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

July 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 2.1.1) 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 e.g. Chopin and Papaspiliopoulos (2020) and Doucet and Johansen (2011) for more background. Here it will suffice to include a motivation in the context of state space models (Section 2.1.1) followed by the formalism of Feynman-Kac models (Section 2.1.3).

2.1.1 State space models

⟨sec:SSMs⟩ 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

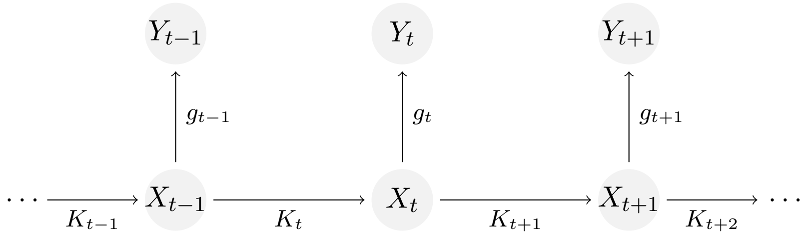


Figure 2.1: Graphical depiction of a general state space model. (X_t) is a Markov process with transition kernels (K_t) representing the underlying state of the system. Y_t is a noisy observation of X_t for each t.

⟨fig:SSM⟩

model,

$$\begin{aligned} X_0 &\sim \mu(\cdot) && \\ X_{t+1} \mid X_t &\sim K_{t+1}(\cdot|X_t) && \text{for } t = 0, 1, \ldots \\ Y_t \mid X_t &\sim g_t(\cdot|X_t) && \text{for } t = 0, 1, \ldots \end{aligned} \tag{2.1}$$

eq:SSM_spec

The index t will frequently be referred to as "time", since in most applications the sequence is indeed a time series, but it need not be.

Here X and/or Y may be multivariate and observation times need not be equally spaced. 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 $p(x_t \mid y_{0:t})$: inferring the current state x_t from the observations up to now $y_{0:t}$

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

(complete) smoothing $p(x_{0:t} \mid y_{0:t})$: inferring the sequence of states up to now $x_{0:t}$ from the observations up to now $y_{0:t}$

fixed-lag smoothing $p(x_{t-h:t} \mid y_{0:t})$: 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

$p(y_{0:t})$ of the observed data given a certain value of θ.

In certain cases, these inference problems may be solved analytically (Section 2.1.2), 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 2.1.4, 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

sec:SSM_exact_inference⟩ 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

⟨sec:FKmodels⟩ Example of non-SSM that is FK?

State space models are very natural and intuitive applications, but they do not do justice to the scope of SMC algorithms, which is much wider. On the other hand, the *Feynman-Kac* formalism captures the full scope, since every SMC algorithm is a Monte Carlo approximation of some Feynman-Kac model. Before formally introducing SMC let us therefore define a generic Feynman-Kac model. For a more in-depth study, the reader is directed to the exhaustive books by Del Moral (2004, 2013) or the more accessible Chopin and Papaspiliopoulos (2020, Chapter 5).

Define a state space $\mathcal{X}$ and a time horizon T. The basic components of the Feynman-Kac model are a Markov law, defined by an initial distribution $\mathbb{M}_0$ and transition kernels M_t for $t = 1, \ldots, T$; and a sequence of "potential" functions $G_0 : \mathcal{X} \mapsto [0, \infty)$ and $G_t : \mathcal{X}^2 \mapsto [0, \infty)$ for $t = 1, \ldots, T$. From these we construct a sequence of Feynman-Kac measures $(\mathbb{Q}_t)_{t=0:T}$ defined by the changes of measure

$$\mathbb{Q}_t(dx_{0:T}) = \frac{1}{L_t} G_0(x_0)\mathbb{M}_0(dx_0) \left\{ \prod_{s=1}^{t} G_s(x_{s-1}, x_s) \right\} \left\{ \prod_{s=1}^{T} M_s(x_{s-1}, dx_s) \right\}, \tag{2.2}$$ eq:FKmeasure

where L_t is the normalising constant required to make $\mathbb{Q}_t$ a probability measure. Other quantities such as $\mathbb{Q}_t(dx_{0:t})$ can be obtained as marginals of (2.2), allowing us to treat all of the inference problems described in Section 2.1.1 by approximating $\mathbb{Q}_t$ and then possibly marginalising.

The generic state space model introduced in (2.1) may be described by a Feynman-Kac model where:

$$\begin{aligned} \mathbb{M}_0 &:= \mu \\ M_t(x_{t-1}, dx_t) &:= K_t(dx_t \mid x_{t-1}) && \text{for } t = 1, \ldots, T \\ G_0(x_0) &:= g_0(y_0 \mid x_0) \\ G_t(x_{t-1}, x_t) &:= g_t(y_t \mid x_t) && \text{for } t = 1, \ldots, T. \end{aligned} \tag{2.3}$$ eq:bootstrap

This is not the only Feynman-Kac model for (2.1); this corresponds to the "bootstrap"

SMC algorithm, which is the simplest implementation but may be significantly outperformed in practice by more involved algorithms such as the "guided" [citation] and "auxiliary" (Carpenter, Clifford, and Fearnhead 1999; Pitt and Shephard 1999) variants. Feynman-Kac formalisms for these variants are presented for example in Chopin and Papaspiliopoulos (2020, Section 5.1.2). Say something about the time horizon, which we have now fixed, but was infinite in SSMs section.

It remains to demonstrate that the measures $\mathbb{Q}_t$ arising from (2.3) are sufficient for all the usual inference problems in the corresponding state space model (2.1). By construction, the complete smoothing distribution is precisely

$$\begin{aligned}\mathbb{Q}_t(dx_{0:t}) &= \frac{1}{L_t} G_0(x_0)\mathbb{M}_0(dx_0) \prod_{s=1}^{t} G_s(x_{s-1}, x_s) M_s(x_{s-1}, dx_s) \\ &= g_0(y_0 \mid x_0)\mu(dx_0) \prod_{s=1}^{t} g_s(y_s \mid x_s) K_s(dx_s \mid x_{s-1}) \\ &= p(dx_{0:t} \mid y_{0:t}).\end{aligned}$$

The filtering, prediction and fixed-lag smoothing distributions are all also marginals of $\mathbb{Q}_t(dx_{0:T})$:

$$\begin{aligned}p(x_t \mid y_{0:t}) &= \mathbb{Q}_t(dx_t) \\ p(x_{t+h} \mid y_{0:t}) &= \mathbb{Q}_t(dx_{t+h}) \\ p(x_{t-h:t} \mid y_{0:t}) &= \mathbb{Q}_t(dx_{t-h:t}),\end{aligned}$$

while the likelihood $p(y_{0:t}) = L_t$. The upshot of this is that we have access to a Monte Carlo approximation of $\mathbb{Q}_t(dx_{0:T})$ then we can conduct inference on all of these distributions, since marginalisation of Monte Carlo samples is trivial. The likelihood, on the other hand, is not obtained by marginalisation; nevertheless, approximations of the likelihood can also be obtained "for free". The next section describes how we may obtain samples from $\mathbb{Q}_t(dx_{0:T})$.

2.1.4 Sequential Monte Carlo for Feynman-Kac models

⟨sec:SMC_FK⟩ In order to implement the SMC algorithm corresponding to a given Feynman-Kac model, we need to be able to sample from $\mathbb{M}_0$ and from $M_t(x, \cdot)$ for all x, t; and evaluate $G_t(x, y)$ pointwise for each x, y, t. Under these conditions we may implement Algorithm 1, which describes a generic SMC algorithm. The only free choices are the parameter N which dictates the number of "particles" used, and the RESAMPLE procedure. However, remember that given a particular state space model there is a choice of possible Feynman-Kac descriptions, and this choice can strongly affect performance.

The choice of RESAMPLE procedure can also have a profound effect on performance and is discussed in detail in Section 2.4. To ensure Algorithm 1 is valid, we will assume that

RESAMPLE satisfies the following properties:

- the population size N is conserved;
- given $w_{t-1}^{(1:N)}$, the expected cardinality of $\{j : a_{t-1}^{(j)} = i\}$ is proportional to $w_{t-1}^{(i)}$.

The latter property is known as *unbiasedness*. These requirements are made rigorous in Definition 2.2.

Input: $T, N, \mathbb{M}_0, (M_t)_{t=1}^T, (G_t)_{t=0}^T$
for $i \in \{1, \ldots, N\}$ **do** Sample $X_0^{(i)} \sim \mathbb{M}_0(\cdot)$
for $i \in \{1, \ldots, N\}$ **do** $w_0^{(i)} \leftarrow \left\{\sum_{j=1}^N G_0(X_0^{(j)})\right\}^{-1} G_0(X_0^{(i)})$
for $t \in \{1, \ldots, T\}$ **do**
  Sample $a_{t-1}^{(1:N)} \sim \text{RESAMPLE}(\{1, \ldots, N\}, w_{t-1}^{(1:N)})$
  for $i \in \{1, \ldots, N\}$ **do** Sample $X_t^{(i)} \sim M_t(X_{t-1}^{(a_{t-1}^{(i)})}, \cdot)$
  for $i \in \{1, \ldots, N\}$ **do** $w_t^{(i)} \leftarrow \left\{\sum_{j=1}^N G_t(X_{t-1}^{(a_{t-1}^{(j)})}, X_t^{(j)})\right\}^{-1} G_t(X_{t-1}^{(a_{t-1}^{(i)})}, X_t^{(i)})$
end

⟨alg:SMC⟩ **Algorithm 1:** Sequential Monte Carlo for a generic Feynman-Kac model

Because Algorithm 1 proceeds sequentially, its computational cost is linear in the time horizon T. Furthermore, the bootstrap algorithm, where the Feynman-Kac model is (2.3), processes the data $y_{0:T}$ one observation at a time via $G_t(x_t) = g_t(y_t \mid x_t)$. This means that the bootstrap SMC algorithm can also be run on-line, incorporating each observation as it becomes available. This is in stark contrast to a standard MCMC approach, for example, which would have to process all of the data at once up to a fixed time horizon. Adding one more observation would require running the MCMC algorithm from scratch on the extended target, making the computational cost (of generating samples from the whole sequence of target distributions, which is what SMC achieves) at least quadratic in time and rendering on-line inference infeasible.

The output of Algorithm 1 is, for $i = 1, \ldots, N$ and $t = 0, \ldots, T$, the *states* $X_t^{(i)} \in \mathcal{X}$ in general the state space can vary over time, the *weights* $w_t^{(i)} \in [0, 1]$ and, for $i = 1, \ldots, N$ and $t = 0, \ldots, T-1$, the *parental indices* $a_t^{(i)}$. Depending on the application, one may want to retain only a subset of this output.

This output can be used to construct discrete approximations of the various distributions of interest, which can be used to estimate integrals against test functions in the usual way. The filtering distribution $p(x_t \mid y_{0:t})$ is approximated by

$$\sum_{i=1}^N w_t^{(i)} \delta_{X_t^{(i)}},$$

where δ_x denotes a unit mass at x. That is, expectations of appropriate test functions

$\varphi : \mathcal{X} \mapsto \mathbb{R}$ are approximated by

$$\mathbb{E}[\varphi(x_t) \mid y_{0:t}] \simeq \sum_{i=1}^{N} w_t^{(i)} \varphi(X_t^{(i)}).$$

The precise meaning of approximation (or $\simeq$) will be clarified in Section 2.1.5. To approximate the smoothing distribution, we first define the *trajectories* of states $X_{t,0:t}^{(i)}$ (for each $i \in \{1, \ldots, N\}$) by setting $X_{t,t}^{(i)} := X_t^{(i)}$ and tracing back through the ancestors via the recursion $X_{t,s}(i) = X_{t,s+1}^{(a_t^{(i)})}$ for each $s \in \{0, \ldots, t\}$. We then construct the approximation

$$\sum_{i=1}^{N} w_t^{(i)} \delta_{X_{t,0:t}^{(i)}}$$

of the smoothing distribution $p(x_{0:t} \mid y_{0:t})$. Similar approximations can be constructed for prediction and fixed-lag smoothing ...be explicit or no?. We can also approximate the marginal likelihood using the *unnormalised* weights should I introduce notation e.g. $\tilde{w}$ for the unnormalised weights, and give them a separate line in Algorithm 1, and include them in the possible outputs of the algorithm?:

$$L_t \simeq \frac{1}{N} \sum_{i=1}^{N} G_0(X_0^{(i)}) \prod_{t=1}^{T} \frac{1}{N} \sum_{i=1}^{N} G_t(X_{t-1}^{a_{t-1}^{(i)}}, X_t^{(i)}). \tag{2.4}$$ eq:likeliho

Define the *offspring counts* for each i, t

$$\nu_t^{(i)} := |\{j : a_t^{(j)} = i\}|,$$

the number of copies of particle i of generation t appearing in generation $t+1$ by resampling. Notice that $\nu_t^{(1:N)}$ is expressed as a non-injective function of $a_t^{(1:N)}$, and as such carries less information.

Figure 2.2 shows a section of the conditional dependence graph implied by Algorithm 1.

2.1.5 Theoretical justification

⟨sec:SMC_theory⟩ It can be shown that SMC approximations of expectations of test functions satisfy various desirable properties. For instance, it is quite easy to show that the filtering approximations converge:

$$\sum_{i=1}^{N} w_t^{(i)} \varphi(X_t^{(i)}) \longrightarrow \mathbb{Q}_t(\varphi),$$

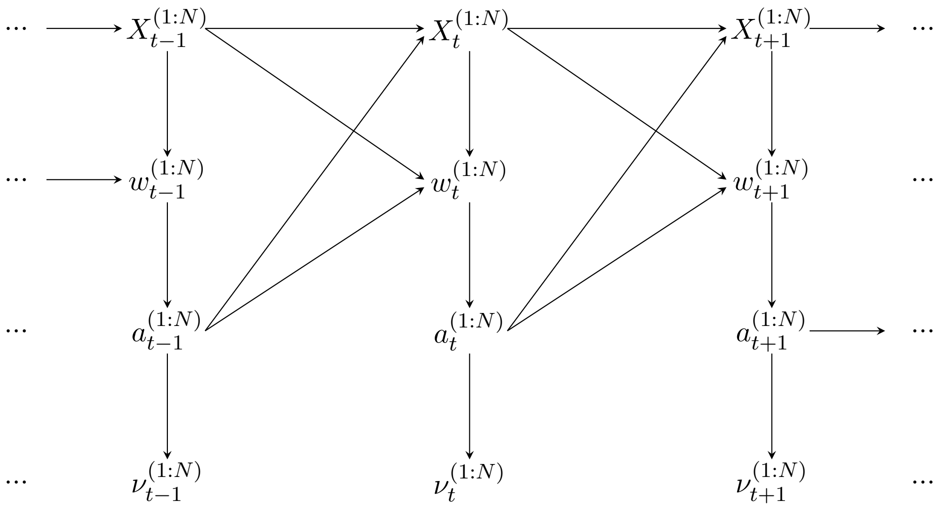


Figure 2.2: Part of the conditional dependence graph implied by Algorithm 1. The direction of time is from left to right.

⟨fig:cond_indep_graph⟩

almost surely and in the L_2 sense, as $N \to \infty$, under some conditions. Moreover, they satisfy a central limit theorem:

$$\sqrt{N}\left(\sum_{i=1}^{N} w_t^{(i)}\varphi(X_t^{(i)}) - \mathbb{Q}_t(\varphi)\right) \longrightarrow \text{Normal}(0, \sigma_t(\varphi))$$

in distribution, as $N \to \infty$. Under additional conditions, the asymptotic variances $\sigma_t(\varphi)$ are stable over t, justifying the use of SMC filtering on-line. It can also easily be shown that the likelihood estimates (2.4) are unbiased (see for example Chopin and Papaspiliopoulos 2020, Proposition 16.3).

There are many other results concerning convergence, stability and error bounds for SMC algorithms. A full exposition of these results and their conditions is beyond the scope of this work, but the book by Del Moral (2013) provides an exhaustive treatment, and some of the key ideas and results are developed in the more accessible book by Chopin and Papaspiliopoulos (2020, Chapter 11). Suffice it to say that SMC algorithms enjoy enough theoretical properties to be useful in practice.

2.2 Coalescent theory ✓

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

2.2.1 Kingman's coalescent

The Kingman coalescent (Kingman 1982a,b,c) 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 King-

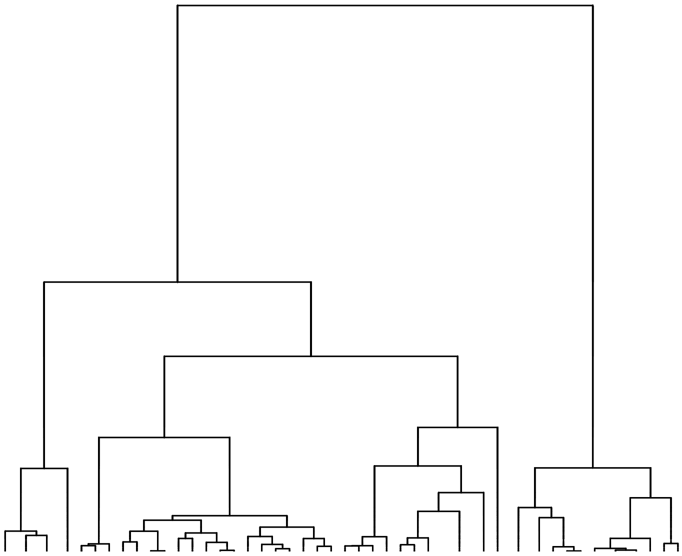


Figure 2.3: A realisation of the n-coalescent with $n = 50$.

man 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) ?

⟨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

$$q_{\xi,\eta} = \begin{cases} 1 & \xi \prec \eta \\ -|\xi|(|\xi|-1)/2 & \xi = \eta \\ 0 & \text{otherwise} \end{cases} \tag{2.5}$$?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).

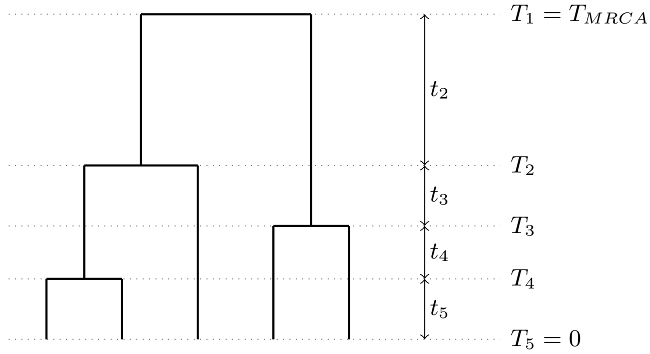


Figure 2.4: Definitions of t_i, T_i in the n-coalescent.

⟨fig:KC_timedefns⟩

Denote by $t_2, t_3 \dots, t_n$ the waiting times between coalescent events, where t_i is the amount of time for which the coalescent has exactly i distinct lineages (see Figure 2.4). A consequence of Definition 2.1 is that these waiting times are independent and have distributions

$$t_i \sim \mathrm{Exp}\left(\binom{i}{2}\right). \qquad (2.6)\ \{?\}$$

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,

$$\mathbb{E}[T_{MRCA}] = \sum_{i=2}^{n} \mathbb{E}[t_i] = \sum_{i=2}^{n} \frac{2}{i(i-1)} = 2\sum_{i=2}^{n} \left\{\frac{1}{i-1} - \frac{1}{i}\right\} = 2\left(1 - \frac{1}{n}\right) \qquad (2.7)\ \{?\}$$

and

$$\mathrm{Var}[T_{MRCA}] = \sum_{i=2}^{n} \mathrm{Var}[t_i] = \sum_{i=2}^{n} \left(\frac{2}{i(i-1)}\right)^2. \qquad (2.8)\ \{?\}$$

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

$$\mathbb{E}[L] = \sum_{i=2}^{n} i\mathbb{E}[t_i] = \sum_{i=2}^{n} \frac{2}{i-1} = \sum_{i=1}^{n-1} \frac{2}{i} \simeq 2\ln(n-1) \qquad (2.9)\ \{?\}$$

and

$$\mathrm{Var}[L] = \sum_{i=2}^{n} i^2\,\mathrm{Var}[t_i] = \sum_{i=2}^{n} \frac{4}{(i-1)^2} = \sum_{i=1}^{n-1} \frac{4}{i^2}. \qquad (2.10)\ \{?\}$$

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, 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

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

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, 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).

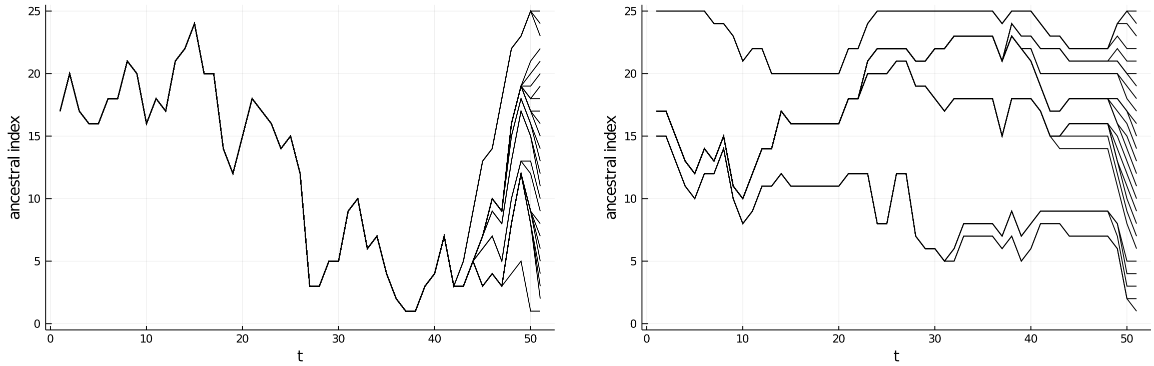
2.3 Sequential Monte Carlo genealogies ✓

⟨sec:SMC_genealogies⟩ We have seen that genetic terminology applies quite naturally to SMC. The resampling step induces parent-offspring relationships, each duplicate of particle i after resampling being considered one of its offspring. Then follows the notion of offspring counts (also known as family sizes), that is, the number of offspring assigned to each parent. Viewed backwards in time, the parent-offspring relationships also imply a genealogy, obtained by tracing the lineages from each terminal particle through its ancestor in each generation. We will see in this section that these genealogies, induced by resampling, are not a mere curiosity but in fact have important implications for the performance of SMC algorithms.

2.3.1 Ancestral degeneracy

Suppose we were using SMC to sample from the smoothing distribution of some state space model. As described in Section 2.1.4, we run our chosen SMC algorithm forwards, then output the N sampled trajectories $X_{t,0:t}^{(i)}$ (for each $i \in \{1, \dots, N\}$). Each trajectory was obtained by tracing back through the parent at each generation, starting from one of the terminal particles. This means that if two terminal particles i and j share a common ancestor at some generation s, then $X_{t,0:s}^{(i)}$ will be exactly equal to $X_{t,0:s}^{(j)}$, because their ancestries coincide from time s to 0.

At every resampling step, some parents may be assigned more than one offspring each, so the further back in time you look, the more of the ancestries of the terminal particles will have coalesced (see Figure 2.5a). The effect of this is that, instead of obtaining N separate sampled trajectories, we actually obtain N sampled trajectories that coalesce backwards in time, which means that the further back in time we look, the fewer distinct samples we have from the corresponding component of the target distribution. Particularly if we are interested in smoothing over a long time horizon, the variance of the SMC estimator is going to blow up.



(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. Each line is the lineage of one of the terminal particles, indicating the index of its ancestor in each generation: (a) with multinomial resampling; (b) the same system with adaptive systematic resampling.

ig:ancestral_degeneracy⟩

On the other hand, ancestral degeneracy actually improves the memory efficiency of SMC. We do not need to store all of the particles generated at each time (at memory cost $O(NT)$), only those that are included in the resulting genealogy. Jacob, Murray, and Rubenthaler (2015) provide an algorithm for efficient storage of the genealogy, reducing the asymptotic memory cost to $O(N \log N + T)$. However, it is certainly still worth trying to reduce ancestral degeneracy because to achieve a given level of error with a highly degenerate system will require such a large N that the memory gains are cancelled.

Mitigating ancestral degeneracy

There are a few possible approaches to mitigating ancestral degeneracy. Firstly, we could try to limit the number of offspring assigned to any one parent during each resampling step. We can only go so far, because we need the resampling procedure to remain unbiased (Section 2.1.4), but we can try to reduce the variance inherent in the resampling procedure. This idea, known as *low-variance resampling* [citation?] is discussed in detail in Section 2.4.

Another idea is to resample less often. Recall that the reason for resampling is to prevent weight degeneracy (that is, one of the weights tending to one while the others tend to zero). Now we see that, while solving one type of degeneracy, resampling creates another. The effect of ancestral degeneracy is essentially the same as that of weight degeneracy: both drastically increase the variance of the resulting SMC estimators. We can therefore consider a trade-off between the two, which is the idea behind *adaptive resampling* [citation]. The trick is to apply the resampling step only at iterations in which a certain criterion is met. The most commonly-used criterion is based on the estimated *effective sample size*,

$$ESS(t) := \left\{ \sum_{i=1}^{N} (w_t^{(i)})^2 \right\}^{-1},$$

which decreases as the weights degenerate. The resampling step is then applied only at iterations t such that $ESS(t)$ is less than some pre-specified threshold, typically $N/2$ [citation].

If adaptive resampling is used, some trivial changes are required to the calculation of the weights in Algorithm 1, to allow for the importance weights to accumulate sequentially until the particles are resampled. See e.g. Chopin and Papaspiliopoulos (2020, Section 10.2) for details.

As well as mitigating ancestral degeneracy, adaptive resampling has the virtue of saving some computation (although the overall asymptotic complexity of the SMC algorithm remains the same). How effective adaptive resampling is depends on the particular application and choice of SMC algorithm. If the proposals (i.e. transition kernels) are not very close to their targets then the weights will degenerate rapidly and the effective sample size criterion (or similar) will not reduce the frequency of resampling very much.

Low-variance resampling is also less effective under poor proposals: the resulting high-variance weights lead to high-variance offspring counts, even under minimum-variance resmapling schemes, because the resampling is required to be unbiased.

Adaptive resampling and low-variance resampling can be combined, and this is widely considered to be the best practice when implementing SMC. Figure 2.5 compares ancestral degeneration under multinomial resampling (a relatively high variance scheme) to the same under adaptive resampling with a minimum-variance resampling scheme. It is easy to see that the degeneration is much more severe in the former case.

There is one technique that completely solves the problem of ancestral degeneracy, namely *backward simulation* [citation]. This involves running an SMC algorithm as usual (the "forward pass"), and then sampling new ancestors for each particle during an additional "backward pass". The backward-simulated parents in each generation are chosen among all N particles, making use of particles that were not included in the forward-sampled trajectories. In terms of genealogies, the effect is striking; the lineages are broken into fragments, so there is no longer any such thing as a lineage or a genealogy.

Since this work concerns genealogies, we will not say much more about backward simulation. There are many situations in which it is impossible to implement and therefore the study of SMC genealogies is still of interest. Firstly, backward simulation inherently requires a forward and backward pass through all of the data, so it cannot be implemented on-line. Secondly, calculating the backward-simulation probabilities requires the Markov kernels M_t of the corresponding Feynman-Kac model to admit densities which can be evaluated pointwise. This is a much stronger requirement than the ability to simulate from M_t, which is the standard requirement for applying SMC algorithms.

2.3.2 Asymptotic genealogies

If we had access to information about the behaviour of SMC genealogies a priori (i.e. without having run the algorithm), we would be in a position to answer many questions of interest. These include practical questions about tuning, for example:

- How many particles should I use in order to maintain (with high enough probability) a given level of error over a time horizon T?
- With N particles, what is the largest lag over which fixed-lag smoothing produces reasonable estimates?
- How many particles should I use within particle Gibbs to ensure that (with high enough probability) at least two distinct trajectories are generated at each iteration?

This last question touches on a critical aspect of the performance of particle Gibbs algorithms, which is discussed in Section 2.5. We could also consider theoretical questions, such as:

- For a given class of models and algorithms, what is the effect of ancestral degeneracy on how the estimators behave over time?
- Which resampling schemes lead to the smallest amount of ancestral degeneracy?
- What is the effect on genealogies of adaptive resampling?

Many of these questions have already been partially addressed, without any explicit analysis of genealogies, by way of variance calculations and simulation experiments. But since these are all genealogical questions by nature, it seems sensible to work directly with the genealogies, if possible. The problem is that the genealogy of particles is a complex object, it is random, and it can depend strongly on the particular choice of Feynman-Kac model and SMC implementation.

It turns out that these problems can be somewhat overcome by considering the genealogies in an asymptotic regime where the number of particles N tends to infinity. In this regime, many different particle systems exhibit genealogies of a common form, namely Kingman's n-coalescent under suitable time-scales. The differences between various algorithms is then encoded in the time-scale function, which is still random but is a less complicated object than the genealogy itself, namely a càdlàg function rather than a labelled weighted tree. In the context of SMC, these asymptotic genealogies were first analysed by Koskela et al. (2018). The simulations therein suggest that such asymptotic results also transfer to finite systems, making them practically useful.

One of the contributions of the current work is to demonstrate that Kingman-type genealogies arise from a wide variety of SMC algorithms, including those most commonly used in practice. This allows, for instance, genealogies of different SMC algorithms to be compared by examining the corresponding time-scale functions.

2.4 Resampling ∼

⟨sec:resampling⟩ As we have seen, resampling is necessary within SMC to "reset" the weights in order to prevent weight degeneracy. Resampling is itself a Monte Carlo procedure: the discrete offspring counts can be viewed as stochastic estimates of the continuous weights. In order to obtain a valid SMC algorithm, these Monte Carlo samples must be unbiased; this and other desirable properties are formalised in Definition 2.2. There is a huge range of resampling procedures that satisfy these properties, some of which perform better than others. Some of the most popular resampling schemes are introduced in Section 2.4.2 and their properties are explored in Section 2.4.3.

2.4.1 Definition ✓

⟨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 (Corenflos et al. 2021; Myers et al. 2021; Reich 2013) 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 (Efron and Tibshirani 1994; Gordon, Salmond, and Smith 1993) is one of the simplest resampling schemes. The parental indices are chosen independently from $\{1, \dots, N\}$, each with probability given by the weight of the corresponding particle $w_t^{(i)}$. That is,

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

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 properties of the Multinomial distribution that this resampling scheme is unbiased.

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

$$\sum_{j=1}^{a_i - 1} w_t^{(j)} \leq U_i \leq \sum_{j=1}^{a_i} w_t^{(j)}. \tag{2.11}$$ eq:syst_str

This is illustrated in Figure 2.6.

Fast implementations of multinomial resampling rely on $U_1, \dots, U_N$ being pre-sorted, which speeds up the search step (2.11). Sorting N numbers requires $O(N \log N)$ computation, but in fact this is not necessary because we can directly sample the order statistics of a Uniform$[0, 1]$ distribution, at $O(N)$ cost. One way to do this (Chopin and Papaspiliopoulos 2020, Proposition 9.1) is to sample $X_i \sim \text{Exp}(1)$ independently for $i = 1, \ldots, N + 1$ and output the normalised sums

$$U_k := \frac{\sum_{i=1}^{k} X_i}{\sum_{i=1}^{N+1} X_i} \tag{2.12} \{?\}$$

for $k = 1, \ldots, N$, which have the same joint distribution as N order statistics from Uniform$[0, 1]$. Another method is to sample $X_i \sim$ Uniform$[0, 1]$ for $i = 1, \ldots, N$ and compute recursively

$$U_N := X_N^{1/N}, \qquad U_k := X_k^{1/k} U_{k+1} \tag{2.13} \{?\}$$

which also gives the correct joint distribution for $U_1, \ldots, U_N$ (Hol, Schön, and Gustafsson 2006). I'm not sure that this really gives the correct distribution...?

This allows multinomial resampling to be implemented at $O(N)$ cost. A side-effect is that the sampled ancestral indices will be ordered and therefore cannot be Categorically distributed, but the offspring counts still have the correct Multinomial distribution. For the purposes of resampling this isn't usually a problem, but the Categorical distribution can anyway be restored at $O(N)$ cost by applying a random permutation to the offspring indices.

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

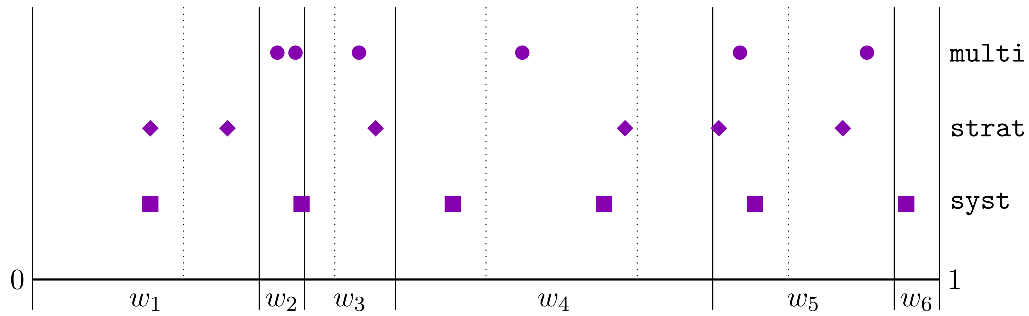


Figure 2.6: Inversion sampling interpretation of multinomial, stratified and systematic resampling. In this example, $N = 6$, $w^{(1:6)} = (0.25, 0.05, 0.1, 0.35, 0.2, 0.05)$ and the uniform random variables input to the resampling schemes are $u_{1:6} = (0.78, 0.29, 0.27, 0.92, 0.54, 0.36)$. The solid vertical lines show the partition of $[0, 1]$ into subintervals of lengths $w^{(1:6)}$. The dotted vertical lines show the partition of $[0, 1]$ into subintervals of length $1/N$, used for stratified and systematic resampling.
Top row (circles): in multinomial resampling, $u_{1:6}$ are fed directly into the inversion sampler. Which subinterval u_i falls into determines the parent of offspring i. The resulting offspring counts in this example are $\nu^{(1:6)} = (0, 2, 1, 1, 2, 0)$.
Middle row (diamonds): in stratified resampling, $u_{1:6}$ are transformed so that one point lies in each subinterval of length $1/N$. The resulting offspring counts are $\nu^{(1:6)} = (2, 0, 1, 1, 2, 0)$.
Bottom row (squares): in systematic resampling, only u_1 is used, being transformed to equally spaced points. The resulting offspring counts are $\nu^{(1:6)} = (1, 1, 0, 2, 1, 1)$.

⟨fig:inv_resampling⟩

$1/N$; that is,

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

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

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

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

The parents are then assigned as in (2.11). This is illustrated in Figure 2.6. 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.

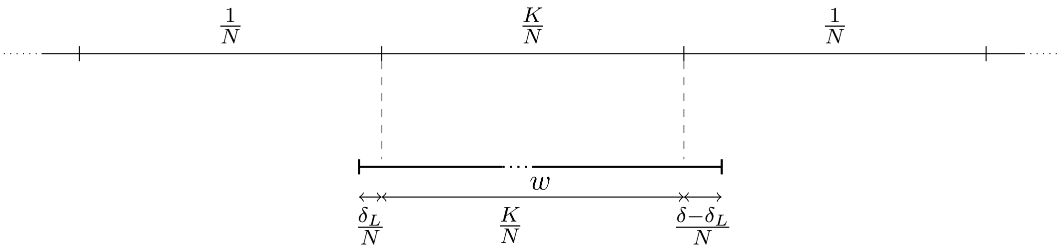
It will be useful later on to have a better idea about the marginal distributions of $\nu_t^{(i)}$ that are induced by stratified resampling. There are complex dependencies between the offspring counts, but we can still find some constraints on the distribution of each count conditional on the corresponding weight. Write the i^{th} weight in the form $w_t^{(i)} = (K+\delta)/N$, where $\delta \in [0,1)$ and $K \in \{0, \ldots, N-1\}$. Considering the illustration Figure 2.6, 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 $j = -1, 0, 1, 2$. 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.

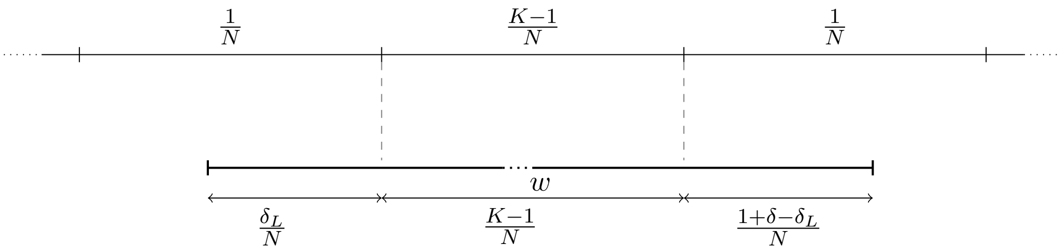
	Case (a)	Case (b)	L.B.	U.B.
p_{-1}	0	$\delta_L(1+\delta-\delta_L)-\delta$	0	$1/4$
p_0	$1-\delta+\delta_L(\delta-\delta_L)$	$1+\delta-2\delta_L(1+\delta-\delta_L)$	$(1-\delta)/2$	$1-3\delta/4$
p_1	$\delta-2\delta_L(\delta-\delta_L)$	$\delta_L(1+\delta-\delta_L)$	$\delta/2$	$(1+\delta)/2$
p_2	$\delta_L(\delta-\delta_L)$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



(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⟩

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 (2.11). The method is illustrated in Figure 2.6.

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. For systematic resampling, Whitley's "roulette wheel" representation is equivalent to that of Figure 2.6.

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

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

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 $\nu_t^{(i)} \in \{0, \ldots, N\}$.

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

Minimal variance branching

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

$$\nu_t^{(i)} \mid w_t^{(i)} \overset{d}{=} \lfloor N w_t^{(i)} \rfloor + \text{Bernoulli}(N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor). \tag{2.14}$$

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 (2.14). 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.

Srinivasan sampling procedure

Gerber, Chopin, and Whiteley (2019) builds on the work of Crisan and Lyons (1999) in that they construct a resampling scheme for which the marginal offspring counts are distributed as (2.14), but the number of particles is held constant and non-negative correlation of offspring counts is enforced. The resulting scheme is termed *Srinivasan sampling procedure resampling*.

The implementation is somewhat complicated compared to the other schemes we have seen (for full details see Gerber, Chopin, and Whiteley 2019, Algorithm 1) but a brief description is given here. The offspring counts are initialised at $Nw_t^{(i)}$, then we iterate through pairs of counts, rounding one of the pair up and the other down by an amount such that at least one of the pair ends up an integer. After at most N such adjustments, all of the counts are integers and can be returned. Each iteration adds and subtracts the same amount so that the sum of the counts is preserved, ensuring that the number of particles remains constant. Which of the selected pair is increased/decreased in each iteration is chosen at random with probabilities that guarantee the resampling is unbiased.

As well as proposing this resampling scheme, Gerber, Chopin, and Whiteley (2019) makes several other contributions to the SMC resampling literature, some of which will be discussed later.

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). 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, as do Srinivasan sampling procedure and minimal variance branching.

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 ✓

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

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

Compare this to star resampling, where the marginal offspring counts

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

having variance

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

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

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

since the stochastic part of $\nu_t^{(i)}$ is a $\mathrm{Bernoulli}(N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)$ random variable (as seen in (2.14)). 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 $R := \sum(N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)$. We have

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

Using the law of total variance,

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

Similarly, for residual resampling with star residuals,

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

and we find

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

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 $w_t^{(i)}$, $\nu_t^{(i)} = \lfloor N w_t^{(i)} \rfloor + j$ with probability p_j for $j = -1, 0, 1, 2$. We can use the expressions for p_{-1}, p_0, p_1, p_2 in the two cases of Figure 2.7, as summarised in Table 2.2, to bound the variance. First write

$$\begin{aligned}
\text{Var}\left[\nu_t^{(i)} \mid w_t^{(i)} \right] &= \mathbb{E}\left[(\nu_t^{(i)} - \lfloor N w_t^{(i)} \rfloor)^2 \mid w_t^{(i)} \right] - \mathbb{E}\left[\nu_t^{(i)} - \lfloor N w_t^{(i)} \rfloor \mid w_t^{(i)} \right]^2 \\
&= p_{-1} + p_1 + 4p_2 - (-p_{-1} + p_1 + 2p_2)^2. \qquad (2.15)
\end{aligned}$$

eq:marg_var

Using the upper and lower bounds in Table 2.2 and then optimising over δ, we obtain the bound

$$\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}] \leq \frac{1}{4} + \frac{1+\delta}{2} + 1 - (0 + \frac{\delta}{2} + 0)^2 = \frac{1}{4}(7 + 2\delta - \delta^2) \leq 2.$$

Optimising the exact expressions in each case (first two columns in Table 2.2) does not improve this overall bound.

Residual-stratified resampling has the further constraint that $p_{-1} = 0$ (i.e. Figure 2.7b doesn't occur) since the residual weights are between 0 and $1/R$. Now the bounds in Table 2.2 are too loose, so we bound the variance by using the exact expressions in each case and optimising over δ_L, δ. By setting p_{-1} to zero in (2.15), substituting the expressions for Case (a) from Table 2.2, and maximising over δ_L and then δ, we get

$$\begin{aligned}
\text{Var}[\nu_t^{(i)} \mid w_t^{(i)}] &= p_1 + 4p_2 - (p_1 + 2p_2)^2 = \delta - 2\delta_L(\delta - \delta_L) + 4\delta_L(\delta - \delta_L) - \delta^2 \\
&= \delta - \delta^2 + 2\delta\delta_L - 2\delta_L^2 \leq \delta - \frac{1}{2}\delta^2 \leq \frac{1}{2}.
\end{aligned}$$

Similarly, for Case (b),

$$\mathrm{Var}[\nu_t^{(i)} \mid w_t^{(i)}] = p_1 + 4p_2 - (p_1 + 2p_2)^2 = \delta_L(1 + \delta - \delta_L) - (\delta_L(1 + \delta - \delta_L))^2 \le \frac{1}{4}.$$

Combining the two cases, we obtain an overall variance bound

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

for residual-stratified resampling.

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. We could also try to bound this variance below, but for every resampling scheme the only lower bound valid for all $w_t^{(i)}$ is zero (consider the case $w_t^{(i)} = 0$).

Contribution to the Monte Carlo variance ✓

While the variance of the offspring counts goes some way towards 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 $(\mathcal{G}_t)_{t\ge0}$ be the filtration generated by the particle positions and weights up to and including time t, so $\mathcal{G}_t$ is the σ-algebra generated by $(X_{0:t}^{(1:N)}, w_{0:t}^{(1:N)})$. Consider the position of the ith particle in generation $t+1$ just after resampling but before mutating, that is $X_t^{(a_t^{(i)})}$. The one-step Monte Carlo variance induced by resampling is defined as

$$\sigma(\varphi) := \mathrm{Var}\left[\frac{1}{N} \sum_{i=1}^{N} \varphi(X_t^{(a_t^{(i)})}) \,\middle|\, \mathcal{G}_t \right] \tag{2.16}$$ 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 ones, are presented in Proposition 2.3. It may be possible to derive similar results regarding residual-stratified and Srinivasan sampling procedure resampling, but such results are hard to obtain due to the strong dependence between parental indices induced by these resampling schemes. This remains an interesting open problem.

In the case of systematic (but not necessarily residual-systematic) resampling, no such variance comparison can be made. Systematic resampling generally yields low variance in practice, but it is possible to construct pathological cases in which it yields higher variance than multinomial resampling (Douc, Cappé, and Moulines 2005, Section 3.4) and it lacks theoretical support more generally (e.g. Gerber, Chopin, and Whiteley 2019, Section 3.3).

hm:resampling_var_compare⟩ **Proposition 2.3** (Variance of resampling schemes)**.** *Let* $\sigma_{\texttt{multi}}$ *etc. denote the variance* (2.16) *under the various resampling schemes, as abbreviated in Table 2.1. For any square-integrable function* φ*,*

⟨item:resampling_var1⟩
(a) $\sigma_{\texttt{multi}}(\varphi) \geq \sigma_{\texttt{res-multi}}(\varphi)$

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

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

⟨item:resampling_var4⟩
(d) $\sigma_{\texttt{res-star}}(\varphi) \geq \sigma_{\texttt{res-multi}}(\varphi) \geq \sigma_{\texttt{res-strat}}(\varphi)$

Proof. **(a)** See Douc, Cappé, and Moulines (2005, Section 3).
(b) See Douc, Cappé, and Moulines (2005, Section 3).
(c) The following expression is derived in Douc, Cappé, and Moulines (2005, Equation (6)):

$$\sigma_{\texttt{multi}}(\varphi) = \frac{1}{N}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})w_t^{(j)} - \frac{1}{N}\left\{\sum_{j=1}^{N}\varphi(X_t^{(j)})w_t^{(j)}\right\}^2.$$

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

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

as required.
(d) The second inequality follows from (b) and is stated in Gerber, Chopin, and Whiteley (2019, p.9). For the first inequality, we use the following expression which is a slight modification of Douc, Cappé, and Moulines (2005, Equation (8)):

$$\sigma_{\texttt{res-multi}}(\varphi) = \frac{R}{N^2}\sum_{j=1}^{N}\varphi^2(X_t^{(j)})r^{(j)} - \frac{R}{N^2}\left(\sum_{j=1}^{N}\varphi(X_t^{(j)})r^{(j)}\right)^2.$$

A derivation similar to theirs can also be used for residual-star resampling. First notice that, conditional on $\mathcal{G}_t$, the Monte Carlo estimate in (2.16) can be decomposed into a sum of conditionally deterministic terms plus a sum of stochastic terms:

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

where the terms in the second sum are all equal, say $\hat{X}_t^{(1)} = \cdots = \hat{X}_t^{(R)} = X_t^\star$. The first sum is conditionally deterministic and hence does not contribute to the Monte Carlo variance (2.16). We have

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

whenever $R \geq 1$. If $R = 0$ then all residual schemes have zero variance and (d) holds trivially. ■

Exchangeability of offspring ✓

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

It is clear that true multinomial resampling satisfies this property since the parental indices are independent and distributed according to the same Categorical distribution. The same goes for star resampling. However, as mentioned earlier, the efficient implementation of multinomial resampling that takes sorted inputs does not leave the offspring exchangeable. Stratified and systematic resampling do not either since, for instance, child 1 is more likely to choose parent 1 than child N is. Residual resampling schemes are also typically implemented in such a way that the offspring are not exchangeable.

Whichever resampling scheme is used, exchangeability of offspring can easily be reintroduced (at $O(N)$ cost) by applying a random permutation to the vector of parental indices after sampling.

Operations in SMC that depend on the ancestral indices are typically independent of ordering, so outputting sorted ancestral indices after resampling is not expected to cause any problem. (?) However, the results of Chapters 3 and 4 rely on assumption (A1) which amounts to exchangeability of offspring, so to be sure that the current genealogical study applies, a permutation should be appended to any non-exchangeable resampling procedure.

Permutation sensitivity ~

Some resampling schmes are sensitive to the order of the weights. That is, permuting the weight vector before resampling can affect the distribution of the resulting offspring counts. Note that this entails a permutation of the parents, whereas the previous section was about permutations of offspring.

For example, consider resampling schemes based on inversion sampling (multinomial, stratified, systematic). Figure 2.8 shows two partitions of $[0, 1]$ constructed from two permutations of the weight vector $w^{(1:6)} = (0.25, 0.05, 0.1, 0.35, 0.2, 0.05)$. This does not affect the distribution of the offspring counts under multinomial resampling, although it will affect the distribution of the parental indices if the fast implementation is used.

On the other hand, under stratified or systematic resampling the distribution of offspring counts is different for the two partitions. To see this, consider parents 2 and 6. When the weights are sorted, the probability that both of these parents are assigned a non-zero number of offspring is zero, because both of their subintervals lie within the same subinterval of length $1/N$, which gets exactly one inversion sampling point. When the weights are in their natural order, as in the top row of Figure 2.8, it is possible under stratified and systematic resampling for both parents 2 and 6 to be assigned one offspring. Clearly, then, the distribution of offspring counts under these resampling schemes differs between the two orderings of the weight vector pictured.

What effect does permutation sensitivity have on performance? Which other schemes do/don't have this property?

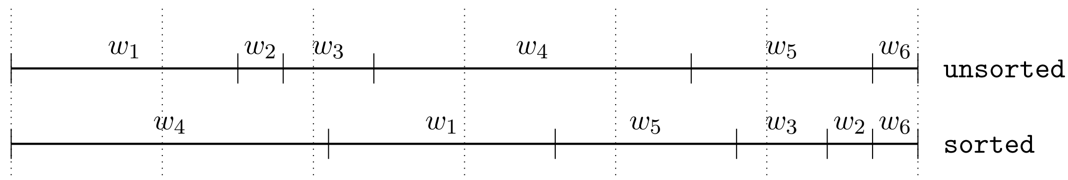


Figure 2.8: An example for which permuting the weights can affect the conditional distribution of offspring counts under some resampling schemes. The example for $N = 6$ used in Figure 2.6. The top row shows the partition into weighted subintervals, in the natural order, as in Figure 2.6. The bottom row shows the partition corresponding to the same weights, but sorted in decreasing order. The dotted lines are spaced $1/N$ apart. Under "permutation-sensitive" resampling schemes, the distribution of offspring counts differs depending on which partition is used.

permutation_sensitivity⟩

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. 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 a random number less than N. Their 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 ✓

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,

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

Gerber, Chopin, and Whiteley (2019) show that negative association of offspring counts is a desirable property which may be used, along with some other machinery, to establish certain weak convergence results for the resampled measures.

Multinomial counts are negatively associated (Joag-Dev and Proschan 1983, Section 3.1), which implies that residual-multinomial resampling also satisfies this property why is that? idea: deterministic parts have 0 correlation and the random parts are NA. same question for the other residual schemes mentioned below. Gerber, Chopin, and Whiteley (2019) construct a counter-example to demonstrate that systematic resampling violates the negative association property. For residual-systematic resampling, we can cook up a counterexample in the same spirit by taking $\varphi(x) = \psi(x) = \mathbb{1}_{\{x=1\}}$, $I = \{1\}$, $J = \{3\}$ and considering a weight vector say $w^{(1:4)} = \frac{1}{8}(1, 1, 1, 5)$ for $N = 4$. Then the residual weights are $r^{(1:4)} = \frac{1}{4}(1,1,1,1)$ with $R = 2$, so

$$\begin{aligned}\mathrm{Cov}\left[\varphi(Z_I), \psi(Z_J)\right] &= \mathbb{E}[\varphi(Z_I)\psi(Z_J)] - \mathbb{E}[\varphi(Z_I)]\mathbb{E}[\psi(Z_J)] \\ &= \mathbb{P}[\nu^{(1)} = 1, \nu^{(3)} = 1] - \mathbb{P}[\nu^{(1)} = 1]\mathbb{P}[\nu^{(3)} = 1] \\ &= \frac{1}{2} - \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} > 0,\end{aligned}$$

since $\nu^{(3)} = 1$ if and only if $\nu^{(1)} = 1$. So residual-systematic resampling also violates the negative association property.

Gerber, Chopin, and Whiteley (2019) also mention some resampling schemes that do result in negatively associated counts: stratified resampling, and by implication residual-stratified resampling; star resampling (see the remark at the end of Gerber, Chopin, and Whiteley (2019, Section 3.2)), and by implication residual-star resampling. The authors go on to introduce the Srinivasan sampling procedure resampling scheme, which by construction yields negatively associated offspring counts.

These results are summarised in Table 2.3. The minimal variance branching algorithm of Crisan and Lyons (1997) does not enforce negative association, so this property depends on the particular implementation, and as such is left blank in Table 2.3.

Star discrepancy ✓

The *star discrepancy* 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, as in Kuipers and Niederreiter (1974, Definition 1.2):

$$D^\star(u_1, \dots, u_N) := \sup_{u\in[0,1]} |d(u)| := \sup_{u\in[0,1]} \left| u - \frac{1}{N}\sum_{i=1}^N \mathbb{1}_{\{u_i \leq u\}} \right|. \tag{2.17}$$ eq:defn_sta

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 $[0,1]$. Thus $D^\star$ measures, in a certain sense, how far the points are from being uniformly spaced.

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 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" point sets to be used in inversion sampling.

In one dimension, the lowest-discrepancy point set is the regular grid $(\frac{1}{2N}, \frac{3}{2N}, \dots, \frac{2N-1}{2N})$, which has star discrepancy $\frac{1}{2N}$ (see for example Kuipers and Niederreiter 1974, Corollary 1.2). However, resampling based on a deterministic point set cannot be unbiased since the resulting parental indices are conditionally deterministic given the weights. Systematic resampling amounts to a randomisation of the regular grid, shifting each grid point by a random amount $u \sim \mathrm{Uniform}[0, 1/N]$. This yields star discrepancy $D^\star = \max\{u, \frac{1}{N} - u\}$,

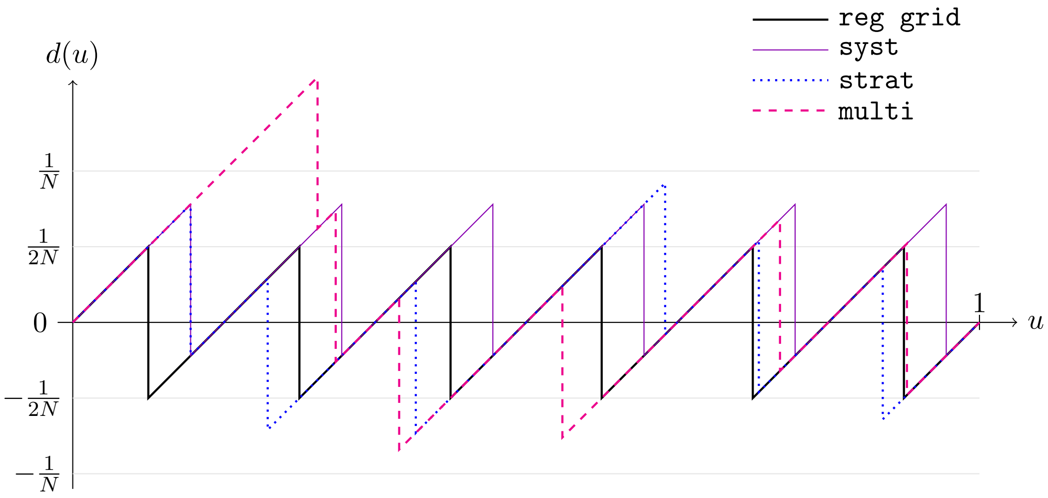


Figure 2.9: Plot of the function inside the supremum in (2.17), for four different point sets. The points $u_{1:6}$ used are the same as in Figure 2.6.
The solid black line corresponds to the regular grid, which achieves the minimal discrepancy $1/(2N)$, but cannot be used for resampling. The star discrepancy of stratified and systematic points varies between $1/(2N)$ and $1/N$ depending on the realisation. In this example, the star discrepancy of the systematic points is $0.78/N$ and of the stratified points is $0.92/N$. The star discrepancy of standard multinomial resampling (that is, i.i.d. Uniform points) can be arbitrarily close to 1 for "bad" realisations; in this example it is $1.62/N$.

⟨fig:star_discrepancy⟩

which is between $1/(2N)$ and $1/N$ almost surely. The point sets generated in stratified resampling also have star discrepancy between $1/(2N)$ and $1/N$, 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. Figure 2.9 illustrates how the star discrepancy is computed, and how it compares between these sampling methods.

	support of $\nu_t^{(i)}$ given $\frac{k}{N} \leq w_t^{(i)} < \frac{k+1}{N}$	$\sup_w$ $\lvert\nu_t^{(i)} - N w_t^{(i)}\rvert$	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	neg. assoc.?
`multi`	$\{0,\ldots,N\}$	N	$N/4$	$\times$	$\times$	$\times$	N	$\checkmark$
`star`	$\{0,N\}$	N	$N^2/4$	$\times$	$\times$	$\times$	1	$\checkmark$?
`strat`	$\{k-1,k,k+1,k+2\}$	2	2	$\times$	$\checkmark$	$\checkmark$	N	$\checkmark$
`syst`	$\{k,k+1\}$	1	1/4	$\checkmark$	$\checkmark$	$\checkmark$	1	$\times$
`res-multi`	$\{k,\ldots,N\}$	N	1	$\times$	$\checkmark$	$\times$	$\leq N$	$\checkmark$
`res-star`	$\{k,N\}$	N	N	$\times$	$\checkmark$	$\times$	1	$\checkmark$?
`res-strat`	$\{k,k+1,k+2\}$	2	1/2	$\times$	$\checkmark$	$\checkmark$	$\leq N$	$\checkmark$
`res-syst`	$\{k,k+1\}$	1	1/4	$\checkmark$	$\checkmark$	$\checkmark$	1	$\times$
`ssp`	$\{k,k+1\}$	1	1/4	$\checkmark$	$\checkmark$	$\checkmark$?	?	$\checkmark$
`branch`	$\{k,k+1\}$	1	1/4	$\checkmark$	$\checkmark$			

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 $X = (X_1, \dots, X_N)$ be a $\mathbb{R}_+^N$-valued random variable. Then $Y = (Y_1, \dots, Y_N) \in \mathbb{N}^N$ 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

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

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, $\mathrm{Var}[\nu_t^{(i)} \mid w_t^{(i)}]$ 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. $w_t^{(1:N)} = (1, \dots, 1)/N$.

2.5 Conditional SMC ✓

⟨sec:condSMC⟩

Be consistent with upper/lower case letters: X, x etc.

Andrieu, Doucet, and Holenstein (2010) propose a number of “particle MCMC” algorithms, which combine SMC with MCMC in order to improve performance in certain situations. One of their algorithms, the *particle Gibbs* sampler (Andrieu, Doucet, and Holenstein 2010, Section 2.4.3), is of particular interest in the current work. For one thing, genealogies are particularly critical to its performance, and for another, the particle step uses a variant SMC algorithm which alters the distribution of genealogies.

In this section, we first introduce the particle Gibbs algorithm and the conditional SMC update, then discuss how ancestral degeneracy impacts the performance of particle Gibbs and how ancestor sampling mitigates this.

2.5.1 Particle Gibbs

To motivate the particle Gibbs algorithm, we introduce a parametrised state space model and explain how combining SMC updates with MCMC sampling allows us to tackle the related inferences effectively. The particle Gibbs algorithm can be applied much more broadly, but this application is particularly intuitive and exhibits all the features of interest to our genealogical study.

Consider a parametrised state space model of the form

$$\begin{aligned}
\theta &\sim p(\cdot) \\
X_0 &\sim \mu^\theta(\cdot) \\
X_{t+1} \mid X_t &\sim K^\theta_{t+1}(\cdot \mid X_t) && \text{for } t = 0, \dots, T-1 \\
Y_t \mid X_t &\sim g^\theta_t(\cdot \mid X_t) && \text{for } t = 0, \dots, T
\end{aligned}$$

exactly like (2.1) except that the specification is now parametrised by θ (which may be multi-dimensional), and we place a prior distribution on θ. As usual, p, μ^θ, (K^θ_t) and (g^θ_t) are part of the model and are assumed to be known but not necessarily tractable.

Suppose that, given some data $y_{0:T}$, we wish to generate Monte Carlo samples from the joint posterior distribution of $X_{0:T}$ and θ. (Even if we are only interested in inferring θ, for instance, it is often more practical to target the joint posterior and then marginalise.) Notice that we are now working with a finite time horizon $T \in \mathbb{N}$. The inference of interest here is not inherently sequential; we are building an MCMC algorithm to sample from a single target distribution which happens to include some sequentially correlated components.

The conditional dependence structure of the model invites the use of a (partially-collapsed) Gibbs sampler, sampling alternately from the conditional distributions $p(\theta \mid x_{0:T}, y_{0:T})$ and $p(x_{0:T} \mid \theta, y_{0:T})$. The θ update,

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

is often quite straightforward, if not analytically then by employing a Metropolis-Hastings step based on the current sampled values of θ and $x_{0:T}$. The X update, meanwhile, is high-dimensional with strong sequential correlations: exactly the situation in which one might use SMC. For the X update, we need a sample from

$$p(dx_{0:T} \mid \theta, y_{0:T}) \propto \mu^\theta(dx_0) g^\theta_0(y_0 \mid x_0) \prod_{s=1}^{T} K^\theta_s(dx_s \mid x_{s-1}) g^\theta_s(y_s \mid x_s), \tag{2.18}$$

eq:PG_Xpost

which can be obtained by running an SMC smoother then sampling one trajectory from its output in proportion to the associated weight.

However, the Markov chain associated to the procedure just described does not admit $p(x_{0:T}, \theta \mid y_{0:T})$ as an invariant distribution. It *approximately* targets this distribution, with some bias. A Gibbs sampler targeting $p(x_{0:T}, \theta \mid y_{0:T})$ exactly can be constructed by replacing the SMC step with a *conditional SMC* step, which takes into account the value of $x_{0:T}$ sampled at the previous iteration, as well as the observations and the current value of θ.

A conditional SMC algorithm for this scenario is presented in Algorithm 2. In contrast to Algorithm 1, the input now includes $x^\star_{0:T}$ and $a^\star_{0:T}$, which encode the states and parental indices, respectively, of the *immortal trajectory* (so called because it "survives" the SMC run with probability one). Within a particle Gibbs algorithm, the immortal trajectory is set to the trajectory sampled at the previous iteration. The resampling step now assigns the immortal offspring to the immortal parent deterministically, and the state of the immortal particle is also updated deterministically rather than via the Markov kernel. As in standard SMC, there is a choice of RESAMPLE procedures, but some care is needed to ensure the correct treatment of the immortal particle (Lee, Murray, and Johansen 2019).

Input: $T, N, \mu^\theta, (K^\theta_t), (g^\theta_t), y_{0:T}, x^\star_{0:T}, a^\star_{0:T}$
Set $X_0^{(a_0^\star)} \leftarrow x_0^\star$
for $i \in \{1, \ldots, N\} \setminus a_0^\star$ **do** Sample $X_0^{(i)} \sim \mu(\cdot)$
for $i \in \{1, \ldots, N\}$ **do** $w_0^{(i)} \leftarrow \left\{\sum_{j=1}^N g_0^\theta(y_0 \mid X_0^{(j)})\right\}^{-1} g_0^\theta(y_0 \mid X_0^{(i)})$
for $t \in \{1, \ldots, T\}$ **do**
  Set $a_{t-1}^{(a_t^\star)} \leftarrow a_{t-1}^\star$, $X_t^{(a_t^\star)} \leftarrow x_t^\star$
  Sample $a_{t-1}^{(1:N)} \setminus a_{t-1}^\star \sim \text{RESAMPLE}(\{1, \ldots, N\}, w_{t-1}^{(1:N)})$
  for $i \in \{1, \ldots, N\} \setminus a_t^\star$ **do** Sample $X_t^{(i)} \sim K_t^\theta(\cdot \mid X_{t-1}^{(a_{t-1}^{(i)})})$
  for $i \in \{1, \ldots, N\}$ **do** $w_t^{(i)} \leftarrow \left\{\sum_{j=1}^N g_t^\theta(y_t \mid X_t^{(j)})\right\}^{-1} g_t^\theta(y_t \mid X_t^{(i)})$
end

⟨alg:condSMC⟩**Algorithm 2:** Conditional sequential Monte Carlo for a parametrised state space model. The immortal particle at each generation has its new state and parental index set deterministically according to the values of $x^\star_{0:T}$ and $a^\star_{0:T}$ given as input.

The complete particle Gibbs algorithm for this example then consists of alternately sampling from the full conditional distribution of θ (e.g. using a Metropolis-Hastings update) and sampling a trajectory $(x_{0:T}, a_{0:T})$ using conditional SMC. See Andrieu, Doucet, and Holenstein (2010, Section 2.4.3) for more details.

2.5.2 Ancestral degeneracy in particle Gibbs

We have seen in Section 2.3 that the phenomenon of ancestral degeneracy can severely affect the performance of SMC algorithms, particularly in smoothing applications. The SMC update of particle Gibbs is a smoothing problem, however it requires only one sampled trajectory from the smoothing distribution, so one might imagine that we are safe from the curse of ancestral degeneracy. In fact, the loss of ancestors causes a different problem for particle Gibbs: it prevents some components of the Markov chain being refreshed, so that the chain mixes slowly.

To see this, consider the illustration in Figure 2.10, which shows the smoothing trajectories generated by a conditional SMC update at some iteration r. The thick black line is the immortal trajectory given as input, that is, the trajectory sampled by the conditional SMC update at iteration $r-1$. Backwards in time, the sampled trajectories quickly coalesce until at time 20 all of the trajectories have coalesced. The common trajectory from time 0 to 20 must necessarily be part of the immortal trajectory. A new trajectory (highlighted in purple) is then sampled among the N generated trajectories. Whichever trajectory we sample, it will certainly overlap with the previously sampled trajectory at least from time 0 to 20.

At the next iteration the newly sampled trajectories will again coalesce onto the immortal trajectory, and this behaviour is repeated. If T is too large with respect to N, the early part of the trajectory will very rarely be updated, so the corresponding states will mix very slowly. For further intuition on this phenomenon the reader is directed to Lindsten and Schön (2013, Section 5.4) which provides a very clear exposition.

The meaning of T "too large" here depends on the model and the type of SMC update used, but typically T is determined by the application and N is limited by computational resources, so we may not be able to control their relative size. The other brute-force approach would be to increase the number of iterations of the MCMC algorithm, but this too is infeasible on a limited computational budget. It is therefore worth investing some effort to find an alternative solution to the problem of ancestral degeneracy within particle Gibbs.

2.5.3 Ancestor sampling

⟨sec:ancsamp⟩ An effective solution (where it is possible to implement it) was proposed by Whiteley (2010) and is known as *ancestor sampling*. It consists of a simple modification to the resampling step within the conditional SMC algorithm. In the basic algorithm with multinomial resampling, at each time step the non-immortal particles are resampled by multinomial resampling according to their weights, while the immortal offspring is deterministically assigned to the immortal parent. That is, at time t, for each $i \in \{1, \ldots, N\}$,

$$\mathbb{P}[a_t^{(j)} = i \mid X_{0:t}^{(1:N)}, x_{0:T}^\star, a_{0:T}^\star] \propto \begin{cases} w_t^{(i)} & j \text{ non-immortal} \\ \mathbb{1}_{\{i=a_t^\star\}} & j \text{ immortal.} \end{cases}$$

Ancestor sampling combines the resampling step with a backward simulation step for the immortal particle. Instead of deterministically inheriting the "correct" parent, the immortal particle samples its parent among all N possible parents. This is justified in the same way as backwards simulation in general (Section ??), provided the ancestor sampling probabilities are chosen correctly, although we now apply the backwards sampling step to the immortal trajectory only. Ancestor sampling can also be implemented for other

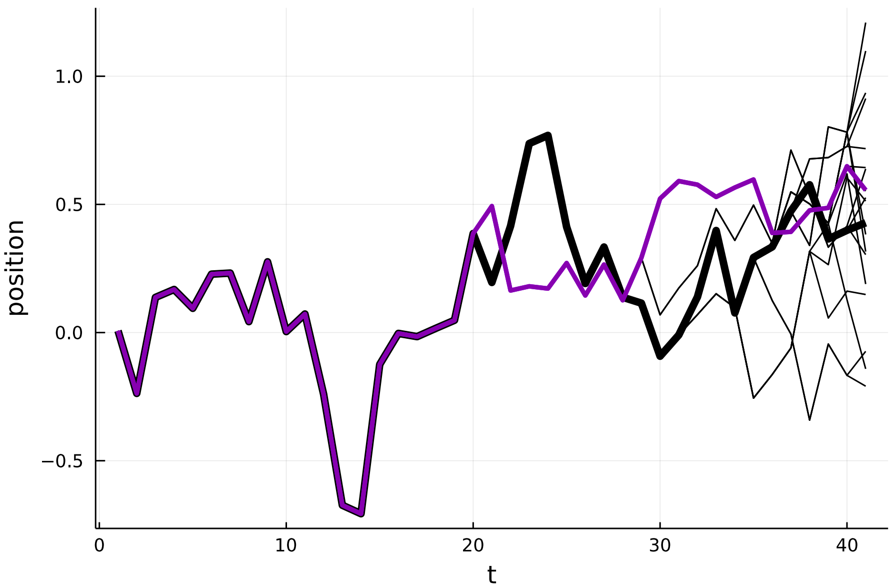


Figure 2.10: Illustration of how ancestral degeneracy causes particle Gibbs to mix slowly on some components. The thick black line is the immortal trajectory, i.e. the sampled trajectory from the previous iteration. Other lines are all of the trajectories generated by conditional SMC. One of these (highlighted in purple) is the sampled trajectory at the current iteration. Due to ancestral degeneracy, the current sample (purple) coincides with the previous sample (thick black) up to time 20, so the components $x_{0:20}$ are not updated in this iteration.

⟨fig:PG_ancdegen⟩

choices of RESAMPLE, using the same backward simulation probabilities (but of course the resampling probabilities for non-immortal particles will be different, and there may be some additional dependence between parental indices). For simplicity we here restrict ourselves to multinomial resampling.

Input: $T, N, \mu^\theta, (K_t^\theta), (g_t^\theta), y_{0:T}, x^\star_{0:T}, a^\star_{0:T}$
Set $X_0^{(a_0^\star)} \leftarrow x_0^\star$
for $i \in \{1, \dots, N\} \setminus a_0^\star$ **do** Sample $X_0^{(i)} \sim \mu^\theta(\cdot)$
for $i \in \{1, \dots, N\}$ **do** $w_0^{(i)} \leftarrow \left\{\sum_{j=1}^N g_0^\theta(y_0 \mid X_0^{(j)})\right\}^{-1} g_0^\theta(y_0 \mid X_0^{(i)})$
for $t \in \{1, \dots, T\}$ **do**
  Set $X_t^{(a_t^\star)} \leftarrow x_t^\star$
  Sample $a_{t-1}^{(a_t^\star)} \sim \text{Categorical}\left(\{1, \dots, N\}, w_{t-1}^{(1:N)} q_t^\theta(x_t^\star \mid X_{t-1}^{(1:N)})\right)$
  Sample $a_{t-1}^{(1:N)} \setminus a_{t-1}^\star \sim \text{RESAMPLE}(\{1, \dots, N\}, w_{t-1}^{(1:N)})$
  for $i \in \{1, \dots, N\} \setminus a_t^\star$ **do** Sample $X_t^{(i)} \sim q_t^\theta(\cdot \mid X_{t-1}^{(a_{t-1}^{(i)})})$
  for $i \in \{1, \dots, N\}$ **do** $w_t^{(i)} \leftarrow \left\{\sum_{j=1}^N g_t^\theta(y_t \mid X_t^{(j)})\right\}^{-1} g_t^\theta(y_t \mid X_t^{(i)})$
end

⟨alg:condSMC_ancsamp⟩**Algorithm 3:** Conditional sequential Monte Carlo with ancestor sampling for a parametrised state space model. The parent of the "immortal particle" is updated at each iteration via an on-line backward simulation step. The second parameter of the Categorical variable should be interpreted element-wise.

Assume that the smoothing distributions admit densities, that is $\mu^\theta(\cdot)$ and $K_t^\theta(\cdot \mid x)$ admit densities for all x, t. Denote the density of K_t^θ by q_t^θ. Define the trajectories $X_{t,0:t}^{(i)}$ (for any t, i) as in Section 2.1.4, starting from $X_{t,t}^{(i)} := X_t^{(i)}$ and tracing back the states of the parents via $X_{t,s}(i) = X_{t,s+1}^{(a_t^{(i)})}$. Then the correct resampling probabilities are, for each i,

$$\mathbb{P}[a_t^{(j)} = i \mid X_{0:t}^{(1:N)}, x^\star_{0:T}, a^\star_{0:T}] \propto \begin{cases} w_t^{(i)} & j \text{ non-immortal} \\ w_t^{(i)} \frac{p((X_{t,0:t}^{(i)}, x^\star_{t+1:T}) \mid \theta, y_{0:T})}{p(X_{t,0:t}^{(i)} \mid \theta, y_{0:t})} & j \text{ immortal.} \end{cases} \tag{2.19}$$ eq:ancsamp_

The ratio of densities can be interpreted as the conditional probability that the whole trajectory is the concatenation of $X_{t,0:t}^{(i)}$ with $x^\star_{t+1:T}$, given that its first $t+1$ states are $X_{t,0:t}^{(i)}$. To simplify the ratio, use (2.18) to write

$$p(X_{t,0:t}^{(i)} \mid \theta, y_{0:t}) \propto \mu^\theta(X_{t,0}^{(i)}) g_0^\theta(y_0 \mid X_{t,0}^{(i)}) \prod_{s=1}^{t} q_s^\theta(X_{t,s}^{(i)} \mid X_{t,s-1}^{(i)}) g_s^\theta(y_s \mid X_{t,s}^{(i)})$$

and

$$p((X^{(i)}_{t,0:t}, x^\star_{t+1:T}) \mid \theta, y_{0:T}) \propto \mu^\theta(X^{(i)}_{t,0}) g^\theta_0(y_0 \mid X^{(i)}_{t,0}) \left\{ \prod_{s=1}^{t} q^\theta_s(X^{(i)}_{t,s} \mid X^{(i)}_{t,s-1}) g^\theta_s(y_s \mid X^{(i)}_{t,s}) \right\}$$
$$\times\, q^\theta_{t+1}(x^\star_{t+1} \mid X^{(i)}_{t,t}) g^\theta_{t+1}(y_{t+1} \mid x^\star_{t+1}) \left\{ \prod_{s=t+2}^{T} q^\theta_s(x^\star_s \mid x^\star_{s-1}) g^\theta_s(y_s \mid x^\star_s) \right\}.$$

The ratio then becomes

$$\begin{aligned} \frac{p((X^{(i)}_{t,0:t}, x^\star_{t+1:T}) \mid \theta, y_{0:T})}{p(X^{(i)}_{t,0:t} \mid \theta, y_{0:t})} &\propto q^\theta_{t+1}(x^\star_{t+1} \mid X^{(i)}_{t,t}) g^\theta_{t+1}(y_{t+1} \mid x^\star_{t+1}) \prod_{s=t+2}^{T} q^\theta_s(x^\star_s \mid x^\star_{s-1}) g^\theta_s(y_s \mid x^\star_s) \\ &\propto q^\theta_{t+1}(x^\star_{t+1} \mid X^{(i)}_{t,t}) \\ &= q^\theta_{t+1}(x^\star_{t+1} \mid X^{(i)}_{t}). \end{aligned}$$

The probabilities in (2.19) become

$$\mathbb{P}[a^{(j)}_t = i \mid X^{(1:N)}_{0:t}, x^\star_{0:T}, a^\star_{0:T}] \propto \begin{cases} w^{(i)}_t & j \text{ non-immortal} \\ w^{(i)}_t q^\theta_{t+1}(x^\star_{t+1} \mid X^{(i)}_t) & j \text{ immortal.} \end{cases} \tag{2.20}$$ eq:ancsamp_

The conditional SMC algorithm with this adaptation is presented in Algorithm 3.

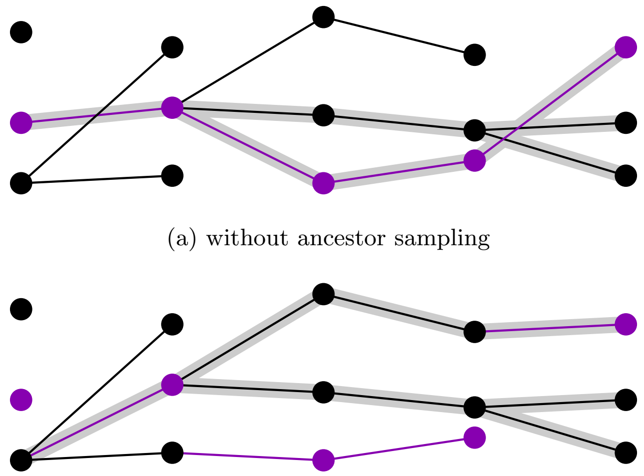
We see that, in order to do ancestor sampling, we need a stronger assumption on the Markov kernels than was required to simply run the conditional SMC algorithm: we now require that, for each t, K^θ_t admits a density q^θ_t and that $q^\theta_t(\cdot \mid x)$ can be evaluated pointwise for any x, whereas previously we only needed to draw samples from $K^\theta_t(\cdot \mid x)$ for any x. This additional requirement rules out ancestor sampling in some applications.

Recall that the usual backward simulation procedure requires a full forward pass to calculate the future states before the backward simulation probabilities can be computed. Ancestor sampling, on the other hand, does not require a forward pass because it only computes backward simulation probabilities for the immortal trajectory, for which all the future states are known in advance. This means that the additional computational cost of implementing ancestor sampling is negligible.

Why ancestor sampling works

We know that complete backward simulation eradicates ancestral degeneracy by breaking up the lineages into disjoint pieces (Section ??). But here we are only backward-simulating one of the N particles, leaving the other $N-1$ lineages to coalesce as usual. So how does this help?

Recall that in particle Gibbs ancestral degeneracy is not itself a problem, because we only require a single sample from the smoothing distribution. The problem is that the consecutive samples are highly correlated, because of the repeated coalescence onto the immortal lineage. The contribution of ancestor sampling is to break up the immortal trajectory so that it no longer appears in the sampled lineages; see Figure 2.11. While the non-immortal trajectories may still coalesce, they no longer preferentially coalesce onto the immortal trajectory. In turn, the sampled trajectory that is output does not overlap unduly with the immortal trajectory that was the previous output, and this completely solves the problem of the slow-mixing particle Gibbs chain.



(a) without ancestor sampling

(b) with ancestor sampling

Figure 2.11: Illustration of how ancestor sampling prevents coalescence onto the immortal trajectory. Immortal particles are highlighted in purple, along with their parent-offspring edges (the given ones in (a) and the ancestor-sampled ones in (b)). The resulting lineages of the terminal particles are highlighted in grey. In (a), the lineages of the terminal particles coalesce onto the immortal trajectory. Imagine time stretching further back: the lineages would continue to coincide with the immortal trajectory forever. In (b), the lineages still coalesce, but not onto the immortal trajectory. The immortal trajectory no longer exists as a lineage.

⟨fig:whyASworks⟩

3 Convergence of Finite-Dimensional Distributions

⟨ch:limits⟩ Add some words here to link from previous chapter once that is written.

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.

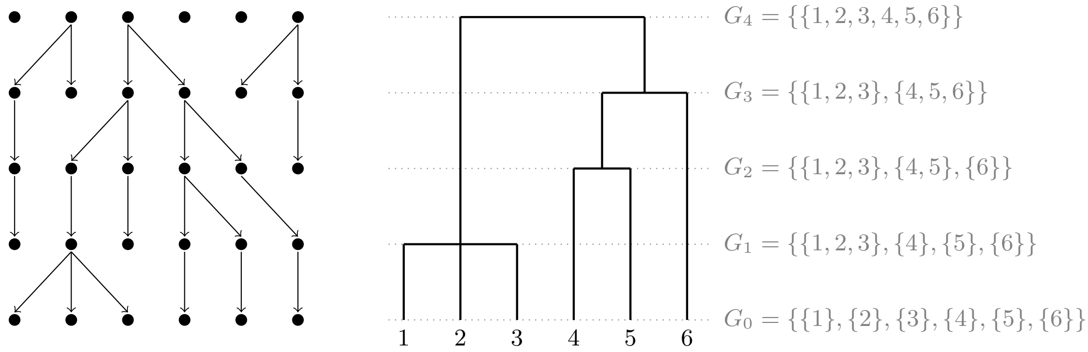
Let $\mathcal{P}_n$ be the space of partitions on $\{1,\dots,n\}$. For convenience, we now label time in reverse, so the terminal particles are at time 0, their parents are at time 1, and so on. Consider a randomly chosen sample of n terminal particles among a total of N particles, and label the sampled 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



(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⟩

merger rate

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

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 $0 \leq f(r) \leq B$. Then*

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

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 $K \geq 1$ and note that $\tau_N(t) \wedge K$ is a bounded $\mathcal{F}_t$-stopping time. Hence we can apply the optional stopping theorem:

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

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

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

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 $\{1, \ldots, n\}$. For any $\xi, \eta \in \mathcal{P}_n$ and $t \in \mathbb{N}$, let $p_{\xi\eta}(t)$ denote the conditional transition probabilities of the genealogical process given $\nu_t^{(1:N)}$. 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 $i \in \{1, \dots, |\eta|\}$. Hence $b_1 + \dots + b_{|\eta|} = |\xi|$. Then the transition probability is given by

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

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

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 $\kappa_1 + 2\kappa_2 + \dots + |\xi|\kappa_{|\xi|} = |\xi|$). 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 $(b_1, b_2, \dots, b_{|\xi|-1}) = (2, 1, \dots, 1)$ 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 $(G_{\tau_N(t)}^{(n,N)})_{t\geq 0}$ 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

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

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:mainthmcon

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

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) ⇒ (3.6) ⇒ (3.7) rather than (3.11) ⇒ (3.6) and (3.11) ⇒ (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.

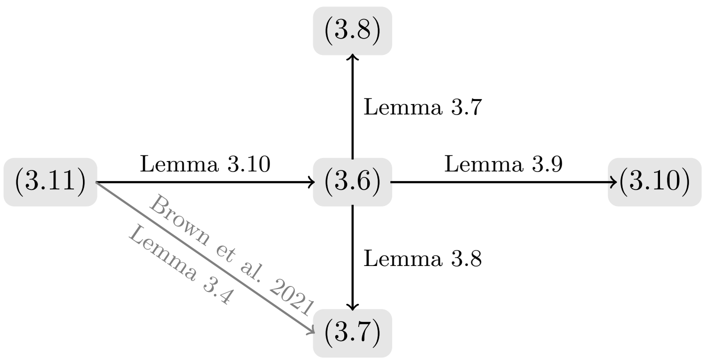


Figure 3.2: Dependencies between conditions of Theorems 3.5 and 3.6. Arrows represent logical implication; labels on arrows indicate the lemma in which the implication is stated. In Brown et al. (2021, Lemma 3.4) the direct implication (3.11) ⇒ (3.7) was proved, but here we will instead show that (3.6)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

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

The second is

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

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

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

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

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

Substituting (3.16) into (3.15) gives

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

and substituting (3.17) into (3.14) gives

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

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

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

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

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

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{E}\left[\prod_{d=1}^{k} \chi_d\right] \geq & \sum_{\beta_1=0}^{\infty} \dots \sum_{\beta_k=0}^{\infty} \sum_{\substack{(\lambda_1,\mu_1)\in\Pi_2([\alpha_1+\beta_1]): \\ |\mu_1|=\beta_1}} \dots \sum_{\substack{(\lambda_k,\mu_k)\in\Pi_2([\alpha_k+\beta_k]): \\ |\mu_k|=\beta_k}} \\
& \left\{ \prod_{d=1}^{k} (\tilde{Q}^{\alpha_d})_{\eta_{d-1}\eta_d} \prod_{r\in\mu_d} - \binom{|\eta_{d-1}| - |\{i \in \lambda_d \cup \pi_d : i < r\}|}{2} \right\} \\
& \times \lim_{N\to\infty} \mathbb{E}\Bigg[\left(\prod_{r=\tau_N(t_{d-1})+1}^{\tau_N(t_d)} \mathbb{1}_{\{c_N(r) > \binom{n-2}{2} D_N(r)\}} \mathbb{1}_{\{c_N(r) < (\tilde{B}_n + \binom{n}{2})^{-1}\}} \right) \\
& \times \sum_{\substack{s_1^{(1)} < \dots < s_{\alpha_1+\beta_1}^{(1)} \\ = \tau_N(t_0)+1}}^{\tau_N(t_1)} \dots \sum_{\substack{s_1^{(k)} < \dots < s_{\alpha_k+\beta_k}^{(k)} \\ = \tau_N(t_{k-1})+1}}^{\tau_N(t_k)} \\
& \prod_{d=1}^{k} \mathbb{1}_{\{\tau_N(t_d)-\tau_N(t_{d-1}) \geq \alpha_d+\beta_d\}} \left\{ \prod_{r\in\lambda_d\cup\mu_d} c_N(s_r^{(d)}) \right\} \Bigg]. \qquad (3.19)\ \boxed{\texttt{eq1}}
\end{aligned}$$

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

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

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

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

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

⟨ch:weakconv⟩ 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

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

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

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

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 $T_i := \varpi_1 + \dots + \varpi_i$. 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 $t_1, \dots, t_k > 0$,

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

Take the basis case $k = 1$. Then

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

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. ∎

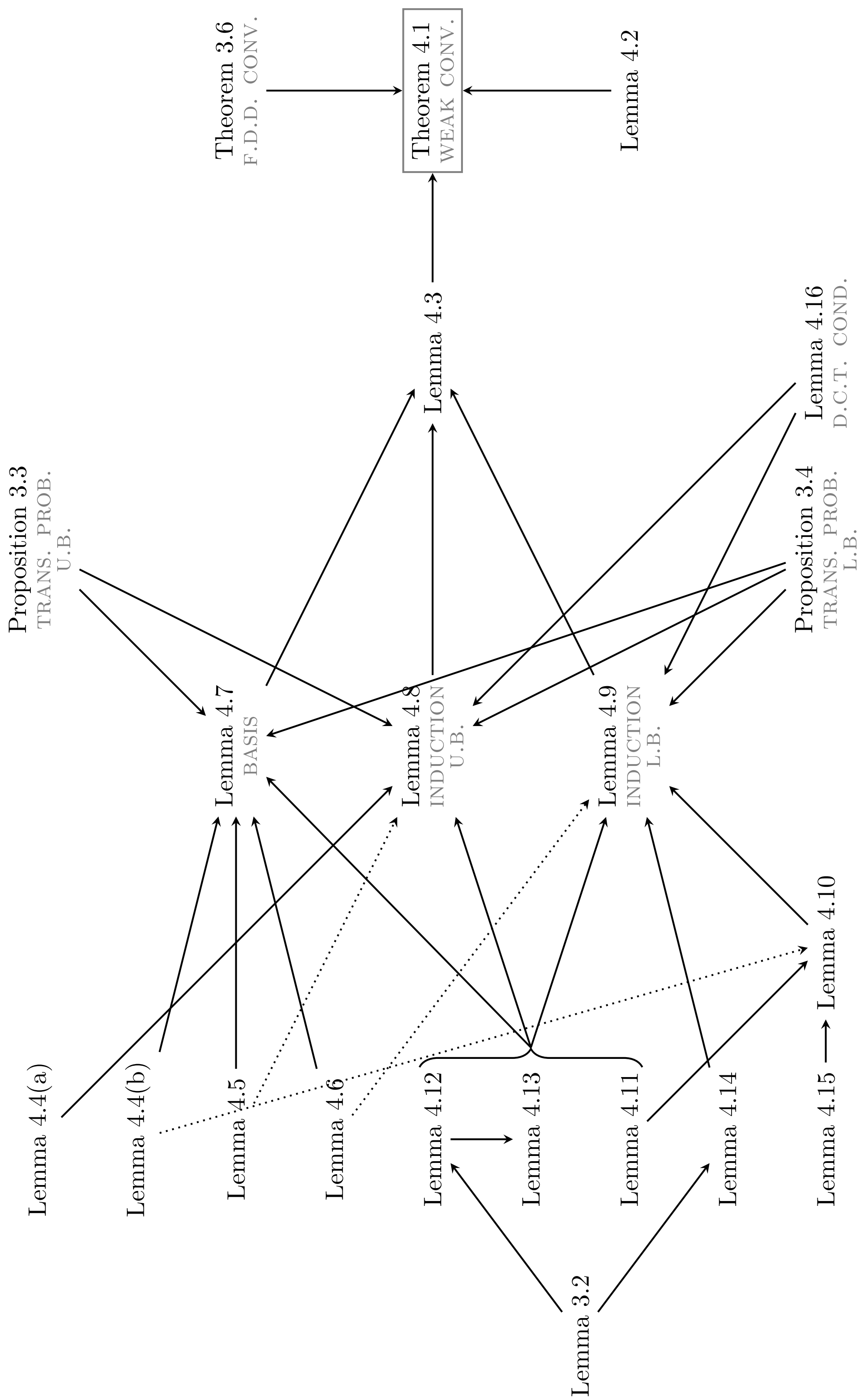


Figure 4.1: Graph showing dependencies between the lemmata used to prove weak convergence. Dotted arrows indicate dependence via a slight modification of the preceding lemma. Dependencies preceding Theorem 3.6 are not shown; these are shown in Figure ?? draw a corresponding figure for fdd proof, or delete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

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

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

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

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

$$\sum_{I=0}^{l-1}\binom{l}{I}B^{l-I}(t+1)^{l-1}\sum_{s=1}^{\tau_N(t)}D_N(s)\leq\left(\sum_{s=1}^{\tau_N(t)}D_N(s)\right)(t+1)^{l-1}(1+B)^{l}, \tag{4.10}$$ 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

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

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):

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

Collecting some terms,

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

eq:021

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,

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

eq:022

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

It remains to show the corresponding lower bound

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

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:

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

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

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

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* $0 = t_0 \le t_1 \le \cdots \le t_k \le t$,

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

Proof. 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:

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

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):

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

For the last line of (4.27),

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

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

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

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,

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

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

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

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

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

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

is bounded above by

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

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

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

Upon taking the expectation and limit, we have

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

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

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

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,

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

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,

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

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

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

(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

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

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

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

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

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

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

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

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,

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

Now, taking the product on the outside,

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

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:

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

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

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

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,

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

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

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

Define the events

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

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

⟨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.

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

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

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

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

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:

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

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.

$$\begin{aligned}\mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \{c_N(r) \geq K D_N(r)\}\right] &\geq \mathbb{P}\left[\bigcap_{r=1}^{\tau_N(t)} \{c_N(r) > K D_N(r)\}\right] \\ &= 1 - \mathbb{P}\left[\bigcup_{r=1}^{\tau_N(t)} \{c_N(r) \leq K D_N(r)\}\right] \\ &= 1 - \mathbb{E}\left[\mathbb{1}_{\bigcup\{c_N(r)\leq K D_N(r)\}}\right] \\ &\geq 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{1}_{\{c_N(r)\leq K D_N(r)\}}\right] \\ &= 1 - \mathbb{E}\left[\sum_{r=1}^{\tau_N(t)} \mathbb{P}[c_N(r) \leq K D_N(r) \mid \mathcal{F}_{r-1}]\right] \qquad (4.39)\ \text{eq:050}\end{aligned}$$

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

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

For the first term,

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

For the second term,

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

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),

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

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

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

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.

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

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.

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

■

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.

For many of the following results it will be necessary to compute filtered expectations $\mathbb{E}_t[\cdot]$, which are generally difficult to compute directly. To simplify the computations we introduce a sequence of σ-algebras $(\mathcal{H}_t)$, defined below, such that filtered expectations can be written in terms of conditional expectations given $\mathcal{H}_t$.

Figure 5.1 shows a section of the conditional dependence graph implied by Algorithm 1, as in Figure 2.2, except that time is now labelled in reverse. The σ-algebra

$$\mathcal{H}_t := \sigma(X_{t-1}^{(1:N)}, X_t^{(1:N)}, w_{t-1}^{(1:N)}, w_t^{(1:N)}), \tag{5.1}$$ eq:defn_Ht

at each time t, forms a separatrix (in the sense of d-separation; see Verma and Pearl (1988)) between the parental indices $a_t^{(1:N)}$ and the previous σ-algebra $\mathcal{F}_{t-1}$ in the filtration. That is, $a_t^{(1:N)}$ is conditionally independent of $\mathcal{F}_{t-1}$ given $\mathcal{H}_t$. The practical upshot of this is that we can use the tower rule along with conditional independence to write filtered expectations as

$$\mathbb{E}_t[f(\nu_t^{(1:N)})] = \mathbb{E}_t\left[\mathbb{E}[f(\nu_t^{(1:N)}) \mid \mathcal{H}_t, \mathcal{F}_{t-1}]\right] = \mathbb{E}_t\left[\mathbb{E}[f(\nu_t^{(1:N)}) \mid \mathcal{H}_t]\right]. \tag{5.2}$$ eq:condexp_

As we will see, this enables us to compute bounds on the filtered expectations of interest relatively easily.

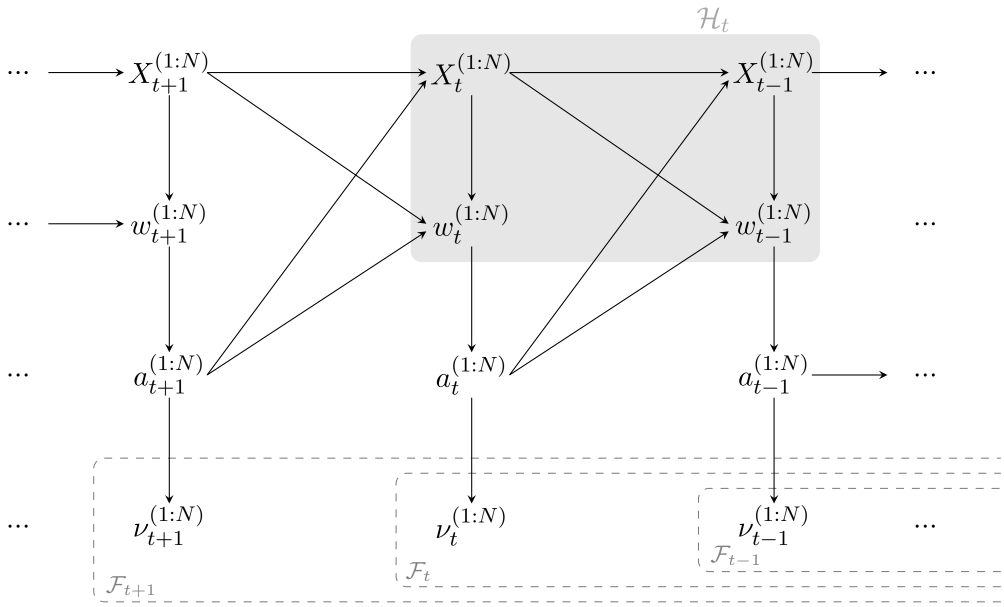


Figure 5.1: Part of the conditional dependence graph implied by Algorithm 1 illustrating the construction of $\mathcal{H}_t$. The direction of time is from left to right. The reverse-time filtration is indicated by the dashed areas. The nodes highlighted in grey generate the separatrix $\mathcal{H}_t$ between $a_t^{(1:N)}$ and $\mathcal{F}_{t-1}$.

⟨fig:cond_indep_graph_Ht⟩

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,*

$$\frac{1}{a} \leq g_t(x,x') \leq a, \quad \varepsilon h(x') \leq q_t(x,x') \leq \frac{1}{\varepsilon} h(x'). \tag{5.3}$$

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 (5.3) 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 (5.3) 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. Define $\mathcal{H}_t$ as in (5.1). Conditional on $\mathcal{H}_t$ the parental indices are independent, with conditional law

$$\mathbb{P}\left[a_t^{(i)} = a_i \mid \mathcal{H}_t\right] \propto g_t(X_{t+1}^{a_{t+1}^{(a_i)}}, X_t^{(a_i)}) q_{t-1}(X_t^{(a_i)}, X_{t-1}^{(i)}) \tag{5.4}$$

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 (5.3) 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

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

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,

$$\frac{1}{(N)_2}\sum_{i=1}^{N}\mathbb{E}[(\nu_t^{(i)})_2 \mid \mathcal{H}_t] \geq \frac{\varepsilon^4}{Na^4} \tag{5.5}$$ eq:mn_cN_LB

and

$$\frac{1}{(N)_3}\sum_{i=1}^{N}\mathbb{E}[(\nu_t^{(i)})_3 \mid \mathcal{H}_t] \leq \frac{a^6}{N^2\varepsilon^6}. \tag{5.6}$$ eq:mn_cN3_U

Applying (5.2) to (5.5) and (5.6) 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 (5.3). 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 (5.5). 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 (5.5) along with (5.2), 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 (5.1). With stratified resampling, conditional on the weights each offspring count almost surely takes one of four values: $\nu_t^{(i)} \in \{\lfloor N w_t^{(i)} \rfloor - 1, \lfloor N w_t^{(i)} \rfloor, \lfloor N w_t^{(i)} \rfloor + 1, \lfloor N w_t^{(i)} \rfloor + 2\}$. Define for each $k \in \mathbb{Z}$

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

eq:pk_defn

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

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

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 it 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_minma

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 (5.3). 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:cNnonzer

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,

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

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):

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

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

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

eq:strat_pb

for each k, which also imply

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

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

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

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*

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

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

$$\mathbb{P}[\max_i w_t^{(i)} - \min_i w_t^{(i)} \ge 2\delta/N \mid \mathcal{F}_{t-1}] \ge \zeta$$

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),

$$p_0^{(i)} \ge (1 - N w_t^{(i)} + \lfloor N w_t^{(i)} \rfloor) \varepsilon^{2N},$$
$$p_1^{(i)} \ge (N w_t^{(i)} - \lfloor N w_t^{(i)} \rfloor) \varepsilon^{2N}. \quad (5.14)$$ eq:SR_pk_LB

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

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

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

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

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 $Nw_t^{(i^\star)} - \lfloor Nw_t^{(i^\star)} \rfloor = 1 - \gamma$, 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 $\nu_t^{(i)} \in \{\lfloor Nw_t^{(i)} \rfloor, \lfloor Nw_t^{(i)} \rfloor + 1, \lfloor Nw_t^{(i)} \rfloor + 2\}$.

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 $i^\star := \mathrm{argmax}_i\{w_t^{(i)}\}$ 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 5.2). Whatever time scale is used, we cannot get away from the fact that this

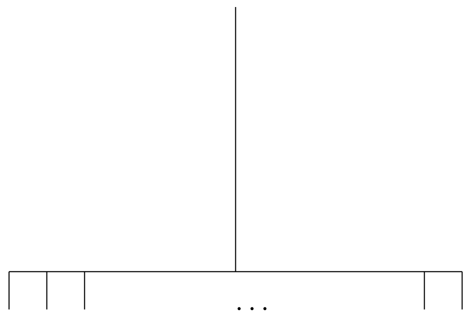


Figure 5.2: Sample genealogy induced by star resampling

⟨fig:star_genealogy⟩

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

The limiting genealogy is more like a *star coalescent* (Griffiths and Mano 2016; Pitman 1999). 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 instantaneously, 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_c

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 (5.1). 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

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

where now

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

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

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

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

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

We therefore have

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

eq:csmc_cN_

and

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

Hence, applying (5.2), we can upper bound the ratio

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

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

⟨sec:ancsamp_genealogy⟩? 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 (2.20), now with time reversed, we obtain add one more step of working below to make it less "magic"?

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

But when i is the index of the immortal particle, $X_{t-1}^{(i)} = x_{t-1}^{\star}$, 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 (5.4), 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 2.10 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

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

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

Berestycki, Nathanaël (2009). *Recent Progress in Coalescent Theory.* 0909.3985. ArXiv.

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

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

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

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

Chopin, Nicolas and Omiros Papaspiliopoulos (2020). *An Introduction to Sequential Monte Carlo.* Springer.

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

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

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

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

– (2013). *Mean Field Simulation for Monte Carlo Integration.* Chapman and Hall/CRC.

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

Doucet, Arnaud and Adam M. Johansen (2011). "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later". In: *Handbook of Nonlinear Filtering*. OUP, pp. 656–704.

Durrett, Richard (2008). *Probability Models for DNA Sequence Evolution*. Springer Science & Business Media.

– (2019). *Probability: Theory and Examples*. 5th ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. DOI: 10.1017/9781108591034.

Efron, Bradley and Robert J. Tibshirani (1994). *An Introduction to the Bootstrap*. CRC press.

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

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

Evensen, Geir (1994). "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, pp. 10143–10162.

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

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

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

Forbes, Catherine et al. (2011). *Statistical Distributions*. John Wiley & Sons.

Fu, Yun-Xin (2006). "Exact Coalescent for the Wright–Fisher Model". In: *Theoretical Population Biology* 69 (4), pp. 385–394.

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

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

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

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

Hol, Jeroen D. (2004). "Resampling in Particle Filters". LiTH-ISY-EX-ET-0283-2004. Internship report. Linköping University.

Hol, Jeroen D., Thomas B. Schön, and Fredrik Gustafsson (2006). "On Resampling Algorithms for Particle Filters". In: *Nonlinear Statistical Signal Processing Workshop*. IEEE, pp. 79–82.

Huang, Chaofan, Vengazhiyil Roshan Joseph, and Simon Mak (2020). *Population Quasi-Monte Carlo*. Tech. rep. 2012.13769. ArXiv.

Jacob, Pierre E., Lawrence M. Murray, and Sylvain Rubenthaler (2015). "Path Storage in the Particle Filter". In: *Statistics and Computing* 25.2, pp. 487–496.

Jazwinski, Andrew H. (2007). *Stochastic Processes and Filtering Theory*. Courier Corporation.

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

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

Kingman, John F. C. (1982a). "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.

– (1982b). "On the Genealogy of Large Populations". In: *Journal of Applied Probability* 19.A, pp. 27–43.

– (1982c). "The Coalescent". In: *Stochastic Processes and Their Applications* 13.3, pp. 235–248.

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

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

Kuipers, Lauwerens and Harald Niederreiter (1974). *Uniform Distribution of Sequences*. John Wiley & Sons.

Lee, Anthony, Lawrence M. Murray, and Adam M. Johansen (2019). "Resampling in Conditional SMC Algorithms". Unpublished.

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

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

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

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

Möhle, Martin and Serik Sagitov (2003). "Coalescent Patterns in Exchangeable Diploid Population Models". In: *Journal of Mathematical Biology* 47, pp. 337–352.

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

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

Myers, Aaron et al. (2021). "Sequential Ensemble Transform for Bayesian Inverse Problems". In: *Journal of Computational Physics* 427, p. 110055.

Notohara, Morihiro (1990). "The Coalescent and the Genealogical Process in Geographically Structured Population". In: *Journal of Mathematical Biology* 29.1, pp. 59–75.

Pitman, Jim (1999). "Coalescents with Multiple Collisions". In: *Annals of Probability*, pp. 1870–1902.

Pitt, Michael K. and Neil Shephard (1999). "Filtering via Simulation: Auxiliary Particle Filters". In: *Journal of the American Statistical Association* 94.446, pp. 590–599.

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

Reich, Sebastian (2013). "A Nonparametric Ensemble Transform Method for Bayesian Inference". In: *SIAM Journal on Scientific Computing* 35.4, A2013–A2024.

Saunders, Ian W., Simon Tavaré, and G. A. Watterson (1984). "On the Genealogy of Nested Subsamples from a Haploid Population". In: *Advances in Applied Probability*, pp. 471–491.

Spouge, John L. (2014). "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, pp. 51–54.

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

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

Wakeley, John (2009). *Coalescent Theory: An Introduction*. Roberts & Co. Publishers.

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

Whiteley, Nick (2010). "Discussion on 'Particle Markov Chain Monte Carlo Methods'". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3, pp. 306–307.

Whitley, Darrell (1994). "A Genetic Algorithm Tutorial". In: *Statistics and Computing* 4.2, pp. 65–85.

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