Resampling and genealogies in sequential Monte Carlo algorithms

Susanna Elizabeth Brown

A thesis submitted for the degree of
Doctor of Philosophy in Statistics

University of Warwick, Department of Statistics

January 2021

Contents

Acknowledgements **v**

Abstract **vi**

List of Acronyms **vii**

Notation **viii**

1 Introduction **1**

2 Background **2**

2.1 Sequential Monte Carlo . . . 2
2.1.1 Motivation . . . 2
2.1.2 Inference in SSMs . . . 2
2.1.3 Exact solutions . . . 2
2.1.4 Feynman-Kac models . . . 3
2.1.5 Sequential Monte Carlo for Feynman-Kac models . . . 3
2.1.6 Theoretical justification . . . 4
2.2 Coalescent theory . . . 4
2.2.1 Kingman's coalescent . . . 4
2.2.2 Properties . . . 5
2.2.3 Models in population genetics . . . 7
2.2.4 Particle populations . . . 8
2.3 Sequential Monte Carlo genealogies . . . 9
2.3.1 From particles to genealogies . . . 9
2.3.2 Performance . . . 9
2.3.3 Mitigating ancestral degeneracy . . . 9
2.3.4 Asymptotics . . . 9
2.4 Resampling . . . 9
2.4.1 Definition . . . 9
2.4.2 What makes a good resampling scheme? . . . 9
2.4.3 Examples . . . 10
2.4.4 Stochastic rounding . . . 10
2.5 Conditional SMC . . . 10
2.5.1 Particle MCMC . . . 10
2.5.2 Particle Gibbs algorithm . . . 10

2.5.3 Ancestor sampling . . . 10

3 Limits **11**

3.1 Encoding genealogies . . . 11

3.1.1 The genealogical process . . . 11

3.1.2 Time scale . . . 11

3.1.3 Transition probabilities . . . 12

3.2 An existing limit theorem . . . 15

3.3 A new limit theorem . . . 15

3.3.1 Proof of theorem . . . 16

4 Applications **17**

4.1 Multinomial resampling . . . 17

4.1.1 Proof of main condition . . . 17

4.1.2 Proof of finite time scale condition . . . 17

4.2 Stochastic rounding . . . 17

4.2.1 Proof of main condition . . . 17

4.2.2 Finite time scale . . . 18

4.3 Stratified resampling . . . 21

4.4 The worst possible resampling scheme . . . 21

4.5 Conditional SMC . . . 21

4.5.1 Proof of main condition . . . 21

4.5.2 Finite time scale . . . 23

4.5.3 Effect of ancestor sampling . . . 24

5 Weak Convergence **25**

5.1 Bounds on sum-products . . . 30

5.2 Main components of weak convergence . . . 33

5.3 Indicators . . . 51

5.4 Other useful results . . . 56

5.5 Dependency graph . . . 57

6 Discussion **59**

List of Figures

2.1 Conditional dependence structure of SMC algorithm 4
2.2 The n-coalescent 5
2.3 Diagram illustrating the definitions of t_i, T_i in the n-coalescent. 6

5.1 Structure of weak convergence proof 58

List of Tables

Acknowledgements

I would like to thank...

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Parts of this thesis have been published by the author:

List of Acronyms

SMC	sequential Monte Carlo
i.i.d.	independent and identically distributed
MRCA	most recent common ancestor

Notation and conventions

$\mathbb{N}$	the natural numbers starting from one, $\{1, 2, \dots\}$
$\mathbb{N}_0$	the natural numbers starting from zero, $\{0, 1, 2, \dots\}$
$[a]$	the set $\{1, 2, \dots, a\}$ where $a \in \mathbb{N}$
$(a)_b$	the falling factorial $a(a-1)\cdots(a-b+1)$ where $a, b \in \mathbb{N}$
$\prod_\emptyset$	the empty product is taken to be 1
$\sum_\emptyset$	the empty sum is taken to be 0, while the sum over an index vector of length zero is the identity operator ?
$\mathcal{F}_t$	the (backward) filtration generated by offspring counts up to time t
$\mathbb{E}$	expectation
$\mathbb{E}_t$	filtered expectation $\mathbb{E}[\cdot \mid \mathcal{F}_{t-1}]$
A^c	denotes the complement of set A

1 Introduction

2 Background

Anyone who considers arithmetical methods of producing random digits is, of course, in a state of sin.

JOHN VON NEUMANN

2.1 Sequential Monte Carlo

2.1.1 Motivation

Being Bayesian. SSMs/HMMs. Example(s) of SSM (1D train?).

2.1.2 Inference in SSMs

What quantities do we want to infer? Why is this generally difficult? Filtering, prediction, smoothing, likelihood/normalising constant.

2.1.3 Exact solutions

This section needs redrafting, but all the content I wanted is here.

In the case of linear Gaussian state space models, the posterior distributions of interest are also Gaussian, with mean and covariance available analytically by way of the Kalman filter (Kalman 1960) and Rauch-Tung-Striebel (RTS) smoother recursions (Rauch, Striebel, and Tung 1965). Recursions are also available for some other conjugate models: see for example Vidoni (1999). Another analytic case occurs if the state space $\mathcal{X}$ is finite, in which case any integrals become finite sums, and the forward-backward algorithm (Baum et al. 1970) yields the exact posteriors.

If the model is Gaussian but non-linear, the posterior filtering distributions can be estimated using the *extended Kalman filter* (see for example Jazwinski (2007)), which applies a first-order linearisation in order to make use of the Kalman filter. This method performs well on models that are "almost linear". The resulting predictor is only *optimal* when the model is actually linear, in which case the extended Kalman filter coincides with the Kalman filter.

For models that are highly non-linear or for which gradients are not readily available, a more suitable method is the *unscented Kalman filter* (Wan and Merwe 2000). This involves taking a representative sample (which is chosen deterministically using the *unscented transformation*) to characterise the distribution at time t, and then propagating these points through the non-linear transition F to obtain a characterisation of the distribution at time $t+1$. This is getting closer to SMC, hmm?

In more complex models such techniques are not feasible, and we are forced to resort to Monte Carlo methods. For state space models, Markov chain Monte Carlo methods perform woefully due to the high dimension of the parameter space and high correlation between dimensions. But we can exploit the sequential nature of the underlying dynamics to decompose the problem into a sequence of inferences of more manageable dimension. This is the motivation behind sequential Monte Carlo (SMC) methods.

2.1.4 Feynman-Kac models

Define a generic FK model. Show that this class includes all SSMs. Example of non-SSM that is FK?

2.1.5 Sequential Monte Carlo for Feynman-Kac models

Present generic algorithm. State the SMC estimators of the quantities of interest. Include the dependence diagram and note that the offspring counts are not independent at each time, but can be made so by conditioning on the separatrix $\mathcal{H}$.

Data: $N, T, \mu, (K_t)_{t=1}^T, (g_t)_{t=0}^T$
for $i \in \{1, \dots, N\}$ **do** Sample $X_0^{(i)} \sim \mu(\cdot)$
for $i \in \{1, \dots, N\}$ **do** $w_0^{(i)} \leftarrow \left\{\sum_{j=1}^N g_0(X_0^{(j)})\right\}^{-1} g_0(X_0^{(i)})$
for $t \in \{0, \dots, T-1\}$ **do**
  Sample $a_t^{(1:N)} \sim \text{RESAMPLE}(\{1, \dots, N\}, w_t^{(1:N)})$
  for $i \in \{1, \dots, N\}$ **do** Sample $X_{t+1}^{(i)} \sim K_{t+1}(X_t^{(a_t^{(i)})}, \cdot)$
  for $i \in \{1, \dots, N\}$ **do** $w_{t+1}^{(i)} \leftarrow \left\{\sum_{j=1}^N g_{t+1}(X_t^{(a_t^{(j)})}, X_{t+1}^{(j)})\right\}^{-1} g_{t+1}(X_t^{(a_t^{(i)})}, X_{t+1}^{(i)})$
end

Algorithm 1: Sequential Monte Carlo

Figure 2.1 shows part of the conditional dependence graph implied by Algorithm 1. Our aim is to find a σ-algebra $\mathcal{H}_t$ at each time t that separates the ancestral process (encoded by $a_t^{(1:N)}$) from the filtration $\mathcal{F}_{t-1}$. That is, $a_t^{(1:N)}$ is conditionally independent of $\mathcal{F}_{t-1}$ given $\mathcal{H}_t$. By a D-separation argument (see Verma and Pearl 1988), the nodes highlighted in grey suffice as the generator of $\mathcal{H}_t$. That is, for each t, we take

$$\mathcal{H}_t = \sigma(X_{t-1}^{(1:N)}, X_t^{(1:N)}, w_{t-1}^{(1:N)}, w_t^{(1:N)}).$$

Notice that $\nu_t^{(1:N)}$ can be expressed as a function of $a_t^{(1:N)}$, and as such carries less information.

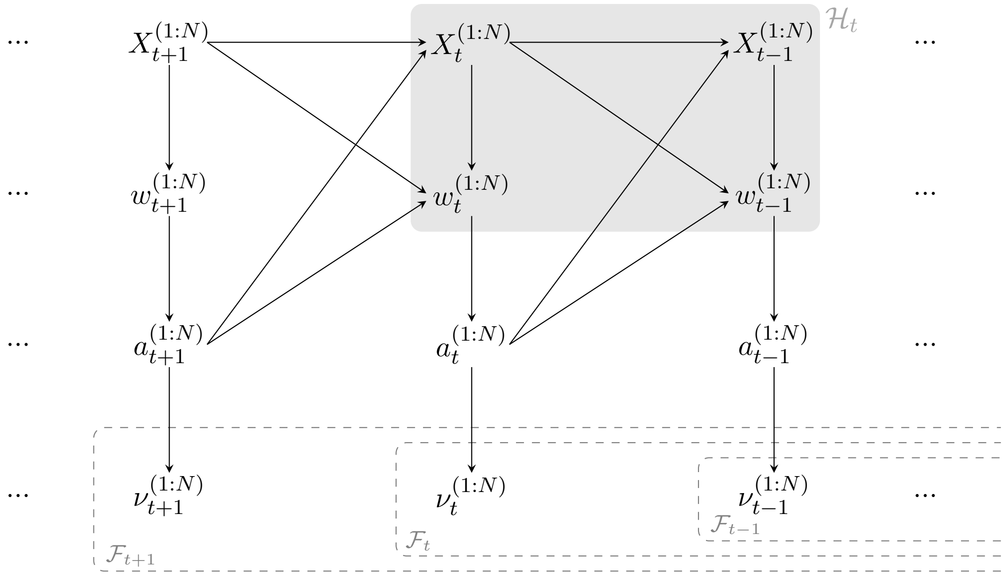

Figure 2.1: Part of the conditional dependence graph implied by Algorithm 1. The direction of time is from left to right. The reverse-time filtration is indicated by the dashed areas. The nodes highlighted in grey generate the separatrix $\mathcal{H}_t$ between $a_t^{(1:N)}$ and $\mathcal{F}_{t-1}$. Use the same shades of grey here as elsewhere

2.1.6 Theoretical justification

How come SMC works? Convergence results (briefly!) e.g. Lp bounds, CLT, stability.

2.2 Coalescent theory

2.2.1 Kingman's coalescent

The Kingman coalescent (Kingman 1982b; Kingman 1982c; Kingman 1982a) is a continuous-time Markov process on the space of partitions of $\mathbb{N}$. For our purposes we need only consider its restriction to $\{1, \ldots, n\}$, termed the n-coalescent (defined below), since we only ever consider finite samples from a population. However, an excellent probabilistic introduction to the Kingman coalescent from the point-of-view of exchangeable random partitions can be found in Berestycki (2009, Chapters 1–2). or Wakeley (2009) ? or Richard Durrett (2008) ?

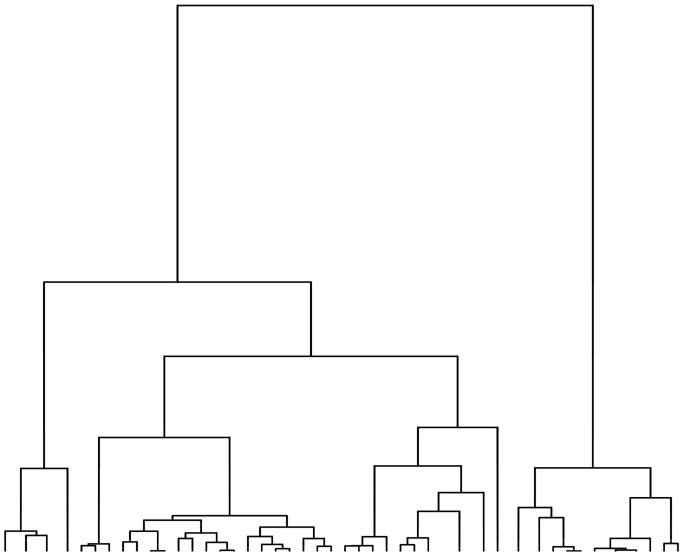

Figure 2.2: A realisation of the n-coalescent with $n = 50$.

Definition 2.1. The *n-coalescent* is the homogeneous continuous-time Markov process on the set of partitions of $\{1, \dots, n\}$ with infinitesimal generator Q having entries

$$q_{\xi,\eta} = \begin{cases} 1 & \xi \prec \eta \\ -|\xi|(|\xi|-1)/2 & \xi = \eta \\ 0 & \text{otherwise} \end{cases} \tag{2.1}$$

where ξ and η are partitions of $\{1, ..., n\}$, $|\xi|$ denotes the number of blocks in ξ, and $\xi \prec \eta$ means that η is obtained from ξ by merging exactly one pair of blocks.

A particularly attractive feature of the n-coalescent is its tractability; its distribution and those of many statistics of interest are available in closed form (Section 2.2.2). It turns out also to be extremely useful as a limiting distribution in population genetics, including the genealogies of a wide range of population models in its domain of attraction (Section 2.2.3).

2.2.2 Properties

The simplicity of Q allows various properties of the n-coalescent to be studied analytically. Refer to more exhaustive studies of the properties in the literature, e.g. Richard Durrett (2008, Section 1.2). Starting with n blocks, exactly $n-1$ coalescences are required to reach the absorbing state where all blocks have coalesced, known in the population genetics literature as the *most recent common ancestor* (MRCA).

Denote by $t_2, t_3 \dots, t_n$ the waiting times between coalescent events, where t_i is the amount of time for which the coalescent has exactly i distinct lineages (see Figure 2.3).

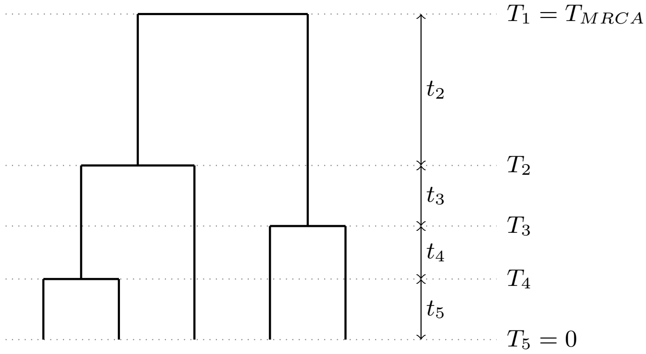

Figure 2.3: Diagram illustrating the definitions of t_i, T_i in the n-coalescent.

A consequence of Definition 2.1 is that these waiting times are independent and have distributions

$$t_i \sim \mathrm{Exp}\left(\binom{i}{2}\right). \tag{2.2}$$

The partial sum $T_k := \sum_{i=k+1}^{n} t_i$ gives the total time up to the $(n-k)^{th}$ coalescence event, i.e. the first time at which there are only k lineages remaining out of the initial n (see Figure 2.3).

Of particular interest is the tree height or time to the most recent common ancestor, $T_{MRCA} := T_1$. With some algebra we find, for instance,

$$\mathbb{E}[T_{MRCA}] = \sum_{i=2}^{n} \mathbb{E}[t_i] = \sum_{i=2}^{n} \frac{2}{i(i-1)} = 2\sum_{i=2}^{n} \left\{\frac{1}{i-1} - \frac{1}{i}\right\} = 2\left(1 - \frac{1}{n}\right) \tag{2.3}$$

and

$$\mathrm{Var}[T_{MRCA}] = \sum_{i=2}^{n} \mathrm{Var}[t_i] = \sum_{i=2}^{n} \left(\frac{2}{i(i-1)}\right)^2. \tag{2.4}$$

The expected tree height converges to 2 as $n \to \infty$, and the variance converges to $4(\pi^2 - 9)/3 \simeq 1.16$. The somewhat surprising fact that the tree height does not diverge with n is a result of the very high rate of coalescence close to the bottom of the tree. This rate is large enough that the full Kingman coalescent (on $\mathbb{N}$) *comes down from infinity*, that is, despite starting with infinitely many blocks, after any positive amount of time these have coalesced into finitely many blocks. Plot mean with sd-ribbon over n for an illustration? SD ribbon isn't the right thing; since we apparently know the actual distribution, plot a high density interval of that. (also for L)

Another quantity of interest is the total branch length, $L := \sum_{i=2}^{n} i t_i$. For instance

$$\mathbb{E}[L] = \sum_{i=2}^{n} i\mathbb{E}[t_i] = \sum_{i=2}^{n} \frac{2}{i-1} = \sum_{i=1}^{n-1} \frac{2}{i} \simeq 2\ln(n-1) \tag{2.5}$$

and

$$\mathrm{Var}[L] = \sum_{i=2}^{n} i^2 \,\mathrm{Var}[t_i] = \sum_{i=2}^{n} \frac{4}{(i-1)^2} = \sum_{i=1}^{n-1} \frac{4}{i^2}. \tag{2.6}$$

Note that although the mean total branch length diverges with n, the variance converges to a constant, $4\pi/6 \simeq 6.6$.

2.2.3 Models in population genetics

The Kingman coalescent is the limiting coalescent process (in the large population limit) for a surprisingly wide range of population models. Some important examples of models in Kingman's "domain of attraction" are introduced in this section. Common to all of these models are the following assumptions:

- The population has constant size N
- Reproduction happens in discrete generations
- The offspring distributions are identical at each generation, and independent between generations
- These models are all *neutral*, i.e. the offspring distribution is exchangeable.

As before section/eq ref?, we define offspring counts in terms of parental indices as $\nu_j := |\{i : a_i = j\}|$. Under the assumption of neutrality, it is sufficient to consider only the offspring counts, rather than the parental indices (which generally carry more information). From a biological perspective, neutrality encodes the absence of natural selection, i.e. no individual in the population is "fitter" than another.

Wright-Fisher model

The neutral Wright-Fisher model (Fisher 1923; Fisher 1930; Wright 1931) is one of the most studied models in population genetics. At each time step the existing generation dies and is replaced by N offspring. The offspring descend from parents $(a_1, \ldots, a_N)$ which are selected according to

$$a_i \overset{iid}{\sim} \mathrm{Categorical}(\{1, \ldots, N\}, (1/N, \ldots, 1/N)).$$

The joint distribution of the offspring counts is therefore

$$(v_1, \ldots, v_N) \sim \mathrm{Multinomial}(N, (1/N, \ldots, 1/N)).$$

Since the Multinomial distribution is exchangeable, this model is neutral. There are several non-neutral variants of the Wright-Fisher model citations?, but they are typically much less tractable than the neutral one.

Kingman showed in his original papers introducing the Kingman coalescent (Kingman 1982b) that, under a certain time-scaling which is what?, genealogies of the neutral Wright-Fisher model converge to the Kingman coalescent as $N \to \infty$.

Cannings model

The neutral Cannings model (Cannings 1974; Cannings 1975) is a more general construction which encompasses the neutral Wright-Fisher model as a special case.

In the Cannings model, the particular offspring distribution is not specified; we only require that it is exchangeable, i.i.d. between generations, and preserves the population size. In particular, the probability of observing offspring counts $(v_1, \dots, v_N)$ must be invariant under permutations of this vector.

Genealogies of the neutral Cannings model also converge to the Kingman coalescent, under some conditions and a suitable time-scaling which is what?, as $N \to \infty$ (see for example Etheridge 2011, Section 2.2). original reference for this?

Moran model

The neutral Moran model (Moran 1958), while perhaps less biologically relevant, is mathematically appealing because its simple dynamics make it particularly tractable.

At each time step, an ordered pair of individuals is selected uniformly at random. The first individual in this pair dies (i.e. leaves no offspring in the next generation), while the other reproduces (leaving two offspring). All of the other individuals leave exactly one offspring. This is another special case of the neutral Cannings model, where the offspring distribution is now uniform over all permutations of $(0, 2, 1, 1, \dots, 1)$. Therefore we know that under a suitable time-scaling which is what?, its genealogies converge to the Kingman coalescent. but it would be good to cite a Moran-specific convegrence result if that preceded the Cannings class. Emphasise the difference between Moran and WF time scales.

2.2.4 Particle populations

Much of the population genetics framework transfers readily to the case of SMC. The population is now a population of particles, with each iteration of the SMC algorithm corresponding to a generation, and resampling playing the part of reproduction. In fact, SMC "populations" are in some ways more suited to these population models than actual populations of organisms. The assumptions that the population has constant size N and that reproduction occurs only at discrete generations are satisfied by construction. However, we cannot assume independence between generations: as seen in Figure 2.1, the offspring counts at subsequent generations are not independent without some conditioning. In fact, after marginalising out the information about the positions of the particles, the genealogical process is not even Markovian. Nor is our model neutral: the resampling distribution depends on the weight of each particle (the weight plays the role of fitness in a non-neutral population model).

2.3 Sequential Monte Carlo genealogies

2.3.1 From particles to genealogies

How does the SMC algorithm induce a genealogy? (resampling = parent-child relationship).

2.3.2 Performance

How do genealogies affect performance? Variance (and variance estimation?), storage cost. Ancestral degeneracy.

2.3.3 Mitigating ancestral degeneracy

Low-variance resampling (save details for next section). Adaptive resampling: idea of balancing weight/ancestral degeneracy; rule of thumb for implementing it; when is it effective or not?; necessary changes to our generic SMC algorithm (calculation of weights in particular). Backward sampling: when is it possible to do this?

2.3.4 Asymptotics

Why are large population asymptotics useful? Existing results (path storage, KJJS).

2.4 Resampling

2.4.1 Definition

The job of resampling (map weights to counts). Define "valid" resampling schemes (the three rules). Counter-examples where these rules are violated (the examples I've mentioned in previous writings, plus optimal transport resampling [see email from James Thornton] and that one FC told me about recently [Huang et al 2020]).

2.4.2 What makes a good resampling scheme?

Low-variance: variance of what? Different criteria/ definitions of optimality. Negative association. Link back to adaptive resampling: interaction between adaptive and low-variance resampling.

2.4.3 Examples

Tour of the key resampling schemes (multinomial, residual-*, stratified, systematic, and the worst possible scheme). Comparison of properties of these, existing results comparing schemes. Implementation considerations. Theoretical justification (or lack of).

2.4.4 Stochastic rounding

Define stochastic rounding. Resampling schemes contained by this class. General properties for this class (marginal distributions, negative association, minimum-variance).

2.5 Conditional SMC

2.5.1 Particle MCMC

Motivate particle MCMC methods.

The idea behind particle MCMC methods is to use SMC steps within the MCMC updates in a way that improves the mixing properties of the Markov chain. In certain models, generally those including some highly correlated sequential components, this strategy can be very effective.

The following scenario illustrates the power of particle MCMC, and is a good model to have in mind as we go on to discuss particle Gibbs and ancestor sampling. Include the model from the start of my ancestor sampling note. Emphasise that the inference itself is not sequential; we are targeting one static posterior distribution, on a fixed time horizon.

2.5.2 Particle Gibbs algorithm

Present particle Gibbs algorithm (for the specific model just introduced?, but note that of course the algorithm is more general). Explain why CSMC is required within particle Gibbs.

2.5.3 Ancestor sampling

Algorithm (or required changes to generic algorithm). Relation to backward sampling. When can it be implemented? Effect on performance (when is it effective?). Maybe illustrate/motivate with some plots as in the ancestor sampling note.

3 Limits

3.1 Encoding genealogies

3.1.1 The genealogical process

Encoding as process on space of partitions $\mathcal{P}_n$. Argue that this encodes everything we need. Initial and absorbing states. Intuit with diagram(s), explain relationship between partition blocks and genealogical tree.

3.1.2 Time scale

Introduce c_N, τ_N, D_N. Contrast to pop gen literature, e.g. our c_N/time scale is random. Properties of these quantities: $c_N, D_N \in [0,1]$, and $D_N \leq c_N$ and $\sum_{r=1}^{\tau_N(t)} c_N \in [t, t+1]$ (or rather the version of that with general start time).

In order to get a continuous limit, we scale time by a function $\tau_N(\cdot)$. In the population genetics literature, a deterministic time scale can be used [citations] and/or this will have been mentioned already in pop gen example models (Section 2.2.3), whereas in our case τ_N depends on the offspring counts and is therefore random. To define the time scale we first define the pair merger rate

$$c_N(t) := \frac{1}{(N)_2} \sum_{i=1}^{N} (\nu_t^{(i)})_2. \tag{3.1}$$

This is the probability, conditional on $\nu_t^{(1:N)}$, that a randomly chosen pair of lineages in generation t merges exactly one generation back. To achieve the limiting pair merger rate of 1, as in the n-coalescent, we rescale time by the generalised inverse

$$\tau_N(t) := \inf \left\{ s \geq 1 : \sum_{r=1}^{s} c_N(r) \geq t \right\}. \tag{3.2}$$

The function τ_N maps continuous to discrete time, providing the link between the discrete-time SMC dynamics and the continuous-time Kingman limit. We will also need the following quantity, which is an upper bound on the rate of multiple mergers (three or more lineages merging, or two or more simultaneous pairwise mergers):

$$D_N(t) := \frac{1}{N(N)_2} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ \nu_t^{(i)} + \frac{1}{N} \sum_{j \neq i} (\nu_t^{(j)})^2 \right\}. \tag{3.3}$$

Some basic properties are given in Proposition 3.1.

Proposition 3.1 (Properties of c_N)**.** *For all $t \in \mathbb{N}$, $t' > s' > 0$,*

(a) $c_N(t), D_N(t) \in [0,1]$

(b) $D_N(t) \leq c_N(t)$

(c) $c_N(t)^2 \leq c_N(t)$

(d) $$t' \leq \sum_{r=1}^{\tau_N(t')} c_N(r) \leq t' + 1.$$

(e) $$t' - s' - 1 \leq \sum_{r=\tau_N(s')+1}^{\tau_N(t')} c_N(r) \leq t' - s' + 1.$$

Proof. **(a)** $c_N(t)$ and $D_N(t)$ are clearly non-negative. Both are maximised when one of the offspring counts is equal to N and the rest are zero, in which case $c_N(t) = D_N(t) = 1$.
(b) As outlined in Koskela et al. (2018, p.9),

$$\begin{aligned}
D_N(t) &:= \frac{1}{(N)_2} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \frac{1}{N} \left\{ \nu_t^{(i)} + \frac{1}{N} \sum_{j \neq i}^{N} (\nu_t^{(j)})^2 \right\} \\
&\leq \frac{1}{(N)_2} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \frac{1}{N} \left\{ \nu_t^{(i)} + \frac{1}{N} \sum_{j \neq i}^{N} N \nu_t^{(j)} \right\} \\
&= \frac{1}{(N)_2} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \frac{1}{N} \left\{ \sum_{j=1}^{N} \nu_t^{(j)} \right\} \leq \frac{1}{(N)_2} \sum_{i=1}^{N} (\nu_t^{(i)})_2 = c_N(t).
\end{aligned}$$

(c) is immediate given (a).
(d) follows directly from the definition of τ_N in (3.2).
(e) Writing

$$\sum_{r=\tau_N(s')+1}^{\tau_N(t')} c_N(r) = \sum_{r=1}^{\tau_N(t')} c_N(r) - \sum_{r=1}^{\tau_N(s')} c_N(r),$$

the result follows by applying (d) to both sums. ■

3.1.3 Transition probabilities

Introduce $p_{\xi\eta}$. Present expression for that (or at least for $p_{\xi\xi}$), and hence the bounds on it that will be used later (keeping big-O terms explicit where possible).

Let $\mathcal{P}_n$ be the space of partitions of $\{1, \ldots, n\}$, and denote by Δ the partition of singletons $\{\{1\}, \ldots, \{n\}\}$. For any $\xi, \eta \in \mathcal{P}_n$ and $t \in \mathbb{N}$, let $p_{\xi\eta}(t)$ denote the conditional transition probabilities of the genealogical process given $\nu_t^{(1:N)}$ ($t \in \mathbb{N}$, $\xi, \eta \in \mathcal{P}_n$). The transition probability $p_{\xi\eta}(t)$ can only be non-zero when η can be obtained from ξ by merging some blocks of ξ. Ordering the blocks by their least element, denote by b_i the number of blocks of ξ that merge to form block i in η ($i \in \{1, \dots, |\eta|\}$). Hence $b_1 + \cdots + b_{|\eta|} = |\xi|$. Then the transition probability is given by

$$p_{\xi\eta}(t) := \frac{1}{(N)_{|\xi|}} \sum_{\substack{i_1 \neq \cdots \neq i_{|\eta|} \\ =1}}^{N} (\nu_t^{(i_1)})_{b_1} \cdots (\nu_t^{(i_{|\eta|})})_{b_{|\eta|}}. \tag{3.4}$$

We will only need to work directly with the *identity* transition probabilities $p_{\xi\xi}(t)$. Upper and lower bounds on these probabilities are presented in Propositions 3.2 and 3.3.

Proposition 3.2 (Lower bound on identity transition probabilities)**.** *Let $\xi \in \mathcal{P}_n$, $N > 2$. Then*

$$p_{\xi\xi}(t) \geq 1 - \binom{|\xi|}{2} \frac{N^{n-2}}{(N-2)_{n-2}} \left[c_N(t) + B_{|\xi|} D_N(t)\right]$$

where $B_{|\xi|} = K(|\xi|-1)!(|\xi|-2)\exp(2\sqrt{2(|\xi|-2)})$ for some $K > 0$ that does not depend on $|\xi|$.

For weak convergence proof, refer to this proposition but rewrite the inequality using $\xi = \Delta$ and α_n, to provide a local target for cross-referencing. Similarly for UB.

Proof. We have the following expression for $p_{\xi\xi}(t)$, by subtracting all possible non-identity transitions (the omitted $k = |\xi|$ term would count identity transitions):

$$p_{\xi\xi}(t) = 1 - \frac{1}{(N)_{|\xi|}} \sum_{k=1}^{|\xi|-1} \sum_{\substack{b_1 \geq \ldots \geq b_k = 1 \\ b_1 + \ldots + b_k = |\xi|}}^{|\xi|} \frac{|\xi|!}{\prod_{j=1}^{|\xi|} (j!)^{\kappa_j} \kappa_j!} \sum_{\substack{i_1 \neq \ldots \neq i_k = 1 \\ \text{all distinct}}}^{N} (\nu_t^{(i_1)})_{b_1} \ldots (\nu_t^{(i_k)})_{b_k},$$

where $\kappa_i = |\{j : b_j = i\}|$ is the multiplicity of mergers of size i (κ_1 counts non-merger events, and we have the identity $\kappa_1 + 2\kappa_2 + \cdots + |\xi|\kappa_{|\xi|} = |\xi|$). The combinatorial factor is the number of partitions of a sequence of length $|\xi|$ having κ_j subsequences of length j for each j (Fu 2006, Equation (11)).

We separate the $k = |\xi|-1$ term (which counts single pair mergers), for which $(b_1, b_2, \ldots, b_{|\xi|-1}) = (2, 1, \ldots, 1)$ and

$$\frac{|\xi|!}{\prod_{j=1}^{|\xi|} (j!)^{\kappa_j} \kappa_j!} = \binom{|\xi|}{2}.$$

For the remaining terms we use

$$\frac{|\xi|!}{\prod_{j=1}^{|\xi|} (j!)^{\kappa_j} \kappa_j!} \leq |\xi|!.$$

Thus

$$p_{\xi\xi}(t) \geq 1 - \frac{1}{(N)_{|\xi|}} \binom{|\xi|}{2} \sum_{\substack{i_1 \neq \ldots \neq i_{|\xi|-1}=1 \\ \text{all distinct}}}^{N} (\nu_t^{(i_1)})_2 \nu_t^{(i_2)} \ldots \nu_t^{(i_{|\xi|-1})}$$
$$- \frac{1}{(N)_{|\xi|}} \sum_{k=1}^{|\xi|-1} \sum_{\substack{b_1 \geq \ldots \geq b_k = 1 \\ b_1 + \ldots + b_k = |\xi|}}^{|\xi|} |\xi|! \sum_{\substack{i_1 \neq \ldots \neq i_k = 1 \\ \text{all distinct}}}^{N} (\nu_t^{(i_1)})_{b_1} \ldots (\nu_t^{(i_k)})_{b_k}$$

Now, for the $k = |\xi| - 1$ term we use the bound

$$\sum_{i_1 \neq \ldots \neq i_{|\xi|-1}=1}^{N} (\nu_t^{(i_1)})_2 \nu_t^{(i_2)} \ldots \nu_t^{(i_{|\xi|-1})} \leq N^{|\xi|-2} \sum_{i=1}^{N} (\nu_t^{(i)})_2$$

while for the other terms we have (similarly to Koskela et al. 2018, Lemma 1 Case 3)

$$\sum_{\substack{i_1 \neq \ldots \neq i_k = 1 \\ \text{all distinct}}}^{N} (\nu_t^{(i_1)})_{b_1} \ldots (\nu_t^{(i_k)})_{b_k} \leq \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left(N^{|\xi|-2} - \sum_{\substack{j_1 \neq \ldots \neq j_{|\xi|-2} = 1 \\ \text{all distinct and } \neq i}}^{N} \nu_t^{(j_1)} \ldots \nu_t^{(j_{|\xi|-2})} \right)$$
$$\leq \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ N^{|\xi|-2} - (N - \nu_t^{(i)})^{|\xi|-2} + \binom{|\xi|-2}{2} \sum_{j \neq i} (\nu_t^{(j)})^2 \left(\sum_{k \neq i} \nu_t^{(k)} \right)^{|\xi|-4} \right\}$$
$$\leq \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ (|\xi| - 2) \nu_t^{(i)} N^{|\xi|-3} + \binom{|\xi|-2}{2} \sum_{j \neq i} (\nu_t^{(j)})^2 N^{|\xi|-4} \right\},$$

where the last step uses $(N - x)^b \geq N^b - bxN^{b-1}$ for $x \leq N$, $b \geq 0$. Hence

$$p_{\xi\xi}(t) \geq 1 - \frac{1}{(N)_{|\xi|}} \binom{|\xi|}{2} N^{|\xi|-2} \sum_{i=1}^{N} (\nu_t^{(i)})_2$$
$$- \frac{N^{|\xi|-3}}{(N)_{|\xi|}} |\xi|! \sum_{k=1}^{|\xi|-1} \sum_{\substack{b_1 \geq \ldots \geq b_k = 1 \\ b_1 + \ldots + b_k = |\xi|}}^{|\xi|} \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ (|\xi| - 2) \nu_t^{(i)} + \binom{|\xi|-2}{2} \frac{1}{N} \sum_{j \neq i} (\nu_t^{(j)})^2 \right\}.$$

The summands in the last line are independent of k, b_i, and the number of terms in the sums over k and $b_1, \ldots, b_k$ is bounded by $\gamma_{|\xi|-2}(|\xi| - 2)$, where γ_n is the number of integer partitions of n. By Hardy and Ramanujan (1918, Section 2), $\gamma_n < Ke^{2\sqrt{2n}}/n$ for a constant

$K > 0$ independent of n. Thus, for $|\xi| > 2$,

$$
\begin{aligned}
p_{\xi\xi}(t) \geq 1 &- \frac{N^{|\xi|-2}}{(N-2)_{|\xi|-2}} \binom{|\xi|}{2} c_N(t) \\
&- \frac{N^{|\xi|-2}}{(N-2)_{|\xi|-2}} K \exp(2\sqrt{2(|\xi|-2)})|\xi|! \frac{1}{N(N)_2} \\
&\qquad \sum_{i=1}^{N} (\nu_t^{(i)})_2 \left\{ (|\xi|-2)\nu_t^{(i)} + \binom{|\xi|-2}{2} \frac{1}{N} \sum_{j \neq i} (\nu_t^{(j)})^2 \right\} \\
\geq 1 &- \frac{N^{|\xi|-2}}{(N-2)_{|\xi|-2}} \binom{|\xi|}{2} c_N(t) \\
&- \frac{N^{|\xi|-2}}{(N-2)_{|\xi|-2}} K \exp(2\sqrt{2(|\xi|-2)})|\xi|! \binom{|\xi|-1}{2} D_N(t) \\
\geq 1 &- \frac{N^{|\xi|-2}}{(N-2)_{|\xi|-2}} \binom{|\xi|}{2} \left[c_N(t) + B_{|\xi|} D_N(t) \right]
\end{aligned}
$$

where

$$
\begin{aligned}
B_{|\xi|} &= \binom{|\xi|}{2}^{-1} K \exp(2\sqrt{2(|\xi|-2)})|\xi|! \binom{|\xi|-1}{2} \\
&= K(|\xi|-1)!(|\xi|-2)\exp(2\sqrt{2(|\xi|-2)}).
\end{aligned}
$$

When $|\xi| \leq 2$, there are no terms with $k \leq |\xi| - 2$, and the result is immediate. ■

Proposition 3.3 (Upper bound on identity transition probabilities)**.** *Let $\xi \in \mathcal{P}_n$, $N >$ ... it came out as $N \geq n^2(n-3)+1$ for the $\xi = \Delta$ case, up to possible errors. Then*

$$
p_{\xi\xi}(t) \leq 1 - \binom{|\xi|}{2} \frac{N^{n-2}}{(N-2)_{n-2}} \left[c_N(t) - B'_{|\xi|} D_N(t) \right]
$$

where $B'_{|\xi|} = \ldots$.

Proof. The proof follows Koskela et al. (2018, Proof of Lemma 1 Case 1) but with the terms in N kept explicit. (where possible/only some of them?)

... ■

3.2 An existing limit theorem

State KJJS theorem. Discuss the conditions in detail. Give outline of proof.

3.3 A new limit theorem

State our limit theorem. Give intuition for the new condition. Compare to KJJS: why our conditions might be considered "weaker" (Moran model example, and whatever else we said to our referee/in the BJJK article); our condition is easier to check (as demonstrated in later corollaries).

Theorem 3.1. *Let $\nu_t^{(1:N)}$ denote the offspring numbers in an IPS satisfying the standing assumption and such that, for any N sufficiently large, $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t. Suppose that there exists a deterministic sequence $(b_N)_{N\geq 1}$ such that $\lim_{N\to\infty} b_N = 0$ and*

$$\frac{1}{(N)_3}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_3] \leq b_N \frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_2] \tag{3.5}$$

for all N, uniformly in $t \geq 1$. Then the rescaled genealogical process $(G^{(n,N)}_{\tau_N(t)})_{t\geq 0}$ converges in the sense of finite-dimensional distributions to Kingman's n-coalescent as $N \to \infty$.

3.3.1 Proof of theorem

Proof that KJJS conditions are implied by ours. Modification of KJJS proof (or even write out a complete proof?) using weaker bound on $p_{\xi\xi}$ (that bound should have been stated and proved already in transition probabilities section).

4 Applications

4.1 Multinomial resampling

This is the easy-to-analyse scheme, because conditionally i.i.d., and was presented in KJJS already. Now (with our simpler conditions) it is easier to show.

4.1.1 Proof of main condition

4.1.2 Proof of finite time scale condition

4.2 Stochastic rounding

4.2.1 Proof of main condition

Corollary 4.1. *Consider an SMC algorithm using any stochastic rounding as its resampling scheme, such that the standing assumption is satisfied. Assume that there exists a constant $a \in [1, \infty)$ such that for all x, x', t,*

$$\frac{1}{a} \leq g_t(x, x') \leq a. \tag{4.1}$$

Assume that $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t. Let $(G_t^{(n,N)})_{t\geq 0}$ denote the genealogy of a random sample of n terminal particles from the output of the algorithm when the total number of particles used is N. Then, for any fixed n, the time-scaled genealogy $(G_{\tau_N(t)}^{(n,N)})_{t\geq 0}$ converges to Kingman's n-coalescent as $N \to \infty$, in the sense of finite-dimensional distributions.

Proof. Using the forward-time Markov property of SMC, and the associated conditional dependence graph, for each N we establish a sequence of σ-algebras

$$\mathcal{H}_t := \sigma(X_{t-1}^{(1:N)}, X_t^{(1:N)}, w_{t-1}^{(1:N)}, w_t^{(1:N)}) \tag{4.2}$$

such that $\nu_t^{(1:N)}$ is conditionally independent of the filtration $\mathcal{F}_{t-1}$ given $\mathcal{H}_t$. The full D-separation argument is presented in Appendix **??**.

Defining the family sizes $\nu_t^{(i)} = |\{j : a_t^{(j)} = i\}|$ as functions of $a_t^{(1:N)}$, we have the almost sure constraint $\nu_t^{(i)} \in \{\lfloor N w_t^{(i)} \rfloor, \lfloor N w_t^{(i)} \rfloor + 1\}$. Denote $p_0^{(i)} := \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor \mid \mathcal{H}_t]$ and $p_1^{(i)} := \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t] = 1 - p_0^{(i)}$.

We obtain the following upper bounds, using the almost sure bounds $w_t^{(i)} \leq a^2/N$ which follow from (4.1) along with the form of the weights in Algorithm 1:

$$\begin{aligned}
\mathbb{E}[(\nu_t^{(i)})_3 \mid \mathcal{H}_t] &= p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor)_3 + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor + 1)_3 \\
&= \lfloor N w_t^{(i)} \rfloor(\lfloor N w_t^{(i)} \rfloor - 1)\{p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor - 2) + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor + 1)\} \\
&= \lfloor N w_t^{(i)} \rfloor(\lfloor N w_t^{(i)} \rfloor - 1)\{\lfloor N w_t^{(i)} \rfloor(p_0^{(i)} + p_1^{(i)}) - 2p_0^{(i)} + p_1^{(i)}\} \\
&= \lfloor N w_t^{(i)} \rfloor(\lfloor N w_t^{(i)} \rfloor - 1)\{\lfloor N w_t^{(i)} \rfloor - 2p_0^{(i)} + p_1^{(i)}\} \\
&\leq a^2(a^2 - 1)(a^2 - 0 + 1)\mathbb{1}_{\lfloor N w_t^{(i)} \rfloor \geq 2} \\
&\leq (a^2 + 1)^3 \mathbb{1}_{\lfloor N w_t^{(i)} \rfloor \geq 2}.
\end{aligned}$$

We also have the lower bounds

$$\begin{aligned}
\mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] &= p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor)_2 + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor + 1)_2 \\
&= \lfloor N w_t^{(i)} \rfloor\{p_0^{(i)}(\lfloor N w_t^{(i)} \rfloor - 1) + p_1^{(i)}(\lfloor N w_t^{(i)} \rfloor + 1)\} \\
&= \lfloor N w_t^{(i)} \rfloor\{\lfloor N w_t^{(i)} \rfloor(p_0^{(i)} + p_1^{(i)}) - p_0^{(i)} + p_1^{(i)}\} \\
&= \lfloor N w_t^{(i)} \rfloor\{\lfloor N w_t^{(i)} \rfloor - p_0^{(i)} + p_1^{(i)}\} \\
&\geq 2(2 - 1 + 0)\mathbb{1}_{\lfloor N w_t^{(i)} \rfloor \geq 2} = 2\mathbb{1}_{\lfloor N w_t^{(i)} \rfloor \geq 2}.
\end{aligned}$$

Applying the tower property and conditional independence,

$$\begin{aligned}
\frac{1}{(N)_2}\sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_2] &= \frac{1}{(N)_2}\mathbb{E}_t\left[\sum_{i=1}^{N} \mathbb{E}\left[(\nu_t^{(i)})_2 \mid \mathcal{H}_t, \mathcal{F}_{t-1}\right]\right] \\
&= \frac{1}{(N)_2}\mathbb{E}_t\left[\sum_{i=1}^{N} \mathbb{E}\left[(\nu_t^{(i)})_2 \mid \mathcal{H}_t\right]\right] \geq \frac{1}{(N)_2}2\mathbb{E}_t\left[|\{i : \lfloor N w_t^{(i)} \rfloor \geq 2\}|\right]
\end{aligned}$$

and similarly

$$\begin{aligned}
\frac{1}{(N)_3}\sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_3] &\leq \frac{1}{(N)_3}(a^2+1)^3\mathbb{E}_t\left[|\{i : \lfloor N w_t^{(i)} \rfloor \geq 2\}|\right] \\
&\leq b_N \frac{1}{(N)_2}\sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_2]
\end{aligned}$$

where

$$b_N := \frac{1}{N-2}\frac{(a^2+1)^3}{2} \underset{N\to\infty}{\longrightarrow} 0$$

is independent of $\mathcal{F}_\infty$, satisfying (3.5). The result follows by applying Theorem ??. ■

4.2.2 Finite time scale

Lemma 4.1. *Consider an SMC algorithm using any stochastic rounding as its resampling scheme. Suppose that $\varepsilon \leq q_t(x, x') \leq \varepsilon^{-1}$ uniformly for some $\varepsilon \in (0, 1]$, and that there exist $\zeta > 0$ and $\delta \in (0, 1)$ such that $\mathbb{P}[\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N \mid \mathcal{F}_{t-1}] \geq \zeta$ for infinitely many t. Then, for all $N > 1$, $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t.*

Proof. Let $\mathcal{H}_t$ be defined as in (4.2). The first step is to show that whenever $\max_i w_t^{(i)} \geq (1+\delta)/N$, $\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] = \mathbb{P}[c_N(t) \neq 0 \mid \mathcal{H}_t]$ is bounded below uniformly in t. For this purpose we need consider only weight vectors such that $w_t^{(i)} \in (0, 2/N)$ for all i; otherwise $\mathbb{P}[c_N(t) \neq 0 \mid \mathcal{H}_t] = 1$ by the definition of stochastic rounding.

Denote $\mathcal{S}_{N-1}^{\delta} = \{w^{(1:N)} \in \mathcal{S}_{N-1} : \forall i,\, 0 < w^{(i)} < 2/N; \max_i w^{(i)} \geq (1+\delta)/N\}$ for any $\delta \in (0,1)$, where $\mathcal{S}_k$ denotes the k-dimensional probability simplex. Fix arbitrary $w_t^{(1:N)} \in \mathcal{S}_{N-1}^{\delta}$. Set $i^\star = \arg\max_i w_t^{(i)}$ and denote $\mathcal{I} = \{i \in \{1, \dots, N\} : w^{(i)} > 1/N\}$. Since all weights are in $(0, 2/N)$, for $i \in \mathcal{I}, \nu_t^{(i)} \in \{1, 2\}$ and for $i \notin \mathcal{I}, \nu_t^{(i)} \in \{0, 1\}$; and since the offspring counts must sum to N, we can write

$$\begin{aligned}
\mathbb{P}[c_N(t) \leq 2/N^2 \mid \mathcal{H}_t] &= \mathbb{P}[\nu_t^{(i)} = 1\, \forall i \in \{1, \dots, N\} \mid \mathcal{H}_t] \\
&= \mathbb{P}[\nu_t^{(i)} = 1\, \forall i \in \mathcal{I} \mid \mathcal{H}_t] \\
&= \prod_{i \in \mathcal{I}} \mathbb{P}[\nu_t^{(i)} = 1 \mid \nu_t^{(j)} = 1\, \forall j \in \mathcal{I} : j < i; \mathcal{H}_t] \\
&= \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t] \prod_{\substack{i \in \mathcal{I} \\ i \neq i^\star}} \mathbb{P}[\nu_t^{(i)} = 1 \mid \nu_t^{(i^\star)} = 1; \nu_t^{(j)} = 1\, \forall j \in \mathcal{I} : j < i; \mathcal{H}_t] \\
&\leq \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t].
\end{aligned} \tag{4.3}$$

The final inequality holds with equality when $|\mathcal{I}| = 1$, i.e. the only weight larger than $1/N$ is $w_t^{(i^\star)}$. Thus $\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t]$ is minimised on $\mathcal{S}_{N-1}^{\delta}$ when only one weight is larger than $1/N$, in which case the values of the other weights do not affect this probability.

Define $w_{\delta'} = \{(1, \dots, 1) + \delta' e_{i^\star} - \delta' e_{j^\star}\}/N$ for fixed $i^\star \neq j^\star$ and $\delta' \in (0, 1)$, where e_i denotes the ith canonical basis vector in $\mathbb{R}^N$. As in the proof of Corollary 4.1, define $p_0^{(i)} = \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor \mid \mathcal{H}_t]$ and $p_1^{(i)} = \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + 1 \mid \mathcal{H}_t]$. Then from (4.3) we have

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t, w_t^{(1:N)} = w_{\delta'}] = 1 - \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t, w_t^{(1:N)} = w_{\delta'}] = p_1^{(i^\star)},$$

evaluated on $w_{\delta'}$. We will need a lower bound on $p_1^{(i^\star)}$ when $w_t^{(1:N)} = w_{\delta'}$. We first derive expressions for $p_0^{(i)}$ and $p_1^{(i)}$ up to a constant, then use $p_0^{(i)} + p_1^{(i)} = 1$ to get a normalised bound. We have

$$
\begin{aligned}
p_0^{(i)} &= C(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor) \\
&\quad \times \sum_{\substack{a_{1:N} \in \{1,\dots,N\}^N: \\ |\{j: a_j = i\}| = \lfloor Nw_t^{(i)} \rfloor}} \mathbb{P}\left[a_t^{(1:N)} = a_{1:N} \mid \nu_t^{(i)}, w_t^{(1:N)}\right] \prod_{k=1}^N q_{t-1}(X_t^{(a_k)}, X_{t-1}^{(k)}), \\
p_1^{(i)} &= C(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor) \\
&\quad \times \sum_{\substack{a_{1:N} \in \{1,\dots,N\}^N: \\ |\{j: a_j = i\}| = \lfloor Nw_t^{(i)} \rfloor + 1}} \mathbb{P}\left[a_t^{(1:N)} = a_{1:N} \mid \nu_t^{(i)}, w_t^{(1:N)}\right] \prod_{k=1}^N q_{t-1}(X_t^{(a_k)}, X_{t-1}^{(k)}).
\end{aligned}
$$

Applying the bounds on q_t, we have

$$
\begin{gathered}
C(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^N \leq p_0^{(i)} \leq C(1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-N}, \\
C(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^N \leq p_1^{(i)} \leq C(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-N}
\end{gathered}
$$

from which we construct the normalised bound

$$
p_1^{(i)} \geq \frac{(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^N}{(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-N} + (1 - Nw_t^{(i)} + \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{-N}} = (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\varepsilon^{2N}.
$$

When $w_t^{(1:N)} = w_{\delta'}$, we have $w_t^{(i^\star)} = (1+\delta')/N$, so $p_1^{(i^\star)} \geq \delta' \varepsilon^{2N}$, which is increasing in δ'. We conclude that $\mathbb{P}[c_N(t) > 2/N^2 | \mathcal{H}_t, \max_i w_t^{(i)} \geq (1+\delta)/N] \geq \min_{\delta' \geq \delta} \delta' \varepsilon^{2N} = \delta \varepsilon^{2N}$.

A slight modification of this argument yields $\mathbb{P}[c_N(t) > 2/N^2 | \mathcal{H}_t, \min_i w_t^{(i)} \leq (1-\delta)/N] \geq \delta \varepsilon^{2N}$. Whenever $\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N$, either $\max_i w_t^{(i)} \geq (1+\delta)/N$ or $\min_i w_t^{(i)} \leq (1-\delta)/N$, so we have $\mathbb{P}[c_N(t) > 2/N^2 | \mathcal{H}_t, \max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N] \geq \delta \varepsilon^{2N}$. Thus

$$
\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \delta \varepsilon^{2N} \mathbb{1}_{\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N}.
$$

Using the D-separation established in Appendix **??** combined with the tower property, we have

$$
\begin{aligned}
\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{F}_{t-1}] = \mathbb{E}_t\left[\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t, \mathcal{F}_{t-1}]\right] &= \mathbb{E}_t\left[\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t]\right] \\
&\geq \delta \varepsilon^{2N} \mathbb{P}[\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N \mid \mathcal{F}_{t-1}],
\end{aligned}
$$

which is bounded below by $\zeta \delta \varepsilon^{2N}$ for infinitely many t. Hence,

$$
\sum_{t=0}^{\infty} \mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{F}_{t-1}] = \infty.
$$

By a filtered version of the second Borel–Cantelli lemma (see for example Rick Durrett 2019, Theorem 4.3.4), this implies that $c_N(t) > 2/N^2$ for infinitely many t, almost surely.

This ensures, for all $t < \infty$, that $\mathbb{P}\left[\exists s < \infty : \sum_{r=1}^{s} c_N(r) \geq t\right] = 1$, which by definition of $\tau_N(t)$ is equivalent to $\mathbb{P}[\tau_N(t) = \infty] = 0$. ■

4.3 Stratified resampling

Proof for this one is in progress, but shouldn't be too difficult. Like SRs, there are only finitely many possible counts conditional on weights, so the same kind of proof will work (but with four cases instead of two).

4.4 The worst possible resampling scheme

Remark that this one doesn't converge to KC, but rather to a star-shaped coalescent.

4.5 Conditional SMC

Why CSMC is qualitatively different to, say, standard SMC with multinomial resampling (immortal particle etc.). Reasons for restriction to multinomial resampling, conjecture that limit theorem holds for other schemes in CSMC.

4.5.1 Proof of main condition

Corollary 4.2. *Consider a conditional SMC algorithm using multinomial resampling, such that the standing assumption is satisfied. Assume there exist constants $\varepsilon \in (0,1], a \in [1,\infty)$ and probability density h such that for all x, x', t,*

$$\frac{1}{a} \leq g_t(x,x') \leq a, \quad \varepsilon h(x') \leq q_t(x,x') \leq \frac{1}{\varepsilon} h(x'). \tag{4.4}$$

Let $(G_t^{(n,N)})_{t\geq 0}$ denote the genealogy of a random sample of n terminal particles from the output of the algorithm when the total number of particles used is N. Then, for any fixed n, the time-scaled genealogy $(G_{\tau_N(t)}^{(n,N)})_{t\geq 0}$ converges to Kingman's n-coalescent as $N \to \infty$, in the sense of finite-dimensional distributions.

Proof. Define the conditioning σ-algebra $\mathcal{H}_t$ as in (4.2). We assume without loss of generality that the immortal particle takes index 1 in each generation. This significantly simplifies the notation, but the same argument holds if the immortal indices are taken to be $a^\star_{(0:T)}$ rather than $(1, \dots, 1)$.

The parental indices are conditionally independent, as in standard SMC with multinomial resampling, but we have to treat $i = 1$ as a special case. We have the following conditional law on parental indices

$$\mathbb{P}\left[a_t^{(i)} = a_i \mid \mathcal{H}_t\right] \propto \begin{cases} \mathbb{1}_{a_i=1} & i = 1 \\ w_t^{(a_i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}) & i = 2, \dots, N. \end{cases}$$

The joint conditional law is therefore

$$\mathbb{P}\left[a_t^{(1:N)} = a_{1:N} \mid \mathcal{H}_t\right] \propto \mathbb{1}_{a_1=1} \prod_{i=2}^{N} w_t^{(a_i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}).$$

First we make the following observation, which follows from a balls-in-bins coupling. Assume (4.4). Then for any function $f : \{1, \dots, N\}^N \to \mathbb{R}$ such that (for a fixed i) $f(a_t'^{(1:N)}) \geq f(a_t^{(1:N)})$ whenever $|\{j : a_t'^{(j)} = i\}| \geq |\{j : a_t^{(j)} = i\}|$,

$$\mathbb{E}[f(A_{1,i}^{(1:N)})] \leq \mathbb{E}[f(a_t^{(1:N)}) \mid \mathcal{H}_t] \leq \mathbb{E}[f(A_{2,i}^{(1:N)})] \tag{4.5}$$

where the elements of $A_{1,i}^{(1:N)}, A_{2,i}^{(1:N)}$ are all mutually independent and independent of $\mathcal{F}_\infty$, and distributed according to

$$A_{1,i}^{(j)} \sim \begin{cases} \delta_1 & j = 1 \\ \text{Categorical}\left((\varepsilon/a)^{\mathbb{1}_{i=1} - \mathbb{1}_{i\neq 1}}, \dots, (\varepsilon/a)^{\mathbb{1}_{i=N} - \mathbb{1}_{i\neq N}}\right) & j \neq 1 \end{cases}$$
$$A_{2,i}^{(j)} \sim \begin{cases} \delta_1 & j = 1 \\ \text{Categorical}\left((a/\varepsilon)^{\mathbb{1}_{i=1} - \mathbb{1}_{i\neq 1}}, \dots, (a/\varepsilon)^{\mathbb{1}_{i=N} - \mathbb{1}_{i\neq N}}\right) & j \neq 1 \end{cases}$$

where the vector of probabilities is given up to a constant in the argument of Categorical distributions. We use these random vectors to construct bounds that are independent of $\mathcal{F}_\infty$. Also define the corresponding offspring counts $V_1^{(i)} = |\{j : A_{1,i}^{(j)} = i\}|$, $V_2^{(i)} = |\{j : A_{2,i}^{(j)} = i\}|$, for $i = 1, \dots, N$, which have marginal distributions

$$V_1^{(i)} \overset{d}{=} \mathbb{1}_{i=1} + \text{Binomial}\left(N-1, \frac{\varepsilon/a}{(\varepsilon/a) + (N-1)(a/\varepsilon)}\right),$$
$$V_2^{(i)} \overset{d}{=} \mathbb{1}_{i=1} + \text{Binomial}\left(N-1, \frac{a/\varepsilon}{(a/\varepsilon) + (N-1)(\varepsilon/a)}\right).$$

Now consider the function $f_i(a_t^{(1:N)}) := (\nu_t^{(i)})_2$. We can apply (4.5) to obtain the lower bound

$$\begin{aligned}
\frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] &\geq \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}[(V_1^{(i)})_2] = \frac{1}{(N)_2} \left[\mathbb{E}[(V_1^{(1)})_2] + \sum_{i=2}^{N} \mathbb{E}[(V_1^{(i)})_2]\right] \\
&= \frac{1}{(N)_2} \left[\frac{(N-1)_2(\varepsilon/a)^2}{\{(\varepsilon/a) + (N-1)(a/\varepsilon)\}^2} + \frac{2(N-1)(\varepsilon/a)}{(\varepsilon/a) + (N-1)(a/\varepsilon)}\right. \\
&\qquad \left. + \sum_{i=2}^{N} \frac{(N-1)_2(\varepsilon/a)^2}{\{(\varepsilon/a) + (N-1)(a/\varepsilon)\}^2}\right] \\
&= \frac{1}{(N)_2} \left[\frac{2(N-1)(\varepsilon/a)}{(\varepsilon/a) + (N-1)(a/\varepsilon)} + \sum_{i=1}^{N} \frac{(N-1)_2(\varepsilon/a)^2}{\{(\varepsilon/a) + (N-1)(a/\varepsilon)\}^2}\right]
\end{aligned}$$

using the moments of the Binomial distribution (see Mosimann 1962 for example) along with the identity $(X+1)_2 \equiv 2(X)_1 + (X)_2$. This is further bounded by

$$\begin{aligned}\frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] &\geq \frac{1}{(N)_2}\left\{\frac{2(N-1)(\varepsilon/a)}{N(a/\varepsilon)} + \frac{(N)_3(\varepsilon/a)^2}{N^2(a/\varepsilon)^2}\right\} \\ &= \frac{1}{N^2}\left\{\frac{2\varepsilon^2}{a^2} + \frac{(N-2)\varepsilon^4}{a^4}\right\}. \end{aligned} \tag{4.6}$$

Similarly, we derive an upper bound on $f_i(a_t^{(1:N)}) := (\nu_t^{(i)})_3$, this time using the identity $(X+1)_3 \equiv 3(X)_2 + (X)_3$:

$$\begin{aligned}\frac{1}{(N)_3}\sum_{i=1}^{N}\mathbb{E}[(\nu_t^{(i)})_3 \mid \mathcal{H}_t] &\leq \frac{1}{(N)_3}\left[\mathbb{E}[(V_2^{(1)})_3] + \sum_{i=2}^{N}\mathbb{E}[(V_2^{(i)})_3]\right] \\ &\leq \frac{1}{(N)_3}\left[\frac{3(N-1)_2(a/\varepsilon)^2}{\{(a/\varepsilon)+(N-1)(\varepsilon/a)\}^2} + \sum_{i=1}^{N}\frac{(N-1)_3(a/\varepsilon)^3}{\{(a/\varepsilon)+(N-1)(\varepsilon/a)\}^3}\right] \\ &\leq \frac{1}{(N)_3}\left\{\frac{3(N-1)_2(a/\varepsilon)^2}{N^2(\varepsilon/a)^2} + \frac{(N)_4(a/\varepsilon)^3}{N^3(\varepsilon/a)^3}\right\} \\ &= \frac{1}{(N)_3}\left\{\frac{3(N-1)_2}{N^2}\frac{a^4}{\varepsilon^4} + \frac{(N)_4}{N^3}\frac{a^6}{\varepsilon^6}\right\} \\ &= \frac{1}{N^3}\left\{\frac{3a^4}{\varepsilon^4} + \frac{(N-3)a^6}{\varepsilon^6}\right\}. \end{aligned}$$

We apply the tower property and conditional independence as in Corollary 4.1, upper bounding the ratio by

$$\begin{aligned}\frac{\frac{1}{(N)_3}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_3]}{\frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_2]} &\leq \frac{N^2}{N^3}\frac{\frac{3a^4}{\varepsilon^4} + \frac{(N-3)a^6}{\varepsilon^6}}{\frac{2\varepsilon^2}{a^2} + \frac{(N-2)\varepsilon^4}{a^4}} \leq \frac{1}{N}\frac{a^6}{\varepsilon^6}\frac{3+(N-3)a^2/\varepsilon^2}{2+(N-2)\varepsilon^2/a^2} \\ &\leq \frac{1}{N}\frac{a^6}{\varepsilon^6}\left\{\frac{3}{2} + \frac{N-3}{N-2}\frac{a^4}{\varepsilon^4}\right\} \leq \frac{1}{N}\left\{\frac{3a^6}{2\varepsilon^6} + \frac{a^{10}}{\varepsilon^{10}}\right\} =: b_N \underset{N\to\infty}{\longrightarrow} 0. \end{aligned}$$

Thus (3.5) is satisfied. It remains to show that, for N sufficiently large, $\mathbb{P}[\tau_N(t)=\infty]=0$ for all finite t, a technicality which is proved in Lemma 4.2 in Appendix **??**. Applying Theorem **??** gives the result. ■

4.5.2 Finite time scale

Lemma 4.2. *Consider a conditional SMC algorithm using multinomial resampling, satisfying the standing assumption and* (4.4). *Then, for all* $N>2$, $\mathbb{P}[\tau_N(t)=\infty]=0$ *for all finite* t.

Proof. Since $c_N(t)\in[0,1]$ almost surely and has strictly positive expectation, for any fixed N the distribution of $c_N(t)$ with given expectation that maximises $\mathbb{P}[c_N(t)=0 \mid \mathcal{F}_{t-1}]$ is two atoms, at 0 and 1 respectively. To ensure the correct expectation, the atom at 1 should have mass $\mathbb{P}[c_N(t)=1 \mid \mathcal{F}_{t-1}] = \mathbb{E}_t[c_N(t)]$, which is bounded below by (4.6). If $c_N(t)>0$

then $c_N(t) \geq 2/(N)_2 > 2/N^2$. Hence, in general $\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{F}_{t-1}] \geq \mathbb{E}_t[c_N(t)]$. Applying (4.6), we have for any finite N,

$$\sum_{t=0}^{\infty} \mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{F}_{t-1}] \geq \sum_{t=0}^{\infty} \mathbb{E}_t[c_N(t)] \geq \sum_{t=0}^{\infty} \frac{1}{N^2} \left\{ \frac{2\varepsilon^2}{a^2} + \frac{(N-2)\varepsilon^4}{a^4} \right\} = \infty$$

By an argument analogous to the conclusion of Lemma 4.1, $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all $t < \infty$. ■

4.5.3 Effect of ancestor sampling

Argue that ancestor sampling removes bias towards assigning offspring to immortal line, and leaves exactly the same genealogy as standard SMC with multinomial resampling.

5 Weak Convergence

At the age of twenty-one he wrote a treatise upon the Binomial Theorem, which has had a European vogue. On the strength of it he won the Mathematical Chair at one of our smaller universities, and had, to all appearances, a most brilliant career before him.

Sherlock Holmes

Some motivation/discussion about weak convergence: why it is more useful than FDDs, that the following theorem has the same conditions as the FDDs one...

We start by defining a suitable metric space. Let $\mathcal{P}_n$ be the space of partitions of $\{1, \ldots, n\}$. Denote by $\mathcal{X}$ the set of all functions mapping $[0, \infty)$ to $\mathcal{P}_n$ that are right-continuous with left limits. (Our rescaled genealogical process $(\mathcal{G}^{(n,N)}_{\tau_N(t)})_{t\geq 0}$ and our encoding of the n-coalescent are piecewise-constant functions mapping time $t \in [0, \infty)$ to partitions, and thus live in the space $\mathcal{X}$.) Finally, equip the space $\mathcal{P}_n$ with the zero-one metric,

$$\rho(\xi, \eta) = 1 - \delta_{\xi\eta} := \begin{cases} 0 & \text{if } \xi = \eta \\ 1 & \text{otherwise} \end{cases} \tag{5.1}$$

for any $\xi, \eta \in \mathcal{P}_n$.

Theorem 5.1. *Let $\nu_t^{(1:N)}$ denote the offspring numbers in an interacting particle system satisfying the standing assumption and such that, for any N sufficiently large, for all finite t, $\mathbb{P}\{\tau_N(t) = \infty\} = 0$. Suppose that there exists a deterministic sequence $(b_N)_{N\in\mathbb{N}}$ such that $\lim_{N\to\infty} b_N = 0$ and*

$$\frac{1}{(N)_3} \sum_{i=1}^{N} \mathbb{E}_t\{(\nu_t^{(i)})_3\} \leq b_N \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}_t\{(\nu_t^{(i)})_2\} \tag{5.2}$$

for all N, uniformly in $t \geq 1$. Then the rescaled genealogical process $(G^{(n,N)}_{\tau_N(t)})_{t\geq 0}$ converges weakly in $(\mathcal{X}, \rho)$ to Kingman's n-coalescent as $N \to \infty$.

Proof. The structure of the proof follows Möhle (1999), albeit with considerable technical complication due to the dependence between generations (non-neutrality) in our model. Is this the main/only source of complication? Since we already have convergence of the finite-dimensional distributions (Theorem ?? refers to a previous chapter not yet written), strengthening this to weak convergence requires relative compactness of the sequence of processes $\{(G^{(n,N)}_{\tau_N(t)})_{t\geq 0}\}_{N\in\mathbb{N}}$.

Ethier and Kurtz (2009, Chapter 3, Corollary 7.4) provides a necessary and sufficient condition for relative compactness: $\mathcal{P}_n$ is finite and therefore complete and separable, and the sample paths of $(G^{(n,N)}_{\tau_N(t)})_{t\geq 0}$ live in $\mathcal{X}$, so the conditions of the corollary are satisfied. The corollary states that the sequence of processes $\{(G^{(n,N)}_{\tau_N(t)})_{t\geq 0}\}_{N\in\mathbb{N}}$ is relatively compact if and only if the following two conditions hold:

1. For every $\varepsilon > 0$, $t \geq 0$ there exists a compact set $\Gamma \subseteq \mathcal{P}_n$ such that
$$\liminf_{N\to\infty} \mathbb{P}[G^{(n,N)}_{\tau_N(t)} \in \Gamma] \geq 1-\varepsilon \tag{5.3}$$
2. For every $\varepsilon > 0$, $t > 0$ there exists $\delta > 0$ such that
$$\liminf_{N\to\infty} \mathbb{P}[\omega(G^{(n,N)}_{\tau_N(\cdot)}, \delta, t) < \varepsilon] \geq 1-\varepsilon \tag{5.4}$$
where ω is the modulus of continuity:
$$\omega(G^{(n,N)}_{\tau_N(\cdot)}, \delta, t) := \inf \max_{i\in[K]} \sup_{u,v\in[T_{i-1},T_i)} \rho\left(G^{(n,N)}_{\tau_N(u)}, G^{(n,N)}_{\tau_N(v)}\right) \tag{5.5}$$
with the infimum taken over all partitions of the form $0 = T_0 < T_1 < \cdots < T_{K-1} < t \leq T_K$ such that $\min_{i\in[K]}(T_i - T_{i-1}) > \delta$. Clarify that such a partition with any K is valid, i.e. K is not fixed.

In our case, Condition 1 is satisfied automatically with $\Gamma = \mathcal{P}_n$, since $\mathcal{P}_n$ is finite and hence compact. Intuitively, Condition 2 ensures that the jumps of the process are well-separated. In our case where ρ is the zero-one metric, we see that $\rho(G^{(n,N)}_{\tau_N(u)}, G^{(n,N)}_{\tau_N(v)})$ is equal to 1 if there is a jump between times u and v, and 0 otherwise. Taking the supremum and maximum then indicates whether there is a jump inside any of the intervals of the given partition; this can only be equal to zero if all of the jumps up to time t occur exactly at the times $T_0, \ldots, T_K$. The infimum over all allowed partitions, then, can only be equal to zero if no two jumps occur less than δ (unscaled) time apart, because of the restriction we placed on these partitions.

The proof is concentrated on proving Condition 2. To do this, we use a coupling with another process that contains all of the jumps of the genealogical process, with the addition of some extra jumps. This process is constructed in such a way that it can be shown to satisfy Condition 2, and hence so does the genealogical process.

Define $p_t := \max_{\xi\in\mathcal{P}_n}\{1 - p_{\xi\xi}(t)\} = 1 - p_{\Delta\Delta}(t)$, where Δ denotes the trivial partition of singletons $\{\{1\},\ldots,\{n\}\}$. For a proof that the maximum is attained at $\xi = \Delta$, see Lemma

5.1. Following Möhle (1999), we now construct the two-dimensional conditionally on $\mathcal{F}$? Markov process $(Z_t, S_t)_{t\in\mathbb{N}_0}$ on $\mathbb{N}_0 \times \mathcal{P}_n$ with transition probabilities

$$\mathbb{P}[Z_t = j, S_t = \eta \mid Z_{t-1} = i, S_{t-1} = \xi] = \begin{cases} 1 - p_t & \text{if } j = i \text{ and } \xi = \eta \\ p_{\xi\xi}(t) + p_t - 1 & \text{if } j = i+1 \text{ and } \xi = \eta \\ p_{\xi\eta}(t) & \text{if } j = i+1 \text{ and } \xi \neq \eta \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

and initial state $Z_0 = 0$, $S_0 = \Delta$. The construction is such that the marginal (S_t) has the same distribution as the genealogical process of interest, and (Z_t) has jumps at all the times (S_t) does plus some extra jumps. (The definition of p_t ensures that the probability in the second case is non-negative, attaining the value zero when $\xi = \Delta$.) And the transition probabilities (jump times) of Z do not depend on the current state.

Denote by $0 = T_0^{(N)} < T_1^{(N)} < \dots$ the jump times of the rescaled process $(Z_{\tau_N(t)})_{t\geq 0}$, and by $\varpi_i^{(N)} := T_i^{(N)} - T_{i-1}^{(N)}$ the corresponding holding times.

Suppose that for some $t > 0$, there exists $m \in \mathbb{N}$ and $\delta > 0$ such that $\varpi_i^{(N)} > \delta$ for all $i \in \{1, \dots, m\}$, and $T_m^{(N)} \geq t$. Then $K_N := \min\{i : T_i^{(N)} \geq t\}$ is well-defined with $1 \leq K_N \leq m$, and $T_1^{(N)}, \dots, T_{K_N}^{(N)}$ form a partition of the form required for Condition 2. Indeed $(Z_{\tau_N(\cdot)})$ is constant on every interval $[T_{i-1}^{(N)}, T_i^{(N)})$ by construction, so $\omega((Z_{\tau_N(\cdot)}), \delta, t) = 0$. We therefore have that for each $m \in \mathbb{N}$ and $\delta > 0$,

$$\mathbb{P}[\omega((Z_{\tau_N(\cdot)}), \delta, t) < \varepsilon] \geq \mathbb{P}[T_m^{(N)} \geq t, \varpi_i^{(N)} > \delta \,\forall i \in \{1, \dots, m\}]. \tag{5.7}$$

Thus a sufficient condition for Condition 2 is: for any $\varepsilon > 0$, $t > 0$, there exist $m \in \mathbb{N}$, $\delta > 0$ such that

$$\liminf_{N\to\infty} \mathbb{P}[T_m^{(N)} \geq t, \varpi_i^{(N)} > \delta \,\forall i \in \{1, \dots, m\}] \geq 1 - \varepsilon. \tag{5.8}$$

Since $T_m^{(N)} = \varpi_1^{(N)} + \dots + \varpi_m^{(N)}$, there is a positive correlation between $T_m^{(N)}$ and each of the $\varpi_i^{(N)}$, so

$$\begin{aligned} \mathbb{P}[T_m^{(N)} \geq{} & t, \varpi_i^{(N)} > \delta \,\forall i \in \{1, \dots, m\}] \\ &= \mathbb{P}[T_m^{(N)} \geq t \mid \varpi_i^{(N)} > \delta \,\forall i \in \{1, \dots, m\}]\, \mathbb{P}[\varpi_i^{(N)} > \delta \,\forall i \in \{1, \dots, m\}] \\ &\geq \mathbb{P}[T_m^{(N)} \geq t]\, \mathbb{P}[\varpi_i^{(N)} > \delta \,\forall i \in \{1, \dots, m\}]. \end{aligned} \tag{5.9}$$

Due to Lemma 5.2, the limiting distributions of $\varpi_i^{(N)}$ are i.i.d. $\mathrm{Exp}(\alpha_n)$, so

$$\liminf_{N\to\infty} \mathbb{P}[\varpi_i^{(N)} > \delta \,\forall i \in \{1, \dots, m\}] = (e^{-\alpha_n \delta})^m \tag{5.10}$$

and

$$\liminf_{N\to\infty} \mathbb{P}[T_m^{(N)} \geq t] = \liminf_{N\to\infty} \mathbb{P}[\varpi_1^{(N)} + \cdots + \varpi_m^{(N)} \geq t] = e^{-\alpha_n \delta} \sum_{i=0}^{m-1} \frac{(\alpha_n t)^i}{i!}. \tag{5.11}$$

using the series expansion for the Erlang cumulative distribution function. citation? Hence

$$\liminf_{N\to\infty} \mathbb{P}[T_m^{(N)} \geq t, \varpi_i^{(N)} > \delta \,\forall i \in \{1, \ldots, m\}] \geq (e^{-\alpha_n \delta})^{m+1} \sum_{i=0}^{m-1} \frac{(\alpha_n t)^i}{i!}, \tag{5.12}$$

which can be made $\geq 1 - \varepsilon$ by taking m sufficiently large and δ sufficiently small. Since this argument applies for any ε and t, (5.8) and hence Condition 2 is satisfied, and the proof is complete. ■

Lemma 5.1. $\max_{\xi \in \mathcal{P}_n} (1 - p_{\xi\xi}(t)) = 1 - p_{\Delta\Delta}(t)$.

Proof. Consider any $\xi \in E$ consisting of k blocks ($1 \leq k \leq n-1$), and any $\xi' \in E$ consisting of $k+1$ blocks. From the definition of $p_{\xi\eta}(t)$ (Koskela et al. 2018, Equation (1)),

$$p_{\xi\xi}(t) = \frac{1}{(N)_k} \sum_{\substack{i_1,\ldots,i_k \\ \text{all distinct}}} \nu_t^{(i_1)} \cdots \nu_t^{(i_k)}. \tag{5.13}$$

Similarly,

$$\begin{aligned} p_{\xi'\xi'}(t) &= \frac{1}{(N)_{k+1}} \sum_{\substack{i_1,\ldots,i_k,i_{k+1} \\ \text{all distinct}}} \nu_t^{(i_1)} \cdots \nu_t^{(i_k)} \nu_t^{(i_{k+1})} \\ &= \frac{1}{(N)_k (N-k)} \sum_{\substack{i_1,\ldots,i_k \\ \text{all distinct}}} \left\{ \nu_t^{(i_1)} \cdots \nu_t^{(i_k)} \sum_{\substack{i_{k+1}=1 \\ \text{also distinct}}}^{N} \nu_t^{(i_{k+1})} \right\}. \end{aligned} \tag{5.14}$$

Discarding the zero summands,

$$p_{\xi'\xi'}(t) = \frac{1}{(N)_k (N-k)} \sum_{\substack{i_1,\ldots,i_k \\ \text{all distinct:} \\ \nu_t^{(i_1)},\ldots,\nu_t^{(i_k)} > 0}} \left\{ \nu_t^{(i_1)} \cdots \nu_t^{(i_k)} \sum_{\substack{i_{k+1}=1 \\ \text{also distinct}}}^{N} \nu_t^{(i_{k+1})} \right\}. \tag{5.15}$$

The inner sum is

$$\sum_{\substack{i_{k+1}=1 \\ \text{also distinct}}}^{N} \nu_t^{(i_{k+1})} = \left\{ \sum_{i=1}^{N} \nu_t^{(i)} - \sum_{i \in \{i_1,\ldots,i_k\}} \nu_t^{(i)} \right\} \leq N - k \tag{5.16}$$

since $\nu_t^{(i_1)}, \dots, \nu_t^{(i_k)}$ are all at least 1. Hence

$$p_{\xi'\xi'}(t) \le \frac{N-k}{(N)_k(N-k)} \sum_{\substack{i_1,\dots,i_k \\ \text{all distinct:} \\ \nu_t^{(i_1)},\dots,\nu_t^{(i_k)}>0}} \nu_t^{(i_1)} \cdots \nu_t^{(i_k)} = p_{\xi\xi}(t). \tag{5.17}$$

Thus $p_{\xi\xi}(t)$ is decreasing in the number of blocks of ξ, and is therefore minimised by taking $\xi = \Delta$, which achieves the maximum n blocks. This choice in turn maximises $1 - p_{\xi\xi}(t)$, as required. ∎

Lemma 5.2. *The finite-dimensional distributions of $\varpi_1^{(N)}, \varpi_2^{(N)}, \dots$ converge as $N \to \infty$ to those of $\varpi_1, \varpi_2, \dots$, where the ϖ_i are independent* $\mathrm{Exp}(\alpha_n)$ *distributed random variables.*

Proof. There is a continuous bijection between the jump times $T_1^{(N)}, T_2^{(N)}, \dots$ and the holding times $\varpi_1^{(N)}, \varpi_2^{(N)}, \dots$, so convergence of the holding times to $\varpi_1, \varpi_2, \dots$ is equivalent to convergence of the jump times to $T_1, T_2, \dots$, where $T_i := \varpi_1 + \cdots + \varpi_i$. We will work with the jump times, following the structure of Möhle (1999, Lemma 3.2).

The idea is to prove by induction that, for any $k \in \mathbb{N}$ and $t_1, \dots, t_k > 0$,

$$\lim_{N\to\infty} \mathbb{P}[T_1^{(N)} \le t_1, \dots, T_k^{(N)} \le t_k] = \mathbb{P}[T_1 \le t_1, \dots, T_k \le t_k]. \tag{5.18}$$

Take the basis case $k = 1$. Then

$$\mathbb{P}[T_1 \le t] = \mathbb{P}[\varpi_1 \le t] = 1 - e^{-\alpha_n t} \tag{5.19}$$

and $T_1^{(N)} > t$ if and only if Z has no jumps up to time t: Expectation appears by tower property to remove (implicit) conditioning in transition probabilities?

$$\mathbb{P}[T_1^{(N)} > t] = \mathbb{E}\left[\prod_{r=1}^{\tau_N(t)} (1 - p_r)\right]. \tag{5.20}$$

Lemma 5.6 shows that this probability converges to $e^{-\alpha_n t}$ as required.

For the induction step, assume that (5.18) holds for some k. We have the following decomposition:

$$\mathbb{P}[T_1^{(N)} \le t_1, \dots, T_{k+1}^{(N)} \le t_{k+1}] = \mathbb{P}[T_1^{(N)} \le t_1, \dots, T_k^{(N)} \le t_k] - \mathbb{P}[T_1^{(N)} \le t_1, \dots, T_k^{(N)} \le t_k, T_{k+1}^{(N)} > t_{k+1}] \tag{5.21}$$

The first term on the RHS converges to $\mathbb{P}[T_1 \le t_1, \dots, T_k \le t_k]$ by the induction hypothesis, and it remains to show that

$$\lim_{N\to\infty} \mathbb{P}[T_1^{(N)} \le t_1, \dots, T_k^{(N)} \le t_k, T_{k+1}^{(N)} > t_{k+1}] = \mathbb{P}[T_1 \le t_1, \dots, T_k \le t_k, T_{k+1} > t_{k+1}]. \tag{5.22}$$

As shown in Möhle (1999), the RHS

$$\mathbb{P}[T_1 \le t_1, \dots, T_k \le t_k, T_{k+1} > t_{k+1}] = \alpha_n^k e^{-\alpha_n t} \sum_{\substack{i_1 \le \cdots \le i_{k-1} \\ \in \{0,\dots,k\}: \\ i_j \ge j \forall j}} \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!}. \tag{5.23}$$

The event on the LHS can be written (Möhle 1999)

$$\mathbb{P}[T_1^{(N)} \le t_1, \dots, T_k^{(N)} \le t_k, T_{k+1}^{(N)} > t_{k+1}] = \mathbb{E}\left[\sum_{\substack{r_1 < \cdots < r_k: \\ r_i \le \tau_N(t_i) \forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1 \\ \notin \{r_1,\dots,r_k\}}}^{\tau_N(t)} (1 - p_r)\right)\right], \tag{5.24}$$

that is, there are jumps at some times $r_1, \dots, r_k$ and identity transitions at all other times. Due to Lemmata 5.7 and 5.8, this probability converges to the correct limit. This completes the induction. ∎

5.1 Bounds on sum-products

Lemma 5.3. *Fix $t > 0$, $l \in \mathbb{N}$.*

(a) $\displaystyle\sum_{s_1 \ne \cdots \ne s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \le (t+1)^l$

(b) $\displaystyle t^l - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right)\binom{l}{2}(t+1)^{l-2} \le \sum_{s_1 \ne \cdots \ne s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \le t^l + c_N(\tau_N(t))(t+1)^l$

Proof. **(a)** It is a true fact that

$$\sum_{s_1 \ne \cdots \ne s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \le \left(\sum_{s=0}^{\tau_N(t)} c_N(s)\right)^l, \tag{5.25}$$

as can be seen by considering the multinomial expansion of the RHS. By definition of τ_N,

$$t \le \sum_{s=0}^{\tau_N(t)} c_N(s) \le t + 1, \tag{5.26}$$

hence

$$\sum_{s_1 \ne \cdots \ne s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \le (t+1)^l. \tag{5.27}$$

(b) As pointed out in Koskela et al. (2018, Equation (8)),

$$\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \geq \left(\sum_{s=0}^{\tau_N(t)} c_N(s) \right)^l - \binom{l}{2} \left(\sum_{s=0}^{\tau_N(t)} c_N(s)^2 \right) \left(\sum_{s=0}^{\tau_N(t)} c_N(s) \right)^{l-2} . \tag{5.28}$$

Substituting (5.26) into the RHS of (5.28) yields the lower bound.

For the upper bound we have

$$\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \leq \left(\sum_{s=0}^{\tau_N(t)} c_N(s) \right)^l \leq \left(\sum_{s=0}^{\tau_N(t)-1} c_N(s) + c_N(\tau_N(t)) \right)^l \leq [t + c_N(\tau_N(t))]^l , \tag{5.29}$$

again using the definition of τ_N. A binomial expansion yields

$$[t + c_N(\tau_N(t))]^l = t^l + \sum_{i=0}^{l-1} \binom{l}{i} t^i c_N(\tau_N(t))^{l-i} = t^l + c_N(\tau_N(t)) \sum_{i=0}^{l-1} \binom{l}{i} t^i c_N(\tau_N(t))^{l-1-i}, \tag{5.30}$$

then since $c_N(s) \leq 1$ for all s,

$$\sum_{i=0}^{l-1} \binom{l}{i} t^i c_N(\tau_N(t))^{l-1-i} \leq \sum_{i=0}^{l-1} \binom{l}{i} t^i \leq (t+1)^l. \tag{5.31}$$

Putting this together yields the upper bound. ■

Lemma 5.4. *Fix $t > 0$, $l \in \mathbb{N}$. Let B be a positive constant which may depend on n.*

$$\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) + B D_N(s_j)] \leq \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) + \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) (t+1)^{l-1} (1+B)^l. \tag{5.32}$$

Proof. We start with a binomial expansion:

$$\begin{aligned} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) + B D_N(s_j)] &= \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \sum_{\mathcal{I} \subseteq [l]} B^{l-|\mathcal{I}|} \left(\prod_{i \in \mathcal{I}} c_N(s_i) \right) \left(\prod_{j \notin \mathcal{I}} D_N(s_j) \right) \\ &= \sum_{\mathcal{I} \subseteq [l]} B^{l-|\mathcal{I}|} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i \in \mathcal{I}} c_N(s_i) \right) \left(\prod_{j \notin \mathcal{I}} D_N(s_j) \right) \end{aligned} \tag{5.33}$$

where $[l] := \{1, \ldots, l\}$. Since the sum is over all permutations of $s_1, \ldots, s_l$, we may arbitrarily choose an ordering for $\{1, \ldots, l\}$ such that $\mathcal{I} = \{1, \ldots, |\mathcal{I}|\}$:

$$\sum_{\mathcal{I} \subseteq [l]} B^{l-|\mathcal{I}|} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i \in \mathcal{I}} c_N(s_i) \right) \left(\prod_{j \notin \mathcal{I}} D_N(s_j) \right) = \sum_{I=0}^{l} \binom{l}{I} B^{l-I} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i) \right) \left(\prod_{j=I+1}^{l} D_N(s_j) \right) \tag{5.34}$$

Separating the term $I = l$,

$$\sum_{I=0}^{l} \binom{l}{I} B^{l-I} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i) \right) \left(\prod_{j=I+1}^{l} D_N(s_j) \right)$$
$$= \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) + \sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i) \right) \left(\prod_{j=I+1}^{l} D_N(s_j) \right). \tag{5.35}$$

In the second term on the RHS, there is always at least one D_N term, and $c_N(s) \geq D_N(s)$ for all s (Koskela et al. 2018, p.9), so we can write

$$\begin{aligned}
\sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i) \right) \left(\prod_{j=I+1}^{l} D_N(s_j) \right) &\leq \sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{l-1} c_N(s_i) \right) D_N(s_l) \\
&\leq \sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} \left(\sum_{s_1 \neq \cdots \neq s_{l-1}}^{\tau_N(t)} \prod_{i=1}^{l-1} c_N(s_i) \right) \sum_{s_l=1}^{\tau_N(t)} D_N(\\
&\leq \sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} (t+1)^{l-1} \sum_{s=1}^{\tau_N(t)} D_N(s)
\end{aligned} \tag{5.36}$$

using (5.25) and (5.26). Finally, by the Binomial Theorem,

$$\sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} (t+1)^{l-1} \sum_{s=1}^{\tau_N(t)} D_N(s) \leq \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) (t+1)^{l-1} (1+B)^l, \tag{5.37}$$

which, together with (5.35), concludes the proof. ■

Lemma 5.5. *Fix $t > 0$, $l \in \mathbb{N}$. Let B be a positive constant which may depend on n.*

$$\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) - B D_N(s_j)] \geq \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) - \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) (t+1)^{l-1} (1+B)^l. \tag{5.38}$$

Proof. A binomial expansion and subsequent manipulation as in (5.33)–(5.35) gives

$$\begin{aligned}
\sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} \left[c_N(s_j) - BD_N(s_j)\right] &= \sum_{\mathcal{I}\subseteq[l]} (-B)^{l-|\mathcal{I}|} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \left(\prod_{i\in\mathcal{I}} c_N(s_i)\right)\left(\prod_{j\notin\mathcal{I}} D_N(s_j)\right) \\
&= \sum_{I=0}^{l} \binom{l}{I} (-B)^{l-I} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i)\right)\left(\prod_{j=I+1}^{l} D_N(s_j)\right) \\
&= \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) + \sum_{I=0}^{l-1} \binom{l}{I} (-B)^{l-I} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i)\right)\Bigg(\prod_{j=I+1}^{l} \\
&\geq \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) - \sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i)\right)\Bigg(\prod_{j=I+1}^{l} D_N
\end{aligned} \tag{5.39}$$

where the last inequality just multiplies some positive terms by -1. Then (5.36)–(5.37) can be applied directly (noting that an upper bound on negative terms gives a lower bound overall):

$$-\sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i)\right)\left(\prod_{j=I+1}^{l} D_N(s_j)\right) \geq -\left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right)(t+1)^{l-1}(1+B)^l \tag{5.40}$$

which concludes the proof. ■

5.2 Main components of weak convergence

Lemma 5.6 (Basis step)**.** *For any* $0 < t < \infty$,

$$\lim_{N\to\infty} \mathbb{E}\left[\prod_{r=1}^{\tau_N(t)} (1-p_r)\right] = e^{-\alpha_n t} \tag{5.41}$$

where $\alpha_n := n(n-1)/2$.

Proof. We start by showing that $\lim_{N\to\infty} \mathbb{E}\left[\prod_{r=1}^{\tau_N(t)} (1-p_r)\right] \leq e^{-\alpha_n t}$.
From Koskela et al. (2018, Lemma 1 Case 1), taking $\xi = \Delta$, we have for each r

$$1 - p_r = p_{\Delta\Delta}(r) \leq 1 - \alpha_n(1+O(N^{-1}))\left[c_N(r) - B'_n D_N(r)\right] \tag{5.42}$$

where the $O(N^{-1})$ term does not depend on r. When $N \geq 3$, a sufficient condition to ensure the bound in (5.42) is non-negative is that the event

$$E^1_N(r) := \left\{c_N(r) \leq \frac{(N-2)_{n-2}}{\alpha_n N^{n-2}}\right\} \tag{5.43}$$

occurs. We will also need to control the sign of $c_N(r) - B_n' D_N(r)$, for which we define the event

$$E_N^2(r) := \left\{ c_N(r) \geq B_n' D_N(r) \right\}, \tag{5.44}$$

and we define $E_N^1 := \bigcap_{r=1}^{\tau_N(t)} E_N^1(r)$ and $E_N^2 := \bigcap_{r=1}^{\tau_N(t)} E_N^2(r)$. Then

$$1 - p_r = p_{\Delta\Delta}(r) \leq 1 - \alpha_n (1 + O(N^{-1})) \left[c_N(r) - B_n' D_N(r) \right] \mathbb{1}_{E_N^1 \cap E_N^2}. \tag{5.45}$$

Applying a multinomial expansion and then separating the positive and negative terms,

$$\begin{aligned}
\prod_{r=1}^{\tau_N(t)} (1 - p_r) &\leq 1 + \sum_{l=1}^{\tau_N(t)} (-\alpha_n)^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} \left[c_N(s_j) - B_n' D_N(s_j) \right] \mathbb{1}_{E_N^1 \cap E_N^2} \\
&= 1 + \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} \left[c_N(s_j) - B_n' D_N(s_j) \right] \mathbb{1}_{E_N^1 \cap E_N^2} \\
&\quad - \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} \left[c_N(s_j) - B_n' D_N(s_j) \right] \mathbb{1}_{E_N^1 \cap E_N^2}.
\end{aligned} \tag{5.46}$$

This is further bounded by applying Lemma 5.5 and then both bounds of Lemma 5.3(b):

$$\begin{aligned}
\prod_{r=1}^{\tau_N(t)} (1 - p_r) &\leq 1 + \Bigg\{ \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \\
&\quad - \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \left[\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) - \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) (t+1)^{l-1} (1 + B_n')^l \right] \\
&\leq 1 + \Bigg\{ \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \left\{ t^l + c_N(\tau_N(t))(t+1)^l \right\} \\
&\quad - \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \left[t^l - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right) \binom{l}{2} (t+1)^{l-2} \right] \\
&\quad - \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) (t+1)^{l-1} (1 + B_n')^l \Bigg\} \mathbb{1}_{E_N^1 \cap E_N^2}.
\end{aligned} \tag{5.47}$$

Collecting some terms,

$$
\begin{aligned}
\prod_{r=1}^{\tau_N(t)}(1-p_r) \leq 1 &+ \sum_{l=1}^{\tau_N(t)}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{E_N^1\cap E_N^2} + c_N(\tau_N(t))\sum_{\substack{l=2\\ \text{even}}}^{\tau_N(t)}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}(t+1)^l \\
&+ \left(\sum_{s=1}^{\tau_N(t)}c_N(s)^2\right)\sum_{\substack{l=1\\ \text{odd}}}^{\tau_N(t)}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}\binom{l}{2}(t+1)^{l-2} \\
&+ \left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)\sum_{\substack{l=1\\ \text{odd}}}^{\tau_N(t)}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}(t+1)^{l-1}(1+B_n')^l \\
\leq 1 &+ \sum_{l=1}^{\infty}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{\{\tau_N(t)\geq l\}}\mathbb{1}_{E_N^1\cap E_N^2} + c_N(\tau_N(t))\exp[\alpha_n(1+O(N^{-1}))(t+ \\
&+ \left(\sum_{s=1}^{\tau_N(t)}c_N(s)^2\right)\frac{1}{2}\alpha_n^2\exp[\alpha_n(1+O(N^{-1}))(t+1)] \\
&+ \left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n')].
\end{aligned}
\tag{5.48}
$$

Now, taking the expectation and limit, then applying Brown et al. (2021, Equations (3.3)–(3.5)), and Lemmata 5.11, 5.12 and 5.14 to deal with the indicators,

$$
\begin{aligned}
\lim_{N\to\infty}\mathbb{E}\left[\prod_{r=1}^{\tau_N(t)}(1-p_r)\right] &\leq 1 + \sum_{l=1}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l\lim_{N\to\infty}\mathbb{P}\left[\{\tau_N(t)\geq l\}\cap E_N^1\cap E_N^2\right] + \lim_{N\to\infty}\mathbb{E}\left[c_N(\tau_N(t))\right]\exp \\
&\quad + \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)}c_N(s)^2\right]\frac{1}{2}\alpha_n^2\exp[\alpha_n(t+1)] \\
&\quad + \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)}D_N(s)\right]\exp[\alpha_n(t+1)(1+B_n')] \\
&= 1 + \sum_{l=1}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l = e^{-\alpha_n t}.
\end{aligned}
\tag{5.49}
$$

Passing the limit and expectation inside the infinite sum is justified by dominated convergence and Fubini.

It remains to show the corresponding lower bound $\lim_{N\to\infty}\mathbb{E}\left[\prod_{r=1}^{\tau_N(t)}(1-p_r)\right]\geq e^{-\alpha_n t}$. From Brown et al. (2021, Equation (3.14)), taking $\xi=\Delta$, we have

$$
1-p_t = p_{\Delta\Delta}(t) \geq 1-\alpha_n(1+O(N^{-1}))\left[c_N(t)+B_nD_N(t)\right] \tag{5.50}
$$

where $B_n > 0$ and the $O(N^{-1})$ term does not depend on t. In particular,

$$1 - p_t = p_{\Delta\Delta}(t) \geq 1 - \frac{N^{n-2}}{(N-2)_{n-2}} \alpha_n [c_N(t) + B_n D_N(t)]. \tag{5.51}$$

Since $D_N(s) \leq c_N(s)$ for all s (Koskela et al. 2018, p.9), a sufficient condition for this bound to be non-negative is

$$E_N^3(r) := \left\{ c_N(r) \leq \frac{(N-2)_{n-2}}{N^{n-2}} \alpha_n^{-1} (1 + B_n)^{-1} \right\}, \tag{5.52}$$

and we again define $E_N^3 := \bigcap_{r=1}^{\tau_N(t)} E_N^3(r)$. We now apply a multinomial expansion to the product, and split into positive and negative terms:

$$\begin{aligned}
\prod_{r=1}^{\tau_N(t)} (1 - p_r) &\geq \left\{ 1 + \sum_{l=1}^{\tau_N(t)} (-\alpha_n)^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) + B_n D_N(s_j)] \right\} \mathbb{1}_{E_N^3} \\
&= \Bigg\{ 1 + \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) + B_n D_N(s_j)] \\
&\qquad - \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) + B_n D_N(s_j)] \Bigg\} \mathbb{1}_{E_N^3}
\end{aligned} \tag{5.53}$$

This is further bounded by applying Lemma 5.4 and both bounds in Lemma 5.3(b):

$$\begin{aligned}
\prod_{r=1}^{\tau_N(t)} (1 - p_r) &\geq \Bigg\{ 1 + \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \\
&\qquad - \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \left[\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) + \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) (t+1)^{l-1} (1 + B_n)^l \right] \\
&\geq \Bigg\{ 1 + \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \left[t^l - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right) \binom{l}{2} (t+1)^{l-2} \right] \\
&\qquad - \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l (1 + O(N^{-1})) \frac{1}{l!} \Bigg[t^l + c_N(\tau_N(t))(t+1)^l + \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) (t+1)^{l-1} (1 + B_n
\end{aligned} \tag{5.54}$$

Collecting terms,

$$\begin{aligned}\prod_{r=1}^{\tau_N(t)}(1-p_r) \geq \sum_{l=0}^{\tau_N(t)}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{E_N^3} - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right)\sum_{\substack{l=2\\ \text{even}}}^{\tau_N(t)}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}\binom{l}{2}(t+1)^{l-}\\
&- c_N(\tau_N(t))\sum_{\substack{l=1\\ \text{odd}}}^{\tau_N(t)}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}(t+1)^l\\
&- \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right)\sum_{\substack{l=1\\ \text{odd}}}^{\tau_N(t)}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}(t+1)^{l-1}(1+B_n)^l\\
\geq \sum_{l=0}^{\infty}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{E_N^3}\mathbb{1}_{\{\tau_N(t)\geq l\}} - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right)\frac{1}{2}\alpha_n^2\exp[\alpha_n(1+O(N^{-1}))(t\\
&- c_N(\tau_N(t))\exp[\alpha_n(1+O(N^{-1}))(t+1)]\\
&- \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n)]. \qquad (5.55)\end{aligned}$$

Now, taking the expectation and limit, and applying Brown et al. (2021, Equations (3.3)–(3.5)) to show that all but the first sum vanish, and Lemmata 5.12 and 5.11 to show that $\lim_{N\to\infty}\mathbb{P}[\{\tau_N(t)\geq l\}\cap E_N^3]=1$,

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}\left[\prod_{r=1}^{\tau_N(t)}(1-p_r)\right] &\geq \sum_{l=0}^{\infty}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\lim_{N\to\infty}\mathbb{P}\left[\{\tau_N(t)\geq l\}\cap E_N^3\right]\\
&\quad - \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right]\frac{1}{2}\alpha_n^2\exp[\alpha_n(t+1)]\\
&\quad - \lim_{N\to\infty}\mathbb{E}\left[c_N(\tau_N(t))\right]\exp[\alpha_n(t+1)]\\
&\quad - \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} D_N(s)\right]\exp[\alpha_n(t+1)(1+B_n)]\\
&= \sum_{l=0}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l = e^{-\alpha_n t}. \qquad (5.56)\end{aligned}$$

Again, passing the limit and expectation inside the infinite sum is justified by dominated convergence and Fubini. Combining the upper and lower bounds in (5.49) and (5.56) respectively concludes the proof. ■

Lemma 5.7 (Induction step upper bound)**.** *Fix $k \in \mathbb{N}$, $i_0 := 0$, $i_k := k$. For any sequence of times $0 = t_0 \leq t_1 \leq \cdots \leq t_k \leq t$,*

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\leq\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\ldots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right] \leq \alpha_n^k e^{-\alpha_n t} \sum_{\substack{i_1\leq\cdots\leq i_{k-1}\\ \in\{0,\ldots,k\}:\\ i_j\geq j\forall j}} \prod_{j=1}^{k} \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}. \tag{5.57}$$

Proof. We use the bound on $(1-p_r)$ from (5.42) and apply a multinomial expansion, defining as in (5.43) and (5.44) respectively the sequences of events E_N^1 and E_N^2 which ensure the bounds are non-negative:

$$\begin{aligned}
\prod_{\substack{r=1\\ \notin\{r_1,\ldots,r_k\}}}^{\tau_N(t)} (1-p_r) &\leq \prod_{\substack{r=1\\ \notin\{r_1,\ldots,r_k\}}}^{\tau_N(t)} \left\{1-\alpha_n(1+O(N^{-1}))[c_N(r)-B_n' D_N(r)]\mathbb{1}_{E_N^1\cap E_N^2}\right\} \\
&= 1+\sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l (1+O(N^{-1}))\frac{1}{l!} \sum_{\substack{s_1\neq\cdots\neq s_l\\ \notin\{r_1,\ldots,r_k\}}}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j)-B_n' D_N(s_j)]\mathbb{1}_{E_N^1\cap E_N^2} \\
&= 1+\sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l (1+O(N^{-1}))\frac{1}{l!} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j)-B_n' D_N(s_j)]\mathbb{1}_{E_N^1\cap E_N^2} \\
&\quad - \sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l (1+O(N^{-1}))\frac{1}{l!} \sum_{\substack{s_1\neq\cdots\neq s_l\\ \exists i,i':s_i=r_{i'}}} \prod_{j=1}^{l} [c_N(s_j)-B_n' D_N(s_j)]\mathbb{1}_{E_N^1\cap E_N^2}.
\end{aligned} \tag{5.58}$$

The penultimate line above is exactly the expansion we had in the basis step (5.46), except for the limit on l, and as such following the same arguments gives a bound analogous to that in (5.48):

$$\begin{aligned}
&1+\sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l (1+O(N^{-1}))\frac{1}{l!} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j)-B_n' D_N(s_j)]\mathbb{1}_{E_N^1\cap E_N^2} \\
&\qquad \leq 1+\sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l (1+O(N^{-1}))\frac{1}{l!} t^l \mathbb{1}_{E_N^1\cap E_N^2} + c_N(\tau_N(t)) \exp[\alpha_n(1 \\
&\qquad\quad + \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right) \frac{1}{2}\alpha_n^2 \exp[\alpha_n(1+O(N^{-1}))(t+1)] \\
&\qquad\quad + \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right) \exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n')].
\end{aligned} \tag{5.59}$$

For the last line of (5.58),

$$
\begin{aligned}
-\sum_{l=1}^{\tau_N(t)-k}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}\sum_{\substack{s_1\neq\cdots\neq s_l\\ \exists i,i':s_i=r_{i'}}}\prod_{j=1}^{l}\{c_N(s_j)-B_n'D_N(s_j)\}\mathbb{1}_{E_N^1\cap E_N^2}& \\
\leq \sum_{l=1}^{\tau_N(t)-k}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}\sum_{\substack{s_1\neq\cdots\neq s_l\\ \exists i,i':s_i=r_{i'}}}\prod_{j=1}^{l}\{c_N(s_j)+B_n'D_N(s_j)\}& \\
\leq \sum_{l=1}^{\tau_N(t)-k}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}\sum_{\substack{s_1\neq\cdots\neq s_l\\ \exists i,i':s_i=r_{i'}}}(1+B_n')^l\prod_{j=1}^{l}c_N(s_j)& \\
\leq \sum_{l=1}^{\tau_N(t)-k}\alpha_n^l(1+O(N^{-1}))\frac{1}{(l-1)!}\sum_{s_1\in\{r_1,\ldots,r_k\}}\sum_{s_2\neq\cdots\neq s_l}^{\tau_N(t)}(1+B_n')^l\prod_{j=1}^{l}c_N(&\\
= \sum_{s\in\{r_1,\ldots,r_k\}}c_N(s)\sum_{l=1}^{\tau_N(t)-k}\alpha_n^l(1+O(N^{-1}))\frac{1}{(l-1)!}(1+B_n')^l\sum_{s_1\neq\cdots\neq s_{l-1}}^{\tau_N(t)}& \\
\leq \sum_{j=1}^{k}c_N(r_j)\sum_{l=1}^{\tau_N(t)-k}\alpha_n^l(1+O(N^{-1}))\frac{1}{(l-1)!}(1+B_n')^l(t+1)^{l-1}& \\
\leq \left(\sum_{j=1}^{k}c_N(r_j)\right)\alpha_n(1+B_n')\exp[\alpha_n(1+O(N^{-1}))(1+B_n')(t+1)],&
\end{aligned}
\tag{5.60}
$$

where the penultimate inequality uses Lemma 5.3(a). Putting these together, we have

$$
\begin{aligned}
\prod_{\substack{r=1\\ \notin\{r_1,\ldots,r_k\}}}^{\tau_N(t)}(1-p_r)\leq 1+\sum_{l=1}^{\tau_N(t)-k}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{E_N^1\cap E_N^2}+c_N(\tau_N(t))\exp[\alpha_n(1+O(N^{-1}))(t+1&\\
+\left(\sum_{s=1}^{\tau_N(t)}c_N(s)^2\right)\frac{1}{2}\alpha_n^2\exp[\alpha_n(1+O(N^{-1}))(t+1)]&\\
+\left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n')]&\\
+\left(\sum_{j=1}^{k}c_N(r_j)\right)\alpha_n(1+B_n')\exp[\alpha_n(1+O(N^{-1}))(1+B_n')(t+1)].&
\end{aligned}
\tag{5.61}
$$

Meanwhile, using the bound on p_r from (5.50) then applying a modification of Lemma 5.4

where the sum is over ordered indices rather than distinct indices,

$$\begin{aligned}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}p_{r_i} &\le \alpha_n^k(1+O(N^{-1}))\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}[c_N(r_i)+B_nD_N(r_i)]\\ &\le \alpha_n^k(1+O(N^{-1}))\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)+\left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)\alpha_n^k(1+O(N^{-1}))(t+1)^{k-1}(1\end{aligned} \tag{5.62}$$

A more liberal (but simpler) bound can be arrived at thus:

$$\begin{aligned}\prod_{i=1}^{k}p_{r_i} &\le \alpha_n^k(1+O(N^{-1}))\prod_{i=1}^{k}[c_N(r_i)+B_nD_N(r_i)]\\ &\le \alpha_n^k(1+O(N^{-1}))\prod_{i=1}^{k}c_N(r_i)(1+B_n)\\ &\le \alpha_n^k(1+O(N^{-1}))(1+B_n)^k\prod_{i=1}^{k}c_N(r_i)\end{aligned} \tag{5.63}$$

which, using Lemma 5.3(a), also leads to the deterministic bound

$$\begin{aligned}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}p_{r_i} &\le \alpha_n^k(1+O(N^{-1}))(1+B_n)^k\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\\ &\le \alpha_n^k(1+O(N^{-1}))(1+B_n)^k\frac{1}{k!}\sum_{r_1\ne\cdots\ne r_k}^{\tau_N(t)}\prod_{i=1}^{k}c_N(r_i)\\ &\le \alpha_n^k(1+O(N^{-1}))(1+B_n)^k\frac{1}{k!}(t+1)^k.\end{aligned} \tag{5.64}$$

Combining (5.61) with the other product, the expression inside the expectation in (5.57) is bounded above by

$$
\begin{aligned}
&\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\\
&\quad\le \left\{1+\sum_{l=1}^{\tau_N(t)-k}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{E_N^1\cap E_N^2}\right\}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}p_{r_i}\\
&\qquad+\left\{c_N(\tau_N(t))\exp[\alpha_n(1+O(N^{-1}))(t+1)]+\left(\sum_{s=1}^{\tau_N(t)}c_N(s)^2\right)\frac{1}{2}\alpha_n^2\exp[\alpha_n(1+O(N^{-1}))(\right.\\
&\qquad\qquad\left.+\left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n')]\right\}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod\\
&\qquad+\exp[\alpha_n(1+O(N^{-1}))(1+B_n')(t+1)]\alpha_n(1+B_n')\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\sum_{j=1}^{k}c_N(r_j)\prod_{i=1}^{k}p_{r_i}.
\end{aligned}
\tag{5.65}
$$

Applying the various bounds (5.62)–(5.64), we have

$$
\begin{aligned}
&\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\\
&\quad\le \alpha_n^k(1+O(N^{-1}))\left\{1+\sum_{l=1}^{\tau_N(t)-k}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{E_N^1\cap E_N^2}\right\}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\\
&\qquad+\left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)\alpha_n^k(1+O(N^{-1}))(t+1)^{k-1}(1+B_n)^k\sum_{l=0}^{\tau_N(t)}(\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\\
&\qquad+\left\{c_N(\tau_N(t))\exp[\alpha_n(1+O(N^{-1}))(t+1)]+\left(\sum_{s=1}^{\tau_N(t)}c_N(s)^2\right)\frac{1}{2}\alpha_n^2\exp[\alpha_n(1+O(N^{-1}))(\right.\\
&\qquad\qquad\left.+\left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n')]\right\}\alpha_n^k(1+O(N^{-1}))(1+B\\
&\qquad+\exp[\alpha_n(1+B_n')(t+1)]\alpha_n(1+B_n')\alpha_n^k(1+O(N^{-1}))(1+B_n)^k\\
&\qquad\qquad\times\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\sum_{j=1}^{k}c_N(r_j)\prod_{i=1}^{k}c_N(r_i).
\end{aligned}
\tag{5.66}
$$

Upon taking the expectation and limit, we have

$$
\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\left(\prod_{i=1}^{k}p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)}(1-p_r)\right)\right]\\
&\quad\le \alpha_n^k\lim_{N\to\infty}\mathbb{E}\left[\left(1+\sum_{l=1}^{\tau_N(t)-k}(-\alpha_n)^l\frac{1}{l!}t^l\mathbb{1}_{E_N^1\cap E_N^2}\right)\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\right]\\
&\qquad+\lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)}D_N(s)\right]\alpha_n^k(t+1)^{k-1}(1+B_n)^k\exp[\alpha_n t]\\
&\qquad+\left\{\lim_{N\to\infty}\mathbb{E}\left[c_N(\tau_N(t))\right]\exp[\alpha_n(t+1)]+\lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)}c_N(s)^2\right]\frac{1}{2}\alpha_n^2\exp[\alpha_n(t+1)]\right.\\
&\qquad\quad\left.+\lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)}D_N(s)\right]\exp[\alpha_n(t+1)(1+B_n')]\right\}\alpha_n^k(1+B_n)^k\frac{1}{k!}(t+1)^k\\
&\qquad+\exp[\alpha_n(1+B_n')(t+1)]\alpha_n^{k+1}(1+B_n')(1+B_n)^k\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\sum_{j=1}^{k}c_N(r_j)\prod_{i=1}^{k}c_N(r_i)\right]
\end{aligned}
\tag{5.67}
$$

The middle terms vanish due to Brown et al. (2021, Equations (3.3)–(3.5)) and the expression becomes

$$
\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\left(\prod_{i=1}^{k}p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)}(1-p_r)\right)\right]\le\alpha_n^k\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\right]\\
&\qquad+\alpha_n^k\sum_{l=1}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l\lim_{N\to\infty}\mathbb{E}\left[\mathbb{1}_{\{\tau_N(t)\ge k+l\}}\mathbb{1}_{E_N^1\cap E_N^2}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\right]\\
&\qquad+\exp[\alpha_n(1+B_n')(t+1)]\alpha_n^{k+1}(1+B_n')(1+B_n)^k\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\sum_{j=1}^{k}c_N(r_j)\prod_{i=1}^{k}c_N(r_i)\right]
\end{aligned}
\tag{5.68}
$$

where passing the limit and expectation inside the infinite sum is justified by dominated convergence and Fubini; see Lemma 5.16. To simplify the last line,

$$\begin{aligned}
\sum_{\substack{r_1<\dots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \sum_{j=1}^{k} c_N(r_j)\prod_{i=1}^{k} c_N(r_i) &\le \frac{1}{k!}\sum_{r_1\ne\dots\ne r_k}^{\tau_N(t)} \sum_{j=1}^{k} c_N(r_j)\prod_{i=1}^{k} c_N(r_i)\\
&= \frac{1}{k!}\sum_{r_1\ne\dots\ne r_k}^{\tau_N(t)} \sum_{j=1}^{k} c_N(r_j)^2\prod_{i\ne j} c_N(r_i)\\
&\le \frac{1}{k!}\sum_{j=1}^{k}\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \sum_{r_1\ne\dots\ne r_{k-1}}^{\tau_N(t)} \prod_{i=1}^{k-1} c_N(r_i)\\
&\le \frac{1}{(k-1)!}\sum_{s=1}^{\tau_N(t)} c_N(s)^2(t+1)^{k-1}, \qquad (5.69)
\end{aligned}$$

using Lemma 5.3(a) for the final inequality. Hence

$$\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\dots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \sum_{s\in\{r_1,\dots,r_k\}} c_N(s)\prod_{i=1}^{k} c_N(r_i)\right] \le \frac{1}{(k-1)!}(t+1)^{k-1}\lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right]=0 \tag{5.70}$$

by Brown et al. (2021, Equation (3.5)). By Lemmata 5.12, 5.11 and 5.14, $\lim_{N\to\infty}\mathbb{P}[\{\tau_N(t)\ge k+l\}\cap E_N^1\cap E_N^2]=1$, so we can apply Lemma 5.9 to the remaining expectations in (5.68), yielding

$$\begin{aligned}
\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\dots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right] &\le \alpha_n^k\sum_{l=0}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k}\frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}\\
&= \alpha_n^k e^{-\alpha_n t}\sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k}\frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}
\end{aligned} \tag{5.71}$$

as required. ■

Lemma 5.8 (Induction step lower bound)**.** *Fix $k \in \mathbb{N}$, $i_0 := 0$, $i_k := k$. For any sequence of times* $0 = t_0 \le t_1 \le \dots \le t_k \le t$,

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{\substack{r_1<\dots<r_k:\\ r_i \le \tau_N(t_i) \forall i}} \left(\prod_{i=1}^{k} p_{r_i} \right) \left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r) \right) \right] \ge \alpha_n^k e^{-\alpha_n t} \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j \ge j \forall j}} \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!}. \tag{5.72}$$

Proof. Firstly,

$$\sum_{\substack{r_1<\dots<r_k:\\ r_i \le \tau_N(t_i) \forall i}} \left(\prod_{i=1}^{k} p_{r_i} \right) \left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r) \right) \ge \sum_{\substack{r_1<\dots<r_k:\\ r_i \le \tau_N(t_i) \forall i}} \left(\prod_{i=1}^{k} p_{r_i} \right) \left(\prod_{r=1}^{\tau_N(t)} (1-p_r) \right). \tag{5.73}$$

Now the second product does not depend on $r_1, \dots, r_k$, and we can use the lower bound from (5.55):

$$\begin{aligned}\prod_{r=1}^{\tau_N(t)} (1-p_r) \ge \sum_{l=0}^{\tau_N(t)} (-\alpha_n)^l (1+O(N^{-1})) \frac{1}{l!} t^l \mathbb{1}_{E_N^3} - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right) \frac{1}{2} \alpha_n^2 \exp[\alpha_n (1+O(N^{-1}))(t+1)] \\ - c_N(\tau_N(t)) \exp[\alpha_n(1+O(N^{-1}))(t+1)] \\ - \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) \exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n)]\end{aligned} \tag{5.74}$$

where E_N^3 is defined as in (5.52). We will also need an upper bound on this product, which is formed from (5.48) with a further deterministic bound:

$$\begin{aligned}\prod_{r=1}^{\tau_N(t)} (1-p_r) &\le \sum_{l=0}^{\tau_N(t)} (-\alpha_n)^l (1+O(N^{-1})) \frac{1}{l!} t^l \mathbb{1}_{\{\tau_N(t)\ge l\}} \mathbb{1}_{E_N^1 \cap E_N^2} + c_N(\tau_N(t)) \exp[\alpha_n(1+O(N^{-1}))(t+1) \\ &\quad + \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right) \frac{1}{2} \alpha_n^2 \exp[\alpha_n(1+O(N^{-1}))(t+1)] \\ &\quad + \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) \exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n')] \\ &\le \exp[\alpha_n(1+O(N^{-1}))t] + \exp[\alpha_n(1+O(N^{-1}))(t+1)] \\ &\quad + \frac{1}{2}\alpha_n^2 (t+1) \exp[\alpha_n(1+O(N^{-1}))(t+1)] + (t+1)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B \\ &\le \left(2 + \frac{\alpha_n^2 (t+1)}{2} \right) \exp[\alpha_n(1+O(N^{-1}))(t+1)] + (t+1)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+\end{aligned} \tag{5.75}$$

Now let us consider the remaining sum-product on the RHS of (5.73). We use the same bound on p_r as in (5.42):

$$p_r = 1 - p_{\Delta\Delta}(r) \geq \alpha_n(1 + O(N^{-1})) \left[c_N(r) - B'_n D_N(r)\right] \tag{5.76}$$

where the $O(N^{-1})$ term does not depend on r. When N is large enough for the factor of $(1 + O(N^{-1}))$ to be non-negative, the condition that the bound in (5.76) is non-negative holds on the event E_N^2 that was defined in (5.44). Then

$$\prod_{i=1}^{k} p_{r_i} \geq \alpha_n^k (1 + O(N^{-1})) \prod_{i=1}^{k} \left[c_N(r_i) - B'_n D_N(r_i)\right] \mathbb{1}_{E_N^2}. \tag{5.77}$$

Applying a modification of Lemma 5.5 where the sum is over ordered indices rather than distinct indices,

$$\begin{aligned}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\leq\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right) &\geq \alpha_n^k(1+O(N^{-1})) \sum_{\substack{r_1<\cdots<r_k:\\ r_i\leq\tau_N(t_i)\forall i}} \prod_{i=1}^{k} \left[c_N(r_i) - B'_n D_N(r_i)\right] \mathbb{1}_{E_N^2} \\ &\geq \alpha_n^k(1+O(N^{-1})) \left\{ \sum_{\substack{r_1<\cdots<r_k:\\ r_i\leq\tau_N(t_i)\forall i}} \prod_{i=1}^{k} c_N(r_i) \mathbb{1}_{E_N^2} - \frac{1}{k!} \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right) (t+1)^{k-1}(1+ \right.\end{aligned} \tag{5.78}$$

The above expression is already split into positive and negative terms; a lower bound on (5.73) can be formed by multiplying the positive terms by the lower bound (5.74) and the negative terms by the upper bound (5.75). Thus

$$
\begin{aligned}
&\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^k p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right) \\
&\qquad \ge \alpha_n^k(1+O(N^{-1})) \sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^k c_N(r_i)\mathbb{1}_{E_N^2}\Bigg\{\sum_{l=0}^{\tau_N(t)}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}_{E_N^3} \\
&\qquad\qquad - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right)\frac{1}{2}\alpha_n^2\exp[\alpha_n(1+O(N^{-1}))(t+1)] \\
&\qquad\qquad - c_N(\tau_N(t))\exp[\alpha_n(1+O(N^{-1}))(t+1)] \\
&\qquad\qquad - \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n)]\Bigg\} \\
&\qquad - \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right)\alpha_n^k(1+O(N^{-1}))\frac{1}{k!}(t+1)^{k-1}(1+B_n')^k\Bigg\{ \\
&\qquad\qquad \left(2+\frac{\alpha_n^2(t+1)}{2}\right)\exp[\alpha_n(1+O(N^{-1}))(t+1)] \\
&\qquad\qquad + (t+1)\exp[\alpha_n(1+O(N^{-1}))(t+1)(1+B_n')]\Bigg\}.
\end{aligned}
\tag{5.79}
$$

Due to Brown et al. (2021, Equations (3.3)–(3.5)), all but the first line on the RHS of the above have vanishing expectation, leaving

$$
\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^k p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right] \\
&\qquad \ge \lim_{N\to\infty}\mathbb{E}\left[\alpha_n^k(1+O(N^{-1}))\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^k c_N(r_i)\mathbb{1}_{E_N^2}\sum_{l=0}^{\tau_N(t)}(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l\mathbb{1}\right. \\
&\qquad = \alpha_n^k\sum_{l=0}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l\lim_{N\to\infty}\mathbb{E}\left[\mathbb{1}_{\{\tau_N(t)\ge l\}}\mathbb{1}_{E_N^2\cap E_N^3}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^k c_N(r_i)\right].
\end{aligned}
\tag{5.80}
$$

Passing the limit and expectation inside the infinite sum is justified by dominated convergence and Fubini; see Lemma 5.16. Lemmata 5.11 and 5.14 establish that $\lim_{N\to\infty}\mathbb{P}[E_N^2\cap E_N^3]=1$ and Lemma 5.12 deals with the other indicator. We can therefore apply

Lemma 5.9 to conclude that

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right] \ge \alpha_n^k \sum_{l=0}^{\infty}(-\alpha_n)^l \frac{1}{l!}t^l \sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k} \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}$$
$$= \alpha_n^k e^{-\alpha_n t} \sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k} \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!} \tag{5.81}$$

as required. ∎

Lemma 5.9. *Fix $k\in\mathbb{N}$, $i_0 := 0$, $i_k := k$. Let E_N be a sequence of events such that $\lim_{N\to\infty}\mathbb{P}[E_N]=1$. Then for any sequence of times $0=t_0\le t_1\le\cdots\le t_k\le t$,*

$$\lim_{N\to\infty} \mathbb{E}\left[\mathbb{1}_{E_N} \sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^{k} c_N(r_i)\right] = \sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k} \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}. \tag{5.82}$$

Proof. As pointed out by Möhle (1999, p. 460), the sum-product on the left hand side can be expanded as

$$\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^{k} c_N(r_i) = \sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k} \frac{1}{(i_j-i_{j-1})!} \sum_{\substack{r_{i_{j-1}+1}\ne\cdots\ne r_{i_j}\\ =\tau_N(t_{j-1})+1}}^{\tau_N(t_j)} \prod_{i=i_{j-1}+1}^{i_j} c_N(r_i). \tag{5.83}$$

By a modification of the upper bound in Lemma 5.3(b) where the lower limit of the sum is a general time rather than 1,

$$\sum_{\substack{r_{i_{j-1}+1}\ne\cdots\ne r_{i_j}\\ =\tau_N(t_{j-1})+1}}^{\tau_N(t_j)} \prod_{i=i_{j-1}+1}^{i_j} c_N(r_i) \le (t_j-t_{j-1})^{i_j-i_{j-1}} + c_N(\tau_N(t_j))(t_j-t_{j-1}+1)^{i_j-i_{j-1}} \tag{5.84}$$

Now, taking the product on the outside,

$$\begin{aligned}
\prod_{j=1}^{k} \frac{1}{(i_j - i_{j-1})!} \sum_{\substack{r_{i_{j-1}+1} \neq \cdots \neq r_{i_j} \\ = \tau_N(t_{j-1})+1}}^{\tau_N(t_j)} \prod_{i=i_{j-1}+1}^{i_j} c_N(r_i) &\leq \prod_{j=1}^{k} \left\{ \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} + c_N(\tau_N(t_j)) \frac{(t_j - t_{j-1} + 1)^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \right. \\
&\leq \prod_{j=1}^{k} \left\{ \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} + c_N(\tau_N(t_j))(t_j - t_{j-1} + 1)^{i_j - i_{j-1}} \right. \\
&= \sum_{\mathcal{I} \subseteq [k]} \left(\prod_{j \in \mathcal{I}} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \right) \left(\prod_{j \notin \mathcal{I}} c_N(\tau_N(t_j))(t_j - t_{j-1} \right. \\
&= \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \\
&\quad + \sum_{\mathcal{I} \subset [k]} \left(\prod_{j \in \mathcal{I}} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \right) \left(\prod_{j \notin \mathcal{I}} c_N(\tau_N(t_j))(t_j - \right. \\
&\leq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \\
&\quad + \sum_{\mathcal{I} \subset [k]} \left(\prod_{j \in \mathcal{I}} t^{i_j - i_{j-1}} \right) \left(\prod_{j \notin \mathcal{I}} c_N(\tau_N(t_j))(t+1)^{i_j - i_{j-1}} \right) \\
&\leq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} + \sum_{\mathcal{I} \subset [k]} c_N(\tau_N(t_{j^\star(\mathcal{I})})) \prod_{j=1}^{k} (t+1)^{i_j - } \\
&= \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} + \sum_{\mathcal{I} \subset [k]} c_N(\tau_N(t_{j^\star(\mathcal{I})}))(t+1)^k
\end{aligned} \tag{5.85}$$

where, say, $j^\star(\mathcal{I}) := \min\{j \notin \mathcal{I}\}$. Now we are in a position to evaluate the limit in (5.82):

$$\begin{aligned}
\lim_{N \to \infty} \mathbb{E} \left[\mathbb{1}_{E_N} \sum_{\substack{r_1 < \cdots < r_k: \\ r_i \leq \tau_N(t_i) \forall i}} \prod_{i=1}^{k} c_N(r_i) \right] &\leq \lim_{N \to \infty} \mathbb{E} \left[\sum_{\substack{r_1 < \cdots < r_k: \\ r_i \leq \tau_N(t_i) \forall i}} \prod_{i=1}^{k} c_N(r_i) \right] \\
&\leq \sum_{\substack{i_1 \leq \cdots \leq i_{k-1} \\ \in \{0, \ldots, k\}: \\ i_j \geq j \forall j}} \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} + \sum_{\substack{i_1 \leq \cdots \leq i_{k-1} \\ \in \{0, \ldots, k\}: \\ i_j \geq j \forall j}} \sum_{\mathcal{I} \subset [k]} \lim_{N \to \infty} \mathbb{E} \left[c_N(\tau_N(t_{j^\star(\mathcal{I})})) \right] (t+1)^k \\
&= \sum_{\substack{i_1 \leq \cdots \leq i_{k-1} \\ \in \{0, \ldots, k\}: \\ i_j \geq j \forall j}} \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!}
\end{aligned} \tag{5.86}$$

using Brown et al. (2021, Equation (3.3)).

For the corresponding lower bound, by a modification of the lower bound in Lemma 5.3(b) where the lower limit of the sum is a general time rather than 1,

$$\begin{aligned}\sum_{\substack{r_{i_{j-1}+1}\neq\cdots\neq r_{i_j}\\ =\tau_N(t_{j-1})+1}}^{\tau_N(t_j)} \prod_{i=i_{j-1}+1}^{i_j} c_N(r_i) &\geq (t_j - t_{j-1})^{i_j - i_{j-1}} - \binom{i_j - i_{j-1}}{2} \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) (t_j - t_{j-1} + 1)^{i_j-} \\ &\geq (t_j - t_{j-1})^{i_j - i_{j-1}} - (i_j - i_{j-1})! \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) (t_j - t_{j-1} + 1)^{i_j-}\end{aligned} \tag{5.87}$$

Define the events

$$E_N^4(j) = \left\{ \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) \leq \frac{1}{(i_j - i_{j-1})!} \left(\frac{t_j - t_{j-1}}{t_j - t_{j-1} + 1} \right)^{i_j - i_{j-1}} \right\}, \tag{5.88}$$

which is sufficient to ensure the j^{th} term in the following product is non-negative, and define $E_N^4 := \bigcap_{j=1}^k E_N^4(j)$. (If $t_j = t_{j-1}$ then $E_N^4(j)$ has probability one automatically; otherwise the constant on the right is strictly positive and so satisfies the conditions of

Lemma 5.13.) Now, taking a product over j,

$$
\begin{aligned}
&\prod_{j=1}^{k} \frac{1}{(i_j - i_{j-1})!} \sum_{\substack{r_{i_{j-1}+1} \neq \cdots \neq r_{i_j} \\ = \tau_N(t_{j-1})+1}}^{\tau_N(t_j)} \prod_{i=i_{j-1}+1}^{i_j} c_N(r_i) \\
&\geq \prod_{j=1}^{k} \left\{ \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} - \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) (t_j - t_{j-1} + 1)^{i_j - i_{j-1} - 2} \right\} \mathbb{1}_{E_N^4} \\
&= \sum_{\mathcal{I} \subseteq [k]} (-1)^{k - |\mathcal{I}|} \left(\prod_{j \in \mathcal{I}} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \right) \left(\prod_{j \notin \mathcal{I}} \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) (t_j - t_{j-1} + 1)^{i_j \cdots} \right. \\
&= \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} \\
&\quad + \sum_{\mathcal{I} \subset [k]} (-1)^{k - |\mathcal{I}|} \left(\prod_{j \in \mathcal{I}} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \right) \left(\prod_{j \notin \mathcal{I}} \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) (t_j - t_{j-1} + \cdots \right. \\
&\geq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} \\
&\quad - \sum_{\mathcal{I} \subset [k]} \left(\prod_{j \in \mathcal{I}} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \right) \left(\prod_{j \notin \mathcal{I}} \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) (t_j - t_{j-1} + 1)^{i_j - i_{j-1} \cdots} \right. \\
&\geq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} \\
&\quad - \sum_{\mathcal{I} \subset [k]} \left(\prod_{j \in \mathcal{I}} t^{i_j - i_{j-1}} \right) \left(\prod_{j \notin \mathcal{I}} \left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right) (t+1)^{i_j - i_{j-1} - 2} \right) \\
&\geq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} \\
&\quad - \sum_{\mathcal{I} \subset [k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2 \right) \left(\prod_{j \in \mathcal{I}} t^{i_j - i_{j-1}} \right) \left(\prod_{j \notin \mathcal{I}} (t+1)^{i_j - i_{j-1} - 1} \right) \\
&\geq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} - \sum_{\mathcal{I} \subset [k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2 \right) \prod_{j=1}^{k} (t+1)^{i_j - i_{j-1}} \\
&= \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} - \sum_{\mathcal{I} \subset [k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2 \right) (t+1)^k,
\end{aligned}
\tag{5.89}
$$

where again we have arbitrarily set $j^\star(\mathcal{I}) := \min\{j \notin \mathcal{I}\}$. We can now evaluate the limit:

$$
\begin{aligned}
\lim_{N\to\infty} \mathbb{E}\left[\mathbb{1}_{E_N} \sum_{\substack{r_1<\dots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^k c_N(r_i)\right] &\ge \lim_{N\to\infty} \mathbb{E}\left[\mathbb{1}_{E_N\cap E_N^4} \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^k \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}\right] \\
&\quad - \lim_{N\to\infty} \mathbb{E}\left[\mathbb{1}_{E_N} \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \sum_{\mathcal{I}\subset[k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2\right)(t+1
\right. \\
&\ge \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^k \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!} \lim_{N\to\infty} \mathbb{E}[\mathbb{1}_{E_N\cap E_N^4}] \\
&\quad - \lim_{N\to\infty} \mathbb{E}\left[\sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \sum_{\mathcal{I}\subset[k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2\right)(t+1)^k\right] \\
&= \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^k \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!} \lim_{N\to\infty} \mathbb{P}[E_N\cap E_N^4] \\
&\quad - \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \sum_{\mathcal{I}\subset[k]} \lim_{N\to\infty} \mathbb{E}\left[\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2\right](t+1)^k \\
&= \sum_{\substack{i_1\le\dots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^k \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!} \qquad (5.90)
\end{aligned}
$$

where for the last equality we use Brown et al. (2021, Equation (3.5)) to show that the second sum vanishes and Lemma 5.13 to show that $\lim_{N\to\infty}\mathbb{P}[E_N\cap E_N^4]=1$. We have shown that the upper and lower bounds coincide, so the result follows. ∎

5.3 Indicators

Lemma 5.10. *Let $(A_N),(B_N)$ be sequences of events. If $\lim_{N\to\infty}\mathbb{P}[A_N]=1$ and $\lim_{N\to\infty}\mathbb{P}[B_N]=1$ then $\lim_{N\to\infty}\mathbb{P}[A_N\cap B_N]=1$.*

The above might be so obvious as to go unstated, but it is very important because it means we don't have to deal with intersections of dependent events! Here is a little proof just to be sure:

Proof.

$$\begin{aligned}
&\lim_{N\to\infty} \mathbb{P}[A_N] = 1 \textbf{ and } \lim_{N\to\infty} \mathbb{P}[B_N] = 1 \\
\Leftrightarrow &\lim_{N\to\infty} \mathbb{P}[A_N^c] = 0 \textbf{ and } \lim_{N\to\infty} \mathbb{P}[B_N^c] = 0 \\
\Rightarrow &\lim_{N\to\infty} \{\mathbb{P}[A_N^c] + \mathbb{P}[B_N^c]\} = 0 \\
\Rightarrow &\lim_{N\to\infty} \mathbb{P}[A_N^c \cup B_N^c] = 0 \\
\Leftrightarrow &\lim_{N\to\infty} \mathbb{P}[A_N \cap B_N] = 1. \qquad (5.91)
\end{aligned}$$

The only part of this argument that I find potentially controversial is going from the third to the fourth line, which is an application of the sandwich theorem (since $0 \leq \mathbb{P}[A_N^c \cup B_N^c] \leq \mathbb{P}[A_N^c] + \mathbb{P}[B_N^c]$). ∎

Lemma 5.11. *Let $K > 0$ be a constant which may depend on n, N but not on r, such that $K^{-2} = O(1)$ as $N \to \infty$. Define the events $E_N(r) := \{c_N(r) < K\}$ and denote $E_N := \bigcap_{r=1}^{\tau_N(t)} E_N(r)$. Then $\lim_{N\to\infty} \mathbb{P}[E_N] = 1$.*

Proof.

$$\begin{aligned}
\mathbb{P}[E_N] &= 1 - \mathbb{P}[E_N^c] = 1 - \mathbb{P}\left[\bigcup_{r=1}^{\tau_N(t)} E_N^c(r)\right] = 1 - \mathbb{E}\left[\mathbb{1}_{\bigcup E_N^c(r)}\right] \geq 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{1}_{E_N^c(r)}\right] \\
&= 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{E}\left[\mathbb{1}_{E_N^c(r)} \mid \mathcal{F}_{r-1}\right]\right] = 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{P}\left[E_N^c(r) \mid \mathcal{F}_{r-1}\right]\right] \qquad (5.92)
\end{aligned}$$

where for the second line we apply Lemma 5.15 with $f(r) = \mathbb{1}_{E_N^c(r)}$. By the generalised Markov inequality,

$$\mathbb{P}[E_N^c(r) \mid \mathcal{F}_{r-1}] = \mathbb{P}[c_N(r) \geq K \mid \mathcal{F}_{r-1}] \leq K^{-2}\mathbb{E}[c_N(r)^2 \mid \mathcal{F}_{r-1}]. \qquad (5.93)$$

Substituting this into (5.92) and applying Lemma 5.15 again, this time with $f(r) = c_N(r)^2$,

$$\mathbb{P}[E_N] \geq 1 - K^{-2}\mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{E}[c_N(r)^2 \mid \mathcal{F}_{r-1}]\right] = 1 - K^{-2}\mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} c_N(r)^2\right]. \qquad (5.94)$$

Applying Brown et al. (2021, Equation (3.5)), the limit is

$$\lim_{N\to\infty} \mathbb{P}[E_N] = 1 - O(1) \times 0 = 1 \qquad (5.95)$$

as required. ∎

Lemma 5.12. *Fix $t > 0$. For any $l \in \mathbb{N}$, $\lim_{N\to\infty} \mathbb{P}[\tau_N(t) \geq l] = 1$.*

Proof. We can replace the event $\{\tau_N(t) \geq l\}$ with an event of the form of E_N in Lemma 5.11:

$$\{\tau_N(t) \geq l\} = \left\{ \min\left\{ s \geq 1 : \sum_{r=1}^{s} c_N(r) \geq t \right\} \geq l \right\} = \left\{ \sum_{r=1}^{l-1} c_N(r) < t \right\} \supseteq \bigcap_{r=1}^{l-1} \left\{ c_N(r) < \frac{t}{l} \right\} \supseteq \bigcap_{r=1}^{\tau_N(t)} \left\{ c_N \right. \tag{5.96}$$

Hence

$$\lim_{N\to\infty} \mathbb{P}[\tau_N(t) \geq l] \geq \lim_{N\to\infty} \mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \left\{ c_N(r) < \frac{t}{l} \right\} \right] = 1 \tag{5.97}$$

by applying Lemma 5.11 with $K = t/l$. ■

Lemma 5.13. *Fix $k \in \mathbb{N}$, a sequence of times $0 = t_0 \leq t_1 \leq \cdots \leq t_k \leq t$, and let $K_1, \ldots, K_k$ be strictly positive constants. Define the events*

$$E_N := \bigcap_{j=1}^{k} \left\{ \sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \leq K_j \right\}. \tag{5.98}$$

Then $\lim_{N\to\infty} \mathbb{P}[E_N] = 1$.

Proof.

$$\mathbb{P}[E_N] = 1 - \mathbb{P}[E_N^c] = 1 - \mathbb{P}\left[\bigcup_{j=1}^{k} \left\{ \sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 > K_j \right\} \right] \geq 1 - \sum_{j=1}^{k} \mathbb{P}\left[\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \geq K_j \right]. \tag{5.99}$$

Applying Markov's inequality,

$$\mathbb{P}[E_N] \geq 1 - \sum_{j=1}^{k} K_j^{-1} \mathbb{E}\left[\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)} c_N(s)^2 \right] \underset{N\to\infty}{\longrightarrow} 1 - \sum_{j=1}^{k} O(1) \times 0 = 1 \tag{5.100}$$

by Brown et al. (2021, Equation (3.5)). ■

Lemma 5.14. *Fix $t > 0$. Let K be a constant not depending on N, r, but which may depend on n.*

$$\lim_{N\to\infty} \mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \{c_N(r) \geq K D_N(r)\} \right] = 1. \tag{5.101}$$

Proof.

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \{c_N(r) \geq KD_N(r)\}\right] &\geq \mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \{c_N(r) > KD_N(r)\}\right] \\
&= 1 - \mathbb{P}\left[\bigcup_{r=1}^{\tau_N(t)} \{c_N(r) \leq KD_N(r)\}\right] \\
&= 1 - \mathbb{E}\left[\mathbb{1}_{\bigcup\{c_N(r)\leq KD_N(r)\}}\right] \\
&\geq 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{1}_{\{c_N(r)\leq KD_N(r)\}}\right] \\
&= 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{P}[c_N(r) \leq KD_N(r) \mid \mathcal{F}_{r-1}]\right] \qquad (5.102)
\end{aligned}$$

where the final inequality is an application of Lemma 5.15 with $f(r) = \mathbb{1}_{\{c_N(r)\leq KD_N(r)\}}$.

Fix $0 < \varepsilon < K^{-1}/2$ and assume $N > \max\{\varepsilon^{-1}, (K^{-1} - 2\varepsilon)^{-1}\}$. For each r, i define the event $A_i(r) := \{\nu_r^{(i)} \leq N\varepsilon\}$. Conditional on $\mathcal{F}_{r-1}$, we have

$$D_N(r) = \frac{1}{N(N)_2}\sum_{i=1}^N (\nu_r^{(j)})_2 \left[\nu_r^{(i)} + \frac{1}{N}\sum_{j\neq i}(\nu_r^{(i)})^2\right]\mathbb{1}_{A_i^c(r)} + \frac{1}{N(N)_2}\sum_{i=1}^N (\nu_r^{(i)})_2 \left[\nu_r^{(i)} + \frac{1}{N}\sum_{j\neq i}(\nu_r^{(j)})^2\right]\mathbb{1}_{A_i(r)} \qquad (5.103)$$

For the first term,

$$\frac{1}{N(N)_2}\sum_{i=1}^N (\nu_r^{(i)})_2 \left[\nu_r^{(i)} + \frac{1}{N}\sum_{j\neq i}(\nu_r^{(j)})^2\right]\mathbb{1}_{A_i^c(r)} \leq \sum_{i=1}^N \mathbb{1}_{A_i^c(r)}. \qquad (5.104)$$

For the second term,

$$\begin{aligned}
\frac{1}{N(N)_2}\sum_{i=1}^N (\nu_r^{(i)})_2 \left[\nu_r^{(i)} + \frac{1}{N}\sum_{j\neq i}(\nu_r^{(j)})^2\right]\mathbb{1}_{A_i(r)} &\leq \frac{1}{N(N)_2}\sum_{i=1}^N (\nu_r^{(i)})_2 \nu_r^{(i)}\mathbb{1}_{A_i(r)} + \frac{1}{N^2(N)_2}\sum_{i=1}^N (\nu_r^{(i)})_2 \sum_{j=1}^N (\nu \\
&\leq \frac{1}{N}c_N(r)N\varepsilon + \frac{1}{N^2(N)_2}\sum_{i=1}^N (\nu_r^{(i)})_2 \sum_{j=1}^N (\nu_r^{(j)})_2 \mathbb{1}_{A_i(r)} \\
&\quad + \frac{1}{N^2(N)_2}\sum_{i=1}^N (\nu_r^{(i)})_2 \sum_{j=1}^N (\nu_r^{(j)})\mathbb{1}_{A_i(r)} \\
&\leq \varepsilon c_N(r) + \frac{1}{N^2}\sum_{i=1}^N \nu_r^{(i)} N\varepsilon c_N(r) + \frac{1}{N^2}c_N(r)N \\
&= c_N(r)\left(2\varepsilon + \frac{1}{N}\right). \qquad (5.105)
\end{aligned}$$

Altogether we have

$$D_N(r) \le c_N(r)\left(2\varepsilon + \frac{1}{N}\right) + \sum_{i=1}^{N} \mathbb{1}_{A_i^c(r)}. \tag{5.106}$$

Hence, still conditional on $\mathcal{F}_{r-1}$,

$$\begin{aligned} \{c_N(r) \le K D_N(r)\} &\subseteq \left\{ c_N(r) \le K c_N(r)(2\varepsilon + N^{-1}) + K \sum_{i=1}^{N} \mathbb{1}_{A_i^c(r)} \right\} \\ &= \left\{ K^{-1} - 2\varepsilon - \frac{1}{N} \le \sum_{i=1}^{N} \frac{\mathbb{1}_{A_i^c(r)}}{c_N(r)} \right\} \end{aligned} \tag{5.107}$$

where the ratio $\mathbb{1}_{A_i^c(r)}/c_N(r)$ is well-defined because

$$A_i^c(r) \Rightarrow c_N(r) := \frac{1}{(N)_2}\sum_{j=1}^{N}(\nu_r^{(j)})_2 \ge \frac{1}{(N)_2}(\nu_r^{(i)})_2 \ge \frac{\varepsilon(N\varepsilon - 1)}{N-1} \ge \varepsilon\left(\varepsilon - \frac{1}{N}\right) > 0. \tag{5.108}$$

Hence by Markov's inequality (the conditions on ε, N ensuring the constant is always strictly positive),

$$\begin{aligned} \mathbb{P}\left[c_N(r) \le K D_N(r) \mid \mathcal{F}_{r-1}\right] &\le \mathbb{P}\left[\sum_{i=1}^{N} \mathbb{1}_{A_i^c(r)} \ge \left(K^{-1} - 2\varepsilon - \frac{1}{N}\right)\varepsilon\left(\varepsilon - \frac{1}{N}\right) \middle| \mathcal{F}_{r-1} \right] \\ &\le \frac{1}{\left(K^{-1} - 2\varepsilon - \frac{1}{N}\right)\varepsilon\left(\varepsilon - \frac{1}{N}\right)} \mathbb{E}\left[\sum_{i=1}^{N} \mathbb{1}_{A_i^c(r)} \middle| \mathcal{F}_{r-1} \right] \\ &\le \frac{1}{\left(K^{-1} - 2\varepsilon - \frac{1}{N}\right)\varepsilon\left(\varepsilon - \frac{1}{N}\right)} \mathbb{E}\left[\sum_{i=1}^{N} \frac{(\nu_r^{(i)})_3}{(N\varepsilon)_3} \middle| \mathcal{F}_{r-1} \right] \\ &\le \frac{1}{\left(K^{-1} - 2\varepsilon - \frac{1}{N}\right)\varepsilon\left(\varepsilon - \frac{1}{N}\right)} \mathbb{E}\left[\frac{N(N)_2}{(N\varepsilon)_3} D_N(r) \middle| \mathcal{F}_{r-1} \right]. \end{aligned} \tag{5.109}$$

Applying Lemma 5.15 once more, with $f(r) = D_N(r)$,

$$\begin{aligned} \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{P}[c_N(r) \le K D_N(r) \mid \mathcal{F}_{r-1}] \right] &\le \frac{1}{\left(K^{-1} - 2\varepsilon - \frac{1}{N}\right)\varepsilon\left(\varepsilon - \frac{1}{N}\right)} \frac{N(N)_2}{(N\varepsilon)_3} \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{E}[D_N(r) \mid \mathcal{F}_{r-1}] \right] \\ &= \frac{1}{\left(K^{-1} - 2\varepsilon - \frac{1}{N}\right)\varepsilon\left(\varepsilon - \frac{1}{N}\right)} \frac{N(N)_2}{(N\varepsilon)_3} \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} D_N(r) \right] \\ &\xrightarrow[N\to\infty]{} \frac{1}{(K^{-1} - 2\varepsilon)\varepsilon^5} \times 0 = 0. \end{aligned} \tag{5.110}$$

Substituting this back into (5.102) concludes the proof. ■

5.4 Other useful results

The following Lemma is taken from Koskela et al. (2018, Lemma 2), where the function is set to $f(r) = c_N(r)$, but the authors remark that the result holds for other choices of function.

Lemma 5.15. *Fix $t > 0$. Let $(\mathcal{F}_r)$ be the backwards-in-time filtration generated by the offspring counts $\nu_r^{(1:N)}$ at each generation r, and let $f(r)$ be any deterministic function of $\nu_r^{(1:N)}$ that is non-negative and bounded. In particular, for all r there exists $B < \infty$ such that $0 \leq f(r) \leq B$. Then*

$$\mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} f(r)\right] = \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\right]. \tag{5.111}$$

Proof. Define

$$M_s := \sum_{r=1}^{s} \{f(r) - \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\}. \tag{5.112}$$

It is easy to establish that (M_s) is a martingale with respect to $(\mathcal{F}_s)$, and $M_0 = 0$. Now fix $K \geq 1$ and note that $\tau_N(t) \wedge K$ is a bounded $\mathcal{F}_t$-stopping time. Hence we can apply the optional stopping theorem:

$$\mathbb{E}[M_{\tau_N(t)\wedge K}] = \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} \{f(r) - \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\}\right] = \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} f(r)\right] - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\right] = \tag{5.113}$$

Since this holds for all $K \geq 1$,

$$\lim_{K\to\infty} \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} f(r)\right] = \lim_{K\to\infty} \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\right]. \tag{5.114}$$

The monotone convergence theorem allows the limit to pass inside the expectation on each side (since increasing K can only increase each sum, by possibly adding non-negative terms). Hence

$$\mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} f(r)\right] = \mathbb{E}\left[\lim_{K\to\infty} \sum_{r=1}^{\tau_N(t)\wedge K} f(r)\right] = \mathbb{E}\left[\lim_{K\to\infty} \sum_{r=1}^{\tau_N(t)\wedge K} \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\right] = \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\right] \tag{5.115}$$

which concludes the proof. ■

There are a few instances where Fubini's Theorem and the Dominated Convergence Theorem are needed in order to pass a limit and expectation through an infinite sum. Now we verify that the conditions of these theorems indeed hold. This result, analogous to that in Koskela et al. (2018, Appendix), is used once in Lemma 5.7 at (5.67) and once in Lemma 5.8 at (5.80).

Lemma 5.16. *For any fixed $t > 0$,*

$$\mathbb{E}\left[\sum_{l=0}^{\infty}\left|(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l \sum_{\substack{r_1<\cdots<r_k:\\ r_i\leq\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\right|\right] < \infty. \tag{5.116}$$

Proof.

$$\begin{aligned}\mathbb{E}\left[\sum_{l=0}^{\infty}\left|(-\alpha_n)^l(1+O(N^{-1}))\frac{1}{l!}t^l \sum_{\substack{r_1<\cdots<r_k:\\ r_i\leq\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\right|\right] &\leq \mathbb{E}\left[\sum_{l=0}^{\infty}\alpha_n^l(1+O(N^{-1}))\frac{1}{l!}t^l(t+1)^k\right] \\ &= \mathbb{E}[\exp\{\alpha_n t(1+O(N^{-1}))\}(t+1)^k] = \exp\{\alpha_n t(1+O(N^{-1}))\}(t+1)^k\end{aligned} \tag{5.117}$$

■

5.5 Dependency graph

Missing links since this graph was updated:

- Lemma 5.3(a) is used three times in Lemma 5.7, but not anywhere else.
- Lemma 5.3 in the current dependency graph is really referring to Lemma 5.3(b)
- Lemma 5.16 is used in Lemmata 5.8 and 5.7.
- Lemma 5.10 is used in Lemmata 5.6, 5.7, 5.8 and 5.9

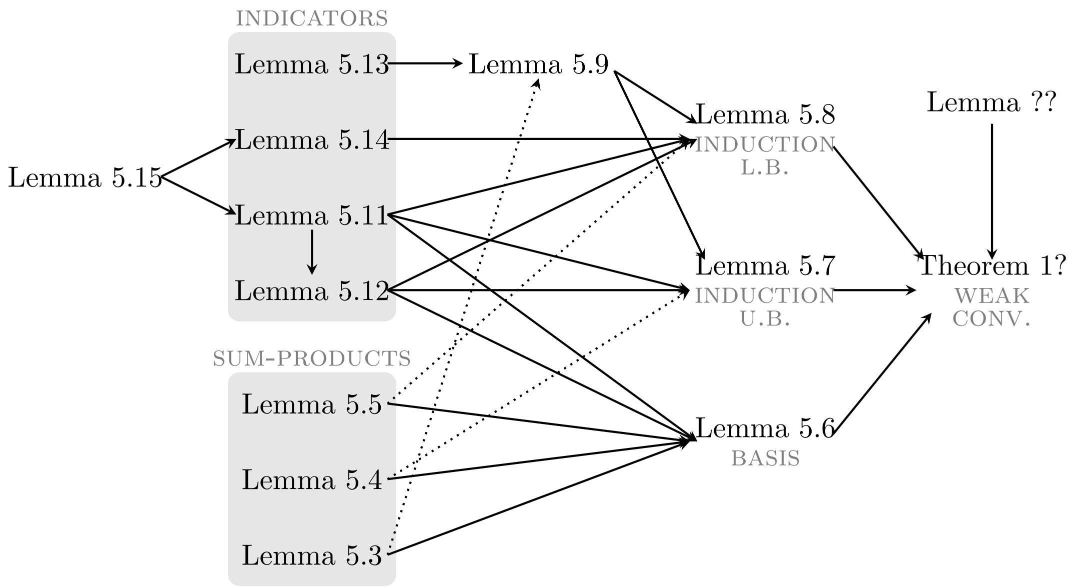

Figure 5.1: Graph showing dependencies between the lemmata used to prove weak convergence. Dotted arrows indicate dependence via a slight modification of the preceding lemma.

6 Discussion

Bibliography

[1] Leonard E. Baum et al. "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains". In: *The Annals of Mathematical Statistics* 41 (1970), pp. 164–171.

[2] Nathanaël Berestycki. *Recent Progress in Coalescent Theory.* 0909.3985. ArXiv, 2009.

[3] Suzie Brown et al. "Simple Conditions for Convergence of Sequential Monte Carlo Genealogies with Applications". In: *Electronic Journal of Probability* 26.1 (2021), pp. 1–22. ISSN: 1083-6489. DOI: 10.1214/20-EJP561.

[4] C. Cannings. "The Latent Roots of Certain Markov Chains Arising in Genetics: A New Approach, I. Haploid Models". In: *Advances in Applied Probability* 6.2 (1974), pp. 260–290.

[5] C. Cannings. "The Latent Roots of Certain Markov Chains Arising in Genetics: A New Approach, II. Further Haploid Models". In: *Advances in Applied Probability* 7.2 (1975), pp. 264–282.

[6] Richard Durrett. *Probability Models for DNA Sequence Evolution.* Springer Science & Business Media, 2008.

[7] Rick Durrett. *Probability: Theory and Examples.* 5th ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2019. DOI: 10.1017/9781108591034.

[8] Alison Etheridge. *Some Mathematical Models from Population Genetics: École D'Été de Probabilités de Saint-Flour XXXIX-2009.* Springer, 2011.

[9] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence.* John Wiley & Sons, 2009.

[10] Ronald Aylmer Fisher. "On the Dominance Ratio". In: *Proceedings of the Royal Society of Edinburgh* 42 (1923), pp. 321–341.

[11] Ronald Aylmer Fisher. "The Distribution of Gene Ratios for Rare Mutations". In: *Proceedings of the Royal Society of Edinburgh* 50 (1930), pp. 205–220.

[12] Y.-X. Fu. "Exact Coalescent for the Wright–Fisher Model". In: *Theoretical Population Biology* 69 (4 2006), pp. 385–394.

[13] Godfrey Harold Hardy and Srinivasa Aaiyangar Ramanujan. "Asymptotic Formulae in Combinatory Analysis". In: *Proceedings of the London Mathematical Society* s2-17.1 (1918), pp. 75–115.

[14] Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory.* Courier Corporation, 2007.

[15] Rudolph Emil Kalman. "A New Approach to Linear Filtering and Prediction Problems". In: *Journal of Basic Engineering* 82.1 (1960), pp. 35–45.

[16] John F. C. Kingman. "Exchangeability and the Evolution of Large Populations". In: *Proceedings of the International Conference on Exchangeability in Probability and Statistics, Rome, 6th-9th April, 1981, in Honour of Professor Bruno de Finetti.* North-Holland, Amsterdam, 1982.

[17] John F. C. Kingman. "On the Genealogy of Large Populations". In: *Journal of Applied Probability* 19.A (1982), pp. 27–43.

[18] John F. C. Kingman. "The Coalescent". In: *Stochastic Processes and Their Applications* 13.3 (1982), pp. 235–248.

[19] Jere Koskela et al. *Asymptotic genealogies of interacting particle systems with an application to sequential Monte Carlo.* Mathematics e-print 1804.01811. ArXiv, 2018.

[20] Martin Möhle. "Weak Convergence to the Coalescent in Neutral Population Models". In: *Journal of Applied Probability* 36.2 (1999), pp. 446–460.

[21] Patrick Alfred Pierce Moran. "Random Processes in Genetics". In: *Mathematical Proceedings of the Cambridge Philosophical Society.* Vol. 54. 1. Cambridge University Press, 1958, pp. 60–71.

[22] James E. Mosimann. "On the Compound Multinomial Distribution, the Multivariate β-Distribution, and Correlations among Proportions". In: *Biometrika* 49.1/2 (1962), pp. 65–82.

[23] Herbert E. Rauch, C. T. Striebel, and F. Tung. "Maximum Likelihood Estimates of Linear Dynamic Systems". In: *AIAA Journal* 3.8 (1965), pp. 1445–1450.

[24] Thomas Verma and Judea Pearl. "Causal Networks: Semantics and Expressiveness". In: *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence.* Minneapolis, MN, Mountain View, CA, 1988, pp. 352–359.

[25] Paolo Vidoni. "Exponential Family State Space Models Based on a Conjugate Latent Process". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.1 (1999), pp. 213–221.

[26] John Wakeley. *Coalescent Theory: An Introduction.* Roberts & Co. Publishers, 2009.

[27] Eric A. Wan and Rudolph van der Merwe. "The Unscented Kalman Filter for Nonlinear Estimation". In: *Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium.* IEEE. 2000, pp. 153–158.

[28] Sewall Wright. "Evolution in Mendelian Populations". In: *Genetics* 16.2 (1931), pp. 97–159.