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

List of Figures

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:

Abstract

List of Abbreviations

SMC	sequential Monte Carlo
i.i.d.	independent and identically distributed
MRCA	most recent common ancestor
PRNG	pseudo-random number generator
CDF	cumulative distribution function
LHS	left hand side
RHS	right hand side

Notation and Conventions

Also include indexing notation $X_{a:b}$, X_{-a}, and X_A where A is a set of indices. And big-O notation. And $\mathbb{Z}$ and $\mathbb{R}$?

$\mathbb{N}$	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 $a \in \mathbb{N}$ also allow $a = 0$ in which case $[a] = \emptyset$?
$\mathcal{S}_k$	the k-dimensional unit simplex $\{x_{1:k+1} \geq 0 : \sum_{i=1}^{k+1} x_i = 1\}$
$(a)_b$	the falling factorial $a(a-1)\cdots(a-b+1)$ where $a \in \mathbb{N}_0, b \in \mathbb{N}$, and define $(a)_0 = 1$. could even allow $a \in \mathbb{R}$ but I don't think I ever use it in that setting
$\binom{a}{b}$	binomial coefficient where $a, b \in \mathbb{N}_0$, defined to be 0 when $a < b$
$\prod_\emptyset$	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 $\mathbb{E}[\cdot \mid \mathcal{F}_{t-1}]$
Var	variance
Cov	covariance
A^c	the complement of set A
$\lvert A \rvert$	the cardinality of set A
1_N	asymptotic notation for a function that converges to 1 as $N \to \infty$

1 Introduction

2 Background

⟨ch:bg⟩

> 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

The idea of Monte Carlo is to use (pseudo-)random numbers to approximate expectations under an intractable probability distribution of interest. Sequential Monte Carlo (SMC) is a class of Monte Carlo algorithms which are implemented sequentially, allowing efficient sampling from *sequences* of distributions. SMC was developed for inference in intractable state space models (details in Section ??) and introduced to the statistics community by Gordon, Salmond, and Smith (1993). The basic idea behind SMC is sequential importance sampling, whereby the posterior importance samples from one target distribution are used to generate proposal samples for the next. A full derivation of the SMC recursions is beyond the scope of this work, but the reader is referred to [citations] for more background. Here it will suffice to include a motivation in the context of state space models (Section ??) followed by the formalism of Feynman-Kac models (Section ??).

2.1.1 State space models ✓

State space models (sometimes called hidden Markov models) are a flexible class of statistical models which are suitable in all sorts of applications where observations appear sequentially. The general model has two components: a Markov process $(X_t)_{t\in\mathbb{N}_0}$ representing the (unobservable) underlying state of the system, and a sequence $(Y_t)_{t\in\mathbb{N}_0}$ of noisy observations of the underlying state. The model is characterised by its conditional independence structure (Figure 2.1) along with an initial distribution μ, the Markov "transition" kernels $(K_t)_{t\in\mathbb{N}}$ and the "emission" distributions $(g_t)_{t\in\mathbb{N}_0}$. Written as a hierarchical

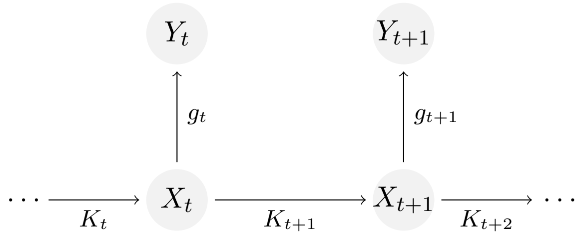


Figure 2.1: Graphical depiction of a general state space model. (X_t) is a Markov process with transition kernels (K_t) representing the underlying state of the system. Y_t is a noisy observation of X_t for each t.

⟨fig:SSM⟩

model,

$$
\begin{aligned}
X_0 &\sim \mu(\cdot) && \\
X_{t+1} \mid X_t &\sim K_{t+1}(\cdot|X_t) && \text{for } t = 0, 1, \ldots \\
Y_t \mid X_t &\sim g_t(\cdot|X_t) && \text{for } t = 0, 1, \ldots
\end{aligned}
$$

Here X and/or Y may be multivariate, observation times need not be equally spaced, and straightforward generalisations of the stated model can allow for situations in which observations are not available as often as the state is updated (up to and including the extreme where the state is a continuous-time Markov process but the observations are available only at discrete times), or on the other hand where observations are observed more frequently than the state is updated.

Applications include: target tracking, where X is the true position of some object and Y encodes some measurements from sensors e.g. radar; stochastic volatility, where X is the volatility and Y is the observed value e.g. the price of a stock; change-point detection; and pretty much any other application in which there is an observed time series from which one would like to infer underlying states.

The principal inferences of interest in state space models are:

filtering inferring the current state x_t from the observations up to now $y_{0:t}$

prediction inferring a future state x_{t+h} from the observations up to now $y_{0:t}$

smoothing inferring the sequence of states up to now $x_{0:t}$ from the observations up to now $y_{0:t}$

fixed-lag smoothing inferring the last h states $x_{t-h:t}$ from the observations up to now $y_{0:t}$

If the dynamics of the state space model are parametrised by some θ, i.e. g_t, K_t depend on θ, we may also be interested in parameter inference and/or computing the likelihood of the observed data given a certain value of θ.

In certain cases, these inference problems may be solved analytically (Section ??), but this is not typically the case. For intractable models we must resort to Monte Carlo methods. However, state space inference is problematic even with Monte Carlo. The main difficulties are that the dimension of the target distribution increases along the sequence, and there is strong dependence between consecutive distributions. Markov chain Monte Carlo (MCMC), for instance, is known to struggle with highly correlated targets [citation] and its performance drops drastically as dimension increases, despite convergence rates that are supposedly independent of dimension [citation].

As we will see in Section ??, sequential Monte Carlo overcomes these problems, turning the problematic properties of the target distribution to its benefit. Correlation between consecutive targets is exploited for sequential updating, which takes in its stride the incrementing dimensionality. The resulting linear-in-t computational complexity also allows inference to be performed on-line, that is, as observations arrive.

2.1.2 Exact inference in state space models ✓

If the state space model has linear dynamics with Gaussian errors, the posterior distributions of interest are also Gaussian with mean and covariance satisying recursions, implemented by the Kalman filter (Kalman 1960) and Rauch-Tung-Striebel smoother (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. However, if the state space becomes large (albeit finite), exact computation becomes infeasible.

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 approximation 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 high-dimensional or highly non-linear or for which gradients are not readily available, the exact Kalman filter updates can be replaced by sample approximations. The *ensemble Kalman filter* (Evensen 1994) uses a Monte Carlo sample from the current time, propagates these points through the transition dynamics, and uses the sample covariance as an estimator of the updated covariance matrix. The means (which are cheaper to evaluate and more stable than the covariances) are still updated using the Kalman filter recursion, based on the estimated covariance. The *unscented Kalman filter* (Wan and Merwe 2000) uses a deterministic sample chosen via the *unscented transformation*, which is then propagated through the non-linear transition to obtain a characterisation of the distribution at the next time step. The sample consists of $2d + 1$ points, where d is the dimension of the state space, and is a sufficient characterisation of a Gaussian distribution. The sample points define a Gaussian approximation to the updated distribution.

In complex or high-dimensional models, such techniques may not be feasible, in which case we must resort to Monte Carlo methods. Markov chain Monte Carlo performs woefully on state space models 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 fixed dimension. This is the motivation behind sequential Monte Carlo (SMC).

2.1.3 Feynman-Kac models

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

2.1.4 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

⟨alg:SMC⟩ **Algorithm 1:** Sequential Monte Carlo

Figure 2.2 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.

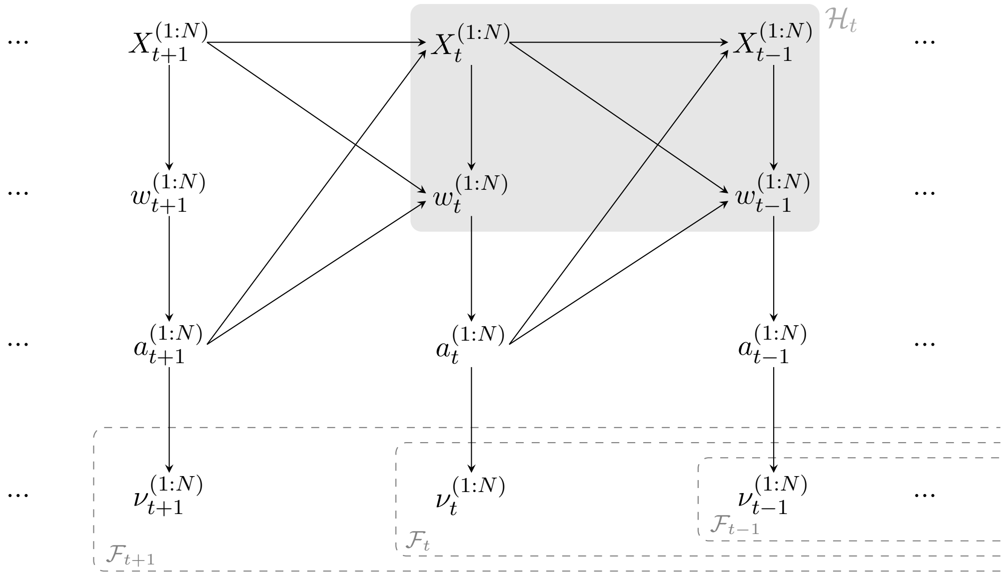


Figure 2.2: Part of the conditional dependence graph implied by Algorithm 1. The direction of time is from left to right. The reverse-time filtration is indicated by the dashed areas. The nodes highlighted in grey generate the separatrix $\mathcal{H}_t$ between $a_t^{(1:N)}$ and $\mathcal{F}_{t-1}$. Use the same shades of grey here as elsewhere

⟨fig:cond_indep_graph⟩

2.1.5 Theoretical justification

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

2.2 Coalescent theory ✓

⟨sec:coaltheory⟩ Write a paragraph introducing the section.

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 Durrett (2008) ?

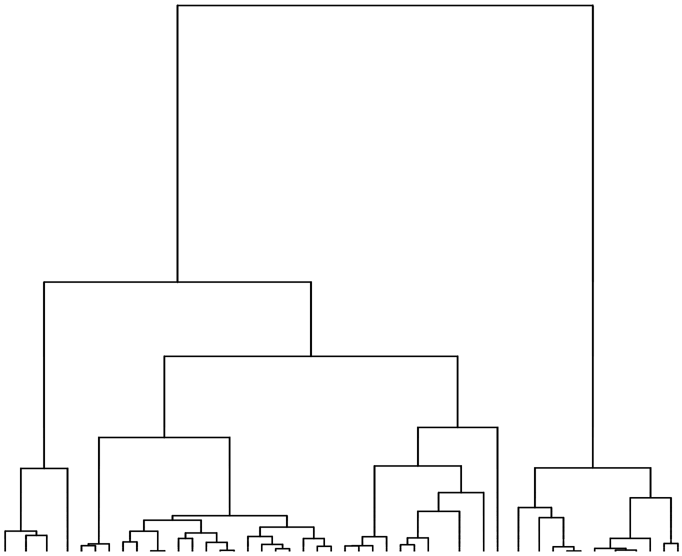


Figure 2.3: A realisation of the n-coalescent with $n = 50$.

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

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

?eq:KCgenerato

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 of Kingman's coalescent

⟨sec:KCproperties⟩ Possibly also include a section about coming down from infinity (just define it basically).

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. 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).

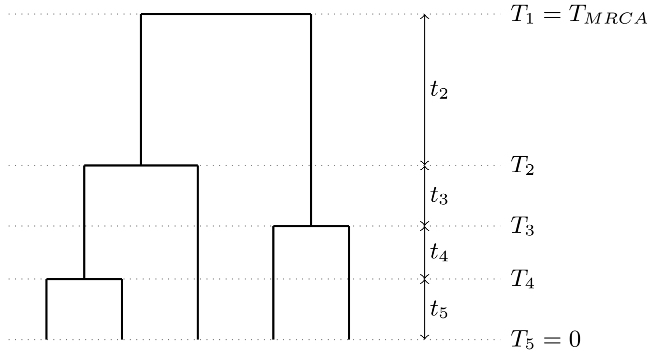


Figure 2.4: Diagram illustrating the definitions of t_i, T_i in the n-coalescent.

⟨fig:KC_timedefns⟩

Denote by $t_2, t_3 \ldots, t_n$ 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 2.4). A consequence of Definition 2.1 is that these waiting times are independent and have distributions

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

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 2.4). The partial sums, being sums of independent Exponential random variables, have HyperExponential distributions.

Refer back to the following three properties later on with reference to their relevance in SMC.

Time to MRCA

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

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

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)

Total branch length

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

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

and

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

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

Probability that sample MRCA equals population MRCA

One other interesting quantity is the probability that the MRCA of k random lineages coincides with the population MRCA (e.g. Durrett 2008, Theorem 1.7). Denote by S_k the relevant event: that a random sample of k lineages has the same as the MRCA as the population. Consider the two subtrees produced by cutting the tree just below the population MRCA. The sample of k lineages coalesces before the population MRCA if and only if all k sampled leaves lie in just one of these two subtrees. A basic consequence of the exchangeability of the n-coalescent is that, in the limit $N \to \infty$, the proportion of leaves in the left subtree is uniformly distributed on $[0, 1]$. Calling this proportion X, we have

$$\mathbb{P}[S_k^c \mid X = x] = x^k + (1-x)^k$$

Integrating against the distribution of X, the probability of interest is

$$\mathbb{P}[S_k] = 1 - \int_0^1 [x^k + (1-x)^k]dx = \frac{k-1}{k+1}$$

as required.

The above is based on properties of the full Kingman coalescent, but similar results are available for the n-coalescent. Consider now a subsample of size k among n lineages that follow the n-coalescent. Denote by $S_{k,n}$ the event that these k lineages have the same MRCA as all n lineages. This probability of this event is calculated in Saunders, Tavaré, and Watterson (1984, Example 1) and again in Spouge (2014, Equation (3)), in both cases arising as a special case of more general results. A direct proof is given below.

Let X be the number of leaves in the left subtree. So $X \in \{1, \ldots, n-1\}$ and, like before, a consequence of exchangeability is that X is uniformly distributed on that set. Now that the total number of branches is finite, we have to count more carefully. Conditional on X

we have

$$\mathbb{P}[S_{k,n}^c \mid X = x] = \left[\binom{x}{k} + \binom{n-x}{k}\right]\binom{n}{k}^{-1}.$$

Integrating against the distribution of X gives

$$\begin{aligned}\mathbb{P}[S_{k,n}] &= 1 - \frac{1}{n-1}\binom{n}{k}^{-1}\sum_{x=1}^{n-1}\left[\binom{x}{k} + \binom{n-x}{k}\right] \\ &= 1 - \frac{1}{n-1}\binom{n}{k}^{-1}\left[\binom{n}{k+1} + \binom{n}{k+1}\right] \\ &= \frac{k-1}{k+1}\frac{n+1}{n-1}\end{aligned}$$

using binomial identities and some algebra. As $n \to \infty$ this agrees with the population-level result above.

2.2.3 Models in population genetics

⟨sec:popgenmodels⟩ 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). Crucially, in the neutral case, offspring counts carry all the information about the distribution of the genealogy that is contained in the parental indices. 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

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

The joint distribution of the offspring counts is therefore

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

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, when time is scaled by a factor of N, 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 $(v_1, \ldots, v_N)$ 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? is not any Kingman 1982 papers, and certainly not Cannings 1974/5 which predates KC

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 $(0, 2, 1, 1, \ldots, 1)$. Therefore we know that under a suitable time-scaling, its genealogies converge to the Kingman coalescent. The time scale in this case is N^2, because reproduction happens at a rate N times or is it technically N-1 times? lower than in the Wright-Fisher model. also cite a Moran-specific convergence result: not sure where (it isn't in Kingman 1982* or in Moran 1958 which predates KC)

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 2.2, 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).

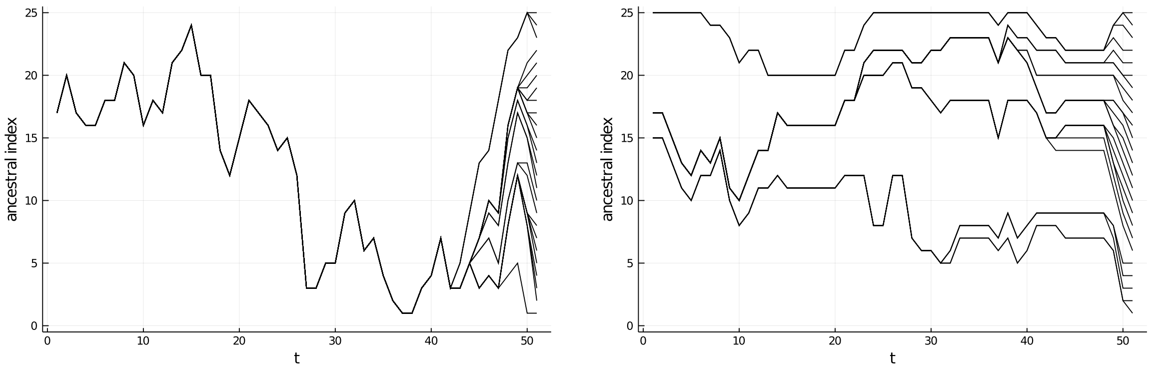
2.3 Sequential Monte Carlo genealogies

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.



(a) Multinomial resampling

(b) Adaptive minimum-variance resampling

Figure 2.5: Illustration of ancestral degeneracy and the mitigating effect of low-variance and adpative resampling. (a) with multinomial resampling, (b) the same system with adaptive systematic resampling.

g: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 ~

As we have seen, resampling is necessary within SMC to "reset" the weights in order to prevent weight degeneracy. The basic role of a resampling scheme is to map the continuous weights to discrete offspring counts, in some "sensible" way (Definition 2.2). The choice of resampling scheme is explored in detail in this section.

2.4.1 Definition ✓

Also say that resampling is itself a Monte Carlo procedure.

⟨defn:resampling⟩ **Definition 2.2.** For our purposes, a valid resampling scheme is a stochastic function mapping weights $w_t^{(1:N)} \in \mathcal{S}_{N-1}$ to offspring counts $\nu_t^{(1:N)} \in \{0, \ldots, N\}^N$ that satisfies the following properties:

⟨item:resampling_property1⟩
1. the population size is conserved: $\sum_{i=1}^N \nu_t^{(i)} = N$ for all N

⟨item:resampling_property2⟩
2. the weights are uniform after resampling: $w_{t+}^{(i)} = 1/N$ for all i

⟨item:resampling_property3⟩
3. the resampling is unbiased: $\mathbb{E}[\nu_t^{(i)} \mid w_t^{(i)}] = N w_t^{(i)}$ for all i.

It is possible to design resampling schemes that violate these properties. For example, a scheme of Liu and Chen (1998) uses the square roots of the weights for resampling, then corrects by setting non-uniform weights after resampling (violating conditions 2 and 3). Fearnhead and Clifford (2003, p.890, point (d)) also appears to resample such that the weights are not uniform after resampling. Resampling different numbers of particles in different iterations (violating condition 1) is of course possible, but we typically have a fixed limit on computational resources, in which case it makes sense to simulate the maximum feasible number of particles N at every iteration. Deterministic resampling schemes (which cannot generally be unbiased, violating condition 3) have been used by some authors. These include schemes based on optimal transport (Reich 2013; Myers et al. 2021; Corenflos et al. 2021) and the importance support points resampling of Huang, Joseph, and Mak (2020). However, the majority of resampling schemes in the literature fit within Definition 2.2, and it is not typically advantageous to violate the properties 1–3.

Within Definition 2.2 there is still a great deal of flexibility. Many different resampling schemes have been proposed in the literature, some of which perform better than others.

Section 2.4.2 introduces some important resampling schemes, and their properties are discussed in Section 2.4.3. These are summarised in Table 2.3.

2.4.2 Examples ~

mples_resamplingschemes⟩ Argue in each case that the scheme is unbiased.

Abbreviation	Description
`multi`	multinomial resampling
`star`	star resampling
`strat`	stratified resampling
`syst`	systematic resampling
`res-multi`	residual resampling with multinomial residuals
`res-star`	residual resampling with star residuals
`res-strat`	residual resampling with stratified residuals
`res-syst`	residual resampling with systematic residuals
`ssp`	Srinivasan sampling procedure resampling
`branch`	minimal variance branching algorithm

Table 2.1: Abbreviations for resampling schemes

⟨tab:resampling_abbrevs⟩

Multinomial resampling ✓

Multinomial resampling (Gordon, Salmond, and Smith 1993; Efron and Tibshirani 1994) is one of the simplest resampling schemes. The parental indices are chosen independently from $\{1, \ldots, N\}$, each with probability given by the weight of the corresponding particle $w_t^{(i)}$. That is,

$$a_t^{(1:N)} \sim \text{Categorical}(\{1, \ldots, N\}, w_t^{(1:N)}).$$

This implies the joint distribution of the offspring counts is

$$\nu_t^{(1:N)} \overset{d}{=} \text{Multinomial}(N, w_t^{(1:N)}).$$

It follows from the mean of the Multinomial distribution that this resampling scheme is unbiased. Although the parental indices are chosen independently, the resulting offspring counts are negatively correlated. — link to GCW19's negative association?

A simple way to sample the parental indices is to use inversion sampling: partition the unit interval into N subintervals each of which will correspond to a certain index i and has length equal to the weight $w_t^{(i)}$; then draw N samples $U_i \sim \text{Uniform}(0, 1)$ and classify them according to which of these subintervals they fall in. Explicitly, the parental index assigned to child i is the index a_i satisfying

$$\sum_{j=1}^{a_i-1} w_t^{(j)} \leq U_i \leq \sum_{j=1}^{a_i} w_t^{(j)}. \tag{2.7}$$ eq:syst_str

This is illustrated in Figure 2.6a.

Fast implementations of multinomial resampling rely on $U_1, \ldots, U_N$ being pre-sorted, which speeds up the search step (2.7). Sorting N numbers is an $O(N \log N)$ operation, but in fact this is not necessary because we can directly sample the order statistics of a Uniform$[0,1]$ distribution [citations: Chopin and Papaspiliopoulos (2020), and a different (or possibly equivalent) method in Hol, Schön, and Gustafsson (2006)] —explore whether these methods are equivalent. This allows multinomial resampling to be implemented at $O(N)$ cost, with the side-effect that the sampled ancestral indices will be ordered. And therefore the sampled parental indices cannot be Cat(N,w) distributed. But the counts are still Multinomial? And anyway for the purposes of resampling this isn't a problem; it might even improve performance?

Residual resampling ✓

Residual resampling is described in Liu and Chen (1998) and also in Whitley (1994) where it is called "remainder stochastic sampling".

Each particle $X_t^{(i)}$ is deterministically assigned $\lfloor N w_t^{(i)} \rfloor$ offspring and the remaining $R := \sum_{i=1}^N (N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor) = nN - \sum_{i=1}^N \lfloor N w_t^{(i)} \rfloor$ offspring are assigned stochastically according to the residual weights

$$r^{(i)} := (N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)/R.$$

Notice that each $r^{(i)}$ lies in the interval $[0, 1/R)$, and $R \in \{0, \ldots, N-1\}$ with $R = 0$ only if all weights are multiples of $1/N$ in which case all residual weights are zero.

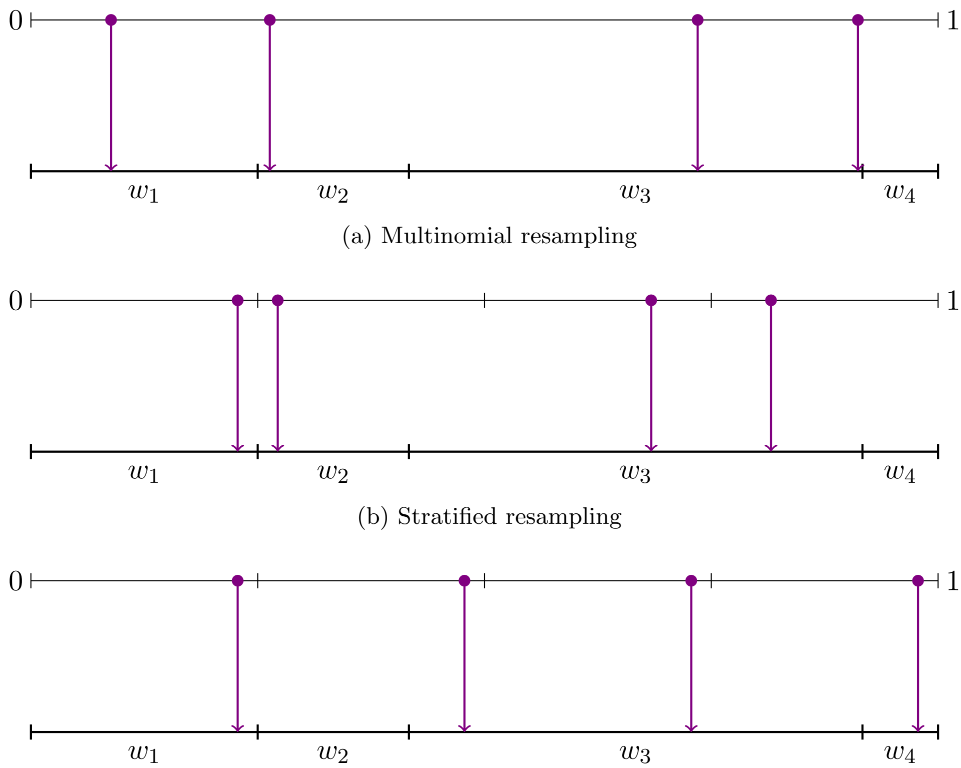
The stochastic part can be done using any of the other basic resampling schemes (e.g. multinomial, stratified, systematic). Most presentations focus on the case where multinomial resampling is used for the residuals, which is by no means the most sensible option. We will explore several different options in what follows.

Stratified resampling ✓

Stratified resampling is introduced in Kitagawa 1996.

As in multinomial resampling, stratified resampling uses inversion sampling to dample the parental indices. However, the samples used for inversion sampling are no longer i.i.d. Uniform$[0,1]$ samples. Instead, one number is sampled from each subinterval of length $1/N$; that is,

$$U_i \sim \text{Uniform}\left(\frac{i-1}{N}, \frac{i}{N}\right).$$

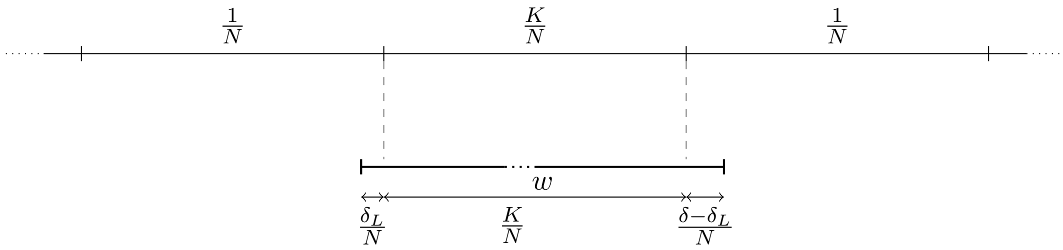


(a) Multinomial resampling

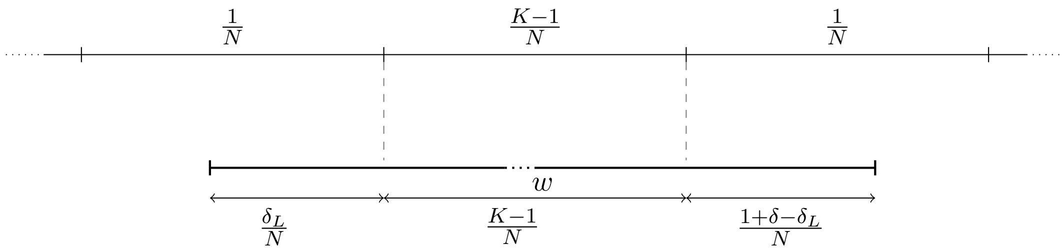
(b) Stratified resampling

(c) Systematic resampling

Figure 2.6: Inversion sampling to obtain Multinomial offspring counts, where the (marginally) Uniform variables for inversion are sampled in different ways. For this example $N = 4$ and the weights are $w_{(1:4)} = \frac{1}{N}(1, \frac{2}{3}, 2, \frac{1}{3})$. Also make the same diagram in Whitley's "roulette wheel" style, to illustrate the difference. Or maybe make it for the "degeneracy under equal weights" section just illustrating the difference it makes in stratified resampling.



(a) The overhang is less than one; $\delta_L \in [0, \delta]$. The parent under consideration is automatically assigned K offspring, plus up to two more.



(b) The overhang is greater than one (this case can only occur when $K \geq 1$); $\delta_L \in (\delta, 1)$. The parent under consideration is automatically assigned $K - 1$ offspring, plus up to two more.

Figure 2.7: Cases for stratified resampling with a fixed weight $w = (K + \delta)/N$.

⟨fig:strat_cases⟩

Alternatively, one may think of standard Uniform samples $u_1, \dots, u_N \sim^{iid}$ Uniform$[0, 1]$ with the transformation

$$U_i = \frac{u_i + i - 1}{N}$$

to give the stratified samples $U_1, \dots, U_N$.

The parents are then assigned as in (2.7). This is illustrated in Figure 2.6b. The offspring distribution is no longer Multinomial, since parental indices are not chosen independently. This scheme ensures that the samples are "well spread out", which reduces the probability of randomly losing high-weight particles or duplicating low-weight particles.

It will be useful later on to have a better idea about the marginal distributions of $\nu_t^{(i)}$ that are induced by stratified resampling. There are complex dependencies between the offspring counts, but we can still find some constraints on the distribution of each count conditional on the corresponding weight. Write the i^{th} weight in the form $w_t^{(i)} = (K+\delta)/N$, where $\delta \in [0, 1)$ and $K \in \{0, \dots, N-1\}$. Considering the illustration Figure 2.6b, the distribution of $\nu_t^{(i)}$ depends not only on $w_t^{(i)}$ but also on where the i^{th} weight interval falls with respect to the length-$(1/N)$ intervals. Denote the left overhang δ_L. There are two cases to consider, which are illustrated in Figure 2.7. In Case (a) the total overhang is less than $1/N$ and $\delta_L \in [0, \delta]$. In Case (b) the total overhang is greater than $1/N$ and $\delta_L \in (\delta, 1)$. Arrangements such that one or both ends have no overhang are special cases of Case (a), with the $\delta_L \in \{0, \delta\}$. Note that Case (b) cannot occur if $K = 0$.

In any case $\nu_t^{(i)} \in \{K-1, K, K+1, K+2\}$ almost surely. To define a probability distribution over these four values, we introduce the notation $p_j := \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + j \mid w_t^{(i)}]$, for $j = -1, 0, 1, 2$. Change notation to avoid confusion with $p_k^{(i)}$'s in Chapter 5?

Since the sample within each interval of length $1/N$ is uniform over that interval, we find the probabilities given in Table 2.2, in terms of δ and δ_L. The probabilities do not depend on K, but of course the corresponding values of $\nu_t^{(i)}$ do.

	Case (a)	Case (b)	L.B.	U.B.
p_{-1}	0	$\delta_L(1+\delta-\delta_L)-\delta$	0	$1/4$
p_0	$1-\delta+\delta_L(\delta-\delta_L)$	$1+\delta-2\delta_L(1+\delta-\delta_L)$	$(1-\delta)/2$	$1-3\delta/4$
p_1	$\delta-2\delta_L(\delta-\delta_L)$	$\delta_L(1+\delta-\delta_L)$	$\delta/2$	$(1+\delta)/2$
p_2	$\delta_L(\delta-\delta_L)$	0	0	$1/4$

Table 2.2: Marginal probability distribution of $\nu_t^{(i)}$ conditional on $w_t^{(i)}=(K+\delta)/N$, in terms of δ and the "left overhang" δ_L, along with upper and lower bounds on these in terms of δ only, which hold in both cases.

⟨tab:strat_probs⟩

Systematic resampling ✓

Systematic resampling is described in Carpenter, Clifford, and Fearnhead (1999) and also in Whitley (1994) where it is called "stochastic universal sampling".

Like stratified resampling, it uses the inversion sampler of multinomial resampling but starts with a more regular set of points in $[0,1]$. In this scheme, only one standard Uniform sample is drawn, $u\sim\text{Uniform}[0,1]$, from which the N samples are generated by via the transformation

$$U_i=\frac{u+i-1}{N}$$

for $i=1,\ldots,N$. The parental indices are again selected according to (2.7). The method is illustrated in Figure 2.6c.

Kitagawa (1996) suggests a deterministic scheme in which the random u is replaced by a fixed $\alpha\in[0,1]$; but, being deterministic, this scheme does not satisfy the unbiasedness property (condition 1 in Definition 2.2). Whitley (1994) describes systematic resampling using a different picture, whereby the interval $[0,1]$ is joined up into a circle, and the systematic samples are evenly spaced pointers on an outer ring, which is spun around like a roulette wheel to sample a random phase which, modulo 1, is equal to u. figure please For systematic resampling, Whitley's "roulette wheel" representation is equivalent to that of Figure 2.6c.

Like stratified resampling, systematic resampling ensures the random numbers are "well spread out"; the resulting samples are even more constrained than with stratified resampling. Systematic resampling also has the advantage of being extremely easy to implement and also computationally efficient, requiring only one sample from a pseudo-random number generator (PRNG) followed by $O(N)$ elementary operations.

However, this scheme is known to exhibit pathological behaviour in some cases because its performance depends on the ordering of the weights. A simple example of this phenomenon is presented in Douc, Cappé, and Moulines (2005). Such behaviour can be avoided by randomly permuting the weights before resampling, and this is the recommended practice.

Star resampling ✓

For the sake of comparison, we also construct a resampling scheme which is the worst possible (in some sense). Sample

$$a_t \sim \mathrm{Categorical}(\{1, \dots, N\}, w_t^{(1:N)})$$

and set $a_t^{(i)} = a_t$ for all i. The resulting offspring counts are all equal to zero except for $\nu_t^{(a_t)}$, which is equal to N. This resampling scheme is indeed unbiased, since each offspring count has marginal distribution

$$\nu_t^{(i)} \mid w_t^{(1:N)} = \begin{cases} 0 & \text{w.p. } 1 - w_t^{(i)} \\ N & \text{w.p. } w_t^{(i)}. \end{cases}$$

We also see these offspring counts have the highest possible marginal variance, subject to $\mathbb{E}[\nu_t^{(i)} \mid w_t^{(i)}] = N w_t^{(i)}$ and $\nu_t^{(i)} \in \{0, \dots, N\}$.

I call this scheme *star resampling* because the parent-offspring relationships at each iteration form a star graph.

Minimum-variance resampling

The minimal variance branching algorithm of Crisan and Lyons (1999) provides a framework for minimal-variance resampling. The idea is to enforce minimal variance by resampling such that each offspring count $\nu_t^{(i)}$, conditionally on $w_t^{(i)}$, has marginal distribution

$$\nu_t^{(i)} \mid w_t^{(i)} \overset{d}{=} \lfloor N w_t^{(i)} \rfloor + \mathrm{Bernoulli}(N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor). \tag{2.8}$$ eq:branchin

We will see later on that this is exactly the framework of *stochastic rounding*. The set-up of Crisan and Lyons (1999) does not require the number of particles to remain constant from one generation to the next (Property 1 in Definition 2.2), so their minimal variance branching algorithm could be implemented for instance by sampling each $\nu_t^{(i)}$ independently from (2.8). The authors remark that enforcing strictly negative correlation between the offspring counts can improve the rate of convergence, but they do not specify how this might be achieved.

Also write about Gerber, Chopin, and Whiteley (2019), which in some sense extends/formalises the notions of Crisan and Lyons (1999).

2.4.3 Properties ∼

c:resampling_properties⟩ Low-variance: variance of what? Different criteria/ definitions of optimality. Link back to adaptive resampling: interaction between adaptive and low-variance resampling. Comparison of properties of these, existing results comparing schemes. Implementation considerations. Theoretical justification (or lack of). Mention computational complexity. This section was dumped from elsewhere and most of its subsections need redrafting. Also add a paragraph here to introduce it, saying that everything is summarised in the table.

Support of offspring numbers ✓

Let us consider the support of the marginal offspring distributions in each scheme, conditional on the weights. Suppose that the i^{th} weight lies in the interval $w_t^{(i)} \in [k/N, (k+1)/N]$.

Under multinomial resampling, it is possible for $\nu_t^{(i)}$ to take any value from 0 to N (although some values are of course more likely than others). Thus it is possible for a high-weight particle to have zero offspring, or a low-weight particle to have many offspring, simply by chance. Recall that the weights give an indication of how "useful" each particle is for the approximation. Thus killing a high-weight particle is likely to increase the variance of the SMC estimates, while duplicating a low-weight particle wastes computational resources on propagating particles that will not contribute much to reducing that variance.

Residual resampling ensures that every particle with above-average (i.e. $> 1/N$) weight has at least one offspring, avoiding the loss of high-weight particles. If the residuals are sampled using multinomial resampling then the duplication of low-weight particles is not avoided, $\nu_t^{(i)} \in \{k, \dots, k+R\} \subseteq \{k, \dots, N\}$, but this can be addressed by using a lower-variance scheme for the residual offspring. Various choices are included in Table 2.3.

Stratified resampling is more restrictive, $\nu_t^{(i)} \in \{k-1, k, k+1, k+2\}$, but allows the possibility of a particle with above-average weight having no offspring. Systematic resampling has the smallest support, $\nu_t^{(i)} \in \{k, k+1\}$, that is possible whilst maintaining unbiasedness.

Another way to quantify this property is by considering the maximum possible difference between the offspring count $\nu_t^{(i)}$ and its expected value $Nw_t^{(i)}$. This is also presented in Table 2.3.

Degeneracy under equal weights ✓

In the case where all of the weights are multiples of $1/N$, low-variance schemes such as residual and systematic resampling become fully deterministic. Since $\lfloor Nw_t^{(i)} \rfloor = Nw_t^{(i)}$ for each i, residual resampling will have $R = 0$ leaving no remainder to be assigned stochastically. In systematic resampling exactly $\lfloor Nw_t^{(i)} \rfloor = Nw_t^{(i)}$ samples will fall in the i^{th} interval. In particular, if $w_t^{(1:N)} = (1, \dots, 1)/N$ then each parent is assigned exactly one offspring deterministcially, so there is effectively no resampling.

The same phenomenon occurs with stratified resampling, but not if one uses Whitley's roulette wheel description (Figure ??). The random phase shift introduced by "spinning the wheel" prevents the inversion sampling intervals from lining up exactly with the weight intervals, so the resampled offspring counts may vary from their means by one either side. Whitley (1994) does not describe stratified resampling, but we see that unlike with systematic resampling, the roulette wheel description is not equivalent to the standard inversion sampling description. For stratified resampling, the roulette wheel adds some extra randomness, so the straightforward inversion sampler is preferred.

If the state space is continuous, the event that all weights are multiples of $1/N$ typically has zero measure, but with non-zero probability we can get arbitrarily close to this regime in which resampling becomes deterministic.

Marginal variance of offspring counts ✓

Mention negative association? = teaser for later, which has to do with covariance between counts rather than marginal variance.

One indication of the performance could be the variance of the resampled offspring counts. For instance we might ask what is the marginal variance of $\nu_t^{(i)}$, conditional on the corresponding weight $w_t^{(i)}$.

In multinomial resampling, the marginal distributions are

$$\nu_t^{(i)} \mid w_t^{(i)} \sim \text{Binomial}(N, w_t^{(i)})$$

so the variance is

$$\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}] = N w_t^{(i)}(1 - w_t^{(i)}).$$

Compare this to star resampling, where the marginal offspring counts

$$\nu_t^{(i)} \mid w_t^{(i)} \stackrel{d}{=} N \, \text{Bernoulli}(w_t^{(i)})$$

having variance

$$\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}] = N^2 w_t^{(i)}(1 - w_t^{(i)}),$$

N times larger than in the multinomial case.

As pointed out in Crisan and Lyons (1999, p.557), their minimal variance branching process yields offspring variance

$$\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}] = (N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)(1 - N w_t^{(i)} + \lfloor N w_t^{(i)} \rfloor) \leq \frac{1}{4},$$

since the stochastic part of $\nu_t^{(i)}$ is a Bernoulli$(N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)$ random variable (as seen in (2.8)). The same marginal variance appears from systematic, residual-systematic and SSP resampling, since these all share the same marginal offspring distributions. We will see in Section 2.4.4 that all of these schemes fall within the *stochastic rounding* class, and the marginal offspring variance is a property shared by all stochastic roundings.

The marginal variance is harder to calculate for other schemes such as residual-multinomial and stratified resampling because these were not defined in terms of marginal distributions, nor are the offspring counts independent conditional on the weights. However, it is possible in some cases to find upper bounds on the variance, and some such bounds are derived below.

Residual-multinomial: $\nu_t^{(i)}$ depends on all of the other weights, as well as $w_t^{(i)}$, but only through the statistic $R := \sum(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)$. We have

$$\nu_t^{(i)} \mid w_t^{(i)}, R \overset{d}{=} \lfloor Nw_t^{(i)} \rfloor + \text{Binomial}\left(R, \frac{Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor}{R}\right).$$

Using the law of total variance,

$$\begin{aligned}
\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}] &= \mathbb{E}\left[\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}, R] \mid w_t^{(i)}\right] + \text{Var}\left[\mathbb{E}[\nu_t^{(i)} \mid w_t^{(i)}, R] \mid w_t^{(i)}\right] \\
&= \mathbb{E}\left[(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\left(1 - \frac{Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor}{R}\right) \mid w_t^{(i)}\right] \\
&\quad + \text{Var}\left[Nw_t^{(i)} \mid w_t^{(i)}\right] \\
&= Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor - (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)^2\, \mathbb{E}[R^{-1} \mid w_t^{(i)}] \\
&\leq Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor.
\end{aligned}$$

Similarly, for residual resampling with star residuals,

$$\nu_t^{(i)} \mid w_t^{(i)}, R \overset{d}{=} \lfloor Nw_t^{(i)} \rfloor + R\,\text{Bernoulli}\left(\frac{Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor}{R}\right).$$

and we find

$$\begin{aligned}
\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}] &= \mathbb{E}\left[\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}, R] \mid w_t^{(i)}\right] + \text{Var}\left[\mathbb{E}[\nu_t^{(i)} \mid w_t^{(i)}, R] \mid w_t^{(i)}\right] \\
&= \mathbb{E}\left[R(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\left(1 - \frac{Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor}{R}\right) \mid w_t^{(i)}\right] \\
&\quad + \text{Var}\left[Nw_t^{(i)} \mid w_t^{(i)}\right] \\
&= \mathbb{E}\left[R(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\left(1 - \frac{Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor}{R}\right) \mid w_t^{(i)}\right] \\
&= (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)\mathbb{E}\left[R \mid w_t^{(i)}\right] - (Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor)^2 \\
&\leq N(Nw_t^{(i)} - \lfloor Nw_t^{(i)} \rfloor).
\end{aligned}$$

For stratified resampling, we can use the constraints on the marginal offspring distribution that were derived in Section 2.4.2. Recall that, conditional on $w_t^{(i)}$, $\nu_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + j$ with probability p_j for $j = -1, 0, 1, 2$. We can use the values of p_{-1}, p_0, p_1, p_2 in the two cases of Figure 2.7 these calculations will need updating since Table 2.2 is now more informative, as summarised in Table 2.2, to bound the variance. First write

$$\begin{aligned}\mathrm{Var}[\nu_t^{(i)} \mid w_t^{(i)}] &= \mathbb{E}[(\nu_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)^2 \mid w_t^{(i)}] - \mathbb{E}[\nu_t^{(i)} - \lfloor N w_t^{(i)} \rfloor \mid w_t^{(i)}]^2 \\ &= p_{-1} + p_1 + 4p_2 - (-p_{-1} + p_1 + 2p_2)^2.\end{aligned}$$

In Case (a), we have

$$\mathrm{Var}[\nu_t^{(i)} \mid w_t^{(i)}] = (\delta - 2\delta_L\delta_R + 4\delta_L\delta_R) - (\delta - 2\delta_L\delta_R + 2\delta_L\delta_R)^2 = \delta + 2\delta_L\delta_R - \delta^2$$

which is maximised at $\delta_L = \delta_R = \delta/2$ for a maximum variance of $\delta(1-\delta/2)$, which is at most $1/2$. In Case (b), we have

$$\mathrm{Var}[\nu_t^{(i)} \mid w_t^{(i)}] = (x_L x_R - \delta + x_L x_R) - (-x_L x_R + \delta + x_L x_R)^2 = \delta + 2x_L x_R - \delta^2$$

which is maximised at $x_L = x_R = (1+\delta)/2$ for a maximum variance of $(1-\delta^2)/2$, which is at most $1/2$. Overall we have the bound

$$\mathrm{Var}[\nu_t^{(i)} \mid w_t^{(i)}] \leq \frac{1}{2}$$

for any $w_t^{(1:N)}$.

Residual-stratified resampling has the further constraint that $p_{-1} = 0$ (i.e. Case 3 of Figure 2.7 doesn't occur) since the residual weights are between 0 and $1/R$. However this does not give an improvement on the stratified bound:

$$\mathrm{Var}[\nu_t^{(i)} \mid w_t^{(i)}] \leq 1/2.$$

Table 2.3 includes upper bounds on $\mathrm{Var}[\nu_t^{(i)}]$ for various resampling schemes, independent of $w_t^{(i)}$. Those general bounds are derived from the results of this section, bounded above independently of the weights. Some of the bounds are certainly not tight.

Contribution to the Monte Carlo variance

Finish the proof. Remark that we can't do a formal variance comparison for syst (and others?).

While the variance of the offspring counts goes some way to providing a comparison between the various resampling schemes, a more relevant property is the contribution of the resampling step to the Monte Carlo variance. This quantifies directly the effect that a certain choice of resampling scheme has on the variance of the Monte Carlo estimators.

Let $(\mathcal{G}_t)_{t\geq 0}$ be the filtration generated by the particle positions and weights up to and including time t. Let $\tilde{X}_t^{(i)}$ denote position of the ith resampled particle. We consider the one-step Monte Carlo variance induced by resampling, that is

$$\rho(\varphi) := \operatorname{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(\tilde{X}_t^{(i)})\,\middle|\,\mathcal{G}_t\right] \tag{2.9}$$ eq:resampli

where φ is an arbitrary test function.

Some results comparing this variance across different resampling schemes are presented in Douc, Cappé, and Moulines (2005). Their results, plus some additional results, are presented in Proposition 2.3.

hm:resampling_var_compare⟩ **Proposition 2.3** (Variance of resampling schemes)**.** *Let $\rho_{\mathtt{multi}}$ etc. denote the variance (2.9) under the various resampling schemes, as abbreviated in Table 2.1. For any square-integrable ? function φ,*

⟨item:resampling_var1⟩
(a) $\rho_{\mathtt{multi}}(\varphi) \geq \rho_{\mathtt{res\text{-}multi}}(\varphi)$

⟨item:resampling_var2⟩
(b) $\rho_{\mathtt{multi}}(\varphi) \geq \rho_{\mathtt{strat}}(\varphi)$

⟨item:resampling_var3⟩
(c) $\rho_{\mathtt{star}}(\varphi) = N\rho_{\mathtt{multi}}(\varphi)$

⟨item:resampling_var4⟩
(d) $\rho_{\mathtt{res\text{-}star}}(\varphi) \geq \rho_{\mathtt{res\text{-}multi}}(\varphi) \geq \rho_{\mathtt{res\text{-}strat}}(\varphi)$

Parts (a) and (b) were proved in Douc, Cappé, and Moulines (2005, Section 3). The second inequality in (d) is stated in Gerber, Chopin, and Whiteley (2019, p.9) and follows from (b), as shown below. Don't include proofs for things that were already proved by someone else, unless this really adds value.

Proof. **multinomial resampling:** the resampled indices are conditionally i.i.d., so

$$\begin{aligned}
\rho_{\mathtt{multi}}(\varphi) &= \operatorname{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(\tilde{X}_t^{(i)})\,\middle|\,\mathcal{G}_t\right] = \frac{1}{N}\operatorname{Var}\left[\varphi(\tilde{X}_t^{(i)})\,\middle|\,\mathcal{G}_t\right] \\
&= \frac{1}{N}\left\{\mathbb{E}\left[\varphi^2(\tilde{X}_t^{(i)})\,\middle|\,\mathcal{G}_t\right] - \mathbb{E}\left[\varphi(\tilde{X}_t^{(i)})\,\middle|\,\mathcal{G}_t\right]^2\right\} \\
&= \frac{1}{N}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})\mathbb{P}[\tilde{X}_t^{(i)} = X_t^{(j)} \mid \mathcal{G}_t] - \frac{1}{N}\left\{\sum_{j=1}^{N}\varphi(X_t^{(j)})\mathbb{P}[\tilde{X}_t^{(i)} = X_t^{(j)} \mid \mathcal{G}_t]\right\}^2 \\
&= \frac{1}{N}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})w_t^{(j)} - \frac{1}{N}\left\{\sum_{j=1}^{N}\varphi(X_t^{(j)})w_t^{(j)}\right\}^2.
\end{aligned}$$

star resampling: all of the resampled indices are equal, say $\tilde{X}_t^{(1)} = \cdots = \tilde{X}_t^{(N)} = X_t^{\star}$, so

$$
\begin{aligned}
\rho_{\texttt{star}}(\varphi) &= \operatorname{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\varphi(\tilde{X}_t^{(i)})\,\middle|\,\mathcal{G}_t\right] = \operatorname{Var}\left[\varphi(X_t^\star)\mid\mathcal{G}_t\right] \\
&= \mathbb{E}\left[\varphi^2(X_t^\star)\mid\mathcal{G}_t\right] - \mathbb{E}\left[\varphi(X_t^\star)\mid\mathcal{G}_t\right]^2 \\
&= \sum_{j=1}^{N}\varphi^2(X_t^{(j)})\mathbb{P}[X_t^\star = X_t^{(j)}\mid\mathcal{G}_t] - \left\{\sum_{j=1}^{N}\varphi(X_t^{(j)})\mathbb{P}[X_t^\star = X_t^{(j)}\mid\mathcal{G}_t]\right\}^2 \\
&= \sum_{j=1}^{N}\varphi^2(X_t^{(j)})w_t^{(j)} - \left\{\sum_{j=1}^{N}\varphi(X_t^{(j)})w_t^{(j)}\right\}^2 \\
&= N\rho_{\texttt{multi}}(\varphi).
\end{aligned}
$$

This proves part (c). Here we see the same factor of N as we had with the marginal variance of offspring counts, due to the variance reduction achieved by taking N independent copies (multinomial resampling) as opposed to N identical copies (star resampling).

residual-multinomial resampling: the Monte Carlo estimate in (2.9) can be decomposed into a sum of conditionally deterministic terms plus a sum of conditionally i.i.d. terms: conditional on $\mathcal{G}_t$,

$$
\frac{1}{N}\sum_{i=1}^{N}\varphi(\tilde{X}_t^{(i)}) = \frac{1}{N}\sum_{i=1}^{N}\lfloor Nw_t^{(i)}\rfloor\varphi(X_t^{(i)}) + \frac{1}{N}\sum_{i=1}^{R}\varphi(\hat{X}_t^{(i)})
$$

where $\hat{X}_t^{(i)} \sim^{\text{iid}} \text{Multinomial}(R, r^{(1:N)})$. The first sum is conditionally deterministic and hence does not contribute to the Monte Carlo variance (2.9). By a similar calculation to that for multinomial resampling,

$$
\begin{aligned}
\rho_{\texttt{res-multi}}(\varphi) &= \operatorname{Var}\left[\frac{1}{N}\sum_{i=1}^{R}\varphi(\hat{X}_t^{(i)})\,\middle|\,\mathcal{G}_t\right] \\
&= \frac{R}{N^2}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})r^{(j)} - \frac{R}{N^2}\left(\sum_{j=1}^{N}\varphi(X_t^{(j)})r^{(j)}\right)^2 \\
&= \frac{1}{N}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})w_t^{(j)} - \frac{1}{N^2}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})\lfloor Nw_t^{(j)}\rfloor - \frac{R}{N^2}\left(\sum_{j=1}^{N}\varphi(X_t^{(j)})r^{(j)}\right)^2.
\end{aligned}
$$

By a similar argument, it can be shown that

$$
\rho_{\texttt{res-star}}(\varphi) = R\,\rho_{\texttt{res-multi}}(\varphi) \geq \rho_{\texttt{res-multi}}(\varphi),
$$

whenever $R \geq 1$, proving the first inequality in (d) (which holds trivially when $R = 0$ because both residual schemes then have zero variance). Maybe I should do res-star explicitly actually; if I'm including proofs that have already been published then I ought to include proofs that haven't. To prove (a), write

$$\begin{aligned}
\rho_{\texttt{res-multi}}(\varphi) &= \frac{1}{N}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})w_t^{(j)} - \frac{1}{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})\lfloor Nw_t^{(j)}\rfloor + \frac{R}{N}\left(\sum_{j=1}^{N}\varphi(X_t^{(j)})r^{(j)}\right)^2\right\} \\
&\leq \frac{1}{N}\sum_{j=1}^{N}\varphi(X_t^{(j)})w_t^{(j)} - \frac{1}{N}\left\{\frac{1}{N}\sum_{j=1}^{N}\varphi(X_t^{(j)})\lfloor Nw_t^{(j)}\rfloor + \frac{R}{N}\sum_{j=1}^{N}\varphi(X_t^{(j)})r^{(j)}\right\}^2 \\
&= \frac{1}{N}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})w_t^{(j)} - \frac{1}{N}\left\{\sum_{j=1}^{N}\varphi(X_t^{(j)})w_t^{(j)}\right\}^2 = \rho_{\texttt{multi}}(\varphi).
\end{aligned}$$

The inequality is an application of Jensen's inequality, since

$$\sum_{j=1}^{N}\frac{\lfloor Nw_t^{(j)}\rfloor}{N} + \frac{R}{N} = 1.$$

... ∎

Exchangeability ∼

We will call a resampling scheme exchangeable if the resulting distribution of parental indices is invariant under permutations of the children. To put it another way, each child chooses its parent from the same marginal distribution.

It is clear that multinomial resampling is exchangeable since in this case the parental indices are independent and identically distributed. However the efficient implementation of multinomial sampling that takes sorted inputs does not preserve exchangeability.

Stratified and systematic resampling are clearly not exchangeable since, for instance, child 1 is more likely to choose parent 1 than child N is. However, this is merely a feature of the arbitrary ordering of the sampling steps: exchangeability can easily be reintroduced (at $O(N)$ cost) by applying a random permutation to the vector of parental indices after sampling. The same goes for residual resampling. This property will not appear in Table 2.3 since it depends upon the particular implementation.

Permutation invariance ✓

A strange property of stratified and systematic resampling is that they are sensitive to the order in which the subintervals are placed. For example, in Figures 2.6b and 2.6c if the intervals w_2 and w_4 were swapped, the number of offspring assigned to particles 2 and 4 would be swapped in each case. Better to use an example where *distribution* of offspring counts (conditional on the weights but not on the Uniform samples) differs depending on order. Such an example is included in my YRM19 presentation on resampling. We can also see that because w_1 has weight $\geq 1/N$ and is placed first, it is guaranteed at least one offspring.

This property can lead to pathological behaviour, but is easily avoided by applying a random permutation to the order of the subintervals. The SSP resampling scheme of Gerber, Chopin, and Whiteley (2019) is intended to share the benefits of systematic resampling whilst avoiding this property.

Sorting

Results from Gerber, Chopin, and Whiteley (2019) about benefits of sorting. What about sorting instead by weights?

Computational complexity ✓

All of the resampling algorithms discussed in Section 2.4.2 can be implemented in $O(N)$ operations. Even `star` and `SSP` and `branching`? If it turns out to differ depending on resampling scheme then include it as a column in Table 2.3. — I think we can't say for `branching` becaue it dpeends on implementation, but Crisan and Lyons (1999, Corollary 18) seems to imply $O(N^2)$...? SSP is definitely $O(N)$. Considering the complexity of each operation, Hol (2004) and Hol, Schön, and Gustafsson (2006) suggest that systematic resampling is fastest because it only requires one pseudo-random number generation, and multinomial resampling is slower than stratified resampling because of the transformations required. Residual resampling is hard to compare directly because a random fraction of the operations are deterministic, so the number of pseudo-random numbers required is less than N. This analysis was backed up by simulation experiments. However, the analysis of per-particle cost is sensitive to the particular implementation of each resampling scheme, the system implementation of pseudo-random number generation and arithmetic operations, and the hardware used.

Negative association

Definition from Gerber, Chopin, and Whiteley (2019). Why is it a good criterion? Which resampling schemes do/don't satisfy it? Also, this could be added as a column in Table 2.3.

Following Gerber, Chopin, and Whiteley (2019), we use the definition of negative association from Joag-Dev and Proschan (1983).

Definition 2.4. Let $(Z_1, \dots, Z_n)$ be a collection of random variables. $Z_{1:n}$ are said to be *negatively associated* if, for every partition of $\{1, \dots, n\}$ into subsets I and J, for all real-valued coordinatewise non-decreasing functions φ, ψ for which the covariance is well defined,

$$\mathrm{Cov}\left[\varphi(Z_I), \psi(Z_J)\right] \leq 0.$$

Star discrepancy ✓

(See Hol, Schön, and Gustafsson (2006) for inspiration.) Include a diagram showing the quantity inside the $D^\star$ supremum, plotted over u, for mn/strat/syst? (sketched in sora — use same weights and samples as in Figure 2.6a.) Does not appear in Table 2.3 since it only makes sense for resampling schemes based on inversion sampling.

The *star discrepancy* [citation] is a measure of the regularity of a given set of points $u_{1:N}$ in the unit hypercube. For our purposes it is sufficient to define the star discrepancy in one dimension:

$$D^\star(u_1, \ldots, u_N) := \sup_{u \in [0,1]} \left| \frac{1}{N} \sum_{i=1}^{N} \mathbb{1}_{\{u_i \leq u\}} - u \right|.$$

The quantity inside the supremum is the difference between the empirical CDF of the observed points $u_{1:N}$ and the CDF of the Uniform distribution on $[0, 1]$. The star discrepancy is used in quasi-Monte Carlo, where "low-discrepancy" points are used in place of uniform samples to decrease the variance of Monte Carlo estimates.

We have noted already we haven't actually, but we should that resampling can itself be viewed as a Monte Carlo procedure. From this point-of-view, stratified and systematic resampling are quasi-Monte Carlo versions of multinomial resampling, since they provide "more regular" points to be used in inversion sampling.

In one dimension, the lowest-discrepancy point set is the regular grid $(\frac{1}{2N}, \frac{3}{2N}, \ldots, \frac{2N-1}{2N})$, which has star discrepancy $1/(2N)$ [citation]. However, to maintain unbiasedness of resampling, the points must have marginal Uniform$[0, 1]$ distributions erm, they don't in e.g. strat and syst. what is the actual requirement?, which the regular grid points clearly do not. The point sets generated in stratified and systematic resampling both have star discrepancy between $1/(2N)$ and $1/N$ almost surely, where the exact value depends on the realisation. This certainly seems to improve on independent uniform points which can have star discrepancy arbitrarily close to 1, the maximum possible value, albeit with diminishing probability as N increases.

	support of $\nu_t^{(i)}$ given $\frac{k}{N} \leq w_t^{(i)} < \frac{k+1}{N}$	$\sup_w$ $\lvert\nu_t^{(i)} - Nw_t^{(i)}\rvert$	upper bound on $\mathrm{Var}[\nu_t^{(i)}]$	stochastic rounding?	degenerate if $w_t^{(1:N)} = \frac{1}{N}(1,\ldots,1)$?	sensitive to permutations of weights?	PRNG calls
`multi`	$\{0,\ldots,N\}$	N	$N/4$	$\times$	$\times$	$\times$	N
`star`	$\{0,N\}$	N	$N^2/4$	$\times$	$\times$	$\times$	1
`strat`	$\{k-1,k,k+1,k+2\}$	2	$1/2$	$\times$	✓	✓	N
`syst`	$\{k,k+1\}$	1	$1/4$	✓	✓	✓	1
`res-multi`	$\{k,\ldots,N\}$	N	1	$\times$	✓	$\times$	$\leq N$
`res-star`	$\{k,N\}$	N	N	$\times$	✓	$\times$	1
`res-strat`	$\{k,k+1,k+2\}$	2	$1/2$	$\times$	✓	✓	$\leq N$
`res-syst`	$\{k,k+1\}$	1	$1/4$	✓	✓	✓	1
`ssp`	$\{k,k+1\}$?	1?	$1/4$?	✓?	✓?	✓?	?
`branch`	$\{k,k+1\}$	1	$1/4$?	✓	✓		

Table 2.3: Summary of some of the properties of resampling schemes explored in Section 2.4.3. The abbreviated names for the resampling schemes are explained in Table 2.1. I need to include an explanation of the column titles in the caption too. Some properties are not specified for `branching` bacuase they will depend on the particular implementation.

mpling_properties⟩

2.4.4 Stochastic rounding ✓

⟨sec:SRs⟩

?⟨defn:stochround⟩? **Definition 2.5.** Let $X = (X_1, \ldots, X_N)$ be a $\mathbb{R}_+^N$-valued random variable. Then $Y = (Y_1, \ldots, Y_N) \in \mathbb{N}^N$ is a *stochastic rounding* of X if each element Y_i takes values

$$Y_i \mid X_i = \begin{cases} \lfloor X_i \rfloor & \text{with probability } 1 - X_i + \lfloor X_i \rfloor \\ \lfloor X_i \rfloor + 1 & \text{with probability } X_i - \lfloor X_i \rfloor. \end{cases}$$

By construction, $\mathbb{E}(Y_i) = X_i$ for each i. Taking X to be N times the vector of particle weights, we can therefore use stochastic rounding to construct a valid resampling scheme, under the further constraint that $Y_1 + \cdots + Y_N = N$. Several ways to enforce this constraint on the joint distribution have been proposed, including systematic resampling, residual resampling with systematic residuals, the minimal variance branching system of Crisan and Lyons (1997), and the Srinivasan sampling process resampling introduced in Gerber, Chopin, and Whiteley (2019).

Explicitly, the offspring counts are marginally distributed according to

$$\nu_t^{(i)} \mid w_t^{(i)} \overset{d}{=} \lfloor N w_t^{(i)} \rfloor + \text{Bernoulli}(N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor).$$

Some of the properties discussed earlier are common to every stochastic rounding scheme. Since all such schemes give offspring counts with the same marginal distributions, properties such as the marginal offspring variance are common to all stochastic roundings. Indeed it is easy to see that the marginal variance of the offspring counts, $\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}]$ is as small as possible under the constraint of unbiasedness (refer to the property in Defintion 2.2?), and as such this is sometimes referred to as minimal-variance resampling. By definition the support of an offspring count $\nu_t^{(i)}$, if the associated weight lies in the interval $k/N \leq w_t^{(i)} < (k+1)/N$, is $\{k, k+1\}$. All stochastic roundings are also degenerate by definition when the weights are all equal, i.e. $w_t^{(1:N)} = (1, \ldots, 1)/N$.

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.

One popular particle MCMC algorithm is particle marginal Metropolis-Hastings (Andrieu, Doucet, and Holenstein 2010)[Section 2.4.2], a pseudo-marginal MCMC algorithm in which SMC provides an unbiased likelihood estimate with which to compute the Metropolis-Hastings acceptance probability. The following exposition will focus on another particle

MCMC algorithm, namely the particle Gibbs sampler (Andrieu, Doucet, and Holenstein 2010)[Section 2.4.3], which is more interesting from the point-of-view of SMC genealogies.

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. 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. The following was dumped from elsewhere and NEEDS REDRAFTING.

The scenario we present is a particle Gibbs algorithm for filtering with unknown parameters. The method applies more generally to particle Gibbs (for which the reader is directed to Lindsten and Schön (2013, Chapter 5)), but we find this particular scenario to be simple and instructive. To generalise to Del Moral's SMC framework basically requires just the change of notation $g \to G_t, f \to M_t$.

Consider the following hidden Markov model, parametrised by a constant parameter θ which may be multidimensional. Define the spaces in which theta,X,Y live?

$$
\begin{aligned}
Y_t \mid X_t &\sim g_\theta(\cdot \mid X_t) \\
X_t \mid X_{t-1} &\sim f_\theta(\cdot \mid X_{t-1}) \\
X_0 &\sim \mu_\theta(\cdot) \\
\theta &\sim p(\cdot)
\end{aligned}
$$

Add ranges for t in first two lines. Use q instead of f for better congruity with other sections? We work on a fixed time horizon $T \in \mathbb{N}$, which is necessary to implement the particle Gibbs algorithm. Can be artificially imposed by block sampling if doing online-ish inference? The measures corresponding to μ_θ, f_θ and g_θ are assumed to be known and to admit densities, and we are also given a fixed sequence of observations $y_{1:T}$. Either mention that f,g can depend on t, or make this explicit in the notation. Don't really need to include prior on theta; it's not relevant to CSMC step.

Our aim is to generate Monte Carlo samples from the joint distribution of $X_{0:T}$ and θ conditional on $y_{1:T}$. Outside of models admitting closed-form solutions, this is typically the most practical way to draw samples from the marginal distributions of either θ or any subset of the states $X_{0:T}$, by marginalising the Monte Carlo samples.

The structure of the model invites Gibbs sampling: alternating between updating θ conditional on $X_{0:T}$, and updating $X_{0:T}$ conditional on θ. "These conditionals are typically much easier to sample from that the corresponding marginals." (due to the dependence structure in the HMM). The θ update consists of sampling from

$$p(\theta \mid x_{0:T}, y_{1:T}) \propto p(\theta)p(x_{0:T}, y_{1:T} \mid \theta),$$

which can be achieved quite easily with a Metropolis–Hastings step. The X update is the 'difficult' part, requiring a sample from

$$p(x_{0:T} \mid \theta, y_{1:T}) =: \gamma_T^\theta(x_{0:T}).$$

Define gamma for general s too:

$$\gamma_s^\theta(x_{0:s}) \propto \mu_\theta(x_0)g_\theta(y_0 \mid x_0)\prod_{r=1}^{s} f_\theta(x_r \mid x_{r-1})g_\theta(y_r \mid x_r).$$

This target distribution is suited to sequential Monte Carlo, and this is where the 'particle' part of particle Gibbs comes in. We update all of the hidden states $X_{0:T}$ in one Gibbs step, which consists of drawing one sample from a particle filter. To target the correct distribution, we use conditional SMC for these updates, conditional on the sample of $X_{0:T}$ from the previous sweep.

Refer back to the CSMC algorithm which I've already written down somewhere, explaining how the inputs to the algorithm correspond to functions/quantities introduced in this setting.

2.5.3 Ancestor sampling ~

⟨sec:ancsamp⟩ 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. The following was dumped from elsewhere and NEEDS REDRAFTING.

Ancestor sampling was first suggested by Nick Whiteley in the discussion on Andrieu, Doucet, and Holenstein (2010). Its contribution is to reduce autocorrelation between samples obtained using the particle Gibbs algorithm (Andrieu, Doucet, and Holenstein 2010). Proper way to cite discussion of paper?

Ancestral degeneracy leads to poor mixing

Particle Gibbs runs into problems when the time horizon T is large compared to the number of particles N. T is determined by the application at hand, and N is limited by computational resources, so we may not be able to control their relative size. The source of the problem is ancestral degeneracy. We know that in standard SMC algorithms this problem exists and its effect is to increase the variance of our Monte Carlo estimates. In particle Gibbs, the N simulated trajectories are not used to estimate anything; only one trajectory is sampled at each step, and becomes one state a Markov chain Monte Carlo

110 *Backward Simulation in Particle MCMC*

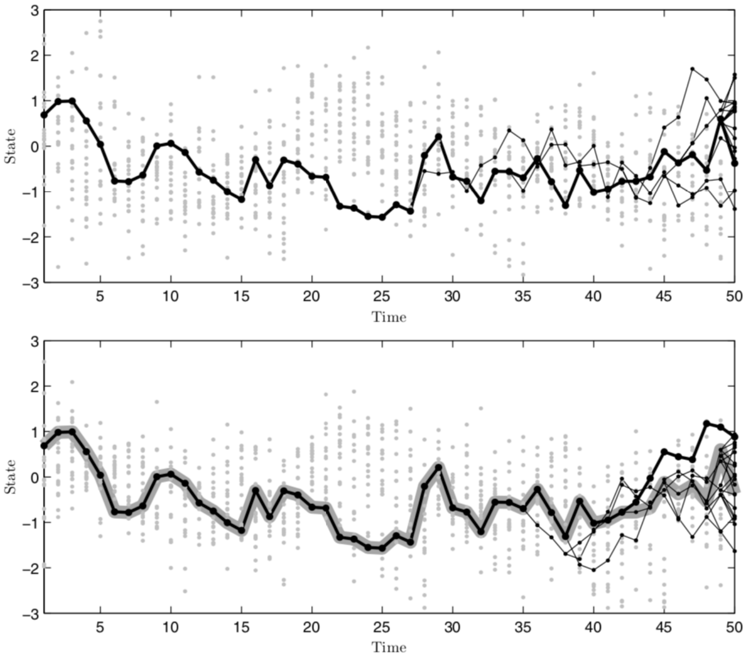


Fig. 5.3 Particle system generated by CSMC at iterations r (top) and $r+1$ (bottom) of the PG sampler. The dots show the particle positions, the thin black lines show the ancestral dependence of the particles and the thick black lines show the sampled trajectories $x_{1:T}[r]$ and $x_{1:T}[r+1]$, respectively. In the bottom pane, the thick gray line illustrates the conditioned path at iteration $r+1$, which by construction equals $x_{1:T}[r]$. Note that, due to path degeneracy, the particles shown as gray dots are not reachable by tracing any of the ancestral lineages from time T and back.

Figure 2.8: PLACEHOLDER. Copied from Lindsten and Schön (2013).

⟨fig:PG_ancdegen⟩

estimate. Ancestral degeneracy now has a less direct effect: it causes the Markov chain to mix slowly.

To see why, take a look at Figure 2.8. In Figure 2.8a we have just completed the r^{th} Gibbs sweep, sampling $\theta[r]$ and $x_{0:T}[r]$. For the $(r+1)^{th}$ sweep, we take as immortal trajectory $x^*_{0:T} = x_{0:T}[r]$, and run conditional SMC. Due to ancestral degeneracy, many of the resulting trajectories coalesce, and since the immortal trajectory must survive across the whole time window, they tend to coalesce onto the immortal trajectory (Figure 2.8b). Now we obtain the next sample $x_{0:T}[r+1]$ by sampling a trajectory among the N we have just simulated. Whichever one we choose, it has a high amount of overlap with the immortal trajectory, i.e. the previous sample $x_{0:T}[r]$. This behaviour tends to repeat at every iteration, meaning the early X coordinates are getting 'stuck' (rarely being updated). This is clearly a problem for the mixing of the Markov chain. Another way to explain this is that the variables defining the immortal trajectory (indices and states) are never refreshed during the Gibbs sweep - this is the explanation given in Lindsten Ch5. It renders the particle Gibbs algorithm impractical for any such model where the time horizon T is too large: either we must run the Markov chain for longer, or increase the number of particles N in the conditional SMC step, neither of which is feasible on a limited computational

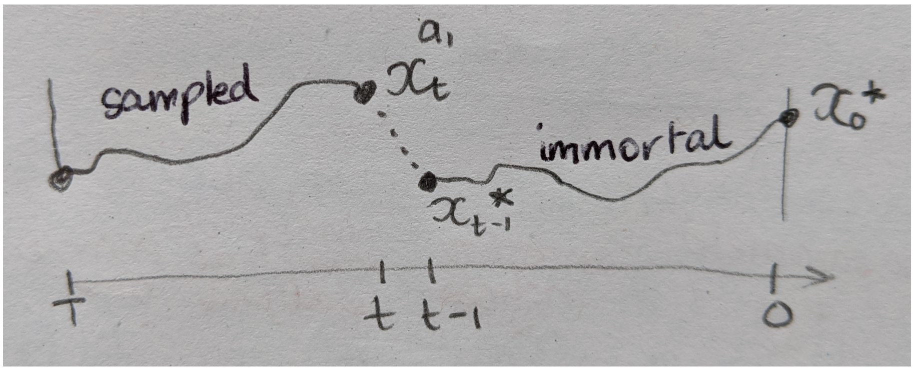


Figure 2.9: PLACEHOLDER. Interpretation of a resampling weight for the immortal offspring.

⟨fig:resample_immortal⟩

budget.

The solution: ancestor sampling

An effective solution (where it is possible to implement it) was proposed by Nick Whiteley and is known as ancestor sampling. It consists of a simple modification to the resampling step within the conditional SMC algorithm. In the basic CSMC algorithm, at each time step the particles are resampled by multinomial resampling according to their weights. That is, at each time t, each non-immortal offspring is assigned a parent as so:

$$\mathbb{P}[a_t^{(j)} = i] \propto w_t^{(i)},$$

while the immortal offspring is deterministically assigned to the immortal parent.

In ancestor sampling we do the same thing, except that the immortal particle is also resampled, rather than being deterministically assigned:

$$\mathbb{P}[a_t^{(j)} = i] \propto \begin{cases} w_t^{(i)} & \text{non-immortal particles} \\ w_t^{(i)} \frac{\gamma_T^\theta((x_{0:t-1}^{(i)}, x_{t:T}^*))}{\gamma_{t-1}^\theta(x_{0:t-1}^{(i)})} & \text{immortal particle.} \end{cases} \tag{2.10}$$ eq:ancsamp_

The ratio of γs can be interpreted as the conditional probability of the trajectory continuing with $x_{t:T}^*$ given is starts with $x_{0:t-1}^{(i)}$ (see Figure 2.9). Using the structure of the hidden Markov model defined earlier, we can rewrite the ratio

$$\frac{\gamma_T^\theta((x_{0:t-1}^{(i)}, x_{t:T}^*))}{\gamma_{t-1}^\theta(x_{0:t-1}^{(i)})} \propto f_\theta(x_t^* \mid x_{t-1}^{(i)}) g_\theta(y_t \mid x_t^*) \prod_{s=t}^{T} f_\theta(x_r^* \mid x_{r-1}^*) g_\theta(y_r \mid x_r^*) \propto f_\theta(x_t^* \mid x_{t-1}^{(i)}).$$

Should the first propto actually be equality? So it looks like it should also be pretty easy to implement ancestor sampling for our model. Write down the pseudocode to prove it? The only catch is that we need to be able to evaluate f_θ pointwise, whereas in the basic algorithm we only need to draw samples from f_θ. This will rule out ancestor sampling in some applications.

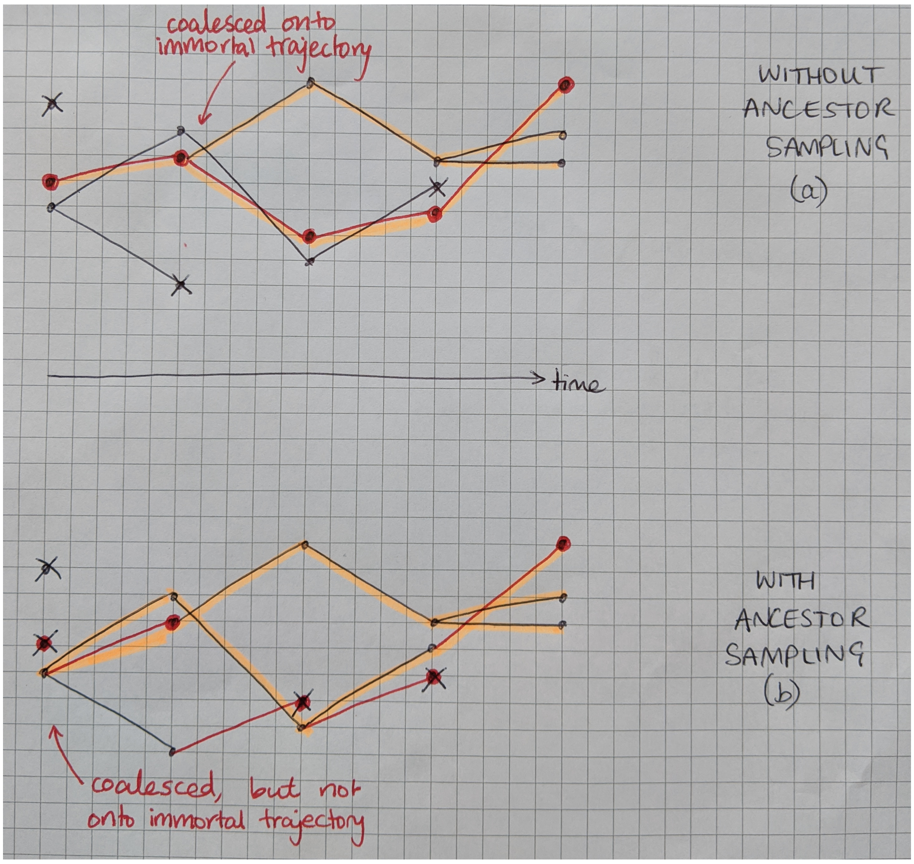


Figure 2.10: PLACEHOLDER. Illustration of how ancestor sampling prevents coalescence onto the immortal trajectory.

⟨fig:whyASworks⟩

Why ancestor sampling works

Ancestor sampling is backward sampling, but only for the immortal trajectory. (It isn't possible to do backward sampling during the forward sweep for any except the immortal trajectory - see that we can't evaluate the required γs without knowing the future states, which are known only for the immortal trajectory.) We know that backward sampling (on all trajectories, in a separate backward sweep) eradicates ancestral degeneracy. But we've only backward-sampled one trajectory, leaving the other $N-1$ trajectories to do their coalescing thing.

The important point is that ancestor sampling does not prevent ancestral degeneracy (it mitigates it a tiny bit like $1/N$). Ancestral degeneracy is pretty much as severe as ever; the difference is that the trajectories no longer coalesce preferentially onto the immortal trajectory. There is no longer an immortal trajectory to coalesce onto. An illustration of this can be seen in Figure 2.10.

Remember, the problem with ancestral degeneracy in particle Gibbs was that it induced strong autocorrelation among consecutive samples of $X_{0:T}$. It isn't a problem that the trajectories coalesce, as long as the thing they coalesce to is able to move, readily exploring the state space. This is achieved with ancestor sampling.

Why is ancestor sampling even a valid thing to do (i.e. still targetting the right thing)? Extended state space argument?(I think there is one in Lindsten Ch5). No need to go into it here unless there is a simple intuitive explanation to put the reader's mind at rest. — Just mention that it's a special application of backward sampling. Is that enough?

3 Limits ✓

3.1 Encoding genealogies

3.1.1 The genealogical process

Before we can analyse genealogies, we need a way to encode them. The encoding will only include the information relevant to the sample genealogy, namely which lineages coalesce at which times. Information about particle positions and "killed" particles is ignored.

Let $\mathcal{P}_n$ be the space of partitions on $\{1, \dots, n\}$. For convenience, we now label time in reverse, so the terminal particles are at time 0, their parents are at time 1, and so on. Consider a randomly chosen sample of n terminal particles among a total of N particles, and label the sampled particles $1, \dots, n$. The genealogical process $(G_t^{(n,N)})_{t\in\mathbb{N}_0}$ for this sample is the $\mathcal{P}_n$-valued stochastic process such that labels i and j are in the same block of the partition $G_t^{(n,N)}$ if and only if terminal particles i and j have a common ancestor at time t (i.e. t generations back).

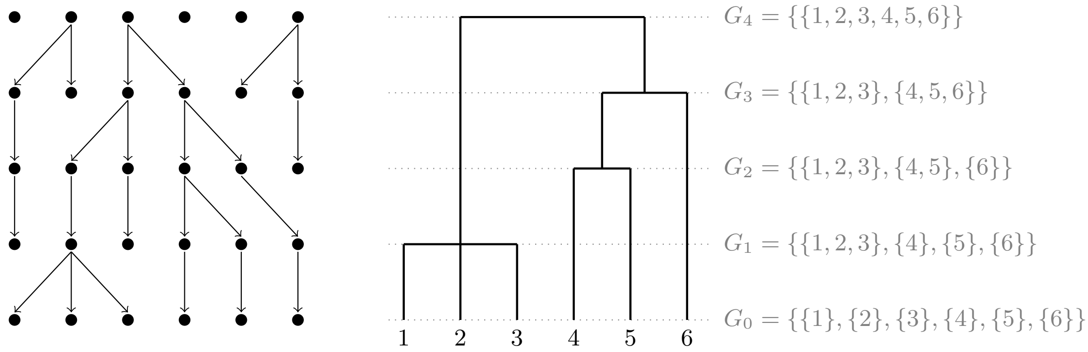
A formulation where $G_t^{(n,N)}$ takes values in the space of equivalence relations from $[n]$ to $[n]$ is sometimes used (e.g. Möhle 1999); interpreting partition blocks as equivalence classes, this formulation is equivalent to ours.

The initial value of the process is the partition of singletons $G_0^{(n,N)} = \{\{1\}, \dots, \{n\}\}$, since all of the terminal particles are in separate lineages. The only possible non-identity transitions are those that merge some blocks of the partition, encoding the coalescence of the corresponding lineages. The trivial partition $\{\{1, \dots, n\}\}$ is therefore an absorbing state, corresponding to all lineages in the sample having coalesced (i.e. the MRCA has been reached). The construction of the genealogical process from the resampling relationships (i.e. the vector of parental indices at each generation) is illustrated in Figure 3.1.

3.1.2 Time scale

In order to have a well-defined limit for the genealogical process as $N \to \infty$, we must scale time by a suitable function $\tau_N(\cdot)$. In the population genetics literature the time scale function is typically deterministic (Section 2.2.3), but 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

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



(a) Resampling relationships (b) Genealogy of terminal particles

Figure 3.1: Illustration of how the sample genealogy is encoded. (a) Relationships induced by resampling in a sample of $n = 6$ particles over four iterations. (b) The genealogy of these six particles, labelled with the value of the genealogical process G_t at each time.

⟨fig:encoding_genealogy⟩

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 a limiting pair merger rate of 1, as in the n-coalescent, we rescale time by the generalised inverse

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

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 conditional probability of a multiple merger (three or more lineages merging, or two or more simultaneous pairwise mergers):

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

This will be used to control the rate of multiple mergers, which must be dominated by the pair-merger rate as $N \to \infty$ if we are to recover a Kingman limit (in which almost surely the only non-identity transitions are pair mergers). Some basic properties of c_N, D_N and τ_N are stated in Proposition 3.1.

⟨thm:cN_properties⟩ **Proposition 3.1.** *For all $t \in \mathbb{N}$, $t' > s' > 0$,*

⟨item:cN_property1⟩

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

⟨item:cN_property2⟩

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

⟨item:cN_property3⟩

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

⟨item:cN_property4⟩

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

⟨item:cN_property5⟩

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

⟨item:cN_property6⟩

(f) $\tau_N(t') \geq t'$.

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),

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

(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.
(f) follows from (a) and the definition of τ_N in (3.2). ■

Another useful property is the following, based on Koskela et al. (2018, Lemma 2). There the special case $f(r) \equiv c_N(r)$ is proved, but the authors remark that the result also holds for other choices of f. Here we state the result in full generality.

⟨thm:kjjslemma2⟩ **Lemma 3.2.** *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. Let $f(r)$ be any deterministic function of $\nu_r^{(1:N)}$ such that for all r there exists $B < \infty$ for which $0 \le f(r) \le B$. Then*

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

Proof. Define

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

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

$$\begin{aligned}\mathbb{E}[M_{\tau_N(t)\wedge K}] &= \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} \{f(r) - \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\}\right] \\ &= \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} f(r)\right] - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)\wedge K} \mathbb{E}[f(r) \mid \mathcal{F}_{r-1}]\right] = 0.\end{aligned}$$

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

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

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 some non-negative terms). Hence

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

which concludes the proof. ■

3.1.3 Transition probabilities

Recall that $\mathcal{P}_n$ denotes the space of partitions of $\{1, \dots, n\}$. 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 $\nu_t^{(1:N)}$. The transition probability $p_{\xi\eta}(t)$ can only be non-zero when η is obtained from ξ

by merging some blocks of ξ (i.e. some lineages coalescing). 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, for each $i \in \{1, \ldots, |\eta|\}$. Hence $b_1 + \cdots + b_{|\eta|} = |\xi|$. Then the transition probability is given by

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

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 3.3 and 3.4.

⟨thm:pDelta_LB⟩ **Proposition 3.3.** *Let $\xi \in \mathcal{P}_n$, $N > 2$. Then*

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

where

$$B_{|\xi|} = K(|\xi|-1)!(|\xi|-2)\exp(2\sqrt{2(|\xi|-2)})$$

for some $K > 0$ that does not depend on $|\xi|$.

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

$$p_{\xi\xi}(t) \geq 1 - \frac{1}{(N)_{|\xi|}}\binom{|\xi|}{2} \sum_{\substack{i_1 \neq \ldots \neq i_{|\xi|-1}=1 \\ \text{all distinct}}}^{N} (\nu_t^{(i_1)})_2 \nu_t^{(i_2)} \ldots \nu_t^{(i_{|\xi|-1})}$$
$$- \frac{1}{(N)_{|\xi|}} \sum_{k=1}^{|\xi|-1} \sum_{\substack{b_1 \geq \ldots \geq b_k = 1 \\ b_1+\ldots+b_k=|\xi|}}^{|\xi|} |\xi|! \sum_{\substack{i_1 \neq \ldots \neq i_k=1 \\ \text{all distinct}}}^{N} (\nu_t^{(i_1)})_{b_1} \ldots (\nu_t^{(i_k)})_{b_k} \tag{3.5}$$

eq:pDeltaLB

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 $k, b_1, \ldots, b_k$, 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 $|\xi| = 2$, (3.5) becomes

$$p_{\xi\xi}(t) \geq 1 - c_N(t)$$

and when $|\xi| = 1$, (3.5) becomes

$$p_{\xi\xi}(t) \geq 1;$$

in both cases the result is immediate. ■

⟨thm:pDelta_UB⟩ **Proposition 3.4.** *Let $\xi \in \mathcal{P}_n$. Then, for N sufficiently large,*

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

where $B'_{|\xi|} = \binom{|\xi|-1}{2}$.

A proof is given in Koskela et al. (2018, Lemma 1 Case 1). refer to the erratum once available, which is more explicit about this proof.

3.2 An existing limit theorem

Under the assumption (A1) stated below, it is sufficient for our purposes to consider only offspring counts $\nu_t^{(1:N)} = (\nu_t^{(1)}, \dots, \nu_t^{(N)})$, where $\nu_t^{(i)} = |\{j : a_t^{(j)} = i\}|$, rather than the parental indices $a_t^{(1:N)}$ which are generally more informative.

⟨standing_assumption⟩

(A1) The conditional distribution of parental indices $a_t^{(1:N)}$ given offspring counts $\nu_t^{(1:N)}$ is uniform over all assignments such that $|\{j : a_t^{(j)} = i\}| = \nu_t^{(i)}$ for all i.

As we saw in Section 2.2, the n-coalescent is *exchangeable*, so for instance the pair of lineages merging at each event is chosen uniformly. (A1) is a weaker condition than exchangeability of the particles within a generation which is sufficient to admit an exchangeable process in the limit. Exchangeability of the particles would imply neutrality, an unreasonable assumption in the setting of SMC. In contrast, (A1) can easily be enforced upon any SMC algorithm by applying a random permutation to the offspring indices immediately after resampling.

Koskela et al. (2018) proved the following theorem which gives sufficient conditions under which sampled genealogies of (non-neutral) interacting particle systems converge to the n-coalescent as $N \to \infty$. Naturally, such a result can only be expected to hold for genealogies of finite samples ($n << N$), and not for the entire genealogy of the N particles. For instance the genealogies arising in SMC algorithms are not restricted to single pair mergers only, although within a sparse sample we may, under mild conditions, see only single pair mergers. That is to say, there is not an extension of this result whereby the whole-population genealogy converges to the Kingman coalescent as $N \to \infty$, unless very restrictive conditions are imposed.

⟨thm:kjjs_mainthm⟩ **Theorem 3.5** (Koskela et al. 2018)**.** *Fix $n \leq N$ as the observed number of particles from the output of an interacting particle system with N particles which satisfies (A1). Suppose that for any $0 \leq s < t < \infty$, we have*

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r)\right] = 0, \tag{3.6}$$ eq:kjjs_big_m

$$\lim_{N\to\infty} \mathbb{E}[c_N(t)] = 0, \tag{3.7}$$ eq:kjjs_binar

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} c_N(r)^2\right] = 0, \tag{3.8}$$ eq:kjjs_binar

$$\textit{and} \qquad \mathbb{E}[\tau_N(t) - \tau_N(s)] \leq C_{t,s} N, \tag{3.9}$$ eq:kjjs_tau_b

for some constant $C_{t,s} > 0$ that is independent of N. Then $(G^{(n,N)}_{\tau_N(t)})_{t\geq 0}$ converges to the Kingman n-coalescent in the sense of finite-dimensional distributions as $N \to \infty$.

To ensure samples of size n have Kingman genealogies in the limit, with pair mergers only, we require that multiple mergers (that is, where more than two lineages merge into one, or where two or more mergers happen simultaneously) occur on a slower time scale than pair mergers. This is the role of condition (3.6).

Conditions (3.7) and (3.8) ensure that the limiting process is continuous and has the required unit pair merger rate. For (3.7) to fail to hold, the expected number of mergers at some generation would have to be $O(N^2)$. This can only happen if the resampling scheme is very bad (e.g. star resampling) or the weights are particularly badly-behaved. The latter is ruled out in the corollaries of Chapter 5 by imposing bounds on the potential functions; this is discussed further in Section 5.1.

Condition (3.9) specifies that the time scale must be $O(N)$. As we saw in Section 2.2.3, this is the correct time scale for the Wright-Fisher model, but for instance the Moran model has time scale $O(N^2)$ and hence violates this condition. Since we know that the neutral Moran model also has Kingman genealogies in the limit, condition (3.9) clearly is not necessary. The simplified statement in Theorem 3.6 does not impose any such condition on the time scale.

The proof of Koskela et al. (2018) does not explicitly use (3.7) but rather the similar condition

$$\lim_{N\to\infty} \mathbb{E}[c_N(\tau_N(t))] = 0. \tag{3.10}$$

eq:kjjs_bin

However, as we will see in the next section (Lemmata 3.8 and 3.9), both (3.7) and (3.10) are implied by (3.6), so the theorem is correct. Such redundancies in the statement of Theorem 3.5 are removed in Theorem 3.6.

The proof of Theorem 3.5 (i.e. Koskela et al. 2018, Theorem 1) proceeds in three parts. The first is a vanishing upper bound on finite-dimensional distributions of the genealogical process when the path of the process involves any multiple mergers. The second is showing that, when the path of the genealogy consists of only single pair mergers, the finite-dimensional distributions of the n-coalescent upper bound those of the genealogical process in the limit $N \to \infty$. The final piece is a similar lower bound, which together with the upper bound establishes convergence of the finite-dimensional distributions.

3.3 A new limit theorem

We now present a related theorem, having the same conclusion but with conditions that are more tractable and remove some redundancies in the statement of Theorem 3.5. While we do not prove that this is a strict generalisation, there are certainly systems which satisfy the conditions of Theorem 3.6 but not of Theorem 3.5.

⟨thm:FDDconv⟩ **Theorem 3.6.** *Let $\nu_t^{(1:N)}$ denote the offspring numbers in an interacting particle system satisfying (A1) 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*

$$\frac{1}{(N)_3} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_3] \leq b_N \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_2] \tag{3.11}$$ eq:mainthmcond

for all N, uniformly in $t \geq 1$. Fix $n \leq N$ and consider a randomly chosen sample of n terminal particles. Then the resulting rescaled genealogical process $(G_{\tau_N(t)}^{(n,N)})_{t\geq 0}$ converges in the sense of finite-dimensional distributions to Kingman's n-coalescent as $N \to \infty$.

On the RHS of (3.11) is the filtered expectation of $c_N(t)$, i.e. the expected pair merger rate, and the LHS is the corresponding rate of triple mergers. Intuitively, (3.11) says that pair mergers dominate triple mergers, the expected rate of which vanishes as $N \to \infty$. As we will see, this implies that pair mergers also dominate all other larger mergers, such as simultaneous pair mergers.

Our result improves on Theorem 3.5 by eliminating the restrictive condition (3.9), which we know is unnecessary. This allows our result to apply to some models not previously included; for example the neutral Moran model. Although we do not prove that Theorem 3.6 is a true generalisation of Theorem 3.5, Möhle and Sagitov (2003, Theorem 5.4) showed that in neutral models the straightforward analogue of (3.11) is both necessary and sufficient, suggesting that in general this condition is not significantly stronger than (3.6)–(3.8) combined.

Our conditions are also significantly easier to verify than those of Theorem 3.5. Not only are four conditions replaced with one, but the condition (3.11) only involves marginal moments of the offspring counts, whereas (3.6) and (3.8) involve mixed moments. As we will see in Chapter 4, once we move beyond conditionally independent resampling schemes such as multinomial resampling, the joint distributions of offspring counts become complex and it may only be feasible to calculate their moments marginally. As such, we are able to verify the conditions of Theorem 3.6 in several cases, including for resampling schemes that induce strong correlations between offspring counts, whereas Koskela et al. (2018) apply their theorem only to multinomial resampling.

Our condition on the time scale, $\mathbb{P}[\tau_N(t) = \infty] = 0$, is not very restrictive. Essentially, it rules out systems in which coalescences occur at only finitely many generations. This condition is not actually necessary for Theorem 3.6 to hold, as such, but if it is violated then the limiting object is an n-coalescent under an infinite time-scaling, that is n lineages never coalescing. This would constitute a qualitatively different result and one that is of little interest for SMC, so we follow Möhle (1998) in excluding it.

3.3.1 Proof of theorem

First we prove that (3.10) and the assumptions (3.6)–(3.8) of Theorem 3.5 all follow from (3.11). Figure 3.2 illustrates how the following Lemmata 3.7–3.10 fit together. The argument differs slightly from that presented in Brown et al. (2021) in that we will here show (3.11) $\Rightarrow$ (3.6) $\Rightarrow$ (3.7) rather than (3.11) $\Rightarrow$ (3.6) and (3.11) $\Rightarrow$ (3.7). This highlights the redundancy in Theorem 3.5, where condition (3.6) directly implies two of the other stated conditions.

The second step in the proof is to show that condition (3.9) is not necessary. In particular, the parts of the proof of Koskela et al. (2018) which relied on (3.9) are rewritten using Proposition 3.3 instead. Proposition 3.3 is a lower bound on the probability of an identity transition, which holds in general without the need for further conditions, so we really are removing condition (3.9) rather than substituting it for a different condition.

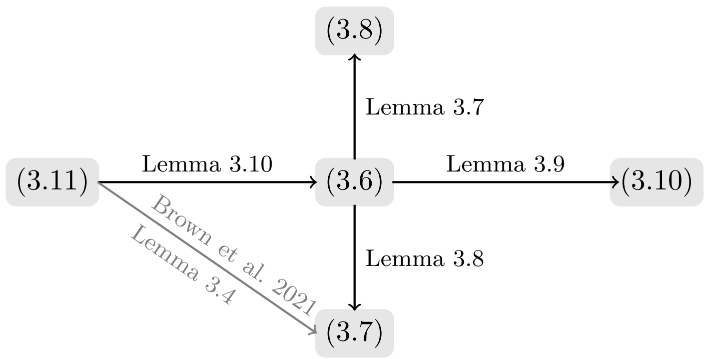


Figure 3.2: Dependencies between conditions of Theorems 3.5 and 3.6. Arrows represent logical implication; labels on arrows indicate the lemma in which the implication is stated. In Brown et al. (2021, Lemma 3.4) the direct implication (3.11) $\Rightarrow$ (3.7) was proved, but here we will instead show that (3.6) $\Rightarrow$ (3.7).

:FDD_proof_dependencies⟩

⟨lem:removeass3⟩ **Lemma 3.7.** *If for all* $0 \le s < t < \infty$

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r)\right] = 0$$

then for all $0 \le s < t < \infty$

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} c_N(r)^2\right] = 0.$$

Proof. We have

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

which is sufficient for the result. ■

⟨thm:DNimpliescN⟩ **Lemma 3.8.** *If for all* $0 \leq s < t < \infty$

$$
\lim_{N\to\infty}\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r)\right] = 0
$$

then for all $t \in \mathbb{N}$

$$
\lim_{N\to\infty}\mathbb{E}[c_N(t)] = 0.
$$

Proof. Fix $\epsilon > 0$, and assume $N > 2/\epsilon$. Following Möhle and Sagitov (2003), define the events

$$
A_i(t) := \{\nu_t^{(i)} \leq N\epsilon\}. \tag{3.12}
$$

eq:define_A

Then

$$
\begin{aligned}
c_N(t) &= \frac{1}{(N)_2}\sum_{i=1}^{N}(\nu_t^{(i)})_2[\mathbb{1}_{A_i(t)}+\mathbb{1}_{A_i(t)^c}] \\
&\leq \frac{N\epsilon}{(N)_2}\sum_{i=1}^{N}\nu_t^{(i)}+\sum_{i=1}^{N}\mathbb{1}_{A_i(t)^c} \\
&= \frac{N\epsilon}{N-1}+\sum_{i=1}^{N}\mathbb{1}_{A_i(t)^c}.
\end{aligned}
$$

Taking expectations and applying the generalised Markov inequality,

$$
\begin{aligned}
\mathbb{E}[c_N(t)] &\leq \epsilon 1_N + \sum_{i=1}^{N} \mathbb{P}[\nu_t^{(i)} > N\epsilon] \\
&\leq \epsilon 1_N + \sum_{i=1}^{N} \frac{\mathbb{E}[(\nu_t^{(i)})_3]}{(N\epsilon)_3} \\
&\leq \epsilon 1_N + \frac{N(N)_2}{(N\epsilon)_3}\mathbb{E}[D_N(t)] \\
&= \epsilon 1_N + \epsilon^{-3} 1_N \mathbb{E}[D_N(t)] \\
&\leq \epsilon 1_N + \epsilon^{-3} 1_N \mathbb{E}\left[\sum_{r=1}^{t} D_N(r)\right] \\
&\leq \epsilon 1_N + \epsilon^{-3} 1_N \mathbb{E}\left[\sum_{r=\tau_N(0)+1}^{\tau_N(t)} D_N(r)\right].
\end{aligned}
$$

Taking limits,

$$
\lim_{N\to\infty} \mathbb{E}[c_N(t)] \leq \epsilon.
$$

Since ϵ was arbitrary this concludes the proof. ■

⟨thm:DNimpliescN_2⟩ **Lemma 3.9.** *If for all* $0 \leq s < t < \infty$

$$
\lim_{N\to\infty} \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r)\right] = 0
$$

then for all $0 < t < \infty$

$$
\lim_{N\to\infty} \mathbb{E}[c_N(\tau_N(t))] = 0.
$$

Proof. Analogously to the proof of Lemma 3.8, we find

$$
\begin{aligned}
\mathbb{E}[c_N(\tau_N(t))] &\leq \epsilon 1_N + \sum_{i=1}^{N} \mathbb{P}[\nu_{\tau_N(t)}^{(i)} > N\epsilon] \\
&\leq \epsilon 1_N + \epsilon^{-3} 1_N \mathbb{E}\left[D_N(\tau_N(t))\right] \\
&\leq \epsilon 1_N + \epsilon^{-3} 1_N \mathbb{E}\left[\sum_{r=\tau_N(0)+1}^{\tau_N(t)} D_N(r)\right] \\
&\xrightarrow[N\to\infty]{} \epsilon
\end{aligned}
$$

which concludes the proof. ■

⟨lem:removeass2⟩ **Lemma 3.10.** *If there exists a deterministic sequence* $(b_N)_{N\geq 1}$ *such that* $\lim_{N\to\infty} b_N = 0$ *and*

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

for all N*, uniformly in* $t\in\mathbb{N}$*, then for all* $0\leq s<t<\infty$

$$\lim_{N\to\infty}\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r)\right] = 0.$$

Proof. We decompose $D_N(t)$ as the sum of two terms and consider their filtered expectations. The first is

$$\begin{aligned}
\frac{1}{N(N)_2}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_2\nu_t^{(i)}] &= \frac{1}{N(N)_2}\sum_{i=1}^{N}\mathbb{E}_t[2(\nu_t^{(i)})_2 + (\nu_t^{(i)})_3] \\
&\leq \frac{2}{N}\mathbb{E}_t[c_N(t)] + \frac{1}{(N)_3}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_3] \\
&\leq \left(\frac{2}{N} + b_N\right)\mathbb{E}_t[c_N(t)]. \qquad (3.13)
\end{aligned}$$ DN_part_1

The second is

$$\begin{aligned}
\frac{1}{N^2(N)_2}\sum_{j=1}^{N}\sum_{i\neq j}\mathbb{E}_t[(\nu_t^{(i)})_2(\nu_t^{(j)})^2] &= \frac{1}{N^2(N)_2}\sum_{j=1}^{N}\sum_{i\neq j}\mathbb{E}_t[(\nu_t^{(i)})_2(\nu_t^{(j)})_2 + (\nu_t^{(i)})_2\nu_t^{(j)}] \\
&\leq \frac{1}{N^2(N)_2}\sum_{j=1}^{N}\sum_{i\neq j}\mathbb{E}_t[(\nu_t^{(i)})_2(\nu_t^{(j)})_2] + \frac{\mathbb{E}_t[c_N(t)]}{N}. \qquad (3.14)
\end{aligned}$$ DN_part_2

Now, with the events $A_i(t)$ defined as in (3.12),

$$\begin{aligned}
\sum_{j=1}^{N}\sum_{i\neq j}\mathbb{E}_t\{(\nu_t^{(i)})_2(\nu_t^{(j)})_2\} &= \sum_{j=1}^{N}\sum_{i\neq j}\left\{\mathbb{E}_t[(\nu_t^{(i)})_2(\nu_t^{(j)})_2\mathbb{1}_{A_i(t)}] + \mathbb{E}_t[(\nu_t^{(i)})_2(\nu_t^{(j)})_2\mathbb{1}_{A_i(t)^c}]\right\} \\
&\leq N\epsilon\sum_{j=1}^{N}\sum_{i\neq j}\mathbb{E}_t[\nu_t^{(i)}(\nu_t^{(j)})_2\mathbb{1}_{A_i(t)}] + N^3\sum_{j=1}^{N}\sum_{i\neq j}\mathbb{E}_t[\nu_t^{(j)}\mathbb{1}_{A_i(t)^c}] \\
&\leq N^2(N)_2\epsilon\mathbb{E}_t[c_N(t)] + N^4\sum_{i=1}^{N}\mathbb{P}[\nu_t^{(i)} > N\epsilon \mid \mathcal{F}_{t-1}]. \qquad (3.15)
\end{aligned}$$ DN_part_3

For $N \geq 3/\epsilon$, by the generalised Markov inequality,

$$\sum_{i=1}^{N} \mathbb{P}(\nu_t^{(i)} > N\epsilon \mid \mathcal{F}_{t-1}) \leq \frac{1}{(N\epsilon)_3} \sum_{i=1}^{N} \mathbb{E}_t\{(\nu_t^{(i)})_3\} = \frac{\{1 + O(N^{-1})\}}{\epsilon^3 (N)_3} \sum_{i=1}^{N} \mathbb{E}_t\{(\nu_t^{(i)})_3\}$$
$$\leq \{1 + O(N^{-1})\} \frac{b_N}{\epsilon^3} \mathbb{E}_t\{c_N(t)\}. \tag{3.16}$$ markovs_ine

Substituting (3.16) into (3.15) gives

$$\sum_{j=1}^{N} \sum_{i \neq j} \mathbb{E}_t[(\nu_t^{(i)})_2 (\nu_t^{(j)})_2] \leq N^4 (1 + O(N^{-1})) \left(\epsilon + \frac{b_N}{\epsilon^3} \right) \mathbb{E}_t[c_N(t)] \tag{3.17}$$ DN_part_4

and substituting (3.17) into (3.14) gives

$$\frac{1}{N^2 (N)_2} \sum_{j=1}^{N} \sum_{i \neq j} \mathbb{E}_t[(\nu_t^{(i)})_2 (\nu_t^{(j)})^2] \leq \left[(1 + O(N^{-1})) \left(\epsilon + \frac{b_N}{\epsilon^3} \right) + \frac{1}{N} \right] \mathbb{E}_t[c_N(t)]. \tag{3.18}$$ DN_last

Combining (3.13) and (3.18), we have that

$$\mathbb{E}_t[D_N(t)] = \left[(1 + O(N^{-1})) \left(\epsilon + \frac{b_N}{\epsilon^3} \right) + \frac{3}{N} + b_N \right] \mathbb{E}_t[c_N(t)].$$

Finally, invoking Lemma 3.2 twice gives

$$\begin{aligned}
\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} D_N(r) \right] &= \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} \mathbb{E}_r[D_N(r)] \right] \\
&\leq \left\{ (1 + O(N^{-1})) \left(\epsilon + \frac{b_N}{\epsilon^3} \right) + \frac{3}{N} + b_N \right\} \mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)} c_N(r) \right] \\
&\leq \left\{ (1 + O(N^{-1})) \left(\epsilon + \frac{b_N}{\epsilon^3} \right) + \frac{3}{N} + b_N \right\} (t - s + 1) \\
&\underset{N \to \infty}{\longrightarrow} \epsilon(t - s + 1),
\end{aligned}$$

and recalling that $\epsilon > 0$ was arbitrary concludes the proof. ■

Update KJJS references in the following to point to relevant places in the erratum. The proof is similar to `Dropbox/SMC_Genealogies/Asymptotic_genealogies_of_interacting_particle_` `Post_acceptance_correction/Erratum/AOS/Round_3/Indicators/full_redraft.pdf`, from page 22. We need to refer to the KJJS erratum (once published), rather than the arXiv version which is not correct (lacking some indicators) and doesn't contain all the required results, then update equation and lemma numbers etc. as required.

To complete the proof of Theorem 3.6 it remains to show that condition (3.9) is unnecessary. We will show that Proposition 3.3 can be used instead of (3.9) to obtain the same result. The only part of Koskela et al. (2018, Proof of Theorem 1) making use of condition

(3.9) is the lower bound on finite-dimensional distributions of the genealogical process for paths involving single pair mergers only. A slight modification of the argument allows a similar lower bound to be obtained via Proposition 3.3 such that as $N \to \infty$ the bound coincides with the corresponding finite-dimensional distributions of the n-coalescent, as required. The modified section of the proof is presented below, using the notation of Koskela et al. (2018) for ease of comparison. Maybe I should actually write out a full proof of the new theorem?

Proof. Let $\chi_d^\star$ be the conditional transition probability of a transition from state η_{d-1} to state η_d at times $\tau_N(t_{d-1})$ and $\tau_N(t_d)$ respectively, conditional on the offspring counts between those times $\nu^{(1:N)}_{\tau_N(d-1)+1}, \dots, \nu^{(1:N)}_{\tau_N(d)}$. This transition can happen via any valid path of merger events, but we restrict to paths involving binary mergers only, and denote by χ_d the conditional transition probability subject to this restriction. Compared to Koskela et al. (2018, Proof of Theorem 1), the derivation of an upper bound on χ_d holds without modification, while the first step in the derivation of a lower bound (Koskela et al. 2018, p.14) involves the application of Koskela et al. (2018, Lemma 1 Case 1) to bound χ_d from below and the subsequent application of (3.9). Instead, we apply Proposition 3.3 to obtain, for sufficiently large N,

$$\chi_d \geq \sum_{\substack{s_1 < \ldots < s_\alpha \\ = \tau_N(t_{d-1})+1}}^{\tau_N(t_d)} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \left(\prod_{r=1}^{\alpha} \mathbb{1}_{\{c_N(s_r) > \binom{n-2}{2} D_N(s_r)\}} \left[c_N(s_r) - \binom{n-2}{2} 1_N D_N(s_r) \right] \right)$$
$$\times \prod_{\substack{r=\tau_N(t_{d-1})+1 \\ r \neq s_1, \ldots, r \neq s_\alpha}}^{\tau_N(t_d)} \left[1 - \tilde{B}_n 1_N D_N(r) - \binom{|\eta_{d-1}| - |\{i : s_i < r\}|}{2} 1_N c_N(r) \right]$$
$$\times \mathbb{1}_{\{c_N(r) < (\tilde{B}_n + \binom{n}{2})^{-1}\}}.$$

Here $\tilde{Q}$ is the matrix obtained from the generator Q of Kingman's n-coalescent (see Definition 2.1) by setting the diagonal entries to 0. The number of pair-merger steps required to transition from η_{d-1} to η_d is $\alpha = |\eta_{d-1}| - |\eta_d|$. The sequences $s_1, \dots, s_\alpha$ denote the times at which these pair-mergers happen. At the remaining times r the partition is unchanged, and the bound of Proposition 3.3 has been applied to the one-step transition probabilities corresponding to these identity transitions. The constant is $\tilde{B}_n := B_n \binom{n}{2}$ where B_n is the constant defined in Proposition 3.3, and we have replaced $|\eta_d|$ by its upper bound n.

The rest of the proof proceeds as in Koskela et al. (2018), albeit from this modified initial lower bound. A multinomial expansion of the product on the second line, noting that $(1_N)^a = 1_N$ for any $a \in \mathbb{R}$, yields

$$\chi_d \geq \left(\prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} \mathbb{1}_{\{c_N(r) > \binom{n-2}{2} D_N(r)\}} \mathbb{1}_{\{c_N(r) < (\tilde{B}_n + \binom{n}{2}))^{-1}\}} \right)$$
$$\times \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \sum_{\substack{(\lambda,\mu)\in\Pi_2([\alpha+\beta]):\\ |\lambda|=\alpha}} 1_N$$
$$\times \sum_{\substack{s_1<\ldots<s_{\alpha+\beta}\\ =\tau_N(t_{d-1})+1}}^{\tau_N(t_d)} \left(\prod_{r\in\lambda} \left[c_N(s_r) - \binom{n-2}{2} 1_N D_N(s_r) \right] \right)$$
$$\times \prod_{r\in\mu} \left\{ -\binom{|\eta_{d-1}| - |\{i\in\lambda : i<r\}|}{2} c_N(s_r) - \tilde{B}_n D_N(s_r) \right\}$$

where $\Pi_i([n])$ denotes the set of partitions of $\{1,\ldots,n\}$ into exactly i blocks. Expanding the product over λ gives

$$\chi_d \geq \left(\prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} \mathbb{1}_{\{c_N(r) > \binom{n-2}{2} D_N(r)\}} \mathbb{1}_{\{c_N(r) < (\tilde{B}_n + \binom{n}{2}))^{-1}\}} \right)$$
$$\times \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \sum_{\substack{(\lambda,\mu,\pi)\in\Pi_3([\alpha+\beta]):\\ |\mu|=\beta}} \binom{n-2}{2}^{|\pi|} (-1)^{|\pi|} 1_N$$
$$\times \sum_{\substack{s_1<\ldots<s_{\alpha+\beta}\\ =\tau_N(t_{d-1})+1}}^{\tau_N(t_d)} \left\{ \prod_{r\in\lambda} c_N(s_r) \right\} \left\{ \prod_{r\in\pi} D_N(s_r) \right\}$$
$$\times \prod_{r\in\mu} \left\{ -\binom{|\eta_{d-1}| - |\{i\in\lambda\cup\pi : i<r\}|}{2} c_N(s_r) - \tilde{B}_n D_N(s_r) \right\}$$

and expanding the product over μ results in

$$\chi_d \geq \left(\prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} \mathbb{1}_{\{c_N(r) > \binom{n-2}{2} D_N(r)\}} \mathbb{1}_{\{c_N(r) < (\tilde{B}_n + \binom{n}{2}))^{-1}\}} \right)$$
$$\times \sum_{\beta=0}^{\tau_N(t_d)-\tau_N(t_{d-1})-\alpha} (\tilde{Q}^\alpha)_{\eta_{d-1}\eta_d} \sum_{\substack{(\lambda,\mu,\pi,\sigma)\in\Pi_4([\alpha+\beta]):\\ |\mu|+|\sigma|=\beta}} \tilde{B}_n^{|\sigma|} \binom{n-2}{2}^{|\pi|} (-1)^{|\pi|+|\sigma|}$$
$$\times 1_N \left\{ \prod_{r\in\mu} - \binom{|\eta_{d-1}| - |\{i\in\lambda\cup\pi : i<r\}|}{2} \right\}$$
$$\times \sum_{\substack{s_1<\ldots<s_{\alpha+\beta}\\ =\tau_N(t_{d-1})+1}}^{\tau_N(t_d)} \left\{ \prod_{r\in\lambda\cup\mu} c_N(s_r) \right\} \prod_{r\in\pi\cup\sigma} D_N(s_r).$$

Via a further multinomial expansion, the lower bound for the k-step transition probability can be written as

$$
\begin{aligned}
\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \geq \lim_{N\to\infty} \mathbb{E}\Bigg[& \left(\prod_{r=\tau_N(t_0)+1}^{\tau_N(t_k)} \mathbb{1}_{\{c_N(r) > \binom{n-2}{2} D_N(r)\}} \mathbb{1}_{\{c_N(r) < (\tilde{B}_n + \binom{n}{2})^{-1}\}} \right) \\
& \times \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \sum_{\substack{(\lambda_1,\mu_1,\pi_1,\sigma_1)\in\Pi_4([\alpha_1+\beta_1]):\\ |\mu_1|+|\sigma_1|=\beta_1}} \cdots \sum_{\substack{(\lambda_k,\mu_k,\pi_k,\sigma_k)\in\Pi_4([\alpha_k+\beta_k]):\\ |\mu_k|+|\sigma_k|=\beta_k}} \\
& \tilde{B}_n^{\sum_{d=1}^{k}|\sigma_d|} \binom{n-2}{2}^{\sum_{d=1}^{k}|\pi_d|} (-1)^{\sum_{d=1}^{k}|\pi_d|+|\sigma_d|} 1_N \\
& \times \left\{ \prod_{d=1}^{k} (\tilde{Q}^{\alpha_d})_{\eta_{d-1}\eta_d} \prod_{r\in\mu_d} -\binom{|\eta_{d-1}| - |\{i \in \lambda_d \cup \pi_d : i < r\}|}{2} \right\} \\
& \times \sum_{\substack{s_1^{(1)}<\ldots<s_{\alpha_1+\beta_1}^{(1)}\\ =\tau_N(t_0)+1}}^{\tau_N(t_1)} \cdots \sum_{\substack{s_1^{(k)}<\ldots<s_{\alpha_k+\beta_k}^{(k)}\\ =\tau_N(t_{k-1})+1}}^{\tau_N(t_k)} \\
& \prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}} \left\{ \prod_{r\in\lambda_d\cup\mu_d} c_N(s_r^{(d)}) \right\} \prod_{r\in\pi_d\cup\sigma_d} D_N(s_r^{(d)}) \Bigg].
\end{aligned}
$$

An argument completely analogous to that in Koskela et al. (2018, Appendix) shows that passing the expectation and the limit through the infinite sums is justified, whereupon the contribution of terms with $\sum_{d=1}^{k}(|\pi_d| + |\sigma_d|) > 0$ vanishes. To see why, follow the argument used to show that the contribution of multiple merger trajectories vanishes in the corresponding upper bound in Koskela et al. (2018). That leaves

$$
\begin{aligned}
\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \geq & \sum_{\beta_1=0}^{\infty} \cdots \sum_{\beta_k=0}^{\infty} \sum_{\substack{(\lambda_1,\mu_1)\in\Pi_2([\alpha_1+\beta_1]):\\ |\mu_1|=\beta_1}} \cdots \sum_{\substack{(\lambda_k,\mu_k)\in\Pi_2([\alpha_k+\beta_k]):\\ |\mu_k|=\beta_k}} \\
& \left\{ \prod_{d=1}^{k} (\tilde{Q}^{\alpha_d})_{\eta_{d-1}\eta_d} \prod_{r\in\mu_d} -\binom{|\eta_{d-1}| - |\{i \in \lambda_d \cup \pi_d : i < r\}|}{2} \right\} \\
& \times \lim_{N\to\infty} \mathbb{E}\Bigg[\left(\prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} \mathbb{1}_{\{c_N(r) > \binom{n-2}{2} D_N(r)\}} \mathbb{1}_{\{c_N(r) < (\tilde{B}_n + \binom{n}{2})^{-1}\}} \right) \\
& \times \sum_{\substack{s_1^{(1)}<\ldots<s_{\alpha_1+\beta_1}^{(1)}\\ =\tau_N(t_0)+1}}^{\tau_N(t_1)} \cdots \sum_{\substack{s_1^{(k)}<\ldots<s_{\alpha_k+\beta_k}^{(k)}\\ =\tau_N(t_{k-1})+1}}^{\tau_N(t_k)} \\
& \prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}} \left\{ \prod_{r\in\lambda_d\cup\mu_d} c_N(s_r^{(d)}) \right\} \Bigg].
\end{aligned}
\tag{3.19}
$$

eq1

Recall (Koskela et al. 2018, Eq (11)):

$$\sum_{\substack{(\lambda,\mu)\in\Pi_2([\alpha+\beta]):\\ |\mu|=\beta}} (\tilde{Q}^{\alpha})_{\eta_{d-1}\eta_d} \prod_{r\in\mu} -\binom{|\eta_{d-1}| - |\{i\in\lambda\cup\pi : i<r\}|}{2} = (Q^{\alpha+\beta})_{\eta_{d-1}\eta_d}.$$

Applying this k times in (3.19) yields

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}\left[\prod_{d=1}^{k}\chi_d\right] \geq \sum_{\beta_1=0}^{\infty}\cdots\sum_{\beta_k=0}^{\infty}&\left\{\prod_{d=1}^{k}(Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d}\right\}\\ &\times\lim_{N\to\infty}\mathbb{E}\Bigg\{\left(\prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)}\mathbb{1}_{\{c_N(r)>\binom{n-2}{2}D_N(r)\}}\mathbb{1}_{\{c_N(r)<(\tilde{B}_n+\binom{n}{2})^{-1}\}}\right)\\ &\qquad\times\left(\prod_{d=1}^{k}\mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\right)\\ &\qquad\times\sum_{\substack{s_1^{(1)}<\ldots<s^{(1)}_{\alpha_1+\beta_1}\\=\tau_N(t_0)+1}}^{\tau_N(t_1)}\cdots\sum_{\substack{s_1^{(k)}<\ldots<s^{(k)}_{\alpha_k+\beta_k}\\=\tau_N(t_{k-1})+1}}^{\tau_N(t_k)}\prod_{d=1}^{k}\prod_{r\in\lambda_d\cup\mu_d}c_N(s_r^{(d)})\Bigg\}.\end{aligned}$$

We now apply equations (14) and (15), respectively, of Koskela et al. (2018), to those terms with a negative ($|\boldsymbol{\beta}|$ odd) and positive ($|\boldsymbol{\beta}|$ even) sign, respectively, to obtain

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}\left[\prod_{d=1}^{k}\chi_d\right] &\geq \sum_{\beta_1=0}^{\infty}\cdots\sum_{\beta_k=0}^{\infty}\left\{\prod_{d=1}^{k}(Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d}\frac{(t_d-t_{d-1})^{\alpha_d+\beta_d}}{(\alpha_d+\beta_d)!}\right\}\\ &\qquad\times\lim_{N\to\infty}\mathbb{E}\Bigg[\left(\prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)}\mathbb{1}_{\{c_N(r)>\binom{n-2}{2}D_N(r)\}}\mathbb{1}_{\{c_N(r)<(\tilde{B}_n+\binom{n}{2})^{-1}\}}\right)\\ &\qquad\qquad\times\left(\prod_{d=1}^{k}\mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1})\geq\alpha_d+\beta_d\}}\right)\Bigg]\\ &\geq \sum_{\beta_1=0}^{\infty}\cdots\sum_{\beta_k=0}^{\infty}\left\{\prod_{d=1}^{k}(Q^{\alpha_d+\beta_d})_{\eta_{d-1}\eta_d}\frac{(t_d-t_{d-1})^{\alpha_d+\beta_d}}{(\alpha_d+\beta_d)!}\right\}\end{aligned}$$

where the expectation of the indicators converges to 1 due to Koskela et al. (2018, Equation (16)) and Lemma 4.12 and Lemma 4.11. Or refer to Koskela et al. (2018, Equation (16)) and Lemma 4 in the appendix of `full_redraft.pdf`. ■

4 Weak Convergence ✓

In this chapter we present a weak convergence result which is identical to Theorem 3.6 except that the mode of convergence is strengthened from convergence of the finite-dimensional distributions to weak convergence. Weak convergence is desirable because it implies convergence of a strictly larger class of functions of genealogies, granting access to the distributions of statistics such as the time to the sample MRCA, the total branch length, and the probability that the MRCA of a subsample is equal to the sample MRCA okay, technically if this one is going to be a "statistic", I'm talking about the indicator on this event.

The extension from Theorem 3.6 to weak convergence requires an additional tightness argument. The proof is rather long-winded since we do not make such strong simplifying assumptions on the dynamics of the interacting particle system as are seen for example in Möhle (1999) and others...?. The proof is broken down into a series of technical results which culminate in Theorem 4.1. The overall structure of the proof is depicted graphically in Figure 4.1.

We start by defining a suitable metric space. Let $\mathcal{P}_n$ be the space of partitions of $\{1, \dots, n\}$. 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 discrete metric,

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

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

⟨thm:weakconv⟩ **Theorem 4.1.** *Let $\nu_t^{(1:N)}$ denote the offspring numbers in an interacting particle system satisfying (A1) 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*

$$\frac{1}{(N)_3} \sum_{i=1}^{N} \mathbb{E}_t \left[(\nu_t^{(i)})_3 \right] \leq b_N \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}_t \left[(\nu_t^{(i)})_2 \right] \tag{4.1}$$ eq:mainthmcon

almost surely 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 of Theorem 4.1. 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? To make it digestible, the proof is broken down into a number of results which are organised into sections; the relationships between these are shown in Figure 4.1.

Since we already have convergence of the finite-dimensional distributions (Theorem 3.6), 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) provide 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 their 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:

⟨item:relcomp1⟩

1. For every $\epsilon > 0$, $t \geq 0$ there exists a compact set $\Gamma \subseteq \mathcal{P}_n$ such that

$$\liminf_{N\to\infty} \mathbb{P}[G^{(n,N)}_{\tau_N(t)} \in \Gamma] \geq 1 - \epsilon$$

⟨item:relcomp2⟩

2. For every $\epsilon > 0$, $t > 0$ there exists $\delta > 0$ such that

$$\liminf_{N\to\infty} \mathbb{P}[\omega(G^{(n,N)}_{\tau_N(\cdot)}, \delta, t) < \epsilon] \geq 1 - \epsilon$$

   where ω is the modified modulus of continuity:

$$\omega(G^{(n,N)}_{\tau_N(\cdot)}, \delta, t) := \inf \max_{i\in[K]} \sup_{u,v\in[T_{i-1},T_i)} \rho\left(G^{(n,N)}_{\tau_N(u)}, G^{(n,N)}_{\tau_N(v)}\right)$$

   with the infimum taken over all partitions of the form $0 = T_0 < T_1 < \cdots < T_{K-1} < t \leq T_K$ (for some K) such that $\min_{i\in[K]}(T_i - T_{i-1}) > \delta$.

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 $\{\{1\}, \dots, \{n\}\}$. For a proof that the maximum is attained at $\xi = \Delta$, see Lemma 4.2. Following Möhle (1999), we now construct the two-dimensional Markov process $(Z_t, S_t)_{t\in\mathbb{N}_0}$ on $\mathbb{N}_0 \times \mathcal{P}_n$ with transition probabilities

$$\mathbb{P}[Z_t = j, S_t = \eta \mid Z_{t-1} = i, S_{t-1} = \xi, \mathcal{F}_\infty] \\ = \begin{cases} 1 - p_t & \text{if } j = i \text{ and } \xi = \eta \\ p_{\xi\xi}(t) + p_t - 1 & \text{if } j = i+1 \text{ and } \xi = \eta \\ p_{\xi\eta}(t) & \text{if } j = i+1 \text{ and } \xi \neq \eta \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$ eq:construct

and initial state $Z_0 = 0$, $S_0 = \Delta$. Unlike the corresponding process in Möhle (1999), in our case the transition probabilities depend on offspring counts, thus the process is only Markovian conditional on $\mathcal{F}_\infty$. It can be thought of as a Markov process in a random environment.

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 of (4.2) 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 fixed $\varpi_1^{(N)}, \dots, \varpi_m^{(N)}$ and $t > 0$, 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 $\delta > 0$,

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

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

$$\liminf_{N\to\infty} \mathbb{P}[T_m^{(N)} \geq t, \varpi_i^{(N)} > \delta \, \forall i \in \{1, \dots, m\}] \geq 1 - \epsilon. \tag{4.3}$$ eq:conditio

Since $T_m^{(N)} = \varpi_1^{(N)} + \dots + \varpi_m^{(N)}$, there is a positive correlation between $T_m^{(N)}$ and each of the $\varpi_i^{(N)}$, and the $\varpi_i^{(N)}$'s are independent conditionally... should all these probabilities be conditioned on $\mathcal{F}_\infty$?, so

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

Due to Lemma 4.3, the limiting distributions of $\varpi_i^{(N)}$ are i.i.d. $\mathrm{Exp}(\alpha_n)$, so

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

and

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

using the series expansion for the Erlang CDF (see for example Forbes et al. 2011, Chapter 15). Hence

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

which can be made $\geq 1-\epsilon$ by taking m sufficiently large and δ sufficiently small. Since this argument applies for any ϵ and t, (4.3) and hence Condition 2 is satisfied, and the proof is complete. ■

⟨thm:maximum_pr⟩ **Lemma 4.2.** $\max_{\xi \in \mathcal{P}_n} (1 - p_{\xi\xi}(t)) = 1 - p_{\Delta\Delta}(t)$.

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. Setting $\eta = \xi$ in (3.4),

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

Similarly,

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

Discarding the zero summands,

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

The inner sum is

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

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

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

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. ∎

⟨thm:holdingtimes_distn⟩ **Lemma 4.3.** *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 $t_1, \dots, t_k > 0$,

$$\lim_{N\to\infty} \mathbb{P}[T_1^{(N)} \le t_1, \dots, T_k^{(N)} \le t_k] = \mathbb{P}[T_1 \le t_1, \dots, T_k \le t_k]. \tag{4.4}$$ eq:518

Take the basis case $k = 1$. Then

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

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

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

Lemma 4.7 shows that this probability converges to $e^{-\alpha_n t}$ as required.

For the induction step, assume that (4.4) holds for some k. We have the following decomposition:

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

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

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

As shown in Möhle (1999), the RHS

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

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

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

that is, there are jumps at some times $r_1, \dots, r_k$ and identity transitions at all other times. Lemmata 4.8 and 4.9 show that this probability converges to the correct limit. This completes the induction. ■

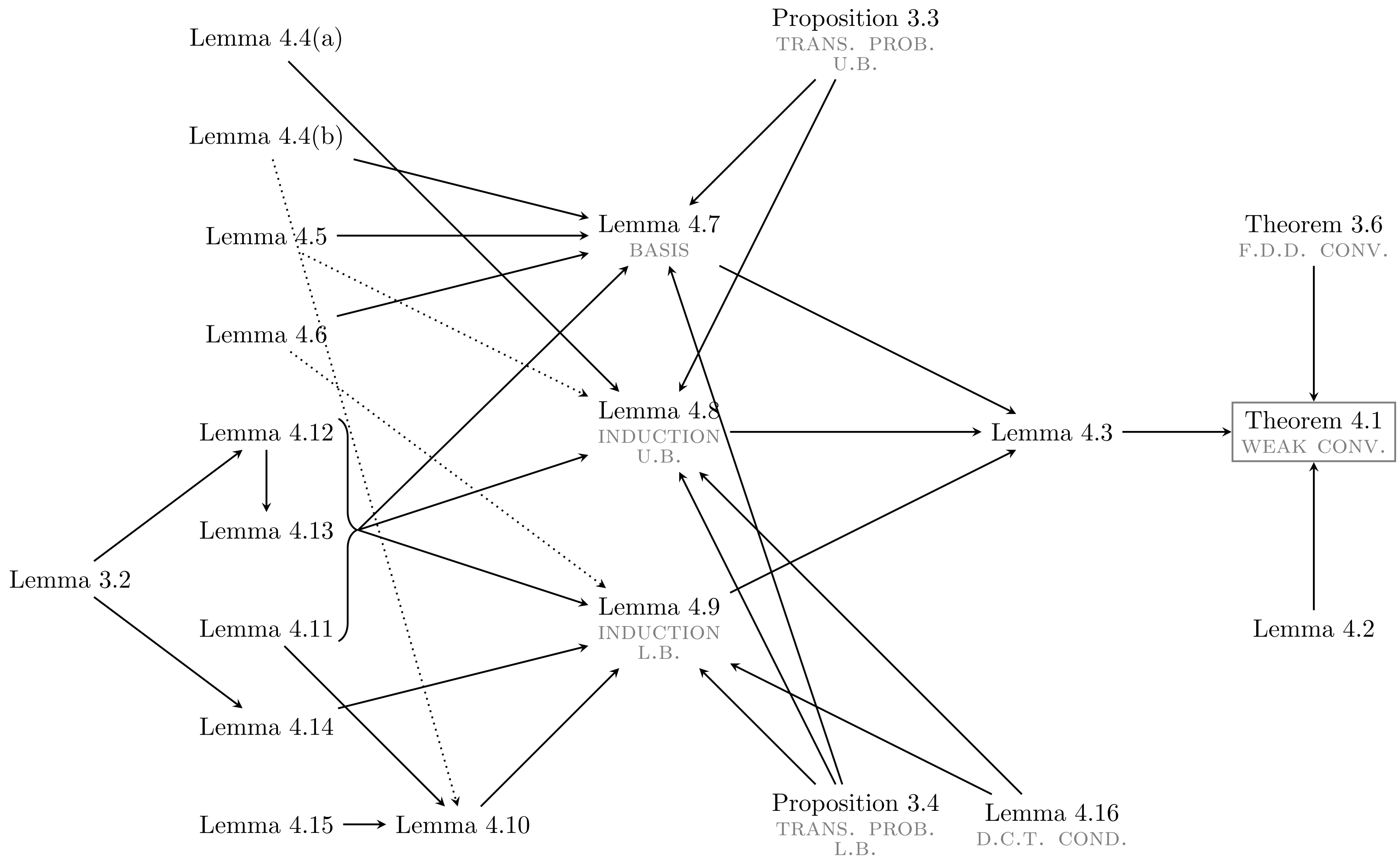


Figure 4.1: Graph showing dependencies between the lemmata used to prove weak convergence. Dotted arrows indicate dependence via a slight modification of the preceding lemma. Dependencies preceding Theorem 3.6 are not shown; these are shown in Figure ?? draw a corresponding figure for fdd proof, or delete this sentence?.

eakconv_structure⟩

4.1 Bounds on sum-products

⟨thm:sumprod1⟩ **Lemma 4.4.** *Fix $t > 0$, $l \in \mathbb{N}$.*

⟨thm:sumprod1_a⟩
(a)
$$\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \leq (t+1)^l$$

⟨thm:sumprod1_b⟩
(b)
$$t^l - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right) \binom{l}{2} (t+1)^{l-2} \leq \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \leq t^l + c_N(\tau_N(t))(t+1)^l$$

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

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

as can be seen by considering the multinomial expansion of the RHS. Applying (4.16),

$$\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \leq (t+1)^l. \tag{4.5}$$ eq:039

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

$$\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) \geq \left(\sum_{s=0}^{\tau_N(t)} c_N(s) \right)^l - \binom{l}{2} \left(\sum_{s=0}^{\tau_N(t)} c_N(s)^2 \right) \left(\sum_{s=0}^{\tau_N(t)} c_N(s) \right)^{l-2}. \tag{4.6}$$ eq:002

Applying (4.16) on the RHS of (4.6) yields the lower bound.

For the upper bound we have

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

using the definition of τ_N. A binomial expansion yields

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

then by (4.14),

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

Putting this together yields the upper bound. ■

⟨thm:sumprod2⟩ **Lemma 4.5.** *Fix $t > 0$, $l \in \mathbb{N}$. Let B be a positive constant which may depend on n.*

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

Proof. We start with a binomial expansion:

$$\begin{aligned}\sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) + BD_N(s_j)] &= \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \sum_{\mathcal{I} \subseteq [l]} B^{l-|\mathcal{I}|} \left(\prod_{i \in \mathcal{I}} c_N(s_i) \right) \left(\prod_{j \notin \mathcal{I}} D_N(s_j) \right) \\ &= \sum_{\mathcal{I} \subseteq [l]} B^{l-|\mathcal{I}|} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i \in \mathcal{I}} c_N(s_i) \right) \left(\prod_{j \notin \mathcal{I}} D_N(s_j) \right) \end{aligned} \tag{4.7}$$

eq:010

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 $\mathcal{I} = \{1, \ldots, |\mathcal{I}|\}$:

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

Separating the term $I = l$,

$$\begin{aligned}&\sum_{I=0}^{l} \binom{l}{I} B^{l-I} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i) \right) \left(\prod_{j=I+1}^{l} D_N(s_j) \right) \\ &\quad = \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} c_N(s_j) + \sum_{I=0}^{l-1} \binom{l}{I} B^{l-I} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \left(\prod_{i=1}^{I} c_N(s_i) \right) \left(\prod_{j=I+1}^{l} D_N(s_j) \right).\end{aligned} \tag{4.8}$$

eq:012

In the second term on the RHS, there is always at least one D_N term, so using (4.15) we can write

$$\begin{aligned}
\sum_{I=0}^{l-1}\binom{l}{I}B^{l-I}\sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)}\left(\prod_{i=1}^{I}c_N(s_i)\right)&\left(\prod_{j=I+1}^{l}D_N(s_j)\right)\\
&\leq\sum_{I=0}^{l-1}\binom{l}{I}B^{l-I}\sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)}\left(\prod_{i=1}^{l-1}c_N(s_i)\right)D_N(s_l)\\
&\leq\sum_{I=0}^{l-1}\binom{l}{I}B^{l-I}\left(\sum_{s_1\neq\cdots\neq s_{l-1}}^{\tau_N(t)}\prod_{i=1}^{l-1}c_N(s_i)\right)\sum_{s_l=1}^{\tau_N(t)}D_N(s_l)\\
&\leq\sum_{I=0}^{l-1}\binom{l}{I}B^{l-I}(t+1)^{l-1}\sum_{s=1}^{\tau_N(t)}D_N(s) \qquad (4.9)\ \boxed{\texttt{eq:013}}
\end{aligned}$$

using (4.5). Finally, by the Binomial Theorem,

$$\sum_{I=0}^{l-1}\binom{l}{I}B^{l-I}(t+1)^{l-1}\sum_{s=1}^{\tau_N(t)}D_N(s)\leq\left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)(t+1)^{l-1}(1+B)^l, \qquad (4.10)\ \boxed{\texttt{eq:014}}$$

which, together with (4.8), concludes the proof. ■

⟨thm:sumprod3⟩ **Lemma 4.6.** *Fix $t>0$, $l\in\mathbb{N}$. Let B be a positive constant which may depend on n.*

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

Proof. A binomial expansion and subsequent manipulation as in (4.7)–(4.8) gives

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

where the last inequality just multiplies some positive terms by -1. Then (4.9)–(4.10) can be applied directly (noting that an upper bound on negative terms gives a lower bound overall):

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

which concludes the proof. ■

4.2 Main components of induction argument

Recall that the following conditions are all consequences of (4.1): for all $t>s>0$,

$$\mathbb{E}\left[c_N(\tau_N(t))\right]\to 0 \tag{4.11}$$
eq:BJJK_eq3

$$\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)}c_N(r)^2\right]\to 0 \tag{4.12}$$
eq:BJJK_eq3

$$\mathbb{E}\left[\sum_{r=\tau_N(s)+1}^{\tau_N(t)}D_N(r)\right]\to 0 \tag{4.13}$$
eq:BJJK_eq3

as $N\to\infty$. Also recall the following properties from Proposition 3.1:

$$c_N(t),D_N(t)\in[0,1] \tag{4.14}$$
eq:cN_prope

$$D_N(t)\leq c_N(t) \tag{4.15}$$
eq:cN_prope

$$t'\leq\sum_{r=1}^{\tau_N(t')}c_N(r)\leq t'+1. \tag{4.16}$$
eq:cN_prope

⟨thm:basis⟩ **Lemma 4.7** (Basis step)**.** *Assume* (4.1) *holds. For any* $0<t<\infty$*,*

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

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

Proof. We start by showing that $\lim_{N\to\infty}\mathbb{E}\left[\prod_{r=1}^{\tau_N(t)}(1-p_r)\right]\leq e^{-\alpha_n t}$. Setting $\xi=\Delta$ in Proposition 3.4, we have for each r

$$1-p_r=p_{\Delta\Delta}(r)\leq 1-\alpha_n 1_N\left[c_N(r)-B_n'D_N(r)\right]. \tag{4.17}$$
eq:018

When $N\geq 3$, a sufficient condition to ensure the bound in (4.17) is non-negative is that the event

$$E_N^1(r):=\left\{c_N(r)<\alpha_n^{-1}A_N\right\} \tag{4.18}$$
eq:defn_E1

occurs, where $A_N = 1_N$ as $N \to \infty$ and is independent of r but will not be specified explicitly. We will also need to control the sign of $c_N(r) - B_n' D_N(r)$, for which we define the event

$$E_N^2(r) := \left\{ c_N(r) \geq B_n' D_N(r) \right\}, \tag{4.19}$$ eq:defn_E2

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

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

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

$$\begin{aligned}
\prod_{r=1}^{\tau_N(t)} (1 - p_r) &\leq 1 + \sum_{l=1}^{\tau_N(t)} (-\alpha_n)^l 1_N \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} \left[c_N(s_j) - B_n' D_N(s_j) \right] \mathbb{1}_{E_N^1 \cap E_N^2} \\
&= 1 + \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} \left[c_N(s_j) - B_n' D_N(s_j) \right] \mathbb{1}_{E_N^1 \cap E_N^2} \\
&\quad - \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!} \sum_{s_1 \neq \cdots \neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} \left[c_N(s_j) - B_n' D_N(s_j) \right] \mathbb{1}_{E_N^1 \cap E_N^2}.
\end{aligned} \tag{4.20}$$ eq:019

This is further bounded by applying Lemma 4.6 and then both bounds of Lemma 4.4(b):

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

Collecting some terms,

$$
\begin{aligned}
\prod_{r=1}^{\tau_N(t)} (1-p_r) \le 1 &+ \sum_{l=1}^{\tau_N(t)} (-\alpha_n)^l 1_N \frac{1}{l!} t^l \mathbb{1}_{E_N^1 \cap E_N^2} + c_N(\tau_N(t)) \sum_{\substack{l=2 \\ \text{even}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!} (t+1)^l \\
&+ \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right) \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!} \binom{l}{2} (t+1)^{l-2} \\
&+ \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) \sum_{\substack{l=1 \\ \text{odd}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!} (t+1)^{l-1} (1+B_n')^l \\
\le 1 &+ \sum_{l=1}^{\infty} (-\alpha_n)^l 1_N \frac{1}{l!} t^l \mathbb{1}_{\{\tau_N(t) \ge l\}} \mathbb{1}_{E_N^1 \cap E_N^2} + c_N(\tau_N(t)) \exp[\alpha_n 1_N (t+1)] \\
&+ \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right) \frac{1}{2} \alpha_n^2 \exp[\alpha_n 1_N (t+1)] \\
&+ \left(\sum_{s=1}^{\tau_N(t)} D_N(s) \right) \exp[\alpha_n 1_N (t+1)(1+B_n')]. \qquad (4.21) \ \boxed{\texttt{eq:021}}
\end{aligned}
$$

Now, taking the expectation and limit, then applying (4.11)–(4.13), and Lemmata 4.12, 4.13 and 4.14 to deal with the indicators,

$$
\begin{aligned}
\lim_{N\to\infty} \mathbb{E}\left[\prod_{r=1}^{\tau_N(t)} (1-p_r) \right] &\le 1 + \sum_{l=1}^{\infty} (-\alpha_n)^l \frac{1}{l!} t^l \lim_{N\to\infty} \mathbb{P}\left[\{\tau_N(t) \ge l\} \cap E_N^1 \cap E_N^2 \right] \\
&\quad + \lim_{N\to\infty} \mathbb{E}\left[c_N(\tau_N(t)) \right] \exp[\alpha_n (t+1)] \\
&\quad + \lim_{N\to\infty} \mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} c_N(s)^2 \right] \frac{1}{2} \alpha_n^2 \exp[\alpha_n (t+1)] \\
&\quad + \lim_{N\to\infty} \mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} D_N(s) \right] \exp[\alpha_n (t+1)(1+B_n')] \\
&= 1 + \sum_{l=1}^{\infty} (-\alpha_n)^l \frac{1}{l!} t^l = e^{-\alpha_n t}. \qquad (4.22) \ \boxed{\texttt{eq:022}}
\end{aligned}
$$

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

It remains to show the corresponding lower bound

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

Setting $\xi = \Delta$ in Proposition 3.3, we have

$$1 - p_t = p_{\Delta\Delta}(t) \geq 1 - \frac{N^{n-2}}{(N-2)_{n-2}} \alpha_n [c_N(t) + B_n D_N(t)] \tag{4.23}$$ eq:pDeltaDe

where $B_n > 0$. Due to (4.15), a sufficient condition for this bound to be non-negative is

$$E_N^3(r) := \left\{ c_N(r) \leq \frac{(N-2)_{n-2}}{N^{n-2}} \alpha_n^{-1} (1 + B_n)^{-1} \right\}, \tag{4.24}$$ eq:defn_E3

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:

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

This is further bounded by applying Lemma 4.5 and both bounds in Lemma 4.4(b):

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

Collecting terms,

$$
\begin{aligned}
\prod_{r=1}^{\tau_N(t)}(1-p_r) \geq \sum_{l=0}^{\tau_N(t)}(-\alpha_n)^l 1_N \frac{1}{l!}t^l \mathbb{1}_{E_N^3} &- \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right) \sum_{\substack{l=2\\ \text{even}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!}\binom{l}{2}(t+1)^{l-2} \\
&- c_N(\tau_N(t)) \sum_{\substack{l=1\\ \text{odd}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!}(t+1)^l \\
&- \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right) \sum_{\substack{l=1\\ \text{odd}}}^{\tau_N(t)} \alpha_n^l 1_N \frac{1}{l!}(t+1)^{l-1}(1+B_n)^l \\
\geq \sum_{l=0}^{\infty}(-\alpha_n)^l 1_N \frac{1}{l!}t^l \mathbb{1}_{E_N^3}\mathbb{1}_{\{\tau_N(t)\geq l\}} &- \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right)\frac{1}{2}\alpha_n^2 \exp[\alpha_n 1_N(t+1)] \\
&- c_N(\tau_N(t)) \exp[\alpha_n 1_N(t+1)] \\
&- \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right) \exp[\alpha_n 1_N(t+1)(1+B_n)]. \qquad (4.25)\ \boxed{\texttt{eq:028}}
\end{aligned}
$$

Now, taking the expectation and limit, and applying (4.11)–(4.13) to show that all but the first sum vanish, and Lemmata 4.13 and 4.12 to show that $\lim_{N\to\infty}\mathbb{P}[\{\tau_N(t)\geq l\}\cap E_N^3] = 1$,

$$
\begin{aligned}
\lim_{N\to\infty}\mathbb{E}\left[\prod_{r=1}^{\tau_N(t)}(1-p_r)\right] \geq \sum_{l=0}^{\infty}(-\alpha_n)^l 1_N \frac{1}{l!}t^l &\lim_{N\to\infty}\mathbb{P}\left[\{\tau_N(t)\geq l\}\cap E_N^3\right] \\
&- \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right]\frac{1}{2}\alpha_n^2\exp[\alpha_n(t+1)] \\
&- \lim_{N\to\infty}\mathbb{E}\left[c_N(\tau_N(t))\right]\exp[\alpha_n(t+1)] \\
&- \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} D_N(s)\right]\exp[\alpha_n(t+1)(1+B_n)] \\
= \sum_{l=0}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l = e^{-\alpha_n t}. &\qquad (4.26)\ \boxed{\texttt{eq:029}}
\end{aligned}
$$

Again, passing the limit and expectation inside the infinite sum is justified by dominated convergence and Fubini. Combining the upper and lower bounds in (4.22) and (4.26) respectively concludes the proof. ■

⟨thm:inductionUB⟩ **Lemma 4.8** (Induction step upper bound)**.** *Assume* (4.1) *holds. Fix* $k \in \mathbb{N}$*,* $i_0 := 0$*,* $i_k := k$*. For any sequence of times* $0 = t_0 \le t_1 \le \cdots \le t_k \le t$*,*

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

Proof. We use the bound on $(1-p_r)$ from (4.17) and apply a multinomial expansion, defining as in (4.18) and (4.19) respectively the sequences of events E_N^1 and E_N^2 which ensure the bounds are non-negative:

$$\begin{aligned}
\prod_{\substack{r=1 \\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r) &\le \prod_{\substack{r=1 \\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} \left\{ 1 - \alpha_n 1_N [c_N(r) - B_n' D_N(r)] \mathbb{1}_{E_N^1 \cap E_N^2} \right\} \\
&= 1 + \sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l 1_N \frac{1}{l!} \sum_{\substack{s_1\neq\cdots\neq s_l \\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) - B_n' D_N(s_j)] \mathbb{1}_{E_N^1 \cap E_N^2} \\
&= 1 + \sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l 1_N \frac{1}{l!} \sum_{s_1\neq\cdots\neq s_l}^{\tau_N(t)} \prod_{j=1}^{l} [c_N(s_j) - B_n' D_N(s_j)] \mathbb{1}_{E_N^1 \cap E_N^2} \\
&\quad - \sum_{l=1}^{\tau_N(t)-k} (-\alpha_n)^l 1_N \frac{1}{l!} \sum_{\substack{s_1\neq\cdots\neq s_l \\ \exists i,i': s_i = r_{i'}}} \prod_{j=1}^{l} [c_N(s_j) - B_n' D_N(s_j)] \mathbb{1}_{E_N^1 \cap E_N^2}.
\end{aligned} \tag{4.27}$$

eq:031

The penultimate line above is exactly the expansion we had in the basis step (4.20), except for the limit on l, and as such following the same arguments gives a bound analogous to that in (4.21):

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

For the last line of (4.27),

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

where the penultimate inequality uses Lemma 4.4(a). Putting these together, we have

$$
\begin{aligned}
\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)}(1-p_r) \leq 1 + \sum_{l=1}^{\tau_N(t)-k}(-\alpha_n)^l 1_N \frac{1}{l!} t^l \mathbb{1}_{E_N^1\cap E_N^2} + c_N(\tau_N(t))\exp[\alpha_n 1_N(t+1)]& \\
+ \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right)\frac{1}{2}\alpha_n^2 \exp[\alpha_n 1_N(t+1)]& \\
+ \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right)\exp[\alpha_n 1_N(t+1)(1+B_n')]& \\
+ \left(\sum_{j=1}^{k} c_N(r_j)\right)\alpha_n(1+B_n')\exp[\alpha_n 1_N(1+B_n')(t+1)].& \qquad (4.28)
\end{aligned}
$$

eq:034b

Meanwhile, using the bound on p_r from (4.23) then applying a modification of Lemma 4.5

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

$$\begin{aligned}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^k p_{r_i} &\le \alpha_n^k 1_N \sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^k [c_N(r_i)+B_nD_N(r_i)]\\ &\le \alpha_n^k 1_N \sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^k c_N(r_i) + \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right)\alpha_n^k 1_N (t+1)^{k-1}(1+B_n)^k.\end{aligned} \tag{4.29}$$ eq:035

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

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

which, using Lemma 4.4(a), also leads to the deterministic bound

$$\begin{aligned}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^k p_{r_i} &\le \alpha_n^k 1_N (1+B_n)^k \sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^k c_N(r_i)\\ &\le \alpha_n^k 1_N (1+B_n)^k \frac{1}{k!}\sum_{r_1\ne\cdots\ne r_k}^{\tau_N(t)}\prod_{i=1}^k c_N(r_i)\\ &\le \alpha_n^k 1_N (1+B_n)^k \frac{1}{k!}(t+1)^k.\end{aligned} \tag{4.30}$$ eq:037

Combining (4.28) with the other product, the expression inside the expectation in Lemma 4.8

is bounded above by

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

Applying the various bounds (4.29)–(4.30), we have

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

Upon taking the expectation and limit, we have

$$
\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right] \\
&\qquad \le \alpha_n^k \lim_{N\to\infty} \mathbb{E}\left[\left(1+\sum_{l=1}^{\tau_N(t)-k}(-\alpha_n)^l \frac{1}{l!}t^l \mathbb{1}_{E_N^1\cap E_N^2}\right)\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k} c_N(r_i)\right] \\
&\qquad\quad + \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} D_N(s)\right]\alpha_n^k(t+1)^{k-1}(1+B_n)^k\exp[\alpha_n t] \\
&\qquad\quad + \Bigg\{\lim_{N\to\infty}\mathbb{E}\left[c_N(\tau_N(t))\right]\exp[\alpha_n(t+1)] + \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right]\frac{1}{2}\alpha_n^2\exp[\alpha_n(t+1)] \\
&\qquad\qquad + \lim_{N\to\infty}\mathbb{E}\left[\sum_{s=1}^{\tau_N(t)} D_N(s)\right]\exp[\alpha_n(t+1)(1+B_n')]\Bigg\}\alpha_n^k(1+B_n)^k\frac{1}{k!}(t+1)^k \\
&\qquad\quad + \exp[\alpha_n(1+B_n')(t+1)]\alpha_n^{k+1}(1+B_n')(1+B_n)^k \\
&\qquad\qquad \times \lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\sum_{j=1}^{k} c_N(r_j)\prod_{i=1}^{k} c_N(r_i)\right]. \qquad (4.31)\ \boxed{\texttt{eq:043}}
\end{aligned}
$$

The middle terms vanish due to (4.11)–(4.13) and the expression becomes

$$
\begin{aligned}
&\lim_{N\to\infty} \mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right] \le \alpha_n^k \lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k} c_N(r_i)\right] \\
&\qquad + \alpha_n^k\sum_{l=1}^{\infty}(-\alpha_n)^l\frac{1}{l!}t^l \lim_{N\to\infty}\mathbb{E}\left[\mathbb{1}_{\{\tau_N(t)\ge k+l\}}\mathbb{1}_{E_N^1\cap E_N^2}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k} c_N(r_i)\right] \\
&\qquad + \exp[\alpha_n(1+B_n')(t+1)]\alpha_n^{k+1}(1+B_n')(1+B_n)^k \\
&\qquad\quad \times \lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\sum_{j=1}^{k} c_N(r_j)\prod_{i=1}^{k} c_N(r_i)\right], \qquad (4.32)\ \boxed{\texttt{eq:040}}
\end{aligned}
$$

where passing the limit and expectation inside the infinite sum is justified by dominated convergence and Fubini; see Lemma 4.16. To simplify the last line,

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

using Lemma 4.4(a) for the final inequality. Hence

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

by (4.12). By Lemmata 4.13, 4.12 and 4.14, $\lim_{N\to\infty}\mathbb{P}[\{\tau_N(t)\ge k+l\}\cap E_N^1\cap E_N^2] = 1$, so we can apply Lemma 4.10 to the remaining expectations in (4.32), yielding

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

as required. ■

⟨thm:inductionLB⟩ **Lemma 4.9** (Induction step lower bound)**.** *Assume* (4.1) *holds. Fix* $k \in \mathbb{N}$, $i_0 := 0$, $i_k := k$. *For any sequence of times* $0 = t_0 \le t_1 \le \cdots \le t_k \le t$,

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

Proof. Firstly,

$$\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right) \left(\prod_{\substack{r=1\\ \notin\{r_1,\dots,r_k\}}}^{\tau_N(t)} (1-p_r) \right) \ge \sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right) \left(\prod_{r=1}^{\tau_N(t)} (1-p_r) \right). \tag{4.33}$$ eq:032a

Now the second product does not depend on $r_1,\dots,r_k$, and we can use the lower bound from (4.25):

$$\begin{aligned}\prod_{r=1}^{\tau_N(t)} (1-p_r) \ge \sum_{l=0}^{\tau_N(t)} (-\alpha_n)^l 1_N \frac{1}{l!} t^l \mathbb{1}_{E_N^3} - \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right) \frac{1}{2}\alpha_n^2 \exp[\alpha_n 1_N (t+1)] \\ - c_N(\tau_N(t)) \exp[\alpha_n 1_N(t+1)] \\ - \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right) \exp[\alpha_n 1_N (t+1)(1+B_n)]\end{aligned} \tag{4.34}$$ eq:033a

where E_N^3 is defined as in (4.24). We will also need an upper bound on this product, which is formed from (4.21) with a further deterministic bound:

$$\begin{aligned}\prod_{r=1}^{\tau_N(t)} (1-p_r) &\le \sum_{l=0}^{\tau_N(t)} (-\alpha_n)^l 1_N \frac{1}{l!} t^l \mathbb{1}_{\{\tau_N(t)\ge l\}} \mathbb{1}_{E_N^1\cap E_N^2} + c_N(\tau_N(t)) \exp[\alpha_n 1_N (t+1)] \\ &\quad + \left(\sum_{s=1}^{\tau_N(t)} c_N(s)^2\right) \frac{1}{2}\alpha_n^2 \exp[\alpha_n 1_N(t+1)] \\ &\quad + \left(\sum_{s=1}^{\tau_N(t)} D_N(s)\right) \exp[\alpha_n 1_N(t+1)(1+B_n')] \\ &\le \exp[\alpha_n 1_N t] + \exp[\alpha_n 1_N (t+1)] \\ &\quad + \frac{1}{2}\alpha_n^2 (t+1)\exp[\alpha_n 1_N(t+1)] + (t+1)\exp[\alpha_n 1_N (t+1)(1+B_n')] \\ &\le \left(2 + \frac{\alpha_n^2(t+1)}{2}\right) \exp[\alpha_n 1_N (t+1)] + (t+1)\exp[\alpha_n 1_N(t+1)(1+B_n')].\end{aligned} \tag{4.35}$$ eq:034a

Now let us consider the remaining sum-product on the RHS of (4.33). We use the same bound on p_r as in (4.17):

$$p_r = 1 - p_{\Delta\Delta}(r) \geq \alpha_n 1_N \left[c_N(r) - B_n' D_N(r)\right] \tag{4.36}$$ eq:050a

where the $O(N^{-1})$ term does not depend on r. When N is large enough for the factor of 1_N to be non-negative, the condition that the bound in (4.36) is non-negative holds on the event E_N^2 that was defined in (4.19). Then

$$\prod_{i=1}^{k} p_{r_i} \geq \alpha_n^k 1_N \prod_{i=1}^{k} \left[c_N(r_i) - B_n' D_N(r_i)\right] \mathbb{1}_{E_N^2}.$$

Applying a modification of Lemma 4.6 where the sum is over ordered indices rather than distinct indices,

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

The above expression is already split into positive and negative terms; a lower bound on (4.33) can be formed by multiplying the positive terms by the lower bound (4.34) and the negative terms by the upper bound (4.35). Thus

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

Due to (4.11)–(4.13), all but the first line on the RHS of the above have vanishing expectation, leaving

$$\begin{aligned}
&\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\ldots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right]\\
&\quad\ge \lim_{N\to\infty}\mathbb{E}\left[\alpha_n^k \mathbb{1}_N \sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^{k} c_N(r_i)\mathbb{1}_{E_N^2}\sum_{l=0}^{\tau_N(t)}(-\alpha_n)^l \mathbb{1}_N \frac{1}{l!}t^l \mathbb{1}_{E_N^3}\right]\\
&\quad= \alpha_n^k \sum_{l=0}^{\infty}(-\alpha_n)^l \frac{1}{l!}t^l \lim_{N\to\infty}\mathbb{E}\left[\mathbb{1}_{\{\tau_N(t)\ge l\}}\mathbb{1}_{E_N^2\cap E_N^3}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \prod_{i=1}^{k} c_N(r_i)\right].
\end{aligned} \tag{4.37}$$

eq:056

Passing the limit and expectation inside the infinite sum is justified by dominated convergence and Fubini; see Lemma 4.16. Lemmata 4.12 and 4.14 establish that $\lim_{N\to\infty}\mathbb{P}[E_N^2\cap E_N^3] = 1$ and Lemma 4.13 deals with the other indicator. We can therefore apply

Lemma 4.10 to conclude that

$$\lim_{N\to\infty} \mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}} \left(\prod_{i=1}^{k} p_{r_i}\right)\left(\prod_{\substack{r=1\\ \notin\{r_1,\ldots,r_k\}}}^{\tau_N(t)} (1-p_r)\right)\right]$$

$$\ge \alpha_n^k \sum_{l=0}^{\infty}(-\alpha_n)^l \frac{1}{l!}t^l \sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\ldots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k} \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}$$

$$= \alpha_n^k e^{-\alpha_n t} \sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\ldots,k\}:\\ i_j\ge j\forall j}} \prod_{j=1}^{k} \frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}$$

as required. ■

⟨thm:induction_sumprodcN⟩ **Lemma 4.10.** *Assume (4.1) holds. 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 $0=t_0\le t_1\le\cdots\le t_k\le t$,*

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

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

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

By a modification of the upper bound in Lemma 4.4(b) where the lower limit of the sum is a general time rather than 1,

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

Now, taking the product on the outside,

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

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

Lemma 4.10:

$$\lim_{N\to\infty}\mathbb{E}\left[\mathbb{1}_{E_N}\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\right]\le\lim_{N\to\infty}\mathbb{E}\left[\sum_{\substack{r_1<\cdots<r_k:\\ r_i\le\tau_N(t_i)\forall i}}\prod_{i=1}^{k}c_N(r_i)\right]$$

$$\le\sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}}\prod_{j=1}^{k}\frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}+\sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}}\sum_{\mathcal{I}\subset[k]}\lim_{N\to\infty}\mathbb{E}\left[c_N(\tau_N(t_{j^\star(\mathcal{I})}))\right](t+1)^k$$

$$=\sum_{\substack{i_1\le\cdots\le i_{k-1}\\ \in\{0,\dots,k\}:\\ i_j\ge j\forall j}}\prod_{j=1}^{k}\frac{(t_j-t_{j-1})^{i_j-i_{j-1}}}{(i_j-i_{j-1})!}$$

using (4.11)

For the corresponding lower bound, by a modification of the lower bound in Lemma 4.4(b) where the lower limit of the sum is a general time rather than 1,

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

$$\ge(t_j-t_{j-1})^{i_j-i_{j-1}}-\binom{i_j-i_{j-1}}{2}\left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)}c_N(s)^2\right)(t_j-t_{j-1}+1)^{i_j-i_{j-1}-2}$$

$$\ge(t_j-t_{j-1})^{i_j-i_{j-1}}-(i_j-i_{j-1})!\left(\sum_{s=\tau_N(t_{j-1})+1}^{\tau_N(t_j)}c_N(s)^2\right)(t_j-t_{j-1}+1)^{i_j-i_{j-1}-2}.$$

Define the events

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

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 4.15.) Now, taking a product over j,

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

$$
\begin{aligned}
&\geq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} \\
&\qquad - \sum_{\mathcal{I} \subset [k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2 \right) \left(\prod_{j \in \mathcal{I}} t^{i_j - i_{j-1}} \right) \left(\prod_{j \notin \mathcal{I}} (t+1)^{i_j - i_{j-1} - 1} \right) \\
&\geq \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} - \sum_{\mathcal{I} \subset [k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2 \right) \prod_{j=1}^{k} (t+1)^{i_j - i_{j-1}} \\
&= \prod_{j=1}^{k} \frac{(t_j - t_{j-1})^{i_j - i_{j-1}}}{(i_j - i_{j-1})!} \mathbb{1}_{E_N^4} - \sum_{\mathcal{I} \subset [k]} \left(\sum_{s=\tau_N(t_{j^\star(\mathcal{I})-1})+1}^{\tau_N(t_{j^\star(\mathcal{I})})} c_N(s)^2 \right) (t+1)^k,
\end{aligned}
$$

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

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

where for the last equality we use (4.12) to show that the second sum vanishes and Lemma 4.15 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. ■

4.3 Indicators

⟨thm:lim_AandB⟩ **Lemma 4.11.** *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.

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

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 $0 \leq \mathbb{P}[A_N^c \cup B_N^c] \leq \mathbb{P}[A_N^c] + \mathbb{P}[B_N^c]$). ■

⟨thm:indicators_cN⟩ **Lemma 4.12.** *Assume* (4.12) *holds. 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.

$$\begin{aligned}
\mathbb{P}[E_N] &= 1 - \mathbb{P}[E_N^c] = 1 - \mathbb{P}\left[\bigcup_{r=1}^{\tau_N(t)} E_N^c(r)\right] = 1 - \mathbb{E}\left[\mathbb{1}_{\bigcup E_N^c(r)}\right] \geq 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{1}_{E_N^c(r)}\right] \\
&= 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{E}\left[\mathbb{1}_{E_N^c(r)} \mid \mathcal{F}_{r-1}\right]\right] = 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{P}\left[E_N^c(r) \mid \mathcal{F}_{r-1}\right]\right]
\end{aligned} \tag{4.38}$$ eq:034

where for the second line we apply Lemma 3.2 with $f(r) = \mathbb{1}_{E_N^c(r)}$. By the generalised Markov inequality,

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

Substituting this into (4.38) and applying Lemma 3.2 again, this time with $f(r) = c_N(r)^2$,

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

Applying (4.12), the limit is

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

as required. ∎

⟨thm:indicators_tau⟩ **Lemma 4.13.** *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 4.12:

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

Hence

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

by applying Lemma 4.12 with $K = t/l$. ∎

⟨thm:indicators_DN⟩ **Lemma 4.14.** *Assume (4.13) holds. Fix $t > 0$. Let K be a constant not depending on N, r, but which may depend on n.*

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

Proof.

$$\begin{aligned}
\mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \{c_N(r) \geq K D_N(r)\} \right] &\geq \mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \{c_N(r) > K D_N(r)\} \right] \\
&= 1 - \mathbb{P}\left[\bigcup_{r=1}^{\tau_N(t)} \{c_N(r) \leq K D_N(r)\} \right] \\
&= 1 - \mathbb{E}\left[\mathbb{1}_{\bigcup \{c_N(r) \leq K D_N(r)\}} \right] \\
&\geq 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{1}_{\{c_N(r) \leq K D_N(r)\}} \right] \\
&= 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{P}[c_N(r) \leq K D_N(r) \mid \mathcal{F}_{r-1}] \right] \qquad (4.39)
\end{aligned}$$

eq:050

where the final inequality is an application of Lemma 3.2 with $f(r) = \mathbb{1}_{\{c_N(r) \le K D_N(r)\}}$.

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

$$
\begin{aligned}
D_N(r) = & \frac{1}{N(N)_2} \sum_{i=1}^{N} (\nu_r^{(j)})_2 \left[\nu_r^{(i)} + \frac{1}{N} \sum_{j \ne i} (\nu_r^{(i)})^2 \right] \mathbb{1}_{A_i^c(r)} \\
& + \frac{1}{N(N)_2} \sum_{i=1}^{N} (\nu_r^{(i)})_2 \left[\nu_r^{(i)} + \frac{1}{N} \sum_{j \ne i} (\nu_r^{(j)})^2 \right] \mathbb{1}_{A_i(r)}.
\end{aligned}
$$

For the first term,

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

For the second term,

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

Altogether we have

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

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

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

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

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

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

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

Applying Lemma 3.2 once more, with $f(r) = D_N(r)$,

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

due to (4.13). Substituting this back into (4.39) concludes the proof. ■

⟨thm:indicators_c2⟩ **Lemma 4.15.** *Assume* (4.12) *holds. 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*

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

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

Proof.

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

Applying Markov's inequality,

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

by (4.12). ■

4.4 Fubini & dominated convergence conditions

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 4.8 at (4.31) and once in Lemma 4.9 at (4.37).

⟨thm:DCT_Fubini⟩ **Lemma 4.16.** *For any fixed $t > 0$,*

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

Proof.

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

■

5 Applications ✓

⟨ch:appl⟩ Theorem 4.1 gives verifiable conditions under which interacting particle systems with dynamics in the form of Algorithm 1 have asymptotically Kingman genealogies. The work was motivated by SMC algorithms, which have the required form. However, certain choices of state space and dynamics within the context of Algorithm 1 yield systems that are not very SMC-like but may have applications in other fields such as population genetics. For instance, we have generally imagined that the resampling scheme is unbiased, but this is by no means necessary for Theorem 4.1 (or indeed Theorem 3.6); it is just that biased resampling schemes are of little use in SMC.

The applications presented in this chapter are all motivated by SMC, but an interesting area of future research would be to explore the implications of Theorem 4.1 in other contexts. From the population genetics point-of-view, Theorem 4.1 may be seen as a complement to the convergence criteria for neutral models (e.g. Möhle 1999) discussed in Section ?? add the section reference once that part of Chapter 2 is written, so it would be interesting to construct some corollaries for classical non-neutral population models.

5.1 Multinomial resampling

⟨sec:corol_mn⟩ Multinomial resampling is often preferred in theoretical studies of SMC, because it renders the parental indices conditionally i.i.d. given the weights, making it relatively simple to analyse the resulting algorithm. The convergence of finite-dimensional distributions for multinomial resampling was proved in Koskela et al. (2018, Corollary 1), but we are now able to prove an analogous weak convergence result. The following proof also demonstrates the relative ease with which we can verify Theorem 3.6 as opposed to Koskela et al. (2018, Theorem 1).

⟨thm:multinomial⟩ **Corollary 5.1.** *Consider an SMC algorithm using multinomial resampling, such that (A1) 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,*

$$\frac{1}{a} \leq g_t(x,x') \leq a, \quad \varepsilon h(x') \leq q_t(x,x') \leq \frac{1}{\varepsilon} h(x'). \tag{5.1}$$ eq:gq_bounds_

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 among a total of N particles. Then, for any fixed n, the time-scaled genealogy $(G_{\tau_N(t)}^{(n,N)})_{t\geq 0}$ converges weakly to Kingman's n-coalescent as $N \to \infty$.

The bounds on g_t and q_t in (5.1) are rather strong; they can only reasonably be expected to hold if the state space is compact. However, they are widespread in the literature, where they are known as the *strong mixing conditions* (Del Moral 2004, Section 3.5.2), because they greatly facilitate the theoretical analysis of SMC algorithms. It is often possible to relax these conditions at the expense of considerable technical complication. The conditions on g_t in (5.1) ensure that the weights are all $O(N^{-1})$, none of them being too close to zero or one. Together with the bounds on q_t, this is enough to control the relative rate of multiple mergers, as seen in the following proof.

Proof. Recall that the sequence of σ-algebras

$$\mathcal{H}_t := \sigma(X_{t-1}^{(1:N)}, X_t^{(1:N)}, w_{t-1}^{(1:N)}, w_t^{(1:N)}) \tag{5.2}$$ eq:defn_Ht

are such that $\nu_t^{(1:N)}$ is conditionally independent of the filtration $\mathcal{F}_{t-1}$ given $\mathcal{H}_t$. Conditional on $\mathcal{H}_t$ the parental indices are independent, with conditional law

$$\mathbb{P}\left[a_t^{(i)} = a_i \mid \mathcal{H}_t\right] \propto g_t(X_{t+1}^{a_{t+1}^{(a_i)}}, X_t^{(a_i)}) q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}) \tag{5.3}$$ eq:parentsl

for each i, so the joint law is

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

Using the bounds (5.1) and the balls-in-bins coupling of Koskela et al. (2018, Proof of Lemma 3), we can obtain bounds on expectations of functions of $a_t^{(1:N)}$. For any $k \in \mathbb{N}$ the function $a_t^{(1:N)} \to (\nu_t^{(i)})_k$ is $\{i\}$-increasing in the sense of Koskela et al. (2018), so we may apply the bounds

$$\mathbb{E}[(V_1^{(i)})_k] \leq \mathbb{E}[(\nu_t^{(i)})_k \mid \mathcal{H}_t] \leq \mathbb{E}[(V_2^{(i)})_k],$$

where

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

independently for each i and independently of $\mathcal{F}_\infty$. Furthermore, using the moments of the Binomial distribution (see for example Mosimann 1962, p. 67)

$$\mathbb{E}[(V_1^{(i)})_k] = (N)_k\left(\frac{\varepsilon/a}{(\varepsilon/a)+(N-1)(a/\varepsilon)}\right)^k \geq (N)_k\left(\frac{\varepsilon/a}{N(a/\varepsilon)}\right)^k = \frac{(N)_k}{N^k}\frac{\varepsilon^{2k}}{a^{2k}}.$$

Similarly,

$$\mathbb{E}[(V_2^{(i)})_k] \leq \frac{(N)_k}{N^k}\frac{a^{2k}}{\varepsilon^{2k}}.$$

We therefore have the bounds

$$\frac{(N)_k}{N^k}\frac{\varepsilon^{2k}}{a^{2k}} \leq \mathbb{E}[(\nu_t^{(i)})_k \mid \mathcal{H}_t] \leq \frac{(N)_k}{N^k}\frac{a^{2k}}{\varepsilon^{2k}}.$$

for each k. Consequently,

$$\frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] \geq \frac{\varepsilon^4}{Na^4} \tag{5.4}$$ eq:mn_cN_LB

and

$$\frac{1}{(N)_3}\sum_{i=1}^{N}\mathbb{E}[(\nu_t^{(i)})_3 \mid \mathcal{H}_t] \leq \frac{a^6}{N^2\varepsilon^6}. \tag{5.5}$$ eq:mn_cN3_U

The definition of $\mathcal{H}_t$ is such that, for any suitable function f, by the tower property and conditional independence we have

$$\mathbb{E}_t[f(\nu_t^{(1:N)})] = \mathbb{E}_t\left[\mathbb{E}[f(\nu_t^{(1:N)}) \mid \mathcal{H}_t, \mathcal{F}_{t-1}]\right] = \mathbb{E}_t\left[\mathbb{E}[f(\nu_t^{(1:N)}) \mid \mathcal{H}_t]\right]. \tag{5.6}$$ eq:condexp_

Applying this identity to (5.4) and (5.5) we find

$$\frac{\frac{1}{(N)_3}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_3]}{\frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}_t[(\nu_t^{(i)})_2]} \leq \frac{a^6/(N^2\varepsilon^6)}{\varepsilon^4/(Na^4)} = \frac{a^{10}}{N\varepsilon^{10}} =: b_N \underset{N\to\infty}{\longrightarrow} 0.$$

Thus (3.11) 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 5.2. Applying Theorem 4.1 then yields the result. ■

⟨thm:mn_nontriviality⟩ **Lemma 5.2.** *Consider an SMC algorithm using multinomial resampling, satisfying (A1) and* (5.1)*. 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 (5.4). If $c_N(t) > 0$ then $c_N(t) \geq 2/(N)_2 > 2/N^2$. Hence, in general $\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{F}_{t-1}] \geq \mathbb{E}_t[c_N(t)]$ the above explanation could be a bit more verbose/explicit. Applying (5.4) along with (5.6), we have for any finite N

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

By a filtered version of the second Borel–Cantelli lemma (see for example 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 $\mathbb{P}[\tau_N(t) = \infty] = 0$. ∎

5.2 Stratified resampling

⟨thm:stratified⟩ **Corollary 5.3.** *Consider an SMC algorithm using stratified resampling, such that (A1) is satisfied. Assume that there exists a constant $a \in [1, \infty)$ such that for all x, x', t,*

$$\frac{1}{a} \leq g_t(x, x') \leq a. \tag{5.7}$$

eq:gq_bounds_s

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 among a total of N particles. Then, for any fixed n, the time-scaled genealogy $(G_{\tau_N(t)}^{(n,N)})_{t \geq 0}$ converges weakly to Kingman's n-coalescent as $N \to \infty$.

Stratified resampling is, by design, much more restrictive than multinomial resampling. Once the weights are known there is little freedom in the offspring counts, so it is not surprising that control over the weights such as (5.7) provides is sufficient without any additional control over the transition densities q_t. This is in contrast to multinomial resampling (Corollary 5.1), where g_t and q_t are more or less on an equal footing, and we require both to be bounded.

It is not immediately clear that the finite time scale condition $\mathbb{P}[\tau_N(t) = \infty] = 0$ holds under conditions (5.7), so it is included in the statement of the corollary. Proposition 5.6 presents some sufficient conditions for the finite time scale, but these are by no means necessary.

By the way, does the lack of conditions of q_t here imply that we do not even need the transition kernels to admit densities?

Proof. Define the σ-algebras $\mathcal{H}_t$ as in (5.2). With stratified resampling, conditional on the weights each offspring count almost surely takes one of four values: $\nu_t^{(i)} \in \{\lfloor Nw_t^{(i)} \rfloor - 1, \lfloor Nw_t^{(i)} \rfloor, \lfloor Nw_t^{(i)} \rfloor + 1, \lfloor Nw_t^{(i)} \rfloor + 2\}$. Define for each $k \in \mathbb{Z}$

$$p_k^{(i)} := \mathbb{P}\left[\nu_t^{(i)} = \lfloor Nw_t^{(i)} \rfloor + k \,\middle|\, \mathcal{H}_t\right]. \tag{5.8}$$ eq:pk_defn

Then $p_k^{(i)} \equiv 0$ for $k \notin \{-1, 0, 1, 2\}$. Now

$$\begin{aligned}\mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] = p_{-1}^{(i)}(\lfloor Nw_t^{(i)} \rfloor - 1)_2 + p_0^{(i)}(\lfloor Nw_t^{(i)} \rfloor)_2 + p_1^{(i)}(\lfloor Nw_t^{(i)} \rfloor + 1)_2 \\ + p_2^{(i)}(\lfloor Nw_t^{(i)} \rfloor + 2)_2\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[(\nu_t^{(i)})_3 \mid \mathcal{H}_t] &= p_{-1}^{(i)}(\lfloor Nw_t^{(i)} \rfloor - 1)_3 + p_0^{(i)}(\lfloor Nw_t^{(i)} \rfloor)_3 + p_1^{(i)}(\lfloor Nw_t^{(i)} \rfloor + 1)_3 \\ &\qquad + p_2^{(i)}(\lfloor Nw_t^{(i)} \rfloor + 2)_3 \\ &= p_{-1}^{(i)}(\lfloor Nw_t^{(i)} \rfloor - 3)(\lfloor Nw_t^{(i)} \rfloor - 1)_2 + p_0^{(i)}(\lfloor Nw_t^{(i)} \rfloor - 2)(\lfloor Nw_t^{(i)} \rfloor)_2 \\ &\qquad + p_1^{(i)}(\lfloor Nw_t^{(i)} \rfloor - 1)(\lfloor Nw_t^{(i)} \rfloor + 1)_2 + p_2^{(i)}\lfloor Nw_t^{(i)} \rfloor(\lfloor Nw_t^{(i)} \rfloor + 2)_2 \\ &\leq \lfloor Nw_t^{(i)} \rfloor\Big\{p_{-1}^{(i)}(\lfloor Nw_t^{(i)} \rfloor - 1)_2 + p_0^{(i)}(\lfloor Nw_t^{(i)} \rfloor)_2 + p_1^{(i)}(\lfloor Nw_t^{(i)} \rfloor + 1)_2 \\ &\qquad + p_2^{(i)}(\lfloor Nw_t^{(i)} \rfloor + 2)_2\Big\} \\ &= \lfloor Nw_t^{(i)} \rfloor \mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] \\ &\leq a^2 \mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t]. \end{aligned} \tag{5.9}$$ eq:strat_v2

The last line uses the almost sure bound $w_t^{(i)} \leq a^2/N$ which follows from (5.7) along with the form of the weights in Algorithm 1. Note that some terms in the above expressions may be equal to zero when $w_t^{(i)}$ is small enough, but the bound still holds in these cases. Since (5.9) holds for all i, applying the tower rule we have

$$\frac{1}{(N)_3}\sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_3] \leq \frac{a^2}{N-2}\frac{1}{(N)_2}\sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_2],$$

satisfying (3.11) with $b_N := a^2/(N-2) \to 0$. The result follows by applying Theorem 4.1. ■

⟨thm:strat_nontriviality⟩ **Proposition 5.4.** *Consider an SMC algorithm using stratified resampling. Suppose that for each t? — actually, isn't is sufficient that these bounds exist for infinitely many t? there exists a constant $\varepsilon \in (0,1]$ and a probability density h such that*

$$\varepsilon h(x') \leq q_t(x,x') \leq \varepsilon^{-1} h(x')$$

uniformly in x, and that there exist $\zeta > 0$ and $\delta > 0$ such that

$$\mathbb{P}[\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N \mid \mathcal{F}_{t-1}] \geq \zeta \tag{5.10}$$ eq:strat_minmax

for infinitely many t. Then, for all $N > 1$, $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t.

We now assume q_t is bounded above and away from zero, as in (5.1). We saw that such a condition was not necessary for Corollary 5.3, and we do not believe it to be necessary here either; it is merely a convenient way to control the contributions from the transition density. The bounds established in the following proof are rather crude, particularly the terms in ε; it may well be possible to achieve similar bounds under less restrictive conditions.

The second condition (5.10) is required to ensure that, at least infinitely often, the weights are not equal to $(1,\dots,1)/N$, since stratified resampling is degenerate under equal weights, which could cause the time scale to explode. It is hardly conceivable that any real SMC algorithm would fail to satisfy this very mild condition, which effectively ensures that the weights cannot be "too well-behaved".

Proof. As argued in Lemma 5.2, it is sufficient to prove that under the stated conditions

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

Firstly,

$$\begin{aligned}\mathbb{P}[c_N(t) \leq 2/N^2 \mid \mathcal{H}_t] &= \mathbb{P}[c_N(t) = 0 \mid \mathcal{H}_t] = \mathbb{P}[\nu_t^{(i)} = 1 \,\forall i \in \{1,\dots,N\} \mid \mathcal{H}_t] \\ &\leq \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t],\end{aligned} \tag{5.11}$$ eq:cNnonzero

where $i^\star := \operatorname{argmax}_i\{w_t^{(i)}\}$ (but note that the inequality holds when $i^\star$ is taken to be any particular index). Define $p_k^{(i)}$ as in (5.8) and recall that, under stratified resampling, $p_k^{(i)} \equiv 0$ for $k \notin \{-1,0,1,2\}$ and

$$\sum_{k=-1}^{2} p_k^{(i)} = \sum_{k=-1}^{2} \mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right] = 1.$$

Up to a proportionality constant C,

$$p_k^{(i)} = C\,\mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right]$$
$$\times \sum_{\substack{a_{1:N}\in\{1,\dots,N\}^N:\\ |\{j:a_j=i\}|=\lfloor N w_t^{(i)}\rfloor + k}} \mathbb{P}\left[a_t^{(1:N)} = a_{1:N} \,\middle|\, \nu_t^{(i)}, w_t^{(1:N)}\right] \prod_{j=1}^N q_{t-1}(X_t^{(a_j)}, X_{t-1}^{(j)})$$

for each $k \in \{-1,0,1,2\}$. We can bound each probability above and below using the almost sure bounds on q_{t-1} from the statement of the Proposition (once the bounds on q_{t-1} are brought outside, the remaining sum of probabilities is equal to one):

$$p_k^{(i)} \geq C\,\mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right] \varepsilon^N \prod_{j=1}^N h(X_{t-1}^{(j)}),$$
$$p_k^{(i)} \leq C\,\mathbb{P}\left[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \,\middle|\, w_t^{(1:N)}\right] \varepsilon^{-N} \prod_{j=1}^N h(X_{t-1}^{(j)}).$$

We then eliminate the proportionality constant C by normalising, to obtain lower bounds

$$\begin{aligned} p_k^{(i)} &\geq \frac{C\,\mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \mid w_t^{(1:N)}]\varepsilon^N \prod_{j=1}^N h(X_{t-1}^{(j)})}{\sum_{j=-1}^2 C\,\mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + j \mid w_t^{(1:N)}]\varepsilon^{-N} \prod_{j=1}^N h(X_{t-1}^{(j)})} \\ &= \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \mid w_t^{(1:N)}]\varepsilon^{2N} \end{aligned} \tag{5.12}$$

eq:strat_pb

for each k, which also imply

$$1 - p_k^{(i)} \geq \left(1 - \mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + k \mid w_t^{(1:N)}]\right) \varepsilon^{2N}. \tag{5.13}$$

eq:strat_no

Suppose that $\max_i w_t^{(i)} - \min_i w_t^{(i)} \geq 2\delta/N$. Then that at least one of $\{\max_i w_t^{(i)} \geq (1+\delta)/N\}$ and $\{\min_i w_t^{(i)} \leq (1-\delta)/N\}$ occurs. We will now examine each of these possibilities.

We can always write the maximum weight as $w_t^{(i^\star)} = \frac{1+\gamma}{N}$ for some $\gamma \geq 0$. Then, using (5.11),

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 - p_0^{(i^\star)} & \text{if } \gamma \in (0,1) \\ 1 - p_{-1}^{(i^\star)} & \text{if } \gamma \in [1,2) \\ 1 & \text{if } \gamma \geq 2. \end{cases}$$

If $\gamma \in (0,1)$ then the "overhang" in the sense of Figure 2.7 is γ, and

$$1 - p_0^{(i^\star)} \geq \frac{3\gamma}{4}\varepsilon^{2N}$$

using Table 2.2 (upper bound on p_0) and (5.13). Similarly, if $\gamma \in [1, 2)$ then the overhang is $\gamma - 1$ and by Table 2.2 (upper bound on p_{-1}),

$$1 - p_{-1}^{(i^\star)} \geq \left(1 - \frac{1}{4}\right) \varepsilon^{2N} \geq \frac{3}{4}\varepsilon^{2N}.$$

Overall, under the constraint $\max_i w_t^{(i)} \geq (1+\delta)/N$, we have

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma \geq \delta} \left\{ \frac{3\gamma}{4}\varepsilon^{2N}\mathbb{1}_{\{\gamma\in[0,1)\}} + \frac{3}{4}\varepsilon^{2N}\mathbb{1}_{\{\gamma\in[1,2)\}} + \mathbb{1}_{\{\gamma\geq 2\}} \right\} = \frac{3}{4}\delta\varepsilon^{2N}.$$

We now construct a similar argument for the minimum weight. Let $j^\star := \operatorname{argmin}_i\{w_t^{(i)}\}$ and write $w_t^{(j^\star)} = \frac{1-\gamma}{N}$, for some $\gamma \in [0, 1]$. Then by (5.11) we have

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(j^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 1 - p_1^{(j^\star)} & \text{if } \gamma \in (0,1] \\ 0 & \text{if } \gamma = 0. \end{cases}$$

If $\gamma \in (0, 1]$ then the "overhang" in the sense of Figure 2.7 is $1 - \gamma$, and

$$1 - p_1^{(j^\star)} \geq \left(1 - \frac{1 + (1-\gamma)}{2}\right) \varepsilon^{2N} = \frac{\gamma}{2}\varepsilon^{2N},$$

using Table 2.2 (upper bound on p_1). Therefore, under the constraint $\min_i w_t^{(i)} \leq (1 - \delta)/N$, we have

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma\geq\delta} \left\{ \frac{\gamma}{2}\varepsilon^{2N} \right\} = \frac{1}{2}\delta\varepsilon^{2N}.$$

Combining both cases, we find for arbitrary r

$$\mathbb{P}[c_N(r) > 2/N^2 \mid \mathcal{H}_r] \geq \frac{1}{2}\delta\varepsilon^{2N}\mathbb{1}_{\{\max_i w_r^{(i)} - \min_i w_r^{(i)} \geq 2\delta/N\}}$$

so, by the tower rule and conditional independence,

$$\begin{aligned} \mathbb{P}[c_N(r) > 2/N^2 \mid \mathcal{F}_{r-1}] &= \mathbb{E}_r\left[\mathbb{P}[c_N(r) > 2/N^2 \mid \mathcal{H}_r]\right] \\ &\geq \frac{1}{2}\delta\varepsilon^{2N}\mathbb{P}[\max_i w_r^{(i)} - \min_i w_r^{(i)} \geq 2\delta/N \mid \mathcal{F}_{r-1}] \\ &\geq \frac{1}{2}\delta\varepsilon^{2N}\zeta > 0 \end{aligned}$$

for infinitely many r. Hence

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

as required. ■

5.3 Stochastic rounding

?⟨thm:stochrounding⟩? **Corollary 5.5.** *Consider an SMC algorithm using any stochastic rounding as its resampling scheme, such that (A1) is satisfied. Assume that there exists a constant* $a \in [1, \infty)$ *such that for all* x, x', t,

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

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 among a total of* N *particles. Then, for any fixed* n, *the time-scaled genealogy* $(G_{\tau_N(t)}^{(n,N)})_{t \geq 0}$ *converges weakly to Kingman's* n*-coalescent as* $N \to \infty$.

Proof. We can apply exactly the proof of Corollary 5.3, except that stochastic rounding is more restrictive than stratified resampling, so that conditional on $w_t^{(1:N)}$ the only possible offspring counts (almost surely) are $\nu_t^{(i)} \in \{\lfloor N w_t^{(i)} \rfloor, \lfloor N w_t^{(i)} \rfloor + 1\}$. We simply set $p_{-1}^{(i)} = p_2^{(i)} = 0$ in the proof of Corollary 5.3 to see that

$$\frac{1}{(N)_3} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_3] \leq \frac{a^2}{N-2} \frac{1}{(N)_2} \sum_{i=1}^{N} \mathbb{E}_t[(\nu_t^{(i)})_2]$$

as required. The result then follows by applying Theorem 4.1. ■

We can also show, under additional conditions, that the assumption $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t holds.

⟨thm:SR_nontriviality⟩ **Proposition 5.6.** *Consider an SMC algorithm using any stochastic rounding as its resampling scheme. Suppose that there exists a constant* $\varepsilon \in (0, 1]$ *and a probability density* h *such that*

$$\varepsilon h(x') \leq q_t(x, x') \leq \varepsilon^{-1} h(x')$$

uniformly in x, *and that there exist* $\zeta > 0$ *and* $\delta > 0$ *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.

This result was published in Brown et al. (2021, Lemma B.1) with the slightly stronger conditions where the bounds on q_t are also uniform in x'. It has since been noted that the conditions given here are sufficient; the h terms can be cancelled as in (5.12).

Proof. Define $p_k^{(i)}$ for $k \in \mathbb{Z}$ as in (5.8). In the case of stochastic rounding, $p_k^{(i)} \equiv 0$ for all

$k \notin \{0,1\}$, and we also have

$$\mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor \mid w_t^{(1:N)}] = 1 - N w_t^{(i)} + \lfloor N w_t^{(i)} \rfloor,$$
$$\mathbb{P}[\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + 1 \mid w_t^{(1:N)}] = N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor.$$

Combining this with (5.12),

$$\begin{aligned} p_0^{(i)} &\geq (1 - N w_t^{(i)} + \lfloor N w_t^{(i)} \rfloor)\varepsilon^{2N}, \\ p_1^{(i)} &\geq (N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)\varepsilon^{2N}. \end{aligned} \tag{5.14}$$ eq:SR_pk_LB

Define $i^\star := \operatorname{argmax}_i\{w_t^{(i)}\}$ and write $w_t^{(i^\star)} = \frac{1+\gamma}{N}$, for some $\gamma \geq 0$. Then, using (5.11),

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 1 - p_0^{(i^\star)} & \text{if } \gamma \in [0,1) \\ 1 & \text{if } \gamma \geq 1. \end{cases}$$

In the case $\gamma \in [0,1)$ we have $N w_t^{(i^\star)} - \lfloor N w_t^{(i^\star)} \rfloor = \gamma$, so

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - p_0^{(i^\star)} = p_1^{(i^\star)} \geq \gamma\varepsilon^{2N},$$

due to (5.14). Therefore, subject to $\max_i w_t^{(i)} \geq (1+\delta)/N$,

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma \geq \delta} \left\{\gamma\varepsilon^{2N}\right\} = \delta\varepsilon^{2N}.$$

Similarly, write $j^\star := \operatorname{argmin}_i\{w_t^{(i)}\}$ and $w_t^{(j^\star)} = \frac{1-\gamma}{N}$, for some $\gamma \in [0,1]$. Then, again using (5.11),

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(j^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 - p_1^{(j^\star)} & \text{if } \gamma \in (0,1) \\ 1 & \text{if } \gamma = 1. \end{cases}$$

If $\gamma \in (0,1)$ then $N w_t^{(i^\star)} - \lfloor N w_t^{(i^\star)} \rfloor = 1 - \gamma$, so

$$1 - p_1^{(j^\star)} = p_0^{(j^\star)} \geq (1 - (1-\gamma))\varepsilon^{2N} = \gamma\varepsilon^{2N}.$$

Therefore, subject to $\min_i w_t^{(i)} \leq (1-\delta)/N$,

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma \geq \delta} \left\{\gamma\varepsilon^{2N}\right\} = \delta\varepsilon^{2N}.$$

Combining the cases for the maximum and minimum weight we have that

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

and we conclude as in Proposition 5.4. ■

5.4 Residual resampling with stratified residuals

⟨sec:corol_resstrat⟩

⟨thm:residual_stratified⟩ **Corollary 5.7.** *Consider an SMC algorithm using residual resampling with stratified residuals, such that (A1) is satisfied. Assume that there exists a constant* $a \in [1, \infty)$ *such that for all* x, x', t,

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

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 among a total of* N *particles. Then, for any fixed* n, *the time-scaled genealogy* $(G_{\tau_N(t)}^{(n,N)})_{t \geq 0}$ *converges weakly to Kingman's* n*-coalescent as* $N \to \infty$.

Proof. We can apply exactly the proof of Corollary 5.3, except that residual-stratified resampling is more restrictive than stratified resampling, so that conditional on $w_t^{(1:N)}$ the only possible offspring counts (almost surely) are $\nu_t^{(i)} \in \{\lfloor N w_t^{(i)} \rfloor, \lfloor N w_t^{(i)} \rfloor + 1, \lfloor N w_t^{(i)} \rfloor + 2\}$. We simply set $p_{-1}^{(i)} = 0$ in the proof of Corollary 5.3 to see that

$$\frac{1}{(N)_3} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_3] \leq \frac{a^2}{N-2} \frac{1}{(N)_2} \sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_2]$$

as required. The result then follows by applying Theorem 4.1. ■

We can also show, under additional conditions, that the assumption $\mathbb{P}[\tau_N(t) = \infty] = 0$ for all finite t holds.

m:resstrat_nontriviality⟩? **Proposition 5.8.** *Consider an SMC algorithm using residual resampling with stratified residuals. Suppose that there exists a constant* $\varepsilon \in (0, 1]$ *and a probability density* h *such that*

$$\varepsilon h(x') \leq q_t(x, x') \leq \varepsilon^{-1} h(x')$$

uniformly in x, *and that there exist* $\zeta > 0$ *and* $\delta > 0$ *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. Define $p_k^{(i)}$ for $k \in \mathbb{Z}$ as in (5.8). In the case of residual resampling with stratified residuals, $p_k^{(i)} \equiv 0$ for all $k \notin \{0, 1, 2\}$. Define $i^\star := \operatorname{argmax}_i \{w_t^{(i)}\}$ and write $w_t^{(i^\star)} = \frac{1+\gamma}{N}$, for some $\gamma \geq 0$. Then, using (5.11),

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(i^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 - p_0^{(i^\star)} & \text{if } \gamma \in (0,1) \\ 1 & \text{if } \gamma \geq 1. \end{cases}$$

In the case $\gamma \in (0,1)$ we have

$$1 - p_0^{(i^\star)} = p_1^{(i^\star)} + p_2^{(i^\star)} \geq p_1^{(i^\star)} \geq \mathbb{P}[\nu_t^{(i^\star)} = \lfloor N w_t^{(i^\star)} \rfloor + 1 \mid w_t^{(1:N)}]\varepsilon^{2N}$$

by (5.12). Also, the residual weight in this case is $r_{i^\star} = \gamma/R$, for some $R \in \{1, \dots, N-1\}$ (since $\gamma > 0$, $R \neq 0$). Therefore $\mathbb{P}[\nu_t^{(i^\star)} = \lfloor N w_t^{(i^\star)} \rfloor + 1 \mid w_t^{(1:N)}]$ is the probability that stratified resampling with R individuals assigns exactly 1 offspring to a parent with weight γ/R. According to Table 2.2 (lower bound on p_1), this probability is at least $\gamma/2$. Hence

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \frac{\gamma}{2}\varepsilon^{2N}.$$

This means that, subject to $\max_i w_t^{(i)} \geq (1+\delta)/N$,

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma \geq \delta} \left\{ \frac{\gamma}{2}\varepsilon^{2N} \right\} = \frac{1}{2}\delta\varepsilon^{2N}.$$

Now a similar calculation for the minimum weight: let $j^\star := \operatorname{argmin}_i\{w_t^{(i)}\}$ and write $w_t^{(j^\star)} = \frac{1-\gamma}{N}$, for some $\gamma \in [0,1]$. Using (5.11),

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq 1 - \mathbb{P}[\nu_t^{(j^\star)} = 1 \mid \mathcal{H}_t] = \begin{cases} 0 & \text{if } \gamma = 0 \\ 1 - p_1^{(j^\star)} & \text{if } \gamma \in (0,1) \\ 1 & \text{if } \gamma = 1. \end{cases}$$

If $\gamma \in (0,1)$ then $r_{j^\star} = (1-\gamma)/R$, for some $R \in \{1, \dots, N-1\}$, and

$$1 - p_1^{(j^\star)} = p_0^{(j^\star)} + p_2^{(j^\star)} \geq p_0^{(j^\star)} \geq \mathbb{P}[\nu_t^{(j^\star)} = \lfloor N w_t^{(j^\star)} \rfloor \mid w_t^{(1:N)}]\varepsilon^{2N}$$

by (5.12). Now $\mathbb{P}[\nu_t^{(j^\star)} = \lfloor N w_t^{(j^\star)} \rfloor \mid w_t^{(1:N)}]$ is the probability that stratified resampling with R individuals assigns exactly 0 offspring to a parent with weight $(1-\gamma)/R$. According to Table 2.2 (lower bound on p_0), this probability is at least $\gamma/2$. Hence

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \frac{\gamma}{2}\varepsilon^{2N}.$$

Therefore, using (5.12), we have that subject to $\min_i w_t^{(i)} \leq (1-\delta)/N$

$$\mathbb{P}[c_N(t) > 2/N^2 \mid \mathcal{H}_t] \geq \min_{\gamma \geq \delta} \left\{ \frac{\gamma}{2}\varepsilon^{2N} \right\} = \frac{1}{2}\delta\varepsilon^{2N}.$$

Combining the cases for the maximum and minimum weight we have

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

and we conclude as in Proposition 5.4. ∎

5.5 Residual resampling with multinomial residuals

We believe that an analogous result holds when the resampling scheme used is residual resampling with multinomial residuals. Considering the ordering by variance presented in Proposition 2.3, the residual-multinomial scheme sits between the multinomial scheme and the residual-stratified scheme, both of which admit the desired convergence result (Corollaries 5.1 and 5.7).

However, we have so far been unable to prove a similar corollary for the residual-multinomial scheme. The techniques used for other residual schemes (see Section 5.4) fail here because the number of offspring assigned to each individual is not upper bounded by $\lfloor N w_t^{(i)} \rfloor$ plus a constant; as many as $R = O(N)$ residual offspring may be assigned to a single individual. The technique used for multinomial resampling (Section 5.1) also fails here: although we have a closed-form expression for the joint distribution of parental indices, it is not a straightforward product form because of the additional dependence between offspring counts induced by the deterministic assignments, so it is unclear how to recover the marginal distributions.

If I manage to prove this corollary, it would make this chapter satisfyingly complete :-) Res-star might prove an easier starting point.

5.6 Star resampling

One might ask the question: is it possible to construct an SMC algorithm whose genealogies converge to some non-trivial limit other than the n-coalescent? The answer is yes, as we now illustrate.

Recall that star resampling assigns all of the offspring to a single parent which is sampled from the Categorical distribution parametrised by $w_t^{(1:N)}$. It is easy enough to show that such a resampling scheme does not satisfy (3.11). The vector of offspring counts is at every generation some permutation of $(N, 0, \dots, 0)$, and hence we calculate

$$\frac{1}{(N)_2}\sum_{i=1}^N \mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] = \frac{1}{(N)_2}(N)_2 = 1,$$
$$\frac{1}{(N)_3}\sum_{i=1}^N \mathbb{E}[(\nu_t^{(i)})_3 \mid \mathcal{H}_t] = \frac{1}{(N)_3}(N)_3 = 1,$$

so no suitable sequence b_N can be found. Now we know that Theorem 3.6 does not apply, but this is not enough because condition (3.11) was not proved to be necessary. But in fact we know exactly what the genealogy of n particles from this SMC algorithm looks like (Figure 5.1). Whatever time scale is used, we cannot get away from the fact that this

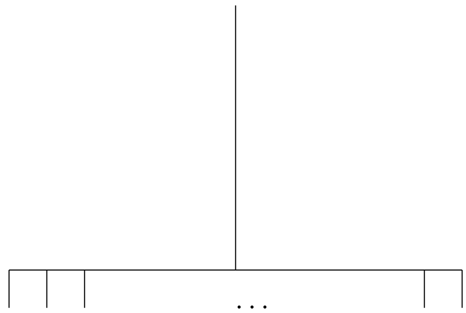


Figure 5.1: Sample genealogy induced by star resampling

⟨fig:star_genealogy⟩

genealogy involves multiple mergers; it cannot converge to the n-coalescent.

The limiting genealogy is more like a *star coalescent* (Pitman 1999; Griffiths and Mano 2016). This is the coalescent process comprising an Exp(1)-distributed event time at which all of the lineages merge into one.

In the case of star resampling we have $c_N(t) \equiv 1$, so the time-scaling function $\tau_N(t)$ defined in (3.2) converges pointwise to the identity function $\tau(t) \equiv t$ as $N \to \infty$, and this does not yield a continuous-time limit. Under any time scale that results in a continuous-time limiting process, the coalescent event time converges to 0, rather than the usual Exp(1)-distributed random variable. The resulting genealogy is a variant star coalescent where the distribution of the event time is a point mass at 0. A fun consequence of this is that this coalescent comes down from infinity, while the star coalescent does not if I decide not to write about cdf∞ in ch2 then remove this last sentence.

5.7 Conditional SMC

In conditional SMC, one "immortal" particle is treated differently to the others when it comes to assigning offspring to parents. The immortal particle is guaranteed at least one offspring, and has on average one more offspring than each of the other parents in each generation. This results in genealogies that are qualitatively different to those of a corresponding standard SMC algorithm. For one thing, the population MRCA is *guaranteed* to be an immortal particle; there is a sense in which the immortal lineage *attracts* coalescence events.

Given this, we should not have been surprised if conditional SMC genealogies converged to a quite different coalescent process, perhaps a *structured coalescent* (Notohara 1990). As it turns out, we still recover Kingman's n-coalescent in the large population limit (Corollary 5.9). A possible explanation for this is that, as $N \to \infty$, the probability of a given sample of size n interacting with the immortal lineage (before its within-sample MRCA) vanishes, leaving a process that looks very much like the one induced by the corresponding standard SMC algorithm.

⟨thm:CSMC⟩ **Corollary 5.9.** *Consider a conditional SMC algorithm using multinomial resampling, such that (A1) is satisfied. Assume there exist constants $\varepsilon \in (0,1]$ and $a \in [1,\infty)$ and probability density h such that for all x, x', t,*

$$\frac{1}{a} \leq g_t(x,x') \leq a, \quad \varepsilon h(x') \leq q_t(x,x') \leq \frac{1}{\varepsilon} h(x'). \tag{5.15}$$

eq:gq_bounds_

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 among a total of N particles. Then, for any fixed n, the time-scaled genealogy $(G_{\tau_N(t)}^{(n,N)})_{t\geq 0}$ converges weakly to Kingman's n-coalescent as $N \to \infty$.

We restrict here to the case of multinomial resampling, which seems to be the most commonly-used resampling scheme within conditional SMC. Implementing other resampling schemes while maintaining the immortal lineage is more involved, though by no means impossible (Lee, Murray, and Johansen 2019). We conjecture that similar results hold for conditional SMC with other resampling schemes, as in the preceding corollaries.

The conditions (5.15) are, as one might expect, identical to those assumed in the case of standard SMC with multinomial resampling (Corollary 5.1).

Proof. Assume, without loss of generality, that the immortal particle takes index 1 in each generation. This assumption is valid due to (A1), and significantly lightens the notation, but the same argument holds if the immortal indices are taken to be $a^\star_{(0:T)}$ rather than $(1,\dots,1)$.

Define $\mathcal{H}_t$ as in (5.2). The parental indices are conditionally independent given $\mathcal{H}_t$, as in standard SMC with multinomial resampling, but we have to treat $i=1$ as a special case. The conditional law on the i^{th} parental index is

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

resulting in the joint law

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

As in Corollary 5.1, under (5.15) we have bounds

$$\mathbb{E}[(V_1^{(i)})_k] \leq \mathbb{E}[(\nu_t^{(i)})_k \mid \mathcal{H}_t] \leq \mathbb{E}[(V_2^{(i)})_k],$$

where now

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

independently for each i and independently of $\mathcal{F}_\infty$. Furthermore, using the Binomial moments and the identity $(X+1)_2 \equiv 2(X)_1 + (X)_2$, one can show that

$$\mathbb{E}[(V_1^{(i)})_2] \geq \begin{cases} \frac{(N-1)_2}{N^2}\frac{\varepsilon^4}{a^4} + \frac{2(N-1)}{N}\frac{\varepsilon^2}{a^2} & \text{if } i=1 \\ \frac{(N-1)_2}{N^2}\frac{\varepsilon^4}{a^4} & \text{if } i \neq 1. \end{cases}$$

Using the identity $(X+1)_3 \equiv 3(X)_2 + (X)_3$, we also have

$$\mathbb{E}[(V_2^{(i)})_3] \leq \begin{cases} \frac{(N-1)_3}{N^3}\frac{a^6}{\varepsilon^6} + \frac{3(N-1)_2}{N^2}\frac{a^4}{\varepsilon^4} & \text{if } i=1 \\ \frac{(N-1)_3}{N^3}\frac{a^6}{\varepsilon^6} & \text{if } i \neq 1. \end{cases}$$

We therefore have

$$\begin{aligned} \frac{1}{(N)_2}\sum_{i=1}^N \mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] \geq \frac{1}{(N)_2}\sum_{i=1}^N \mathbb{E}[(V_1^{(i)})_2] &\geq \frac{1}{(N)_2}\left[\frac{2(N-1)}{N}\frac{\varepsilon^2}{a^2} + \sum_{i=1}^N \frac{(N-1)_2}{N^2}\frac{\varepsilon^4}{a^4}\right] \\ &= \frac{1}{N^2}\left[2\frac{\varepsilon^2}{a^2} + (N-2)\frac{\varepsilon^4}{a^4}\right] \geq \frac{\varepsilon^4}{Na^4} \end{aligned} \tag{5.16}$$

eq:csmc_cN_

and

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

Hence, applying (5.6), we can upper bound the ratio

$$\frac{\frac{1}{(N)_3}\sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_3]}{\frac{1}{(N)_2}\sum_{i=1}^N \mathbb{E}_t[(\nu_t^{(i)})_2]} \leq \frac{a^{10}}{N\varepsilon^{10}} =: b_N \xrightarrow[N\to\infty]{} 0$$

so (3.11) is satisfied. Proof that the time scale is finite is relegated to Lemma 5.10, whence we conclude by applying Theorem 4.1.

■

⟨thm:CSMC_nontriviality⟩ **Lemma 5.10.** *Consider a conditional SMC algorithm using multinomial resampling, satisfying (A1) and* (5.15)*. Then, for all* $N>2$, $\mathbb{P}[\tau_N(t)=\infty]=0$ *for all finite* t*.*

Proof. The proof is identical to that of Lemma 5.2, since (5.16) gives us exactly the same lower bound on $\mathbb{E}_t[c_N(t)]$ that we had in standard SMC with multinomial resampling. ■

5.7.1 Effect of ancestor sampling

Ancestor sampling breaks up the immortal lineage into sections, so it is not really a lineage anymore, and we do not really have a pure coalescent process backwards in time. Regardless, we shall throw caution to the wind and examine the resulting "genealogies".

Using the parent sampling probabilities specified in (2.10), now with time reversed and the notation made to fit preferably the presentation in Chapter 2 should use the notation we want to use here with this study of genealogies, we obtain add one more step of working below to make it less "magic"?

$$\mathbb{P}[a_t^{(i)} = a_i \mid \mathcal{H}_t] \propto \begin{cases} w_t^{(a_i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}) & i \in \text{non-immortal particles} \\ w_t^{(a_i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{*}) & i = \text{immortal particle.} \end{cases}$$

But when i is the index of the immortal particle, $X_{t-1}^{(i)} = X_{t-1}^{*}$, so the above simplifies to

$$\mathbb{P}[a_t^{(i)} = a_i \mid \mathcal{H}_t] \propto w_t^{(a_i)} q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)})$$

for each i, which is exactly (5.3), the law on parental indices under standard SMC with multinomial resampling. In other words, when parental indices are chosen, the immortal particle is treated exactly like all of the other particles; it has completely lost its "reproductive advantage". This means it is no more likely for lineages to coalesce onto the "immortal" lineage than onto any other lineage, so we do not see the behaviour of Figure 2.8 which caused the particle Gibbs chain to mix slowly over the sequential component. This supports the claim of Section 2.5.3: particle Gibbs with ancestor sampling still experiences ancestral degeneracy, but this no longer causes the sequential component to get stuck.

6 Discussion

Bibliography

[1] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. "Particle Markov Chain Monte Carlo Methods". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3 (2010), pp. 269–342.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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.

[7] James Carpenter, Peter Clifford, and Paul Fearnhead. "Improved Particle Filter for Nonlinear Problems". In: *IEE Proceedings — Radar, Sonar and Navigation* 146.1 (1999), pp. 2–7.

[8] Nicolas Chopin and Omiros Papaspiliopoulos. *An Introduction to Sequential Monte Carlo*. Springer, 2020.

[9] Adrien Corenflos et al. *Differentiable Particle Filtering via Entropy-Regularized Optimal Transport*. Tech. rep. 2102.07850. ArXiv, 2021.

[10] Dan Crisan and Terry Lyons. "Nonlinear Filtering and Measure-Valued Processes". In: *Probability Theory and Related Fields* 109.2 (1997), pp. 217–244.

[11] Dan Crisan and Terry Lyons. "A Particle Approximation of the Solution of the Kushner–Stratonovitch Equation". In: *Probability Theory and Related Fields* 115.4 (1999), pp. 549–578.

[12] Pierre Del Moral. *Feynman–Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, 2004.

[13] Randal Douc, Olivier Cappé, and Eric Moulines. "Comparison of Resampling Schemes for Particle Filtering". In: *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*. IEEE. 2005, pp. 64–69.

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

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

[16] Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. CRC press, 1994.

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

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

[19] Geir Evensen. "Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics". In: *Journal of Geophysical Research: Oceans* 99.C5 (1994), pp. 10143–10162.

[20] Paul Fearnhead and Peter Clifford. "On-line Inference for Hidden Markov Models via Particle Filters". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65.4 (2003), pp. 887–899.

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

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

[23] Catherine Forbes et al. *Statistical Distributions*. John Wiley & Sons, 2011.

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

[25] Mathieu Gerber, Nicolas Chopin, and Nick Whiteley. "Negative Association, Ordering and Convergence of Resampling Methods". In: *The Annals of Statistics* 47.4 (2019), pp. 2236–2260.

[26] Neil J. Gordon, David J. Salmond, and Adrian F. M. Smith. "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation". In: *IEE Proceedings F (Radar and Signal Processing)*. Vol. 140. 2. IET. 1993, pp. 107–113.

[27] Robert Griffiths and Shuhei Mano. "The Star-shaped Λ-coalescent and Fleming–Viot Process". In: *Stochastic Models* 32.4 (2016), pp. 606–631.

[28] 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.

[29] Jeroen D. Hol. “Resampling in Particle Filters”. LiTH-ISY-EX-ET-0283-2004. Internship report. Linköping University, 2004.

[30] Jeroen D. Hol, Thomas B. Schön, and Fredrik Gustafsson. “On Resampling Algorithms for Particle Filters”. In: *Nonlinear Statistical Signal Processing Workshop*. IEEE. 2006, pp. 79–82.

[31] Chaofan Huang, V. Roshan Joseph, and Simon Mak. *Population Quasi-Monte Carlo*. Tech. rep. 2012.13769. ArXiv, 2020.

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

[33] Kumar Joag-Dev and Frank Proschan. “Negative Association of Random Variables with Applications”. In: *The Annals of Statistics* (1983), pp. 286–295.

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

[35] 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.

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

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

[38] Genshiro Kitagawa. “Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models”. In: *Journal of Computational and Graphical Statistics* 5.1 (1996), pp. 1–25.

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

[40] Anthony Lee, Lawrence Murray, and Adam M. Johansen. “Resampling in Conditional SMC Algorithms”. Unpublished. 2019.

[41] Fredrik Lindsten and Thomas B Schön. “Backward Simulation Methods for Monte Carlo Statistical Inference”. In: *Foundations and Trends in Machine Learning* 6.1 (2013), pp. 1–143.

[42] Jun S Liu and Rong Chen. “Sequential Monte Carlo Methods for Dynamic Systems”. In: *Journal of the American Statistical Association* 93.443 (1998), pp. 1032–1044.

[43] Martin Möhle. “Robustness Results for the Coalescent”. In: *Journal of Applied Probability* 35.2 (1998), pp. 438–447.

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

[45] Martin Möhle and Serik Sagitov. “A Classification of Coalescent Processes for Haploid Exchangeable Population Models”. In: *The Annals of Probability* 29.4 (2001), pp. 1547–1562.

[46] Martin Möhle and Serik Sagitov. “Coalescent Patterns in Exchangeable Diploid Population Models”. In: *Journal of Mathematical Biology* 47 (2003), pp. 337–352.

[47] 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.

[48] 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.

[49] Aaron Myers et al. “Sequential Ensemble Transform for Bayesian Inverse Problems”. In: *Journal of Computational Physics* 427 (2021), p. 110055.

[50] M. Notohara. “The Coalescent and the Genealogical Process in Geographically Structured Population”. In: *Journal of Mathematical Biology* 29.1 (1990), pp. 59–75.

[51] Jim Pitman. “Coalescents with Multiple Collisions”. In: *Annals of Probability* (1999), pp. 1870–1902.

[52] 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.

[53] Sebastian Reich. “A Nonparametric Ensemble Transform Method for Bayesian Inference”. In: *SIAM Journal on Scientific Computing* 35.4 (2013), A2013–A2024.

[54] Ian W. Saunders, Simon Tavaré, and G. A. Watterson. “On the Genealogy of Nested Subsamples from a Haploid Population”. In: *Advances in Applied Probability* (1984), pp. 471–491.

[55] John L. Spouge. “Within a Sample from a Population, the Distribution of the Number of Descendants of a Subsample’s Most Recent Common Ancestor”. In: *Theoretical Population Biology* 92 (2014), pp. 51–54.

[56] 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.

[57] 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.

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

[59] 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.

[60] Darrell Whitley. "A Genetic Algorithm Tutorial". In: *Statistics and Computing* 4.2 (1994), pp. 65–85.

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