniform convergence on compact sets.
4. The definition of \mathcal{R}^N together with the assumptions yields a weak convergence $\mathcal{R}_x^N \rightarrow \mathcal{R}_x$ first in terms of finite dimensional distributions and then in terms of uniform convergence on compact sets.
5. We can pose an even weaker condition on the convergence of the flows \mathcal{F}^N . It suffices that only a dense subset of the set $\{t \geq 0 \mid \mathcal{R}_{t;x} > 0\}$ needs to be taken into account for the convergence of the finite dimensional distributions of $\mathcal{F}_{k_t^N;x,y}^N$.
6. The Equation (3.7) can be rewritten as $\mathcal{F}'_{t;x,y} = \mathcal{F}'_{t;y,x} = 0$ for Lebesgue almost all t with $\mathcal{R}_{t;x} = 0$.

3.2 Examples of Models Meeting the Assumptions

In this section we will present examples for which the assumptions in 3.6 hold. First, we note that the reproduction mechanism defines our time rescaling but that the migration and coalescence mechanisms can otherwise be considered separately.

As we already noted in Examples 1.6: one way of modeling migration of individuals in the population model is to consider independent random walks on G . For the backwards-in-time spatial coalescent process we can just reverse those random walks to get new random walks which describe the migration of individuals backwards in time (special case of Remark 1.9).

Thus, for $N \in \mathbb{N}_+$ we consider a family of i.i.d. random walks $(X^{i,N})_{i \in [N]}$ on G with transition matrix $P^N = (P^N(x, y))_{x, y \in G}$. We assume that the random walks are irreducible and positively recurrent with stationary distribution $\pi^N = (\pi_x^N)_{x \in G}$. Furthermore we assume that the random walks are in equilibrium.

In this setting \mathcal{M}^N and \mathcal{N}_0^N are given in the following way:

$$\mathcal{N}_{0;x}^N := \sum_{i \in [N]} \mathbb{1}_{\{X_0^{i,N} = x\}}$$

and for $x, y \in G$ with $x \neq y$ and $k \in \mathbb{N}_0$:

$$\mathcal{M}_{k;x,y}^M := \sum_{i \in [N]} \mathbb{1}_{\{X_k^{i,N} = x, X_{k+1}^{i,N} = y\}}.$$

Proposition 3.8 (Migration Via Independent Random Walkers).

Set $c^N := N^{-\alpha}$ and $k_t^N := \lceil t/c^N \rceil$. Let

$$\frac{P^N - I}{c^N} \rightarrow Q, \quad \pi^N \rightarrow \pi \quad (3.8)$$

as entrywise convergence of real numbers, with $Q = (Q(x, y))_{x, y \in G}$ a stable (i.e., finite diagonal entries), conservative (i.e. row sums are zero) generator matrix (i.e. nonnegative off diagonal entries) with $\sum_{x \in G} -Q(x, x)\pi_x < \infty$ where π is a distribution on G (represented as a row vector) with $\pi \cdot Q = 0$.

Then we have for all $t \in \mathbb{R}_+$ and $x, y \in G$:

$$\mathcal{R}_{t;x}^N \rightarrow \pi_x \quad \text{and} \quad \mathcal{F}_{k_t^N;x,y}^N \rightarrow t\pi_x Q(x, y) \quad (3.9)$$

almost surely as $N \rightarrow \infty$. Define $\mathcal{R}_{t;x} := \pi_x$ and $\mathcal{F}_{t;x,y} := t\pi_x Q(x, y)$ for all $t \in \mathbb{R}_+$, $x \neq y \in G$. Then 1. and 3. to 7. of Assumptions 3.6 are fulfilled. We even achieve almost sure pointwise convergence instead of just convergence of the finite dimensional distributions of the considered processes.

Proof. We first show the convergence of $\mathcal{R}_{t;x}^N$ for $x \in G$, $t \geq 0$. Note that by definition

$$\mathcal{R}_{t;x}^N = \frac{\mathcal{N}_{t;x}^N}{N} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}_{\{X_t^{i,N}=x\}}$$

The random variables $\mathbb{1}_{\{X_t^{i,N}=x\}}$ are mutually independent $\text{Bin}(1, \pi_x^N)$ distributed for $i \in [N]$ since the random walks are in equilibrium. We consider the array $(E_{N,i})_{N \in \mathbb{N}, i \in [N]}$ given by

$$E_{N,i} := \mathbb{1}_{\{X_t^{i,N}=x\}} - \pi_x^N.$$

Then the array is row-wise independent with $\mathbb{E}(E_{N,i}) = 0$ and $|E_{N,i}| \leq 1$. Thus we have by a strong law of large numbers for such arrays (see Theorem 4 in [28]) that

$$\mathcal{R}_{t;x}^N - \pi_x^N = \frac{1}{N} \sum_{i=1}^N E_{N,i} \rightarrow 0 \quad \text{for } N \rightarrow \infty \text{ almost surely.}$$

Together with $\pi_x^N \rightarrow \pi_x$ for $N \rightarrow \infty$ this shows the first part of (3.9).

Next we show the convergence of $\mathcal{F}_{k_t^N;x,y}^N$ for $x, y \in G$ and $t \in \mathbb{R}_+$. We have

$$\mathcal{F}_{k_t^N;x,y}^N = \frac{\sum_{j=1}^{k_t^N} \mathcal{M}_{j;x,y}^N}{N} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^{k_t^N} \mathbb{1}_{\{X_{j-1}^{i,N}=x, X_j^{i,N}=y\}}.$$

It is noteworthy, that heuristically, due to the ergodic theorem for Markov chains, the inner sum will behave asymptotically like $k_t^N \pi_x^N P^N(x, y)$ which would then converge to $t \pi_x Q(x, y)$ for $N \rightarrow \infty$. The issue with this approach is that we cannot apply the ergodic theorem since the transition matrix P^N and the stationary distribution π^N depend on N . Even worse, for large N the Markov chain X^N will move slower and therefore we cannot expect that the ergodic theorem could be applied uniformly in N . Alternatively we could try to apply the strong law for row-wise independent arrays again. Consider the array $(Z_{N,i})_{N \in \mathbb{N}, i \in [N]}$ given by

$$Z_{N,i} := \sum_{j=1}^{k_t^N} \left(\mathbb{1}_{\{X_{j-1}^{i,N}=x, X_j^{i,N}=y\}} - \pi_x^N P^N(x, y) \right).$$

Due to independence of $X^{i,N}$ the rows of $(Z_{N,i})$ are again independent and since the chains are in equilibrium:

$$\mathbb{P}(X_{j-1}^{i,N} = x, X_j^{i,N} = y) = \pi_x^N P^N(x, y) \quad (3.10)$$

and thus $\mathbb{E}(Z_{N,i}) = 0$ for all $N \in \mathbb{N}$, $i \in [N]$. But we can no longer guarantee the uniform stochastic boundedness of $Z_{N,i}$ which was necessary to apply Theorem 4 in [28].

Thus, we have to estimate the probabilities explicitly and show convergence from scratch. To ease notation we define for $j \in \{1, \dots, k_t^N\}$:

$$A_{j;x,y}^{i,N} := \{X_{j-1}^{i,N} = x, X_j^{i,N} = y\}.$$

Note that if $j_1, \dots, j_m \in \{1, \dots, k_t^N\}$ are mutually different the event $\bigcap_{l=1}^m A_{j_l;x,y}^{i,N}$ entails that the Markov chain $X^{i,N}$ jumps from x to y on at least m different occasions and thus

$$\mathbb{P} \left(\bigcap_{l=1}^m A_{j_l;x,y}^{i,N} \right) \leq P^N(x, y)^m.$$

We also define for $j_1, \dots, j_m \in \{1, \dots, k_t^N\}$

$$Y_{j_1, \dots, j_m; x, y}^{i, N} := \prod_{l=1}^m \left(\mathbb{1}_{A_{j_l; x, y}^{i, N}} - \pi_x^N P^N(x, y) \right).$$

Moreover, if we expand this expression we get summands of the form

$$\pm \left(\prod_{l \in I} \mathbb{1}_{A_{j_l; x, y}^{i, N}} \right) (\pi_x^N P^N(x, y))^k$$

with $k \in [m]$ and $I \subseteq [m]$ a subset with $|I| = m - k$. We first consider the case, that the j_l are mutually different. Taking expectations and applying the above estimation yields that each summand is bound by $P^N(x, y)^{|I|} P^N(x, y)^k = P^N(x, y)^m$. Since there are m^2 summands we get

$$\mathbb{E}(|Y_{j_1, \dots, j_m; x, y}^{i, N}|) \leq m^2 P^N(x, y)^m.$$

In the case that the summands are not mutually different we assume $j_1 = j_m$ without loss of generality. Then, since

$$\begin{aligned} & |\mathbb{1}_{\{A_{j_1; x, y}^{i, N}\}} - \pi_x^N P^N(x, y)| \cdot |\mathbb{1}_{\{A_{j_m; x, y}^{i, N}\}} - \pi_x^N P^N(x, y)| \\ &= |\mathbb{1}_{\{A_{j_1; x, y}^{i, N}\}} - \pi_x^N P^N(x, y)|^2 \\ &\leq |\mathbb{1}_{\{A_{j_1; x, y}^{i, N}\}} - \pi_x^N P^N(x, y)| \end{aligned}$$

(note that the expression inside the square is a random variable taking values in $[0, 1]$) and thus

$$|Y_{j_1, \dots, j_m; x, y}^{i, N}| \leq |Y_{j_1, \dots, j_{m-1}; x, y}^{i, N}|$$

we get $\mathbb{E}(|Y_{j_1, \dots, j_m}^{i, N}|) \leq (m-1)^2 P^N(x, y)^{m-1}$. Applying this consecutively until there are only mutually different indices we get

$$\mathbb{E}(|Y_{j_1, \dots, j_m}^{i, N}|) \leq a^2 P^N(x, y)^a.$$

where $a \in [m]$ is the number of mutually different j_l .

We will now estimate the L^4 distance between $\mathcal{F}_{k_t^N; x, y}^N$ and $k_t^N \pi_x^N P^N(x, y)$. We have using that the $Z_{N, i}$ are i.i.d. in $i \in [N]$ and have mean zero. We use this to expand the 4th moment:

$$\begin{aligned} & \mathbb{E} \left(\left(\mathcal{F}_{k_t^N; x, y}^N - k_t^N \pi_x^N P^N(x, y) \right)^4 \right) \\ &= \mathbb{E} \left(\left(\frac{1}{N} \sum_{i=1}^N Z_{N, i} \right)^4 \right) \\ &= \frac{N}{N^4} \mathbb{E}((Z_{N, 1})^4) + \frac{(N)_2}{N^4} \mathbb{E}(Z_{N, 1}) \cdot \mathbb{E}(Z_{N, 1}^3) + \frac{(N)_2}{N^4} \mathbb{E}(Z_{N, 1}^2)^2 \\ &+ \frac{(N)_3}{N^4} \mathbb{E}(Z_{N, 1})^2 \cdot \mathbb{E}(Z_{N, 1}^2) + \frac{N(N)_4}{N^4} \mathbb{E}(Z_{N, 1})^4 \\ &= \frac{N}{N^4} \mathbb{E}((Z_{N, 1})^4) + \frac{(N)_2}{N^4} \mathbb{E}(Z_{N, 1}^2)^2. \end{aligned}$$

Our goal is to show that the expectations above are bounded for $N \rightarrow \infty$ which then shows that the fourth moment goes to 0 for $N \rightarrow \infty$ with order $1/N^2$.

We have for the first expectation where the sums are always taken over mutually different indexes in $[k_t^N]$ (we use multinomial coefficients):

$$\begin{aligned} \mathbb{E}(Z_{N, 1}^4) &= \binom{4}{1, 1, 1, 1} \sum_{j_1, j_2, j_3, j_4} \mathbb{E}(Y_{j_1, j_2, j_3, j_4}^{1, N}) \\ &+ \binom{4}{2, 1, 1} \sum_{j_1, j_2, j_3} \mathbb{E}(Y_{j_1, j_2, j_3, j_3}^{1, N}) + \binom{4}{2, 2} \sum_{j_1, j_2} \mathbb{E}(Y_{j_1, j_1, j_2, j_2}^{1, N}) \\ &+ \binom{4}{3, 1} \sum_{j_1, j_2} \mathbb{E}(Y_{j_1, j_1, j_1, j_2}^{1, N}) + \binom{4}{4} \sum_{j_1} \mathbb{E}(Y_{j_1, j_1, j_1, j_1}^{1, N}) \\ &\leq \binom{4}{1, 1, 1, 1} k_t^N (k_t^N - 1)(k_t^N - 2)(k_t^N - 3) P^N(x, y)^4 \\ &+ \binom{4}{2, 1, 1} k_t^N (k_t^N - 1)(k_t^N - 2) P^N(x, y)^3 + \binom{4}{2, 2} k_t^N (k_t^N - 1) P^N(x, y)^2 \\ &+ \binom{4}{3, 1} k_t^N (k_t^N - 1) P^N(x, y)^2 + \binom{4}{4} k_t^N P^N(x, y). \end{aligned}$$

The right hand side of this inequality is bounded in $N \in \mathbb{N}_+$ due to $c^N \rightarrow 0$ and since by (3.8):

$$k_t^N P^N(x, y) = \left[\frac{t}{c^N} \right] c^N \frac{P^N(x, y)}{c^N} \rightarrow tQ(x, y) \quad \text{for } N \rightarrow \infty. \quad (3.11)$$

Thus,

$$\limsup_{N \rightarrow \infty} \mathbb{E} \left(\left(\sum_{j=1}^{k_t^N} \mathbb{1}_{A_{j;x,y}^{1,N}} - \pi_x^N P^N(x, y) \right)^4 \right) < \infty.$$

Next, we have to estimate the second moment. We get similar to the fourth moment (again the sums are taken over mutually different indices):

$$\begin{aligned} & \mathbb{E} (Z_{N,1}^2) \\ &= \binom{2}{1} \sum_{j_1, j_2} \mathbb{E} (Y_{j_1, j_2; x, y}^{1, N}) + \binom{2}{2} \sum_{j_1} \mathbb{E} (Y_{j_1, j_1; x, y}^{1, N}) \\ &\leq 2k_t^N (k_t^N - 1) P^N(x, y)^2 + k_t^N P^N(x, y). \end{aligned}$$

Again, the right-hand side is bounded in $N \in \mathbb{N}_+$. Thus there is a constant $D \geq 0$ that does not depend on N such that

$$\mathbb{E} \left(\left(\mathcal{F}_{k_t^N; x, y}^N - k_t^N \pi_x^N P^N(x, y) \right)^4 \right) \leq \frac{D}{N^2}.$$

The Markov inequality yields for $\varepsilon > 0$:

$$\mathbb{P}(|\mathcal{F}_{k_t^N; x, y}^N - k_t^N \pi_x^N P^N(x, y)| > \varepsilon) \leq \frac{D}{N^2 \varepsilon^4}.$$

Since the probability is summable the Borel-Cantelli Lemma yields for $N \rightarrow \infty$:

$$|\mathcal{F}_{k_t^N; x, y}^N - k_t^N \pi_x^N P^N(x, y)| \rightarrow 0 \quad \text{almost surely.}$$

We have already seen in (3.11) that $k_t^N P^N(x, y) \rightarrow tQ(x, y)$ and together with $\pi_x^N \rightarrow \pi_x$ we get $\mathcal{F}_{k_t^N; x, y}^N \rightarrow t\pi_x Q(x, y)$ for $N \rightarrow \infty$ almost surely. This shows the second part of (3.9).

We show that the Assumptions 3.6 are fulfilled as claimed. 1. of Assumptions 3.6 holds trivially. The functions $t \mapsto \mathcal{F}_{t;x,y} = t\pi_x Q(x, y)$ are linear in t and thus absolutely continuous with differential $\mathcal{F}'_{t;x,y} = \pi_x Q(x, y)$. Together with (3.9) we get 3. of Assumptions 3.6 with almost sure pointwise convergence of processes. Note that we have for all $t \geq 0$:

$$\sum_{x \in G} \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;x,y} = \sum_{x \in G} \sum_{y \neq x} \pi_x Q(x, y) = \sum_{x \in G} -\pi_x Q(x, x) < \infty.$$

Thus 4. of Assumptions 3.6 is fulfilled. Since $\mathcal{R}_{0;x} = \pi_x$ is a distribution on G we have $\sum_{x \in G} \mathcal{R}_{0;x} = 1$ and thus 5. of Assumptions 3.6 is fulfilled. Let $x \in G$, $t \geq 0$. Next, we show the almost sure convergence of the sum:

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}^N \rightarrow t \sum_{y \neq x} \pi_x Q(x, y) = -t \pi_x Q(x, x).$$

Note that

$$\sum_{y \in G \setminus \{x\}} \mathbf{1}_{A_{j;x,y}^{i,N}} = \{X_{j-1}^{i,N} = x, X_j^{i,N} \neq x\}$$

and

$$\sum_{y \in G \setminus \{x\}} k_t^N \pi_x^N P^N(x, y) = k_t^N \pi_x^N (1 - P^N(x, x)).$$

We can argue completely analogously to the proof of the almost sure convergence $\mathcal{F}_{t;x,y}^N \rightarrow t \pi_x Q(x, y)$ by simply replacing $P^N(x, y)$ with $1 - P^N(x, x)$ and replacing “hitting y ” with “avoiding x ” in all events, to get $D \geq 0$ such that for all $\varepsilon > 0$:

$$\mathbb{P} \left(\left| \sum_{y \in G \setminus \{x\}} \mathcal{F}_{k_t^N;x,y}^N - k_t^N \pi_x^N (1 - P^N(x, x)) \right| > \varepsilon \right) \leq \frac{D}{N^2 \varepsilon^4}.$$

And again by Borel-Cantelli and due to

$$k_t^N \pi_x^N (1 - P^N(x, x)) \rightarrow -t \pi_x Q(x, x)$$

we get for all $x \in G$ and $t \geq 0$:

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}^N \rightarrow -t \pi_x Q(x, x) \quad \text{for } N \rightarrow \infty \text{ almost surely.}$$

Due to

$$\mathcal{R}_{t;x}^N = \mathcal{R}_{0;x}^N + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N - \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N$$

the almost sure convergence of $\mathcal{R}_{t;x}^N$ now implies for $N \rightarrow \infty$:

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N = \mathcal{R}_{t;x}^N - \mathcal{R}_{0;x}^N + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N \rightarrow -t \pi_x Q(x, x) \quad \text{almost surely.}$$

Since $\pi \cdot Q = 0$ we have

$$-t \pi_x Q(x, x) = t \sum_{y \in G \setminus \{x\}} \pi_y Q(y, x)$$

and thus we have almost surely for $N \rightarrow \infty$:

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}^N \rightarrow -t\pi_x Q(x, x) = \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}$$

and

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N \rightarrow -t\pi_x Q(x, x) = \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,y}.$$

Thus 6. of Assumptions 3.6 holds in terms of almost sure pointwise convergence. Due to $\mathcal{R}_{t;x} = \pi_x > 0$ for all $x \in G$, $t \in \mathbb{R}_+$ 7. of Assumptions 3.6 is also fulfilled. \square

One way to ensure that the requirements of Proposition 3.8 are met, especially if we want to attain a specific random walk in the limit, is by using lazy walkers.

Example 3.9 (Independent Lazy Walkers). Let $Q = (Q(x, y))_{x,y \in G}$ be a conservative, stable generator matrix. More precisely Q is a matrix with nonnegative off-diagonal entries (interpretation as rates), row sums equal 0 (conservative) and finite, nonpositive diagonal entries (stable). and π a distribution on G (represented as a row vector) with $\pi \cdot Q = 0$ and $\sup_{x \in G} -Q(x, x) < \infty$. Let $\alpha > 0$. For large $N \in \mathbb{N}_+$ we can define a stochastic matrix (a matrix with nonnegative entries and row sums equal 1) $P^N = (P^N(x, y))_{x,y \in G}$ via

$$P^N(x, y) = \frac{Q(x, y)}{N^\alpha} \leq 1 \quad \text{for } x \neq y$$

and

$$P^N(x, x) = 1 - \sum_{y \neq x} P^N(x, y) = 1 + \frac{Q(x, x)}{N^\alpha} \geq 0.$$

The stochastic matrix P^N describes a discrete-time Markov chain and thus a random walk on G . We call this random walk *lazy* since the probability $P(x, x)$ to remain at a given state x converges to 1 for $N \rightarrow \infty$. In words: for large N the random walk will stay in its current state for long periods of time. Furthermore the random walks for different N essentially only differ in this holding probability in the sense that the conditional probability of moving from x to y given that the walker does not stay in x is the same for all N . This Markov chain has the equilibrium distribution π since we have

for $x \in G$ and $N \in \mathbb{N}$:

$$\begin{aligned} (\pi \cdot P^N)_x &= \sum_{y \neq x} \pi_y P^N(y, x) + \pi_x \left(1 - \sum_{y \in G \setminus \{x\}} P^N(x, y) \right) \\ &= \pi_x + N^{-\alpha} \sum_{y \in G \setminus \{x\}} (\pi_y Q(y, x) - \pi_x Q(x, y)) \\ &= \pi_x + N^{-\alpha} \sum_{y \in G} \pi_y Q(y, x) = \pi_x. \end{aligned}$$

Note that $N^\alpha P^N(x, y) = Q(x, y)$ and

$$N^\alpha (P^N(x, x) - 1) = - \sum_{y \neq x} N^\alpha P^N(x, y) = Q(x, x).$$

Thus the assumptions of Proposition 3.8 are fulfilled since we even have $N^\alpha (P^N - I) = Q$ and $\pi_x^N = \pi_x$ for all $N \in \mathbb{N}_+$.

3.3 Definition of the Limiting Coalescent

We will now define the limiting coalescent using the limiting processes \mathcal{R} and \mathcal{F} given by Assumptions 3.6.

This definition will use the concept of infinitesimal rates to describe a specific type of time-inhomogeneous Markov chain which we call a *time-inhomogeneous regular jump process*. Since the concept of rates can be well enough understood on a heuristic level we will postpone some of the technical details of how the process is derived from the rates. For the rigorous construction see Definition 6.2.

Definition 3.10 (The Limiting Coalescent). We consider a sample of n ancestral lines at time zero. Let $x_i \in G$ be the starting location of the i -th ancestral line of the sample. We assume $\mathcal{R}_{0;x_i} > 0$ for all $i \in [n]$.

Let \mathcal{F} and \mathcal{R} be given as in the Assumptions 3.6. We define the limiting coalescent $\Pi = (\Pi_t)_{t \in \mathbb{R}_+}$ conditioned on \mathcal{R} and \mathcal{F} as a time-inhomogeneous regular jump process (see Definition 6.2) in the space of labeled partitions $\mathcal{P}_{G,n}$ with the following properties:

1. Let $S := \{(t, x) \in \mathbb{R}_+ \times G \mid \mathcal{R}_{t;x} > 0\}$ and $G_t := \{x \in G \mid (t, x) \in S\}$ for $t \in \mathbb{R}_+$. Note that it will turn out later (see Proposition 7.2 or, if G finite, 4 of Assumptions 3.6) that the mapping $t \mapsto \mathcal{R}_{t;x}$ is continuous. Thus S is an open subset of $\mathbb{R}_+ \times G$. We restrict the state space $\mathcal{P}_{G,n}$ at time t to labeled partitions with labels in G_t . We refer to this new

state space as the *restricted state space*. We will also later see that with the proposed rates Π will be a process in the restricted state space, which is necessary for the infinitesimal rates to be finite. In the context of Definition 6.2 the restricted state space will be the set of admissible states for our process and S will be the set of admissible states if we only consider the migration of the label of an individual line.

2. Π_0 has the form $((\{1\}, x_1), \dots, (\{n\}, x_n))$. we assume that $(0, x_i) \in S$ for all $i \in [n]$ in order to ensure that Π_0 takes values in the restricted state space.
3. Let $x \neq y \in G$ and $t \in \mathbb{R}_+$ with $(t, x) \in S$. The infinitesimal rate with which one ancestral line changes its label from x to y at time t is given by

$$q_{x,y}^{\text{Mig}} := \frac{\mathcal{F}'_{t;x,y}}{\mathcal{R}_{t;x}}. \quad (3.12)$$

4. Let $x \in G$, $t \in \mathbb{R}_+$ with $(t, x) \in S$ and let there be i ancestral lines in x at time t . Any coalescence event A for the lines at x can be considered as at most i simultaneous mergers of lines. Let $l_1, \dots, l_i \in \mathbb{N}_0$ be the size of the mergers. (So l_1 lines find a common ancestor simultaneously another l_2 find a different common ancestor and so on.) By construction $l_1 + \dots + l_i = i$. The infinitesimal rate with which such a coalescence happens is given by

$$q_{x,A}^{\text{Coal}} := \frac{\beta_x}{\mathcal{R}_{t;x}^\alpha} \phi_{x,i}(l_1, \dots, l_i). \quad (3.13)$$

Remark 3.11. By the consistency property of the ϕ (Proposition 3.4) together with the representation of consistent rates with a measure Ξ (Theorem 2.3) the limiting coalescent has the form of a spatial Ξ - n -coalescent, with potentially different Ξ_x at different sites, which additionally runs on a random timescale sped up by $\beta_x \mathcal{R}_x^{-\alpha}$. Also, the migration of ancestral lines conditioned on \mathcal{F} is independent.

One of the key features of our limiting process is that it is, in general, not described by a strongly continuous semigroup. In particular it may not be a Feller process.

Proposition 3.12. *Consider the time-inhomogeneous Markov chain Y taking values in G which describes the motion of a single ancestral line in the limiting coalescent from Definition 3.10. We define the set of real-valued, continuous functions on S , vanishing at infinity as*

$$C_0(S) := \{f: S \rightarrow \mathbb{R} \mid \text{continuous with } |f|^{-1}([\varepsilon, \infty)) \text{ compact for all } \varepsilon > 0\}.$$

Let $f \in C_0(S)$ and $t \geq 0$. We define the operator semigroup $(T_t)_{t \geq 0}$ via

$$T_t f: S \rightarrow \mathbb{R}, \quad (w, x) \mapsto \mathbb{E}(f(w+t, Y_{w+t}) | Y_w = x).$$

Furthermore let there be $x \in G$ and $0 \leq s \leq u$ with $(s, x) \in S$ but $(u, x) \notin S$. Then $(T_t)_{t \in \mathbb{R}_+}$ is not strongly continuous.

Proof. Since S is open we can assume without loss of generality, that u is chosen minimal under the constraints $s \leq u$ and $(u, x) \notin S$. We show the claim by contradiction. Assume that $(T_t)_{t \in \mathbb{R}_+}$ is strongly continuous. Then for all $f \in C_0(S)$ we have $T_t f \rightarrow f$ for $t \rightarrow 0$ in the strong topology. Let $y \in G$ with $(u, y) \in S$ and define

$$S_y := \{w \in \mathbb{R}_+ | (w, y) \in S\}.$$

Furthermore we define the following function:

$$f_y: S \rightarrow \mathbb{R}, \quad (w, z) \mapsto \begin{cases} 0 & \text{if } z \neq y \\ \frac{(1+w)(1 \wedge d(w, S_y^c))}{(1+w)(1 \wedge d(u, S_y^c))} & \text{if } z = y \end{cases}$$

where $d(w, S_y^c)$ is the Euclidean distance between the point $w \in \mathbb{R}_+$ and the closed set $S_y^c \subseteq \mathbb{R}_+$. Note that since $d(\bullet, S_y^c)$ is continuous the function f_y is also continuous. Also since $d(v, S_y^c) \rightarrow 0$ whenever $v \rightarrow w \in S$ and since $(1+v)^{-1} \rightarrow 0$ for $v \rightarrow \infty$ we have $f_y \in C_0(S)$. Furthermore we have $f_y(u, y) = 1$. Consider $n \in \mathbb{N}_+$ with $s < u - 1/n$ and thus $(u - n^{-1}, x) \in S$ due to minimality of u . We have $T_0 f(u, x) = 0$ and thus by strong continuity for $n \rightarrow \infty$

$$\mathbb{P}(Y_u = y | Y_{u-n^{-1}} = x) \leq \mathbb{E}(f_y(u, Y_u) | Y_{u-n^{-1}} = x) = T_{n^{-1}} f(u - n^{-1}, x) \rightarrow 0.$$

Since this holds for all $y \in G$ with $(u, y) \in S$ we get by dominated convergence:

$$\mathbb{P}((u, Y_u) \in S | Y_{u-n^{-1}} = x) \rightarrow 0.$$

But since S is the admissible set for Y the probability on the left hand side is always equal 1 which yields a contradiction. \square

Remark 3.13. Note that finding $(u, x) \notin S$, $(s, x) \in S$ with $s \leq u$ in Proposition 3.12 is equivalent to the existence of a site x which carries positive mass $\mathcal{R}_{s;x} > 0$ at some time s but has lost all of its mass at a later time u . This situation will usually arise if mass is not ensured to be positive at all times and sites.

3.4 The Limiting result

We can now state our main theorem.

Theorem 3.14. *Let $n \in \mathbb{N}$ and $\pi \in \mathcal{P}_{G,n}$ such that π has no ancestral lines labeled x with $\mathcal{R}_{0;x} = 0$. Let $\Pi^N = (\Pi_k^N)_{k \in \mathbb{N}_0}$ be the coalescent of the population as defined in Section 1.3. Define $(\hat{\Pi}_t^N)_{t \in \mathbb{R}_+} := (\Pi_{k_t^N}^N)_{t \in \mathbb{R}_+}$. Furthermore, let $\Pi = (\Pi_t)_{t \in \mathbb{R}_+}$ be the limiting coalescent as defined in Definition 3.10 and let $\Pi_0^N = \pi = \Pi_0$ for large N . Moreover, assume that Assumptions 3.6 hold.*

Then we get the following weak convergence of processes in the Skorohod topology (for a detailed description of the Skorohod topology see [8])

$$(\hat{\Pi}_t^N)_{t \in \mathbb{R}_+} \rightarrow (\Pi_t)_{t \in \mathbb{R}_+} \quad \text{for } N \rightarrow \infty.$$

Remark 3.15.

1. For given $n, N \in \mathbb{N}_+$ and $\pi \in \mathcal{P}_{G,n}$ with no blocks with label $x \in G$ if $\mathcal{R}_{0;x} = 0$ it may not be possible to start Π^N in π . But due to $\mathcal{R}_{0;x} > 0$ for all blocks in π with label x and due to $\mathcal{N}_{0;x}^N \sim N\mathcal{R}_{0;x}$ we can start Π^N in π for sufficiently large $N \in \mathbb{N}_+$. Thus, this poses no issue for the limiting result in Theorem 3.14.
2. Let 2. of Assumptions 3.6 hold for all $n \in \mathbb{N}_+$. Let $N \in \mathbb{N}_+$ and let $\iota_{G,N}: \mathcal{P}_{G,N} \rightarrow \mathcal{P}_G$ be a mapping with $\tau_{G,N} \circ \iota_{G,N} = \text{id}_{\mathcal{P}_{G,N}}$. Consider the starting condition $\pi_N \in \mathcal{P}_{G,N}$ given by having blocks of type $\{k\}$ which carry the location of individual $k \in [N]$ at time 0 as label and stipulate $\iota_{G,N}(\pi_N) \rightarrow \pi_\infty \in \mathcal{P}_G$ for $N \rightarrow \infty$. Note that for sample size $n = N$ we can start $\hat{\Pi}$ in π_N . Then by definition of the topology of \mathcal{P}_G and since Theorem 3.14 holds for all $n \in \mathbb{N}$ with starting condition $\tau_{G,n}^N(\pi_N)$ we have

$$\iota_{G,N}(\hat{\Pi}) \rightarrow \Pi$$

weakly with respect to the vague topology, where Π is the spatial Ξ -coalescent started in π_∞ . This means that essentially convergence holds also for large sample sizes $n = O(N)$ though this is simply due to the topology of \mathcal{P}_G which only requires all restrictions to small sample sizes n to converge.

3. Note that in 1. of Assumptions 3.6 we stipulated that the rescaling c_x^m behaves polynomially in m for $m \rightarrow \infty$ which led to the global rescaling $k_t^N = \lfloor tN^\alpha \rfloor$. This case is particularly interesting since, on one hand this covers the usual population models (e.g., Wright-Fisher-Model: $c_x^m = 1/m$, Moran-Model: $c_x^m = 1/m^2$) and on the other hand the local

mass $\mathcal{R}_{t;x}$ will influence the rate of coalescence in the limiting process. In fact by (3.13) the rate of coalescence in the limiting process will increase at sites which feature a small local mass.

Note that if the c_x^m would behave logarithmically we would not see such an influence of the local mass $\mathcal{R}_{t;x}$ on the rate of coalescence in the limit and if c_x^m would behave exponentially the influence would be so strong, that the limiting process would only allow for coalescence at the site which carries the least non-zero mass.

For now we will give a sketch for the proof of Theorem 3.14. For the rigorous proof see Chapter 8. The limiting process is a time-inhomogeneous Markov chain and thus we can consider the process $(t, \Pi_t)_{t \in \mathbb{R}_+}$ which is a time-homogeneous Markov process taking values in $\mathbb{R}_+ \times \mathcal{P}_{G,n}$. The problem with this approach is that if we allow for the mass \mathcal{R}_x to drop from some positive value to zero over time then the semigroup of the Markov process ceases to be strongly continuous (see Proposition 3.12). In particular most of the standard theory for showing convergence by considering generators or martingale problems is not applicable. At the same time we want to use the fact that we are considering a Markov chain albeit time-inhomogeneous. Since we could not find a satisfying theory which specifically treats convergence of time-inhomogeneous Markov chains we developed our own approach to this problem.

To explain said approach we first introduce the concept of graphical representations. We consider a general discrete state space E and let O be an open subset of $\mathbb{R}_+ \times E$. Also, we set $E_\infty = E \cup \{\infty\}$ to be the one-point-compactification of E if E is infinite (and $E_\infty = E$ if E is finite). We use E_∞ instead of E only for topological reasons and will not explicitly use the added point ∞ . For our purposes a graphical representation is an integer valued random measure η on $O \times E_\infty$ which is almost surely finite on compact subsets of $O \times E_\infty$, almost surely only has atoms of mass 1 and $\eta(O \times \{\infty\}) = 0$. Each atom, also called point, of η has a location $(t, x, y) \in O \times E$. We now visualize (see Figure 5.1 later on) such a point for $x \neq y$ as an arrow pointing from $(t, x) \in \mathbb{R}_+ \times E$ to $(t, y) \in \mathbb{R}_+ \times E$. We then can use such a graphical representation to define a piecewise constant process $Y = (Y_t)_{t \in \mathbb{R}_+}$ taking values in E by tracking along those arrows (see Figure 5.2 later on). More precisely, if $Y_{t-} = x$ and if there is an atom in η at (t, x, y) , then we require Y to jump to y at time t . Assume that Y is a time-discrete Markov chain taking values in E (for instance Π^N) and define \hat{Y} using a time rescaling $t \mapsto k_t^N$. We can define a graphical representation η for \hat{Y} by using the transition probabilities of Y :

$$p_{k;x,y} := \mathbb{P}(Y_{k+1} = y | Y_k = x).$$

For each $k_t^N \in \mathbb{N}_+$, $x \in E$ there is at most one arrow out of x at time k_t^N and we use the distribution on E given by the vector $(p_{k;x,y})_{y \in E}$ to determine the state which the arrow points to. In the case that the distribution yields x we have no arrow pointing out of x at time k_t^N . Moreover, we consider the distributions for different k_t^N and x to be independent.

For a Markov chain in continuous time (like Π) we can also use graphical representations. Indeed, we even use graphical representations to define Π properly in the first place (see Definition 6.2). In this case the graphical representation η is given as a Poisson point process on $O \times E_\infty$ and the infinitesimal rates $q_{t;x,y}$ of our MCCT are densities of the intensity measure of η on $\mathbb{R}_+ \times \{x\} \times \{y\}$.

These graphical representations are often used in population genetics to describe and couple stochastic processes. The relevant observation in our case is that convergence already holds in terms of the graphical representations. A sensible choice for the topology on the space of graphical representations is the so called vague topology. The vague topology can be defined on the larger space of measures on $O \times E_\infty$ which are finite on compact subsets of $O \times E_\infty$. Such a sequence of measures μ_n converges to μ if and only if

$$\int f \, d\mu_n \rightarrow \int f \, d\mu$$

for all $f: O \times E_\infty \rightarrow \mathbb{R}$ continuous with compact support.

We start by using graphical representations for $\hat{\Pi}^N$ (as described above) and Π (as by Definition 6.2. Here

$$O := \{(t, \pi) \in \mathbb{R}_+ \times \mathcal{P}_{G,n} \mid \pi \text{ has only labels in } G_t\}$$

is our restricted state space (the labeled partitions which) and $E := \mathcal{P}_{G,n}$. We consider the mapping that maps a graphical representation to a stochastic process. We show that this mapping is continuous at points where the resulting process is non-explosive with respect to the vague topology on the space of graphical representations (as a subspace of measures on $O \times E_\infty$ which are finite on compact subsets) and the Skorohod topology on the space of processes (see Proposition 5.5).

A (random) graphical representation η also defines an intensity measure ν via

$$\nu(A) = \mathbb{E}(\eta(A)) \quad \text{for all } A \subseteq O \times E_\infty \text{ measurable.}$$

We show that vague convergence of the intensity measures already implies weak convergence of the graphical representations with respect to the vague topology (see Proposition 4.19). Thus it is left to show that the limiting process is non-explosive, which is the case if and only if the migration of an

ancestral line in the limit is non-explosive, and that the intensity measures converge in the vague topology.

For the non-explosiveness we look at general time-inhomogeneous Markov chains and show that we can ensure this property by finding a certain solution to the differential equation $\pi'_t = \pi_t \cdot Q_t$ with π_t a distribution for all $t \in \mathbb{R}_+$ (see Theorem 6.19). This result is of interest in its own right. For a time-homogeneous setting $Q = Q_t$ in which Q is a bounded operator it is known that the differential equation $\pi'_t = \pi_t \cdot Q$ determines the distribution $\pi_t = \pi_0 \exp(tQ)$ of a time-homogeneous MCCT. But the generalization to the time-inhomogeneous setting introduces new issues since Q_s and Q_t may not commute. Also, in the time-homogeneous setting this is a generalization of the fact that an equilibrium distribution π is uniquely determined by the property $\pi \cdot Q = 0$.

Note that for a time-inhomogeneous Markov chain there are two different types of explosions. On one hand, the process can explode by diverging to infinity in a finite amount of time. This type of explosion is also possible in the time-homogeneous setting. On the other hand a time-inhomogeneous process allows for the possibility of explosions due to rates becoming locally infinite. This can even occur if the state space is finite. Solving the differential equation $\pi'_t = \pi_t \cdot Q_t$ not only allows us to show non-explosiveness for both cases, but we were also able to show that, for a finite G and thus finite E , the solution π_t captures explosions as discontinuities of the solution $t \mapsto \pi_t$ (see Theorem 6.18).

To show convergence of the intensity measures we have to improve the convergence of \mathcal{F}^N to \mathcal{F} to hold uniformly on compact sets (and the same for \mathcal{R}^N and \mathcal{R}). This also allows us to assume that \mathcal{F}^N and \mathcal{R}^N converge almost surely (see Lemma 7.8 and then Lemma 8.1).

Using the Continuous Mapping Theorem we can then show that $\hat{\Pi}^N$ conditioned on $\mathcal{F}^N, \mathcal{R}^N$ almost surely converges weakly with respect to the Skorohod topology to Π conditioned on \mathcal{F}, \mathcal{R} . By dominated convergence we can get rid of the conditioning and get that $\hat{\Pi}^N$ converges to Π weakly with respect to the Skorohod topology.

3.5 Discussion of the Result

We start by comparing with results about large population limits for spatial coalescents. Notahara [21] considered a situation in which individuals in the population migrate independently and at each site $x \in G$ a Wright-Fisher-type reproduction mechanism was employed. This means, that each individual chooses its parent independently uniformly at random from the previous

generation. Furthermore the system was considered to be time-homogeneous. Our scheme covers the case of independent migrators (see Proposition 3.8) and a Wright-Fisher reproduction (see Example 1.6) mechanism corresponds to $c_x^m = 1/m$ and thus $\alpha = 1$ and $\beta_x = 1$ in our setting. One difference is that in Notahara's setting the reproduction step rebalanced the local population size whereas we could have a situation in which the local population size is asymptotically balanced but the reproduction keeps the number of individuals constant. As in the nonspatial case the limiting coalescent will be a spatial Kingman-coalescent, also referred to as the structured coalescent.

Herbots [11] and later Heuer and Sturm [12] considered balanced migration on a finite set G . In [12] the number of migrants from x to y was $n_{xy} \in \mathbb{N}_0$ with

$$\sum_{y \in G \setminus \{x\}} n_{xy} = \sum_{y \in G \setminus \{x\}} n_{yx}.$$

In this case the local population size N_x is constant in t as well as deterministic. Furthermore the local population size N_x was assumed to also be constant in x . It was assumed that $N_x/N \rightarrow 1/|G| > 0$ as well as $n_{xy}/c^N \rightarrow p(x, y) \geq 0$ where $c^N \rightarrow 0$ was a global time rescaling. Note that our result covers this situation since for $t \in \mathbb{R}_+$, $x \neq y \in G$ we have for $N \rightarrow \infty$:

$$\mathcal{R}_{t;x}^N = N_x/N \rightarrow 1/|G| = \mathcal{R}_{t;x}, \quad \mathcal{F}_{t;x,y}^N = \left\lfloor \frac{t}{c^N} \right\rfloor n_{xy} \rightarrow tp(x, y) = \mathcal{F}_{t;x,y}.$$

There are also convergence results in which the spatial information does not impact the coalescent in the limit, so called fast migration results. Heuristically, if migration happens on a faster timescale than coalescence it is possible for the spatial distribution of the ancestral lines to become almost stationary before coalescence happens. In such a case the limiting coalescent does no longer see the underlying spatial structure.

Hössjer [13] provided such a limiting result for the fast migration case on a finite space $\{1, \dots, L\}$. Similar to [12] a balanced migration and exchangeable, conservative reproduction was assumed but instead of migration happening on the slow timescale $O(c_N)$ the migration happened on the timescale $O(1)$. To attain a limit only the coalescent component of the spatial coalescent is considered. Hössjer [13] showed that under certain conditions on the moments of the offspring distribution (similar to the convergence of the $\phi_{x,i}(l_1, \dots, l_i)$ in 2 of our Assumptions 3.6) the coalescent component of the genealogy converges to a Kingman-coalescent with a time rescaling $c > 0$.

Another fast migration result on a finite space was given by Sagitov et al. [24] who considered the migration backwards in time at time $u \in \mathbb{N}$ to be governed by random transition matrices $B_1^{(u)}$ which converge for $u \rightarrow \infty$ to a

random stationary transition matrix P_1 . It is noteworthy that even though the migration in [24] is time-inhomogeneous this is not due to change in the local population size, which was still assumed to remain constant in time. This is similar to the case in our setting, in which the flows \mathcal{F} may keep the local total mass \mathcal{R} constant but may themselves vary, thus introducing a time-inhomogeneous component without changing the local population size. More precisely, the (random) transition matrices for a population of N individuals $B_1^{(u)}(N)$ are assumed to have the following asymptotics:

$$B_1^{(u)}(N) = B_1^{(u)} + N^{-1}D_1^{(u)}(N)$$

with $D_1^{(u)}(N)$ uniformly bounded in $u, N \in \mathbb{N}$. Furthermore some assumptions on the mixing property of the transition matrices $B_1^{(u)}$ were made. Sagitov et al. [24] showed that under these assumptions the genealogy of a spatial Wright-Fisher model, with a suitable time rescaling, converges to the standard (nonspatial) Kingman-Coalescent.

We will now talk about applications of our result. Our scaling regime is of particular interest since it retains the spatial information of the system. This is useful for studying the influence of spatial structure on genealogies. There are plenty of recent results looking at spatial Λ -Coalescents in particular.

In the previously mentioned paper by Heuer and Sturm [12] the resulting large population limit was then used to derive a robustness result for the coalescent of a population living on a large, discrete 2-dimensional torus. More precisely it was shown that the coalescence component of a suitably time-rescaled spatial Λ -coalescent in which lines migrate using simply symmetric random walks on the discrete torus $\{-L, \dots, L\}^2$ converges for $L \rightarrow \infty$ to a time-rescaled Kingman coalescent if the lines were sampled suitably far apart at time 0.

Using the result from this thesis we can extend the application of the paper greatly. It is no longer required for the number of migrants n_{xy} to be deterministic or balanced and the N_x to be constant in x , instead we only require the limiting process to be well behaved. For instance modeling migration using independent migrators (see Examples 1.6) still allows for the same type of large population limit.

Since the core argument in [12] was that with probability 1 for $L \rightarrow \infty$ only at most two ancestral lines could meet at the same time the robustness result in [12] could be extended to a setting where instead of a spatial Λ -coalescent one could consider a spatial Ξ -coalescent since both coalescents have the same behavior for sample size 2.

Results similar to [12] but for a spatial Λ -coalescent on the d -dimensional discrete torus with $d \geq 3$ were already derived by Limic and Sturm [19]. The

transience of the simple symmetric random walk in \mathbb{Z}^d for $d \geq 3$ allowed for nontrivial results even if the lines are not sampled far apart initially.

There are also results considering a spatial Λ -coalescent in a continuum. Barton et al. [2] showed that the genealogy of a suitably time-rescaled spatial Λ -coalescent on the 2-dimensional torus $[-L/2, L/2]$ converges to the Kingman-coalescent. In order to work with the spatial continuum small and large reproduction events affecting the population were considered. Small events would only require repopulation of an area with a radius $r \ll L$ but they would happen frequently. Large events would affect an area of radius Ψ_L with $\Psi_L \rightarrow \infty$ for $L \rightarrow \infty$ but they would be rare (rate $\rho_L \psi_L^{-2} \rightarrow 0$ for $L \rightarrow \infty$). These rare, large scale events lead to a Λ -coalescent in the limit whereas the small events only yield a Kingman-component (an atom at 0 in Λ) in the limit.

Greven et al. [9] considered a spatial Λ -Cannings model subject to a mean field migration mechanism. The coalescent of such a model is a spatial Λ -coalescent with migration rates $q_{t;x,y}^{\text{Mig}} = c$ for all $x, y \in G$, $t \geq 0$. Thus this coalescent fits into our framework if we consider a symmetric, balanced setting $\mathcal{R}_{t;x} = 1/|G|$, $\mathcal{F}_{t;x,y} = ct/|G|$ for all $x, y \in G$ and assume that no simultaneous mergers in our Ξ -coalescent can happen (making it a Λ -coalescent).

Greven et al. [10] considered a spatial coalescent on a countable, additive group V (later $V = \mathbb{Z}$). Individuals were assumed to migrate (backwards in time) with rate 1 using a transition kernel $a(x, y) = a(0, y - x)$ and to reproduce with rate γ according to a time-continuous Moran model. More precisely every pair of individuals would, with rate γ , be replaced with two offspring of one individual of the pair, chosen uniformly at random. Note that the total rate with which the considered Markov chain moves out of $x \in V$ does not depend on V . By Kolmogorov's backwards equation (for time-homogeneous Markov chains in continuous time) the assumptions of Theorem 6.19 are fulfilled and thus we can apply Proposition 7.5 to see that there exists a system of flows which yields this particular migration mechanism. In [10] it is assumed that each step of the random walk given by $a(x, y)$ has zero mean and finite variance $\sigma^2 \in (0, \infty)$. For a given $\varepsilon > 0$ time gets rescaled via $t \mapsto \varepsilon^2 t$ and space gets rescaled via $z \mapsto \varepsilon \sigma^{-1} z$. For $\varepsilon \rightarrow 0$ a limit to the genealogy of a so called continuum-sites stepping stone model on \mathbb{R} is attained. This genealogy can be described as a system of coalescing Brownian motions on \mathbb{R} .

It should be noted that all these results consider usually at least a time-homogeneous migration for the ancestral lines. Our result of course allows even for time-inhomogeneous limits which will usually happen due to large, migration induced fluctuations in the local population size. Furthermore we have seen in Proposition 3.12 that severe bottlenecks can even lead to loss

of the Feller property of the limiting coalescent. Thus our result can also help understanding the limits in which the previously discussed papers should be applied. Genealogies of populations which were subject to substantial migration in the past, for instance using migration to populate a previously vacant location, will look significantly different than genealogies of populations with relatively constant local population size.

3.6 Overview Over the Rest of the Thesis

The remainder of the thesis is focused on proving our main result, Theorem 3.14. In Chapter 4 we introduce the concept of vague convergence and conditions for weak convergence of integer valued random measures with respect to the vague topology. Furthermore we show that vague convergence of intensity measures of certain point processes implies weak convergence of the point processes themselves to a Poisson point process.

In Chapter 5 we introduce the concept of (proper) graphical representations. Furthermore, we describe how a process can be defined by tracking through a graphical representation and show continuity of this procedure at certain, good representations.

In Chapter 6 we define time-inhomogeneous regular jump processes with rate matrix Q_t which are time-inhomogeneous Markov chains that can be defined using graphical representations. We also develop a method to identify the distribution and in particular non-explosive behavior of these jump processes by solving the differential equation $\pi'_t = \pi_t \cdot Q_t$.

In Chapter 7 we introduce the system of flows which is our limiting migration mechanism. We show that the rate matrix Q_t for a system of flows yields a proper graphical representation and find a solution for $\pi'_t = \pi_t \cdot Q_t$ in order to show the non-explosiveness of our migration mechanism. Moreover, we show that the weak convergence of the finite dimensional distributions of the flow processes \mathcal{F}^N and the mass processes \mathcal{R}^N can be refined to weak uniform convergence on compact sets.

In Chapter 8 we prove our main Theorem 3.14 and in Chapter 9 we give an outlook and ideas for future research and possible generalizations of our results.

Chapter 4

The Vague Topology

4.1 Definition and Basic Theory

As was outlined in the sketch of the proof for the main result the vague topology will be important. Thus we want to give some well known results for integer valued random measures and derive some more specific convergence results we will need later on. This chapter mostly uses the groundwork done by Kallenberg [16]. We start by introducing some notations.

Notation 4.1.

1. Let A_i be disjoint sets with $i \in I$ for some index set I . We write

$$\sum_{i \in I} A_i := \bigcup_{i \in I} A_i.$$

2. Let $A \subset E$ with E a topological space. We write \bar{A} for the *closure* of A , A° for the *interior* of A and $\partial A := \bar{A} \setminus A^\circ$ for the *boundary* of A .
3. Let $A \subset E$ with E a topological space. Recall that A is called *compact* if every open cover of A allows for a finite subcover and *relatively compact* if \bar{A} is compact.

The following lemma shows some basic properties of the boundary.

Lemma 4.2. *Let E be a topological space. Let A, B be subsets of E . The following claims hold.*

1. *We have*

$$\overline{A \cup B} = \bar{A} \cup \bar{B}, \quad (A \cap B)^\circ = A^\circ \cap B^\circ \quad (4.1)$$

as well as

$$\overline{A \cap B} \subseteq \bar{A} \cap \bar{B}, \quad A^\circ \cup B^\circ \subseteq (A \cup B)^\circ. \quad (4.2)$$

If additionally E is separable and $\overline{A \cap B} \subseteq A^\circ \cup B^\circ$, then equality holds in (4.2).

2. It holds: $\partial A = \partial A^c$.

3. It holds:

$$\partial(A \setminus B) \subseteq \partial A \cup \partial B. \quad (4.3)$$

Furthermore, if E is separable, $B \subseteq A$ and either B is closed and A is open or A is closed and B is open, then equality holds in (4.3).

4. It holds:

$$\partial(A \cup B) \subseteq \partial A \cup \partial B \quad \text{and} \quad \partial(A \cap B) \subseteq \partial A \cup \partial B. \quad (4.4)$$

Proof.

1. We start by showing the first equality in (4.1). We have on one hand:

$$\overline{(A \cup B)} \subseteq \overline{(\overline{A} \cup \overline{B})} = \overline{A} \cup \overline{B}$$

and on the other hand:

$$A \subseteq \overline{(A \cup B)} \quad \text{and} \quad B \subseteq \overline{(A \cup B)}$$

which yields

$$\overline{A} \subseteq \overline{(A \cup B)} \quad \text{and} \quad \overline{B} \subseteq \overline{(A \cup B)}.$$

Thus we get $\overline{A} \cup \overline{B} \subseteq \overline{(A \cup B)}$. The second equality of (4.1) can be shown by applying complements on both sides and using $\overline{A^c} = (A^\circ)^c$.

Now we show (4.2). We have

$$\overline{(A \cap B)} \subseteq \overline{(\overline{A} \cap \overline{B})} = \overline{A} \cap \overline{B}.$$

The second inclusion follows by applying complements.

Now let

$$x \in \overline{A} \cap \overline{B} \subseteq A^\circ \cup B^\circ.$$

Without loss of generality let $x \in A^\circ$. Since $x \in \overline{B}$ and E separable there exists a sequence $(x_n)_{n \in \mathbb{N}_+} \subseteq B$ which converges to x .

Since A° is an open neighborhood of x there is an $m \in \mathbb{N}$ with $x_n \in A^\circ$ for all $n \geq m$. Thus the subsequence $(x_n)_{n \geq m}$ lives in $A \cap B$ and converges to x . This implies $x \in \overline{A \cap B}$. Thus we have $\overline{A} \cap \overline{B} \subseteq \overline{A \cap B}$ which shows equality. The second equality in (4.2) follows by applying

complements to both sides of the first equality using A^c and B^c instead of A , B and using that

$$\begin{aligned}\overline{A^c} \cap \overline{B^c} &\subseteq (A^c)^\circ \cup (B^c)^\circ \\ \Leftrightarrow (A^\circ)^c \cap (B^\circ)^c &\subseteq \overline{A^c} \cup \overline{B^c} \\ \Leftrightarrow \overline{A} \cap \overline{B} &\subseteq A^\circ \cup B^\circ.\end{aligned}$$

2. We have due to $\overline{A^c} = (A^\circ)^c$:

$$\partial A^c = \overline{A^c} \setminus (A^c)^\circ = \overline{A^c} \cap ((A^c)^\circ)^c = (A^\circ)^c \cap \overline{(A^c)^c} = \overline{A} \cap (A^\circ)^c = \partial A.$$

3. We start by showing (4.3). We have due to (4.1) and (4.2):

$$\begin{aligned}\partial(A \setminus B) &= \overline{(A \setminus B)} \setminus (A \setminus B)^\circ \subseteq \overline{A} \cap \overline{B^c} \cap (A^\circ \cap (B^c)^\circ)^c \\ &= \overline{A} \cap \overline{B^c} \cap ((A^\circ)^c \cup ((B^c)^\circ)^c) \\ &= (\overline{A} \cap \overline{B^c} \cap (A^\circ)^c) \cup (((B^c)^\circ)^c \cap \overline{A} \cap \overline{B^c}) \\ &\subseteq (\overline{A} \cap (A^\circ)^c) \cup (\overline{B^c} \cap ((B^c)^\circ)^c) = \partial A \cup \partial B^c.\end{aligned}$$

Now, 2. implies $\partial B = \partial B^c$ which shows the first claim.

We continue with the second claim. Let $B \subseteq A \subseteq E$ with B open and A closed. Note that due to 2. we have $\partial B = \partial B^c$. Thus we get, since $A \setminus B$ is closed and due to (4.1):

$$\begin{aligned}\partial(A \setminus B) &= (A \setminus B) \setminus (A \setminus B)^\circ \\ &= A \cap B^c \cap ((A \cap B^c)^\circ)^c \\ &= A \cap B^c \cap (A^\circ \cap (B^c)^\circ)^c \\ &= A \cap B^c \cap (A^\circ)^c \cup ((B^c)^\circ)^c \\ &= (A \cap B^c \cap (A^\circ)^c) \cup (A \cap B^c \cap (B^c)^\circ)^c \\ &= (B^c \cap (A \setminus A^\circ)) \cup (A \cap (B^c \setminus (B^c)^\circ)) \\ &= (B^c \cap \partial A) \cup (A \cap \partial B^c) \\ &= (B^c \cap \partial A) \cup (A \cap \partial B).\end{aligned}$$

Moreover, we have $\partial B \subseteq \overline{B} \subseteq A$ and $\partial A \subseteq \overline{A^c} \subseteq B^c$ and thus the right hand side of the equation simplifies to $\partial A \cup \partial B$.

Next we assume that A is open and B is closed. We first note that $\overline{A} = A \cup \partial A$ and due to $B \subseteq A$ and 2:

$$\partial A = \partial A^c \subseteq A^c \subseteq B^c = (B^c)^\circ$$

as well as $\overline{B^c} = B^c \cup \partial B^c$ and

$$\partial B^c = \partial B \subseteq B \subseteq A = A^\circ.$$

Thus, it holds:

$$\overline{A} \cap \overline{B^c} = (A \cup \partial A) \cap (B^c \cup \partial B^c) \subseteq A \cup B^c = A^\circ \cup B^c.$$

Therefore, equality holds in (4.2) and we get

$$\begin{aligned} \partial(A \setminus B) &= (\overline{A \setminus B}) \setminus (A \setminus B) \\ &= \overline{A \setminus B} \cap (A \setminus B)^c \\ &= \overline{A} \cap B^c \cap (A^c \cup B) \\ &= (\overline{A} \cap A^c \cap B^c) \cup (B^c \cap B \cap \overline{A}) \\ &= (\partial A \cap B^c) \cup (\partial B^c \cap \overline{A}). \end{aligned}$$

The right-hand side simplifies to $\partial A \cup \partial B$ since

$$\partial A = \partial A^c \subseteq A^c \subseteq B^c \quad \text{and} \quad \partial B^c = \partial B \subseteq B \subseteq A.$$

4. We start by showing the first inclusion. We have due to (4.3) and 2.:

$$\partial(A \cup B) = \partial(A^c \cap B^c) = \partial(A^c \setminus B) \subseteq \partial A^c \cup \partial B = \partial A \cup \partial B.$$

The second inclusion follows in similar fashion:

$$\partial(A \cap B) = \partial(A \setminus B^c) \subseteq \partial A \cup \partial B^c = \partial A \cup \partial B.$$

□

Remark 4.3. The separability of E in 2. and 3. of Lemma 4.2 can be dropped. The proof works analogously for general topological spaces if one replaces the concept of sequences with the concept of nets in the proof. We only presented the results in the separable case since our spaces are always separable and nets are not common knowledge outside of topology.

We now introduce the class of spaces considered by Kallenberg [16].

Definition 4.4. We call a topological space E equipped with its Borel σ -field $\mathcal{B}(E)$ an *lscH space* if:

1. E is locally compact. This means that every point $x \in E$ allows for a relatively compact, open neighborhood.

2. E is second countable. This means that the topology has a countable basis, or more precisely, a countable collection of open sets $\{O_n | n \in \mathbb{N}\}$ such that for each $U \subseteq E$ open there exists $I \subseteq \mathbb{N}$ with

$$U = \bigcup_{i \in I} O_i.$$

It should be noted that for metrizable spaces this is equivalent to separability.

3. E is a Hausdorff space. This means that for any two points $x \neq y \in E$ there exist disjoint open neighborhoods of x and y .

Moreover, we define $\mathcal{M}_c(E)$ as the set of measures on $(E, \mathcal{B}(E))$ which are finite on (relatively) compact subsets of E .

The following proposition introduces some basic properties of lscH spaces:

Proposition 4.5 (Properties of lscH Spaces). *Let E be an lscH space and $\mu \in \mathcal{M}_c(E)$. The following claims hold:*

1. E is metrizable and separable.
2. The system of open, relatively compact sets $B \subseteq E$ with $\mu(\partial B) = 0$ is a π -system (i.e., closed under finite intersections) which generates the topology of E and in particular $\mathcal{B}(E)$.
3. E is σ -compact, which means there exists a sequence $C_n \nearrow E$ with C_n compact for all $n \in \mathbb{N}_+$. We can choose this sequence with the property $\mu(\partial C_n) = 0$ for all $n \in \mathbb{N}_+$.
4. μ is σ -finite.

Proof.

1. For this part of the proof we will refer to the notation and results by Joshi [15]. We give a short description of the nomenclature used in [15]. A space E is called *regular* if, for any point $x \in E$ and any closed subset $C \subseteq E$ with $x \notin C$, we can find disjoint, open sets U, V with $x \in U$ and $C \subseteq V$. E is called T_1 if for any $x \neq y \in E$ we can find an open neighborhood U of x which does not contain y . E is called T_3 if it is T_1 and regular.

We first note that every Hausdorff space is T_1 . Due to Corollary (3.4) in [15] every locally compact Hausdorff space is regular and thus T_3 . Now Urysohn's Metrization Theorem (Corollary (3.2) in [15]) yields that a

second countable space is metrizable if and only if it is T_3 . In particular E is metrizable. Let $d: E \times E \rightarrow \mathbb{R}_+$ be a corresponding metric. To conclude the claim we note that every second countable space has to be separable (Theorem (1.5) in [15]).

2. To show that the system is a π -system consider $B, C \subseteq E$ open, relatively compact with $\mu(\partial B) = \mu(\partial C) = 0$. Due to $D := B \cap C \subseteq B$ the intersection D is again relatively compact as well as open and (4.4) yields $\partial D \subseteq \partial B \cup \partial C$. Thus we have $\mu(\partial D) \leq \mu(\partial B) + \mu(\partial C) = 0$. Therefore the system is closed under finite intersections and therefore a π -system.

Now we show that the system generates $\mathcal{B}(E)$. We can use the metric d to define the open ball with radius $r > 0$ and center $x \in E$:

$$B_r(x) := \{y \in E \mid d(x, y) < r\}.$$

Since E is locally compact for every $x \in E$ there exists an $r_x > 0$ such that $B_{r_x}(x)$ is relatively compact and thus $B_\varepsilon(x)$ is relatively compact for all $\varepsilon \in (0, r_x)$. Furthermore, we have for the closure and the boundary of the ball:

$$\overline{B_r(x)} = \{y \in E \mid d(x, y) \leq r\} \quad \partial B_r(x) = \{y \in E \mid d(x, y) = r\}.$$

In particular the boundaries $\partial B_r(x)$ are mutually disjoint for different $r > 0$. Since the set $(0, r_x]$ is uncountable and $\mu(\overline{B_{r_x}(x)}) < \infty$ we can find $r_{x,n} > 0$ with $B_{r_x}(x)$ relatively compact and $\mu(\partial B_{r_{x,n}}(x)) = 0$ for all $x \in E$, $n \in \mathbb{N}_+$ such that $r_{x,n} \rightarrow 0$ for $n \rightarrow \infty$. Now let $U \subseteq E$ be open. We have

$$U = \bigcup_{x,n} B_{r_{x,n}}(x)$$

where the union is taken over all $x \in E$ and all $n \in \mathbb{N}_+$ with $B_{r_{x,n}}(x) \subseteq U$. Thus the topology of E is generated by relatively compact sets such that μ has no mass on the boundary and in particular the Borel- σ -algebra $\mathcal{B}(E)$ is generated by these sets.

3. Any second countable space is also Lindelöf (see Theorem (1.5) in [15]). This means that any open cover allows for a countable subcover. We use the construction from the previous part of the proof to write

$$E = \bigcup_{x \in E, n \in \mathbb{N}_+} B_{r_{x,n}}(x).$$

Since E is a Lindelöf space this open cover can be reduced to a countable subcover

$$E = \bigcup_{n \in \mathbb{N}_+} B_n$$

with B_n open, relatively compact and $\mu(\partial B_n) = 0$. We define

$$C_n := \bigcup_{i=1}^n \bar{B}_i.$$

As a finite union of compact sets C_n is compact and $C_n \nearrow E$ by construction. Furthermore we have by (4.4): $\partial C_n \subseteq \bigcup_{i=1}^n \partial B_i$. Thus E is σ -compact and due to $\mu(\partial B_i) = 0$ for all $i \in \mathbb{N}_+$ we have $\mu(\partial C_n) = 0$ for all $n \in \mathbb{N}_+$.

4. We use the sequence C_n of compact sets constructed in the previous part of the proof. Thus $\mu(C_n) < \infty$ by definition and $C_n \nearrow E$. This shows that μ is σ -finite.

□

Remark 4.6. Note that every locally compact, separable metric space is also an lscH space. Thus 1. of Proposition 4.5 yields that lscH spaces are essentially the same as locally compact, separable metric spaces. The concept of lscH spaces is helpful if it would be cumbersome to construct a metric on a given space. For our purposes E will have the form of a countable disjoint union of copies of open subsets of \mathbb{R}_+ or \mathbb{R} which is clearly an lscH space.

We now introduce a topology on $\mathcal{M}_c(E)$.

Definition 4.7 (The Vague Topology). Let E be an lscH space. We say that a sequence $(\mu_n)_{n \in \mathbb{N}_+} \subseteq \mathcal{M}_c(E)$ converges *vaguely* to $\mu \in \mathcal{M}_c(E)$ if and only if, for any functions $f \in C_c(E)$ (continuous mapping $E \rightarrow \mathbb{R}$ with compact support), we have

$$\int f \, d\mu_n \rightarrow \int f \, d\mu \quad \text{as } n \rightarrow \infty.$$

In this case we write $\mu_n \xrightarrow{v} \mu$. The induced topology on $\mathcal{M}_c(E)$ is called the *vague topology*.

Definition 4.8. A subset $\mathcal{U} \subseteq \mathcal{B}(E)$ is called *separating* if for any compact set K and any open set G with $K \subseteq G$ there is a $U \in \mathcal{U}$ with $K \subseteq U \subseteq G$.

The following proposition gives some important examples of separating sets.

Proposition 4.9.

1. Let $E = \mathbb{R}^d$ with $d \in \mathbb{N}_+$ and \mathcal{U} the set of all disjoint finite unions of half open, bounded rectangles of the form $(a, b]$. Then \mathcal{U} is separating.
2. Let $F \subseteq E$ be a subset equipped with the induced topology. Then the compact sets in F are precisely the compact sets in E which are contained in F .
3. Let $O \subseteq E$ be an open subspace and O locally compact. Given a separating set \mathcal{U} with respect to E the set

$$\mathcal{U}_O := \{U \in \mathcal{U} \mid \bar{U} \subseteq O \text{ compact}\}$$

is separating with respect to O .

4. Let $C \subseteq E$ be a closed subspace. Given a separating set \mathcal{U} with respect to E the set

$$\mathcal{U}_C := \{U \cap C \mid U \in \mathcal{U}\}$$

is separating with respect to C .

5. Let $E = \mathbb{R} \times F \times F_\infty$ with F an at most countable, discrete set and $F_\infty = F \cup \{\infty\}$ the one-point compactification of F if F is infinite and $F_\infty = F$ if F is finite. Let \mathcal{U} be the set of all finite, disjoint unions of sets of the form $(a, b] \times F^{(1)} \times F^{(2)} \subseteq \mathbb{R} \times F \times F_\infty$ with $a < b$, $F^{(1)} \subseteq F$ finite and $F^{(2)} \subseteq F_\infty$ compact. Then \mathcal{U} is separating.

Proof.

1. Let $C \subseteq U \subseteq E$ with C compact and U open. Let $x \in C$. Since U is open we can find an open rectangle set (a_x, b_x) such that $x \in (a_x, b_x)$ and $(a_x, b_x] \subseteq U$. Since C is compact and $\{(a_x, b_x) \mid x \in C\}$ is an open cover of C we can find an $n \in \mathbb{N}$ and $x_k \in C$ for $k \in [n]$ such that $\{(a_{x_1}, b_{x_1}), \dots, (a_{x_n}, b_{x_n})\}$ defines a finite subcover of C . Thus we have

$$C \subseteq \bigcup_{k=1}^n (a_{x_k}, b_{x_k}] \subseteq U.$$

Since half open rectangles in \mathbb{R}^d form a semiring we can refine the collection of rectangles to a disjoint collection.

2. Let $K \subseteq F$ be a compact set in F . Any cover $\{O_i\}_{i \in I}$ of K with open sets (in E) $O_i \subseteq E$ yields an open cover with open sets (in F) $O_i \cap F \subseteq F$ which has a finite subcover $(O_{i_n} \cap F)_{n=1}^m$ in F . Then $(O_{i_n})_{n=1}^m$ is a finite subcover in E . Thus we have that any compact set in the subspace F is also a compact set in E .

Now let $K \subseteq E$ be compact (in E) and $K \subseteq F$. Let $(O_i \cap F)_{i \in I}$ be an open cover of K in F such that $O_i \subseteq E$ is open for all $i \in I$. Thus $(O_i)_{i \in I}$ is an open cover of K in E and has a finite subcover $(O_{i_n})_{n=1}^m$. But since $K \subseteq F$ we have that $(O_{i_n} \cap F)_{n=1}^m$ is a finite subcover in F . Thus K is compact in F .

3. Let $K \subseteq G \subseteq O$ with K compact and G open in O . By the previous part K is also compact in E and since O is open G is open in E . We could now already separate K and G in E but we want the additional condition that the separating set has a compact closure. We first note that since O is locally compact we can find an open set $H \subseteq G$ with $K \subseteq H$ and $\overline{H} \subseteq O$ compact. To construct H in detail we use local compactness to find for any given $x \in K$ a neighborhood $B_x \subseteq G$ with $\overline{B_x} \subseteq O$ compact. The B_x yield an open cover of K and thus we can find a finite subcover. Since the union of finitely many compact sets is again compact this subcover has the desired property. Since \mathcal{U} is separating in E we can find a $U \in \mathcal{U}$ such that $K \subseteq U \subseteq H \subseteq G$. Due to $\overline{U} \subseteq \overline{H} \subseteq O$ and compactness of \overline{H} we have \overline{U} compact and thus $U \in \mathcal{U}_O$.
4. Let $K \subseteq O \subseteq C$ with K compact (in C) and O open (in C). Due to 2. C is compact as a subset of E . Moreover, since $O \subseteq C$ is open there exists $\hat{O} \subseteq E$ open with $O = \hat{O} \cap C$. Thus we have $K \subseteq O \subseteq \hat{O} \subseteq E$ with K compact in E and \hat{O} open in E . Since \mathcal{U} is separating for E there exists $U \in \mathcal{U}$ with

$$K \subseteq U \subseteq \hat{O} \quad \text{and thus} \quad K \cap C \subseteq U \cap C \subseteq \hat{O} \cap C.$$

Since $K \cap C = K$ and $\hat{O} \cap C = O$ this shows the claim.

5. Let F be finite with cardinality $n \in \mathbb{N}_+$. The topological space $\mathbb{R} \times F \times FF_\infty$ is coordinatewise isomorphic to

$$E := \mathbb{R} \times [n] \times [n] \subseteq \mathbb{R}^3.$$

Furthermore E is a closed subspace of \mathbb{R}^3 . Thus we can apply 1. together with 4. to conclude that finite disjoint unions of bounded rectangle sets in \mathbb{R}^3 intersected with E are a separating system for E .

Now, let F be infinite. The topological space $\mathbb{R} \times F \times F_\infty$ is coordinate-wise isomorphic to

$$E := \mathbb{R} \times \mathbb{N}_+ \times (\{-1/n | n \in \mathbb{N}_+\} \cup \{0\}) \subseteq \mathbb{R}^3.$$

Furthermore E is a closed subspace of \mathbb{R}^3 . Thus we can apply 1. together with 4. to conclude that finite disjoint unions of bounded rectangle sets in \mathbb{R}^3 intersected with E are a separating system for E . Note that for $a < b \in \mathbb{R}$ the set $(a, b] \cap (\{-1/n | n \in \mathbb{N}_+\} \cup \{0\})$ is always compact and the set $(a, b] \cap \mathbb{N}_+$ is always finite. Thus the intersections of rectangles with \mathbb{R}^3 produces exactly sets of the claimed form.

□

Theorem 4.10 (Some Basic Facts about the Vague Topology). *Let E be an $lcscH$ space.*

1. $\mathcal{M}_c(E)$ endowed with the vague topology is a Polish space. This means that $\mathcal{M}_c(E)$ is metrizable in a fashion that makes it a separable, complete metric space.
2. $\mu_n \xrightarrow{v} \mu$ if and only if for all $B \subseteq E$ relatively compact with $\mu(\partial B) = 0$ we have $\mu_n(B) \rightarrow \mu(B)$.
3. The space of integer valued measures on E that are finite on compact sets is a closed subspace of $\mathcal{M}_c(E)$.

4. Let $\mu_n \xrightarrow{v} \mu$ then

$$\int f \, d\mu_n \rightarrow \int f \, d\mu$$

for all bounded, measurable f which are μ -almost everywhere continuous and have compact support.

5. Let $E = \mathbb{R}_+$ and let μ_n, μ be measures in $\mathcal{M}_c(\mathbb{R}_+)$ with measure generating functions F_n, F with $F_n(0) = 0 = F(0)$. The convergence of $F_n(t) \rightarrow F(t)$ for all $t \in \mathbb{R}_+$ at which F is continuous is equivalent to $\mu_n \xrightarrow{v} \mu$ in $\mathcal{M}_c(\mathbb{R}_+)$.

Proof.

1. This is part of Theorem A 2.3 in [16]. Note that in particular E itself is a Polish space. Using this we can choose a metric d on E for which (E, d) is separable and complete.

2. This is also part of Theorem A 2.3 in [16]. Technically, only one direction is stated in the theorem so alternatively consider Theorem 14.16 in [16] which shows it for the more general case of random measures.
3. Let $\mu_n \rightarrow \mu$ with μ_n integer valued measures. By the previous claim we have $\mu_n(B) \rightarrow \mu(B)$ for all $B \subseteq E$ relatively compact with $\mu(\partial B) = 0$. Furthermore we have $\mu_n(B) \in \mathbb{N}_0$ for all $n \in \mathbb{N}_+$ and thus $\mu(B) \in \mathbb{N}_0$. Let $C \subseteq E$ be relatively compact with $\mu(\partial C) = 0$. We now use Dynkin's π - λ -Theorem to conclude that the restricted measure $\mu|_C$ has to be integer valued on all measurable sets. Consider the system

$$\mathcal{D} := \{B \in \mathcal{B}(C) \mid \mu(B) \in \mathbb{N}_0\}.$$

we show that \mathcal{D} is a Dynkin system over C :

- Since C is relatively compact with $\mu(\partial C) = 0$ we have already shown $\mu(C) \in \mathbb{N}_0$ and thus $C \in \mathcal{D}$.
- Let $A \in \mathcal{D}$. Then $\mu(C \setminus A) = \mu(C) - \mu(A) \in \mathbb{Z}$. Now $\mu(C \setminus A) \geq 0$ implies $C \setminus A \in \mathcal{D}$.
- Let $A_1, A_2, \dots \in \mathcal{D}$ pairwise disjoint. Then

$$\mu \left(\bigcup_{n \in \mathbb{N}_+} A_n \right) = \sum_{n \in \mathbb{N}_+} \mu(A_n) \in \mathbb{N}_0 \cup \{\infty\}.$$

But due to $\mu(C) < \infty$ the above sum has to converge and we get

$$\bigcup_{n \in \mathbb{N}_+} A_n \in \mathcal{D}.$$

Thus \mathcal{D} is a Dynkin system. We have already seen that all relatively compact subsets $B \subseteq C$ with $\mu(\partial B) = 0$ are contained in C . By 2. of 4.5 these sets are a π -system generating $\mathcal{B}(C)$. We apply Dynkin's π - λ -Theorem to conclude $\mathcal{D} = \mathcal{B}(C)$ and thus $\mu|_C(B) \in \mathbb{N}_0$ for all $B \in \mathcal{B}(C)$. To extend this result to E we apply 3. of 4.5 to find a sequence C_n of compact sets with $\mu(\partial C_n) = 0$ such that $C_n \nearrow E$. We have for every measurable set $B \in \mathcal{B}(E)$ by the continuity of the measure μ from below:

$$\mu(B) = \lim_{n \rightarrow \infty} \mu(B \cap C_n) = \mu|_{C_n}(B \cap C_n) \in \mathbb{N}_0 \cup \{\infty\}.$$

Thus μ is an integer valued measure on E .

4. This follows analogously to the proof in case of weak convergence of probability measures. Let $f: E \rightarrow \mathbb{R}$ bounded, measurable with compact support C and $N_f \subseteq C$ its set of discontinuities. By assumption we have $\mu(N_f) = 0$. For now we consider the case $\mu(\partial C) = 0$. We start by restricting f to C . It suffices to show convergence for the restricted function. Let $\varepsilon > 0$. Since f is bounded we can find $l \in \mathbb{N}$ and a finite sequence $a_0 \leq \dots \leq a_l$ with $a_0 < \inf_{x \in C} f(x) \leq \sup_{x \in C} f(x) < a_k$ and $|a_k - a_{k-1}| \leq \varepsilon$ for $k \in \{1, \dots, l\}$. Furthermore since the sets $f^{-1}(\{a\}) \subseteq C$ are disjoint for $a \in \mathbb{R}$ we can choose the a_k such that $\mu(f^{-1}(\{a_k\})) = 0$ for all $k \in \{0, \dots, l\}$. We now define $f_\varepsilon(x) := a_k$ for $x \in C$ with $f(x) \in (a_{k-1}, a_k]$. Note that

$$\partial\mu(f^{-1}((a_{k-1}, a_k])) \subseteq f^{-1}(\{a_{k-1}, a_k\}) \cup N_f$$

is a μ -nullset. Thus we have due to 2. for $n \rightarrow \infty$:

$$\begin{aligned} \int f_\varepsilon d\mu_n &= \sum_{k=1}^l a_k \mu_n(f^{-1}((a_{k-1}, a_k])) \\ &\rightarrow \sum_{k=1}^l a_k \mu(f^{-1}((a_{k-1}, a_k])) = \int f_\varepsilon d\mu. \end{aligned}$$

Therefore, we have

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \left| \int f d\mu_n - \int f d\mu \right| \\ &\leq \limsup_{n \rightarrow \infty} \int |f_\varepsilon - f| d\mu_n + \int |f_\varepsilon - f| d\mu + \left| \int f_\varepsilon d\mu_n - \int f_\varepsilon d\mu \right| \\ &\leq \limsup_{n \rightarrow \infty} \varepsilon(\mu_n(C) + \mu(C)) + \left| \int f_\varepsilon d\mu_n - \int f_\varepsilon d\mu \right| \\ &= 2\mu(C)\varepsilon. \end{aligned}$$

We conclude the claim for $\varepsilon \rightarrow 0$.

It is left to show that it suffices to consider C with $\mu(\partial C) = 0$. Since E is locally compact we have that the ε -extension $C^\varepsilon := \{x \in E \mid d(x, C) \leq \varepsilon\}$ is compact for ε sufficiently small. Furthermore, the boundaries $\partial C^\varepsilon \subseteq \{x \in E \mid d(x, C) = \varepsilon\}$ are disjoint for different ε . Thus there exists an $\varepsilon > 0$ sufficiently small with C^ε compact and $\mu(\partial C^\varepsilon) = 0$.

5. The pointwise convergence of measure generating functions in case of vague convergence of the measures is a direct consequence of 2.

Now we show that convergence of the measure generating functions implies vague convergence of the measures. We have for $s \leq t$ with s, t continuity points of μ :

$$\mu_n((s, t]) = F_n(t) - F_n(s) \rightarrow F(t) - F(s) = \mu((s, t]).$$

We use a standard approximation argument to show convergence of integrals over continuous functions with compact support. Consider $f \in C_c(\mathbb{R}_+)$ supported on a compact set $[0, T]$ with $\mu(\{T\}) = 0$. Since f is uniformly continuous we can approximate it uniformly with a sequence $f_k \rightarrow f$ where f_k is piecewise constant on intervals for which μ has no mass on the boundary. The above convergence on half open intervals yields

$$\int f_k \, d\mu_n \rightarrow \int f_k \, d\mu \quad \text{for } n \rightarrow \infty.$$

The argument is from now on a similar approximation to the previous part of the proof. Let $\varepsilon > 0$ and $k \in \mathbb{N}$ be large enough such that $\|f - f_k\|_\infty \leq \varepsilon$. We have

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \left| \int f \, d\mu_n - \int f \, d\mu \right| \\ & \leq \limsup_{n \rightarrow \infty} \int |f - f_k| \, d\mu_n + \left| \int f_k \, d\mu_n - \int f_k \, d\mu \right| + \int |f - f_k| \, d\mu \\ & \leq \limsup_{n \rightarrow \infty} \|f - f_k\|_\infty (\mu_n((0, T]) + \mu((0, T])) + \left| \int f_k \, d\mu_n - \int f_k \, d\mu \right| \\ & \leq \limsup_{n \rightarrow \infty} \varepsilon (F_n(T) - F_n(0) + F(T) - F(0)) + \left| \int f_k \, d\mu_n - \int f_k \, d\mu \right| \\ & = 2\varepsilon(2F(T) - 2F(0)). \end{aligned}$$

For $\varepsilon \rightarrow 0$ this yields the claim. □

4.2 Convergence of Integer Valued Random Measures

We will need conditions for weak convergence of integer valued random measures with respect to the vague topology. This section will show this convergence using the results of Kallenberg [16] about weak convergence in the vague topology. At the end of this section we will also present an alternative approach for showing the convergence using more modern techniques involving Wasserstein metrics.

Definition 4.11. Let $\mu \in \mathcal{M}_c(E)$ be an integer valued random measure on an lscH space E . This means that $\mu(B) \in \mathbb{N}_0 \cup \{\infty\}$ for all $B \in \mathcal{B}(E)$ almost surely. Note that μ is necessarily almost surely atomic. We call μ simple if and only if

$$\mathbb{P}(\text{Each atom of } \mu \text{ has mass } 1) = 1.$$

Theorem 4.12 (Convergence of Integer Valued Random Measures). *Let $(\mu_n)_{n \in \mathbb{N}} \subseteq \mathcal{M}_c(E)$ be a sequence of integer valued random measures on E and $\mu \in \mathcal{M}_c(E)$ a simple integer valued random measure. Let for a separating set \mathcal{U} :*

1. $\lim_{n \rightarrow \infty} \mathbb{P}(\mu_n(U) = 0) = \mathbb{P}(\mu(U) = 0)$ for all $U \in \mathcal{U}$.
2. $\limsup_{n \rightarrow \infty} \mathbb{E}(\mu_n(K)) \leq \mathbb{E}(\mu(K)) < \infty$ for all $K \subseteq E$ compact.

Then $\mu_n \rightarrow \mu$ weakly with respect to the vague topology on $\mathcal{M}_c(E)$.

Proof. This is Proposition 14.17 in [16]. □

It is well known that a Binomial distribution $\text{Bin}(n, p)$ for $n \rightarrow \infty$ and $np \rightarrow \lambda \geq 0$ converges weakly to a Poisson distribution $\text{Pois}(\lambda)$. This result is sometimes called the *law of small numbers*. The above results allow us to show a similar result for random measures. Instead of the Binomial distribution we consider so called *Bernoulli processes* and the Poisson distribution is replaced with a *Poisson point process*.

Proposition 4.13. *Let $(c_n)_{n \in \mathbb{N}}$ be a sequence of positive real numbers with $c_n \rightarrow 0$ for $n \rightarrow \infty$. For $n \in \mathbb{N}$ let $p_n \in [0, 1]$ such that $p_n/c_n \rightarrow \lambda \geq 0$ for $n \rightarrow \infty$. Let $(B_n^{(m)})_{m \in \mathbb{Z}}$ be an i.i.d. sequence of $\text{Bin}(1, p_n)$ distributed random variables. We define an integer valued random measure on \mathbb{R} , called Bernoulli process, via*

$$\mu_n := \sum_{m \in \mathbb{Z}} B_n^{(m)} \delta_{m \cdot c_n}.$$

Then $\mu_n \rightarrow \mu$ weakly, where μ is the Poisson point process on \mathbb{R} with constant intensity λ .

Proof. We check the conditions in Theorem 4.12. Since the intensity measure given by the constant density $t \mapsto \lambda$ is continuous we have that the Poisson point process μ almost surely has only atoms of mass 1. Let \mathcal{U} be the set of finite, disjoint unions of half open intervals in \mathbb{R} . We have seen in Proposition 4.9 that \mathcal{U} is indeed separating.

1. We show the result for a single half open interval $(a, b]$ with $a < b$. We have:

$$\mu_n((a, b]) = \sum_{m=[a/c_n]+1}^{[b/c_n]} B_n^{(m)} \sim \text{Bin}([b/c_n] - [a/c_n], p_n)$$

Now the law of small numbers implies $\mu_n((a, b]) \rightarrow \text{Pois}(\lambda(b - a))$ weakly and in particular $\lim_{n \rightarrow \infty} \mathbb{P}(\mu_n((a, b]) = 0) = \mathbb{P}(\mu((a, b]) = 0)$. We can then use independence of μ_n over disjoint sets (since the $B_n^{(m)}$ are independent over m) and independence of the Poisson point process over disjoint sets to extend the result to disjoint unions of half-open intervals.

2. Let $K \subseteq \mathbb{R}$ be compact and $\varepsilon > 0$. Due to the continuity of the Lebesgue measure $\text{Vol}(\bullet)$ from above and using compactness of K we can find a finite cover of K with open intervals $\{(a_k, b_k)\}_{k=1}^l$ such that

$$\sum_{k=1}^l \text{Vol}((a_k, b_k)) \leq \text{Vol}(K) + \varepsilon.$$

We have already seen in the first part that

$$\mu_n((a_k, b_k)) \rightarrow \text{Pois}(\lambda(b_k - a_k)) \quad \text{weakly for } n \rightarrow \infty.$$

In fact, the convergence also holds for the expectations. Thus we have

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbb{E}(\mu_n(K)) &\leq \limsup_{n \rightarrow \infty} \sum_{k=1}^l \mathbb{E}(\mu_n((a_k, b_k))) \\ &= \sum_{k=1}^l \mathbb{E}(\text{Pois}(\lambda(b_k - a_k))) \\ &= \lambda \sum_{k=1}^l \text{Vol}((a_k, b_k)) \\ &\leq \lambda(\text{Vol}(K) + \varepsilon) = \mathbb{E}(\mu(K)) + \lambda\varepsilon. \end{aligned}$$

Letting $\varepsilon \rightarrow 0$ yields $\limsup_{n \rightarrow \infty} \mathbb{E}(\mu_n(K)) \leq \mathbb{E}(\mu(K))$. Thus the assumptions of Theorem 4.12 are fulfilled and the calim follows.

□

In the above proposition we assume that the possible locations for atoms of μ_n are on an equidistant lattice with mesh size c_n . (In other words we rescale time with a global factor $1/c_n$). We can even generalize this to allow for non-equidistant meshes. Furthermore in the above proposition the probabilities p_n attached to each point in the mesh did not depend on the location of the point resulting in a time-homogeneous situation. We also want to generalize this to allow for time-inhomogeneous setups. We first need some notation.

Definition 4.14. For $n \in \mathbb{N}$ let $A^{(n)}: \mathbb{Z} \rightarrow \mathbb{R}, z \mapsto A_z^{(n)}$ be a strictly increasing map such that $A_z^{(n)} \rightarrow \pm\infty$ for $z \rightarrow \pm\infty$. We may also refer to this map as a *mesh* in \mathbb{R} . Define $c_{n,z} := A_z^{(n)} - A_{z-1}^{(n)} > 0$ the *mesh size*. Let $p_z^{(n)} \in [0, 1]$ for $n \in \mathbb{N}, z \in \mathbb{Z}$. Consider an independent (in $z \in \mathbb{Z}$) sequence of $\text{Bin}(1, p_z^{(n)})$ distributed random variables $(B_z^{(n)})_{n \in \mathbb{N}}$ and define the Bernoulli process μ_n on \mathbb{R} via

$$\mu^{(n)} := \sum_{z \in \mathbb{Z}} B_z^{(n)} \delta_{A_z^{(n)}}$$

Furthermore let $\nu^{(n)}$ be the measure on \mathbb{R} given by

$$\nu^{(n)} = \sum_{z \in \mathbb{Z}} p_z^{(n)} \delta_{A_z^{(n)}}.$$

We call $\nu^{(n)}$ the *intensity measure or expectation measure* of $\mu^{(n)}$. Note that for $s < t$ we have

$$\nu^{(n)}((s, t]) = \mathbb{E}(\mu^{(n)}((s, t])).$$

We start by showing some general result for the above objects.

Lemma 4.15. Consider the situation of Definition 4.14. For $s < t$ and $n \in \mathbb{N}$ define

$$p_{\max}^{(n)}((s, t]) := \max\{p_z^{(n)} \mid A_z^{(n)} \in (s, t]\}.$$

Now let $O \subseteq \mathbb{R}$ be an open set. Assume that $\nu_{|O}^{(n)} \xrightarrow{v} \nu \in \mathcal{M}_c(O)$ for $n \rightarrow \infty$ and ν a continuous measure. Then for all $s < t$ with $[s, t] \subseteq O$ for $n \rightarrow \infty$ we have $p_{\max}^{(n)}((s, t]) \rightarrow 0$.

Proof. Let $s < t$ with $[s, t] \subseteq O$. The vague convergence $\nu_{|O}^{(n)} \xrightarrow{v} \nu$ implies $\nu^{(n)}((s, t]) \rightarrow \nu((s, t])$ (see 2. in Theorem 4.10). We show $p_{\max}^{(n)}((s, t]) \rightarrow 0$ by contradiction. If the convergence does not hold, then there exists an $\varepsilon > 0$ and a subsequence $(n_k)_{k \in \mathbb{N}}$ and for each k a $z_k \in \mathbb{Z}$ with $A_{z_k}^{(n_k)} \in (s, t]$ such that $p_{z_k}^{(n_k)} > \varepsilon$ for all $k \in \mathbb{N}$. Since $A_{z_k}^{(n_k)} \in [s, t]$ for all $n \in \mathbb{N}$ we can choose a further

subsequence such that $A_{z_k}^{(n_k)}$ converges to some $u \in [s, t]$ for $k \rightarrow \infty$. Note that $u \in [s, t] \subseteq O$ with O open. Thus we find $\delta > 0$ such that $[u - \delta, u + \delta] \subseteq O$. For large k the set $[A_{z_k}^{(n_k)} - \delta/2, A_{z_k}^{(n_k)} + \delta/2]$ is contained in $(u - \delta, u + \delta)$ and $p_{z_k}^{(n_k)} > \varepsilon$ for all $k \in \mathbb{N}_+$. Thus we get for $k \rightarrow \infty$:

$$\varepsilon < p_{z_k}^{(n_k)} \leq \nu^{(n_k)}((u - \delta, u + \delta)) \rightarrow \nu((u - \delta, u + \delta)).$$

This implies $\nu((u - \delta, u + \delta)) > \varepsilon$ for all $\delta > 0$ and thus $\nu(\{u\}) \geq \varepsilon > 0$, which is a contradiction to the continuity of ν . \square

We can now show the generalization of Proposition 4.13 for a possibly time-inhomogeneous time-rescaling and time-inhomogeneous probabilities.

Proposition 4.16. *We consider the situation of Lemma 4.15 and thus in particular of Definition 4.14. Let μ be a Poisson point process on O with intensity measure ν .*

Then we have $\mu^{(n)} \rightarrow \mu$ weakly in the vague topology.

Proof. Since ν is a continuous measure we have

$$\mathbb{P}(\text{Each atom of } \mu \text{ has mass } 1) = 1.$$

Hence it suffices to check the conditions of Theorem 4.12. Our separating set will be the set of all disjoint finite unions of half open intervals which closures are contained in O . Note that due to the results in Proposition 4.9 this set is indeed separating. We proceed showing the remaining conditions of Theorem 4.12.

1. It suffices to conclude the convergence of the probabilities for half open intervals $U := (s, t] \subseteq O$ since both the random measures μ_n and the random measure μ are independent over disjoint intervals and thus equality for half open intervals yields equality for finite disjoint unions of half open intervals. We have, using the definition of $\mu^{(n)}$:

$$\mathbb{P}(\mu^{(n)}(U) = 0) = \prod_{A_z^{(n)} \in U} (1 - p_z^{(n)}).$$

We already know from Lemma 4.15 that the probabilities $1 - p_z^{(n)}$ in the product converge to 1 for $n \rightarrow \infty$ uniformly in z . In particular we have for large n that all the probabilities are positive. Thus applying

the mean value theorem to $\frac{d}{dx} \log(x) = 1/x$ yields:

$$\begin{aligned} -\log\left(\mathbb{P}(\mu^{(n)}(U) = 0)\right) &= \sum_{A_z^{(n)} \in U} -\log(1 - p_z^{(n)}) \\ &= \sum_{A_z^{(n)} \in U} p_z^{(n)} \frac{\log(1) - \log(1 - p_z^{(n)})}{p_z^{(n)}} \\ &= \sum_{A_z^{(n)} \in U} p_z^{(n)} \frac{1}{\xi_z^{(n)}} \end{aligned}$$

for some value $\xi_z^{(n)} \in [1 - p_z^{(n)}, 1]$. As noted before Lemma 4.15 yields $p_z^{(n)} \rightarrow 0$ uniformly in z for $n \rightarrow \infty$ and thus $1/\xi_z^{(n)} \rightarrow 1$ uniformly in z for $n \rightarrow \infty$. Furthermore, we have due to $\mathbb{E}(B_z^{(n)}) = p_z^{(n)}$ and 2. of Theorem 4.10 that for $n \rightarrow \infty$:

$$\mathbb{E}\left(\mu^{(n)}((s, t])\right) = \sum_{A_z^{(n)} \in (s, t]} p_z^{(n)} = \nu^{(n)}(s, t] \rightarrow \nu((s, t]). \quad (4.5)$$

Thus we get

$$-\log\left(\mathbb{P}(\mu^{(n)}(U) = 0)\right) \rightarrow \nu((s, t]) \quad \text{for } n \rightarrow \infty.$$

Multiplying both sides with -1 and applying the exponential function gives

$$\mathbb{P}(\mu^{(n)}(U) = 0) \rightarrow \exp\left(-\nu((s, t])\right) = \mathbb{P}(\mu(U) = 0) \quad \text{for } n \rightarrow \infty.$$

2. Let $C \subseteq O$ be compact and consider the Euclidean metric d on O . Also, let $\varepsilon > 0$. We can find $\delta > 0$ such that

$$C^\delta := \{x \in O \times E_\infty \mid d(x, C) < \delta\}$$

is relatively compact with $\nu(\partial C^\delta) = 0$ such that $\nu(C^\delta) - \nu(C) < \varepsilon$ (for the last step we use the continuity of the measure ν from above). We have, since $\nu_n \xrightarrow{v} \nu$ on O (see Theorem 4.10):

$$\limsup_{n \rightarrow \infty} \mathbb{E}(\mu_n(C)) \leq \lim_{n \rightarrow \infty} \mathbb{E}(\mu_n(C^\delta)) = \lim_{n \rightarrow \infty} \nu_n(C^\delta) = \nu(C^\delta) \leq \nu(C) + \varepsilon.$$

Now we letting $\varepsilon \rightarrow 0$ shows the second condition of Theorem 4.12. \square

Remark 4.17.

1. Note that the vague convergence of the measures $\nu^{(n)} \xrightarrow{v} \nu$ holds if and only if

$$\nu^{(n)}((s, t]) \rightarrow \nu((s, t])$$

for all $[s, t] \subseteq O$. To show vague convergence here we can use the same approximation argument as in the proof of 5. in Theorem 4.10 on each connected component of O . This suffices since every compact set $C \subseteq O$ is only contained in finitely many of the connected components of O . (The connected components form an open cover of C .) Moreover, it even suffices to check the convergence for $[s, t] \subseteq O$ with $s, t \in \mathbb{Q}$ by using monotonicity of the measures $\nu^{(n)}, \nu$ and continuity of ν .

2. Note that we consider open subsets of \mathbb{R} which is not the same as starting with \mathbb{R} and then restricting to an open subset. Notably the limiting measure ν does not need to be finite on bounded subsets of O but it has to be finite on compact subsets of O as a limit in the vague topology. If we consider closed subspaces this problem does not occur since vague convergence of measures on a closed subspace is equivalent to vague convergence of the measures extended to the original space (by setting them equal to zero on the complement of the open subspace). Thus it is not necessary to restrict to closed subsets (or here in general intersections of open and closed subsets of \mathbb{R}). This is in particular noteworthy since we will later consider \mathbb{R}_+ instead of \mathbb{R} and open subspaces of \mathbb{R}_+ .

We now want to generalize the previous result to a discrete, spatial situation. First we need to extend Definition 4.14 to a spatial situation.

Definition 4.18. Let E be a countable set and let $E_\infty := E \cup \{\infty\}$ be the one-point compactification of E if E is infinite and equal E if E is finite. For $n \in \mathbb{N}_0$ and $x \in E$ consider a mesh $(A_{z,x}^{(n)})_{z \in \mathbb{Z}}$ in \mathbb{R} as defined in Definition 4.14. For $n \in \mathbb{N}_0$, $z \in \mathbb{Z}$ and $x \in E$ let $X_{z,x}^{(n)}$ be E valued random variables independent in z . Define for $z \in \mathbb{Z}$ and $x, y \in E$ the probability $p_{z,x,y}^{(n)} := \mathbb{P}(X_{z,x}^{(n)} = y)$. We now define the *Bernoulli process* $\mu^{(n)}$ on $\mathbb{R} \times E \times E_\infty$ via

$$\mu^{(n)} := \sum_{z \in \mathbb{Z}} \sum_{x \in E} \sum_{y \in E \setminus \{x\}} \mathbb{1}_{\{X_{z,x}^{(n)} = y\}} \delta_{(A_{z,x}^{(n)}, x, y)}.$$

Furthermore, we define the *intensity measure (or expectation measure)* $\nu^{(n)}$ of $\mu^{(n)}$ as the following discrete measure on $\mathbb{R} \times E \times E_\infty$:

$$\nu^{(n)} := \sum_{z \in \mathbb{Z}} \sum_{x \in E} \sum_{y \in E \setminus \{x\}} p_{z,x,y}^{(n)} \delta_{(A_{z,x}^{(n)}, x, y)}.$$

Analogous to the nonspatial case we have for $s < t$:

$$\nu^{(n)}((s, t] \times \{x\} \times \{y\}) = \mathbb{E}\left(\mu^{(n)}((s, t] \times \{x\} \times \{y\})\right).$$

Proposition 4.19. *Consider the situation of Definition 4.18. Let $O \subseteq \mathbb{R} \times E$ be open. Let $\nu_{|_{O \times E_\infty}}^{(n)} \rightarrow \nu$ in the vague topology for measures on $O \times E_\infty$ with $\nu \in \mathcal{M}_c(O \times E_\infty)$ continuous. Then $\mu_n \rightarrow \mu$ weakly with respect to the vague topology, where μ is the Poisson point process on $O \times E_\infty$ with intensity measure ν . In particular we have $\nu(O \times \{\infty\}) = 0$.*

Proof. We show that the assumptions of Theorem 4.12 are fulfilled. Let $x \neq y \in E$. Since its intensity measure ν is continuous the Poisson point process ν is simple.

1. Let $O_x := \{t \in \mathbb{R} | (t, x) \in O\}$ be the x -fiber of O for a given $x \in E$. Since O was open in $\mathbb{R} \times E$ we have that O_x is an open subset of \mathbb{R} . We use 5. of Proposition 4.9 and 3. of Proposition 4.9 to see that sets of the form

$$\sum_{x \in F^{(1)}} \sum_{k=1}^n (((t_{k-1}, t_k] \cap O_x) \times \{x\}) \times F_k^{(2)}$$

with $n \in \mathbb{N}$, $t_0 \leq \dots \leq t_n$, $F^{(1)} \subseteq E$ finite, $F_k^{(2)} \subseteq E_\infty$ yield a separating class in $O \times E_\infty$. Note that $\mu_n(\mathbb{R} \times E \times \{\infty\}) = 0$ by definition. Since μ is independent over disjoint sets and μ_n is independent over disjoint sets in time and the first space component (the $X_{z,x}^{(n)}$ are independent in z) it suffices to check the first assumption of Theorem 4.12 for sets of the form $((a, b] \cap O_x) \times \{x\} \times F$ for $a < b$, $x \in E$, $F \subseteq E \setminus \{x\}$. We can now define the random measure $\tilde{\mu}_n$ on O_x induced by the projection $(O_x \times \{x\}) \times F \rightarrow O_x$ applied to μ_n and analogously we define $\tilde{\mu}$ as the measure induced by μ in this manner. Furthermore, we note that the intensity measures $\tilde{\nu}^{(n)}$ and $\tilde{\nu}$ of these Bernoulli processes are also the measures induced by the projection and the intensity measures of μ_n and μ .

Note that since E_∞ is compact we have $\tilde{\mu}_n \in \mathcal{M}_c(O_x)$ and it is given by

$$\tilde{\mu}_n = \sum_{z \in \mathbb{Z}} \mathbb{1}_{\{X_{z,x}^{(n)} \in F\}} \mathbb{1}_{O_x}(A_{z,x}^{(n)}) \delta_{A_{z,x}^{(n)}}.$$

We note that the family $(\mathbb{1}_{\{X_{z,x}^{(n)} \in F\}})_{z \in \mathbb{Z}}$ is an independent family of random variables with $\mathbb{1}_{\{X_{z,x}^{(n)} \in F\}} \sim \text{Bin}(1, \sum_{y \in F} p_{z,x,y}^{(n)})$. Furthermore,

since $\nu_{|_{O \times E_\infty}}^{(n)} \xrightarrow{v} \nu$ we have for $C \subseteq O$ compact for $n \rightarrow \infty$ (using 2. of Theorem 4.10):

$$\tilde{\nu}^{(n)}(C) = \nu^{(n)}(C \times \{x\} \times F) \rightarrow \nu(C \times \{x\} \times F) = \tilde{\nu}(C).$$

Hence, the intensity measures of $\tilde{\mu}_n$ also converges vaguely to the intensity measure of $\tilde{\mu}$. Thus, we can use our nonspatial result in Proposition 4.16 to get $\tilde{\mu}_n \rightarrow \tilde{\mu}$ weakly in the vague topology where $\tilde{\mu}$ is a Poisson point process on O_x with intensity measure $\tilde{\nu}$. In particular we had shown the following convergence for $n \rightarrow \infty$ in the proof of Proposition 4.16:

$$\begin{aligned} \mathbb{P}(\mu_n(((a, b] \cap O_x) \times \{x\}) \times F) = 0) \\ &= \mathbb{P}(\tilde{\mu}_n((a, b]) = 0) \\ &\rightarrow \mathbb{P}(\tilde{\mu}((a, b]) = 0) \\ &= \mathbb{P}(\mu(((a, b] \cap O_x) \times \{x\}) \times F) = 0). \end{aligned}$$

2. The argument is analogous to the nonspatial case. Let $C \subseteq O \times E_\infty$ be compact and choose a suitable metric d on $O \times E_\infty$. Let $\varepsilon > 0$. We use continuity from above of ν to find $\delta > 0$ such that

$$C^\delta := \{x \in O \times E_\infty \mid d(x, C) < \delta\}$$

is relatively compact with $\nu(\partial C^\delta) = 0$ and $\nu(C^\delta) - \nu(C) < \varepsilon$. We have since $\nu_n \xrightarrow{v} \nu$ on $O \times E_\infty$ (see Theorem 4.10):

$$\limsup_{n \rightarrow \infty} \mathbb{E}(\mu_n(C)) \leq \lim_{n \rightarrow \infty} \mathbb{E}(\mu_n(C^\delta)) = \lim_{n \rightarrow \infty} \nu_n(C^\delta) = \nu(C^\delta) \leq \nu(C) + \varepsilon.$$

Letting $\varepsilon \rightarrow 0$ shows the second condition of Theorem 4.12.

□

Using the previous proposition we can show weak convergence of integer valued random measures in good situations by showing vague convergence of their intensity measures. The following proposition will show that if we consider two sequences of intensity measures which describe independent events and which converge vaguely then the sequence of intensity measures which describes the simultaneous occurrence of the two events will converge to zero. In other words independent events can not occur simultaneously in the limit.

Proposition 4.20. *Consider two sequences of intensity measures $(\nu_1^{(n)})_{n \in \mathbb{N}}$ and $(\nu_2^{(n)})_{n \in \mathbb{N}}$ on $\mathbb{R} \times E \times E_\infty$ as in Definition 4.18 using the same mesh $(A_{z,x}^{(n)})_{z \in \mathbb{Z}}$. In particular let*

$$\nu_1^{(n)} = \sum_{z \in \mathbb{Z}} \sum_{x \in E} \sum_{y \in E \setminus \{x\}} p_{z,x,y}^{(n)} \delta_{(A_{z,x}^{(n)}, x, y)}$$

and

$$\nu_2^{(n)} = \sum_{z \in \mathbb{Z}} \sum_{x \in E} \sum_{y \in E \setminus \{x\}} q_{z,x,y}^{(n)} \delta_{(A_{z,x}^{(n)}, x, y)}.$$

We can define a third sequence of intensity measures $(\nu_3^{(n)})_{n \in \mathbb{N}}$:

$$\nu_3^{(n)} := \sum_{z \in \mathbb{Z}} \sum_{x \in E} \sum_{y \in E \setminus \{x\}} p_{z,x,y}^{(n)} q_{z,x,y}^{(n)} \delta_{(A_{z,x}^{(n)}, x, y)}.$$

Let $O \subset \mathbb{R} \times E$ be open and restrict the measures to $O \times E_\infty$. If there are $\nu_1, \nu_2 \in \mathcal{M}_c(O \times E_\infty)$ continuous, such that $\nu_1^{(n)} \rightarrow \nu_1$ as well as $\nu_2^{(n)} \rightarrow \nu_2$ in the vague topology, then $\nu_3^{(n)} \rightarrow 0$ in the vague topology.

In particular, if we have independent Bernoulli processes $\mu_1^{(n)}, \mu_2^{(n)}$ with intensities $\nu_1^{(n)}, \nu_2^{(n)}$ then the Bernoulli process which has atoms exactly at locations where $\mu_1^{(n)}$ and $\mu_2^{(n)}$ have both an atom will have the intensity measure $\nu_3^{(n)}$. In the case of vague convergence of intensity measures this Bernoulli process will converge to 0 weakly with respect to the vague topology.

Proof. Let $x \in E$ and $F \subseteq E_\infty$. Consider the set $O_x := \{u \in \mathbb{R} \mid (u, x) \in O\}$ and let $[s, t] \subseteq O_x$ be a compact interval. It holds:

$$\begin{aligned} \nu_3^{(n)}((s, t] \times \{x\} \times F) &= \sum_{z \in \mathbb{Z}} \mathbf{1}_{(s,t]}(A_{z,x}^{(n)}) \sum_{y \in F} p_{z,x,y}^{(n)} q_{z,x,y}^{(n)} \\ &\leq p_{\max}^{(n)}((s, t]) \sum_{z \in \mathbb{Z}} \mathbf{1}_{(s,t]}(A_{z,x}^{(n)}) \sum_{y \in F} q_{z,x,y}^{(n)} \\ &= p_{\max}^{(n)}((s, t]) \nu_2^{(n)}((s, t] \times \{x\} \times F). \end{aligned}$$

where we set

$$p_{\max}^{(n)}((s, t]) := \max \left\{ \sum_{y \in E_\infty} p_{z,x,y}^{(n)} \mid A_{z,x}^{(n)} \in (s, t] \right\}.$$

Applying Lemma 4.15 to the measure (on the open set $O_x \subseteq \mathbb{R}$) given by $B \mapsto \nu_1^{(n)}(B \times \{x\} \times E_\infty)$ we get $p_{\max}^{(n)}((s, t]) \rightarrow 0$ for $n \rightarrow \infty$ and using the vague convergence of $\nu_2^{(n)}$ we get

$$\nu_2^{(n)}((s, t] \times \{x\} \times F) \rightarrow \nu_2((s, t] \times \{x\} \times F) < \infty.$$

Thus we have for $n \rightarrow \infty$:

$$\nu_3^{(n)}((s, t] \times \{x\} \times F) \rightarrow 0. \quad (4.6)$$

Let $C \subseteq O \times E_\infty$ be compact. By 5. of Proposition 4.9 we can separate C and O with a finite union of sets of the type $(s, t] \times \{x\} \times F$. Thus (4.6) implies $\nu_3^{(n)}(C) \rightarrow 0$ for all C compact (which implies the same for all relatively compact sets) and thus $\nu_3^{(n)} \xrightarrow{v} 0$ using 2. of Theorem 4.10. \square

We will now give an alternative proof for Proposition 4.19 using Wasserstein metrics which was suggested to us by Dominic Schumacher. We start by restricting our point processes and their intensity measures to a compact set $C \times F \times E_\infty \subseteq O \times E_\infty$ with $F \subseteq E$ finite. Furthermore since the point processes are independent on disjoint time intervals and since C is contained in finitely many connected components of O we can assume without loss of generality $C = [a, b] \subseteq O$ with a, b not equal to $A_{z,x}^{(n)}$ for all $z \in \mathbb{Z}$, $x \in E$, $n \in \mathbb{N}$. Our goal is to show weak convergence of the integer valued random measures $\mu_n \rightarrow \mu$ restricted to $[a, b] \times F \times E_\infty$ with respect to the weak topology. We start by considering the Wasserstein metrics d_2 and \bar{d}_2 . Since the definitions of d_2 and \bar{d}_2 are a bit lengthy and we do not actually utilise the definitions in our arguments we will only provide a reference where the metrics are defined (see [26]). It holds $\bar{d}_2 \leq d_2$ and \bar{d}_2 metricizes weak convergence (see Proposition 2.3. (ii) and (iii) in [26]). We consider the Bernoulli process μ_n with intensity measure ν_n and the Poisson point process μ with intensity measure ν as given in Definition 4.18. Furthermore let $\hat{\mu}_n$ be a Poisson point process with intensity measure ν_n . We have

$$\bar{d}_2(\mu_n, \mu) \leq \bar{d}_2(\mu_n, \hat{\mu}_n) + \bar{d}_2(\hat{\mu}_n, \mu) \leq d_2(\mu_n, \hat{\mu}_n) + \bar{d}_2(\hat{\mu}_n, \mu). \quad (4.7)$$

For the first summand we want to apply Theorem 10.F in [1]. We first need some notation, we set:

$$\alpha := (A_{z,x}^{(n)}, x, y) \in [a, b] \times F \times E_\infty, \quad I_\alpha := \mathbb{1}_{\{X_{z,x}^{(n)}=y\}}, \quad p_\alpha := \mathbb{E}(I_\alpha).$$

We define the index sets

$$\Gamma := \{\alpha | x \in F, y \in E_\infty \setminus \{x\}, z \in \mathbb{Z} \text{ with } A_{z,x}^{(n)} \in [a, b]\}$$

and

$$\Gamma_{(t,x,y)}^s := \{(t, x, w) \in \Gamma | w \in E_\infty \setminus \{x, y\}\}, \quad \Gamma_\alpha^w := \Gamma \setminus \Gamma_\alpha^s.$$

We define the random variables

$$Z_\alpha := \sum_{\beta \in \Gamma_\alpha^s} I_\beta$$

and note that

$$\mathbb{E}(Z_\alpha) = \sum_{w \in E_\infty \setminus \{x, y\}} p_{z, x, w}^n, \quad Z_\alpha I_\alpha = 0.$$

Proposition 4.21. *We can estimate the d_2 distance between the Bernoulli process μ_n and the Poisson point process $\hat{\mu}_n$ in the following fashion:*

$$d_2(\mu_n, \hat{\mu}_n) \leq \sum_{\alpha \in \Gamma} ((p_\alpha)^2 + p_\alpha \mathbb{E}(Z_\alpha) + \mathbb{E}(I_\alpha Z_\alpha)). \quad (4.8)$$

Proof. This is a consequence of Theorem 10.F in [1]. Note that our right-hand side is essentially the first summand in the right-hand side of the Theorem. Also, the η_α which show up in the second summand in the Theorem are equal zero since I_α is independent of $\{I_\beta | \beta \in \Gamma_\alpha^w\}$. Furthermore since the intensity measures of μ_n and $\hat{\mu}_n$ coincide the last summand in the Theorem also vanishes. Furthermore, we can estimate the coefficient in front of the first summand in the Theorem by 1 which yields the claim. \square

Now we use

$$p_\alpha = \nu_n(\{A_{z, x}^{(n)}\} \times \{x\} \times \{y\})$$

and

$$\mathbb{E}(Z_\alpha) = \nu_n(\{A_{z, x}^{(n)}\} \times \{x\} \times E_\infty \setminus \{x, y\})$$

to estimate:

$$\begin{aligned} & \left(\sum_{\alpha \in \Gamma} (p_\alpha)^2 + p_\alpha \mathbb{E}(Z_\alpha) + \mathbb{E}(I_\alpha Z_\alpha) \right) \\ &= \left(\sum_{\alpha \in \Gamma} (p_\alpha)^2 + p_\alpha \mathbb{E}(Z_\alpha) \right) \\ &\leq \left(\max_{A_{z, x}^{(n)} \in [a, b], x \in E} \sum_{y \in E_\infty \setminus \{x\}} p_{z, x, y}^{(n)} \right) 2\nu_n([a, b] \times F \times E_\infty). \end{aligned}$$

The maximum over the probabilities converges to zero due to Lemma 4.15 and due to continuity of ν the second factor converges to $\nu([a, b] \times F \times E_\infty)$. Thus the the right hand side of (4.8) and therefore the right hand side of (4.7) converges to zero. For the second summand we first note that due to independence in a Poisson point process we can further restrict to the set $[a, b] \times \{x\} \times \{y\}$ for given $x \in E$ and $y \in E_\infty \setminus \{x\}$. In that case the Poisson point processes can be interpreted as Poisson point processes $\hat{\mu}_n$ and μ on $[a, b]$ and their intensity measures ν_n and ν yield measure generating functions F_n and F on $[a, b]$ with $F_n(a) = 0 = F(a)$.

Proposition 4.22. *We can estimate the \bar{d}_2 distance between the Poisson point processes $\hat{\mu}_n$ and μ in the following way:*

$$\bar{d}_2(\hat{\mu}_n, \mu) \leq \int_a^b |F_n(s) - F(s)| ds + |F_n(b) - F(b)|. \quad (4.9)$$

Proof. This is Lemma B.1. in [25]. □

Due to $\nu_n \xrightarrow{v} \nu$ and using 5. of Theorem 4.10 we get $F_n(s) \rightarrow F(s)$ for $n \rightarrow \infty$ for all $s \in [a, b]$. Together with dominated convergence this implies that the right-hand side of (4.9) converges to zero and thus the second summand on the right-hand side in (4.7) converges to zero. This implies weak convergence with respect to the weak topology. Since weak convergence implies vague convergence and since it suffices to check vague convergence on arbitrarily large compact sets (using 2. of Theorem 4.10) we can conclude vague convergence on the whole space.

Chapter 5

The Graphical Representation and the Tracking Map

5.1 The Space of Graphical Representations

Definition 5.1 (The Space of Graphical Representations). Let E be a countable set and $S \subseteq \mathbb{R}_+ \times E$ an open set. Let $E_\infty = E \cup \{\infty\}$ the one-point compactification if E is infinite and $E_\infty = E$ if E is finite. Also, let $S_T := S \cap [0, T] \times E$ for $T \geq 0$. We define REP (short for representation) to be the space of all $\mathbb{N}_0 \cap \{\infty\}$ valued measures on $S \times E_\infty$ which are finite on compact sets. Any $\eta \in \text{REP}$ can be written as

$$\eta = \sum_{j=1}^n \delta_{(t_j, x_j, y_j)}$$

with $n \in \mathbb{N}_0 \cup \{\infty\}$, $(t_j, x_j) \in S$ and $y_j \in E_\infty$ for $j \in [n]$. We call $\eta \in \text{REP}$ a *graphical representation on E with admissible states S* . For $(t, x, y) \in S \times E_\infty$ such that η has an atom at (t, x, y) we say that there is an *arrow pointing from x to y at time t* . As the name suggests, we represent these atoms indeed as arrows in a diagram. Figure 5.1 shows an example for a graphical representation with $E = \{1, 2, 3, 4, 5, 6\}$. The horizontal lines in the diagram represent the set $S \subseteq \mathbb{R}_+ \times E$ (which is why there are sometimes holes in the horizontal lines) and the arrows represent atoms of η .

Analogously, we define REP_T as the space of integer valued measures on $S_T \times E_\infty$ which are finite on compact sets. We define $\text{REP}^* \subseteq \text{REP}$ as the integer valued measures for which no two arrows happen at the same time and $\text{REP}_T^* \subseteq \text{REP}_T$ as the integer valued measures which do not feature arrows at the time T and for which no two arrows happen at the same time. Note

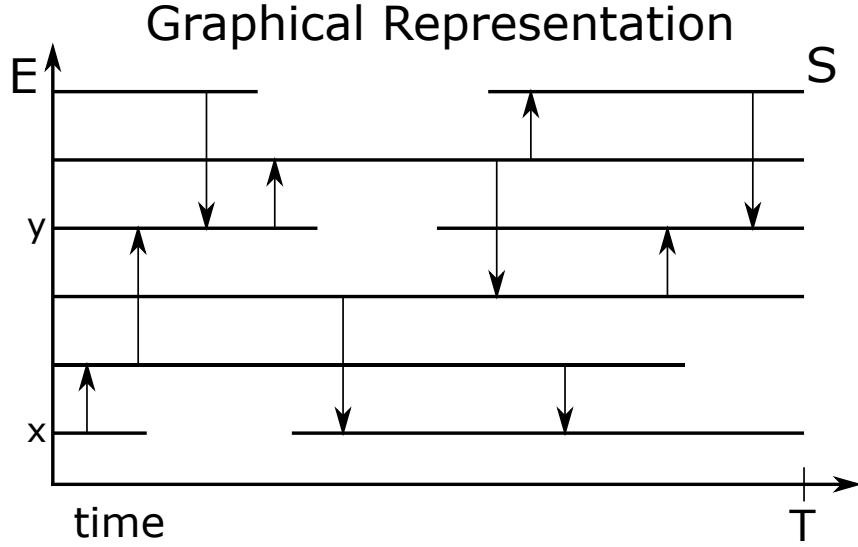


Figure 5.1: A graphical representation for a finite space E consisting of 6 elements

that the graphical representation in Figure 5.1 also corresponds to a measure in $\eta \in \text{REP}_T^*$.

We now want to define a topology on REP . It turns out that we just need the concept of vague convergence of measures that are finite on compact sets.

Definition 5.2 (The Topology on the Space of Graphical Representations). We choose the topology on $\text{REP} \subseteq \mathcal{M}_c(S \times E_\infty)$ to be induced by the vague topology. Due to Theorem 4.10 and since $S \times E_\infty$ is a Polish space we have that $\mathcal{M}_c(S \times E_\infty)$ is Polish and REP is a closed subspace and thus also Polish. We define the topology on REP_T analogously.

5.2 Tracking a Graphical Representation

Given a graphical representation we can define a process by tracking along the arrows of the representation.

Definition 5.3 (The Tracking Process). Let E be a countable set and $S \subseteq \mathbb{R}_+ \times E$ open. Let $\partial \notin E$ be a graveyard state and $\eta \in \text{REP}$ a graphical representation on E with admissible states S with almost surely no arrows pointing to ∞ . For a starting point $x \in E$ we define a map

$$\Phi(x) = \Phi: \text{REP} \rightarrow (E \cup \{\partial\})^{\mathbb{R}_+}, \quad \eta \mapsto (t \mapsto x(t)).$$

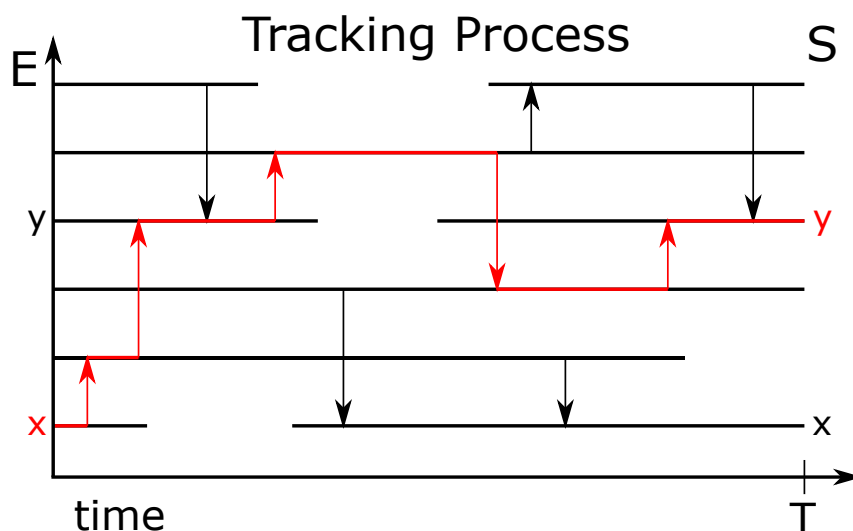


Figure 5.2: The tracking process applied to Figure 5.1 started at x at time 0

We define $t \mapsto x(t)$ as a piecewise constant map with jumptimes T_0, T_1, \dots using the following recursion:

1. Set $x(0) = x$ and $T_0 = 0$.
2. If for $n \in \mathbb{N}_0$ $T_n \in \mathbb{R}_+$ is defined and $x(T_n) = y \in E$ set T_{n+1} as the infimum over all times $t > T_n$ where arrows point out of y in the graphical representation or the times with $(t, y) \notin S$ and set $x(t) = y$ for $t \in [T_n, T_{n+1})$.
3. If $(T_{n+1}, x(T_n)) \notin S$ set $x(t) = \partial$ for all $t \geq T_{n+1}$.
4. If $T_{n+1} = T_n$ set $x(t) = \partial$ for all $t > T_n$.
5. If $T_{n+1} > T_n$ but the infimum is not uniquely given by one arrow (either by it not being attained or by there being multiple arrows at the same time) set $x(t) = \partial$ for all $t \geq T_{n+1}$.
6. If $T_{n+1} > T_n$ and there exists a unique first arrow out of y , let z be the site to which the arrow points and set $x(T_{n+1}) = z$. Then go back to step 2.
7. If Steps 3 to 5 never occur but we have $\tau := \sup_{n \in \mathbb{N}} T_n < \infty$ set $x(t) = \partial$ for $t \geq \tau$.

We call the function $t \mapsto x(t)$ the *tracking process of the graphical representation* η started in x . For a given $(s, x) \in S$ we can generalize this procedure by shifting the graphical representation by $-s$ along the time axis and shifting the resulting process by s along time. This allows us to also define the tracking process $[s, \infty) \rightarrow E \cup \{\partial\}$ for the starting condition $x(s) = x$.

The case that we reach ∂ via step 7. is called an *explosion* and τ the *explosion time*. If for any admissible starting point $(s, x) \in S$ the tracking process can only reach ∂ via explosion we call the graphical representation *proper*. We call the set

$$\text{Tr} := \bigcup_{n \in \mathbb{N}_0} [T_n, T_{n+1}] \times \{x(T_n)\} \subseteq \mathbb{R}_+ \times (E \cup \{\partial\})$$

the *track* of η started in x .

For the case $E = \{1, 2, 3, 4, 5, 6\}$ and for the starting condition $x(0) = x$ we visualized the tracking procedure applied to the graphical representation shown in Figure 5.1 with the red line in Figure 5.2. In fact the (horizontal parts of the) red line are the track of η started in x . In this particular example we can see that the tracking process will be in state y at time T ($x(T) = y$). Furthermore, the graphical representation featured in this example is not proper, since it is possible to choose a starting point $(s, x) \in S$ right before a hole in the horizontal line corresponding to x such that there will not be an arrow out of x before the tracking process reaches the hole. In fact the only way how a graphical representation featuring such holes can be proper is, if arrows cluster to the left of any hole.

Let $T > 0$. We also define $\Psi_T: \text{REP}_T \rightarrow (E \cup \{\partial\})^{[0, T]}$ as the restriction of Ψ to REP_T (alternatively due to construction we could also use the above construction just up to time T). We also define $\text{Tr}_T := \text{Tr} \cap [0, T] \times (E \cup \{\partial\})$.

We need some notation for upcoming function spaces.

Notation 5.4. Let E be a topological space and $T \geq 0$. We use the following notations:

1. We write $D_E([0, \infty))$ for the set of all càdlàg functions $f: [0, \infty) \rightarrow E$. This means that for all strictly increasing sequences $(x_n)_{n \in \mathbb{N}_+}$ in $[0, \infty)$ with $x_n \nearrow x \in [0, \infty)$ the sequence $(f(x_n))_{n \in \mathbb{N}_+}$ converges to some limit $f(x-) \in E$ and for all decreasing sequences $(x_n)_{n \in \mathbb{N}_+}$ in $[0, \infty)$ with $x_n \searrow x \in [0, \infty)$ we have $f(x_n) \rightarrow f(x)$ in E .
2. We endow $D_E([0, \infty))$ with the Skorohod topology as defined by Ethier and Kurtz [8].

3. We write $D_E([0, T])$ for the set of all càdlàg functions $f: [0, T] \rightarrow E$ which do not have a jump at time T .
4. We endow $D_E([0, T])$ with the Skorohod topology induced by the inclusion map $\iota: D_E([0, T]) \rightarrow D_E([0, \infty))$ which maps $f \in D_E([0, T])$ to its constant extension to $[0, \infty)$.
5. We write $C_E([0, \infty))$ for the set of all continuous functions $f: [0, \infty) \rightarrow E$.
6. We endow $C_E([0, \infty))$ with the topology induced by uniform convergence on compact subsets of $[0, \infty)$.

Proposition 5.5. *Let $T > 0$, $x \in E$. Consider a sequence of graphical representations $(\eta_n)_{n \in \mathbb{N}_+} \subseteq \text{REP}_T$ and $\eta \in \Psi_T^{-1}(D_E([0, T]))$ with $\eta_n \rightarrow \eta$ for $n \rightarrow \infty$ in the vague topology. Then $\Psi_T(\eta_n) \in D_E([0, T])$ for n sufficiently large and $\Psi_T(\eta_n) \rightarrow \Psi_T(\eta)$ in the Skorohod topology on $D_E([0, T])$. Let $t \in [0, T]$, then the mapping*

$$f_t: \text{REP}_T^* \rightarrow E \cup \{\partial\} \quad \eta \mapsto \Psi_T(\eta)(t)$$

is measurable.

Proof. Consider the track Tr_T of η for the starting condition $x \in E$. It helps to refer to Figure 5.2 for a visualization of the track. Since the tracking process $\psi_T(\eta)$ is a càdlàg function which does not jump at T and does not visit the graveyard state ∂ we have for $(t, y) \in \text{Tr}_T$ that $(t, y) \in S_T$. Thus Tr_T and $S_T^c := [0, T] \times E \setminus S_T$ are disjoint, compact subsets of $[0, T] \times E$. In particular Tr_T and S_T^c have positive distance from another. Thus we can find an open set $A \subseteq S_T$ such that $\text{Tr}_T \subseteq A \subseteq \bar{A} \subseteq S_T$. Note that by 2 of Proposition 4.9 \bar{A} is a compact subset of S_T . by the continuity of the (integer valued) measure η from above we can choose A small enough such that $\eta(\bar{A} \times E_\infty) = \eta(\text{Tr} \times E_\infty)$ and thus in particular $\eta(\partial(A \times E_\infty)) = 0$ (note that by definition this would not be the case for Tr_T itself). Consider η restricted to $\bar{A} \times E$. Since $\psi_T(\eta) \in D_E([0, T])$ and since all arrows pointing out of \bar{A} correspond to jumps of $\psi(\eta)$ the arrows pointing out of \bar{A} can be uniquely ordered and there is no arrow at time 0 or at time T . Also, since $\Psi_T(\eta)(t) \neq \infty$ for all $t \in [0, T]$ there is no arrow pointing out of \bar{A} to ∞ . Since $\eta_n \xrightarrow{v} \eta$ and since $\bar{A} \times E_\infty$ is compact and has no mass on the boundary with respect to η we have in particular weak convergence of the integer valued measures restricted to $\bar{A} \times E$. Thus, for n large enough, we will have the same number of arrows going out of \bar{A} in η_n as in η . By choosing n even larger we can assume that the arrows can be ordered and the order coincides

with the order of the arrows pointing out of \bar{A} in η . By the same argument we can also achieve that there is no arrow at time 0 or time T . Furthermore, since the arrows in η which point out of Tr_T by definition also point into Tr_T and $\text{Tr}_T \subseteq A$ with A open we can choose n even larger such that each arrow in η_n that converges to an arrow in Tr_T pointing from y to z at some time t will point into a neighborhood of $(y, z, t) \in \text{Tr}_T$ which is contained in A . Now we note that if all these conditions are fulfilled the tracking process for η_n will visit the same states in the same order as the tracking process for η up to time T since only the order of arrows matters and we always only consider arrows going out of \bar{A} . In particular we never visit the graveyard state ∂ . Thus we have $\Psi_T(\eta_n) \in D_E([0, T])$ for n large. To see continuity we now note that the tracking process of η_n already performs the same jumps as the tracking process of η and for $n \rightarrow \infty$ the jump times converge as well. Additionally there are no jump at times T and 0. This implies that the tracking process converges in the Skorohod topology. In detail: for $n \in \mathbb{N}_+$ sufficiently large, we can find a strictly increasing, piecewise linear mapping $\lambda_n: [0, T] \rightarrow [0, T]$ such that $\lambda_n(0) = 0$, $\lambda_n(T) = T$ and λ_n maps the jumptimes of $\Psi_T(\eta)$ to the jumptimes of $\Psi_T(\eta_n)$. Thus the Skorohod distance (see (5.2) in [8]) of the functions is bounded by the maximal, logarithmic distance between the corresponding jumptimes of $\Psi_T(\eta)$ and $\Psi_T(\eta_n)$ which converges to zero for $n \rightarrow \infty$. This implies $\Psi_T(\eta_n) \rightarrow \Psi_T(\eta)$ in the Skorohod topology.

We now show that f is measurable. Let $A \subseteq E$. Since $\partial \notin A$ and since the domain of f is REP_T^* we have that $f^{-1}(A) \subseteq \Psi_T^{-1}(D_E([0, T]))$. Let $\rho_t: D_E([0, T]) \rightarrow E$ be the evaluation map $\rho(x) = x(t)$. We have shown in the previous part of the proof that Ψ_T is continuous on $\Psi_T^{-1}(D_E([0, T]))$. The projection ρ_t is measurable and $f_t = \rho_t \circ \Psi_T$ is therefore measurable if restricted to $\Psi_T^{-1}(D_E([0, T])) \cap \text{REP}_T^*$. Thus we can conclude that $f^{-1}(A)$ is a measurable set. In particular we have that $f^{-1}(E)$ is a measurable set and thus $f^{-1}(\{\partial\}) = \text{REP}_T^* \setminus f^{-1}(E)$ is a measurable set as well. Thus we can conclude measurability of $f^{-1}(A)$ for all $A \subseteq E \cup \{\partial\}$. Therefore f is measurable which shows the second claim. \square

Chapter 6

Time-Inhomogeneous Regular Jump Processes

In this chapter we will define time-inhomogeneous regular jump processes, which are a class of time-inhomogeneous Markov chains which can be described via graphical representations. Later in the chapter we will provide conditions under which such a process does not explode.

6.1 Definitions and Basic Properties

We will use the following notation:

Notation 6.1. As before let E be a countable discrete space and S be an open subset of $\mathbb{R}_+ \times E$. We define for $x \in E$ and $t \in \mathbb{R}_+$:

$$T_x := \{s \in \mathbb{R}_+ \mid (s, x) \in S\} \subseteq \mathbb{R}_+, \quad E_t := \{y \in E \mid (t, y) \in S\} \subseteq E.$$

Definition 6.2 (Time-Inhomogeneous Regular Jump Process).

For $u \in \mathbb{R}_+$ let $Q_u = (q_{u;x,y})_{x,y \in E} \in \mathbb{R}^{E \times E}$ be a matrix with the following properties:

1. We have $q_{u;x,y} \in \mathbb{R}_+$ for all $x, y \in E$ with $x \neq y$. We call $q_{u;x,y}$ the *(infinitesimal) rate of jumps from x to y at time u* .
2. The matrix Q_u is *stable*. More precisely $q_{u;x} := -q_{u;x,x} < \infty$ for Lebesgue almost all $u \in T_x$.
3. The matrix Q_u is *conservative*. More precisely $q_{u;x} = \sum_{y \in E \setminus \{x\}} q_{u;x,y}$ for all $x \in E$ and Lebesgue almost all $u \in T_x$.

4. $q_{u;x,y} = 0$ if $(u, x) \notin S$ or $(u, y) \notin S$.
5. Let $x \in E$, $y \neq x$. The mapping $u \mapsto q_{u,x,y}$ is measurable on \mathbb{R}_+ and the mapping $u \mapsto q_{u,x}$ is integrable on compact subsets of T_x . Also, for all $t \in T_x^c$ such that there is a sequence $(t_n)_{n \in \mathbb{N}} \subseteq T_x$ with $t_n \nearrow t$ we stipulate

$$\int_{t-\delta}^t q_{u;x} du = \infty \quad \text{for all } \delta \in (0, t].$$

We can define a random graphical representation in REP by considering the Poisson point processes η on $\mathbb{R}_+ \times E \times E_\infty$ with intensity measure ν given by

$$\nu((s, t] \times F_1 \times F_2) = \sum_{x \in F_1} \sum_{y \in F_2 \setminus \{x\}} \int_s^t q_{u;x,y} du.$$

Given this representation and any $(s, x) \in S$ we can now define the tracking process $X := (X_t)_{t \geq s}$ started in x at time s as defined in Definition 5.3. We call X the *time-inhomogeneous regular jump process defined by Q* . Note that X will take values in $\tilde{E} := E \cup \{\partial\}$ where ∂ is an additional absorbing graveyard state and that (X_t, t) will either be in S or $X_t = \partial$. The properties 4. and 5. of Q ensure that the graphical representation is proper. Thus X can only reach ∂ by explosion. We call S the set of *admissible states of X* . Furthermore note that X depends on x and s but we will usually omit this in the notation for X . Instead, when considering probabilities or expectations, we will use the notation $\mathbb{P}^{(s,x)}$ and $\mathbb{E}^{(s,x)}$ if we consider X to be started in x at time s . Given $(s, x) \in S$ and $(t, y) \in S$ we define the transition matrix $P(s, t) := (p_{x,y}(s, t))_{x,y \in E}$ given by

$$p_{x,y}(s, t) := \mathbb{P}^{(s,x)}(X_t = y).$$

Note that since we did not allow $y = \partial$ the probabilities may not add up to 1. In this case there is a positive probability for explosion.

In addition, given a starting time $s \in \mathbb{R}_+$ and an initial distribution $\pi_s = (\pi_{s;x})_{x \in E}$ with $\pi_{s;x} = 0$ if $x \in E_s^c$ we define for $t \geq s$ the subdistribution $\pi_t = (\pi_{t;x})_{x \in E}$ via

$$\pi_{t;x} = \sum_{y \in E} \pi_{s;y} p_{y,x}(s, t). \quad (6.1)$$

Note that $\pi_{t;x}$ is the probability for the Markov chain X to be in x at time t if it is started in the distribution π_s at time s .

Proposition 6.3. *Consider the situation of Definition 6.2. Then the graphical representation defined by Q is indeed proper. Furthermore, the time-inhomogeneous regular jump process defined by Q is a strong, right-continuous (time-inhomogeneous) Markov chain with respect to the filtration $(\mathcal{F}_t)_{t \geq s}$ where \mathcal{F}_t is generated by the random graphical representation restricted to the time interval $[s, t]$.*

Proof. Since the rate with which arrows point in or out of $S^c \times E$ is 0 there are almost surely no such arrows.

Furthermore since the intensity measures are finite on compact subsets of $S \times E_\infty$ the points of the Poisson point process almost surely have no clusterpoints in $S \times E$.

Additionally, since the intensity measures are continuous the arrows in the graphical representation are almost surely at mutually different times.

Let $x \in E$. Since the rate for arrows to point out of x integrates to infinity when approaching the boundary of T_x (from the left) we have almost surely that such arrows cluster to the left of any time $t \in \partial T_x$.

Thus it is almost surely impossible for the process X to reach the graveyard state in finitely many steps (this corresponds to steps 3 - 5 in Definition 5.3) since there is always either a unique next arrow pointing out of the current location of X before the time reaches the boundary of T_x or no arrow pointing out of x after some time s and $[s, \infty) \subseteq T_x$. Therefore the graphical representation is almost surely proper.

We now show that X is a time-inhomogeneous Markov chain. Since the intensity measure of the Poisson point process is continuous we have that the graphical representation is almost surely in REP^* . We consider a version which is surely in REP^* .

Let $T \geq 0$. We have shown in Proposition 6.3 that X_T is a measurable function of the graphical representation up to T and thus X_T is \mathcal{F}_T -measurable. It is left to show the Markov property.

Let $0 \leq s \leq t$ and $x \in E$. Let η be the graphical representation restricted to $(s, t] \times E^2$ and shifted by s to the left. Note that we have that $\eta \in \text{REP}_{t-s}^*$ almost surely. For fixed s, t we can again find a version of our representation such that this holds surely. Furthermore let $\Psi(x)(\eta)$ be the tracking process which tracks through η and starts at state x . Again by Proposition 6.3 we have that the map $\eta \mapsto \Psi(x)(\eta)(t-s)$ is measurable. Since η is independent from \mathcal{F}_s and X_s is \mathcal{F}_s -measurable we have:

$$\begin{aligned} \mathbb{P}(X_t = x | \mathcal{F}_s) &= \mathbb{P}(\Psi(\eta, X_s)(t-s) = x | \mathcal{F}_s) \\ &= \mathbb{P}(\Psi(\eta, X_s)(t-s) = x | X_s) = \mathbb{P}(X_t = x | X_s). \end{aligned}$$

Thus X is a time-inhomogeneous Markov chain.

Right continuity of the paths follows directly from the definition of the tracking map Ψ and the fact that our graphical representation is almost surely proper. In particular X is càdlàg up to a possible explosion time and constant after the explosion time.

Let $(s, x) \in S$. Consider a time-inhomogeneous Markov chain X with $X_s = x$. Ethier and Kurtz [8] point out that we can always consider the (time-homogeneous) Markov process $(t+s, X_{t+s})_{t \in \mathbb{R}_+}$ with state space $S \cup (\mathbb{R}_+ \times \{\partial\})$. For our purposes it suffices to show that this Markov process is indeed strongly Markov and right-continuous. We first note that we can use our graphical representation to construct this Markov process directly for any starting condition $(s, x) \in S$. As usual we denote the starting condition as a superscript in the probabilities and expectations. Since X is right continuous this process is right continuous as well. Now let $k \in \mathbb{N}$, $s \leq s + r_1 < \dots < s + r_k$ and ϕ_1, \dots, ϕ_k bounded, continuous functionals on E . Define the mapping

$$g: s \mapsto \mathbb{E}^{(s,x)}(\phi_1(X_{s+r_1}) \cdot \dots \cdot \phi_k(X_{s+r_k})). \quad (6.2)$$

We want to show right-continuity of g . Note that in (6.2) the expectation on the right-hand side depends on s in two ways. On one hand there is a dependence due to the starting condition and on the other hand a dependence of the times at which X is evaluated inside the expectation. We will show right-continuity of g by showing that it is continuous in the starting condition and right-continuous with respect to the evaluation of X . We start by showing continuity for $s + r_1 = t_1, \dots, s + r_k = t_k$ fixed. Let $\varepsilon > 0$. We can find a neighborhood $(s - \delta, s + \delta) \subseteq T_x$ of $s \in T_x$ such that the probability for an arrow out of x at any time $u \in (s - \delta, s + \delta)$ is smaller than ε . But if there is no arrow pointing out of x at any time in $u \in (s - \delta, s + \delta)$ the tracking map produces the same result for any starting condition (u, x) . Thus we have for all $u \in (s - \delta, s + \delta)$:

$$|\mathbb{E}^{(s,x)}(\phi_1(X_{t_1}) \cdot \dots \cdot \phi_k(X_{t_k})) - \mathbb{E}^{(u,x)}(\phi_1(X_{t_1}) \cdot \dots \cdot \phi_k(X_{t_k}))| \leq 2 \|\phi_1\| \cdot \dots \cdot \|\phi_k\| \varepsilon.$$

Here $\|\bullet\|$ denotes the sup-norm of a function. This shows the continuity.

Now we fix the starting condition $(s, x) = (t, x)$ with $t < r_1$. Since the paths of X are almost surely right-continuous and since the functionals ϕ_1, \dots, ϕ_k are continuous and bounded we can use dominated convergence to see that the mapping

$$[0, r_1 - t] \rightarrow \mathbb{R}, \quad s \mapsto \mathbb{E}^{(t,x)}(\phi_1(X_{s+r_1}) \cdot \dots \cdot \phi_k(X_{s+r_k}))$$

is right-continuous. Thus g is also right-continuous. We now show the strong Markov property of $(t + s, X_{t+s})_{t \in \mathbb{R}_+}$. The strategy is analogous to the case

for Feller processes (see the proof of Theorem 2.7 in [8]). We first consider a stopping time τ which only takes countably many values $(t_n)_{n \in \mathbb{N}}$. We have for $0 \leq r_1 \cdots \leq r_k$:

$$\begin{aligned} & \mathbb{E}^{(s,x)}(\phi_1(X_{\tau+r_1}) \cdots \phi_k(X_{\tau+r_k}) | \mathcal{F}_\tau) \\ &= \sum_{n \in \mathbb{N}} \mathbf{1}_{\{\tau=t_n\}} \mathbb{E}^{(s,x)}(\phi_1(X_{t_n+r_1}) \cdots \phi_k(X_{t_n+r_k}) | \mathcal{F}_{t_n}) \\ &= \sum_{n \in \mathbb{N}} \mathbf{1}_{\{\tau=t_n\}} \mathbb{E}^{(t_n, X_{t_n})}(\phi_1(X_{r_1}) \cdots \phi_k(X_{r_k})) \\ &= \mathbb{E}^{(\tau, X_\tau)}(\phi_1(X_{r_1}) \cdots \phi_k(X_{r_k})). \end{aligned}$$

Now let τ be an arbitrary almost surely finite stopping time. We can approximate τ from above by a sequence of stopping times $\tau_n \searrow \tau$ where τ_n only takes countably many values. Thus we have

$$\mathbb{E}^{(s,x)}(\phi_1(X_{\tau_n+r_1}) \cdots \phi_k(X_{\tau_n+r_k}) | \mathcal{F}_{\tau_n}) = \mathbb{E}^{(\tau_n, X_{\tau_n})}(\phi_1(X_{r_1}) \cdots \phi_k(X_{r_k})).$$

Now we use right continuity of the paths to conclude $(\tau_n, X_{\tau_n}) \rightarrow (\tau, X_\tau)$ almost surely. Thus using dominated convergence on the left hand side and the shown right-continuity on the right hand side (and the fact that E is discrete) as well as $\mathcal{F}_\tau \subseteq \mathcal{F}_{\tau_n}$ we get

$$\mathbb{E}^{(s,x)}(\phi_1(X_{\tau+r_1}) \cdots \phi_k(X_{\tau+r_k}) | \mathcal{F}_\tau) = \mathbb{E}^{(\tau, X_\tau)}(\phi_1(X_{r_1}) \cdots \phi_k(X_{r_k})).$$

Using the monotone class theorem we can lift this property to bounded measurable functionals of the whole path. Going back to the Markov chain X we can now conclude that it too has to be strongly Markov. \square

Our first goal is to show Kolmogorov's backwards equation for P . Since we will work with the concept of absolute continuity and weak differentials we give a short definition.

Definition 6.4. Let $A \subseteq \mathbb{R}$ be an arbitrary subset and $f: A \rightarrow \mathbb{R}$. We call f absolutely continuous on A , if there exists a so called *weak differential* $g: A \rightarrow \mathbb{R}$, which is integrable on compact subsets of A , such that we have for any $[s, t] \subseteq A$:

$$f(t) - f(s) = \int_s^t g(x) dx.$$

Remark 6.5. It is commonly known in measure theory that absolute continuity on $A = [s, t]$ can be represented using an ε - δ -criterion in the following way.

$f: [s, t] \rightarrow \mathbb{R}$ is absolutely continuous if and only if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for any collection $s \leq x_1 \leq y_1 \leq \dots \leq x_k \leq y_k \leq t$ with

$$\sum_{j=1}^k (y_j - x_j) \leq \delta$$

we have

$$\sum_{j=1}^k |f(y_j) - f(x_j)| \leq \delta.$$

Furthermore, the main theorem of calculus implies that if f is absolutely continuous on an open set $O \subseteq \mathbb{R}$ then f is Lebesgue almost everywhere differentiable (in the sense of classical calculus) and the differential coincides Lebesgue almost everywhere with the weak differential.

Lemma 6.6. *Consider the situation of Definition 6.2. Let $(s, x) \in S$ as well as $y \in E \setminus \{x\}$. Let τ_y be the first time after s for which there is an arrow pointing from x to y in the graphical representation. τ_y is an $(\mathcal{F}_t)_{t \geq s}$ -stopping time. The times are mutually independent for different y and we have for distribution function $F_y: [s, \infty) \rightarrow [0, 1]$ and density f_y of τ_y :*

$$F_y(r) = 1 - \exp\left(-\int_s^r q_{u;x,y} \, du\right), \quad f_y(r) = q_{r;x,y} \exp\left(-\int_s^r q_{u;x,y} \, du\right).$$

Now define $\tau := \inf_{z \in E \setminus \{x\}} \tau_z$, we have for $r \geq s$:

$$\mathbb{P}^{(x,s)}(\tau > r) = \exp\left(-\int_s^r q_{u;x} \, du\right)$$

and for $y \in E$:

$$\mathbb{P}^{(x,s)}(\tau_y = \tau \leq r) = \int_s^r q_{u;x,y} \exp\left(-\int_s^u q_{u';x} \, du'\right) \, du$$

Finally we get for the density of τ_y on the event that there is at least one arrow pointing out of x in $[s, r]$ and that the first arrow is pointing to y :

$$g: [s, r] \rightarrow \mathbb{R}_+, \quad u \mapsto q_{u;x,y} \exp\left(-\int_s^u q_{u';x} \, du'\right).$$

Proof. Let $r \geq 0$. The event $\{\tau_y \leq s + r\}$ is the same as there being at least one arrow in the graphical representation pointing from x to y at some time in $(s, s + r]$. In particular the event is in \mathcal{F}_{s+r} and thus τ_y is a stopping time. The independence of the times is a direct consequence of the independence

of the corresponding Poisson point processes in the graphical representation. Since the number of arrows pointing from x to y in the time interval $[s, s+r]$ is Poisson distributed with parameter $\int_s^{s+r} q_{u;x,y} du$ we get the result for F_y since it is exactly the probability for such a Poisson distribution to take a value greater than zero.

We now apply the fundamental theorem of calculus to derive the density. In detail we observe that the function $r \mapsto \int_s^r q_{u;x,y} du$ is absolutely continuous on $T_x \cap [s, \infty)$, nondecreasing and that it is infinite for $r \in T_x^c$, $r \geq s$. Let $t \in (s, \infty]$ be given by $t := \inf T_x^c \cap [s, \infty)$. Then F_y is absolutely continuous on $[s, t)$ with $F_y(r) \nearrow 1$ for $r \nearrow t$ and $F_y(l) = 1$ for all $l \geq t$. Thus $F_y: [s, \infty) \rightarrow [0, 1]$ is an absolutely continuous distribution function. By the fundamental theorem of calculus we can calculate the Lebesgue density f_y by differentiating F_y which yields the claimed result. To derive the distribution function of τ we use independence to get for $r \geq s$

$$\begin{aligned} \mathbb{P}^{(s,x)}(\tau > r) &= \mathbb{P}^{(s,x)}(\tau_z > r \text{ for all } z \in E \setminus \{x\}) \\ &= \prod_{z \in E \setminus \{x\}} \mathbb{P}^{(s,x)}(\tau_z > r) \\ &= \exp\left(-\int_0^r q_{u;x} du\right). \end{aligned}$$

Note that in similar fashion we also get for $y \neq x$, $r \geq s$:

$$\mathbb{P}^{(s,x)}\left(\inf_{z \in E \setminus \{x,y\}} \tau_z > r\right) = \exp\left(-\int_0^r q_{u;x} - q_{u;x,y} du\right).$$

We use this and get by the strong Markov property:

$$\begin{aligned} \mathbb{P}^{(s,x)}(\tau_y = \tau \leq r) &= \mathbb{P}^{(s,x)}(\tau_y \leq r, \tau_y < \inf_{z \in E \setminus \{x,y\}} \tau_z) \\ &= \int_s^r f_y(u) \cdot \mathbb{P}^{(s,x)}\left(\inf_{z \in E \setminus \{x,y\}} \tau_z > u\right) du \\ &= \int_s^r q_{u;x,y} \exp\left(-\int_s^u q_{u';x} du'\right) du. \end{aligned}$$

With the same argument as for F_y we can show that this function is absolutely continuous on $[s, \infty)$ and by the fundamental theorem of calculus g is the density of τ_y on the stipulated event. \square

We can now use these results to show Kolmogorov's backwards equation.

Theorem 6.7 (Kolmogorov's Backwards Equation). *Consider the situation of Definition 6.2 and let $(s, x) \in S$ and $t \geq s$. The following claims hold:*

1. The mapping

$$h_y: [s, t] \rightarrow [0, 1], \quad u \mapsto p_{x,y}(u, t)$$

is absolutely continuous on $T_x \cap [s, t]$.

2. We have for the Lebesgue almost everywhere defined differential of h_y :

$$\frac{d}{du} h_y(u) = - \sum_{z \in E} q_{u;x,z} \cdot p_{z,y}(u, t).$$

Note that for $z \in E_u^c$ we have $q_{u;x,z} = 0$ and thus we can set the summand to be zero even though $p_{z,y}(u, t)$ was not defined for such z .

Proof.

1. Let $s \leq v \leq w \leq t$ and v, w in the same connected component of T_x . Let τ be the time of the first arrow out of x after time v and τ_z the first arrow pointing from x to z after time v . Consider the Markov chain started in $x \in E$ at time v . On the event $\{\tau > w\}$ it will stay in x until time w and on the event $\{\tau \leq w\} \cap \{\tau = \tau_z\}$ it will jump to $z \in E \setminus \{x\}$ at time τ . Thus we have by the strong Markov property:

$$\begin{aligned} \mathbb{P}^{(v,x)}(X_t = y) &= \mathbb{P}^{(v,x)}(X_t = y, \tau > w) \\ &+ \sum_{z \in E \setminus \{x\}} \mathbb{P}^{(v,x)}(X_t = y, \tau \leq w, \tau_z = \tau) \\ &= \mathbb{P}^{(w,x)}(X_t = y) \mathbb{P}^{(v,x)}(\tau > w) \\ &+ \sum_{z \in E \setminus \{x\}} \mathbb{E}^{(v,x)} \left(\mathbb{P}^{(\tau,z)}(X_t = y) \mathbf{1}_{\{\tau \leq w, \tau_z = \tau\}} \right) \\ &= p_{x,y}(w, t) \mathbb{P}^{(v,x)}(\tau > w) + \sum_{z \in E \setminus \{x\}} \mathbb{E}^{(v,x)} \left(p_{z,y}(\tau, t) \mathbf{1}_{\{\tau \leq w, \tau_z = \tau\}} \right) \end{aligned}$$

We have by Lemma 6.6:

$$\begin{aligned} p_{x,y}(v, t) &= p_{x,y}(w, t) \exp \left(- \int_v^w q_{u;x} du \right) \\ &+ \sum_{z \in E \setminus \{x\}} \int_v^w p_{z,y}(u, t) q_{u;x,z} \exp \left(- \int_v^u q_{u';x} du' \right) du. \quad (6.3) \end{aligned}$$

Thus we get

$$\begin{aligned}
& |p_{x,y}(w, t) - p_{x,y}(v, t)| \\
& \leq p_{x,y}(w, t) \left(1 - \exp \left(- \int_v^w q_{u;x} du \right) \right) \\
& + \sum_{z \in E \setminus \{x\}} \int_v^w p_{z,y}(u, t) q_{u;x,z} \exp \left(- \int_v^u q_{u';x} du' \right) du \\
& \leq \left(1 - \exp \left(- \int_v^w q_{u;x} du \right) \right) + \sum_{z \in E \setminus \{x\}} \int_v^w q_{u;x,z} \exp \left(- \int_v^u q_{u';x} du' \right) du \\
& \leq \int_v^w q_{u;x} du + \int_v^w q_{u;x} \exp \left(- \int_v^u q_{u';x} du' \right) du \\
& \leq 2 \int_v^w q_{u;x} du.
\end{aligned}$$

Since $u \mapsto q_{u;x}$ is integrable on compact subsets of T_x we have that $u \mapsto p_{x,y}(u, t)$ is absolutely continuous on $T_x \cap [s, t]$ (see Remark 6.5).

2. Absolute continuity implies differentiability Lebesgue almost everywhere. Let the function be differentiable at $w \in [s, t] \subseteq T_x$. It suffices to calculate the left-hand differential. We observe that we can apply Tonelli's Theorem to exchange sum and integration in the right-hand side of (6.3). Since $u \mapsto q_{u;x}$ is integrable on the compact set $[s, t]$ the the right-hand side is differentiable at $v = w$ and we get by the fundamental theorem of calculus:

$$\begin{aligned}
\frac{d}{dv} p_{x,y}(v, t) \Big|_{v=w} &= p_{x,y}(w, t) q_{w;x} + \sum_{z \in E \setminus \{x\}} -p_{z,y}(w, t) q_{w;x,z} \\
&= - \sum_{z \in E \setminus \{x\}} q_{w;x,z} p_{z,y}(w, t).
\end{aligned}$$

□

6.2 Some Set Theory

As the title may suggest this section is mostly about a general concept we will need in the following section where we return to analysing time-inhomogeneous Markov chains. It should be noted that our only application of the results in this section are 4. of Theorem 6.16 and Theorem 6.18 which treat the case that E is finite with great generality. The proof of our main result

(Theorem 3.14) is using Theorem 6.19 instead, which does not require this section. Hence, a reader who is only interested in the main result may skip this section.

Consider a property that holds for almost all real numbers in the sense that it holds for $\mathbb{R} \setminus A$ for A closed and countable. Furthermore assume that we can extend the property over isolated points of A . The goal of this section is to show that the property holds for all real numbers. (For the actual mathematical formulation of this see Theorem 6.12.) The concept of removing isolated points from a set is not new in set theory. Removing all isolated points from a set will yield the so called *Cantor-Bendixson derivative* and doing so recursively yields the *Cantor-Bendixson sequence*. Furthermore it is known that this recursion does terminate. It does so not necessarily after finitely many steps but at a countable ordinal (see for example the proof of the Cantor-Bendixson Theorem 4.6 in [14]). In order to avoid using the set theoretical concept of ordinals we will use an alternative line of proof which ultimately relies on Zorn's Lemma instead. We start with a couple of definitions and lemmas which are already rather well known in set theory. We provide proofs for the convenience of the reader nonetheless in particular to avoid delving deeper than necessary into set theoretical concepts.

Definition 6.8. Let $C \subseteq \mathbb{R}$ be a subset.

1. We call $x \in C$ an isolated point if there exists an $\varepsilon > 0$ such that

$$(x - \varepsilon, x + \varepsilon) \cap C = \{x\}.$$

2. Let C be a closed set with no isolated points then we call C *perfect*.

Example 6.9. Here are two examples for perfect sets:

1. Let $C = [a, b] \subseteq \mathbb{R}$ with $a < b$, then C is perfect.
2. Let $C \subseteq [0, 1]$ be the Cantor set, then C is perfect.

The following lemma corresponds to Theorem 4.5 in [14].

Lemma 6.10 (Perfect Sets are Uncountable). *Let $C \subseteq \mathbb{R}$ be perfect then C is either empty or uncountable.*

Proof. Let C be nonempty then there is a $y \in C$ and y can not be isolated. Thus there has to be a sequence $(y_n)_{n \in \mathbb{N}_+} \subseteq C \setminus \{y\}$ with $y_n \rightarrow y$. In particular S can not be finite.

We now show the claim by contradiction. Assume that C is not uncountable and not empty then it has to be an infinite, but countable, set. Let $C = (x_n)_{n \in \mathbb{N}_+}$ be an enumeration of C . We now construct a falling sequence of closed subsets $C_n \subseteq C$ by the following recursion:

1. To define C_1 consider $x_1 \in C$. Since x_1 is not isolated there is $0 < \varepsilon_1 < 1$ such that $C \cap \text{Ann}_{x_1}(\varepsilon_1/2, \varepsilon_1)$ is not empty where for $x \in \mathbb{R}$, $0 < \delta < \varepsilon$ we define the annulus:

$$\text{Ann}_x(\delta, \varepsilon) := (x - \varepsilon, x + \varepsilon) \setminus [x - \delta, x + \delta].$$

Furthermore, since C is countable we can choose ε_1 such that

$$\partial \text{Ann}_{x_1}(\varepsilon_1/2, \varepsilon_1) \cap C = \emptyset.$$

We now define $C_1 := C \cap \text{Ann}_{x_1}(\varepsilon_1/2, \varepsilon_1)$. Note that, since the annulus is an open set, C_1 has no isolated points and, since there is no point of C on the boundary of the annulus, C_1 is a closed set. Thus C_1 is a new countable, perfect set with $x_1 \notin C_1$. We can get an enumeration $(x_n^1)_{n \in \mathbb{N}}$ of C_1 by removing the elements of $C \setminus C_1$ from our original enumeration $(x_n)_{n \in \mathbb{N}}$.

2. Given $C_k \subseteq C$ perfect and countable with enumeration $(x_n^k)_{n \in \mathbb{N}}$ we apply the same procedure as before. We choose x_1^k and an annulus $\text{Ann}_{x_1^k}(\varepsilon_k/2, \varepsilon_k)$ with $\varepsilon_k < 1/k$ such that

$$C_{k+1} := C_k \cap \text{Ann}_{x_1^k}(\varepsilon_k/2, \varepsilon_k)$$

is again countably infinite and perfect. We update our enumeration $(x_n^k)_{n \in \mathbb{N}}$ to $(x_n^{k+1})_{n \in \mathbb{N}}$ by removing all elements that are not in C_{k+1} .

This recursion yields a falling sequence of closed sets C_k with diameter $2/k \rightarrow 0$. Since \mathbb{R} is complete there is an $x \in \mathbb{R}$ in the intersection of all C_k . But by construction for any $n \in \mathbb{N}_+$ there is a $k \in \mathbb{N}_+$ with $x_n \notin C_k$ and thus $x \neq x_n$ for all $n \in \mathbb{N}$. Thus we have found an element of C which is not part of the enumeration and therefore shown a contradiction. \square

The first part of the following lemma is the well known intersection theorem by Cantor (see for example Section 7.8 in [18]). The second part yields a variation of Cantor's intersection theorem for totally ordered collections of compact subsets of a compact, countable set.

Lemma 6.11 (Cantor's Intersection Theorem).

1. Let $(C_k)_{k \in \mathbb{N}_+}$ be a decreasing sequence of compact subsets of \mathbb{R} with $C_k \searrow \emptyset$. Then there is a $k \in \mathbb{N}_+$ with $C_k = \emptyset$.
2. Let $C \subseteq \mathbb{R}$ be compact and at most countable and $\{C_i\}_{i \in I}$ a collection of compact subsets of C (indexed over some index set I) which is totally ordered with respect to the inclusion. Furthermore let

$$\bigcap_{i \in I} C_i = \emptyset.$$

Then there is an $i \in I$ with $C_i = \emptyset$.

Proof.

1. We show this by contradiction. Let for all $k \in \mathbb{N}_+$ C_k be nonempty and choose $x_k \in C_k$. The sequence $(x_k)_{k \in \mathbb{N}_+} \subseteq C$ lives in a compact set and thus has a convergent subsequence. Without loss of generality we assume that $(x_k)_{k \in \mathbb{N}_+}$ is already convergent to some $x \in C$. Since C_k contains all x_n with $n \geq k$ we have $x \in C_k$ for all $k \in \mathbb{N}_+$. This is a contradiction to $C_k \searrow \emptyset$.
2. We want to apply the first part of the lemma and thus we have to find an actual sequence of C_i which falls to \emptyset . We construct this sequence $(C^k)_{k \in \mathbb{N}_+}$ by the following recursion. Note that in the finite setting the recursion will simply terminate after finitely many steps so we consider a countably infinite setting. Since C is countable there is an enumeration $(c_n)_{n \in \mathbb{N}_+}$ of C . We start by choosing some $i \in I$ and setting $C^1 := C_i$. Let now $C^k \subseteq C$ be given for some $k \in \mathbb{N}_+$ with $C^k \in \{C_i | i \in I\}$ and such that $\{c_1, \dots, c_k\} \cap C^k = \emptyset$. If $c_{k+1} \notin C^k$ we just set $C^{k+1} := C^k$. Let $c_{k+1} \in C^k$. Since the intersection over all C_i is empty we can find a $j \in I$ with $c_{k+1} \notin C_j$ and since the C_i are totally ordered and $c_{k+1} \in C^k$ we have $C_j \subseteq C^k$. Now we set $C_{k+1} := C_j$. By construction we have $c_k \notin C^k$ and thus $C^k \searrow \emptyset$. By the first part of the lemma we have a $k \in \mathbb{N}_+$ with $C^k = \emptyset$ and by construction there is an $i \in I$ with $C^k = C_i$ which shows the claim.

□

The following theorem corresponds to the concept of applying transfinite induction to the Cantor-Bendixson sequence (see Theorem 2.14 in [14] for the concept of transfinite induction over ordinals and the proof of Theorem 4.6 in [14] for the Cantor-Bendixson sequence).

Theorem 6.12. *Let $K \subseteq \mathbb{R}$ be closed and $O \subseteq K$ be open (in K). Let $C \subseteq K$ be closed and countable and set $U := O \setminus C$. Let \sim be an equivalence relation on O with the following properties:*

1. *If $x, y \in U$ are in the same connected component of U then $x \sim y$.*
2. *Let $z \in O$. If there is an $\varepsilon > 0$ with $x \sim y$ for all $x, y \in (z - \varepsilon, z)$ then $x \sim y$ for all $x, y \in (z - \varepsilon, z]$ and if there is an $\varepsilon > 0$ with $x \sim y$ for all $x, y \in (z, z + \varepsilon)$ then $x \sim y$ for all $x, y \in [z, z + \varepsilon)$.*

Then $x \sim y$ for all x, y in the same connected component of O .

Proof. We will show the result using Zorn's Lemma. Let A be the set of all countable and closed subsets $D \subseteq C$ with the property that for any x, y in the same connected component of $O \setminus D$ we have $x \sim y$. The set A is partially ordered with respect to the inclusion of sets and by assumption $C \in A$. Zorn's Lemma states that if every totally ordered subset of A has a minimal element then A itself has a minimal element (minimal in the sense that there is no strictly smaller element in A)

Thus we have to consider a totally ordered subset $\{C_i\}_{i \in I} \subseteq A$. Our first claim is that

$$E := \bigcap_{i \in I} C_i \in A.$$

Clearly E is an at most countable and closed subset of C since all these properties are inherited by intersections. Let $x \leq y$ be in the same connected component of $O \setminus E$. Set $\tilde{C}_i := C_i \cap [x, y]$. Then the set $\{\tilde{C}_i\}_{i \in I}$ is still totally ordered and the sets C_i are compact and at most countable subsets of $[x, y]$. Furthermore since $[x, y] \subseteq E^c$ we have

$$\bigcap_{i \in I} \tilde{C}_i \subseteq E \cap E^c = \emptyset.$$

Thus by 2. of Lemma 6.11 we can find an $i \in I$ with $\tilde{C}_i = \emptyset$. Therefore $[x, y] \subseteq O \setminus C_i$ and hence $x \sim y$ due to $C_i \in A$.

This shows $E \in A$ and E as the intersection over all $\{C_i\}_{i \in I}$ is a minimal element for $\{C_i\}_{i \in I}$.

By Zorn's Lemma we can now find an $F \in A$ minimal. Now we use the second property of our equivalence relation to show that this minimal element has no isolated points.

Assume that $O \setminus F$ has an isolated point $z \in O$. Consider $F' := F \setminus \{z\}$. Since z was isolated in $O \setminus F$ and since F is closed the set F' is still closed. Thus F' is an at most countable and closed subset of C . Let $x \leq y$ be in the same connected component of $O \setminus F'$. If x, y are in the same connected

component of $O \setminus F$ then $x \sim y$ due to $F \in A$. Otherwise we have $x \leq z \leq y$. If $x < z$ then x has to be in a connected component of $O \setminus F$ neighboring z on the left and if $z < y$ then y has to be in a connected component of $O \setminus F$ neighboring z to the right. Otherwise we have $x = z$ or $y = z$ respectively. In any case we have, by the second assumption for the equivalence relation, that $x \sim z \sim y$ and thus $F' \in A$. But since F was minimal this is impossible. Thus F has no isolated points and is a perfect subset of \mathbb{R} . But since F is at most countable Lemma 6.10 yields $F = \emptyset \in A$ and the equivalence relation holds on all connected components of O . \square

Remark 6.13. Note that stipulating $O \subseteq K$ open in K with K closed is equivalent to O being an intersection of an open and a closed subset of \mathbb{R} . But Theorem 6.12 can indeed be applied to a general subset O . The reason for this is that we only make statements about connected components of O . However, any such component $U \subseteq O$ will be an open subset of \bar{U} (the closure of U in \mathbb{R}). Since U was a component of O we have $\bar{U} \cap O = U$. Thus we can replace K by \bar{U} and O by U to get the result for each specific connected component of O .

The following example showcases a possible application of Theorem 6.12 and also the necessity for C to be countable and closed.

Example 6.14.

1. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be càdlàg and let the set C of discontinuities of f be a closed subset of \mathbb{R}_+ . If f is nonincreasing on each connected component of C^c and if for any jump of f at some x we have $f(x) \leq f(x-)$ then f is nonincreasing on \mathbb{R} . We can show this using Theorem 6.12 by setting $S = O = \mathbb{R}$ and $x \sim y$ for $x < y$ if and only if $f(x) \geq f(y)$. Note that if f is nonincreasing on (a, b) and on (b, c) and if $f(b-) \geq f(b)$ then f is nonincreasing on (a, c) . Hence \sim fulfills the assumptions of the theorem.
2. Let $f: \mathbb{R} \rightarrow [0, 1]$ be the Cantor function and C the cantor set. The function f is nonincreasing on connected components of C^c (even constant) and $f(x) = f(x-)$ for any $x \in C$. However f is not nonincreasing on \mathbb{R} , in fact it is nondecreasing. Of course the problem is that C is not countable thus Theorem 6.12 can not be applied.
3. Let $f: \mathbb{R} \rightarrow \mathbb{R}$ be a càdlàg function with set of discontinuities $C = \mathbb{Q}$ and with $f(q) \leq f(q-)$ for all $q \in \mathbb{Q}$. Then, the connected components of C^c are singletons $\{x\}$ with $x \in \mathbb{Q}^c$. So trivially f would be nonincreasing on connected components of C^c but of course f does not have to be

nonincreasing itself. Here the problem is that C is not closed and thus Theorem 6.12 can not be applied.

In applications of Theorem 6.12 U is usually given and $C = \partial U$ is the boundary of U . Thus we need a way to ensure that the boundary of the open set U is countable.

Proposition 6.15. *Let $U \subseteq \mathbb{R}_+$ be open. Let there be a function $f: U \rightarrow \mathbb{R}_+$ such that:*

1. $\limsup_{u \searrow t} f(u) < \infty$ for all $t \in \partial U$.
2. $\limsup_{u \nearrow t} f(u) = \infty$ for all $t \in \partial U$.

Then the boundary ∂U is at most countable.

Proof. It suffices to show that the boundary of $U \cap [0, n]$ is countable for all $n \in \mathbb{N}_+$. It suffices to consider the case $n = 1$, since the argument for a general $n \in \mathbb{N}_+$ works analogously. Thus we can assume without loss of generality $U \subseteq [0, 1]$. We start by showing that ∂U does not contain a strictly decreasing sequence. Let $x < y < z \in \partial U$. Since $y \in \partial U$ but $y \notin U$ we have $(x, y) \cap U \neq \emptyset$ or $(y, z) \cap U \neq \emptyset$. Thus there is an $u \in U$ with $u \in (x, z)$. Let $y' \in \partial U$ be the right boundary of the connected component of U which contains u . We have $y' > x$ and due to $z \notin U$ we also have $y' \leq z$. The second property of f with $t = y'$ yields that for any $c > 0$ we can find $u' \in (x, z) \cap U$ with $f(u') \geq c$.

Now consider a strictly decreasing sequence $(x_n)_{n \in \mathbb{N}} \subseteq \partial U$. The sequence is bounded from below by 0 and thus convergent. Since ∂U is closed we have $x_n \searrow x \in \partial U$ for $n \rightarrow \infty$. Using the previous argument we can find a decreasing sequence $(u'_n)_{n \in \mathbb{N}} \subseteq U$ with $f(u'_n) \geq n$ and $u'_n \rightarrow x$ for $n \rightarrow \infty$. But this is a contradiction to the first property of f since $(f(u'_n))_{n \in \mathbb{N}}$ is not a bounded sequence even though $(u'_n)_{n \in \mathbb{N}}$ is a strictly decreasing sequence in U .

Now we show that any set $\partial U \subseteq [0, 1]$ which contains no strictly decreasing sequences has to be countable. Let $x \in \partial U$. Since ∂U contains no decreasing sequences there is an $\varepsilon_x > 0$ with $\partial U \cap (x, x + \varepsilon_x) = \emptyset$ and $x + \varepsilon_x \leq 2$. By construction the open intervals $(x, x + \varepsilon_x)$ are disjoint for different x . Thus if there were uncountably many different $x \in \partial U$ then we would get uncountably many $\varepsilon_x > 0$. Thus we could find $\varepsilon > 0$ and infinitely many $x \in \partial U$ with $\varepsilon_x \geq \varepsilon > 0$. But $[0, 2]$ can not contain infinitely many disjoint intervals of length $\varepsilon > 0$ thus we get a contradiction. Therefore ∂U has to be countable which shows the claim. \square

6.3 Identifying the Distribution of a Time-Inhomogeneous Regular Jump Process

We now want to extend the backwards equation (Theorem 6.7) to the distribution π_t of our time-homogeneous Markov chain X as defined in (6.1).

Theorem 6.16 (Properties of π_t). *Consider the situation of Definition 6.2 and recall the definition of π_t (6.1). Let $(s, X_s) \in S$ be a random variable. Let $x \in E$. The function $g_x: [s, \infty) \rightarrow [0, 1]$ given by*

$$g_x(u) = \pi_{u;x} := \mathbb{E}(\mathbb{P}^{(s, X_s)}(X_u = x))$$

has the following properties:

1. g_x is right continuous on $[s, \infty)$ and continuous on $T_x \cap [s, \infty)$. The function $\sum_{x \in E} g_x$ is càdlàg and nonincreasing.
2. $g_x(u) = 0$ for $u \in T_x^c$.
3. Let E be finite, $x \in E$ and g_y be differentiable at $u > s$ for all $y \in E$. Then

$$g'_x(u) = \sum_{y \in E} \pi_{u;y} q_{u;y,x}. \quad (6.4)$$

4. Let E be finite and ∂T_x be countable for all $x \in E$. Then g_x is absolutely continuous on T_x with Lebesgue almost everywhere existing differential given by (6.4).

Proof.

1. Note that X has right-continuous paths. Thus dominated convergence yields that $g_x(u) = \pi_{u;x} = \mathbb{E}(\mathbb{P}^{(s, X_s)}(X_u = x))$ is right continuous. Let $u \in T_x$. Since the rate function $t \mapsto q_{t;x}$ is integrable in a neighborhood of u paths will almost surely only visit x finitely many times in such a neighborhood. Furthermore almost surely no arrow in our graphical representation is located at time u . Thus the paths of the process given by $\mathbb{1}_{\{X_t=x\}}$ are almost all continuous at u . By dominated convergence we get the continuity of g_x on T_x .

Furthermore, by Proposition 4.5 the graphical representation is almost surely proper and thus the left-sided limits for a path of X almost surely only fail at an explosion (see Definition 5.3). Note that $\sum_{x \in E} \mathbb{1}_{\{X_u=x\}}$

has càdlàg paths even if X explodes and dominated convergence yields that $\sum_{x \in E} g_x$ is càdlàg. We have

$$\sum_{x \in E} g_x(u) = 1 - \mathbb{P}^{(s, X_s)}(X_u = \partial).$$

Since ∂ is absorbing this function is nonincreasing.

2. This follows since (u, X_u) only takes values in S or $[s, \infty) \times \{\partial\}$.
3. Let E be finite, $x \in E$ and let g_y be differentiable at $u > s$ for all $y \in E$. We have for $s < v \leq u$, by using Theorem 6.7:

$$\begin{aligned} 0 &= \left. \frac{d}{dv} \pi_{u;x} \right|_{v=u} \\ &= \left(\left. \frac{d}{dv} \sum_{y \in E} \pi_{v;y} p_{y,x}(v, u) \right) \right|_{v=u} \\ &= \sum_{y \in E} \left. \frac{d}{dv} \pi_{v;y} \right|_{v=u} \delta_{x,y} + \sum_{y \in E} \pi_{u;y} \left. \frac{d}{dv} p_{y,x}(v, u) \right|_{v=u} \\ &= g'_x(u) - \sum_{y \in E} \pi_{u;y} \sum_{z \in E} q_{u;y,z} \delta_{z,x} \\ &= g'_x(u) - \sum_{y \in E} \pi_{u;y} q_{u;y,x}. \end{aligned}$$

Solving for $g'_x(u)$ yields the claim.

4. We first show absolute continuity on the set $A := \bigcap_{y \in E} (\partial T_y)^c$. Note that, since E is finite, A is open in \mathbb{R}_+ (as a finite intersection of open sets).

Let $s \leq v \leq w$ with $[v, w] \subseteq A \cap [s, \infty)$. This implies that for any $y \in E$ we have $[v, w] \subseteq T_y$ or $\pi_{v;y} = 0$ and $q_{u;y} = 0$ for all $u \in [v, w]$. We have due to Theorem 6.7:

$$\begin{aligned} |\pi_{w;x} - \pi_{v;x}| &\leq \sum_{y \in E} \pi_{v;y} |p_{y,x}(v, w) - \delta_{x,y}| \\ &= \sum_{y \in E} \pi_{v;y} |p_{y,x}(v, w) - p_{y,x}(w, w)| \\ &\leq \sum_{y \in E} \pi_{v;y} \int_v^w \sum_{z \in E} |q_{u;y,z}| p_{z,x}(u, w) du \\ &\leq \int_v^w 2 \sum_{y \in E} q_{u;y} du. \end{aligned}$$

Since E is finite the integrand on the right hand side is integrable and thus we get absolute continuity of g_x on $A \cap [s, \infty)$.

It is left to extend the absolute continuity to points in T_x . We will use the set theoretical Theorem 6.12 to show this. We consider $K := [s, \infty)$, $O := T_x \cap K$, $C := \bigcup_{y \in E} \partial T_y \cap K$ (which has countable boundary due to 4. of Lemma 4.2) and $v \sim w$ if and only if g_x is absolutely continuous on $[v, w]$ (if $v \leq w$) or $[w, v]$ (if $w \leq v$) and the differential of g_x on $[v, w]$ (or $[w, v]$ respectively) is Lebesgue almost everywhere given by (6.4). We have already shown that the equivalence relation holds on the connected components of $O \setminus C$. It is left to show that we can extend it to isolated points.

Let $z \in T_x$ and $\varepsilon > 0$ with g_x absolutely continuous on $(z - \varepsilon, z)$ as well as $(z, z + \varepsilon)$ and $[z - \varepsilon, z + \varepsilon] \subseteq T_x$ (this can be achieved since $z \in T_x$ and T_x open). In particular the function $u \mapsto q_{u;x}$ is integrable on $[z - \varepsilon, z + \varepsilon]$. Furthermore, let g'_x be Lebesgue almost everywhere given by (6.4) on the intervals $(z - \varepsilon, z)$ and $(z, z + \varepsilon)$. Let $v \in (z - \varepsilon, z)$, $w \in (z, z + \varepsilon)$ and $w_n \nearrow z$ as well as $v_n \searrow z$. Since $u \mapsto \pi_{u;x} q_{u;x}$ is integrable on $(z - \varepsilon, z + \varepsilon)$ the Lebesgue integral $\int_v^z g'_x(u) du$ is defined and has finite negative part. We have since g_x is continuous on $(z - \varepsilon, z + \varepsilon)$, by the fundamental theorem of calculus and applying monotone convergence to the positive and negative parts of g'_x respectively:

$$g_x(z) - g_x(v) = \lim_{n \rightarrow \infty} g_x(w_n) - g_x(v) = \lim_{n \rightarrow \infty} \int_v^{w_n} g'_x(u) du = \int_v^z g'_x(u) du.$$

Analogously we have

$$g_x(w) - g_x(z) = \lim_{n \rightarrow \infty} g_x(w) - g_x(v_n) = \lim_{n \rightarrow \infty} \int_{v_n}^w g'_x(u) du = \int_z^w g'_x(u) du.$$

In particular the positive part of g'_x also has to be Lebesgue integrable on $[v, z]$. Thus $g_x(w) - g_x(v) = \int_v^w g'_x(u) du$ for $[v, w] \subseteq (z - \varepsilon, z + \varepsilon)$ which shows absolute continuity and (6.4) on $(z - \varepsilon, z + \varepsilon)$ and therefore $v \sim w$. Theorem 6.12 implies that g_x is absolutely continuous on T_x . □

Remark 6.17. Note that since $t \mapsto q_{t;x}$ integrates to infinity when approaching the boundary ∂T_x from the left we have in particular $\limsup_{u \nearrow t, u \in T_x} q_{u;x} = \infty$ for $t \in \partial T_x$. If additionally $\limsup_{u \searrow t, u \in T_x} q_{u;x} < \infty$ for all $t \in \partial T_x$, for instance if $u \mapsto q_{u;x}$ has a right-continuous extension to \overline{T}_x , we can apply Proposition 6.15 to conclude that ∂T_x has to be countable.

Our aim is now to show that the properties of $t \mapsto \pi_t$ shown in Theorem 6.16 already determine it uniquely.

Theorem 6.18. *Let Q and π_s (for a fixed $s \in \mathbb{R}_+$) be as in Definition 6.2. Let E be finite and ∂T_x be countable for all $x \in E$. For $x \in E$ let $g_x: [s, \infty) \rightarrow \mathbb{R}$ be a function with the following properties:*

1. g_x is right continuous on $[s, \infty)$.
2. $g_x(t) = 0$ for $t \in T_x^c$.
3. g_x is absolutely continuous on T_x .
4. The differential of g_x fulfills the following differential equation Lebesgue almost everywhere:

$$g'_x(t) = \sum_{y \in E} g_y(t) q_{t;y,x}.$$

5. We have $g_x(s) \geq 0$ for all $x \in E$ and $\sum_{x \in E} g_x(s) < \infty$.

Then g_x fulfills $g_x(t) \geq 0$ for all $t \in [s, \infty)$, $x \in E$. Furthermore there exists exactly one family of functions with these properties and it is given by $\pi_{t;x}$ (see (6.1)) after rescaling the starting condition $g_x(s)$ to be a distribution (or it is given by the constant zero function, if $g_x(s) = 0$ for all $x \in E$).

Proof. We will refer to a family of functions $(g_x)_{x \in G}$ which fulfills 1. - 5. as a *solution* with initial condition $(g_x(s))_{x \in E}$.

We first note that showing $g_x(t) \geq 0$ for all $t \in [s, \infty)$, $x \in E$ implies uniqueness: Let g_x^1 and g_x^2 be two solutions then $g_x^1 - g_x^2$ and $g_x^2 - g_x^1$ are solutions (for the zero starting condition) since it is a linear problem. Thus we have $g_x^1 - g_x^2 \geq 0$ and $g_x^2 - g_x^1 \geq 0$ which implies $g_x^1 = g_x^2$. This shows uniqueness of solutions. If the starting condition is zero then obviously the constant zero functions solve the problem. Otherwise we note that the properties of g_x are well behaved under multiplication with positive scalars and thus we can assume $\sum_{x \in E} g_x(s) = 1$. (We can normalize due to 5.) Now Theorem 6.16 shows that the distribution of the time-inhomogeneous regular jump process generated by Q has the appropriate properties and thus has to be the unique solution to the problem.

It is left to show that solutions have to be nonnegative for a nonnegative initial condition. Consider a solution g_x with $g_x(s) \geq 0$ for all $x \in E$. Every g_x is absolutely continuous on $T_x \cap [s, \infty)$ and on the inner part of $T_x^c \cap [s, \infty)$ (due to 2.). Thus $g_x \wedge 0$ is also absolutely continuous on these sets with weak differential $g'_x \mathbf{1}_{\{g_x < 0\}}$. Since E is finite we have that $h := \sum_{x \in E} g_x \wedge 0$ is absolutely continuous on $\bigcap_{x \in E} (\partial T_x)^c \cap [s, \infty)$.

We have for the weak differential of h :

$$\begin{aligned}
h'(t) &= \sum_{x \in E} g'_x(t) \mathbf{1}_{\{g_x(t) < 0\}} \\
&\stackrel{4.}{=} \sum_{x \in E} \sum_{y \in E} g_y(t) q_{t;y,x} \mathbf{1}_{\{g_x(t) < 0\}} \\
&= \sum_{x \in E} \sum_{y \neq x} (g_y(t) q_{t;y,x} - g_x(t) q_{t;x,y}) \mathbf{1}_{\{g_x(t) < 0\}} \\
&\geq \sum_{x \in E} \sum_{y \neq x} (g_y(t) q_{t;y,x} - g_x(t) q_{t;x,y}) \mathbf{1}_{\{g_x(t) < 0\} \cap \{g_y(t) < 0\}} \\
&= 0.
\end{aligned}$$

Thus h is nondecreasing on connected components of $\bigcap_{x \in E} (\partial T_x)^c \cap [s, \infty)$.

Furthermore we note that if g_x is not continuous at some $t \in [s, \infty)$ then by 2. $g_x(t) = 0 = g_x(t) \wedge 0$ and thus

$$\limsup_{u \nearrow t} g_x(u) \wedge 0 \leq 0 = g_x(t) \wedge 0 \stackrel{1.}{=} \lim_{u \searrow t} g_x(u) \wedge 0.$$

Applying the (finite) sum over E on both sides yields:

$$\limsup_{u \nearrow t} h(u) \leq h(t) = \lim_{u \searrow t} h(u) = \liminf_{u \searrow t} h(u).$$

Consider a discontinuity t of h such that there is an $\varepsilon > 0$ with h being nondecreasing on $(t - \varepsilon, t)$ and $(t, t + \varepsilon)$, then h has to be nondecreasing on $(t - \varepsilon, t + \varepsilon)$. We now apply Theorem 6.12, setting

$$K := O := [s, \infty), \quad C := \bigcup_{x \in E} \partial T_x \cap K$$

and $x \sim y$ if h is nondecreasing on $[x, y]$ (if $x \leq y$) or on $[y, x]$ (if $y \leq x$). Note that since ∂T_x is countable by assumption the set C is also countable (and closed by definition).

The theorem yields that h is nondecreasing on $[s, \infty)$. But $h(s) = 0$ since $g_x(s) \geq 0$ for all $x \in E$ and $h \leq 0$ by definition. This implies $h(t) = 0$ for all $t \geq s$. But $h(t) = 0$ can only hold if $g_x(t) \geq 0$ for all $x \in E$ which shows the claim. □

In the case that E is countably infinite several problems arise with this approach. Sums can not necessarily be exchanged and $\bigcup_{x \in E} \partial T_x$ may no longer be closed. But under additional assumptions we can still identify the distribution of a time inhomogeneous Markov chain by solving the differential equation $\pi'_t = \pi_t \cdot Q$.

Theorem 6.19. *Let E be countable. Let Q be given as in Definition 6.2. Let $g_x(t) := \pi_{t;x}$ fulfill the assumptions of Theorem 6.18 with the additional conditions that $t \mapsto \sum_{x \in E} \pi_{t;x} q_{t;x}$ is integrable on compact subsets of $[s, \infty)$ and $t \mapsto \pi_{t;x}$ is absolutely continuous on $[s, \infty)$. Let $y \in E$, then the mapping*

$$h_y: [0, t] \rightarrow \mathbb{R}, \quad s \mapsto \sum_{x \in E} \pi_{s;x} p_{x,y}(s, t)$$

is constant. In particular: If we start the Markov chain X at time s in the distribution given by π_s it will have distribution π_t at time t . Furthermore, the mapping

$$g: \mathbb{R}_+ \rightarrow \mathbb{R}, \quad s \mapsto \sum_{x \in E} \pi_{s;x}$$

is also constant. If we can find a solution with $\pi_{x;0} > 0$ for a given $(0, x) \in S$ then X started in x at time 0 will almost surely never hit ∂ and thus it will almost surely not explode.

Proof. We have for $s \leq v \leq w \leq t$ by Theorem 6.7 and since π_s fulfills 4. of Theorem 6.18:

$$\begin{aligned} h_y(w) - h_y(v) &= \sum_{x \in E} \pi_{w;x} p_{x,y}(w, t) - \pi_{v;x} p_{x,y}(v, t) \\ &= \sum_{x \in E} (\pi_{w;x} - \pi_{v;x}) p_{x,y}(w, t) - \pi_{v;x} (p_{x,y}(v, t) - p_{x,y}(w, t)) \\ &= \sum_{x \in E} \left(\int_v^w \pi'_{u;x} p_{x,y}(w, t) \, du - \pi_{v;x} \int_v^w \sum_{z \in E} q_{u;x,z} p_{z,y}(u, w) \, du \right) \\ &= \sum_{x \in E} \left(\int_v^w \sum_{z \in E} \pi_{u;z} q_{u;z,x} p_{x,y}(w, t) \, du - \int_v^w \sum_{z \in E} \pi_{v;x} q_{u;x,z} p_{z,y}(u, w) \, du \right) \\ &= \sum_{x \in E} \int_v^w \sum_{z \in E} \pi_{u;z} q_{u;z,x} p_{x,y}(w, t) - \pi_{v;x} q_{u;x,z} p_{z,y}(u, w) \, du. \end{aligned}$$

Since $\sum_{x \in E} \pi_{u;x} q_{u;x}$ is integrable and nonnegative we can apply Fubini's Theorem to exchange the above sums which yields

$$h_y(w) - h_y(v) = h_y(v) - h_y(w) = 0.$$

Thus h_y is a constant function on $[s, t]$ and we get (using the notation for

matrix multiplication):

$$\begin{aligned}\pi_s \cdot P(s, t) &:= \left(\sum_{x \in E} \pi_{s;x} p_{x,y}(s, t) \right)_{y \in E} = (h_y(s))_{y \in E} \\ &= (h_y(t))_{y \in E} = \left(\sum_{x \in E} \pi_{t;x} p_{x,y}(t, t) \right)_{y \in E} \\ &= \pi_t P(t, t) = \pi_t.\end{aligned}$$

But the left hand side is exactly the distribution of X at time t if we start in the distribution π_s at time s .

Now we show that the function g is also constant. Since $u \mapsto \pi_{u;x}$ is absolutely continuous on \mathbb{R}_+ we have by 4. of Theorem 6.18 for all $0 \leq s \leq t$:

$$\begin{aligned}g(t) - g(s) &= \sum_{x \in E} \pi_{t;x} - \pi_{s;x} = \sum_{x \in E} \int_s^t \pi'_{u;x} \, du \\ &= \sum_{x \in E} \int_s^t \sum_{y \in E} \pi_{u;y} q_{u;y,x} \, du \\ &= \sum_{x \in E} \int_s^t \sum_{y \in E \setminus \{x\}} (\pi_{u;y} q_{u;y,x} - \pi_{u;x} q_{u;x,y}) \, du.\end{aligned}$$

Since by assumption the mapping

$$u \mapsto \sum_{x \in E} \pi_{u;x} q_{u;x} = \sum_{x \in E} \sum_{y \in E \setminus \{x\}} \pi_{u;x} q_{u;x,y}$$

is integrable on $[s, t]$ we can apply Fubini's theorem to exchange sums thus showing

$$g(t) - g(s) = g(s) - g(t) = 0$$

which implies that g is constant. Furthermore if π_0 is a distribution on E we have for all $t \in \mathbb{R}_+$:

$$\mathbb{P}(X_t \neq \partial) = \sum_{x \in E} \pi_{t;x} = g(t) = g(0) = \sum_{x \in E} \pi_{0;x} = 1.$$

This shows nonexplosiveness of the Markov chain X if started with the initial distribution π_0 . Now let $(0, x) \in S$ with $\pi_{0;x} > 0$. We can condition X started in the initial distribution π_0 on the event $\{X_0 = x\}$. Due to the Markov property of X this yields X started in the initial distribution $(\delta_{x,y})_{y \in E}$. Since the probability for the process started in π_0 to hit the graveyard state ∂ is zero the same has to be the case for the process conditioned on $\{X_0 = x\}$. Since an explosion would yield a jump to the graveyard state it will almost surely not happen. \square

Chapter 7

System of Flows

7.1 Flows and Particle Tracking

Two of our goals are to ensure that the rates of the limiting process given in Definition 3.10 fulfill the conditions of Definition 6.2 and that the process is non-explosive.

Consider the migration of a single ancestral line in the limiting process. If S is an admissible set for the migration, then it is sensible to consider the space of partitions with labels in S as the admissible set for the spatial coalescent. Furthermore, we note that by the nature of coalescence processes the number of ancestral lines can only decrease over time. Therefore non-explosiveness of the migration of a single ancestral line implies non-explosiveness of the whole limiting process. For that reason it makes sense to focus on the migration of a single ancestral line. It turns out that the migration, due to the way the rates are derived from \mathcal{F} and \mathcal{R} , follows a general concept which we will call a system of flows.

One visualization of this, which also explains the name, is the following. Consider a countable set G of tanks holding water. Any tank $x \in G$ holds some nonnegative amount of water $\mathcal{R}_{t,x}$ at any time $t \in \mathbb{R}_+$. Additionally water can be pumped from one tank x to another $y \neq x$. The total amount of water pumped from x to y up to time t is some nonnegative number $\mathcal{F}_{t,x,y}$, the *total flow from x to y up to time t* . Our migration now is modeled by tracking the path of a single water molecule through this system of tanks.

Of course, we cannot choose \mathcal{F} and \mathcal{R} arbitrarily and we will need some conditions to allow for a proper setup later on.

Definition 7.1 (A System of Flows). Let G be a countable set. We call a triplet $(G, \mathcal{R}, \mathcal{F})$, where \mathcal{R} and \mathcal{F} denote collections of functions $t \mapsto \mathcal{F}_{t,x,y}$ for $x \neq y \in G$, $t \geq 0$ and $t \mapsto \mathcal{R}_{t,x}$ for $x \in G$, $t \geq 0$, a *proper system of flows*

if the following properties hold for all $t \in \mathbb{R}_+$, $x \in G$, $y \in G \setminus \{x\}$:

1. Only nonnegative mass: $\mathcal{R}_{t;x} \geq 0$.
2. Only nonnegative flows: $\mathcal{F}_{t;x,y} \geq 0$ and the function $t \mapsto \mathcal{F}_{t;x,y}$ is nondecreasing.
3. Conservation of mass:

$$\mathcal{R}_{t;x} = \mathcal{R}_{0;x} + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} - \mathcal{F}_{t;x,y}.$$

4. The mass is normalized (at time 0): $\sum_{x \in G} \mathcal{R}_{0;x} = 1$.
5. Regularity of flows: the function $t \mapsto \mathcal{F}_{t;x,y}$ is absolutely continuous with weak differential $\mathcal{F}'_{t;x,y}$ wherever it exists.
6. There is no flow through sites with zero mass: $\mathcal{F}'_{t;x,y} = 0$ and $\mathcal{F}'_{t;y,x} = 0$ for all $y \in G$ and Lebesgue almost all t with $\mathcal{R}_{t;x} = 0$.
7. Globally finite flow in finite time: the function

$$\mathbb{R}_+ \rightarrow \mathbb{R}_+, \quad t \mapsto \sum_{x \in G} \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;x,y}$$

is integrable on compact subsets of \mathbb{R}_+ .

We call $\mathcal{R}_{t;x}$ the *mass at x at time t* and $\mathcal{F}_{t;x,y}$ the *cumulative flow from x to y up to time t* .

We now show, that conservation of mass implies that the system stays normalized and that absolutely continuous flows imply that the mass process \mathcal{R}_x is also absolutely continuous.

Proposition 7.2. *Given a proper system of flows $(G, \mathcal{R}, \mathcal{F})$ the following properties hold:*

1. *The mass stays conserved: $\sum_{x \in G} \mathcal{R}_{t;x} = 1$ for any $t \in \mathbb{R}_+$.*
2. *Let $x \in G$. The mapping $t \mapsto \mathcal{R}_{t;x}$ is absolutely continuous with weak differential:*

$$\mathcal{R}'_{t;x} = \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;y,x} - \mathcal{F}'_{t;x,y}.$$

Proof.

1. We have by the conservation of mass (see 3. of Definition 7.1):

$$\begin{aligned} \sum_{x \in G} \mathcal{R}_{t;x} &= \sum_{x \in G} \left(\mathcal{R}_{0;x} + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} - \mathcal{F}_{t;x,y} \right) \\ &= \sum_{x \in G} \mathcal{R}_{0;x} + \sum_{x \in G \setminus \{x\}} \mathcal{F}_{t;y,x} - \mathcal{F}_{t;x,y}. \end{aligned}$$

The first sum is equal 1 by the normalization property (4. of Definition 7.1). The second sum is zero since by Fubini's Theorem and by integrability (see 7. of Definition 7.1):

$$\begin{aligned} \sum_{x \in G} \sum_{y \neq x} \mathcal{F}_{t;x,y} &= \sum_{x \in G} \sum_{y \neq x} \int_0^t \mathcal{F}'_{u;x,y} \, du \\ &= \sum_{y \in G} \sum_{x \neq y} \int_0^t \mathcal{F}'_{u;x,y} \, du \\ &= \sum_{x \in G} \sum_{y \neq x} \mathcal{F}_{t;y,x}. \end{aligned}$$

Thus $\sum_{x \in G} \mathcal{R}_{t;x} = 1$ for all $t \geq 0$ which shows the first claim.

2. Let $0 \leq s \leq t$, $x \in G$. We have by the conservation of mass and absolute continuity of the flows:

$$\begin{aligned} \mathcal{R}_{t;x} - \mathcal{R}_{s;x} &= \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} - \mathcal{F}_{s;y,x} - (\mathcal{F}_{t;x,y} - \mathcal{F}_{s;x,y}) \\ &= \sum_{y \in G \setminus \{x\}} \int_s^t \mathcal{F}'_{u;y,x} - \mathcal{F}'_{u;x,y} \, du. \end{aligned}$$

Property 7. of Definition 7.1 in particular yields absolute integrability in the above equation and thus we can apply Fubini's Theorem, which yields:

$$\mathcal{R}_{t;x} - \mathcal{R}_{s;x} = \int_s^t \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{u;y,x} - \mathcal{F}'_{u;x,y} \, du.$$

Thus $t \mapsto \mathcal{R}_{t;x}$ is absolutely continuous with the claimed weak differential.

□

We go back to the interpretation at the beginning of the chapter. We consider the following question: If we sample one water molecule at time 0 in tank $x \in G$ (assuming we actually have water in that tank at time 0, of course). How will this molecule move through the tanks? Mathematically we want to model the movement as a stochastic process $(X_t)_{t \in \mathbb{R}_+}$ in G . It makes sense to assume the process $(X_t)_{t \in \mathbb{R}_+}$ to be Markov albeit not time-homogeneous since our macroscopic system of flows may change in time. What should the infinitesimal rate for moving from x to y be? Considering a small time segment $[t, t + \Delta t]$ the amount of mass traveling from x to y in that segment is $\mathcal{F}_{t+\Delta t; x, y} - \mathcal{F}_{t; x, y}$ the probability for our molecule to be part of this mass is approximately $(\mathcal{F}_{t+\Delta t; x, y} - \mathcal{F}_{t; x, y})/\mathcal{R}_{t; x} \sim \Delta t \mathcal{F}'_{t; x, y}/\mathcal{R}_{t; x}$. By the Poisson approximation of Binomial trials with small success probability it makes sense to assume that the infinitesimal rate is $\frac{\mathcal{F}'_{t; x, y}}{\mathcal{R}_{t; x}}$. Indeed it turns out, that choosing these rates will give us the rates of a time-inhomogeneous regular jump process as given in Definition 6.2.

Proposition 7.3 (Rates of the Particle Tracking Process). *Let $(G, \mathcal{R}, \mathcal{F})$ be a proper system of flows. We define for $x \in G$:*

$$S := \{(t, y) \in \mathbb{R}_+ \times G \mid \mathcal{R}_{t; y} > 0\}, \quad T_x := \{t \in \mathbb{R}_+ \mid (t, x) \in S\}.$$

Furthermore, for $t \geq 0$ we consider the matrix $Q_t := (q_{t; x, y})_{x, y \in G}$ defined via:

$$q_{t; x, y} := \frac{\mathcal{F}'_{t; x, y}}{\mathcal{R}_{t; x}} \quad \text{for } x \neq y \quad \text{and} \quad -q_{t; x, x} := q_{t; x} := \sum_{y \in G \setminus \{x\}} q_{t; x, y}.$$

If $(t, x) \notin S$ we formally set $q_{t; x, y} = 0$.

Then $t \mapsto Q_t$ fulfills the properties of Definition 6.2 for the admissible set of states S and thus defines a time-inhomogeneous regular jump process X for any starting condition $x \in G$ with $\mathcal{R}_{0; x} > 0$.

Proof. We show the five properties required by Definition 6.2.

1. Since $u \mapsto \mathcal{F}_{u; x, y}$ is nondecreasing and due to $\mathcal{R}_{u; t} \geq 0$ we have $q_{u; x, y} \geq 0$.
2. Note that by 2. of Proposition 7.2 the mapping $t \mapsto \mathcal{R}_{t; x}$ is continuous. Thus, for any compact set $K \subseteq T_x$ there is an $\varepsilon > 0$ such that $\mathcal{R}_{u; x} \geq \varepsilon$ for all $u \in K$. Therefore, we have by 7. of Definition 7.1:

$$\int_K q_{u; x} \, du \leq \frac{1}{\varepsilon} \int_K \sum_{y \in G} \mathcal{F}'_{u; x, y} \, du \leq \frac{1}{\varepsilon} \sum_{x \in G} \sum_{y \neq x} \int_K \mathcal{F}'_{u; x, y} \, du < \infty.$$

Thus $u \mapsto q_{u; x}$ is integrable on compact subsets of T_x and $u \mapsto q_{u; x, y}$ is integrable as well. In particular this implies stability almost everywhere.

3. The matrix Q_t is conservative by definition of the diagonal entries $q_{t;x,x}$.
4. Whenever $(u, x) \notin S$ or $(u, y) \notin S$ we have $\mathcal{F}'_{u;x,y} = 0$ and the same holds for the rates (using the formal convention in the definition).
5. Let $\mathcal{R}_{s;x} > 0$ and $\mathcal{R}_{t;x} = 0$. We have by 2. of Proposition 7.2:

$$\begin{aligned}
\int_s^t q_{u;x} \, du &= \int_s^t \sum_{y \neq x} \frac{\mathcal{F}'_{u;x,y}}{\mathcal{R}_{u;x}} \, du \\
&\geq \int_s^t \sum_{y \neq x} \frac{\mathcal{F}'_{u;x,y} - \mathcal{F}'_{u;y,x}}{\mathcal{R}_{u;x}} \, du \\
&= \int_s^t -\frac{\mathcal{R}'_{u;x}}{\mathcal{R}_{u;x}} \, du \\
&= \log(\mathcal{R}_{s;x}) - \log(\mathcal{R}_{t;x}) = \infty.
\end{aligned}$$

Since $(u, x) \in S$ is equivalent to $\mathcal{R}_{u;x} > 0$ which shows the claimed property.

□

We can now use our general theory for time-inhomogeneous regular jump processes to show that if we start the particle tracking process in the probability distribution given by the mass distribution $\pi_{0;x} := \mathcal{R}_{0;x}$ then the process will stay in the mass distribution: $\pi_{t;x} = \mathcal{R}_{t;x}$ for all $t \in \mathbb{R}_+$.

Proposition 7.4. *Let Q_t be given by a proper system of flows (see Definition 3.1) and let S and T_x be given as in Proposition 7.3. Define for $x \in G$ and $t \geq 0$: $g_x(t) := \pi_{t;x} := \mathcal{R}_{t;x}$. Then the assumptions of Theorem 6.19 (and in particular Theorem 6.18) are fulfilled for the initial condition $s = 0$, $\pi_{0;x} := \mathbb{R}_{0;x}$. Moreover, the particle tracking process conditioned on \mathcal{R} and \mathcal{F} has distribution $(\mathcal{R}_{t;x})_{x \in G}$ at time t and is almost surely non-explosive.*

Proof. First we have to show that $g_x(u) := \mathcal{R}_{u;x}$ fulfills the conditions of Theorem 6.18.

1. Due to 2. of Proposition 7.2 the mapping $u \mapsto \mathcal{R}_{u;x}$ is absolutely continuous on \mathbb{R}_+ and in particular right-continuous.
2. We have $\mathcal{R}_{u;x} = 0$ for $u \in T_x^c$ by the definition of S .
3. We have shown absolute continuity of g_x on \mathbb{R}_+ in 2. of Proposition 7.2.

4. We calculated the weak differential of g_x in 2. of Proposition 7.2. This yields for $u \in \mathbb{R}_+$, $x \in G$:

$$\begin{aligned} \sum_{y \in G} g_y(u) q_{u;y,x} &= \sum_{y \in G \setminus \{x\}} \mathcal{R}_{u;y} \frac{\mathcal{F}'_{u;y,x}}{\mathcal{R}_{u;y}} - \sum_{y \in G \setminus \{x\}} \mathcal{R}_{u;x} \frac{\mathcal{F}'_{u;x,y}}{\mathcal{R}_{u;x}} \\ &= \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{u;y,x} - \mathcal{F}'_{u;x,y} \\ &= \mathcal{R}'_{u;x} = g_x(u)'. \end{aligned}$$

5. We have $\mathcal{R}_{u;x} \geq 0$ by definition and $\sum_{x \in G} \mathcal{R}_{u;x} = 1$ for all $u \geq 0$ due to 1. of Proposition 7.2.

Since we already noted that g_x is absolutely continuous on \mathbb{R}_+ the only additional assumption of Theorem 6.19 is the integrability of $u \mapsto \sum_{x \in G} \pi_{u;x} q_{u;x}$ on compact subsets of \mathbb{R}_+ . We have for $u \geq 0$:

$$\sum_{x \in G} \pi_{u;x} q_{u;x} = \sum_{x \in G} \mathcal{R}_{u;x} \sum_{y \neq x} \frac{\mathcal{F}'_{u;x,y}}{\mathcal{R}_{u;x}} = \sum_{x \in G} \sum_{y \neq x} \mathcal{F}'_{u;x,y}.$$

By 7. of Definition 7.1 the right-hand side is integrable on compact subsets of \mathbb{R}_+ . Thus all conditions of Theorem 6.19 are fulfilled and application of the theorem yields that $(\mathcal{R}_{t;x})_{x \in G}$ is the distribution of the particle tracking process at time t and almost sure non-explosiveness of the process. \square

It turns out, that rewriting the rate matrix Q_t in terms of a system of flows can be done whenever we can identify the distribution of the time-inhomogeneous regular jump process via the differential equation $\pi'_t = \pi_t \cdot Q_t$.

Proposition 7.5. *Let $S \subseteq \mathbb{R}_+ \times G$ be an open set and let $Q_t = (q_{t;x,y})_{x,y \in G}$ be the rate matrix of a time-inhomogeneous regular jump process $X = (X_t)_{t \in \mathbb{R}_+}$ taking values in G with admissible set S . We define for an admissible state $(t, x) \in S$ the probability $\pi_{t;x} := \mathbb{P}(X_t = x)$. Assume that $t \mapsto \pi_{t;x}$ is absolutely continuous on \mathbb{R}_+ and that we have $\pi'_t = \pi_t \cdot Q_t$ where π' is the weak differential of π . Then X can be modeled using a system of flows.*

Proof. We consider the system of flows defined by:

$$\mathcal{F}'_{t;x,y} := \pi_{t;x} q_{t;x,y}, \quad \mathcal{R}_{0;x} := \pi_{0;x} \quad \text{for } x \neq y.$$

By definition the flows $\mathcal{F}_{\bullet;x,y}$ are absolutely continuous. We have for $\mathcal{R}_{t;x}$ using Tonelli's theorem:

$$\begin{aligned}
\mathcal{R}_{t;x} &= \mathcal{R}_{0;x} + \sum_{y \neq x} \mathcal{F}_{t;y,x} - \mathcal{F}_{t;x,y} \\
&= \mathcal{R}_{0;x} + \sum_{y \neq x} \int_0^t \mathcal{F}'_{u;y,x} - \mathcal{F}'_{u;x,y} \, du \\
&= \mathcal{R}_{0;x} + \sum_{y \neq x} \int_0^t \pi_{u;y} q_{u;y,x} - \pi_{u;x} q_{u;x,y} \, du \\
&= \mathcal{R}_{0;x} + \int_0^t (\pi_u \cdot Q_u)_x \, du \\
&= \mathcal{R}_{0;x} + \int_0^t \pi'_{u;x} \, du \\
&= \mathcal{R}_{0;x} + \pi_{t;x} - \pi_{0;x} = \pi_{t;x}.
\end{aligned}$$

Thus we have

$$\frac{\mathcal{F}'_{t;x,y}}{\mathcal{R}_{t;x}} = \frac{\pi_{t;x} q_{t;x,y}}{\pi_{t;x}} = q_{t;x,y}$$

as claimed. \square

Remark 7.6. Consider the situation of Proposition 7.5 and note that if $Q_t = Q$ is time-homogeneous and we are in equilibrium $\pi_t = \pi$ and thus $\pi'_t = 0 = \pi Q$ then the flows are time-homogeneous themselves and explicitly given as

$$\mathcal{F}'_{t;x,y} = \pi_x q_{xy}, \quad \mathcal{R}_{0;x} = \pi_x.$$

For instance, we could represent the rate 1, simple, symmetric random walk on the discrete torus $\mathbb{T}_L^d := \{-L, \dots, L\}^d$ for $L, d \in \mathbb{N}_+$ with time-homogeneous flows

$$\mathcal{R}_{0;x} = \frac{1}{(2L+1)^d}, \quad \mathcal{F}'_{t;x,y} = \frac{1}{(2(2L+1))^d} \quad \text{if } x, y \in \mathbb{T}_L^d \text{ are next neighbors.}$$

But Proposition 7.5 also shows, that even when we do not have an equilibrium distribution, for instance in the case of the simple symmetric random walk on \mathbb{Z}^d , we can still represent our time-homogeneous Markov chain using a system of flows. However, in that case the system of flows itself can no longer be chosen in a time-homogeneous manner even though the original Markov chain is time-homogeneous.

It is also noteworthy, that the generality in which Proposition 7.5 holds suggests, that considering a time-inhomogeneous regular jump process given by a system of flows is barely a restriction at all but rather a slightly different point of view.

7.2 Uniform Convergence of Flows

We now go back to the main Assumptions 3.6 in which we stipulated convergence of the finite dimensional distributions of \mathcal{F}^N to \mathcal{F} . The goal of this section is to improve this to weak convergence with respect to uniform convergence on compact sets. This is possible using that the processes $(\mathcal{F}_{t;x,y}^N)_{t \in \mathbb{R}_+}$ and $(\mathcal{F}_{t;x,y})_{t \in \mathbb{R}_+}$ are nondecreasing and that $(\mathcal{F}_{t;x,y})_{t \in \mathbb{R}_+}$ is continuous.

Our first goal is to show that the concept of pointwise convergence and uniform convergence on compact sets coincide in this case.

Proposition 7.7. *Let $O \subseteq \mathbb{R}_+$ be open. Let $f_n, f: \mathbb{R}_+ \rightarrow \mathbb{R}$ with:*

1. f_n and f are nondecreasing on connected components of O .
2. f is continuous on O .
3. There is a dense subset $Q \subseteq O$, with $0 \in Q$ if $0 \in O$, such that

$$f_n(q) \rightarrow f(q) \quad \text{for all } q \in Q.$$

Then $f_n \rightarrow f$ uniformly on compact subsets of O .

Proof. Since any compact subset of O is contained in a finite union of compact intervals contained in O it suffices to show uniform convergence on any compact interval $[s, t] \subseteq O$. Furthermore since $Q \subseteq O$ is dense and contains 0 whenever O contains 0 we can enlarge the interval such that $s, t \in Q$. Due to $[s, t] \subseteq O$ the functions f_n, f are nondecreasing on $[s, t]$ and f is uniformly continuous on $[s, t]$. Let $\varepsilon > 0$. Since f is uniformly continuous on $[s, t]$ there is a $\delta > 0$ such that $|f(v) - f(w)| < \varepsilon$ for $|v - w| < \delta$. Since Q is dense we can find $k \in \mathbb{N}$ and a partition $s = u_0 < u_1 < \dots < u_{k-1} < u_k = t$ of $[s, t]$ with $u_i \in Q$ and $|u_i - u_{i-1}| < \delta$ for all $i \in \{1, \dots, k\}$. Since we have $f_n(u_i) \rightarrow f(u_i)$ for all $i \in \{0, \dots, k\}$ we can find $N_\varepsilon \in \mathbb{N}$ such that for all $n \geq N_\varepsilon$, we have

$$\max_{i \in \{0, \dots, k\}} |f_n(u_i) - f(u_i)| < \varepsilon.$$

Now consider $u \in [s, t]$. We can find $i \in \{1, \dots, k\}$ such that $u \in [u_{i-1}, u_i]$. We have $f_n(u) \in [f_n(u_{i-1}), f_n(u_i)]$ and $f_n(u_i) \in (f(u_i) - \varepsilon, f(u_i) + \varepsilon)$ as well as $f_n(u_{i-1}) \in (f(u_{i-1}) - \varepsilon, f(u_{i-1}) + \varepsilon)$. We get $f_n(u) \in (f(u_{i-1}) - \varepsilon, f(u_i) + \varepsilon)$. But we also have $f(u) \in [f(u_{i-1}), f(u_i)]$ and $|f(u_i) - f(u_{i-1})| < \varepsilon$ and thus $f(u_i), f(u_{i-1}) \in (f(u) - \varepsilon, f(u) + \varepsilon)$. Therefore we get

$$f_n(u) \in (f(u_{i-1}) - \varepsilon, f(u_i) + \varepsilon) \subseteq (f(u) - 2\varepsilon, f(u) + 2\varepsilon).$$

Thus we have $|f_n(u) - f(u)| < 2\varepsilon$ for $n \geq N_\varepsilon$ uniformly for $u \in [s, t]$, which shows the claim. \square

We can now apply this proposition to improve the convergence of the processes \mathcal{F}^N and \mathcal{R}^N ,

Lemma 7.8. *Let G be at most countable. For $x \neq y \in G$ consider real-valued stochastic processes $\mathcal{F}_{x,y}^N := (\mathcal{F}_{t;x,y}^N)_{t \in \mathbb{R}_+}$, $\mathcal{F}_{x,y} := (\mathcal{F}_{t;x,y})_{t \in \mathbb{R}_+}$, $\mathcal{R}_x^N := (\mathcal{R}_{t;x}^N)_{t \in \mathbb{R}_+}$, $\mathcal{R}_x := (\mathcal{R}_{t;x})_{t \in \mathbb{R}_+}$ with the following properties:*

1. $\mathcal{F}_{x,y}^N \in D_{\mathbb{R}_+}[0, \infty)$, $\mathcal{F}_{x,y} \in C_{\mathbb{R}_+}[0, \infty)$.
2. $\mathcal{F}_{0;x,y}^N = \mathcal{F}_{0;x,y} = 0$ for all $x \neq y$.
3. $\mathcal{F}_{x,y}^N$ and $\mathcal{F}_{x,y}$ have nondecreasing paths.
4. For $x \in G$, $t \in \mathbb{R}_+$ we have almost surely

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N \in D_{\mathbb{R}_+}[0, \infty), \quad \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} \in D_{\mathbb{R}_+}[0, \infty)$$

and

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} \in C_{\mathbb{R}_+}[0, \infty), \quad \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y} \in C_{\mathbb{R}_+}[0, \infty).$$

5. We have for $x \in G$:

$$\mathcal{R}_{t;x}^N = \mathcal{R}_{0;x}^N + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N - \mathcal{F}_{t;x,y}^N \quad (7.1)$$

as well as

$$\mathcal{R}_{t;x} = \mathcal{R}_{0;x} + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} - \mathcal{F}_{t;x,y}. \quad (7.2)$$

6. Let for $x \in G$: $\mathcal{R}_{0;x}^N \rightarrow \mathcal{R}_{0;x}$ weakly.
7. For $x \in G$, $y \in G$ the finite dimensional distributions of $\mathcal{F}_{x,y}^N$ converge weakly to the finite dimensional distributions of $\mathcal{F}_{x,y}$.
8. For $x, y \in G$ we have

$$\sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x}^N \rightarrow \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;y,x} \quad \text{and} \quad \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}^N \rightarrow \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}$$

as weak convergence of finite dimensional distributions.

9. All weak convergences hold in a joint fashion for all $x, y \in G$

Then we have $(\mathcal{F}^N, \mathcal{R}^N) \rightarrow (\mathcal{F}, \mathcal{R})$ as weak convergence with respect to uniform convergence on compact subsets of $\mathbb{R}_+ \times G \times G$ for $\mathcal{F}^N \rightarrow \mathcal{F}$ and of $\mathbb{R}_+ \times G$ for $\mathcal{R}^N \rightarrow \mathcal{R}$.

Proof. We start off, by showing weak convergence with respect to uniform convergence on compact subsets for $\mathcal{F}_{x,y}^N$. Let $D_{\mathbb{R}_+}^{\nearrow}[0, \infty) \subseteq D_{\mathbb{R}_+}[0, \infty)$ be the subspace of nondecreasing càdlàg functions. Consider the mapping

$$\text{id}: D_{\mathbb{R}_+}^{\nearrow}[0, \infty) \rightarrow D_{\mathbb{R}_+}^{\nearrow}[0, \infty), \quad f \mapsto f$$

where we use the product topology (which corresponds to pointwise convergence) on the left hand side and topology of uniform convergence on compact sets on the right hand side. By Proposition 7.7 we know that this mapping is continuous at any $f \in D_{\mathbb{R}_+}^{\nearrow}[0, \infty)$ which is continuous. Thus, by the Continuous Mapping Theorem, weak convergence with respect to the product topology (which is weak convergence of finite dimensional distributions) yields weak convergence with respect to uniform convergence on compact sets as long as the processes are nondecreasing and the limiting process is almost surely continuous. This shows the claim for $\mathcal{F}_{x,y}^N$. If all weak convergences hold jointly for $x, y \in G$ so does the weak convergence with respect to the uniform convergence on compact sets and we get $\mathcal{F}^N \rightarrow \mathcal{F}$ weakly with respect to uniform convergence on compact subsets of $\mathbb{R}_+ \times G^2$.

Since the processes $\sum_{y \in G \setminus \{x\}} \mathcal{F}_{\bullet, y, x}^N, \sum_{y \in G \setminus \{x\}} \mathcal{F}_{\bullet, x, y}^N$, are also nondecreasing and their finite dimensional distributions converge to a nondecreasing, continuous limit we can also achieve weak convergence with respect to uniform convergence on compact sets of $\mathbb{R}_+ \times G \times G$ for these processes. Since the convergence holds jointly for all $x \in G$ and jointly with the convergence $\mathcal{R}_0^N \rightarrow \mathcal{R}_0$ we get by (7.1) and (7.2):

$$\begin{aligned} \mathcal{R}_{\bullet, x}^N &= \mathcal{R}_{0; x}^N + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{\bullet, y, x}^N - \sum_{y \in G \setminus \{x\}} \mathcal{F}_{\bullet, x, y}^N \\ &\rightarrow \mathcal{R}_{0; x} + \sum_{y \in G \setminus \{x\}} \mathcal{F}_{\bullet, y, x} - \sum_{y \in G \setminus \{x\}} \mathcal{F}_{\bullet, x, y} \\ &= \mathcal{R}_{\bullet, x}. \end{aligned}$$

as weak convergence with respect to uniform convergence on compact subsets of $\mathbb{R}_+ \times G$. Since this convergence holds jointly with the convergence of the \mathcal{F}^N we get $(\mathcal{F}^N, \mathcal{R}^N) \rightarrow (\mathcal{F}, \mathcal{R})$ weakly with respect to uniform convergence on compact subsets (of $\mathbb{R}_+ \times G^2$ and $\mathbb{R}_+ \times G$ respectively). \square

Remark 7.9. We can consider $(\mathcal{F}^N, \mathcal{R}^N)$ and $(\mathcal{F}, \mathcal{R})$ as elements of the product space $D_{\mathbb{R}_+}[0, \infty)^{G^2} \times D_{\mathbb{R}}[0, \infty)^G$. Note that the weak convergence with respect to uniform convergence on compact sets $(\mathcal{F}^N, \mathcal{R}^N) \rightarrow (\mathcal{F}, \mathcal{R})$ implies weak convergence with respect to the product topology in the space $D_{\mathbb{R}_+}[0, \infty)^{G^2} \times D_{\mathbb{R}}[0, \infty)^G$ (we use the Skorohod topology in each factor). This can be particularly useful since $D_{\mathbb{R}_+}[0, \infty)$ and $D_{\mathbb{R}}[0, \infty)$ are both Polish spaces with respect to the Skorohod topology (see [8]). In particular the countable product space $D_{\mathbb{R}_+}[0, \infty)^{G^2} \times D_{\mathbb{R}}[0, \infty)^G$ is Polish. Thus we can apply Skorohod's Representation Theorem to find a new probability space and versions of $\mathcal{F}^N, \mathcal{F}, \mathcal{R}^N, \mathcal{R}$ such that $(\mathcal{F}^N, \mathcal{R}^N) \rightarrow (\mathcal{F}, \mathcal{R})$ holds almost surely in $D_{\mathbb{R}_+}[0, \infty)^{G^2} \times D_{\mathbb{R}}[0, \infty)^G$. But since the limiting processes are continuous Skorohod convergence $\mathcal{F}_{\bullet;x,y}^N \rightarrow \mathcal{F}_{\bullet;x,y}$ and $\mathcal{R}_{\bullet;x}^N \rightarrow \mathcal{R}_{\bullet;x}$ implies uniform convergence on compact subsets of \mathbb{R}_+ . Thus we have $(\mathcal{F}^N, \mathcal{R}^N) \rightarrow (\mathcal{F}, \mathcal{R})$ almost surely with respect to uniform convergence on compact subsets of $\mathbb{R}_+ \times G \times G$ and $\mathbb{R}_+ \times G$ respectively.

Furthermore, we can even relax the weak convergence of finite dimensional distributions to only hold on a dense subset of \mathbb{R}_+ since this was already sufficient for applying Proposition 7.7.

Chapter 8

Proof of the Limiting Result

In this chapter we will prove our main result, Theorem 3.14, using the Assumptions 3.6.

8.1 Verification of the Properties of \mathcal{F} , \mathcal{R}

Recall Definition 3.1 for the definition of $(\mathcal{R}_k^N)_{k \in \mathbb{N}_0}$, $(\mathcal{F}_k^N)_{k \in \mathbb{N}_0}$. We define for $t \in \mathbb{R}_+$, $N \in \mathbb{N}_+$: $\mathcal{F}_t^N := \mathcal{F}_{k_t^N}^N$ and $\mathcal{R}_t^N := \mathcal{R}_{k_t^N}^N$. Since this is a slight abuse of notation we will make sure to always use indexes u, s, t when talking about the time-continuous processes and indexes k, n when talking about discrete-time processes in this proof. The only case when we plug in an actual number will be for $t = 0$ in which case both notations coincide. Furthermore, when we suppress the time index, writing $\mathcal{F}_{x,y}^N$ or \mathcal{R}_x^N we will always refer to the time-continuous case. Our first goal is to show that the conditions of Lemma 7.8 are fulfilled. Let $x \neq y \in G$.

1. $\mathcal{F}_{x,y}^N$ is càdlàg since it is even piecewise constant by definition and the constant extension to \mathbb{R}_+ is done in a right-continuous fashion. $\mathcal{F}_{x,y}$ is even absolutely continuous almost surely by assumption.
2. $\mathcal{F}_{0;x,y}^N = 0$ by definition and $0 = \mathcal{F}_{0;x,y}^N \rightarrow \mathcal{F}_{0;x,y}$ weakly implies $\mathcal{F}_{0;x,y} = 0$ almost surely.
3. $\mathcal{F}_{x,y}^N$ is nondecreasing by definition (as a cumulative expression). The limiting process $\mathcal{F}_{x,y}$ thus has to be almost surely nondecreasing due to $(\mathcal{F}_{s;x,y}^N, \mathcal{F}_{t;x,y}^N) \rightarrow (\mathcal{F}_{s;x,y}, \mathcal{F}_{t;x,y})$ weakly for $s \leq t$. (Technically we use the almost sure continuity of the paths of the limiting processes to ensure the property holds for all $s \leq t$ almost surely.)

4. The path $t \mapsto \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}^N$ is piecewise constant and right continuous. Furthermore we have that since every generation every individual in our population can migrate at most one time for $k \in \mathbb{N}$:

$$\sum_{x \in G} \sum_{y \in G \setminus \{x\}} \mathcal{F}_{k;x,y}^N - \mathcal{F}_{k-1;x,y}^N \leq 2N.$$

Thus the path is also real-valued. Now we want to show continuity for the limiting sums. We can even show absolute continuity. Let $t \geq 0$. We have, using Tonelli's Theorem and 4. of Assumptions 3.6 almost surely:

$$\begin{aligned} \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y} &= \sum_{y \in G \setminus \{x\}} \int_0^t \mathcal{F}'_{u;x,y} \, du \\ &= \int_0^t \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{u;x,y} \, du. \end{aligned}$$

In particular this holds for all $t \in \mathbb{Q}_+$ almost surely. Thus the paths $t \mapsto \sum_{y \in G \setminus \{x\}} \mathcal{F}_{t;x,y}$ are almost surely absolutely continuous with weak differential $t \mapsto \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;x,y}$. Analogously these arguments work as well if we exchange the roles of x and y inside the sum.

5. The total number of individuals at a given site x and time k can be calculated by taking the number of individuals at x at time 0 and then adding all individuals that ever migrated to x and subtracting all individuals that ever left x . This property still holds after normalization and the time rescaling (here we need that the time rescaling at x is independent from the site x). Thus equation (7.1) holds by definition of the migration. Furthermore, equation (7.2) is simply a consequence of the definition of \mathcal{R} .
6. This is part of 5. of Assumptions 3.6.
7. This is 3. of Assumptions 3.6.
8. This is 6. of Assumptions 3.6.
9. It was noted in the Assumptions 3.6 that all convergences are considered to hold jointly.

Thus we can apply Lemma 7.8 and get $(\mathcal{F}^N, \mathcal{R}^N) \rightarrow (\mathcal{F}, \mathcal{R})$ weakly with respect to uniform convergence on compact sets. Furthermore as noted in Remark 7.9 that we can use Skorohod's Representation Theorem to find a

probability space and versions of the stochastic processes $\mathcal{F}^N, \mathcal{R}^N, \mathcal{F}, \mathcal{R}$ such that the convergence holds almost surely. It suffices to prove the claim of Theorem 3.14 on this new probability space.

We now want to show that the limit $(\mathcal{F}, \mathcal{R})$ defines a proper system of flows as in Definition 7.1. We show the necessary properties. Let $x \neq y \in G$ and $t \in \mathbb{R}_+$:

1. $\mathcal{R}_{t;x} \geq 0$ follows since $\mathcal{R}_{t;x}^N \geq 0$ and $\mathcal{R}_{t;x}^N \rightarrow \mathcal{R}_{t;x}$ almost surely. Using continuity of \mathcal{R}_x we get \mathcal{R}_x nonnegative almost surely.
2. We have already seen that $t \mapsto \mathcal{F}_{t;x,y}$ is almost surely nondecreasing. Due to $\mathcal{F}_{0;x,y} = 0$ we also get that $\mathcal{F}_{t;x,y}$ is almost surely nonnegative.
3. This is just by definition of \mathcal{R} .
4. This is part of 5. of Assumptions 3.6.
5. This is part of 3. of Assumptions 3.6.
6. This is implied by 7. of Assumptions 3.6.
7. This is 4. of Assumptions 3.6.

Thus by Proposition 7.2 the mass is conserved in the limit ($\sum_{x \in G} \mathcal{R}_{t;x} = 1$ for all $t \in \mathbb{R}_+$ almost surely). Furthermore by Proposition 7.3 the migration of a single ancestral line in the limiting process Π is a time-inhomogeneous regular jump process with admissible set of states $S := \{(t, x) \in \mathbb{R}_+ \times G \mid \mathcal{R}_{t;x} > 0\}$.

8.2 Non-Explosiveness of the Limiting Process

We consider our setting conditioned on \mathcal{R}, \mathcal{F} for this section. The limiting process Π is driven by two independent mechanisms:

- Migration of ancestral lines following independent particle tracking processes given by $(\mathcal{R}, \mathcal{F})$ (see Definition 5.3).
- Independent coalescence at every site x with respect to the time rescaled Ξ -coalescent defined by the rates $\phi_{x,i}(l_1, \dots, l_i)$ (given i lines at site x).

Since any coalescence event has to decrease the number of ancestral lines by at least 1 we can see at most n coalescence events in any path of Π . Thus Π can only explode via migration and since migration of ancestral lines is

independent and migration is independent from coalescence the process Π explodes if and only if the particle tracking process for a single line explodes. But due to Proposition 7.4 this is almost surely not the case. Hence, the limiting process Π is non-explosive.

8.3 Modeling Π^N with Graphical Representations

By definition the limiting process Π is obtained using graphical representations but we have not yet obtained the coalescent of the population Π^N in this manner. We condition our setting on $\mathcal{R}^N, \mathcal{F}^N$ for this section.

Our time rescaling $k_t^N = \lfloor N^\alpha t \rfloor$ translates to a mesh $A_{k,x} := k/N^\alpha$. As we pointed out in the definition of the population coalescent Π^N (see after Definition 1.12) the coalescent fulfills the consistency relation. This means, that we can model coalescence and migration for a smaller sample $l \leq n$ by modeling it for n and restricting to l lines. We consider the following possible events for a sample of n ancestral lines located at site $x \in G$ at time k and define graphical representations for them:

1. We first consider the case that there are at least n individuals at site x at time k . Thus we can sample n ancestral lines at x at time k . Those lines will perform a migration in which line $i \in [n]$ migrates to site $y_i \in G$ at the time step $k \rightarrow k + 1$. The possible outcomes of such a migration can be coded using the vector $\vec{y} := (y_1, \dots, y_n)$. We also define $\vec{x} := (x, \dots, x) \in G^n$. We denote the probability of the aforementioned event as $p_{k;x,\vec{y}}$. We can combinatorically calculate the probability conditioned on $\mathcal{F}^N, \mathcal{R}^N$. We consider an urn experiment where we have an urn with $N_{k;x}^N = N \mathcal{R}_{k;x}^N$ balls having colors $y \in G$ such that the number of balls with color $y \neq x$ is given by $N(\mathcal{F}_{k+1;x,y}^N - \mathcal{F}_{k;x,y}^N)$ (the remaining balls have color x). Then $p_{k;x,\vec{y}}$ is the probability that if we draw n balls out of the urn without replacement we get the sequence of colors (y_1, \dots, y_n) . We define an appropriate intensity measure on \mathbb{R}_+ :

$$\nu_{x,\vec{y}}^{N,\text{Mig}} := \sum_{k \in \mathbb{N}} p_{k;x,\vec{y}} \delta_{k/N^\alpha}.$$

To find a Bernoulli process belonging to this intensity measure in the sense of Definition 4.18 we construct random variables $X_{k,x}^{(N,\text{Mig})}$ taking values in G^n independent in $k \in \mathbb{N}_+$ and $x \in G$ (we still work conditioned

on \mathcal{F}^N and \mathcal{R}^N) with

$$\mathbb{P}(X_{k,x}^{(N,\text{Mig})} = \vec{y}) = p_{k;x,\vec{y}}.$$

We call the resulting independent integer valued random measures $\mu_{x,\vec{y}}^{N,\text{Mig}}$. More precisely, we define:

$$\mu_{x,\vec{y}}^{N,\text{Mig}} := \sum_{k \in \mathbb{N}_+} \sum_{x \in G} \sum_{\vec{y} \in G^n} \mathbb{1}_{\{X_{k,x}^{(N,\text{Mig})} = \vec{y}\}} \delta_{(k/N^\alpha, x, \vec{y})}.$$

Now we consider the case that there are only $k < n$ individuals at x at time k . In this case the above approach will yield a measure $\mu_{x,\vec{y}}^{N,\text{Mig}}$ with $\vec{y} \in G^k$. We simply replace \vec{y} with $(y_1, \dots, y_k, x, \dots, x) \in G^n$ in order to achieve the same notation as previously. Later on we will ensure that if there only are k ancestral lines at a location then only the first k entries of \vec{y} are relevant for our construction.

2. We again assume that our sample of size n is fully located at $x \in G$ and that we have at least n individuals at x at time $k+1$ (Note that the migration step already happened thus we have to use $k+1$ instead of k here). Let $A \in \mathcal{P}_n$. The lines can perform an A -collision (see Definition 1.12), meaning that exactly the lines, which indices are contained in a mutual block of A , merge. We denote the probability of the coalescence event by $p_{k;x,A}$. It should be noted that since coalescence technically happens after the migration step of our population (backwards in time) the effective number of individuals for the coalescence is given by $\mathcal{N}_{k+1;x}^N = N\mathcal{R}_{k+1;x}^N$. Thus the coalescence probability at time k is calculated as in the nonspatial case using the population size $\mathcal{N}_{k+1;x}^N$. Now we define the appropriate intensity measure on \mathbb{R}_+ :

$$\nu_{x,A}^{N,\text{Coal}} := \sum_{k \in \mathbb{N}} p_{k;x,A} \delta_{k/N^\alpha}.$$

Again, we consider \mathcal{P}_n -valued random variables $X_{k,x}^{(N,\text{Coal})}$ independent in $k \in \mathbb{N}$ and $x \in G$ and independent from the random variables chosen in the migration step, with

$$\mathbb{P}(X_{k,x}^{(N,\text{Coal})} = A) = p_{k;x,A}$$

to define independent integer valued random measures $\mu_{x,A}^{N,\text{Coal}}$ similar to Definition 4.18:

$$\mu_{x,A}^{N,\text{Coal}} := \sum_{k \in \mathbb{N}_+} \sum_{x \in G} \sum_{A \in \mathcal{P}_n} \mathbb{1}_{\{X_{k,x}^{(N,\text{Coal})} = A\}} \delta_{(k/N^\alpha, x, \vec{y})}.$$

Now we consider the case that there are only $k < n$ individuals at x at time $k + 1$. The above construction yields a random measure $\mu_{x,A}^{N,\text{Coal}}$ with $A \in \mathcal{P}_k$. We now replace $A = (B_1, \dots, B_l)$ with $(B_1, \dots, B_l, \{k + 1\}, \dots, \{n\}) \in \mathcal{P}_n$ in order to achieve the same notation as in the case that we were able to sample n individuals. As is the case for migration it will turn out that if only k ancestral lines are available for coalescence then only $\tau_k^n(A)$ matters for the coalescence.

We can now combine these integer valued random measures into a graphical representation μ^N on $\mathbb{R}_+ \times \mathcal{P}_{G,n} \times \mathcal{P}_{G,n}$ for Π^N using the following procedure:

Consider any grid point $t = k/N^\alpha$ with $k \in \mathbb{N}_+$. For any labeled partition $\pi \in \mathcal{P}_{G,n}$ we consider independent copies of the measures $\mu_{x,A}^{N,\text{Coal}}$ and $\mu_{x,\vec{y}}^{N,\text{Mig}}$ for all $x \in G$, $A \in \mathcal{P}_n$, $\vec{y} \in G^n$. We order the non-empty blocks of π with respect to their smallest elements and give each block an index $i \in [n]$ according to the ordering (The block containing 1 gets index 1, the block which contains the next smallest element not contained in the first block gets index 2 and so on). Note that this particular choice of indexing the blocks will ensure that szenarios, in which we were not able to sample n individuals due to low sample size, will not produce any issues in the way we have treated them above. We now want to combine the migration and the coalescence step and apply them to π . We first consider possible migrations. For any site x there is exactly one $\vec{y} \in G$ such that $\nu_{x,\vec{y}}^{N,\text{Mig}}$ has an atom at (t, x, \vec{y}) . We define the labeled partition $\pi' \in \mathcal{P}_{G,n}$ by having the block with index k at site x move to y_k for all $k \in [n]$ and $x \in G$. Now we consider a possible coalescence. For any site x there is exactly one $A \in \mathcal{P}_n$ such that $\nu_{x,A}^{N,\text{Coal}}$ has an atom at (t, x, A) . We now define π'' by considering the labeled partition π' and for any site x merging all blocks at x which have their index contained in a block of A .

If and only if $\pi'' \neq \pi$ we draw an arrow from π to π'' at time t in our graphical representation μ^N . We define the intensity measure ν^N of μ^N according to Definition 4.18 (by taking the expectation).

We now show, that the graphical representation μ^N is indeed a graphical representation for our time-rescaled population coalescent $\hat{\Pi}^N$ (i.e., tracking through μ^N will give a process which has the same law as $\hat{\Pi}^N$ if the initial distributions coincide). Due to consistency it does not matter that the random measures $\mu_{x,\vec{y}}^{N,\text{Mig}}$, $\mu_{x,A}^{N,\text{Coal}}$ use the probabilities for the case that the full sample of n lines is still intact and at x . Also, note that coalescence of blocks in $\hat{\Pi}^N$ only depends on migration due to the change labels induced by migration. Since the random measures $\nu_{x,A}^{N,\text{Coal}}$ and $\nu_{x,\vec{y}}^{N,\text{Mig}}$ were chosen mutually independent we have that the probability of an arrow pointing from

π to π'' at time $t = k/N^\alpha$ is exactly given by the probability of $\hat{\Pi}^N$ jumping from π to π'' at time t . Additionally the random measures were chosen independently for different x which fits with the concept that migration and coalescence in $\hat{\Pi}^N$ is independent at different locations. This takes care of the joint distribution of jump times (and jump types) of $\hat{\Pi}^N$ and shows that tracking through ν^N will generate a stochastic process with the same law as Π^N .

On a side note: the reason why we also stipulated independence for different labeled partitions π is of a more technical nature and it has no impact on the law of the tracking process. Heuristically it does not matter how the joint distribution of arrows pointing out of different locations at the same time is chosen since the tracking procedure can only visit any time once. For our purposes choosing independence here fits with the concept of using Poisson point processes to define the limit since in such a random measure all disjoint sections are automatically independent.

8.4 Convergence of the Graphical Representations

The reason why we did not define μ^N directly but by using $\mu_{x, \vec{y}}^{N, \text{Mig}}$ and $\mu_{x, A}^{N, \text{Coal}}$ is that we can use independence together with Proposition 4.20 to analyse the limit of ν^N .

Assume, for the time being, that we have shown vague convergence of $\mu_{x, \vec{y}}^{N, \text{Mig}}$ and $\mu_{x, A}^{N, \text{Coal}}$ after removing atoms that belong to the trivial events $\vec{y} = (x, \dots, x)$ and $A = \{\{1\}, \dots, \{n\}\}$ to continuous limiting measures. By Proposition 4.20 and using independence we can conclude that only atoms that belong to either the migration of lines at one particular site x (and no further migration or coalescence) or atoms that belong to a coalescence event at one particular site x (and no other coalescence or migration) can yield a positive intensity in the limit. Furthermore, the limiting intensity for these events is exactly given by the limit of the intensity measure $\nu_{x, \vec{y}}^{N, \text{Mig}}$ or $\nu_{x, A}^{N, \text{Coal}}$ which belongs to said event.

Before we show the convergence we need to consider that our limiting process will only feature blocks with admissible labels corresponding to the set of admissible states $S := \{(t, x) \in \mathbb{R}_+ \times G \mid \mathcal{R}_{t,x} > 0\}$. Thus, for $x \in G$ we will show vague convergence of intensity measures as measures in $\mathcal{M}_c(T_x)$ with $T_x := \{t \in \mathbb{R}_+ \mid (t, x) \in S\}$.

Lemma 8.1. *The following convergences of intensity measures hold almost surely:*

1. *Let $x \in G$ and $\vec{y} \in G^n$ such that there is exactly one entry y_i of \vec{y} which is not equal to x . Then*

$$d\nu_{x, \vec{y}}^{N, \text{Mig}} \rightarrow \frac{\mathcal{F}'_{t;x,y}}{\mathcal{R}_{t;x}} dt$$

as vague convergence in $\mathcal{M}_c(T_x)$.

2. *Let $x \in G$ and $\vec{y} \in G^n$ such that there is more than one entry y_i of \vec{y} which is not equal to x . Then*

$$\nu_{x, \vec{y}}^{N, \text{Mig}} \rightarrow 0$$

as vague convergence in $\mathcal{M}_c(T_x)$.

3. *Let $x \in G$ and $A \in \mathcal{P}_n$, $A \neq \{\{1\}, \dots, \{n\}\}$ and let l_1, \dots, l_n denote the blocksizes of A . Then it holds for $N \rightarrow \infty$ with respect to the vague topology:*

$$d\nu_{x,A}^{N, \text{Coal}} \rightarrow \frac{\beta_x}{\mathcal{R}_{t;x}^\alpha} \phi_{x,n}(l_1, \dots, l_n) dt.$$

Proof.

1. Let $y \neq x$. For combinatorial reasons we start by showing the result for the following sum:

$$\sum_{\vec{y} \in A_y} \nu_{x, \vec{y}}^{N, \text{Mig}}$$

with $A_y \subseteq G^n$ being the set of all vectors for which the first entry is equal to $y = y_1$ (the other entries can be arbitrary).

Since the migration is done using an urn experiment we can calculate probabilities explicitly. The total number of individuals migrating from x to y at time k is given by $N\Delta \mathcal{F}_{k;x,y}^N := N(\mathcal{F}_{k;x,y}^N - \mathcal{F}_{k-1;x,y}^N)$ and the total number of individuals at site x is $N\mathcal{R}_{k;x}^N$. Thus, we have at time $t = k/N^\alpha$:

$$\mathbb{P}\left(\text{First entry of } X_{k_t^N, x}^{N, \text{Mig}} \text{ is equal } y_1\right) = \frac{N\Delta \mathcal{F}_{t;x,y}^N}{N\mathcal{R}_{t;x}^N} = \frac{\Delta \mathcal{F}_{t;x,y}^N}{\mathcal{R}_{t;x}^N}.$$

Thus for $[s, t] \subseteq T_x$:

$$\sum_{\vec{y} \in A_y} \nu_{x, \vec{y}}^{N, \text{Mig}}([s, t]) = \int_s^t \frac{1}{\mathcal{R}_{u;x}^N} d\mathcal{F}_{u;x,y}^N.$$

Since $[s, t] \subseteq T_x$ and since \mathcal{R}_x is continuous we have by definition of T_x that there is an $\varepsilon > 0$ such that $\mathcal{R}_{u;x} \geq \varepsilon$ for all $u \in [s, t]$. Additionally we have shown that $\mathcal{R}_{u;x}^N \rightarrow \mathcal{R}_{u;x}$ uniformly for $u \in [s, t]$ almost surely. Thus we have

$$\frac{1}{\mathcal{R}_{u;x}^N} \rightarrow \frac{1}{\mathcal{R}_{u;x}} \quad \text{uniformly for } u \in [s, t] \text{ almost surely.}$$

We estimate

$$\begin{aligned} & \left| \int_s^t \frac{1}{\mathcal{R}_{u;x}^N} d\mathcal{F}_{u;x,y}^N - \int_s^t \frac{1}{\mathcal{R}_{u;x}} d\mathcal{F}_{u;x,y} \right| \\ & \leq \int_s^t \left| \frac{1}{\mathcal{R}_{u;x}^N} - \frac{1}{\mathcal{R}_{u;x}} \right| d\mathcal{F}_{u;x,y}^N + \left| \int_s^t \frac{1}{\mathcal{R}_{u;x}} d\mathcal{F}_{u;x,y}^N - \int_s^t \frac{1}{\mathcal{R}_{u;x}} d\mathcal{F}_{u;x,y} \right| \end{aligned}$$

The first summand converges to zero due to uniform convergence of the integrand. The second summand converges to zero since $\mathcal{F}_{u;x,y}^N \rightarrow \mathcal{F}_{u;x,y}$ pointwise almost surely using 5. of Theorem 4.10 (convergence of measure generating functions implies vague convergence) and the fact that $u \mapsto \frac{1}{\mathcal{R}_{u;x}} \mathbf{1}_{[s,t]}(u)$ is Lebesgue almost everywhere continuous and supported on a compact set.

Thus we have for all $s, t \in \mathbb{Q}_+$ almost surely:

$$\lim_{N \rightarrow \infty} \sum_{\vec{y} \in A_y} \nu_{x, \vec{y}}^{N, \text{Mig}}((s, t]) = \int_s^t \frac{1}{\mathcal{R}_{u;x}} d\mathcal{F}_{u;x,y}.$$

Note that we will use this result in the next part of the proof in order to show that all summands which correspond to \vec{y} with 2 distinct entries which differ from x converge to zero. After we have shown that we can conclude:

$$\lim_{N \rightarrow \infty} \nu_{x, (y, x, \dots, x)}^{N, \text{Mig}}((s, t]) = \lim_{N \rightarrow \infty} \sum_{\vec{y} \in A_y} \nu_{x, \vec{y}}^{N, \text{Mig}}((s, t]) = \int_s^t \frac{1}{\mathcal{R}_{u;x}} d\mathcal{F}_{u;x,y}$$

almost surely. Due to exchangeability the result is the same if we consider the k -th coordinate to differ from x instead of the first coordinate.

2. Let $x \neq y \neq z \neq x \in G$. We start by showing that the following intensity measures converge to zero:

$$\sum_{\vec{y} \in B_y} \nu_{x, \vec{y}}^{N, \text{Mig}}, \quad \sum_{\vec{y} \in C_{y,z}} \nu_{x, \vec{y}}^{N, \text{Mig}}$$

with $B_y \subseteq G^n$ the set of all vectors for which the first two entries are equal to y and $C_{y,z}$ the set of all vectors for which the first entry is equal to y and the second entry is equal to z .

We consider the first measure. Let $[s, t] \subseteq T_x$ and $u \in [s, t]$. Again, we get an $\varepsilon > 0$ such that $\mathcal{R}_{u;x} \geq \varepsilon$ for all $u \in [s, t]$. We have, using the urn experiment:

$$\begin{aligned} \mathbb{P}\left(X_{k_u^N, x}^{N, \text{Mig}} \text{ has first and second entry } y\right) &= \frac{N\Delta\mathcal{F}_{u;x,y}^N(N\Delta\mathcal{F}_{u;x,y}^N - 1)}{N\mathcal{R}_{u;x}^N(N\mathcal{R}_{u;x}^N - 1)} \\ &= \frac{\Delta\mathcal{F}_{u;x,y}^N(N\Delta\mathcal{F}_{u;x,y}^N - 1)}{\mathcal{R}_{u;x}^N(N\mathcal{R}_{u;x}^N - 1)} \end{aligned}$$

Let N be large enough such that $\varepsilon N \geq 2$, we have

$$\frac{N\Delta\mathcal{F}_{u;x,y}^N - 1}{N\mathcal{R}_{u;x}^N - 1} \leq 2 \frac{N\Delta\mathcal{F}_{u;x,y}^N}{N\mathcal{R}_{u;x}^N} = 2 \frac{\Delta\mathcal{F}_{u;x,y}^N}{\mathcal{R}_{u;x}^N}.$$

Thus we get

$$\mathbb{P}(X_{k_u^N, x}^{N, \text{Mig}} \text{ has first and second entry } y) \leq 2\mathbb{P}(X_{k_u^N, x}^{N, \text{Mig}} \text{ has first entry } y)^2$$

and therefore, using the convergence we have shown in the previous part (for the sum over A_y) together with Proposition 4.20, we get

$$\sum_{\vec{y} \in B_y} \nu_{x, \vec{y}}^{N, \text{Mig}} \rightarrow 0$$

vaguely in $\mathcal{M}_c(T_x)$.

The second case can be treated similarly. With the same choices for $[s, t]$, ε and N as before we have:

$$\begin{aligned} &\mathbb{P}\left(X_{k_u^N, x}^{N, \text{Mig}} \text{ has first entry } y \text{ and second entry } z\right) \\ &= \frac{N\Delta\mathcal{F}_{u;x,y}^N N\Delta\mathcal{F}_{u;x,z}^N}{N\mathcal{R}_{u;x}^N(N\mathcal{R}_{u;x}^N - 1)} = \frac{\Delta\mathcal{F}_{u;x,y}^N N\Delta\mathcal{F}_{u;x,z}^N}{\mathcal{R}_{u;x}^N(N\mathcal{R}_{u;x}^N - 1)} \leq 2 \frac{\Delta\mathcal{F}_{u;x,y}^N \Delta\mathcal{F}_{u;x,z}^N}{\mathcal{R}_{u;x}^N \mathcal{R}_{u;x}^N} \\ &\leq 2\mathbb{P}\left(X_{k_u^N, x}^{N, \text{Mig}} \text{ has first entry } y\right) \mathbb{P}\left(X_{k_u^N, x}^{N, \text{Mig}} \text{ has first entry } z\right) \end{aligned}$$

Again, the previous result (for the sum over A_y) together with Proposition 4.20 yields

$$\sum_{\vec{y} \in C_{y,z}} \nu_{x, \vec{y}}^{N, \text{Mig}} \rightarrow 0$$

vaguely in $\mathcal{M}_c(T_x)$ almost surely. Due to exchangeability this implies that the intensity with which any simultaneous migration happens converges to zero.

3. We first condition on the event $\{m = \mathcal{N}_{k_t^N+1;x}^N\}$ for $m \in \{n, \dots, N\}$. Denote the probability for our sample to perform an A -collision for $A \in \mathcal{P}_n$ as $p_{x,A}(m)$. Since the offspring laws only depend on the site and the total number of individuals at that site we get for the unconditioned probability: $p_{x,A}(\mathcal{N}_{k_t^N+1;x}^N)$. Note that we use $k_t^N + 1$ since the migration of the population happens before the coalescence backwards in time. This will, of course, not make a difference in the limit. For notational convenience we will write $\mathcal{N}_{t+;x}^N := \mathcal{N}_{k_t^N+1;x}^N$ and $\mathcal{R}_{t+;x}^N := \mathcal{R}_{t+1/N^{\alpha};x}^N$.

We first note that we have from the nonspatial result and sample size n (see Equation (28) by [20]) due to 2. of Assumptions 3.6:

$$\frac{p_{x,A}(m)}{c_x^m} \rightarrow \phi_{x,n}(l_1, \dots, l_n) \quad \text{for } m \rightarrow \infty.$$

Now let $[s, t] \subseteq T_x$ and $\varepsilon > 0$ with $\mathcal{R}_{u;x} \geq \varepsilon$ for all $u \in [s, t]$ as before. Since $\mathcal{R}_{u;x}^N \rightarrow \mathcal{R}_{u;x}$ uniformly in $u \in [s, t]$ we have

$$\mathcal{N}_{u+;x}^N = N\mathcal{R}_{u+;x}^N \geq N\varepsilon/2$$

uniformly in $u \in [s, t]$ for N large. In particular we get

$$\frac{p_{x,A}(\mathcal{N}_{u+;x}^N)}{c_x^{\mathcal{N}_{u+;x}^N}} \rightarrow \phi_{x,n}(l_1, \dots, l_n).$$

for $N \rightarrow \infty$ uniformly for $u \in [s, t]$ almost surely. Furthermore, by 1. of Assumptions 3.6 we have as well

$$c_x^{\mathcal{N}_{u+;x}^N} (\mathcal{N}_{u+;x}^N)^{\alpha} \rightarrow \beta_x$$

for $N \rightarrow \infty$ uniformly in $u \in [s, t]$ almost surely. And of course we also have

$$\frac{N^{\alpha}}{(\mathcal{N}_{u+;x}^N)^{\alpha}} = \frac{1}{(\mathcal{R}_{u+;x}^N)^{\alpha}} \rightarrow \frac{1}{\mathcal{R}_{u;x}^{\alpha}}$$

for $N \rightarrow \infty$ uniformly for $u \in [s, t]$ almost surely. Together we get

$$\begin{aligned} N^{\alpha} p_{x,A}(\mathcal{N}_{u+;x}^N) &= \frac{N^{\alpha}}{(\mathcal{N}_{u+;x}^N)^{\alpha}} (\mathcal{N}_{u+;x}^N)^{\alpha} c_x^{\mathcal{N}_{u+;x}^N} \frac{p_{x,A}(\mathcal{N}_{u+;x}^N)}{c_x^{\mathcal{N}_{u+;x}^N}} \\ &\rightarrow \frac{\beta_x}{\mathcal{R}_{u;x}^{\alpha}} \phi_{x,n}(l_1, \dots, l_n) \end{aligned}$$

for $N \rightarrow \infty$ uniformly in $u \in [s, t]$ almost surely.

Now we get for the intensity measure:

$$\nu_{x,A}^{N,\text{Coal}}((s, t]) = \sum_{k=k_s^N+1}^{k_t^N} p_{x,A}(\mathcal{N}_{k+1;x}^N) = \int_{[sN^\alpha]N^{-\alpha+1}}^{\lfloor tN^\alpha \rfloor N^{-\alpha}} N^\alpha p_{x,A}(\mathcal{N}_{u+;x}^N) du$$

Thus, the uniform convergence of the integrand yields

$$\nu_{x,A}^{N,\text{Coal}}((s, t]) \rightarrow \int_s^t \frac{\beta_x}{\mathcal{R}_{u;x}^\alpha} \phi_{x,n}(l_1, \dots, l_n) du$$

for $N \rightarrow \infty$ almost surely. Since it suffices for vague convergence to show convergence of the measures of intervals $(s, t]$ with $s, t \in \mathbb{Q}_+$ this yields almost sure vague convergence. □

8.5 Concluding the Proof

We now show convergence of Π^N conditioned on $(\mathcal{F}^N, \mathcal{R}^N)$ to Π conditioned on $(\mathcal{F}, \mathcal{R})$ as weak convergence in the Skorohod topology. Let $T \geq 0$. Since all processes are in the Skorohod space it suffices to show convergence up to time T .

We use the graphical representation μ^N and the tracking map to construct the spatial coalescent Π^N . Note that if we consider the set of admissible states (for the labels) to be S instead of $\mathbb{R}_+ \times G$ we will only get the process Π^N if our tracking process never visits ∂ . We have shown that the intensity measure ν^N of μ^N converges vaguely to the intensity measure of the graphical representation of our limiting process. By Proposition 4.19 this implies weak convergence of the graphical representations with respect to the vague topology. We have already seen that the graphical representation of our limit process yields a non-explosive, time-inhomogeneous, regular jump process for any admissible starting condition. Furthermore, since the intensity measures of the limiting process are continuous, we can apply the Continuous Mapping Theorem together with Proposition 5.5 to show weak convergence of the respective tracking processes with respect to the Skorohod topology up to time T . In particular the tracking process for μ^N will eventually (for N large) become a $\mathcal{P}_{G,n}$ -valued càdlàg process up to time T and then coincide with Π^N .

Thus we get $\Pi^N \rightarrow \Pi$ up to time T weakly in the Skorohod topology conditioned on $(\mathcal{F}^N, \mathcal{R}^N)$ and $(\mathcal{F}, \mathcal{R})$ respectively almost surely. As usual

considering a sequence $T_N \nearrow \infty$ we can achieve Skorohod convergence of the whole processes. Technically, it is important here that Π^N itself is always a $\mathcal{P}_{G,n}$ -valued càdlàg process. This would not necessarily be the case for the tracking process due to the set of admissible states S (for the labels). However, since our convergence result holds, we can conclude that the labels of blocks in Π^M will stay in the admissible set up to time T for sufficiently large, random $M \in \mathbb{N}_+$. Thus, the probability that the tracking process differs from Π^N up to time T converges to zero for $N \rightarrow \infty$.

It is left to show convergence without the conditioning on the mass and flow processes $(\mathcal{F}^N, \mathcal{R}^N)$ and $(\mathcal{F}, \mathcal{R})$. Let $Y: D_G[0, \infty) \rightarrow \mathbb{R}$ be a continuous, bounded functional. Weak convergence of the conditional distributions implies

$$\mathbb{E}(Y(\Pi^N) | \mathcal{F}^N, \mathcal{R}^N) \rightarrow \mathbb{E}(Y(\Pi) | \mathcal{F}, \mathcal{R}) \quad \text{almost surely.}$$

Taking expectations on both sides and using dominated convergence we get

$$\mathbb{E}(Y(\Pi^N)) \rightarrow \mathbb{E}(Y(\Pi))$$

which shows weak convergence $\Pi^N \rightarrow \Pi$ with respect to the Skorohod topology. This finally completes the proof of Theorem 3.14.

Chapter 9

Outlook

Our main result was Theorem 3.14 which showed convergence of the coalescent of a spatial Cannings model to a spatial Ξ -coalescent. We expect that this result as well as some results that were used in the proof and are of interest in their own right and can be generalized in the following way.

We first consider a generalization of the main result. Our hypothesis is that the global integrability condition in 4. of Assumption 3.6 can be relieved to a local version. In fact this condition was only required for showing non-explosiveness of the limiting process. Thus, the following conjecture would allow us to mitigate the assumption for the main result.

Conjecture 9.1. *Our limiting process as given in Definition 3.10 is non-explosive even if instead of $t \mapsto \sum_{x \in G} \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;x,y}$ being integrable on compact subsets of \mathbb{R}_+ we only have that for each $x \in G$ the functions*

$$t \mapsto \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;x,y} \quad \text{and} \quad t \mapsto \sum_{y \in G \setminus \{x\}} \mathcal{F}'_{t;y,x}$$

are integrable on compact subsets of \mathbb{R}_+ and $\sum_{x \in G} \mathcal{R}_{t;x} = 1$ for all $t \geq 0$.

Intuition for this is provided by the comparison of the particle tracking process with the concept of tracking a water molecule through a system of water tanks (see the beginning of Section 7.1). For this heuristics, having a global condition on the flows like 4. of Assumption 3.6 is not necessary, it suffices that our cumulative flows are locally finite. We build upon the heuristics by choosing a molecule uniformly at random at time zero and tracking it through the system up to some time t . It makes sense, that the location of the molecule at time t has the same distribution as if we simply sample a molecule uniformly at random at time t . If the system is still conservative, in the sense that the total mass stays constant (this no

longer guaranteed in general) this heuristics suggests that the particle tracking process is almost surely non-explosive, since the marginal distributions of the process have no probability on ∂ . We can also give a simple example in which global flow can be infinite but the particle tracking process would still be non-explosive:

Example 9.2. Consider the following system of flows on $G = \mathbb{N}_0$. We start with mass 2^{-k} at even numbers $2k \in \mathbb{N}_0$ and no mass at odd numbers. We also only allow for flow between $2k$ and $2k + 1$. We choose this flow in a way such that the total mass flown from $2k$ to $2k + 1$ and from $2k + 1$ to $2k$ is equal 1 at time $t = 1$. For example having 2^{k-1} full mass exchanges (all mass gets pumped from $2k$ to $2k + 1$ and then all mass gets pumped back) would yield the desired flow. Even though the local flow is still finite, the cumulative, global flow at time 1 would be infinite since every pair $(2k, 2k + 1)$ supplies a cumulative flow equal 1 at time 1. But note that, if we consider the particle tracking process started at $2k \in \mathbb{N}_0$, then most of the behavior of the system is not relevant. In fact, the process can only jump between $2k$ and $2k + 1$. Thus we can consider the setting to be finite for the purposes of our analysis and in particular the process would not explode.

Next, consider that we proved non-explosiveness of our limiting process by identifying the distribution of a time-inhomogeneous Markov chain by solving the differential equation $\pi'_t = \pi_t \cdot Q_t$. In particular Theorem 6.18 was of special interest but it was proven only for the case of a finite state space E . For the infinite case we needed additional assumptions (see Theorem 6.19). This leads us to the following conjecture.

Conjecture 9.3. *The result of Theorem 6.18 should hold for E countable and for arbitrary ∂T_x .*

It should be noted, that this result would have its own merits since it also allows for computation of the explosion probability $1 - \sum_{x \in E} \pi_{t;x}$ in cases when explosions are indeed possible. As intuition we can again look at a system of flows, but include the absorbing graveyard state ∂ as a possible location for a water molecule, which is reached if the path of the water molecule explodes. The previous heuristics still makes sense and it should not matter whether E is infinite or how the boundaries of T_x look like.

We now consider a different type of generalization. It should be noted, that we normalized our total mass at time 0 (see 5. of Assumptions 3.6). This fits together with the idea that the global population size of our spatial Cannings model is considered as finite $N < \infty$. But we could relax this condition and only require local population sizes to be finite. In terms of the masses this would yield a situation in which $\sum_{x \in G} \mathcal{R}_{t;x} = \infty$ but $\mathcal{R}_{t;x} < \infty$ for

$x \in G, t \geq 0$. For example consider $G = \mathbb{Z}$ and assume a transition invariant setting. One possibility would be to start with N individuals at each $x \in G$ and let all individuals perform i.i.d. simply symmetric random walks. In this case the global mass will be infinite and thus it would not be possible to treat such a setting with our current tools. We could, nevertheless, try a similar approach as in the finite case. Again, the main issue comes up when trying to show the non-explosiveness of the limiting process. It should be noted that a system that carries infinite mass will usually also feature infinite global cumulative flows. So it makes sense to prove Conjectures 9.1 and 9.3 for the case of finite total mass before moving on to an infinite total mass situation. Another step that has to be taken in order to generalize to infinite mass would be to find an analogous condition to the conservation of mass in Conjecture 9.3.

We defined our limiting process using a time-dependent rate matrix Q_t (see Definition 3.10. But the definition of a time-inhomogeneous, regular jump process (see Definition 6.2) can be generalized by using continuous intensity measures instead of restricting to absolutely continuous measures. Thus, it would be interesting to examine, whether our result can be generalized to a setting with only continuous intensity measures in the limit. Though this would mainly be of theoretical interest for getting a “bigger picture” and not be relevant for application.

Another possible generalization would be to not consider an exchangeable setting but rather allowing for different types of individuals in the population that have different offspring laws. This is, for example, the case if our population is subjected to natural selection. If we have additional information about genetical advantages and disadvantages of individuals in our sample, then the offspring law is no longer exchangeable. Thus, this would be needed in order to analyse population models that feature selection and mutation in a meaningful way.

Bibliography

- [1] Bardour, A. D., Holst, L., and Janson, S. (1992). *Poisson Approximation*. Clarendon Press.
- [2] Barton, N., Etheridge, A., and Véber, A. (2010). A New Model for Evolution in a Spatial Continuum. *Electronic Journal of Probability*, 15:162–216.
- [3] Cannings, C. (1974). The Latent Roots of Certain Markov Chains Arising in Genetics: A New Approach, I. Haploid Models . *Advances in Applied Probability*, 6(2):260–290.
- [4] Cannings, C. (1975). The Latent Roots of Certain Markov Chains Arising in Genetics: A New Approach. II. Further Haploid Models . *Advances in Applied Probability*, 7(2):264–282.
- [5] Durrett, R. and Schweinsberg, J. (2004). Approximating Selective Sweeps. *Theoretical Population Biology*, 66:129–138.
- [6] Durrett, R. and Schweinsberg, J. (2005). Random Partitions Approximating the Coalescence of Lineages During a Selective Sweep. *The Annals of Applied Probability*, 15:1591–1651.
- [7] Eldon, B. and Wakeley, J. (2006). Coalescent Processes When the Distribution of Offspring Number Among Individuals is Highly Skewed. *Genetics*, 172:2621–2633.
- [8] Ethier, S. and Kurtz, T. (2002). *Markov Processes: Characterization and Convergence*. Wiley.
- [9] Greven, A., den Hollander, F., Kliem, S., and Klimovsky, A. (2014). Renormalisation of Hierarchically Interacting Cannings Processes. *Latin American Journal of Probability and Mathematical Statistics*, 11.

- [10] Greven, A., Sun, R., and Winter, A. (2015). Continuum Space Limit of the Genealogies of Interacting Fleming-Viot Processes on \mathbb{Z} . *arXiv:1508.07169 [math.PR]*.
- [11] Herbots, H. (1997). The Structured Coalescent. *Progress in Population Genetics and Human Evolution*, pages 231–255.
- [12] Heuer, B. and Sturm, A. (2013). On Spatial Coalescents With Multiple Mergers in Two Dimensions. *Theoretical Population Biology*, 87:90–104.
- [13] Hössjer, O. (2011). Coalescence Theory for a General Class of Structured Populations with Fast Migration. *Advances in Applied Probability*, 43(4):1027–1047.
- [14] Jech, T. (2006). *Set Theory*. Springer, 3rd edition.
- [15] Joshi, K. D. (1983). *Introduction to General Topology*. Wiley Eastern Limited.
- [16] Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer, 1st edition.
- [17] Kingman, J. F. C. (1982). On the Genealogy of Large Populations. *Journal of Applied Probability*, 19:27–43.
- [18] Lewin, J. (2003). *An Interactive Introduction to Mathematical Analysis*. Cambridge University Press.
- [19] Limic, V. and Sturm, A. (2006). The Spatial Λ -Coalescent. *Electronic Journal of Probability*, 11:363–393.
- [20] Möhle, M. and Sagitov, S. (2001). A Classification of Coalescent Processes for Haploid Exchangeable Population Models. *The Annals of Probability*, 29(4):1546–1562.
- [21] Notahara, M. (1990). The Coalescent and the Genealogical Process in Geographically Structured Population. *Journal of Mathematical Biology*, 29:59–75.
- [22] Pitman, J. (1999). Coalescents With Multiple Collisions. *The Annals of Probability*, 27(4):1870–1902.
- [23] Sagitov, S. (1999). The General Coalescent With Asynchronous Mergers of Ancestral Lines. *Journal of Applied Probability*, 36(4):1116–1125.

-
- [24] Sagitov, S., Jagers, P., and Vatutin, V. (2010). Coalescent Approximation for Structured Populations in a Stationary Random Environment. *Theoretical Population Biology*, 78(3):192–199.
- [25] Schuhmacher, D. and Ross, N. (2016). Wireless Network Signals with Moderately Correlated Shadowing Still Appear Poisson. *arXiv:1606.05825 [cs.NI]*.
- [26] Schuhmacher, D. and Xia, A. (2008). A New Metric Between Distributions of Point Processes. *Advanced Applications in Probability*, 40:651–672.
- [27] Schweinsberg, J. (2000). Coalescents with Simultaneous Multiple Collisions. *Electronic Journal of Probability*, 5:1–50.
- [28] Taylor, R. L. and Hu, T. (1987). Strong Laws of Large Numbers for Arrays of Rowwise Independent Random Elements. *International Journal of Mathematics and Mathematical Sciences*, 10(4):805–814.
- [29] Wakeley, J. (2009). *Coalescent Theory: an Introduction*. Roberts.