Asymmetric Sequential Monte Carlo

Qiming Du

Abstract

Sequential Monte Carlo is a general framework aiming at sampling a sequence of measures $(\eta_n; n \geq 0)$ connected by some nonlinear operators. In the classical setting, the simulation consists in a multinomial resampling selection step and a Markov mutation step at each iteration of the algorithm. When the potential functions are $[0, 1]$-valued, a well-known variant is to conduct a Bernoulli *survival test* before the multinomial resampling step: the surviving particles will not be resampled, whilst the non-survived particles perform a multinomial resampling. We go one step further, that is, we suppose that the surviving particles and non-surviving particles will mutate according to different Markov kernels. We refer this situation as "asymmetric resampling". The idea is natural in rare-event simulation and particle tempering problems, where the Markov kernel at step n is η_n-invariant. In this scenario, the surviving particles do not perform a Markov transition while the non-surviving ones do. We provide a CLT-type result as well as consistent variance estimators, which allows to conduct statistical inference with a single run of the simulation. We also give some analysis on the behavior of non-asymptotic variance. In particular, we provide an unbiased variance estimator for the unnormalized measures under certain conditions. To do this, we introduce generalized coalescent tree-based measures and their particle approximations as a complement of the ones introduced respectively in [CDMG11] and [DG19](Chapter 2). We firmly believe that they represent an important and natural family of mathematical objects in the general framework of SMC. They are connected respectively to the asymptotic variance and non-asymptotic variance by some nontrivial combinatorial properties that apply to all the one-parent interacting particle systems. We expect the same methodology may also inspire further analysis for the models in a continuous-time setting, such as Fleming-Viot particle systems (see, e.g. [DCGR17]), and more general resampling schemes.

Contents

1 Introduction

Sequential Monte Carlo (SMC) methods are powerful numeric algorithms widely used in many fields in computational statistics, such as Bayesian inference, filtering, rare-events simulations, etc. The reader is referred to [DdFG01] for a larger list of available applications. The basic idea is to simulate an Interacting Particle System (IPS) in order the approximate a sequence of probability measures $(\eta_n)_{n\geq 0}$ or positive finite measures $(\gamma_n)_{n\geq 0}$ connected by some non-linear operators. The estimators are naturally designed as the associated empirical measures at each level of the IPS. The proper mathematical foundation and more theoretical aspects such as convergence results and bias analysis can be found for example in the pair of books [DM04, DM13] and references therein.

Classical SMC methods consist in a multinomial selection step and a Markov mutation step at each step of the algorithm. This resampling strategy is well-understood both in theory and in practice. It corresponds to a natural interpretation of the Boltzmann-Gibbs transformation w.r.t. the potential functions $(G_n; n \geq 0)$ on the empirical measures $(\eta_n^N; n \geq 0)$. There are a lot of variants on this resampling strategy, such as residual resampling, stratified resampling and systematic resampling, etc. The reader is referred to [HSG06] for a quick survey. Some theoretical analysis can be found in the recent work [GCW17], emphasizing on the most important variants of the resampling schemes mentioned above.

In contrast with the standard setting, we study the resampling strategy that uses two different Markov kernels, denoted respectively by $\mathring{M}_n$ and $\dot{M}_n$, at each iteration of the algorithm. In the rest of this article, they are referred to as *mutation kernels*. Roughly speaking, at each step, say, from level $n-1$ to level n, each particle performs a Bernoulli *survival test* w.r.t. the $[0,1]$-valued potential function G_{n-1}: the survived particle mutates according to $\dot{M}_n$ while the non-surviving ones executes a multinomial resampling, also w.r.t. the potential function G_{n-1}, after which a mutation according to the kernel $\mathring{M}_n$ will be executed. The precise mathematical definition will be given in Section 2.4.

The main motivation is from the generalized Adaptive Multilevel Splitting (gAMS) methods introduced in [BGG⁺16], where the kernel $\dot{M}_n$ is designed to be the identity $\delta_x(dy)$. This idea is natural in the applications such as Particle Tempering and Subset Simulation, where $\mathring{M}_n$ is designed to be an η_n-invariant kernel. The original consideration of this particular resampling scheme is to reduce the unnecessary computational costs brought by the mutation kernel $\mathring{M}_n$. Since the invention of Particle Markov Chain Monte Carlo methods (PMCMC, cf. [ADH10]), the design of mutation kernels $\mathring{M}_n$ becomes much easier and more computationally demanding at the same time. One typical example is SMC^2 methods (cf. [CJP13, CRGP15]). The basic idea is to use another SMC-based IPS and freezing techniques to construct an η_n-invariant kernel at each level. One can imagine that the computational costs are mainly from the implementation of the mutation kernels $(\mathring{M}_n; n \geq 1)$, which can be dramatically reduced by the asymmetric resampling scheme in this article if we choose $\dot{M}_n$ to be the identity or some other "cheap" kernel. The same situation can also be found in Adaptive Multilevel Splitting methods (AMS) in rare-event simulation problems (see, e.g., [BGG⁺16] and [CDGR18]), when the mutation kernels are proposed on the path space.

Another motivation is the symmetric sampling, namely, the case where all the particles mutate according to the same Markov kernel at each step. More precisely, it means that $\dot{M}_n \equiv \mathring{M}_n$ for each $n \geq 1$ and this setting enters the classic Feynman-Kac particle models intensely studied in [DM04]. It is well-known that the asymptotic variance is smaller than the one under classical multinomial resampling scheme. The exact difference at each step is presented in (20). As the symmetric resampling can also slightly reduce the computational costs required by the multinomial resampling, there is no practical reason to implement multinomial resampling scheme if an upper bound for G_n is available. In fact, more advanced resampling schemes can still be considered to reduce the variance, since there is still a multinomial resampling step for the non-survival particles in this setting. Due to additional technical complications, they are left for future investigations.

In order to conduct statistical inference, it is important to study the asymptotic behaviors of the empirical measures associated to the IPS (see, e.g., [DM04, Cho04, DM08]). In this respect, if one has a CLT-type theorem for some test function f such as

$$\sqrt{N}\left(\eta_n^N(f) - \eta_n(f)\right) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}(0, \sigma_n(f)^2),$$

it is sufficient to provide a consistent estimator $\sigma_n^N(f)$ of $\sigma_n(f)$ since Slutsky's lemma guarantees that

$$\frac{\sqrt{N}\left(\eta_n^N(f) - \eta_n(f)\right)}{\sigma_n^N(f)} \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}(0, 1).$$

An asymptotic confidential interval can therefore be derived as a by-product of the simulation of IPS. Although SMC methods are intensely studied for over 20 years, the classical way to achieve this is still by resimulating the IPS independently many times and by estimating $\sigma_n(f)^2$ with the crude variance estimator. This is not always practical: a single run of an IPS may take a lot of time, and one also expects that all the computational power is used to improve precision, rather than to estimate the variance. In addition, as the estimator $\eta_n^N(f)$ of $\eta_n(f)$ provided by SMC methods is typically biased, it is also nontrivial to implement parallel computing for a large number of independent IPS with N relatively small. As a consequence, a variance estimator available with a single run of the simulation is of crucial interest for applications.

The breakthrough is due to Chan and Lai in [CL13]. By using the ancestral information encoded in the genealogy of the associated IPS, the first consistent variance estimators are proposed. Then, Lee and Whiteley [LW18] provided an unbiased variance estimator for the unnormalized measures γ_n^N and a term by term estimator, which helps to better understand the role of the genealogy in variance related problems. Then, a more numerically stable variance estimator is provided in [OD19], as a natural fixed-lag version of the original one proposed in [CL13], when more stability properties of the IPS are available. Another recent result is given in [DG19](Chapter 2), by extending the estimator of Lee & Whiteley to the adaptive SMC context (cf. [BJKT16]). All these estimators are studied in the classical SMC framework, meaning under multinomial resampling scheme.

From a theoretical viewpoint, the current setting can be regarded as a "playground" for more sophisticated algorithms in the adaptive context and/or in a continuous-time setting: there is no additional attention required to deal with complicated regularity assumptions, and we can thus focus on the structural properties of the IPS. Similar to the case where the variance estimators provided by Lee & Whiteley in [LW18] are still valid in the adaptive SMC framework with some additional assumptions (cf. Assumption 2, [DG19](Chapter 2)), we expect that the variance estimators provided in this article are still valid in more general settings, and our methodology can also be extended in such scenarios. The rigorous mathematical formulation of the current setting can be seen as a generalization of the discrete-time Feynman-Kac particle models presented in the literature such as [DM04] and [DM13]. Our technical tools consist in a new family of mathematical objects, i.e., the so-called coalescent Feynman-Kac measures and coalescent tree occupation measures. They are introduced in order to apply the same methodology as in the previous work [DG19](Chapter 2), which can potentially be a universal strategy to conduct variance estimation in the one-parent IPS context. We hope these theoretical tools may help the analysis of more complex and advanced models in the IPS context.

1.1 Main results

On one hand, in a very general setting, we provide consistent estimations for the target measures in SMC context with controllable asymptotic uncertainty under our specific asymmetric resampling scheme. Since the computations of the variance estimators are highly nontrivial, we provide detailed and efficient algorithms with time and space complexity analysis in Section A. If there is any ambiguity w.r.t. the notation, the reader is referred to Section 1.2.

Theorem 1.1. *Let $(E_n; n \leq 0)$ be a sequence of Polish state spaces. Given a sequence of $[0,1]$-valued potential functions $(G_n; n \geq 0)$ and a canonical Markov chain $(X_n; n \geq 0)$ taking values in $(E_n; n \geq 0)$, with initial distribution η_0 and transition kernels $(\mathring{M}_n; n \geq 1)$, we define the family of measures $(\gamma_n; n \geq 0)$ by*

$$\gamma_n(f) := \mathbf{E}\left[f(X_n) \prod_{p=0}^{n-1} G_p(X_p) \right].$$

Assuming that $\gamma_n(1) > 0$ for any $n \geq 0$, we also define $\eta_n(f) := \gamma_n(f)/\gamma_n(1)$. For any test function $f \in \mathcal{B}_b(E_n)$, when the number of particle N tends to infinity, the estimators given by Algorithm 2 in Section A, denoted respectively by $\gamma_n^N(f)$ and $\eta_n^N(f)$, converge almost surely to $\gamma_n(f)$ and $\eta_n(f)$ if for any $n \geq 1$, we have

$$\forall \varphi_n \in \mathcal{B}_b(E_n), \quad \gamma_{n-1}(G_{n-1} \times \mathring{M}_n(\varphi_n)) = \gamma_{n-1}(G_{n-1} \times \dot{M}_n(\varphi_n)).$$

Moreover, we also have

$$\frac{\sqrt{N}\left(\gamma_n^N(f) - \gamma_n(f)\right)}{\hat{\sigma}_{\gamma_n^N}(f)} \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}(0,1),$$

as well as

$$\frac{\sqrt{N}\left(\eta_n^N(f) - \eta_n(f)\right)}{\hat{\sigma}_{\eta_n^N}(f - \eta_n^N(f))} \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}(0,1),$$

where the computation of $\hat{\sigma}_{\gamma_n^N}(f)$ and $\hat{\sigma}_{\eta_n^N}(f - \eta_n^N(f))$ are respectively provided in Algorithm 5 and Algorithm 6 in Section A.

On the other hand, under mild assumptions, we provide an unbiased non-asymptotic variance estimator of $\gamma_n^N(f)$, which, again, represents the output of Algorithm 2.

Theorem 1.2. *Assume the same setting as in Theorem 1.1. Under the condition discussed in Section 4.1, which at least contains the case where $\dot{M}_n \equiv \mathring{M}_n$ for any $n \geq 1$, the estimator $\gamma_n^N(f)$ is an unbiased estimator for $\gamma_n(f)$. Moreover, the estimator provided by Algorithm 7 in Section A is an unbiased estimator for the non-asymptotic variance of $\gamma_n^N(f)$.*

1.2 Notation

Before getting into details, let us provide a few notations which are useful in the following.

- The underlying probability space is denoted by $(\Omega, \mathcal{F}, \mathbf{P})$. For σ-fields $\mathcal{E}, \mathcal{G} \subset \mathcal{F}$, $\mathcal{E} \vee \mathcal{G}$ denotes the smallest σ-field on Ω containing $\mathcal{E}$ and $\mathcal{G}$. For any $x, y \in \mathbf{R}$, we denote $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$. We also adopt the standard convention $\inf \emptyset = \infty$.
- Let X be a number, a function or a random variable. We adopt the following convention:
$$\frac{1}{\mathrm{X}}\mathbf{1}_{\mathrm{X}\neq 0} := \begin{cases} \frac{1}{\mathrm{X}} & \text{if} \quad \mathrm{X} \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$
Therefore, we admit the calculation
$$\mathrm{X} \times \frac{1}{\mathrm{X}}\mathbf{1}_{\mathrm{X}\neq 0} = \mathbf{1}_{\mathrm{X}\neq 0}.$$
- Random variables take values in Polish spaces, i.e., a topological space E which is metrizable, separable and complete for some distance d_E. It is endowed with the Borel σ-algebra generated by d_E, denoted by $\mathcal{B}(E)$.
- We denote respectively by $\mathcal{M}(E)$, $\mathcal{M}_+(E)$ and $\mathcal{P}(E)$ the set of all signed finite measures, the subset of all nonnegative finite measures and the subset of all probability measures on $(E, \mathcal{B}(E))$. The set $\mathcal{P}(E)$ is endowed with the Prohorov-Lévy metric, i.e., the weak convergence "$\xrightarrow{\mathrm{d}}$" is tested with continuous bounded functions.
- $\mathcal{B}_b(E)$ denotes the collection of all the bounded measurable functions from $(E, \mathcal{B}(E))$ to $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ equipped with uniform norm $\|\cdot\|_\infty$, among which the constant function will be denoted by 1 with a slight abuse of notation. Given a probability measure η in $\mathcal{P}(E)$ and for all test functions in $\mathcal{B}_b(E)$, we denote η-ess sup(f) the essential supremum of f. It is defined by
$$\eta\text{-ess sup}(f) := \inf \{a \in \mathbf{R} : \eta(x \in E : f(x) > a) = 0\}.$$
- For all $\mu \in \mathcal{M}(E)$ and for all test functions $f \in \mathcal{B}_b(E)$, $\mu(f)$ denotes the integration
$$\int_E f(x)\mu(dx).$$
A finite transition kernel Q from $(E, \mathcal{B}(E))$ to $(F, \mathcal{B}(F))$ is a function
$$Q : E \times \mathcal{B}(F) \mapsto \mathbf{R}_+.$$
More precisely, for all $x \in E$, $Q(x, \cdot)$ is a finite nonnegative measure in $\mathcal{M}_+(F)$ and for all $A \in \mathcal{B}(F)$, $x \mapsto Q(x, A)$ is a $\mathcal{B}(E)$-measurable function. We say that Q is a Markov transition kernel if Q is a finite transition kernel and for all $x \in E$, $Q(x, \cdot)$ is a probability measure in $\mathcal{P}(F)$. For a signed measure $\mu \in \mathcal{M}(E)$ and a test function $f \in \mathcal{B}_b(F)$, we denote respectively $\mu Q \in \mathcal{M}(E)$ and $Qf \in \mathcal{B}_b(E)$ are respectively defined as follows:
$$\forall A \in \mathcal{B}(F), \quad \mu Q(A) := \int_E \mu(dx)Q(x, A),$$

and

$$\forall x \in E, \quad Qf(x) := \int_F Q(x, dy) f(y).$$

Let Q_1 and Q_2 be two finite transition kernels respectively from E_0 to E_1 and from E_1 to E_2. When well-defined, we denote $Q_1 \cdot Q_2$ or simply Q_1Q_2, the transition kernel from E_0 to E_2 defined by

$$\forall (x, A) \in E_0 \times \mathcal{B}(E_2), \quad Q_1Q_2(x, A) := \int_{E_1} Q_1(x, dy) Q_2(y, A).$$

Note that, there is no reason that Q_1Q_2 is still a finite transition kernel in general. We say that Q_1 is a *uniformly* finite transition kernel from space E_0 to E_1 if

$$\sup_{x \in E_0} \int Q_1(x, dy) < +\infty.$$

For example, a Markov transition kernel is a uniformly finite transition kernel. Let Q_2 be a uniformly finite transition kernel from E_1 to E_2, we have that Q_1Q_2 is also a uniformly finite transition kernel from E_0 to E_2.

- For two test functions $f, g \in \mathcal{B}_b(E)$, we denote

$$f \otimes g : E^2 \ni (x, y) \mapsto f(x)g(y) \in \mathbf{R}.$$

In particular, we denote

$$f^{\otimes 2} := f \otimes f.$$

Accordingly, we denote

$$\mathcal{B}_b(E)^{\otimes 2} := \{ f \otimes g : f, g \in \mathcal{B}_b(E) \}.$$

For two finite transition kernels Q and H from $(E, \mathcal{B}(E))$ to $(F, \mathcal{B}(F))$, we denote, for all $(x, y) \in E \times E$ and for all $(A, B) \in \mathcal{B}(F) \otimes \mathcal{B}(F)$,

$$Q \otimes H\left((x, y), (A, B)\right) := Q(x, A) \times H(y, B).$$

Similarly, we also denote

$$Q^{\otimes 2} := Q \otimes Q.$$

- In order to define the coalescent tree-based measures of size 2, we introduce the transition operators C_0 and C_1 as

$$C_0((x, y), (dx', dy')) := \delta_{(x,y)}(dx', dy'),$$

and

$$C_1((x, y), d(x', y')) := \delta_{(x,x)}(dx', dy').$$

In other words, for any measurable function $H : E \times E \mapsto \mathbf{R}$, we have

$$C_0(H)(x, y) = H(x, y) \qquad \text{and} \qquad C_1(H)(x, y) = H(x, x).$$

- For all $\boldsymbol{x} = (x^1, \ldots, x^N) \in E^N$, we define the empirical measure associated to $\boldsymbol{x}$ by

$$m : \boldsymbol{x} \mapsto m(\boldsymbol{x}) := \frac{1}{N} \sum_{i=1}^{N} \delta_{x^i} \in \mathcal{P}(E).$$

We denote

$$m^{\otimes 2} : \boldsymbol{x} \mapsto m^{\otimes 2}(\boldsymbol{x}) := \frac{1}{N^2} \sum_{i,j} \delta_{(x^i, x^j)} \in \mathcal{P}(E),$$

and

$$m^{\odot 2} : \boldsymbol{x} \mapsto m^{\odot 2}(\boldsymbol{x}) := \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(x^i, x^j)} \in \mathcal{P}(E).$$

A straightforward computation shows that

$$m^{\otimes 2}(\boldsymbol{x}) = \frac{N-1}{N} m^{\odot 2}(\boldsymbol{x}) C_0 + \frac{1}{N} m^{\odot 2}(\boldsymbol{x}) C_1. \tag{2}$$

With a slight abuse of notation, considering $[N] := \{1, 2, \ldots, N\}$, we write

$$m([N]) := \frac{1}{N} \sum_{i=1}^{N} \delta_i \quad \text{and} \quad m^{\otimes 2}([N]) := m([N]) \otimes m([N]),$$

as well as

$$m^{\odot 2}([N]) := \frac{1}{N(N-1)} \sum_{i \neq j} \delta_{(i,j)}.$$

2 SMC framework

In this section, we define the SMC framework studied in this article. We present some standard convergence results on the consistency and asymptotic normality of the associated Interactive Particle System (IPS) in the discrete time setting. We mainly use the language of Feynman-Kac particle models, and the reader is referred to the pair of books [DM04] and [DM13] for more details. The main goal is to establish central limit theorems and to specify the asymptotic variance in our specific asymmetric setting.

2.1 Setting

Let $(E_n, \mathcal{B}(E_n))_{n \geq 0}$ be a sequence of Polish spaces and let us fix a probability measure $\eta_0 \in \mathcal{P}(E_0)$. We consider a sequence of $[0,1]$-valued measurable potential functions $(G_n)_{n \geq 0}$ and a sequence of Markov transition kernels $(\mathring{M}_n)_{n \geq 1}$ s.t. $\mathring{M}_n : (E_{n-1}, \mathcal{B}(E_n)) \mapsto [0,1]$. We define the Feynman-Kac kernels as follows

$$\forall (x, A) \in (E_{n-1}, \mathcal{B}(E_n)), \quad \mathring{Q}_n(x, A) := G_{n-1}(x) \mathring{M}_n(x, A).$$

It is readily checked that Q_n is a uniformly finite transition kernel. Therefore, we define the unnormalized Feynman-Kac measure γ_n by

$$\forall n \geq 1, \quad \gamma_n := \eta_0 \mathring{Q}_1 \mathring{Q}_2 \cdots \mathring{Q}_n,$$

with $\gamma_0 := \eta_0$. By definition, for all $n \geq 1$, γ_n is a sub-probability measure. For all $n \geq 0$, we suppose that we have a meaningful sampling problem at each step, i.e., we assume that $\gamma_n(1) > 0$. Therefore, one can define the normalized Feynman-Kac measures

$$\forall n \geq 1, \quad \eta_n := \frac{\gamma_n}{\gamma_n(1)}.$$

We adopt the convention

$$\eta_{-1} = \gamma_{-1} = \eta_0.$$

By standard convention for the product symbol "$\prod$", it is readily checked that

$$\forall n \geq 0, \quad \gamma_n = \left\{ \prod_{p=0}^{n-1} \eta_p(G_p) \right\} \eta_n. \tag{3}$$

Different from the classical framework of SMC methods, we suppose that there exists an additional sequence of Markov transition kernels $(\dot{M}_n)_{n\geq 0}$, such that for the Feynman-Kac kernel defined by

$$\forall (x, A) \in E_{n-1} \times \mathcal{B}(E_n), \quad \dot{Q}_n(x, A) := G_{n-1}(x) \times \dot{M}_n(x, A),$$

we have, for all $n \geq 0$,

$$\gamma_n \dot{Q}_{n+1} = \gamma_n \mathring{Q}_{n+1} = \gamma_{n+1}.$$

Using the Feynman-Kac kernels $\mathring{Q}_n$ and $\dot{Q}_n$, we define the Feynman-Kac kernel Q_n by

$$Q_n := \eta_{n-1}(G_{n-1}) \dot{Q}_n + [1 - \eta_{n-1}(G_{n-1})]\, \mathring{Q}_n.$$

More rigorously, for any $\mu \in \mathcal{M}(E_{n-1})$ and for any $f \in \mathcal{B}_b(E_n)$, we have

$$\mu Q_n(f) := \eta_{n-1}(G_{n-1}) \mu \dot{Q}_n(f) + [1 - \eta_{n-1}(G_{n-1})]\, \mu \mathring{Q}_n(f).$$

Hence, for all $0 \leq p < n < +\infty$, the associated Feynman-Kac partial semigroup is defined as follows:

$$Q_{p,n} := Q_{p+1} \cdots Q_n.$$

The term "partial" comes from the fact that the state spaces E_n may vary w.r.t. the time horizon n. Hence, it is not a semigroup. In particular, the partial unit elements at each step is defined by $Q_{n,n}(x, A) := \delta_x(A)$ on the space E_n.

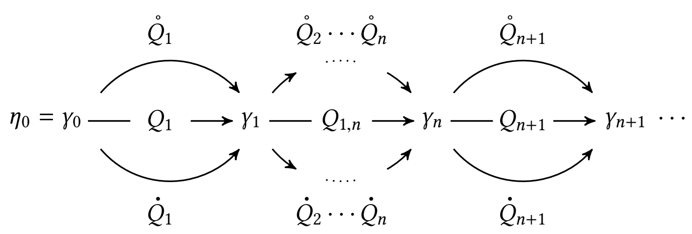

Figure 1: Illustration of the Feynman-Kac measures flow.

Remark. Technically speaking, such $\overset{\bullet}{M}_n$ always exists. For example, one may consider the choice $\overset{\bullet}{M}_n \equiv \overset{\circ}{M}_n$. Even in this simple symmetric setting, the variance related problems are already very challenging. The CLT-type results are well-known (see, e.g., Chapter 7 of [DM04]). However, to the best of our knowledge, there is no consistent asymptotic variance estimators available with a single simulation of the particle system. Meanwhile, it is natural to implement the asymmetric resampling in the case where $\overset{\circ}{M}_n$ is an η_n-invariant kernel: we are interested by the choice $\overset{\bullet}{M}_n(x, dy) := \delta_x(dy)$ since it requires the least computational cost, which is widely-used by the practitioners in tempering and rare-event simulation. Therefore, we combine these two examples and go one step further: we consider the asymmetric resampling scheme and we provide some theoretical analysis. When $\overset{\circ}{M}_n$ is not an η_n-invariant kernel, it is also always possible to construct a nontrivial $\overset{\bullet}{M}_n$ using $\overset{\circ}{M}_n$. In fact, $\overset{\bullet}{M}_n$ can still be thought as some "cheaper" version of the latter: $\overset{\circ}{M}_n$ can be designed as the composition of $\overset{\bullet}{M}_n$ and an η_n-invariant kernel, for which one may consider the PMCMC-type kernel, which is always available with $(G_n; n \geq 0)$ and $(\overset{\circ}{M}_n; n \geq 1)$ under the current setting. Intuitively speaking, this SMC^2-type design can help to reduce the dependence due to the multinomial resampling step.

Definition 2.1. *We introduce the asymmetric McKean kernel $K_{n,\mu}$ from E_{n-1} to E_n, parameterized by some positive finite measure $\mu \in \mathcal{M}_+(E_{n-1})$ such that $\mu(G_{n-1}) > 0$, defined as follows*

$$\forall A \in \mathcal{B}(E_n), \quad K_{n,\mu}(x, A) := G_{n-1}(x)\overset{\bullet}{M}_n(x, A) + (1 - G_{n-1}(x))\frac{\mu(G_{n-1} \times \overset{\circ}{M}_n(A))}{\mu(G_{n-1})}.$$

Accordingly, we also define the McKean-type Feynman-Kac kernel $Q_{n,\mu}$ by

$$Q_{n,\mu} := \mu(G_{n-1})K_{n,\mu}(x, A),$$

with the convention

$$\forall x \in E_0, \quad Q_{0,\mu}(x, A) := \eta_0(A). \tag{4}$$

Remark. Standard calculations show that the McKean-type kernels $K_{n,\eta_{n-1}}$ and $Q_{n,\eta_{n-1}}$ also connect the Feynman-Kac measures flow:

$$\eta_{n-1}K_{n,\eta_{n-1}} = \eta_n \quad \text{and} \quad \gamma_{n-1}Q_{n,\eta_{n-1}} = \gamma_n. \tag{5}$$

Assuming that G_n is upper bounded by 1 rather than a finite positive number $\|G_n\|_\infty$ is purely for technical reasons, in order to simplify the relatively heavy notation. There is no loss of generality for the case where $\|G_n\|_\infty$ is known: we could always consider the "normalized" version of potential function

$$\bar{G}_n := \frac{G_n}{\|G_n\|_\infty},$$

in order to construct a potential function varying on the interval [0,1]. However, when $\|G_n\|_\infty$ is not explicitly tractable, it is not possible to design the asymmetric version of SMC sampler with fixed normalizer. When the normalizer is set to be $+\infty$, we return to the classical multinomial resampling scheme. This is a crucial problem in applications such as tempering, when determining a reasonable upper bound of the potential function is not always trivial. One possible solution is to consider the adaptive normalizer, depending upon the entry measure μ, rather than a prefixed one. An interesting example of the adaptive normalizer is defined by

$$\mu\text{-ess}\sup(G_{n-1}),$$

where μ denotes the entry measure of the McKean kernel. This is the "laziest" resampling scheme we could ever design, which gives potentially the smallest asymptotic variance, and no upper bound of the potential function is required. However, we failed to provide the general analysis for this case since the calculation of the asymptotic variance in the CLT-type results will become more challenging and it is possible that stronger mixing properties for $\mathring{Q}_n$ and $\dot{Q}_n$ have to be assumed. Heuristically speaking, in order to establish the CLT-type results and to conduct the asymptotic variance estimation, one needs a convergence of the following type:

$$\exists \epsilon_n \in \mathbf{R}_+^*, \quad 1/\max_{1\leq i\leq N} G_n(X_n^i) \xrightarrow[N\to\infty]{\mathbf{P}} \epsilon_n.$$

This requires much stronger convergence than the well-known almost sure convergence of the empirical measures. However, if the convergence above holds, we expect that the methodology in this article would still be valid, with only minor notational complications. In a nutshell, one needs to discuss the property above in concrete applications, such as the mixing property of the Markov kernels, etc. Meanwhile, the goal of the present work is to obtain some general structural results without further assumptions. As a consequence, we decide to leave this important case for future research.

2.2 Interacting particle system

The Interacting Particle System (IPS) in this article refers to a Markov chain $(\mathbf{X_n}; n \geq 0)$ with absorption in the product spaces $(E_n^N, \mathcal{B}(E_n)^{\otimes N}; n \geq 0)$. As we have seen in the previous section (5), the normalized Feynman-Kac measures η_n and η_{n+1} are connected by K_{n+1,η_n}, which depends on the measure of the previous step η_n. Hence, it is not possible to simulate directly according to the kernel K_{n+1,η_n}. The idea of the IPS is to simulate N particles $\mathbf{X_n} = (X_n^1, X_n^2, \dots, X_n^N)$ step by step. Therefore, by exploiting the empirical measure $m(\mathbf{X_n})$ to approximate its "limiting" measure η_n, we are able to simulate the next layer of particles $\mathbf{X_{n+1}}$ with the approximated kernel $K_{n+1,m(\mathbf{X_n})}$. In this section, we deal with the version without the genealogy (i.e., the indices of the parent of each particle) and the survival history of IPS. The mechanism of the IPS is defined as follows:

(i) $\mathbf{X_0} \sim \eta_0^{\otimes N}$;

(ii) Stop the algorithm at step $n \geq 0$ if $m(\mathbf{X_n})(G_n) = 0$;

(iii) If not stopped at step $n \geq 0$,

$$\mathbf{X_{n+1}} \sim \bigotimes_{i=1}^{N} K_{n+1,m(\mathbf{X_n})}(X_n^i, \cdot).$$

A more detailed explanation on the algorithm can be found in Section 2.4. The particle approximation of the normalized measure η_n is defined by

$$\eta_n^N := m(\mathbf{X_n}) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X_n^i}.$$

According to (3), the unnormalized version γ_n^N is defined by

$$\gamma_n^N := \left\{ \prod_{p=0}^{n-1} \eta_p^N(G_p) \right\} \eta_n^N .$$

The absorbing time τ_N of the Feynman-Kac IPS is defined by

$$\tau_N := \inf \left\{ n \in \mathbf{N} : m(\mathbf{X_n})(G_n) = 0 \right\} .$$

2.3 Asymptotic results

In this section, we establish some basic convergence results such as law of large numbers and central limit theorem for the empirical Feynman-Kac measures. These results are standard in the case where $\dot{Q}_n \equiv \mathring{Q}_n$ (see, e.g., Chapter 7 of [DM04]) and the proofs are housed respectively in Section C.3 and Section C.4. The goal is to understand the consequences of the introduction of $\dot{Q}_n$, especially on the form of the asymptotic variances.

Theorem 2.1. *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} \xrightarrow[N\to\infty]{a.s.} \gamma_n(f).$$

The almost sure convergence also holds for $\eta_n^N \mathbf{1}_{\tau_N \geq n}$. In particular, by taking the test function 1 for η_n^N, we get

$$\mathbf{1}_{\tau_N \geq n} \xrightarrow[N\to\infty]{a.s.} 1.$$

Moreover, if we assume symmetric resampling, that is $\dot{Q}_n \equiv \mathring{Q}_n$ for any $n \geq 1$, we also have

$$\forall n \geq 0, \quad \mathbf{E}\left[\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}\right] = \gamma_n(f).$$

Theorem 2.2. *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\forall n \geq 0, \quad \sqrt{N}\left(\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\right) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma_{\gamma_n}^2(f)\right),$$

with the asymptotic variance defined as follows:

$$\sigma_{\gamma_n}^2(f) := \sum_{p=0}^{n} \left(\gamma_p^{\otimes 2} C_1 Q_{p,n}^{\otimes 2}(f^{\otimes 2}) - \gamma_{p-1}^{\otimes 2} C_1 Q_{p,\eta_{p-1}}^{\otimes 2} Q_{p,n}^{\otimes 2}(f^{\otimes 2}) \right). \tag{6}$$

Similarly, we also have

$$\forall n \geq 0, \quad \sqrt{N}\left(\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma_{\eta_n}^2(f - \eta_n(f))\right),$$

with $\sigma_{\eta_n}^2$ defined by

$$\forall \varphi \in \mathcal{B}_b(E_n), \quad \sigma_{\eta_n}^2(\varphi) := \sigma_{\gamma_n}^2(\varphi)/\gamma_n(1)^2. \tag{7}$$

Let us emphasize that one of the main goals of this article is to provide consistent estimators w.r.t. the particle numbers N for the asymptotic variances $\sigma_{\gamma_n}^2$ and $\sigma_{\eta_n}^2$ defined above. In practice, thanks to Slutsky's lemma, the consistent variance estimators allow us to deduce confidential intervals with one single simulation of IPS.

2.4 Genealogy and survival history

In this section, we give a more detailed version of the IPS defined in Section 2.2, namely, the actual simulation algorithm we execute in practice. Specifically, we trace two kinds of information: the genealogy $\mathbf{A_n} = (A_n^1, \ldots, A_n^N) \in [N]^N$ and the so-called survival history $\mathbf{B_n} = (B_n^1, \ldots, B_n^N) \in \{0, 1\}^N$. They are both intermediate random variables introduced in the real-world algorithm, so that one can simulate according to an approximated kernel $K_{n+1,m(\mathbf{X_n})}$. Note that

$$A_n^i = j$$

means that the parent of X_{n+1}^i at level n is X_n^j. Besides, $B_n^i = 1$ indicates that the particle X_n^i has survived at step n, i.e., the parent of X_{n+1}^i is X_n^i ($A_n^i = i$) and

$$X_{n+1}^i \sim \dot{M}_n(X_n^i, \cdot).$$

Note that this does not mean that a non-survived particle at step n, i.e., a particle such that $B_n^i = 0$, is disappeared in the IPS after step n: it can still be selected as a parent by the multinomial resampling step. Unlike the multinomial selection scheme, the information encoded in IPS and its genealogy is not enough to conduct the variance estimation. This is the reason why survival history has to be taken into consideration. Another remark is for rare-event simulation, or more generally, the case where $(G_n; n \geq 0)$ are all indicator functions: the survival history is already encoded in $(G(X_n^i), n \geq 0, i \in [N])$. Hence, there is no need to track them separately. Now, let us give the proper definition of the IPS with its genealogy and survival history:

(i) Initial distribution:
At step 0, we let $\mathbf{X_0} \sim \eta_0^{\otimes N}$.

(ii) Stopping criterion:
Stop the algorithm at step $n \geq 0$ if $m(\mathbf{X_n})(G_n) = 0$.

(iii) Transition kernels:
If not stopped at $n \geq 0$, we execute the elementary transition $X_n^i \rightsquigarrow X_{n+1}^i$ for all $1 \leq i \leq N$ conditionally independently, following the three steps:

- Survival test: Let B_n^i be a Bernoulli random variable with probability $G_n(X_n^i)$, that is
$$B_n^i \sim G_n(X_n^i)\delta_1 + (1 - G_n(X_n^i))\delta_0.$$
- Selection: If $B_n^i = 1$, we let $A_n^i = i$. Otherwise, the parent index is selected by the following multinomial selection
$$A_n^i \sim \sum_{k=1}^{N} \frac{G_n(X_n^k)}{\sum_{j=1}^{N} G_n(X_n^j)} \delta_k.$$
Therefore, given $B_n^i = \beta_n^i$, we have
$$A_n^i \sim \beta_n^i \delta_i + (1 - \beta_n^i) \sum_{k=1}^{N} \frac{G_n(X_n^k)}{\sum_{j=1}^{N} G_n(X_n^j)} \delta_k.$$
- Mutation: Given $B_n^i = \beta_n^i$ and $A_n^i = a_n^i$, each particle X_n^i evolves independently from level n to level $n + 1$ according to the following transition kernel:
$$X_{n+1}^i \sim \beta_n^i \dot{M}_{n+1}(X_n^i, \cdot) + (1 - \beta_n^i)\mathring{M}_{n+1}(X_n^{a_n^i}, \cdot).$$

3 Variance estimations

In this section, we provide estimators for the asymptotic variances $\sigma^2_{\gamma_n}(f)$ and $\sigma^2_{\eta_n}(f)$: we provide a term by term asymptotic variance estimator, an unbiased variance estimator under symmetric resampling scheme, and finally, an efficient asymptotic variance estimator. The strategy is almost identical as in [DG19](Chapter 2). First, we give an alternative representation of the asymptotic variance $\sigma^2_{\gamma_n}(f)$ using some generalized coalescent tree-based measures. Next, we provide convergence results of the particle approximations of these generalized coalescent tree-based measures, which gives naturally a term by term variance estimator. Finally, we connect this term by term estimator to the non-asymptotic variance using a nontrivial combinatorial property of the IPS given in Theorem B.1, from which we derive an efficient variance estimator that can be computed with the optimal $\mathcal{O}(nN)$ time complexity.

3.1 Asymptotic variance expansion

In this section, we revisit the asymptotic variance $\sigma^2_{\gamma_n}(f)$ of Theorem 2.2 using some novel coalescent tree-based measures. More precisely, unlike the multinomial case, the form of the asymptotic variance $\sigma^2_{\gamma_n}(f)$ is relatively complex under asymmetric resampling and there is no free coalescent tree-based expansion as in [DG19](Chapter 2). Hence, we need to introduce some generalized coalescent tree-based measures as a supplement of the one introduced in [CDMG11]. The goal is plain and simple: we want to establish an alternative representation of the asymptotic variance based on some coalescent tree-based measures. To begin, let us define the so-called *coalescent Feynman-Kac kernels*:

$$\text{(i)} \begin{cases} \mathbf{Q}_n^{\dagger,0} := Q_n^{\otimes 2}; \\ \mathbf{Q}_n^{\dagger,1} := C_1 Q_n^{\otimes 2} - \eta_{n-1}(G_{n-1})^2 C_1 \dot{Q}_n^{\otimes 2}. \end{cases}$$

$$\text{(ii)} \begin{cases} \widetilde{\mathbf{Q}}_n^{\dagger,0} := Q_n^{\otimes 2}; \\ \widetilde{\mathbf{Q}}_n^{\dagger,1} := \eta_{n-1}(G_{n-1}) \left[(G_{n-1} \times \dot{Q}_n) \otimes \mathring{Q}_n + \mathring{Q}_n \otimes (G_{n-1} \times \dot{Q}_n) \right] \\ \qquad\quad + \eta_{n-1}(G_{n-1}^2) \left[\dot{Q}_n^{\otimes 2} - \dot{Q}_n \otimes \mathring{Q}_n - \mathring{Q}_n \otimes \dot{Q}_n \right]. \end{cases}$$

The full description of this new family of kernels can be found in Section B.2. Using the partial semigroup property of the coalescent Feynman-Kac kernels defined above, we introduce some generalized coalescent tree-based measures. They will be referred to as *coalescent Feynman-Kac measures* in this article. In the next definition, we call $b := (b_0, \ldots, b_n) \in \{0,1\}^{n+1}$ a coalescence indicator where $b_p = 1$ indicates that there is a coalescence at level p.

Definition 3.1. *For any $n \geq 1$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, we define the signed finite measures $\Gamma_n^{\dagger,b}$ and $\widetilde{\Gamma}_n^{\dagger,b}$ by*

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma_n^{\dagger,b}(F) := \eta_0^{\otimes 2}\, \mathbf{Q}_1^{\dagger,b_0}\, \mathbf{Q}_2^{\dagger,b_1} \cdots \mathbf{Q}_n^{\dagger,b_{n-1}} C_{b_n}(F),$$

and

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \widetilde{\Gamma}_n^{\dagger,b}(F) := \eta_0^{\otimes 2}\, \widetilde{\mathbf{Q}}_1^{\dagger,b_0}\, \widetilde{\mathbf{Q}}_2^{\dagger,b_1} \cdots \widetilde{\mathbf{Q}}_n^{\dagger,b_{n-1}} C_{b_n}(F),$$

with the convention

$$\Gamma_0^{\dagger,b}(F) = \widetilde{\Gamma}_0^{\dagger,b}(F) := \eta_0^{\otimes 2} C_{b_0}.$$

When there is only one coalescence at level p, we write respectively $\Gamma_n^{\dagger,(p)}(F)$ and $\widetilde{\Gamma}_n^{\dagger,(p)}(F)$ instead. When there is no coalescence, we denote respectively $\Gamma_n^{\dagger,(\emptyset)}(F)$ and $\widetilde{\Gamma}_n^{\dagger,(\emptyset)}(F)$.

The connection of the original coalescent tree-based measures proposed in [CDMG11] and the generalized version defined above will be discussed in Section B.1 and Section B.2. By exploiting this novel pair of coalescent Feynman-Kac measures, we have the following alternative representation of the asymptotic variance $\sigma^2_{\gamma_n}(f)$. The rigorous verification is housed in Section C.2.

$$\sigma^2_{\gamma_n}(f) := \sum_{p=0}^{n} \left(\Gamma_n^{\dagger,(p)}(f^{\otimes 2}) - \Gamma_n^{\dagger,(\emptyset)}(f^{\otimes 2}) \right) + \sum_{p=0}^{n-1} \widetilde{\Gamma}_n^{\dagger,(p)}(f^{\otimes 2}). \tag{8}$$

3.2 Term by term asymptotic variance estimators

Thanks to the alternative representation (8) given in the last section, the variance estimation problem is reformulated as how we can estimate the corresponding coalescent Feynman-Kac measures. Using the same idea as in [DG19](Chapter 2), we construct the particle approximation of $\Gamma_n^{\dagger,b}$ and $\widetilde{\Gamma}_n^{\dagger,b}$. They will be referred to as *coalescent tree occupation measures* in this article.

In the following, $\tilde{a}_p^{[2]} = (\tilde{a}_p^1, \tilde{a}_p^2)$ and $\ell_p^{[2]} = (\ell_p^1, \ell_p^2)$ denote two couples of indices between 1 and N, while an $(n+1)$-sequence of couples of indices such that $\ell_p^1 \neq \ell_p^2$ for all $0 \leq p \leq n$ is written

$$\ell_{0:n}^{[2]} = (\ell_0^{[2]}, \cdots, \ell_n^{[2]}) \in \left((N)^2\right)^{\times(n+1)},$$

where $(N)^2 := \{(i,j) \in [N]^2 : i \neq j\}$. Especially, we denote

$$\ell_{p:p+1}^{[2]} := (\ell_p^{[2]}, \ell_{p+1}^{[2]}).$$

Additionally, we use the notation $X_n^{\ell_n^{[2]}} = (X_n^{\ell_n^1}, X_n^{\ell_n^2})$ to shorten the writings. One can also find a toy example in [DG19](Chapter 2) in order to get more intuitions for the following definition.

Definition 3.2. *For any $n \geq 0$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, the estimator $\Gamma_{n,N}^{\ddagger,b}$ of $\Gamma_n^{\dagger,b}$ is defined by*

$$\forall F \in \mathcal{B}_b(E_n^2),\ \Gamma_{n,N}^{\ddagger,b}(F) = \frac{1}{N(N-1)} \sum_{\ell_n^{[2]} \in (N)^2} \left\{ \prod_{p=0}^{n-1} \sum_{\ell_p^{[2]} \in (N)^2} \mathbf{G}_\mathrm{p}^{\ddagger}(\mathbf{X}_\mathrm{p}) \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\} C_{b_n}(F)(X_n^{\ell_n^{[2]}}),$$

with $\mathbf{G}_\mathrm{p}^{\ddagger}$ defined by

$$\forall \boldsymbol{x_p} \in E_p^N, \qquad \mathbf{G}_\mathrm{p}^{\ddagger}(\boldsymbol{x_p}) := \frac{N}{N-1} m^{\otimes 2}(\boldsymbol{x_p})(G_p^{\otimes 2}),$$

and $\lambda_p^b(\tilde{a}_p^{[2]}, \ell_p^{[2]}) \in \{0,1\}$ is the indicator function defined by

$$\lambda_p^b(\tilde{a}_p^{[2]}, \ell_p^{[2]}) := \mathbf{1}_{\{b_p=0\}} \mathbf{1}_{\{\tilde{a}_p^1 = \ell_p^1 \neq \tilde{a}_p^2 = \ell_p^2\}} + \mathbf{1}_{\{b_p=1\}} \mathbf{1}_{\{\tilde{a}_p^1 = \ell_p^1 = \tilde{a}_p^2 \neq \ell_p^2\}}.$$

Notice that, by standard convention, we get

$$\Gamma_{0,N}^{\ddagger,b} := \frac{1}{N(N-1)} \sum_{\ell_0^{[2]} \in (N)^2} C_{b_0}(F)(X_0^{\ell_0^{[2]}}) = \frac{1}{N(N-1)} \sum_{i \neq j} C_{b_0}(F)(X_0^i, X_0^j). \tag{9}$$

Remark. In fact, $\Gamma^{\ddagger,b}_{n,N}$ defined above is exactly the same estimator as $\Gamma^{b}_{n,N}$ defined in Definition 3.2 of [DG19](Chapter 2). The change of notation is due to the change of resampling scheme, and the exact reason lies in a technical result (cf. Proposition C.9). The inhomogeneity of the notation w.r.t. "‡" and "†" is due to some nontrivial combinatorial structure of the asymmetric SMC framework. The detailed explanation can be found in Proposition B.3 and other remarks in Section B.2.

Definition 3.3. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator b, the estimator $\widetilde{\Gamma}^{\dagger,b}_{n,N}$ of $\widetilde{\Gamma}^{\dagger,b}_{n}$ is defined by*

$$\widetilde{\Gamma}^{\dagger,b}_{n,N}(F) := \frac{1}{N(N-1)} \sum_{\ell_n^{[2]}\in(N)^2} \left\{ \prod_{p=0}^{n-1} \sum_{\ell_p^{[2]}\in(N)^2} \widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,b_p}(\ell^{[2]}_{p:p+1}, \mathbf{B}_{\mathrm{p}}, \mathbf{X}_{\mathrm{p}}) \lambda_p^{(\varnothing)}(A_p^{\ell^{[2]}_{p+1}}, \ell_p^{[2]}) \right\} F(X_n^{\ell_n^{[2]}})$$

with $\widetilde{\mathbf{G}}_p^{\dagger,b_p}$ defined as follows:

$$\forall (\ell^{[2]}_{p:p+1}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) \in \left((N)^2\right)^{\times 2} \times \{0,1\}^N \times E_p^N,$$

we let

$$\widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,0}(\ell^{[2]}_{p:p+1}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) := \mathbf{G}_{\mathrm{p}}^{\ddagger}(\mathbf{X}_{\mathrm{p}}) - \frac{1}{N-1} \widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,1}(\ell^{[2]}_{p:p+1}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}),$$

and

$$\begin{aligned}
\widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,1}(\ell^{[2]}_{p:p+1}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) :=& \beta_p^{\ell^1_{p+1}} \beta_p^{\ell^2_{p+1}} m(\boldsymbol{x}_{\boldsymbol{p}})(G_p^2) \\
&+ \beta_p^{\ell^1_{p+1}} (1 - \beta_p^{\ell^2_{p+1}}) m(\boldsymbol{x}_{\boldsymbol{p}})(G_p) \frac{G_p(x_p^{\ell^1_p}) m(\boldsymbol{x}_{\boldsymbol{p}})(G_p) - m(\boldsymbol{x}_{\boldsymbol{p}})(G_p^2)}{\sum_{k\neq \ell^1_p} \left(1 - G_p(X_p^k)\right)/N} \\
&+ \beta_p^{\ell^2_{p+1}} (1 - \beta_p^{\ell^1_{p+1}}) m(\boldsymbol{x}_{\boldsymbol{p}})(G_p) \frac{G_p(x_p^{\ell^2_p}) m(\boldsymbol{x}_{\boldsymbol{p}})(G_p) - m(\boldsymbol{x}_{\boldsymbol{p}})(G_p^2)}{\sum_{k\neq \ell^2_p} \left(1 - G_p(X_p^k)\right)/N}.
\end{aligned}$$

We also define

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma^{(\varnothing)}_{n,N}(F) := \widetilde{\Gamma}^{\dagger,(\varnothing)}_{n,N}(F).$$

Remark. In particular, if G_n is an indicator function for all $n \geq 0$, we consider

$$\widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,1}(\ell^{[2]}_{p:p+1}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) := \beta_p^{\ell^1_{p+1}} \beta_p^{\ell^2_{p+1}} m(\mathbf{X}_{\mathrm{p}})(G_n^2),$$

which leads to a simpler form of $\widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,0}$, i.e.,

$$\widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,0}(\ell^{[2]}_{p:p+1}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) := \beta_p^{\ell^1_{p+1}} \beta_p^{\ell^2_{p+1}} m(\boldsymbol{x}_{\boldsymbol{p}})^{\odot 2}(G_p) + \frac{N}{N-1}(1 - \beta_p^{\ell^1_{p+1}} \beta_p^{\ell^2_{p+1}}) m^{\otimes 2}(\boldsymbol{x}_{\boldsymbol{p}})(G_p^{\otimes 2}). \quad (10)$$

Returning to the coalescent tree-based expansion given in (8), it is natural to define the term by term estimators $\sigma^2_{\gamma_n^N}(f)$ as follows:

$$\sigma^2_{\gamma_n^N}(f) := \left(\sum_{p=0}^{n} \left(\Gamma^{\ddagger,(p)}_{n,N}(f^{\otimes 2}) - \Gamma^{\ddagger,(\varnothing)}_{n,N}(f^{\otimes 2}) \right) + \sum_{p=0}^{n-1} \widetilde{\Gamma}^{\dagger,(p)}_{n,N}(f^{\otimes 2}) \right) \mathbf{1}_{\tau_N \geq n}.$$

Then, by (7), it is natural to consider

$$\sigma^2_{\eta_n^N}(f) := \sigma^2_{\gamma_n^N}(f)\,/\gamma_n^N(1)^2.$$

Therefore, thanks to Theorem B.3 and Corollary B.3.1 , we have the consistency of these term by term variance estimators.

Theorem 3.1 (Consistency of $\sigma^2_{\gamma_n^N}$ and $\sigma^2_{\eta_n^N}$)**.** *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\sup_{N\geq 0} \sqrt{N}\mathbf{E}\left[\left|\sigma^2_{\gamma_n^N}(f) - \sigma^2_{\gamma_n}(f)\right|\right] < +\infty,$$

as well as

$$\sigma^2_{\eta_n^N}\left(f - \eta_n^N(f)\right) - \sigma^2_{\eta_n}\left(f - \eta_n(f)\right) = \mathcal{O}_{\mathbf{P}}\left(\frac{1}{\sqrt{N}}\right).$$

Remark. We do not provide the algorithm to compute these estimators, since, to the best of our knowledge, they can only be computed with time complexity $\mathcal{O}(nN^2)$. Therefore, they mainly serve as theoretical handy tools prove the consistency of the efficient estimator given in Algorithm 5 and Algorithm 6. However, with the same techniques as in these two Algorithms, one should be able to design an algorithm such that each term in the asymptotic variance can be evaluated separately, with time complexity $\mathcal{O}(nN)$. The details are given in Section B.6.

3.3 Unbiased non-asymptotic variance estimator

In this section, we provide an unbiased non-asymptotic variance estimator which is only valid under symmetric resampling scheme, i.e., $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 0$. However, we prove that in the general case, this estimator also yields a consistent asymptotic variance estimator. In fact, in order to provide an unbiased non-asymptotic variance estimator, the idea is much more straightforward: thanks to Theorem 2.1, we know that under symmetric sampling scheme, the estimation $\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}$ is unbiased. As a consequence, one has

$$\mathrm{Var}\left[\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}\right] = \mathbf{E}\left[\gamma_n^N(f)^2\mathbf{1}_{\tau_N\geq n}\right] - \underbrace{\mathbf{E}\left[\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}\right]^2}_{=\gamma_n(f)^2=\gamma_n^{\otimes 2}(f^{\otimes 2})}.$$

It is then clear that constructing an unbiased non-asymptotic variance estimator is equivalent to constructing an unbiased estimator for the measure $\gamma_n^{\otimes 2}$. Therefore, the following proposition is a direct consequence of Proposition C.3. The detailed computation is provided in Algorithm 7 in Section A.

Theorem 3.2. *Assume symmetric resampling, that is, $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 0$. For any test function $f \in \mathcal{B}_b(E_n)$, the estimator $V_n^N(f)$ defined below is an unbiased variance estimator of $\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}$:*

$$V_n^N(f) := \left(\gamma_n^N(f)^2 - \Gamma^{(\emptyset)}_{n,N}(f^{\otimes 2})\right)\mathbf{1}_{\tau_N\geq n}. \tag{11}$$

Remark. In fact, this unbiased estimator can also be used by AMS methods if the image of the reaction coordinate is a finite set, and under some regularity assumption on the resampling kernel (e.g. Assumption 1, 2 of [BGG+16]) is satisfied. Although the associated

IPS is not simulated by symmetric resampling, recent results (cf. [CDGR18]) show that one can construct an IPS with particles defined by some level-indexed processes that are "mathematically symmetrically resampled". It can be regarded as an almost sure equivalence between an artificial asymmetric IPS, i.e. the real-world algorithm, and a symmetric IPS constructed by some abstract mathematical objects. More generally, when the reaction coordinate is finite-valued, the AMS method enters the asymmetric SMC framework. More discussions on this topic can be found in Section 4.1. As a consequence, for all the unbiasedness results in this article, we only use the condition *under symmetric resampling* or *assume symmetric resampling* in order to simplify the writings.

3.4 Connection between the estimators

The connection between the term by term estimators and the non-asymptotic variance estimator is based on Theorem B.1. Unfortunately, to the best of our knowledge, both of these estimators can only be computed with $\mathcal{O}(nN^2)$ time complexity. However, this connection inspired the construction of the efficient consistent estimator provided in the next section. The proof is provided in Section C.5.

Proposition 3.1. *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\sup_{N>1} N\mathbf{E}\left[\left|NV_n^N(f) - \sigma^2_{\eta_n^N}(f)\right|\right] < +\infty, \tag{12}$$

and

$$NV_n^N(f - \eta_n^N(f)) - \sigma^2_{\eta_n^N}(f - \eta_n^N(f)) = \mathcal{O}_{\mathbf{P}}\left(\frac{1}{N}\right). \tag{13}$$

3.5 Efficient asymptotic variance estimators

The efficient variance estimator can be regarded as a "mix" of the term by term estimator and the non-asymptotic estimator provided respectively in Section 3.2 and Section 3.3. Inspired by a by-product (41) in the proof of Proposition 3.1, an intermediate estimator can be proposed as follows

$$\left(N\left(\gamma_n^N(f)^2 - \Gamma^{\ddagger,(\emptyset)}_{n,N}(f^{\otimes 2})\right) + \sum_{p=0}^{n-1} \widetilde{\Gamma}^{\dagger,(p)}_{n,N}(f^{\otimes 2})\right)\mathbf{1}_{\tau_N \geq n}.$$

By the same technique as the one proposed in [LW18], the former term

$$N\left(\gamma_n^N(f)^2 - \Gamma^{\ddagger,(\emptyset)}_{n,N}(f^{\otimes 2})\right)$$

can be computed with $\mathcal{O}(nN)$ time complexity, which corresponds to the term $\mathrm{V}^{\ddagger}$ in Algorithm 5. Hence, the design of an efficient variance estimator amounts to constructing an efficient estimator for $\widetilde{\Gamma}^{\dagger,(p)}_n$. Since we failed to provide an efficient algorithm to compute $\widetilde{\Gamma}^{\dagger,(p)}_{n,N}$, we consider some particle approximation $\widetilde{\Gamma}^{\ddagger,(p)}_{n,N}$ of $\widetilde{\Gamma}^{\dagger,(p)}_n$, such that $\widetilde{\Gamma}^{\ddagger,(p)}_{n,N}$ is "close" enough to the original one $\widetilde{\Gamma}^{\dagger,(p)}_{n,N}$, and which, at the same time, can be computed with $\mathcal{O}(nN)$ time complexity.

Definition 3.4. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator b, the efficient estimator $\widetilde{\Gamma}_{n,N}^{\ddagger,b}$ of $\widetilde{\Gamma}_n^{\dagger,b}$ is defined by*

$$\widetilde{\Gamma}_{n,N}^{\ddagger,b}(F) := \frac{1}{N(N-1)} \sum_{\ell_n^{[2]} \in (N)^2} \left\{ \prod_{p=0}^{n-1} \sum_{\ell_p^{[2]} \in (N)^2} \widetilde{\mathbf{G}}_{\mathrm{p}}^{\ddagger,b_p}(\ell_{p:p+1}^{[2]}, \mathbf{B}_{\mathrm{p}}, \mathbf{X}_{\mathrm{p}}) \lambda_p^{(\emptyset)}(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\} F(X_n^{\ell_n^{[2]}})$$

with $\widetilde{\mathbf{G}}_p^{\ddagger,b_p}$ defined by, $\forall(\ell_{p:p+1}^{[2]}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) \in \left((N)^2\right)^{\times 2} \times \{0,1\}^N \times E_p^N$,

$$\begin{cases} \widetilde{\mathbf{G}}_{\mathrm{p}}^{\ddagger,0}(\ell_{p:p+1}^{[2]}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) := \mathbf{G}_{\mathrm{p}}^{\ddagger}(\boldsymbol{x}_{\boldsymbol{p}}); \\ \widetilde{\mathbf{G}}_{\mathrm{p}}^{\ddagger,1}(\ell_{p:p+1}^{[2]}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}) := \widetilde{\mathbf{G}}_{\mathrm{p}}^{\dagger,1}(\ell_{p:p+1}^{[2]}, \boldsymbol{\beta}_{\boldsymbol{p}}, \boldsymbol{x}_{\boldsymbol{p}}). \end{cases}$$

Proposition 3.2. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator $b \in \{0,1\}^{n+1}$, we have*

$$\sup_{N>1} N\mathbf{E}\left[\left|\widetilde{\Gamma}_{n,N}^{\ddagger,b}(F) - \widetilde{\Gamma}_{n,N}^{\dagger,b}(F)\right|\right] < +\infty.$$

Finally, we provide the efficient asymptotic variance estimators, which are respectively the output of Algorithm 5 and Algorithm 6. For any test function $f \in \mathcal{B}_b(E_n)$, we define

$$\hat{\sigma}_{\gamma_n^N}^2(f) := N\left(\gamma_n^N(f)^2 - \Gamma_{n,N}^{\ddagger,(\emptyset)}\right) + \sum_{p=0}^{n-1} \widetilde{\Gamma}_{n,N}^{\ddagger,(p)}(f^{\otimes 2}), \tag{14}$$

and

$$\hat{\sigma}_{\eta_n^N}^2(f) = \hat{\sigma}_{\gamma_n^N}^2(f)/\gamma_n^N(1)^2.$$

Thanks to Proposition B.4, Proposition 3.2, Proposition B.4, Theorem 3.1 and Theorem 2.1, we have the following consistency result.

Theorem 3.3 (Consistency of $\hat{\sigma}_{\gamma_n^N}^2$ and $\hat{\sigma}_{\eta_n^N}^2$)**.** *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\sup_{N>1} \sqrt{N}\mathbf{E}\left[\left|\hat{\sigma}_{\gamma_n^N}^2(f) - \sigma_{\gamma_n}^2(f)\right|\right] < +\infty,$$

as well as

$$\hat{\sigma}_{\eta_n^N}^2\left(f - \eta_n^N(f)\right) - \sigma_{\eta_n}^2(f - \eta_n(f)) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

4 Discussions

4.1 Unbiasedness condition

We discuss the condition such that the estimations for the unnormalized measures are unbiased. More precisely,

$$\mathbf{E}\left[\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}\right] = \gamma_n(f) \quad \text{and} \quad \mathbf{E}\left[\Gamma_{n,N}^{(\emptyset)}(f)\mathbf{1}_{\tau_N \geq n}\right] = \gamma_n(f)^2.$$

The major motivation is to give an intuitive interpretation of Assumption 2 of [BGG+16] in our fixed level setting. Recent works (cf. [DCGR17] and [CDGR18]) show that Assumption

2 of [BGG+16] can eventually be regarded as a requirement to reformulate AMS methods as Fleming-Viot particle systems, which enters the continuous-time generalization of the current framework, with symmetric resampling, namely, the case where $\dot{Q}_n \equiv \mathring{Q}_n$. More concretely, if the real-world simulation is under asymmetric resampling scheme, but we are somehow able to find out an underlying mathematical structure that is symmetrically resampled, the unbiasedness will be recovered. In addition, as discussed in Section B.1, it also gives a smaller asymptotic variance than the multinomial resampling. Unfortunately, the only setting we could provide that allows this property is the family of AMS methods in the dynamical setting. The readers are referred to [DCGR17], [CDGR18] and [BGG+16] for more details. In this respect, since the real mathematical object is simply a symmetric SMC model, we did not emphasize the more complex condition discussed above, which is also due to the fact that the variance estimators are the same in the symmetric and asymmetric resampling schemes. Moreover, it would be interesting to see if we can implement the same technique in Particle Tempering methods, at least in some specific situations.

4.2 Other comments

On the adaptive SMC models First, we want to mention that the current setting covers the Adaptive Multilevel Splitting methods when the image of the reaction coordinate is a finite set. This is due to the fact that under asymmetric resampling scheme, when the potential of a particle is 1, there is no additional computational cost required to evolve the IPS. In practice, it corresponds to the case where the reaction coordinate function is calculated by some prefixed grid. Next, since the ingredients we need to implement asymmetric SMC are nearly the same as for the classical multinomial SMC, we can therefore consider the corresponding adaptive methods as in our previous work in [DG19](Chapter 2). Since all the technical results are done in a similar style, we expect the adaptive version with Assumption 2 of [DG19](Chapter 2) to be a simple generalization from a mathematical point of view, at the price of some notational complications. Therefore, the variance estimators may be used as a reference if the underlying resampling scheme is changed to the asymmetric one. We also expect the asymptotic variance estimators to apply to another family of adaptive SMC models, which contain an online adaptive resampling strategy, such as [GDM17] and [DMDJ12]. Roughly speaking, the resampling is executed when some summary statistics, such as the popular *Effective Sample Size*, attains some prefixed threshold. In this scenario, as the adaptive model and the fixed reference model are connected by a coupling argument, one is also encouraged to use our estimators in the real-world applications as a reference. However, there are situations where we are sure that the estimators provided in this article will fail. The first example is in Section 2 of [BJKT16]. If the stability property given in Theorem 2.3 is not verified: namely, if the "limit" model has different asymptotic variance, then, it is not possible to conduct variance estimations with our estimators. The same argument also applies to the *Adaptive Tempering* introduced in section 3 of [BJKT16]: it is not possible to use our variance estimator as a reference even if we change the underlying resampling scheme. In all cases, the rigorous analysis of the adaptive context requires more attention on the regularity of the adaptive parametrizations.

On the non-asymptotic variance expansion and long-term behaviors There is an angle that we could but we did not touch in the present work: using the finer analysis given in Lemma C.10 and the decomposition in Theorem B.1, we can give a very sharp upper bound of the non-asymptotic variance, w.r.t. both n and N. Since it is well understood that the non-asymptotic variance may also contribute to the bound encountered in the propagation of chaos property of the particle system (cf. [DMPR09]), we expect that one can also derive the sharp propagation of chaos bound for the non-asymptotic empirical measures. In particular, this kind of analysis can provide information on the bias associated to the estimation $\eta_n^N(f)\mathbf{1}_{\tau_N \geq n}$. The same idea can also be applied to obtain sharp $\mathbb{L}^p$-bound estimates. This is a relatively large topic, the rigorous analysis is thus left for future investigations. Returning to the variance estimation problems, we remark that all the consistent variance estimators provided in this article are essentially for "short-term" models, namely, n is set to be finite and N tends to infinity. When more stability properties of the Feynman-Kac kernels are available, it would be interesting to investigate the fixed-lag variance estimator such as the one introduced in [OD19]. The same kind of estimators, i.e., by considering only part of the genealogy, as well as the survival history, is expected to be more numerically stable in the long term. We expect that the regularity requirements will be the same as in the multinomial case.

On the PMCMC-type kernels Another remark is on the PMCMC-type kernels: starting from a trajectory of the particle system, the new sample is constructed by simulating an IPS with this frozen trajectory, and we pick randomly and uniformly an ancestral lineage in the novel IPS, using its terminal point as the new sample. This kind of kernel does not enter the framework of gAMS since Assumption 2 of [BGG+16] is not verified. Therefore, no level-indexed process can be derived. However, it is a widely used kernel in the Particle Filters and Particle Tempering contexts. In rare-event simulation context, it is also promising in resolving the high dimensional multimodal metastable problems. As a complement, the present work can implement this type of kernels: in fact, one may prefix a very fine grid of levels: since the resampling scheme we use does not require additional computation when all the particles have survived, the implementation is very close to the last-particle AMS methods in practice. Moreover, one can also construct the PMCMC-type kernel by using the standard transition kernel that satisfies Assumption 2 of [BGG+16]. It is possible to study the performance of this Markov kernel with the theoretical tools we provided in this article. This kind of connection is also well illustrated in [ALV18]. Nevertheless, the rigorous analysis is also left for future research.

Appendices

A Algorithms to compute variance estimators

We provide all the supporting algorithms in this section. The matrix-type data structures will be denoted by bold abbreviations, such as **IPS**, **GENE** and **SH**. They stand respectively for *Interacting Particle System*, *genealogy* and *survival history*. For any set E, we denote $\mathcal{M}_{n\times N}(E)$ the collection of all the $n \times N$ matrices with elements taking values in E. For example, the notation $\mathbf{SH}[p, i]$ stands for the element at p-th row and i-th column of the matrix **SH**. In particular, since the state space may vary w.r.t. time horizon, **IPS** is not necessarily a matrix, however, we still use the notation $\mathbf{IPS}[p, i]$ to denote the i-th particle at level p of IPS.

Algorithm 1: Simulation of an IPS with genealogy and survival history.

Require: particle number N, time horizon n, potentials $(G_p; 0 \le p \le n-1)$, Markov kernels $(\mathring{M}_p; 1 \le p \le n)$ and $(\dot{M}_p; 1 \le p \le n)$, initial distribution η_0.

Result: absorbing time $\mathrm{T} \in \{0, 1, \dots, n\}$, particle system **IPS** of size $(n+1) \times N$, genealogy **GENE** $\in \mathcal{M}_{n\times N}([N])$, survival history **SH** $\in \mathcal{M}_{n\times N}(\{0,1\})$, ancestor indices **EVE** $\in \mathcal{M}_{(n+1)\times N}([N])$.

Initialization:
Allocate memory for **IPS**, **GENE**, **SH** and **EVE**;
$\mathrm{T} = 0$;
SumG = 0;
for $i \in \{1, 2, \dots, N\}$ **do**
  $\mathbf{IPS}[0, i] \sim \eta_0$;
  SumG = SumG+$G_0(\mathbf{IPS}[0, i])$;
  $\mathbf{EVE}[0, i] = i$;
end
Iteration:
while SumG > 0 *and* $\mathrm{T} < n$ **do**
  SumG = 0;
  for $i \in \{1, 2, \dots, N\}$ **do**
    U $\sim$ Uniform[0, 1];
    if U $\le G_{\mathrm{T}}(\mathbf{IPS}[\mathrm{T}, i])$ **then**
      ParentIndex = i;
      $\mathbf{IPS}[\mathrm{T}+1, i] \sim \dot{M}_{\mathrm{T}+1}(\mathbf{IPS}[\mathrm{T}, \text{ParentIndex}], \cdot)$;
      $\mathbf{SH}[\mathrm{T}, i] = 1$;
    else
      ParentIndex $\sim$ Categorical $\left(G_{\mathrm{T}}(X_{\mathrm{T}}^1), G_{\mathrm{T}}(X_{\mathrm{T}}^2), \dots, G_{\mathrm{T}}(X_{\mathrm{T}}^N)\right)$;
      $\mathbf{IPS}[\mathrm{T}+1, i] \sim \mathring{M}_{\mathrm{T}+1}(\mathbf{IPS}[\mathrm{T}, \text{ParentIndex}], \cdot)$;
      $\mathbf{SH}[\mathrm{T}, i] = 0$;
    end
    $\mathbf{EVE}[\mathrm{T}+1, i] = \mathbf{EVE}[\mathrm{T}, \text{ParentIndex}]$;
    $\mathbf{GENE}[\mathrm{T}, i]$ = ParentIndex;
    SumG = SumG+$G_{\mathrm{T}+1}(\mathbf{IPS}[\mathrm{T}+1, i])$;
  end
  $\mathrm{T} = \mathrm{T} + 1$;
end
$\mathrm{T} = \max\{0, \mathrm{T}-1\}\mathbf{1}_{\mathrm{T}<n} + n\mathbf{1}_{\mathrm{T}=n}$.

Algorithm 2: Computation of $\gamma_n^N(f)$ and $\eta_n^N(f)$.

Require: absorbing time T, the associated interacting particle system **IPS**, test function f.
Result: estimators $\gamma_n^N(f)$ and $\eta_n^N(f)$.

if $\mathrm{T} < n$ **then**
  $\eta_n^N(f) = 0$;
  $\gamma_n^N(f) = 0$;
else
  Normalizer = 1;
  for $p \in \{0, 1, \ldots, n-1\}$ **do**
    Normalizer = Normalizer $\times \frac{1}{N}\sum_{i=1}^N G_p(\mathbf{IPS}[p,i])$;
  end
  $\eta_n^N(f) = \frac{1}{N}\sum_{i=i}^N f(\mathbf{IPS}[n,i])$;
  $\gamma_n^N(f) = \text{Normalizer} \times \eta_n^N(f)$;
end
return $\gamma_n^N(f)$ *and* $\eta_n^N(f)$.

Now, let us provide the efficient algorithms for the consistent asymptotic variance estimators (cf. Algorithm 5, 6). We need some auxiliary steps: generation of backward genealogy tracing matrix Θ and definition of a special "star inner product" on $\mathbf{R}^3$. They are provided respectively in Algorithm 3 and Algorithm 4. With a slight abuse of notation, we use the notation $\mathcal{M}_{n\times 1}(E)$ to denote the collection of all the array of length n on the set E. For $\mathrm{A} \in \mathcal{M}_{n\times 1}(E)$, we use $\mathrm{A}[p]$ to denote the p-th element of A. To simplify the notation, $\mathrm{A} = \mathrm{zeros}(n, N)$ means that we allocate memory for $\mathrm{A} \in \mathcal{M}_{n\times N}(\mathbf{R})$ and let all the elements of A be 0. In addition, the k-th row of A will be denoted by $\mathrm{A}[k, :]$.

Algorithm 3: Generate backward genealogy tracing matrix: Θ.

Require: absorbing time T, genealogy of an IPS $\mathbf{GENE} \in \mathcal{M}_{n\times N}([N])$.
Result: backward genealogy tracing matrix Θ, where $\Theta[p,i]$ stands for the parent index at level p of i-th particle at level **T**.

Initialization:
Allocate memory for $\Theta \in \mathcal{M}_{\mathrm{T}\times N}([N])$;
Iteration:
for $i \in \{1, 2, \ldots, N\}$ **do**
  CurrentIndex = i;
  for $p \in \{1, 2, \ldots, \mathrm{T}\}$ **do**
    ParentIndex = $\mathbf{GENE}[\mathrm{T}-p, \text{CurrentIndex}]$;
    $\Theta[\mathrm{T}-p, i]$ = ParentIndex;
    CurrentIndex = ParentIndex;
  end
end

Algorithm 4: Compute star inner product in $\mathbf{R}^3$: starProduct(X, Y).

Require: vector $X = (X[1], X[2], X[3]) \in \mathbf{R}^3$, vector $Y = (Y[1], Y[2], Y[3]) \in \mathbf{R}^3$.
Result: value of $\langle X, Y\rangle_\star \in \mathbf{R}$.

return starProduct(X, Y) = $\mathrm{X}[1]\times\mathrm{Y}[1] + \mathrm{X}[2]\times\mathrm{Y}[3] + \mathrm{X}[3]\times\mathrm{Y}[2]$.

Algorithm 5: Consistent asymptotic variance estimator for $\gamma_n^N(f)$.

Require: $(G_p; 0 \leq p \leq n-1)$, T, **IPS**, test function f, Θ, **SH**, **EVE** and $\gamma_n^N(f)$.
Result: asymptotic variance estimator $\hat{\sigma}^2_{\gamma_n^N}(f)$.

if $\mathrm{T} < n$ **then**
  return 0;
else
  Allocate memory for MeanG and MeanG2 $\in \mathcal{M}_{n \times 1}(\mathbf{R})$;
  Normalizer = 0;
  for $p \in \{0, 1, \ldots, n-1\}$ **do**
    $\mathrm{MeanG}[p] = \frac{1}{N} \sum_{i=1}^{N} G_p(\mathbf{IPS}[p, i])$;
    $\mathrm{MeanG2}[p] = \frac{1}{N} \sum_{i=1}^{N} G_p(\mathbf{IPS}[p, i])^2$;
    Normalizer = Normalizer × MeanG[p];
  end
  ArrayEve = zeros(N, 1);
  for $i \in \{1, 2, \ldots, N\}$ **do**
    ArrayEve[**EVE**[n, i]] = ArrayEve[**EVE**[n, i]] + f(**IPS**[n, i]) × Normalizer;
  end
  $\mathrm{SumEve} = \sum_{i=1}^{N} \mathrm{ArrayEve}[i]^2$;
  $\mathrm{V}^{\ddagger} = N \times \left(\gamma_n^N(f)^2 - \left[\left(N \times \gamma_n^N(f)\right)^2 - \mathrm{SumEve}\right] \times \frac{N^{n-1}}{(N-1)^{n+1}}\right)$;
  $\widetilde{\mathrm{V}}^{\dagger} = 0$;
  for $p \in \{0, 1, \ldots, n-1\}$ **do**
    MatrixEve = zeros(N, 3);
    for $i \in \{1, 2, \ldots, N\}$ **do**
      Index = $\Theta[p, i]$;
      IndexPrime = $\Theta[p+1, i]$;
      F = f(**IPS**[n, i]) × Normalizer/MeanG[p];
      MatrixEve[i, 1] = **SH**[p, IndexPrime] × $\sqrt{\mathrm{MeanG2}[p]}$ × F;
      MatrixEve[i, 2] = **SH**[p, IndexPrime] × $\frac{N \times \left(G_p(\mathbf{IPS}[p, \mathrm{Index}]) \times \mathrm{MeanG}[p] - \mathrm{MeanG2}[p]\right)}{N - 1 - N \times \mathrm{MeanG}[p] + G_p(\mathbf{IPS}[p, \mathrm{Index}])}$ × F;
      MatrixEve[i, 3] = (1 − **SH**[p, IndexPrime]) × MeanG[p] × F;
    end
    SumMatrixEve = zeros(N, 3);
    for $i \in \{1, 2, \ldots, N\}$ **do**
      SumMatrixEve[**EVE**[n, i], :] = SumMatrixEve[**EVE**[n, i], :] + MatrixEve[i, :];
    end
    SumEve = 0;
    for $i \in \{1, 2, \ldots, N\}$ **do**
      SumEve = SumEve + starProduct(SumMatrixEve[i], SumMatrixEve[i]);
    end
    for $i \in \{2, \ldots, N\}$ **do**
      MatrixEve[1, :] = MatrixEve[1, :] + MatrixEve[i, :];
    end
    SumCurrent = starProduct(MatrixEve[1, :], MatrixEve[1, :]) − SumEve;
    $\widetilde{\mathrm{V}}^{\dagger} = \widetilde{\mathrm{V}}^{\dagger} + \mathrm{SumCurrent} \times \frac{N^{n-3}}{(N-1)^{n-1}}$;
  end
end
return $\mathrm{V}^{\ddagger} + \widetilde{\mathrm{V}}^{\dagger}$.

Algorithm 6: Consistent asymptotic variance estimator for $\eta_n^N(f)$.

Require: $(G_p; 0 \leq p \leq n-1)$, T, **IPS**, test function f, Θ, **SH**, **EVE** and $\eta_n^N(f)$.
Result: asymptotic variance estimator $\hat{\sigma}^2_{\eta_n^N}(f - \eta_n^N(f))$.

```
if T < n then
    return 0;
else
    Allocate memory for MeanG and MeanG2 ∈ M_{n×1}(R);
    for p ∈ {0, 1, ..., n − 1} do
        MeanG[p] = (1/N) ∑_{i=1}^{N} G_p(IPS[p, i]);
        MeanG2[p] = (1/N) ∑_{i=1}^{N} G_p(IPS[p, i])^2;
    end
    ArrayEve = zeros(N, 1);
    for i ∈ {1, 2, ..., N} do
        ArrayEve[EVE[n, i]] = ArrayEve[EVE[n, i]] + f(IPS[n, i]) − η_n^N(f);
    end
    SumEve = ∑_{i=1}^{N} ArrayEve[i]^2;
    V‡ = SumEve × N^n / (N−1)^{n+1};
    Ṽ† = 0;
    for p ∈ {0, 1, ..., n − 1} do
        MatrixEve = zeros(N, 3);
        for i ∈ {1, 2, ..., N} do
            Index = Θ[p, i];
            IndexPrime = Θ[p + 1, i];
            F = ( f(IPS[n, i]) − η_n^N(f) ) /MeanG[p];
            MatrixEve[i, 1] = SH[p, IndexPrime] × √(MeanG2[p]) × F;
            MatrixEve[i, 2] = SH[p, IndexPrime] × N×(G_p(IPS[p, Index])×MeanG[p]−MeanG2[p]) / (N−1−N×MeanG[p]+G_p(IPS[p, Index])) × F;
            MatrixEve[i, 3] = (1 − SH[p, IndexPrime]) × MeanG[p] × F;
        end
        SumMatrixEve = zeros(N, 3);
        for i ∈ {1, 2, ..., N} do
            SumMatrixEve[EVE[n, i], :] = SumMatrixEve[EVE[n, i], :] + MatrixEve[i, :];
        end
        SumEve = 0;
        for i ∈ {1, 2, ..., N} do
            SumEve = SumEve + starProduct(SumMatrixEve[i], SumMatrixEve[i]);
        end
        for i ∈ {2, ..., N} do
            MatrixEve[1, :] = MatrixEve[1, :] + MatrixEve[i, :];
        end
        SumCurrent = starProduct(MatrixEve[1, :], MatrixEve[1, :]) − SumEve;
        Ṽ† = Ṽ† + SumCurrent × N^{n−3} / (N−1)^{n−1};
    end
end
return V‡ + Ṽ†.
```

Algorithm 7: Unbiased non-asymptotic variance estimator for $\gamma_n^N(f)$.

Require: $(G_p; 0 \leq p \leq n-1)$, T, **IPS**, test function f, **GENE**, **SH**, **EVE** and $\gamma_n^N(f)$.
Result: non-asymptotic variance estimator $V_n^N(f)$.

if $\mathrm{T} < n$ **then**
    return 0;
else
    Allocate memory for MeanG, MeanG2 and MeanGdot2 $\in \mathcal{M}_{n\times 1}(\mathbf{R})$;
    for $p \in \{0, 1, \ldots, n-1\}$ **do**
        $\mathrm{MeanG}[p] = \frac{1}{N}\sum_{i=1}^{N} G_p(\mathbf{IPS}[p,i])$;
        $\mathrm{MeanG2}[p] = \frac{1}{N}\sum_{i=1}^{N} G_p(\mathbf{IPS}[p,i])^2$;
        $\mathrm{MeanGdot2}[p] = (\mathrm{MeanG}[p] - \mathrm{MeanG2}[p]/N) \times \frac{N}{N-1}$;
    end
    $\mathrm{V}^{(\varnothing)} = 0$;
    for $i \in \{1, 2, \ldots, N-1\}$ **do**
        for $j \in \{i+1, \ldots, N\}$ **do**
            if $\mathbf{EVE}[n,i] \neq \mathbf{EVE}[n,j]$ **then**
                $\mathrm{ProdCouple} = f(\mathbf{IPS}[n,i]) \times f(\mathbf{IPS}[n,j])$;
                $\mathrm{Index1} = i$, $\mathrm{Index2} = j$;
                for $p \in \{0, 1, \ldots, n-1\}$ **do**
                    $\mathrm{ParentIndex1} = \mathbf{GENE}[n-p-1, \mathrm{Index1}]$;
                    $\mathrm{ParentIndex2} = \mathbf{GENE}[n-p-1, \mathrm{Index2}]$;
                    if $\mathbf{SH}[n-p-1, \mathrm{Index1}] = 1$ & $\mathbf{SH}[n-p-1, \mathrm{Index2}] = 1$ **then**
                        $\mathrm{ProdCouple} = \mathrm{ProdCouple} \times \mathrm{MeanGdot2}[n-p-1]$;
                    else if $\mathrm{SH}[n-p-1, \mathrm{Index1}] = 1$ & $\mathrm{SH}[n-p-1, \mathrm{Index2}] = 0$ **then**
                        $\mathrm{ProdCouple} = \mathrm{ProdCouple} \times \mathrm{MeanG}[n-p-1] \times \Big\{\mathrm{MeanG}[n-p-1] \times \frac{N}{N-1} - N \times \frac{G_{n-p-1}(\mathbf{IPS}[n-p-1,\mathrm{ParentIndex1}]) \times \mathrm{MeanG}[n-p-1] - \mathrm{MeanG2}[n-p-1]}{(N-1)\times\big(N-1-\mathrm{MeanG}[n-p-1]+G_{n-p-1}(\mathbf{IPS}[n-p-1,\mathrm{ParentIndex1}])\big)}\Big\}$;
                    else if $\mathrm{SH}[n-p-1, \mathrm{Index1}] = 0$ & $\mathrm{SH}[n-p-1, \mathrm{Index2}] = 1$ **then**
                        $\mathrm{ProdCouple} = \mathrm{ProdCouple} \times \mathrm{MeanG}[n-p-1] \times \Big\{\mathrm{MeanG}[n-p-1] \times \frac{N}{N-1} - N \times \frac{G_{n-p-1}(\mathbf{IPS}[n-p-1,\mathrm{ParentIndex2}]) \times \mathrm{MeanG}[n-p-1] - \mathrm{MeanG2}[n-p-1]}{(N-1)\times\big(N-1-\mathrm{MeanG}[n-p-1]+G_{n-p-1}(\mathbf{IPS}[n-p-1,\mathrm{ParentIndex2}])\big)}\Big\}$;
                    else
                        $\mathrm{ProdCouple} = \mathrm{ProdCouple} \times \frac{N}{N-1} \times \mathrm{MeanG}[n-p-1]^2$;
                    end
                    $\mathrm{Index1} = \mathrm{ParentIndex1}$;
                    $\mathrm{Index2} = \mathrm{ParentIndex2}$;
                end
                $\mathrm{V}^{(\varnothing)} = \mathrm{V}^{(\varnothing)} + \mathrm{ProdCouple}$;
            end
        end
    end
end
$\mathrm{V}^{(\varnothing)} = 2 \times \mathrm{V}^{(\varnothing)}/(N(N-1))$;
return $\gamma_n^N(f)^2 - \mathrm{V}^{(\varnothing)}$.

Above, we also provide Algorithm 7 to compute unbiased non-asymptotic variance estimator for $\gamma_n^N(f)$. A tremendous amount of effort has been spent in order to find an $\mathcal{O}(nN)$ time complexity algorithm, which, unfortunately, does not pay back. The Algorithm 7 is of $\mathcal{O}(nN^2)$ time complexity, which means that even with a very little n, the computation will be intractable when $N > 10^6$. In fact, the construction for the consistent asymptotic variance estimators, i.e., Algorithm 5 and Algorithm 6, of $\mathcal{O}(nN)$ time complexity is highly nontrivial, and we failed to apply the same technique to reduce the time complexity for the unbiased non-asymptotic variance estimator. Still, this estimator may be useful for rare-event simulation problems or other applications whose target measure is γ_n. One can thus take advantage of the parallel computing for relatively small N. Though, the crude estimator is a by-product in this case, the average of this estimator may still be typically more accurate and the statistical inference for the variance of the variance estimator is also available if an unbiased variance estimator is provided. Note that the *lack-of-bias* is not free in general, and we provide a relatively general condition in Section 4.1. In fact, even without unbiased condition, this estimator multiplied by N is also a consistent asymptotic variance estimator for $\gamma_n^N(f)$. Another remark is that $\mathrm{V}^\ddagger$ found in Algorithm 5 and Algorithm 6 represent respectively the estimators $NV_n^N(f)\gamma_n^N(1)^2$ and $NV_n^N(f - \eta_n^N(f))$ provided in [LW18]. Due to the change of resampling scheme, some modifications, namely, $\widetilde{\mathrm{V}}^\dagger$ have to be taken into consideration. Now, we provide the time complexity and space complexity of the algorithms in the SMC context. The multinomial resampling scheme will be set as benchmark, with the variance estimators provided in [LW18].

Estimation	Time complexity	Space complexity
$\eta_n^N(f)$ or $\gamma_n^N(f)$	$\mathcal{O}(nN)$	$\mathcal{O}(N)$
non-asymptotic variance of $\gamma_n^N(f)$	$\mathcal{O}(nN)$	$\mathcal{O}(N)$
asymptotic variance of $\gamma_n^N(f)$	$\mathcal{O}(nN)$	$\mathcal{O}(N)$
asymptotic variance of $\eta_n^N(f)$	$\mathcal{O}(nN)$	$\mathcal{O}(N)$

Table 1: Time and space complexity under multinomial resampling scheme.

Estimation	Time complexity	Space complexity
$\eta_n^N(f)$ or $\gamma_n^N(f)$	$\mathcal{O}(nN)$	$\mathcal{O}(N)$
non-asymptotic variance of $\gamma_n^N(f)$	$\mathcal{O}(nN^2)$	$\mathcal{O}(nN)$
asymptotic variance of $\gamma_n^N(f)$	$\mathcal{O}(nN)$	$\mathcal{O}(nN)$
asymptotic variance of $\eta_n^N(f)$	$\mathcal{O}(nN)$	$\mathcal{O}(nN)$

Table 2: Time and space complexity under asymmetric resampling scheme.

We remark that we did not provide the algorithm to compute $\gamma_n^N(f)$ and $\eta_n^N(f)$ with $\mathcal{O}(N)$ space complexity, which is readily obtained with some modification of Algorithm 1, since we mainly focus on the variance estimation problems. We can see from Table 1 and Table 2 that the main drawback of the presented setting is the space complexity and the unbiased variance estimator for the non-asymptotic variance of $\gamma_n^N(f)$. However, the main computational consumption, in general, is brought by the resampling kernels $\mathring{M}_n$

at each iteration of the algorithm. If a cheap kernel $\dot{M}_n$ is available, it is expected that the asymmetric resampling scheme would dramatically reduce the computational cost of the simulation of IPS. In fact, in real-world applications, the computational consumption of the asymptotic variance estimators is negligible compared to the simulation of IPS. In addition, the asymmetric setting also gives smaller variance in some specific situations, such as AMS methods.

B Coalescent tree-based expansions

As we have seen in Theorem 2.2, the asymptotic variance becomes sophisticated when the asymmetric resampling is implemented. Therefore, we need to develop a novel mathematical language in order to conduct calculations and eventually, to understand the structures behind. In this section, we give a detailed development of the coalescent tree-based expansions encountered in the asymmetric SMC framework. Before going further, let us list some definitions and properties associated to McKean-type Feynman-Kac kernel $Q_{n,\mu}$ within reach by some straightforward algebraic calculations in the following proposition. For the sake of simplification, these properties will be of constant use in the following sections and may be applied without reference.

Proposition B.1. *For any probability measure $\mu \in \mathcal{P}(E_{n-1})$ and test function $f \in \mathcal{B}_b(E_n)$, we have the following properties:*

(i) $\sup_{x\in E_{n-1}} Q_{n,\mu}(f)(x) \leq 2\,\|f\|_\infty$.

(ii) *We define $R_{n,\mu}$ by*

$$\forall (x,A) \in E_{n-1}\times\mathcal{B}(E_n), \quad R_{n,\mu}(x,A) := \mu(G_{n-1})\dot{Q}_n(x,A) - G_{n-1}(x)\mu\mathring{Q}_n(A), \tag{15}$$

with the convention

$$\forall x \in E_0, \quad R_{0,\mu}(x,A) := \eta_0(A). \tag{16}$$

Then, we have

$$Q_{n,\mu}(f)(x) = \mu\mathring{Q}_n(f) + R_{n,\mu}(f)(x),$$

and

$$\eta_{n-1}R_{n,\eta_{n-1}}(f) = 0.$$

(iii) *We have*

$$Q_{n,\mu}(f)(x)^2 = \mu\mathring{Q}_n(f)^2 + R_{n,\mu}(f)(x)^2 + 2\mu\mathring{Q}_n(f)R_{n,\mu}(f)(x),$$

as well as

$$\eta_{n-1}\left(Q_{n,\eta_{n-1}}(f)^2\right) = \eta_{n-1}\mathring{Q}_n(f)^2 + \eta_{n-1}\left(R_{n,\eta_{n-1}}(f)^2\right). \tag{17}$$

B.1 Original coalescent tree-based measures

First, let us recall the original coalescent tree-based expansion introduced in [CDMG11]. The following definition is adopted from the Definition 3.1 of [DG19](Chapter 2), which is essentially the same as the one introduced in [CDMG11]. A more general version for the particle block of size greater than 2 can be found in [DMPR09].

Definition B.1. *For any $n' \geq n$, we associate with any coalescence indicator $b \in \{0,1\}^{n'+1}$ the nonnegative measures $\Gamma_n^b \in \mathcal{M}_+(E_n^2)$ defined for any $F \in \mathcal{B}_b(E_n^2)$ by*

$$\Gamma_n^b(F) := \eta_0^{\otimes 2} C_{b_0} Q_1^{\otimes 2} C_{b_1} \cdots Q_n^{\otimes 2} C_{b_n}(F).$$

When there is only one coalescence at, say, level p, we write $\Gamma_n^{(p)}(F)$ instead of $\Gamma_n^b(F)$ (see Figure 2). When there is no coalescence at all, that is $b = (\emptyset)$, we have

$$\Gamma_n^{(\emptyset)}(F) = \gamma_n^{\otimes 2}(F).$$

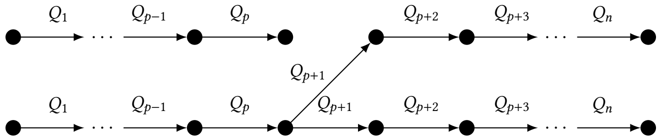

Figure 2: A representation of the original coalescent tree-based measure $\Gamma_n^{(p)}$.

Comparison of asymptotic variance. Now, we suppose that $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 1$. Let us go back to the form of the asymptotic variance $\sigma^2_{\gamma_n}(f)$ defined in (6). It is easy to verify that

$$\Gamma_n^{(p)}(f^{\otimes 2}) = \gamma_p(1)\, \gamma_p\left(Q_{p,n}(f)^2\right). \tag{18}$$

By applying (17) with some standard algebraic manipulations, we have

$$\sigma^2_{\gamma_n}(f) = \sum_{p=0}^{n} \left(\Gamma_n^{(p)}(f^{\otimes 2}) - \Gamma_n^{(\emptyset)}(f^{\otimes 2})\right) - \sum_{p=1}^{n} \gamma_{p-1}(1)\gamma_{p-1}\left(R_{p,\eta_{p-1}} Q_{p,n}(f)^2\right). \tag{19}$$

One may notice that the first term corresponds to the asymptotic variance of the multinomial resampling scheme (see., e.g Theorem 2.1 of [DG19](Chapter 2)). Since the term

$$\gamma_{p-1}(1)\gamma_{p-1}\left(R_{p,\eta_{p-1}} Q_{p,n}(f)^2\right) \tag{20}$$

is nonnegative, we deduce that the choice $\dot{Q}_n \equiv \mathring{Q}_n$ is always better than the multinomial resampling scheme in terms of asymptotic variance. Moreover, we notice that the original coalescent tree-based measures introduced in [CDMG11] and [DG19](Chapter 2) failed to provide a full description of the asymptotic variance, even in this simple symmetric case. This is the main difficulty compared to the multinomial resampling scheme, where the alternative representation is free. Therefore, we need to develop some new tools to understand the term given in (20).

B.2 Coalescent Feynman-Kac kernels

As we have seen in the last section, the original coalescent tree-based measures fail to provide insights on the asymptotic variance $\sigma^2_{\gamma_n}(f)$. In order to go one step further, let us go back to Definition B.1. We consider the following alternative writing

$$\Gamma_n^b(F) := \eta_0^{\otimes 2} \underbrace{C_{b_0} Q_1^{\otimes 2}}_{\mathbf{Q}_1^{b_0}} \underbrace{C_{b_1} Q_2^{\otimes 2}}_{\mathbf{Q}_2^{b_1}} \cdots \underbrace{C_{b_{n-1}} Q_n^{\otimes 2}}_{\mathbf{Q}_n^{b_{n-1}}} C_{b_n}(F),$$

which gives a similar definition as γ_n based on the partial semigroup structure of the Feynman-Kac kernels:

$$\Gamma_n^b(F) := \eta_0^{\otimes 2} \cdot \mathbf{Q}_1^{b_0} \cdot \mathbf{Q}_2^{b_1} \cdots \mathbf{Q}_n^{b_{n-1}} C_{b_n}(F).$$

We say $\mathbf{Q}_n^{b_{n-1}}$ conserves the structure of coalescence if

$$\forall b_{n-1} \in \{0,1\}, \quad C_{b_{n-1}} \mathbf{Q}_n^{b_{n-1}} \equiv \mathbf{Q}_n^{b_{n-1}}.$$

This simple observation gives an interesting idea on how we could possibly overcome the difficulties encountered in the asymptotic variance representation: we change the construction of the partial semigroup according to our asymmetric resampling scheme, in order to establish the coalescent tree-based expansion of the asymptotic variance. The first pair of replacement is for $\mathbf{Q}_n^0$ and $\mathbf{Q}_n^1$. We define

$$\begin{cases} \mathbf{Q}_n^0 := \left(\mathring{Q}_n + \eta_{n-1}^{\otimes 2}(1 \otimes G_{n-1})(\dot{Q}_n - \mathring{Q}_n) \right)^{\otimes 2}; \\ \mathbf{Q}_n^1 := C_1 \mathbf{Q}_n^0. \end{cases} \tag{21}$$

This replacement is compatible with the notation above if $\dot{Q}_n \equiv \mathring{Q}_n$. Note that

$$\eta_{n-1}^{\otimes 2}(1 \otimes G_{n-1}) = \eta_{n-1}(G_{n-1}).$$

Next, we need to introduce several important coalescent Feynman-Kac kernels, which, at the moment, is not as intuitive as the one introduced above. Their introduction is motivated by an observation in the proof of a technical result (cf. Proposition C.9):

(i) $\begin{cases} \mathbf{Q}_n^{\dagger,0} := \mathbf{Q}_n^0; \\ \mathbf{Q}_n^{\dagger,1} := \mathbf{Q}_n^1 - \eta_{n-1}^{\otimes 2}(G_{n-1}^{\otimes 2}) C_1 \dot{Q}_n^{\otimes 2}. \end{cases}$

(ii) $\begin{cases} \widetilde{\mathbf{Q}}_n^{\dagger,0} := \mathbf{Q}_n^0; \\ \widetilde{\mathbf{Q}}_n^{\dagger,1} := \eta_{n-1}^{\otimes 2}(1 \otimes G_{n-1}) \left[(G_{n-1} \times \dot{Q}_n) \otimes \mathring{Q}_n + \mathring{Q}_n \otimes (G_{n-1} \times \dot{Q}_n) \right] \\ \qquad\qquad + \eta_{n-1}^{\otimes 2}(C_1 G_{n-1}^{\otimes 2}) \left[\dot{Q}_n^{\otimes 2} - \dot{Q}_n \otimes \mathring{Q}_n - \mathring{Q}_n \otimes \dot{Q}_n \right]. \end{cases}$

(iii) $\begin{cases} \check{\mathbf{Q}}_n^{\dagger,0} := \mathbf{Q}_n^0; \\ \check{\mathbf{Q}}_n^{\dagger,1} := C_1 \widetilde{\mathbf{Q}}_n^{\dagger,1} - \eta_{n-1}^{\otimes 2}(C_1 G_{n-1}^{\otimes 2}) C_1 \dot{Q}_n^{\otimes 2}. \end{cases}$

(iv) $\forall b_{n-1} \in \{0,1\}$, $\widetilde{\mathbf{Q}}_{n,(N)}^{\dagger,b_{n-1}} := \frac{1}{N-1} \widetilde{\mathbf{Q}}_n^{\dagger,b_{n-1}}$.

(v) $\forall b_{n-1} \in \{0,1\}$, $\check{\mathbf{Q}}_{n,(N)}^{\dagger,b_{n-1}} := \frac{1}{N-1} \check{\mathbf{Q}}_n^{\dagger,b_{n-1}}$.

(vi) $\begin{cases} \mathbf{Q}_{n,(N)}^{\ddagger,0} := \mathbf{Q}_n^{\dagger,0} + \widetilde{\mathbf{Q}}_{n,(N)}^{\dagger,1}; \\ \mathbf{Q}_{n,(N)}^{\ddagger,1} := \mathbf{Q}_n^{\dagger,1} + \check{\mathbf{Q}}_{n,(N)}^{\dagger,1}. \end{cases}$

It is readily checked that they are all uniformly finite transition kernels and that, except the kernels with "$\sim$", namely,

$$\widetilde{\mathbf{Q}}_n^{\dagger,b_{n-1}} \quad \text{and} \quad \widetilde{\mathbf{Q}}_{n,N}^{\dagger,b_{n-1}}$$

all of the other kernels conserve the coalescence structure. This observation may be the intrinsic reason why they play a particularly important role in variance related problems.

No matter how anecdotal it seems, we claim that these kernels are at the core of the analysis of the variance related problems. Although we are not able to clarify the exact purpose of the construction of these coalescent Feynman-Kac kernels at the moment, we can explain, however, the logic of our notation: the number of daggers "†" indicates the number of kernels between $b_{n-1} = 0$ and $b_{n-1} = 1$, that is changed from the original definition (21). At the same time, the kernel for $b_{n-1} = 1$ is always replaced before the kernel for $b_{n-1} = 0$. This is why all the kernels that have only one dagger share the same $\mathbf{Q}_n^0$ for the case $b_{n-1} = 0$. Since the number of particles N is also involved in the definition, we add parenthesis "(N)" to specify the number of particles N, in order to differentiate from the coalescent tree occupation measures. With a slight abuse of notation, when there is no ambiguity, we omit the part "(N)" for simplicity. For example, we may use $\mathbf{Q}_n^{\ddagger,b}$ to denote $\mathbf{Q}_{n,(N)}^{\ddagger,b}$. All the kernels defined from point (i) to point (v) are introduced to describe the binary decomposition w.r.t. "+" in the definition of point (vi). We say that the kernels defined above are in the class $\mathcal{Q}_n^{(2)}$, or the kernels are of $\mathcal{Q}_n^{(2)}$-class.

Next, we define the generalized coalescent tree-based measures. In this article, they will be referred to as the coalescent Feynman-Kac measures, by using the partial semi-group properties of these coalescent Feynman-Kac kernels. For example, we denote

$$\mathbf{Q}_{p,n}^{b} := \mathbf{Q}_{p+1,n}^{b_p}\, \mathbf{Q}_{p+2,n}^{b_{p+1}} \cdots \mathbf{Q}_{p,n}^{b_{n-1}},$$

with the convention

$$\mathbf{Q}_{n,n}^{b_{n-1}} := Q_{n,n}^{\otimes 2}.$$

Definition B.2. *For any $n \geq 1$, $N \geq 2$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, we define the signed finite measures $\Gamma_{n,(N)}^{\ddagger,b}$ by*

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma_{n,(N)}^{\ddagger,b}(F) := \eta_0^{\otimes 2}\, \mathbf{Q}_{1,(N)}^{\ddagger,b_0}\, \mathbf{Q}_{2,(N)}^{\ddagger,b_1} \cdots \mathbf{Q}_{n,(N)}^{\ddagger,b_{n-1}} C_{b_n}(F),$$

with the convention

$$\Gamma_{0,(N)}^{\ddagger,b}(F) = \eta_0^{\otimes 2} C_{b_0}.$$

Similar as in Definition B.1, when there is only one coalescence at level p, we write $\Gamma_{n,(N)}^{\ddagger,(p)}(F)$ instead. When there is no coalescence , we denote $\Gamma_{n,(N)}^{\ddagger,(\emptyset)}(F)$.

Meanwhile, we define the $\mathcal{Q}_{\hat{n}}^{(2)}$-class kernels by replacing all the $\eta_{n-1}^{\otimes 2}$ in the definition above with the empirical sub-probability measure

$$(\eta_{n-1}^N)^{\odot 2}\mathbf{1}_{\tau_N \geq n-1} := m^{\odot 2}(\mathbf{X_{n-1}})\mathbf{1}_{\tau_N \geq n-1}.$$

In regard to the notation, all the "n" in the definition will be replaced by "$\hat{n}$" correspondingly:

(i) $\begin{cases} \mathbf{Q}_{\hat{n}}^0 := \left(\mathring{Q}_n + (\eta_{n-1}^N)^{\odot 2}(1 \otimes G_{n-1})\mathbf{1}_{\tau_N \geq n-1}(\dot{Q}_n - \mathring{Q}_n)\right)^{\otimes 2}; \\ \mathbf{Q}_{\hat{n}}^1 := C_1 \mathbf{Q}_{\hat{n}}^0. \end{cases}$

(ii) $\begin{cases} \mathbf{Q}_{\hat{n}}^{\dagger,0} := \mathbf{Q}_{\hat{n}}^0; \\ \mathbf{Q}_{\hat{n}}^{\dagger,1} := \mathbf{Q}_{\hat{n}}^1 - (\eta_{n-1}^N)^{\odot 2}(G_{n-1}^{\otimes 2})\mathbf{1}_{\tau_N \geq n-1} C_1 \dot{Q}_{\hat{n}}^{\otimes 2}. \end{cases}$

$$\text{(iii)} \begin{cases} \widetilde{\mathrm{Q}}_{\hat{n}}^{\dagger,0} := \mathrm{Q}_{\hat{n}}^{0}; \\ \widetilde{\mathrm{Q}}_{\hat{n}}^{\dagger,1} := (\eta_{n-1}^{N})^{\odot 2}(1 \otimes G_{n-1})\mathbf{1}_{\tau_N \geq n-1}\left[(G_{n-1} \times \dot{Q}_{\hat{n}}) \otimes \mathring{Q}_{\hat{n}} + \mathring{Q}_{\hat{n}} \otimes (G_{n-1} \times \dot{Q}_{\hat{n}})\right] \\ \qquad +(\eta_{n-1}^{N})^{\odot 2}(C_1 G_{n-1}^{\otimes 2})\mathbf{1}_{\tau_N \geq n-1}\left[\dot{Q}_{\hat{n}}^{\otimes 2} - \dot{Q}_{\hat{n}} \otimes \mathring{Q}_{\hat{n}} - \mathring{Q}_{\hat{n}} \otimes \dot{Q}_{\hat{n}}\right]. \end{cases}$$

$$\text{(iv)} \begin{cases} \check{\mathrm{Q}}_{\hat{n}}^{\dagger,0} := \mathrm{Q}_{\hat{n}}^{0}; \\ \check{\mathrm{Q}}_{\hat{n}}^{\dagger,1} := C_1 \widetilde{\mathrm{Q}}_{\hat{n}}^{\dagger,1} - (\eta_{n-1}^{N})^{\odot 2}(C_1 G_{n-1}^{\otimes 2})\mathbf{1}_{\tau_N \geq n-1} C_1 \dot{Q}_{\hat{n}}^{\otimes 2}. \end{cases}$$

(v) $\forall b_{n-1} \in \{0,1\}$, $\widetilde{\mathrm{Q}}_{\hat{n},(N)}^{\dagger,b_{n-1}} := \frac{1}{N-1}\widetilde{\mathrm{Q}}_{\hat{n}}^{\dagger,b_{n-1}}$.

(vi) $\forall b_{n-1} \in \{0,1\}$, $\check{\mathrm{Q}}_{\hat{n},(N)}^{\dagger,b_{n-1}} := \frac{1}{N-1}\check{\mathrm{Q}}_{\hat{n}}^{\dagger,b_{n-1}}$.

$$\text{(vii)} \begin{cases} \mathrm{Q}_{\hat{n},(N)}^{\ddagger,0} := \mathrm{Q}_{\hat{n}}^{\dagger,0} + \widetilde{\mathrm{Q}}_{\hat{n},(N)}^{\dagger,1}; \\ \mathrm{Q}_{\hat{n},(N)}^{\ddagger,1} := \mathrm{Q}_{\hat{n}}^{\dagger,1} + \check{\mathrm{Q}}_{\hat{n},(N)}^{\dagger,1}. \end{cases}$$

We remark that the kernels of $\mathcal{Q}_{\hat{n}}^{(2)}$-class will not be used to define the coalescent tree-based measures, the introduction is purely for technical reasons (cf. Proposition C.9, Proposition C.10, Lemma C.10 and Lemma C.11). They are eventually proved to be very "close" to the $\mathcal{Q}_n$-class kernels (cf. Lemma C.12) by the propagation of chaos property of the IPS (cf. Proposition C.5).

B.3 Binary decompositions

Next, we define some auxiliary coalescent Feynman-Kac kernels using the same idea. Before that, we need some new notation to describe the coalescence structure that is a little bit more complicated than the basic binary structure illustrated in Figure 2. For one coalescence indicator b, we use $|b|$ to denote the number of 1 in b, namely

$$|b| := \sum_{p=0}^{n} |b_p|.$$

Using the same definition as above, for two coalescence indicators b and b' in $\{0,1\}^{n+1}$, the notation $|b - b'|$ denotes the number of different elements between b and b'. More precisely,

$$|b - b'| = \sum_{p=0}^{n} \left|b_p - b'_p\right| = \#\left\{p \in \{0,1,\ldots,n\} : b_p \neq b'_p\right\}.$$

In particular, when $|b| = 0$, we denote $(\varnothing) := (0,\ldots,0)$.

For two coalescence indicator b and b' in $\{0,1\}^{n+1}$, we say $b \leq b'$ if for any $0 \leq p \leq n$, we have $b_p \leq b'_p$. More over, if $b \leq b'$, and $b \neq b'$, we say $b < b'$. We also consider the set of coalescence indicators $\mathcal{S}_n(b)$, $\mathcal{S}_n^{>}(b)$ and $\mathring{\mathcal{S}}_n(b)$ defined as follows:

- $\mathcal{S}_n(b) := \left\{b' \in \{0,1\}^{n+1} \mid b_n = b'_n\right\}$;
- $\mathcal{S}_n^{>}(b) := \left\{b' \in \{0,1\}^{n+1} \mid b' > b \text{ and } b_n = b'_n\right\}$;
- $\mathring{\mathcal{S}}_n(b) := \mathcal{S}_n(b) \backslash \{b\}$.

Let us consider the following auxiliary coalescent Feynman-Kac kernels:

$$\begin{cases} \mathrm{Q}_n^{0|0} := \mathrm{Q}_n^{\dagger,0} \\ \mathrm{Q}_n^{1|0} := \widetilde{\mathrm{Q}}_n^{\dagger,1} \\ \mathrm{Q}_n^{0|1} := \check{\mathrm{Q}}_n^{\dagger,1} \\ \mathrm{Q}_n^{1|1} := \mathrm{Q}_n^{\dagger,1} \end{cases} \quad \text{and} \quad \begin{cases} \mathrm{Q}_{n,(N)}^{0|0} := \mathrm{Q}_n^{\dagger,0} \\ \mathrm{Q}_{n,(N)}^{1|0} := \widetilde{\mathrm{Q}}_{n,(N)}^{\dagger,1} \\ \mathrm{Q}_{n,(N)}^{0|1} := \check{\mathrm{Q}}_{n,(N)}^{\dagger,1} \\ \mathrm{Q}_{n,(N)}^{1|1} := \mathrm{Q}_n^{\dagger,1} \end{cases} \tag{22}$$

We remark that these kernels are constructed in respect of the decomposition of the partial group structure w.r.t. the composition "+", namely, at level n, we have

$$\underbrace{0 : \mathrm{Q}_{n,(N)}^{\ddagger,0}}_{0|0:\mathrm{Q}_n^{\dagger,0}+\widetilde{\mathrm{Q}}_{n,(N)}^{\dagger,1}:1|0} \quad \text{and} \quad \underbrace{1 : \mathrm{Q}_{n,(N)}^{\ddagger,1},}_{1|1:\mathrm{Q}_n^{\dagger,1}+\check{\mathrm{Q}}_{n,(N)}^{\dagger,1}:0|1} \tag{23}$$

whence the four cases which correspond to the four possible choices when the partial semigroup structure is passing through the coalescent Feynman-Kac kernel $\mathrm{Q}_{n,(N)}^{\ddagger,b_{n-1}}$. With this in mind, we define some auxiliary coalescent tree-based measures that are useful for the decomposition mentioned above.

Definition B.3. *For any $n' \geq n \geq 1$, $N \geq 2$ and for any coalescence indicators $b, b' \in \{0,1\}^{n'+1}$, we define the signed finite measures $\Gamma_n^{b'|b}$ and $\Gamma_{n,N}^{b'|b}$ respectively by*

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma_n^{b'|b}(F) := \eta_0^{\otimes 2}\, \mathrm{Q}_1^{b_0'|b_0}\, \mathrm{Q}_2^{b_1'|b_1} \cdots \mathrm{Q}_n^{b_{n-1}'|b_{n-1}} C_{b_n}(F),$$

and

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma_{n,(N)}^{b'|b}(F) := \eta_0^{\otimes 2}\, \mathrm{Q}_{1,(N)}^{b_0'|b_0}\, \mathrm{Q}_{2,(N)}^{b_1'|b_1} \cdots \mathrm{Q}_{n,(N)}^{b_{n-1}'|b_{n-1}} C_{b_n}(F),$$

with the convention

$$\Gamma_0^{b'|b}(F) = \Gamma_{0,(N)}^{b'|b}(F) := \eta_0^{\otimes 2} C_{b_0}.$$

We finally define all the coalescent tree-based measures as a generalization of the work in [CDMG11]. The following proposition is a direct consequence of the introduction of

$$\widetilde{\mathrm{Q}}_{n,(N)}^{\dagger,b_{n-1}} = \frac{1}{N-1}\widetilde{\mathrm{Q}}_n^{\dagger,b_{n-1}} \quad \text{and} \quad \check{\mathrm{Q}}_{n,(N)}^{\dagger,b_{n-1}} = \frac{1}{N-1}\check{\mathrm{Q}}_n^{\dagger,b_{n-1}}.$$

Proposition B.2. *For any $n' \geq n \geq 1$, $N \geq 2$, for any coalescence indicators $b, b' \in \{0,1\}^{n'+1}$ and for any test function $F \in \mathcal{B}_b(E_n^2)$, we have the following equalities:*

(i) $\widetilde{\Gamma}_{n,(N)}^{\dagger,b}(F) = \left(\frac{1}{N-1}\right)^{|b|} \widetilde{\Gamma}_n^{\dagger,b}(F)$;

(ii) $\forall b' \in \mathcal{S}_n^{>}(b)$, $\Gamma_{n,(N)}^{b'|b}(F) = \left(\frac{1}{N-1}\right)^{|b'|-|b|} \widetilde{\Gamma}_n^{b'|b}(F)$;

(iii) $\Gamma_{n,(N)}^{b'|b}(F) = \left(\frac{1}{N-1}\right)^{|b-b'|} \widetilde{\Gamma}_n^{b'|b}(F)$.

Since we have all the necessary ingredients at hand, we provide the most important result of this section. For all test function $F \in \mathcal{B}_b(E_n^2)$, we have

$$\begin{aligned}\Gamma^{\ddagger,b}_{n,(N)}(F) = &\sum_{b' \in \mathcal{S}_n(b)} \Gamma^{b'|b}_{n,(N)}(F)\\ =&\Gamma^{b|b}_{n,(N)}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n(b)} \Gamma^{b'|b}_{n,(N)}(F)\\ =&\Gamma^{b|b}_{n}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n(b)} \left(\frac{1}{N-1}\right)^{|b'-b|} \Gamma^{b'|b}_{n}(F)\\ =&\Gamma^{\dagger,b}_{n}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n(b)} \left(\frac{1}{N-1}\right)^{|b'-b|} \Gamma^{b'|b}_{n}(F).\end{aligned} \tag{24}$$

In particular, for the case where $b = (\emptyset)$, we have

$$\begin{aligned}\Gamma^{\ddagger,(\emptyset)}_{n,(N)}(F) = &\sum_{b' \in \mathcal{S}_n(b)} \Gamma^{b'|b}_{n,(N)}(F)\\ =&\Gamma^{(\emptyset)|(\emptyset)}_{n,(N)}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n((\emptyset))} \Gamma^{b'|(\emptyset)}_{n,(N)}(F)\\ =&\Gamma^{\dagger,(\emptyset)}_{n}(F) + \sum_{b' \in \mathcal{S}^{>}_n((\emptyset))} \left(\frac{1}{N-1}\right)^{|b'|} \widetilde{\Gamma}^{\dagger,b'}_{n}(F)\\ =&\Gamma^{(\emptyset)}_{n}(F) + \frac{1}{N-1}\sum_{p=0}^{n-1} \widetilde{\Gamma}^{\dagger,(p)}_{n}(F) + \sum_{b' \in \mathcal{S}^{>}_n(b), |b'|\geq 2} \left(\frac{1}{N-1}\right)^{|b'|} \widetilde{\Gamma}^{\dagger,b'}_{n}(F).\end{aligned} \tag{25}$$

Taking into account that all the coalescent tree-based meaures are finite signed measures, the calculations above give the following proposition.

Proposition B.3. *For any $n' \geq n \geq 1$ and for any coalescent indicator $b \in \{0,1\}^{n'+1}$, we have*

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma^{\ddagger,b}_{n,(N)}(F) - \Gamma^{\dagger,b}_{n}(F) = \mathcal{O}\left(\frac{1}{N}\right).$$

In particular, we have

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma^{\ddagger,(\emptyset)}_{n,(N)}(F) - \Gamma^{(\emptyset)}_{n}(F) - \frac{1}{N-1}\sum_{p=0}^{n-1} \widetilde{\Gamma}^{\dagger,(p)}_{n}(F) = \mathcal{O}\left(\frac{1}{N^2}\right).$$

Remark. Above lies part of the reason why we have inhomogeneity in the notation w.r.t. "‡" and "†" in Definition 3.2. In fact, the strategy to prove the consistency given in Theorem B.3 is divided into two steps, and the latter is done by Proposition B.3 above:

- $\Gamma^{\ddagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} - \Gamma^{\ddagger,b}_{n,(N)}(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right)$;
- $\Gamma^{\ddagger,b}_{n,(N)}(F) - \Gamma^{\dagger,b}_{n}(F) = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right)$.

B.4 Coalescent tree occupation measures

In this section, we introduce the particle approximations of the coalescent tree-based measures discussed in the last section. To do this, we need to adopt some notation from [DG19](Chapter 2). The readers are also referred to Appendix A.2 of [DG19](Chapter 2) to find more intuitions and the connection between the construction of these particle approximations and the Particle Markov Chain Monte Carlo methods. Due to the fact that the underlying resampling scheme is changed, the analysis also becomes more challenging. Fortunately, the basic idea remains the same: we exploit the information encoded in the genealogy, and in addition, the information encoded in the survival history, to approximate the coalescent Feynman-Kac measures. The key idea is to collect all the corresponding coalescent tree-type forms illustrated in Figure 2, and the coalescent tree occupation measures are constructed as weighted empirical terminal measures of these particle blocks. The intuition of this procedure remains identical to the previous work in [DG19](Chapter 2). More precisely, the major difference is about the weights mentioned above, which correspond to the potential function of the original IPS in many-body Feynman-Kac models (cf. $\mathcal{G}_p^{(q)}(\boldsymbol{x_p})$ defined in A.1 of [DG19](Section 2.5.1)). Under asymmetric resampling scheme, it is necessary to consider the influence of the survival history. Thus, it is expected that the constructions become more sophisticated. Another remark is on the measure $\widetilde{\Gamma}_n^{\dagger,b}$: since the related coalescent Feynman-Kac kernels do not conserve the coalescence structure, its particle approximation also turns out to be a little bit different. Recall that $\lambda_p^b(\tilde{a}_p^{[2]}, \ell_p^{[2]}) \in \{0,1\}$ is an indicator function defined by

$$\lambda_p^b(\tilde{a}_p^{[2]}, \ell_p^{[2]}) := \mathbf{1}_{\{b_p=0\}}\mathbf{1}_{\{\tilde{a}_p^1=\ell_p^1\neq\tilde{a}_p^2=\ell_p^2\}} + \mathbf{1}_{\{b_p=1\}}\mathbf{1}_{\{\tilde{a}_p^1=\ell_p^1=\tilde{a}_p^2\neq\ell_p^2\}}.$$

Definition B.4 (Coalescent tree occupation measures). *For any $n' \geq n \geq 0$ and for any coalescence indicator $b \in \{0,1\}^{n'+1}$, the random measure $\bar{\Gamma}_{n,N}^b$ is defined by*

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \bar{\Gamma}_{n,N}^b(F) := \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_{0:n}^{[2]} \in ((N)^2)^{\times(n+1)}} \left\{ \prod_{p=0}^{n-1} \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\} C_{b_n}(F)(X_n^{\ell_n^{[2]}}).$$

The next theorem is brought from Proposition 4.2 of [DG19](Chapter 2), and the proof is a slightly modified version of the one given in Section 4.5 of [DG19] (Section 2.4.5). It is one of the most important ingredients that connect the coalescent tree occupation measures and non-asymptotic variance of Feynman-Kac IPS. It provides information on the combinatorial structure of the IPS in regard to the coalescent tree occupation measures $\bar{\Gamma}_{n,N}^b$, which does not depend on the resampling scheme and regularity assumptions whilst the IPS is well-defined. Namely, it reveals the essential combinatorial properties that apply to all genealogy tree-based particle systems when each particle has only one parent. This combinatorial property, in particular, is also valid in continuous-time settings and/or in the frameworks with even more complex resampling schemes. A possible variant of the current work is to replace multinomial resampling for the non-survival particles with more advanced methods, such as residual resampling, stratified resampling and systematic resampling, etc. In this article, it is the bridge between the asymptotic variance, non-asymptotic variance, and eventually, the construction of our variance estimators. The proof is given in Section C.6.

Theorem B.1. *For any test function $F \in \mathcal{B}_b(E_n^2)$, we have the following decompositions.*

$$(\eta_n^N)^{\otimes 2}(F)\mathbf{1}_{\tau_N \geq n} = \sum_{b \in \{0,1\}^{n+1}} \left(\frac{N-1}{N}\right)^{n+1-|b|} \left(\frac{1}{N}\right)^{|b|} \bar{\Gamma}^b_{n,N}(F)\mathbf{1}_{\tau_N \geq n}, \quad a.s.$$

and

$$(\gamma_n^N)^{\otimes 2}(F)\mathbf{1}_{\tau_N \geq n} = \sum_{b \in \{0,1\}^{n+1}} \left(\frac{N-1}{N}\right)^{n+1-|b|} \left(\frac{1}{N}\right)^{|b|} {}^{\ddagger}\Gamma^b_{n,N}(F)\mathbf{1}_{\tau_N \geq n}. \quad a.s.$$

Below we list the most important lack-of-bias and convergence results of the coalecent tree occupation measures, serving as the particle approximations of the coalescent Feynman-Kac measures. The proofs are divided into several technical results: they are direct consequences of the combination of Lemma C.1, Lemma C.2, Proposition C.2, Proposition C.3 and Proposition B.3, with some standard manipulations of bounded i.i.d. random variables for the case $n = 0$, which is provided in Lemma C.4. The only remark is that for any coalescent indicator b and for any coalescent Feynman-Kac kernel, e.g., $\widetilde{\mathbf{Q}}_n^{\dagger,b}$, we have, by definition,

$$\forall \varphi, \psi \in \mathcal{B}_b(E_n), \quad \exists f, g \in \mathcal{B}_b(E_{n-1}) \quad s.t. \quad \widetilde{\mathbf{Q}}_n^{\dagger,b}(\varphi \otimes \psi) \equiv f \otimes g.$$

The property above also holds for $\mathbf{Q}_n^{\ddagger,b}$.

Theorem B.2 (Unbiasedness). *Assume symmetric resampling, that is $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 1$, then, we have*

$$\mathbf{E}\left[\Gamma^{(\emptyset)}_{n,N}(F)\mathbf{1}_{\tau_N \geq n}\right] = \Gamma^{(\emptyset)}_n(F) = \gamma_n^{\otimes 2}(F).$$

Theorem B.3 (Consistency). *For any coalescence indicator $b \in \{0,1\}^{n+1}$ and any test function $F \in \mathcal{B}_b(E_n)^{\otimes 2}$, we have*

(i) $\Gamma^{\ddagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} - \Gamma^{\dagger,b}_n(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right)$;

(ii) $\widetilde{\Gamma}^{\dagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} - \widetilde{\Gamma}^{\dagger,b}_n(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right)$.

Remark. The notation

$$\Gamma^{\ddagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} - \Gamma^{\dagger,b}_n(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right),$$

means that

$$\left\|\Gamma^{\ddagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} - \Gamma^{\dagger,b}_n(F)\right\|_{\mathbb{L}^1} = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right).$$

The reader is referred to the beginning of Section C for details.

By linearity of signed measures, we have, on the event $\{\tau_N \geq n\}$,

$$\begin{aligned}&\Gamma^{\ddagger,b}_{n,N}\left(\left[f - \eta_n^N(f)\right]^{\otimes 2}\right)\\ =&\Gamma^{\ddagger,b}_{n,N}(f^{\otimes 2}) - \eta_n^N(f)\left(\Gamma^{\ddagger,b}_{n,N}(1 \otimes f) + \Gamma^{\ddagger,b}_{n,N}(f \otimes 1)\right) + \eta_n^N(f)^2\Gamma^{\ddagger,b}_{n,N}\left(1^{\otimes 2}\right),\end{aligned}$$

as well as

$$
\begin{aligned}
&\widetilde{\Gamma}^{\dagger,b}_{n,N}\left(\left[f-\eta^N_n(f)\right]^{\otimes 2}\right)\\
=&\widetilde{\Gamma}^{\dagger,b}_{n,N}(f^{\otimes 2})-\eta^N_n(f)\left(\widetilde{\Gamma}^{\dagger,b}_{n,N}(1\otimes f)+\widetilde{\Gamma}^{\dagger,b}_{n,N}(f\otimes 1)\right)+\eta^N_n(f)^2\widetilde{\Gamma}^{\dagger,b}_{n,N}(1^{\otimes 2})\,.
\end{aligned}
$$

Then, Theorem 2.1 gives

$$\eta^N_n(f)\mathbf{1}_{\tau_N\geq n}-\eta_n(f)=o_{\mathbf{p}}(1)\,,$$

and

$$1/\gamma^N_n(f)^2\mathbf{1}_{\tau_N\geq n}-1/\gamma_n(f)^2=o_{\mathbf{p}}(1)\,.$$

Finally, combined to Proposition C.7 and Proposition C.8, we have the following corollary.

Corollary B.3.1. *For any coalescence indicator $b\in\{0,1\}^{n+1}$ and any test function $f\in\mathcal{B}_b(E_n)$, we have*

(i) $\Gamma^{\ddagger,b}_{n,N}\left(\left[f-\eta^N_n(f)\right]^{\otimes 2}\right)\mathbf{1}_{\tau_N\geq n}/\gamma^N_n(1)^2-\Gamma^{\dagger,b}_n\left(\left[f-\eta_n(f)\right]^{\otimes 2}\right)/\gamma_n(1)^2=\mathcal{O}_{\mathbf{p}}\left(\frac{1}{\sqrt{N}}\right)$;

(ii) $\widetilde{\Gamma}^{\dagger,b}_{n,N}\left(\left[f-\eta^N_n(f)\right]^{\otimes 2}\right)\mathbf{1}_{\tau_N\geq n}/\gamma^N_n(1)^2-\widetilde{\Gamma}^{\dagger,b}_n\left(\left[f-\eta_n(f)\right]^{\otimes 2}\right)/\gamma_n(1)^2=\mathcal{O}_{\mathbf{p}}\left(\frac{1}{\sqrt{N}}\right)$.

B.5 Feynman-Kac measures flow in a random environment.

In this section, we provide another interpretation on the construction of the coalescent tree occupation measures. One of the main message of [DG19](Chapter 2) is to provide some intuition on the construction of the coalescent tree occupation measures using the many-body Feynman-Kac models introduced in [DMKP16], by considering a Gibbs sampler w.r.t. the *original IPS* and *coupled particle block* on a sophisticated path space. Then, we define the event that traps the desired coalescent particle block and eventually, we construct the estimator given in Definition B.4. This methodology gives the foundation of the present work: Definition 3.3 is also obtained by this procedure, though it is not discussed in detail as in [DG19](Chapter 2). Now, let us look at this family of random measures from a different angle. We begin with some basic observations. To facilitate the writings, let us fix a time horizon $T\in\mathbf{N}^*$. The following discussion is valid on the event $\{\tau_N\geq T\}$ and $0\leq n\leq T$. Given $\mathcal{W}^N_T$ (cf. Section C.1) and fixing $\ell^{[2]}_{n-1}\in(N)^2$, we have

$$\forall b\in\{0,1\}^{T+1},\ \forall n\in[T],\qquad \sum_{\ell^{[2]}_n\in(N)^2}\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})=1.$$

Therefore, let us consider the state space $(N)^2$, the matrix of size $N(N-1)\times N(N-1)$, with some prefixed ordering rule on the set $(N)^2$, is denoted by

$$\left(\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})\right)_{\left(\ell^{[2]}_{n-1},\ell^{[2]}_n\right)\in((N)^2)^{\times 2}},\tag{26}$$

which can then be regarded as a random transition matrix, with

$$\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})$$

denoting the probability of transition from the site $\ell^{[2]}_{n-1}$ to the site $\ell^{[2]}_n$. For the general theory regarding to the Markov chain in a random environment, the readers are referred to [Cog80]. Returning to the definition of coalescent tree occupation measures $\Gamma^{\ddagger,b}_{n,N}$, we can find a similar semigroup structure: the initial distribution on the state space $(N)^2$ is $m^{\odot 2}([N])$ and the potential function on the site $\ell^{[2]}_{n-1}$ is $\mathbf{G}^{\ddagger}_{\mathbf{n}}(\mathbf{X}_{\mathbf{n-1}})$, which is a constant function given $\mathcal{W}^N_T$. Therefore, by denoting

$$\mathbf{H}^{\ddagger,b}_n[\ell^{[2]}_{n-1},\ell^{[2]}_n] := \mathbf{G}^{\ddagger,b}_{\mathbf{n}}(\mathbf{X}_{\mathbf{n-1}})\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1}),$$

and by respecting the composition law of the matrix multiplication, we define

$$\left(\mathbf{H}^{\ddagger,b}_n \cdot \mathbf{H}^{\ddagger,b}_{n+1}\right)[\ell^{[2]}_{n-1},\ell^{[2]}_{n+1}] := \sum_{\ell^{[2]}_n \in (N)^2} \mathbf{H}^{\ddagger,b}_n[\ell^{[2]}_{n-1},\ell^{[2]}_n] \times \mathbf{H}^{\ddagger,b}_{n+1}[\ell^{[2]}_n,\ell^{[2]}_{n+1}].$$

In addition, for any random measure Λ_{n-1} on the state space $(N)^2$, we define

$$\left(\Lambda_{n-1} \cdot \mathbf{H}^{\ddagger,b}_n\right)[\ell^{[2]}_n] := \sum_{\ell^{[2]}_n \in (N)^2} \Lambda_{n-1}[\ell^{[2]}_{n-1}] \times \mathbf{H}^{\ddagger,b}_n[\ell^{[2]}_{n-1},\ell^{[2]}_n],$$

and for any random test function $\mathbf{F}$ on the state space $(N)^2$, we define

$$\mathbf{H}^{\ddagger,b}_n(\mathbf{F})[\ell^{[2]}_{n-1}] := \sum_{\ell^{[2]}_n \in (N)^2} \mathbf{H}^{\ddagger,b}_n[\ell^{[2]}_{n-1},\ell^{[2]}_n] \times \mathbf{F}[\ell^{[2]}_n].$$

In particular, we denote

$$\Lambda_n(\mathbf{F}) := \sum_{\ell^{[2]}_n \in (N)^2} \Lambda_n[\ell^{[2]}_n] \times \mathbf{F}[\ell^{[2]}_n].$$

Obviously, these composition law does not depend on the prefixed ordering rule on the set $(N)^2$ since the sum "+" of the random variables is commutative. Now, we are able to give an alternative representation of $\Gamma^{\ddagger,b}_{n,N}$. We define

$$\Lambda^{\ddagger,b}_n[\ell^{[2]}_n] := m^{\odot 2}([N]) \cdot \mathbf{H}^{\ddagger,b}_1 \cdot \mathbf{H}^{\ddagger,b}_2 \cdots \mathbf{H}^{\ddagger,b}_n[\ell^{[2]}_n], \tag{27}$$

with the convention $\Lambda^{\ddagger,b}_0 := m^{\odot 2}([N])$. Accordingly, the random test function $\mathbf{F}^b_n$ is defined by

$$\mathbf{F}^b_n[\ell^{[2]}_n] := C_{b_n}(F)(X^{\ell^{[2]}_n}_n).$$

Consequently, we have

$$\Lambda^{\ddagger,b}_n(\mathbf{F}^b_n) = \Gamma^{\ddagger,b}_{n,N}(F). \tag{28}$$

The random measure notation above is frequently used in the proof of technical results, an explicit form can be found later in (56). Similarly, we denote

$$\widetilde{\mathbf{H}}^{\dagger,b}_n[\ell^{[2]}_{n-1},\ell^{[2]}_n] := \widetilde{\mathbf{G}}^{\dagger,b}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n},\mathbf{B}_{\mathbf{n-1}},\mathbf{X}_{\mathbf{n-1}})\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1}),$$

and

$$\widetilde{\Lambda}^{\dagger,b}_n[\ell^{[2]}_n] := m^{\odot 2}([N]) \cdot \widetilde{\mathbf{H}}^{\dagger,b}_1 \cdot \widetilde{\mathbf{H}}^{\dagger,b}_2 \cdots \widetilde{\mathbf{H}}^{\dagger,b}_n[\ell^{[2]}_n],$$

with the convention $\widetilde{\Lambda}^{\dagger,b}_0 := m^{\odot 2}([N])$. Apart from the fact that this writing guided and simplified some of the proofs of the technical results, the main motivation is to provide a decomposition result similart to the one in (25), which is essentially due to the partial $\mathcal{R}$-algebra homomorphism. We define $\mathbf{H}_n^0 := \widetilde{\mathbf{H}}_n^{\dagger,0}$. By definition, we have

$$\mathbf{H}_n^0[\ell_{n-1}^{[2]}, \ell_n^{[2]}] + \frac{1}{N-1}\widetilde{\mathbf{H}}_n^{\dagger,1}[\ell_{n-1}^{[2]}, \ell_n^{[2]}] = \mathbf{H}_n^{\ddagger,b}[\ell_{n-1}^{[2]}, \ell_n^{[2]}],$$

which yields, for the associated random matrix,

$$\mathbf{H}_n^0 + \frac{1}{N-1}\widetilde{\mathbf{H}}_n^{\dagger,1} = \mathbf{H}_n^{\ddagger,b}.$$

Therefore, we have the following decomposition:

$$\begin{aligned}\Lambda_n^{\ddagger,(\varnothing)}(\mathbf{F}_\mathbf{n}^b) =& \Lambda_n^{(\varnothing)}(\mathbf{F}_\mathbf{n}^b) + \sum_{b'\in\mathcal{S}_n^{>}((\varnothing))}\left(\frac{1}{N-1}\right)^{|b'|}\widetilde{\Lambda}_n^{\dagger,b'}(\mathbf{F}_\mathbf{n}^b)\\ =& \Lambda_n^{(\varnothing)}(\mathbf{F}_\mathbf{n}^b) + \frac{1}{N-1}\sum_{p=0}^{n-1}\widetilde{\Lambda}_n^{\dagger,(p)}(\mathbf{F}_\mathbf{n}^b) + \sum_{b'\in\mathcal{S}_n^{>}(b),|b'|\geq 2}\left(\frac{1}{N-1}\right)^{|b'|}\widetilde{\Lambda}_n^{\dagger,b'}(\mathbf{F}_\mathbf{n}^b),\end{aligned}$$

which is equivalent to

$$\begin{aligned}\Gamma_{n,N}^{\ddagger,(\varnothing)}(F) =& \Gamma_{n,N}^{(\varnothing)}(F) + \sum_{b'\in\mathcal{S}_n^{>}((\varnothing))}\left(\frac{1}{N-1}\right)^{|b'|}\widetilde{\Gamma}_{n,N}^{\dagger,b'}(F)\\ =& \Gamma_{n,N}^{(\varnothing)}(F) + \frac{1}{N-1}\sum_{p=0}^{n-1}\widetilde{\Gamma}_{n,N}^{\dagger,(p)}(F) + \sum_{b'\in\mathcal{S}_n^{>}(b),|b'|\geq 2}\left(\frac{1}{N-1}\right)^{|b'|}\widetilde{\Gamma}_{n,N}^{\dagger,b'}(F).\end{aligned} \tag{29}$$

Thanks to Proposition C.8 and the decomposition (29) above, we have the following proposition.

Proposition B.4. *For any $n' \geq n \geq 1$ and for any coalescent indicator $b \in \{0,1\}^{n'+1}$, we have*

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \left(\Gamma_{n,N}^{\ddagger,(\varnothing)}(F) - \Gamma_{n,N}^{(\varnothing)}(F) - \frac{1}{N-1}\sum_{p=0}^{n-1}\widetilde{\Gamma}_{n,N}^{\dagger,(p)}(F)\right)\mathbf{1}_{\tau_N\geq n} = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{N^2}\right).$$

B.6 Efficient estimator of $\widetilde{\Gamma}_n^{\dagger,(p)}$

As is mentioned before, we failed to provide an $\mathcal{O}(nN)$ time complexity algorithm to compute the term by term variance estimator and the non-asymptotic variance estimator provided in the previous sections. Therefore, we give a new asymptotic variance estimators that can be computed with $\mathcal{O}(nN)$ time complexity. The idea is to construct some new coalescent tree occupation measures that are very "close" to $\widetilde{\Gamma}_{n,N}^{\dagger,(p)}$, which is easier to obtain by some numerical techniques to reduce the computational costs. First, let us define a new sequence of random matrix on the event $\{\tau_N \geq n\}$. For any $N > 1$, we consider

$$\begin{cases}\widetilde{\mathbf{H}}_n^{\ddagger,0} := \mathbf{H}_n^{\ddagger,0} = \widetilde{\mathbf{H}}_n^{\dagger,0} + \frac{1}{N-1}\widetilde{\mathbf{H}}_n^{\dagger,1};\\ \widetilde{\mathbf{H}}_n^{\ddagger,1} := \widetilde{\mathbf{H}}_n^{\dagger,1}.\end{cases}$$

Next, using the same semigroup property of these random matrix and random test function as in (27) and (28), by consider the initial distribution $m^{\odot 2}([N])$, we define the new coalescent tree occupation measure $\widetilde{\Gamma}^{\ddagger,b}_{n,N}$ for each $b \in \{0,1\}^{n+1}$ by

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \widetilde{\Gamma}^{\ddagger,b}_{n,N}(F) := \widetilde{\Lambda}^{\ddagger,b}_n(\mathbf{F}_n) = m^{\odot 2}([N]) \cdot \widetilde{\mathbf{H}}^{\ddagger,b}_1 \cdot \widetilde{\mathbf{H}}^{\ddagger,b}_2 \cdots \widetilde{\mathbf{H}}^{\ddagger,b}_n(\mathbf{F}_n),$$

with

$$\mathbf{F}_n[\ell_n^{[2]}] := F(X_n^{\ell_n^{[2]}}).$$

Proposition B.5. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and for any $p \in \{0,1,\dots,n-1\}$, we have*

$$\left(\widetilde{\Gamma}^{\ddagger,(p)}_{n,N}(F) - \widetilde{\Gamma}^{\dagger,(p)}_{n,N}(F)\right) \mathbf{1}_{\tau_N \geq n} = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{N}\right).$$

Finally, without loss of generality, we explain why the estimator given in (14) can be computed with $\mathcal{O}(nN)$ time complexity. In fact, the first part of the estimator

$$\sum_{p=0}^{n} \left(\Gamma^{\ddagger,(p)}_{n,N}(f^{\otimes 2}) - \Gamma^{\ddagger,(\emptyset)}_{n,N}(f^{\otimes 2})\right)$$

can be approximated by the variance estimator $NV_n^N(f)\gamma_n^N(1)^2$ proposed by Lee & Whiteley [LW18]. Hence, an $\mathcal{O}(nN)$ algorithm is therefore available. It is then sufficient to provide an $\mathcal{O}(nN)$ algorithm to compute $\widetilde{\Gamma}^{\ddagger,(p)}_{n,N}(f^{\otimes 2})$. This is possible due to the homogeneity of the potential function $\mathbf{G}^{\ddagger}_{\mathbf{n}}(\mathbf{X_n})$ w.r.t. the different indices $\ell_n^{[2]}$ and the following technical lemma.

Lemma B.1. *Let $(\mathcal{R}, +, \star)$ be a ring, and $(\mathcal{E}_i)_{i \in [k]}$ be a disjoint partition of $[N]$ for some $k \geq 1$, then for any sequence $(a_i)_{i \in [N]}$ composed by elements of $\mathcal{R}$, we have the following equality:*

$$\sum_{\substack{i \in \mathcal{E}_p, j \in \mathcal{E}_q \\ 0 \leq p \neq q \leq k}} a_i \star a_j = \left(\sum_{s=1}^{N} a_s\right) \star \left(\sum_{s=1}^{N} a_s\right) - \sum_{r=1}^{k} \sum_{\ell \in \mathcal{E}_r} a_\ell \star a_\ell.$$

In our case, the partition $(\mathcal{E}_i)_{i \in [k]}$ is the divided by the ancestor indices of the particles of level n. The ring $\mathcal{R}$ is $\mathbf{R}^3$ and the "$\star$" product refers to the operation

$$\mathbf{R}^3 \times \mathbf{R}^3 \ni (x,y,z) \star (x',y',z') \mapsto (x_1 \times x', y \times z', z \times y') \in \mathbf{R}^3,$$

which represents an intermediate step in the Algorithm 4, whose final output is

$$\langle (x,y,z), (x',y',z') \rangle_\star := x_1 \times x' + y \times z' + z \times y'.$$

This is useful in calculating the term

$$\widetilde{\Lambda}^{\ddagger,(p)}_n[\ell_n^{[2]}] \mathbf{F}^b_n(\ell_n^{[2]}).$$

In fact, for any test function $F \in \mathcal{B}_b(E_n)^{\otimes 2}$, the term above can be a.s. reformulated as

$$\exists (a_\ell)_{\ell \in [N]} \in \left(\mathbf{R}^3\right)^N, \quad s.t. \quad \forall \ell_n^{[2]} \in (N)^2,\ \widetilde{\Lambda}^{\ddagger,(p)}_n[\ell_n^{[2]}] \mathbf{F}^b_n(\ell_n^{[2]}) = \left\langle a_{\ell_n^1}, a_{\ell_n^2} \right\rangle_\star.$$

Hence, by applying Lemma B.1, one can therefore compute $\widetilde{\Gamma}^{\ddagger,(p)}_{n,N}(f^{\otimes 2})$ with $\mathcal{O}(nN)$ time complexity. The details can be found in Algorithm 5, and the design of the Algorithm 6 is similar. We remark that when the homogeneity w.r.t. the potential functions for the Feynman-Kac chain in a random environment is missing for more than finite levels w.r.t. n and N, we are not able to construct the ring homomorphism discussed above. This is the intrinsic reason why we failed to apply this technique to reduce the time complexity of the non-asymptotic variance estimator $V_n^N(f)$. In the Algorithm 7, the corresponding term

$$m^{\odot 2}([N])\mathbf{H}^{(\emptyset)}_{0,n}[\ell_n^{[2]}]\mathbf{F}_n^b(\ell_n^{[2]})$$

is therefore calculated by violently searching all the possible choices. This is why the computation is of time complexity $\mathcal{O}(nN^2)$.

C Proofs

In this section, we list all the proofs in the present work. Some notation are gathered in Section C.1, such as the formal definitions of the filtrations frequently used in the proofs, along with the most important martingale decompositions. A little plan on the organization of the technical results is also provided. In order to facilitate the writing, the stochastic bounds introduced in [Jan11] are intensely involved in our technical results. More precisely, we use frequently the notation $\mathcal{O}_{\mathbf{p}}$, $\mathcal{O}_{\mathbb{L}^p}$ and $\mathcal{O}_{a.s.}$. Let $(a_N; N \in \mathbf{N})$ be a sequence of natural numbers, where N represents the number of particles in the IPS. The notation

$$X_N = \mathcal{O}_{\mathbf{p}}(a_N)$$

means that the sequence $(X_N/a_N; N \in \mathbf{N})$ is tight, namely, for any $\epsilon > 0$, there exists $0 < M_\epsilon < +\infty$, such that

$$\limsup_{N\in\mathbf{N}} \mathbf{P}\left(|X_N/a_N| > M_\epsilon\right) < \epsilon.$$

In particular, $o_{\mathbf{p}}(1)$ means convergence to 0 in probability. The notation

$$X_N = \mathcal{O}_{\mathbb{L}^p}(a_N)$$

means that the random variable X_N/a_N is uniformly bounded in $\mathbb{L}^p$-norm w.r.t. N. The notation

$$X_N = \mathcal{O}_{a.s.}(a_N)$$

indicates that

$$\mathbf{P}\left(\left\{\omega \in \Omega : \sup_{N\in\mathbf{N}} |X_N(\omega)/a_N| < +\infty\right\}\right) = 1.$$

Thanks to Cauchy-Schwartz inequality and Markov's inequality, we have

$$X_N = \mathcal{O}_{a.s.}(a_N) \implies X_N = \mathcal{O}_{\mathbb{L}^2}(a_N) \implies X_N = \mathcal{O}_{\mathbb{L}^1}(a_N) \implies X_N = \mathcal{O}_{\mathbf{p}}(a_N).$$

We also remark that for all these 4 types of stochastic bounds, they are weaker than the corresponding convergence. For example, if

$$X_N/a_N \xrightarrow[N\to\infty]{\mathbb{L}^1/\mathbf{P}} \text{Const.} < +\infty,$$

one also has

$$X_N = \mathcal{O}_{\mathbb{L}^1/\mathbf{p}}(a_N).$$

C.1 Martingales

We present some important martingales encountered in the analysis of SMC framework. They are crucial to some of the technical resutls in this article. We also hope that the similar construction may inspire the future work in different settings. Before going into details, let us define some filtrations associated to the Feynman-Kac IPS.

Filtrations. $(\mathcal{F}_n^N)_{n\geq 0}$ denotes the filtration that consists the information of the values of particles. More precisely,

$$\mathcal{F}_{-1}^N := \{\emptyset, \Omega\} \quad \text{and} \quad \forall n \geq 0, \quad \mathcal{F}_n^N := \sigma(\mathbf{X_0}, \ldots, \mathbf{X_n}).$$

If we only add one particle at each step, a more refined filtration $(\mathcal{E}_k^N)_{k\geq 0}$ can be defined by

$$\forall k \in [(n+1)N], \quad \mathcal{E}_k^N = \mathcal{F}_{p_k-1}^N \vee \sigma(X_{p_k}^1, \cdots, X_{p_k}^{i_k}),$$

where for any $k \in [(n+1)N]$, we adopt the notaition

$$p_k := \left\lfloor \frac{k}{N} \right\rfloor \qquad \text{and} \qquad i_k := k - p_k \times N.$$

Next, $(\mathcal{G}_n^N)_{n\geq 0}$ denotes the filtration that contains the genealogy of IPS, which is defined by

$$\forall n \in \{-1, 0\}, \quad \mathcal{G}_n^N := \mathcal{F}_n^N \quad \text{and} \quad \forall n \geq 1, \quad \mathcal{G}_n^N := \mathcal{F}_n^N \vee \sigma(\mathbf{A_0}, \ldots, \mathbf{A_{n-1}}).$$

Finally, the filtration that contains all the information including survival history of the particle system are denoted by $(\mathcal{W}_n^N)_{n\geq 0}$, namely,

$$\forall n \in \{-1, 0\}, \quad \mathcal{W}_n^N := \mathcal{F}_n^N \quad \text{and} \quad \forall n \geq 1, \quad \mathcal{W}_n^N := \mathcal{G}_n^N \vee \sigma(\mathbf{B_0}, \ldots, \mathbf{B_{n-1}}).$$

Moreover, as is used several times in some technical results, we also consider an updatated filtration $(\overline{\mathcal{W}}_n^N)_{n\geq 0}$ defined by

$$\overline{\mathcal{W}}_n^N := \mathcal{W}_n^N \vee \sigma(\mathbf{B}_n).$$

Proposition C.1. *For any test function $f \in \mathcal{B}_b(n)$, we define*

$$f_{p,n} := Q_{p,n}(f).$$

Then, $(U_k^N(f))_{k\geq 1}$ defined by

$$U_k^N(f) := \gamma_{p_k}^N(1) f_{p_k,n}(X_{p_k}^{i_k}) \mathbf{1}_{\tau_N \geq p_k} - \gamma_{p_k-1}^N(1) Q_{p_k,\eta_{p_k-1}^N}(f_{p_k,n})(X_{p_k-1}^{i_k}) \mathbf{1}_{\tau_N \geq p_k - 1}$$

is a $(\mathcal{E}_k^N)$-martingale difference array.

Proof. The measurability is clear by definition. Since $\|G_n\|_\infty$ is bounded by 1, we have

$$\left| U_k^N(f) \right| \leq 3 \|f\|_\infty, \quad a.s. \tag{30}$$

which gives the integrability. Then, by the fact that

$$\mathbf{1}_{\tau_N \geq p_k - 1} = \mathbf{1}_{\tau_N \geq p_k} + \mathbf{1}_{\tau_N = p_k - 1},$$

one writes

$$
\begin{aligned}
&\mathbf{E}\left[U_k^N(f) \mid \mathcal{E}_{k-1}^N\right]\\
=&\mathbf{E}\left[\gamma_{p_k}^N(1) f_{p_k,n}(X_{p_k}^{i_k})\mathbf{1}_{\tau_N \geq p_k-1} - \gamma_{p_k-1}^N(1) Q_{p_k,\eta_{p_k-1}^N}(f_{p_k,n})(X_{p_k-1}^{i_k})\mathbf{1}_{\tau_N \geq p_k-1} \,\middle|\, \mathcal{E}_{k-1}^N\right]\\
&-\underbrace{\mathbf{E}\left[\gamma_{p_k}^N(1) f_{p_k,n}(X_{p_k}^{i_k})\mathbf{1}_{\tau_N = p_k-1} \,\middle|\, \mathcal{E}_{k-1}^N\right]}_{=0 \quad a.s.}\\
=&\mathbf{1}_{\tau_N \geq p_k-1}\mathbf{E}\left[\gamma_{p_k}^N(1) f_{p_k,n}(X_{p_k}^{i_k}) - \gamma_{p_k-1}^N(1) Q_{p_k,\eta_{p_k-1}^N}(f_{p_k,n})(X_{p_k-1}^{i_k}) \,\middle|\, \mathcal{E}_{k-1}^N\right]\\
=&0. \quad a.s.
\end{aligned}
$$

This ends the verification of Proposition C.1. □

Recall that, by definition, we have

$$
Q_n = \mathring{Q}_n + \eta_{n-1}(G_{n-1})(\dot{Q}_n - \mathring{Q}_n).
$$

Hence,

$$
\gamma_{n-1} Q_n = \gamma_n,
$$

which yields

$$
\gamma_0(f_{0,n}) = \gamma_n(f).
$$

We denote

$$
D_{p,n}^N(f) := \gamma_{p-1}^N(Q_{\hat{p}} - Q_p)(f_{p,n})\mathbf{1}_{\tau_N \geq p-1},
$$

with

$$
\forall p \geq 1, \quad Q_{\hat{p}} := \mathring{Q}_p + \eta_{p-1}^N(G_{p-1})\mathbf{1}_{\tau_N \geq p-1}(\dot{Q}_p - \mathring{Q}_p).
$$

The interest of the martingale difference sequence defined above lies in the following decomposition:

$$
\begin{aligned}
&\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\\
=&\sum_{p=0}^{n}\left(\gamma_p^N(f_{p,n})\mathbf{1}_{\tau_N \geq p} - \gamma_{p-1}^N Q_{p,\eta_{p-1}^N}(f_{p,n})\mathbf{1}_{\tau_N \geq p-1}\right)\\
=&\frac{1}{N}\sum_{k=1}^{(n+1)N} U_k^N(f) + \sum_{p=1}^{n} D_{p,n}^N(f),
\end{aligned}
\tag{31}
$$

taking into account the convention

$$
\gamma_{-1}^N = \gamma_0 = \eta_0 \quad \text{and} \quad Q_{0,\eta_{-1}^N}(f_{0,n})(x) \equiv \eta_0(f_{0,n}) = \gamma_n(f).
$$

Note that, for the case $\dot{Q}_p \equiv \mathring{Q}_p$, we have almost surely $D_{p-1,n}^N(f) \equiv 0$. In this case,

$$
\left(\gamma_p^N Q_{p,n}(f)\right)_{0 \leq p \leq n}
$$

is a $(\mathcal{F}_p; 0 \leq p \leq n)$-martingale.

Now, to facilitate the writing, we fix a finite time horizon $T \in \mathbf{N}^*$, and a test function $F \in \mathcal{B}_b(E_T)$. As a natural extension, we discuss a similar family of bias-martingales decomposition brought by the partial semigroup structure of coalescent Feynman-Kac kernels. Let us consider the term defined as follows:

$$\mathbb{X}_n^{\ddagger,b}(F) := \Gamma_{n,N}^{\ddagger,b}\ \mathbf{Q}_{n,T}^{\ddagger,b}(F)\mathbf{1}_{\tau_N \geq n}.$$

First, the integrability is guaranteed by Proposition C.7. Then, thanks to Proposition C.9, we get a almost sure equality which is very "close" to a martingale structure:

$$\mathbf{E}\left[\Gamma_n^{\ddagger,b}\ \mathbf{Q}_{n,T}^{\ddagger,b}(F)\mathbf{1}_{\tau_N \geq n}\ \middle|\ \mathcal{G}_{n-1}^N\right] = \Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}\ \mathbf{Q}_{n,T}^{\ddagger,b}(F)\mathbf{1}_{\tau_N \geq n-1}.$$

Note that, since $\mathbf{Q}_{n+1}^{\ddagger,b_n}$ conserves the coalescence structure, the term C_{b_n} disappears. We denote

$$^{\sharp}\mathbb{D}_n^{\ddagger,b}(F) := \mathbb{X}_n^{\ddagger,b}(F) - \mathbf{E}\left[\mathbb{X}_n^{\ddagger,b}(F)\ \middle|\ \mathcal{G}_{n-1}^N\right],$$

as well as

$$^{\flat}\mathbb{D}_n^{\ddagger,b}(F) := \mathbf{E}\left[\mathbb{X}_n^{\ddagger,b}(F)\ \middle|\ \mathcal{G}_{n-1}^N\right] - \mathbb{X}_{n-1}^{\ddagger,b}(F).$$

Thanks to Lemma C.10, Lemma C.12, the Minkowski's inequality and conservation of coalescence structure, we deduce the following Lemma C.1. Then, the Proposition C.2 is a direct application of Doob decomposition theorem.

Lemma C.1. *For any test function $F \in \mathcal{B}_b(E_T^2)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, we have*

$$\sum_{p=1}^{n} {}^{\sharp}\mathbb{D}_n^{\ddagger,b}(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right),$$

and

$$\sum_{p=1}^{n} {}^{\flat}\mathbb{D}_n^{\ddagger,b}(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right).$$

Proposition C.2. *For any test function $F \in \mathcal{B}_b(E_T^2)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, the integrable process $(\mathbb{X}_n^{\ddagger,b}(F); 0 \leq n \leq T)$ can be decomposed to a $(\mathcal{G}_n^N; 0 \leq n \leq T)$-martingale $(\mathbb{M}_n^{\ddagger,b}(F); 0 \leq n \leq T)$, and a integrable predictable process $(\mathbb{A}_n^{\ddagger,b}(F); 0 \leq n \leq T)$, respectively defined by*

$$\mathbb{M}_n^{\ddagger,b}(F) := \mathbb{X}_0^{\ddagger,b}(F) + \sum_{p=1}^{n} {}^{\sharp}\mathbb{D}_n^{\ddagger,b}(F),$$

and

$$\mathbb{A}_n^{\ddagger,b}(F) := \sum_{p=1}^{n} {}^{\flat}\mathbb{D}_n^{\ddagger,b}(F).$$

Similarly, we also discuss the martingale decomposition associated to the measure $\widetilde{\Gamma}_{n,N}^{\dagger,b}$. We define

$$\widetilde{\mathbb{X}}_n^{\dagger,b}(F) := \widetilde{\Gamma}_{n,N}^{\dagger,b}\ \widetilde{\mathbf{Q}}_{n,T}^{\dagger,b}(F)\mathbf{1}_{\tau_N \geq n},$$

as well as

$$^{\sharp}\widetilde{\mathbb{D}}_n^{\dagger,b}(F) := \widetilde{\mathbb{X}}_n^{\dagger,b}(F) - \mathbf{E}\left[\widetilde{\mathbb{X}}_n^{\dagger,b}(F)\ \middle|\ \mathcal{W}_{n-1}^N\right],$$

and

$$^{\flat}\widetilde{\mathbb{D}}_n^{\dagger,b}(F) := \mathbf{E}\left[\widetilde{\mathbb{X}}_n^{\dagger,b}(F)\ \middle|\ \mathcal{W}_{n-1}^N\right] - \widetilde{\mathbb{X}}_{n-1}^{\dagger,b}(F).$$

Thanks to Proposition C.10, Lemma C.11 and Lemma C.12, we have the following results. The unbiasedness given in Proposition C.3 is a direct consequence of the definition of $\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,(\emptyset)}$ and $\widetilde{\mathbf{Q}}_{n}^{\dagger,(\emptyset)}$.

Lemma C.2. *For any test function $F \in \mathcal{B}_b(E_T^2)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, we have*

$$\sum_{p=1}^{n} {}^{\sharp}\widetilde{\mathbb{D}}_n^{\dagger,b}(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right),$$

and

$$\sum_{p=1}^{n} {}^{\flat}\widetilde{\mathbb{D}}_n^{\dagger,b}(F) = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right).$$

Proposition C.3. *For any test function $F \in \mathcal{B}_b(E_T^2)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, the integrable process $(\widetilde{\mathbb{X}}_n^{\dagger,b}(F); 0 \le n \le T)$ can be decomposed to a $(\mathcal{W}_n^N; 0 \le n \le T)$-martingale $(\widetilde{\mathbb{M}}_n^{\dagger,b}(F); 0 \le n \le T)$, and a integrable predictable process $(\widetilde{\mathbb{A}}_n^{\dagger,b}(F); 0 \le n \le T)$, respectively defined by*

$$\widetilde{\mathbb{M}}_n^{\dagger,b}(F) := \widetilde{\mathbb{X}}_0^{\dagger,b}(F) + \sum_{p=1}^{n} {}^{\sharp}\widetilde{\mathbb{D}}_n^{\dagger,b}(F),$$

and

$$\widetilde{\mathbb{A}}_n^{\dagger,b}(F) := \sum_{p=1}^{n} {}^{\flat}\widetilde{\mathbb{D}}_n^{\dagger,b}(F).$$

In particular, under symmetric resampling scheme, that is $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \in [T]$, we also have

$${}^{\flat}\widetilde{\mathbb{D}}_n^{\dagger,(\emptyset)}(F) \equiv 0, \quad a.s.$$

which yields

$$\left(\widetilde{\mathbb{X}}_n^{\dagger,(\emptyset)}(F)\right)_{0 \le n \le T}$$

is a $(\mathcal{W}_n^N; 0 \le n \le T)$-martingale.

C.2 Verification of asymptotic variance expansion

In this section, we verify the asymptotic variance expansion (8) given in Section 3.1. Recall that, with the introduction of coalescent Feynman-Kac kernels, we have

$$\sigma_{\gamma_n}^2(f) = \sum_{p=0}^{n} \left(\gamma_p^{\otimes 2} C_1 \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) - \gamma_{p-1}^{\otimes 2} C_1 Q_{p,\eta_{p-1}}^{\otimes 2} \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2})\right).$$

By definition, since

$$\gamma_{n-1}^{\otimes 2}\mathbf{Q}_n^0 = \left(\gamma_{n-1}\mathring{Q}_n + \underbrace{\eta_{n-1}(G_{n-1})\gamma_{n-1}(\dot{Q}_n - \mathring{Q}_n)}_{\equiv 0.}\right)^{\otimes 2} = \gamma_n^{\otimes 2},$$

we have

$$\gamma_p^{\otimes 2} C_1 \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) = \Gamma_n^{(p)}(f^{\otimes 2}).$$

Next, for the latter term, since

$$\mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) = \left(Q_{p,n}(f)\right)^{\otimes 2},$$

by applying (17), we deduce that

$$\begin{aligned}\forall p \in [n], \quad & \gamma_{p-1}^{\otimes 2} C_1 Q_{p,\eta_{p-1}}^{\otimes 2} \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) \\ &= \underbrace{\gamma_{p-1}^{\otimes 2}(C_1(1))\eta_{p-1}^{\otimes 2} \mathring{Q}_p^{\otimes 2}\ \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2})}_{\Gamma_n^{(\emptyset)}(f^{\otimes 2})} + \gamma_{p-1}^{\otimes 2} C_1 R_{p,\eta_{p-1}}^{\otimes 2} \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2})\end{aligned}$$

Note that

$$\begin{aligned}\forall \varphi \in \mathcal{B}_b(E_p), \quad & R_{p,\eta_{p-1}}(\varphi)(x)^2 \\ &= \eta_{p-1}(G_{p-1})^2 \dot{Q}_n(\varphi)(x)^2 + G_{p-1}(x)^2 \eta_{p-1}\mathring{Q}_p(\varphi)^2 \\ &\quad - 2\eta_{p-1}(G_{p-1})\eta_{p-1}\mathring{Q}_p(\varphi)(G_{p-1} \times \dot{Q}_p)(\phi)(x),\end{aligned}$$

whence

$$\begin{aligned}& \gamma_{p-1}^{\otimes 2} C_1 R_{p,\eta_{p-1}}^{\otimes 2}(\varphi^{\otimes 2}) \\ &= \eta_{p-1}(G_{p-1})^2 \gamma_{p-1}^{\otimes 2} C_1 \dot{Q}_n^{\otimes 2}(\varphi^{\otimes 2}) \\ &\quad - \underbrace{\left(\underbrace{-\eta_{p-1}(G_{p-1}^2)\gamma_{p-1}^{\otimes 2}\mathring{Q}_p^{\otimes 2}}_{\eta_{p-1}(G_{p-1}^2)\gamma_{p-1}^{\otimes 2}[\mathring{Q}_p^{\otimes 2} - \dot{Q}_p \otimes \mathring{Q}_p - \mathring{Q}_p \otimes \dot{Q}_p]} + 2\eta_{p-1}(G_{p-1})\gamma_{p-1}^{\otimes 2}[\mathring{Q}_p \otimes (G_{p-1} \times \dot{Q}_p)] \right)(\varphi^{\otimes 2})}_{\gamma_{p-1}^{\otimes 2}\widetilde{\mathbf{Q}}_p^{\dagger,1}(\varphi^{\otimes 2}) = \widetilde{\Gamma}_{p-1}^{\dagger,(\emptyset)}\widetilde{\mathbf{Q}}_p^{\dagger,1}(\varphi^{\otimes 2})}.\end{aligned}$$

Replacing $\varphi^{\otimes 2}$ by $\mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) = \mathbf{Q}_{p,n}^{\dagger,(\emptyset)}(f^{\otimes 2})$, we get

$$\begin{aligned}\forall p \in [n], \quad & \gamma_{p-1}^{\otimes 2} C_1 Q_{p,\eta_{p-1}}^{\otimes 2} \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) \\ &= \Gamma_n^{(\emptyset)}(f^{\otimes 2}) - \widetilde{\Gamma}_{p-1}^{\dagger,(\emptyset)}\ \widetilde{\mathbf{Q}}_p^{\dagger,1}\ \mathbf{Q}_{p,n}^{\dagger,(\emptyset)}(f^{\otimes 2}) + \eta_{p-1}(G_{p-1})^2 \gamma_{p-1}^{\otimes 2} C_1 \dot{Q}_p^{\otimes 2}\ \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) \\ &= \Gamma_n^{(\emptyset)}(f^{\otimes 2}) - \widetilde{\Gamma}_n^{\dagger,(p-1)}(f^{\otimes 2}) + \eta_{p-1}(G_{p-1})^2\ \Gamma_{p-1}^{\dagger,(\emptyset)} C_1 \dot{Q}_p^{\otimes 2}\ \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}).\end{aligned}$$

Taking into account that

$$\forall p \in [n], \quad \Gamma_n^{(p-1)}(f^{\otimes 2}) - \eta_{p-1}(G_{p-1})^2\ \Gamma_{p-1}^{\dagger,(\emptyset)} C_1 \dot{Q}_p^{\otimes 2} \mathbf{Q}_{p,n}^{(\emptyset)}(f^{\otimes 2}) = \Gamma_n^{\dagger,(p-1)}(f^{\otimes 2}),$$

and

$$\Gamma_n^{(n)}(f^{\otimes 2}) = \Gamma_n^{\dagger,(n)}(f^{\otimes 2}),$$

as well as the convention (4) for the case $p = 0$, that writes

$$\gamma_{-1}^{\otimes 2} C_1 Q_{0,\eta_{-1}}^{\otimes 2} \mathbf{Q}_{0,n}^{(\emptyset)}(f^{\otimes 2}) = \Gamma_n^{(\emptyset)}(f^{\otimes 2}),$$

we finally obtain the coalescent tree-based asymptotic variance expansion:

$$\sigma_{\gamma_n}^2(f) := \sum_{p=0}^{n} \left(\Gamma_n^{\dagger,(p)}(f^{\otimes 2}) - \Gamma_n^{(\emptyset)}(f^{\otimes 2}) \right) + \sum_{p=0}^{n-1} \widetilde{\Gamma}_n^{\dagger,(p)}(f^{\otimes 2}).$$

C.3 Proof of Theorem 2.1

For any test function $f \in \mathcal{B}_b(E_n)$, the unbiased property

$$\mathbf{E}\left[\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}\right] = \gamma_n(f)$$

for the case $\dot{Q}_n \equiv \mathring{Q}_n$ is a direct consequence of the martingale decomposition (31). For the almost sure convergence, the proof is done by induction. For the step 0, the almost sure convergence of η_0^N is a direct consequence of law of large numbers for i.i.d. random variables. For step $n \geq 1$, we suppose that for each $0 \leq p \leq n-1$, we have

$$\forall \varphi_p \in \mathcal{B}_b(E_p), \quad \eta_p^N(\varphi_p) \xrightarrow[N\to\infty]{a.s.} \eta_p^N(\varphi_p).$$

We first check that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N} U_k^N(f) \xrightarrow[N\to\infty]{a.s.} 0. \tag{32}$$

Taking into account that

$$\left|U_k^N(f)\right| \leq 3\,\|f\|_\infty, \quad a.s.$$

we have, thanks to Azuma-Hoeffding inequality, for any $\alpha > 0$,

$$\mathbf{P}\left(\left|\sum_{k=1}^{(n+1)N} U_k^N(f)\right| > N\alpha\right) \leq 2\exp\left\{\frac{-2N\alpha^2}{9(n+1)\,\|f\|_\infty^2}\right\}.$$

Hence, the almost sure convergence (32) is then ensured by Borel-Cantelli lemma. On the other hand, the induction hypothesis gives

$$\forall p \in [n], \quad D_{p,n}^N(f) \xrightarrow[N\to\infty]{a.s.} 0,$$

which yields

$$\sum_{p=1}^{n} D_{p,n}^N(f) \xrightarrow[N\to\infty]{a.s.} 0.$$

The verification of the almost sure convergence for $\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}$ is then complete. The almost sure convergence of η_n^N is then trivial since for any test function $f \in \mathcal{B}_b(E_n)$, the convention (1) allows the writing

$$\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} = \frac{\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}}{\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n}}.$$

C.4 Proof of Theorem 2.2

Lemma C.3. *Let μ^N be an empirical on E_{n-1}, we suppose that there exists a probability measure μ on E_{n-1}, such that for any test function $\phi \in \mathcal{B}_b(E_{n-1})$, one has*

$$\mu^N(\phi) \xrightarrow[N\to\infty]{a.s.} \mu(\phi).$$

Then, for any test function $f \in \mathcal{B}_b(E_n)$, we have the following almost sure convergence:

(i) $\mu^N \left(R_{n,\mu^N}(f)^2\right) \xrightarrow[N\to\infty]{a.s.} \mu\left(R_{n,\mu}(f)^2\right)$;

(ii) $\mu^N \left(Q_{n,\mu^N}(f)^2\right) \xrightarrow[N\to\infty]{a.s.} \mu\left(Q_{n,\mu}(f)^2\right)$.

Proof. Before starting the proof, let us recall that for any probability measure $\mu \in \mathcal{P}(E_{n-1})$ and for any test function $f \in \mathcal{B}_b(E_n)$, we have

$$Q_{n,\mu}(f)(x) = \mu\mathring{Q}_n(f) + R_{n,\mu}(f)(x)$$

with $R_{n,\mu}$ defined in (15). Basic algebraic mulipulation gives

$$Q_{n,\mu}(f)(x)^2 = \mu\mathring{Q}_n(f)^2 + R_{n,\mu}(f)(x)^2 + 2\mu\mathring{Q}_n(f)R_{n,\mu}(f)(x).$$

Recall that, by definition, we have

$$R_{n,\mu^N}(f)(x) = \mu^N(G_{n-1})\dot{Q}_n(f)(x) - G_{n-1}(x)\mu^N\mathring{Q}_n(f)$$

and

$$\begin{aligned} &R_{n,\mu^N}(f)(x)^2 \\ =&\mu^N(G_{n-1})^2\dot{Q}_n(f)(x)^2 + G_{n-1}(x)^2\mu^N\mathring{Q}_n(f)^2 - 2\mu^N(G_{n-1})\mu^N\mathring{Q}_n(f)G_{n-1}(x)\dot{Q}_n(f)(x), \end{aligned}$$

whence we deduce that

$$\begin{aligned} &\mu^N\left(R_{n,\mu^N}(f)^2\right) \\ =&\mu^N(G_{n-1})^2\mu^N\left(\dot{Q}_n(f)^2\right) + \mu^N(G_{n-1}^2)\mu^N\left(\mathring{Q}_n(f)\right)^2 - 2\mu^N(G_{n-1})\mu^N\mathring{Q}_n(f)\mu^N\left(G_{n-1}\dot{Q}_n(f)\right). \end{aligned}$$

Since $G_{n-1}, \mathring{Q}_n(f) \in \mathcal{B}_b(E_{n-1})$, Theorem 2.1 gives that

$$\begin{aligned} &\mu^N(G_{n-1})^2\mu_N\left(\dot{Q}_n(f)^2\right) + \mu_N(G_{n-1}^2)\mu^N\mathring{Q}_n(f)^2 - 2\mu^N(G_{n-1})\mu^N\mathring{Q}_n(f)\mu_N\left(G_{n-1}\dot{Q}_n(f)\right). \\ &\xrightarrow[N\to\infty]{a.s.} \mu(G_{n-1})^2\mu\left(\dot{Q}_n(f)^2\right) + \mu(G_{n-1}^2)\mu\mathring{Q}_n(f)^2 - 2\mu(G_{n-1})\mu\mathring{Q}_n(f)\mu\left(G_{n-1}\dot{Q}_n(f)\right). \end{aligned}$$

On the other hand, as

$$\mu\left(R_{n,\mu}(f)^2\right) = \mu(G_{n-1})^2\mu\left(\dot{Q}_n(f)^2\right) + \mu(G_{n-1}^2)\mu\mathring{Q}_n(f)^2 - 2\mu(G_{n-1})\mu\mathring{Q}_n(f)\mu\left(G_{n-1}\dot{Q}_n(f)\right).$$

we safely deduce that

$$\mu^N\left(R_{n,\mu}(f)^2\right) \xrightarrow[N\to\infty]{a.s.} \mu\left(R_{n,\mu}(f)^2\right),$$

which terminates the verification for the point (i). Next, by standard calculation, we obtain

$$Q_{n,\mu^N}(f)(x)^2 = \mu^N\mathring{Q}_n(f)^2 + R_{n,\mu^N}(f)(x)^2 + 2\mu^N\mathring{Q}_n(f)R_{n,\mu^N}(f)(x),$$

whence

$$\begin{aligned} &\mu^N\left(Q_{n,\mu^N}(f)^2\right) \\ =&\mu^N(1)\mu^N\mathring{Q}_n(f)^2 + \mu^N\left(R_{n,\mu^N}(f)^2\right) + 2\mu^N\mathring{Q}_n(f)\mu^N(G_{n-1})\mu^N(\dot{Q}_n - \mathring{Q}_n)(f). \end{aligned}$$

Finally, as $\mathring{Q}_n(f), \dot{Q}_n(f) \in \mathcal{B}_b(E_{n-1})$, point (i) and Theorem 2.1 combined with the fact that

$$\mu\left(Q_{n,\mu}(f)^2\right) = \mu(1)\mu\mathring{Q}_n(f)^2 + \mu\left(R_{n,\mu}(f)^2\right) + 2\mu\mathring{Q}_n(f)\mu(G_{n-1})\mu(\dot{Q}_n - \mathring{Q}_n)(f)$$

ensure the desired convergence in point (ii). This closes the proof of Lemma C.3. $\square$

Now, let us start the proof of the CLT-type result for $\gamma_n^N \mathbf{1}_{\tau_N \geq n}$ and $\eta_n^N \mathbf{1}_{\tau_N \geq n}$. The proof is done by induction. The verification of step 0 is trivial by the central limit theorem for i.i.d. random variables. For step $n \geq 1$, we suppose that, for any test function $\varphi_p \in \mathcal{B}_b(E_p)$, we have

$$\forall 0 \leq p \leq n-1, \quad \sqrt{N}\left(\eta_p^N(\varphi_p)\mathbf{1}_{\tau_N \geq p} - \eta_p(\varphi_p)\right) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma_{\eta_p}^2(\varphi_p - \eta_p(\varphi_p))\right).$$

Again, let us go back to the decomposition (31). First, we prove that

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{(n+1)N} U_k^N(f) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma_{\gamma_n}^2\right). \tag{33}$$

In order to apply Theorem 2.3 in [McL74], one needs to verify that

- The boundness of G_n gives that

$$\max_{1 \leq k \leq (n+1)N} \left| \frac{1}{\sqrt{N}} U_k^N(f) \right| \leq \frac{3}{\sqrt{N}} \|f\|_\infty, \tag{34}$$

  which shows that $\max_{1\leq k\leq (n+1)N} \left|\frac{1}{\sqrt{N}} U_k^N(f)\right|$ is uniformly bounded in $\mathbb{L}^2$-norm.

- From (34), one also gets that

$$\max_{1 \leq k \leq (n+1)N} \left| \frac{1}{\sqrt{N}} U_k^N(f) \right| \xrightarrow[N\to\infty]{\mathrm{P}} 0.$$

- For the asymptotic variance, we deduce that

$$\begin{aligned}\left(U_k^N(f)\right)^2 = & \underbrace{\gamma_{p_k}^N(1)^2 f_{p_k,n}(X_{p_k}^{i_k})^2 \mathbf{1}_{\tau_N \geq p_k}}_{P_1^N(k)} + \underbrace{\gamma_{p_k-1}^N(1)^2 Q_{p_k,\eta_{p_k-1}^N}(f_{p_k,n})(X_{p_k-1}^{i_k})^2 \mathbf{1}_{\tau_N \geq p_k-1}}_{P_2^N(k)} \\ & - \underbrace{2\,\gamma_{p_k-1}^N(1)^2 \eta_{p_k-1}^N(G_{p_k-1}) f_{p_k,n}(X_{p_k}^{i_k}) Q_{p_k,\eta_{p_k-1}^N}(f_{p_k,n})(X_{p_k-1}^{i_k}) \mathbf{1}_{\tau_N \geq p_k}}_{P_3^N(k)}\end{aligned}$$

  First, let us prove that

$$\frac{1}{N} \sum_{k=1}^{(n+1)N} P_1^N(k) \xrightarrow[N\to\infty]{a.s.} \sum_{p=0}^{n} \gamma_p(1)\gamma_p(f_{p,n}^2). \tag{35}$$

  In fact, by the construction of the Feynman-Kac IPS, we have

$$\frac{1}{N} \sum_{k=1}^{(n+1)N} P_1^N(k) = \sum_{p=0}^{n} \gamma_p^N(1)\gamma_p^N(f_{p,n}^2).$$

  Hence, Theorem 2.1 gives the desired convergence (35). Second, for the term concerning $P_2(k)$, we would like to show that

$$\frac{1}{N} \sum_{k=1}^{(n+1)N} P_2^N(k) \xrightarrow[N\to\infty]{a.s.} \sum_{p=0}^{n} \gamma_{p-1}(1)\gamma_{p-1}\left(Q_{p,\eta_{p-1}}(f_{p,n})^2\right). \tag{36}$$

Similar to the previous case, we deduce that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N} P_2^N(k) = \sum_{p=0}^{n} \gamma_{p-1}^N(1)\gamma_{p-1}^N\left(Q_{p,\eta_{p-1}^N}(f_{p,n})^2\right).$$
$$= \sum_{p=0}^{n} \gamma_{p-1}^N(1)^2\eta_{p-1}^N\left(Q_{p,\eta_{p-1}^N}(f_{p,n})^2\right).$$

The convergence (36) is then obtained by combining Theorem 2.1 and the point (ii) of Lemma C.3. Then, for the term concerning $P_3(k)$, we prove that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N} P_3^N(k) \xrightarrow[N\to\infty]{a.s.} \sum_{p=0}^{n} \gamma_{p-1}(1)\gamma_{p-1}\left(Q_{p,\eta_{p-1}}(f_{p,n})^2\right). \tag{37}$$

Notice that

$$\begin{aligned}
&\mathbf{E}\left[P_3^N(k)\,\middle|\,\mathcal{F}_{p_k-1}^N\right]\\
=&\gamma_{p_k-1}^N(1)^2\eta_{p_k-1}^N(G_{p_k-1})Q_{p_k,\eta_{p_k-1}^N}(f_{p,n})(X_{p_{k-1}}^{i_k})\mathbf{E}\left[f(X_{p_k}^{i_k})\,\middle|\,\mathcal{F}_{p_k-1}^N\right]\\
=&\gamma_{p_k-1}^N(1)^2\,\underbrace{\eta_{p_k-1}^N(G_{p_k-1})K_{p_k,\eta_{p_k-1}^N}(f_{p,n})(X_{p_{k-1}}^{i_k})}_{Q_{p_k,\eta_{p_k-1}^N}(f_{p,n})(X_{p_{k-1}}^{i_k})}\,Q_{p_k,\eta_{p_k-1}^N}(f_{p,n})(X_{p_{k-1}}^{i_k})\\
=&P_2^N(k).
\end{aligned}$$

Hence, by exploiting the already proved convergence (36), it is sufficient to verify that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N}\left(P_3^N(k)-P_2^N(k)\right) \xrightarrow[N\to\infty]{a.s.} 0. \tag{38}$$

Recall the filtration $(\mathcal{E}_k^N; k\geq 0)$ defined by

$$\forall k\in[(n+1)N],\quad \mathcal{E}_k^N = \mathcal{F}_{p_k}^N \vee \sigma(X_{p_k}^1,\cdots,X_{p_k}^{i_k}).$$

It is readily checked that $(P_3^N(k)-P_2^N(k))$ is a $(\mathcal{E}_k^N)$-martingale difference array. In addition, the boundness of G_n ensures that

$$\left|P_3^N(k)-P_2^N(k)\right| \leq 8\left\|f\right\|_\infty. \qquad a.s.$$

Thanks to Heoffding-Azuma inequality, one obtains

$$\forall\alpha>0,\quad \mathbf{P}\left(\left|\sum_{k=1}^{(n+1)N} P_3^N(k)-P_2^N(k)\right|\geq N\alpha\right)\leq 2\exp\left(\frac{-\alpha^2 N}{32(n+1)\left\|f\right\|_\infty^2}\right).$$

The almost sure convergence (38) is then followed from Borel-Cantelli lemma. In conclusion, by combining (35),(36) and (37), one gets

$$\frac{1}{N}\sum_{k=1}^{(n+1)N}\left(U_k^N(f)\right)^2 \xrightarrow[N\to\infty]{a.s./\mathbf{P}} \sigma_{\gamma_n}^2(f). \tag{39}$$

Next, the induction hypothesis, Lemma C.5 and Theorem 2.1 ensure that

$$D_{p,n}^N(f) = \underbrace{\left[\eta_{p-1}^N(G_{p-1}) - \eta_{p-1}(G_{p-1})\right] \mathbf{1}_{\tau_N \geq p-1}}_{\mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right)} \underbrace{\gamma_{p-1}^N(\dot{Q}_p - Q_p)(f_{p,n})\mathbf{1}_{\tau_N \geq p-1}}_{o_{\mathrm{p}}(1)}, \tag{40}$$

whence

$$\sum_{p=0}^{n-1} D_{p,n}^N(f) = o_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

Slutsky's lemma then gives the CLT-type convergence for $\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}$. The CLT-type result for $\eta_n^N(f)\mathbf{1}_{\tau_N \geq n}$ is a direct consequence of Slutsky's lemma and the following decomposition

$$\begin{aligned} &\sqrt{N}\left(\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right)\mathbf{1}_{\tau_N \geq n} \\ =&\frac{1}{\gamma_n^N(1)}\sqrt{N}\left(\gamma_n^N(f - \eta_n(f))\mathbf{1}_{\tau_N \geq n} - \gamma_n(f - \eta_n(f))\right)\mathbf{1}_{\tau_N \geq n}. \end{aligned}$$

This ends the proof of Theorem 2.2.

C.5 Proof of Proposition 3.1

First, we notice that

$$V_n^N(f) = \left(\gamma_n^N(1)^2 - \Gamma_{n,N}^{\ddagger,(\emptyset)}\right)\mathbf{1}_{\tau \geq n} + \left(\Gamma_{n,N}^{\ddagger,(\emptyset)} - \Gamma_{n,N}^{(\emptyset)}\right)\mathbf{1}_{\tau \geq n}. \quad a.s.$$

We start by study the first term on the right-hand side of the equality above. Thanks to Theorem B.1, and by considering the stochastic bound given in Proposition C.7, we have, on the event $\{\tau_N \geq n\}$,

$$(\gamma_n^N)^{\otimes 2}(f^{\otimes 2}) = \left(\frac{N-1}{N}\right)^{n+1}\Gamma_{n,N}^{\ddagger,(\emptyset)}(f^{\otimes 2}) + \frac{1}{N}\left(\frac{N-1}{N}\right)^n\sum_{p=0}^{n}\Gamma_{n,N}^{\ddagger,(p)}(f^{\otimes 2}) + \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{N^2}\right).$$

Notice that

$$\left(\frac{N-1}{N}\right)^n = 1 - \mathcal{O}\left(\frac{1}{N}\right) \quad \text{and} \quad \left(\frac{N-1}{N}\right)^{n+1} - 1 = -\frac{n+1}{N} + \mathcal{O}\left(\frac{1}{N^2}\right),$$

which yields

$$N\left(\gamma_n^N(1)^2 - \Gamma_{n,N}^{\ddagger,(\emptyset)}\right)\mathbf{1}_{\tau \geq n} = \sum_{p=0}^{n}\left(\Gamma_{n,N}^{\ddagger,(p)}(f^{\otimes 2}) - \Gamma_{n,N}^{(\emptyset)}(f^{\otimes 2})\right) + \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{N}\right). \tag{41}$$

Next, by applying the decomposition given in Proposition B.4, we deduce that

$$N\left(\Gamma_{n,N}^{\ddagger,(\emptyset)} - \Gamma_{n,N}^{(\emptyset)}\right)\mathbf{1}_{\tau \geq n} = \left(1 + \frac{1}{N-1}\right)\sum_{p=0}^{n-1}\widetilde{\Gamma}_{n,N}^{\dagger,(p)}(F)\mathbf{1}_{\tau \geq n}.$$

Combining the two parts, we finally obtain the desired stochastic bound in (12). For (13), the reasoning is similar by the same algebraic manipulations. The only remark is that due to the "normalization" procedure, the stochastic bound w.r.t. $\mathbb{L}^2$-norm given by Proposition C.7 and Proposition C.8 will be replaced by a weaker version, namely,

$$\Gamma_{n,N}^{\ddagger,b}(F)/\gamma_n^N(1)^2 = \mathcal{O}_{\mathrm{p}}(1) \quad \text{and} \quad \widetilde{\Gamma}_{n,N}^{\dagger,b}(F)/\gamma_n^N(1)^2 = \mathcal{O}_{\mathrm{p}}(1).$$

This is ensured by Theorem 2.1.

C.6 Proof of Theorem B.1

On the event $\{\tau_N < n\}$, it is clear that both equalities hold. On the event $\{\tau_N \geq n\}$, the particle system is well-defined from level 0 to level n. Since

$$\left(\frac{N-1}{N}\right)^{n+1-|b|}\left(\frac{1}{N}\right)^{|b|} = \prod_{p=0}^{n}\frac{(N-1)^{1-b_p}}{N},$$

we have

$$\begin{aligned}&\left(\frac{N-1}{N}\right)^{n+1-|b|}\left(\frac{1}{N}\right)^{|b|}\bar{\Gamma}^{b}_{n,N}(F)\\ &= \frac{N^{n-1}}{(N-1)^{n+1}}\sum_{\ell^{[2]}_{0:n}\in((N)^2)^{\times(n+1)}}\left\{\prod_{p=0}^{n}\frac{(N-1)^{1-b_p}}{N}\right\}\left\{\prod_{p=0}^{n-1}\lambda^{b}_{p}(A_p^{\ell^{[2]}_{p+1}},\ell^{[2]}_p)\right\}C_{b_n}(F)(X_n^{\ell^{[2]}_n}).\end{aligned}$$

Enumerating all the possibilities for the coalescence indicator $b \in \{0,1\}^{n+1}$ leads to

$$\begin{aligned}&\sum_{b\in\{0,1\}^{n+1}}\left\{\prod_{p=0}^{n}\frac{(N-1)^{1-b_p}}{N}\right\}\bar{\Gamma}^{b}_{n,N}(F)\\ &= \sum_{\ell^{[2]}_0\in(N)^2}\cdots\sum_{\ell^{[2]}_{n-1}\in(N)^2}\left\{\prod_{p=0}^{n-1}\left(\frac{1}{N}\mathbf{1}_{\{A_p^{\ell^1_{p+1}}=A_p^{\ell^2_{p+1}}=\ell^1_p\neq\ell^2_p\}}+\frac{N-1}{N}\mathbf{1}_{\{A_p^{\ell^1_{p+1}}=\ell^1_p\neq A_p^{\ell^2_{p+1}}=\ell^2_p\}}\right)\right\}\\ &\left(\frac{N}{N-1}\right)^{n}\left\{\frac{N-1}{N}m^{\odot 2}(\mathbf{X_n})C_0(F)+\frac{1}{N}m^{\odot 2}(\mathbf{X_n})C_1(F)\right\}.\end{aligned}$$

To conclude, one just has to observe that, for each $0 \leq p \leq n-1$,

$$\sum_{\ell^{[2]}_p\in(N)^2}\left(\frac{1}{N}\mathbf{1}_{\{A_p^{\ell^1_{p+1}}=A_p^{\ell^2_{p+1}}=\ell^1_p\neq\ell^2_p\}}+\frac{N-1}{N}\mathbf{1}_{\{A_p^{\ell^1_{p+1}}=\ell^1_p\neq A_p^{\ell^2_{p+1}}=\ell^2_p\}}\right)=\frac{N-1}{N},\quad a.s.$$

while, by (2),

$$\frac{N-1}{N}m^{\odot 2}(\mathbf{X_n})C_0(F)+\frac{1}{N}m^{\odot 2}(\mathbf{X_n})C_1(F) = m^{\otimes 2}(\mathbf{X_n})(F) = (\eta_n^N)^{\otimes 2}(F).$$

Multiplying both sides by $\gamma_n^N(1)^2$ gives the corresponding relation for $(\gamma_n^N)^{\otimes 2}(F)$.

C.7 Proof of Proposition B.5

For all $N \geq 2$, let us consider the following auxiliary random matrix:

$$\begin{cases}\widetilde{\mathbf{H}}_n^{\ddagger,0|0} := \widetilde{\mathbf{H}}_n^{\dagger,0};\\ \widetilde{\mathbf{H}}_n^{\ddagger,1|0} := \widetilde{\mathbf{H}}_n^{\dagger,1};\\ \widetilde{\mathbf{H}}_n^{\ddagger,0|1} := 0;\\ \widetilde{\mathbf{H}}_n^{\ddagger,1|1} := \widetilde{\mathbf{H}}_n^{\dagger,1}.\end{cases}\quad\text{and}\quad\begin{cases}\widetilde{\mathbf{H}}_{n,(N)}^{\ddagger,0|0} := \widetilde{\mathbf{H}}_n^{\dagger,0};\\ \widetilde{\mathbf{H}}_{n,(N)}^{\ddagger,1|0} := \frac{1}{N-1}\widetilde{\mathbf{H}}_n^{\dagger,1};\\ \widetilde{\mathbf{H}}_{n,(N)}^{\ddagger,0|1} := 0;\\ \widetilde{\mathbf{H}}_{n,(N)}^{\ddagger,1|1} := \widetilde{\mathbf{H}}_n^{\dagger,1}.\end{cases}$$

Using the partial semigroup structure, we define, for any coalescence indicators b and b',

$$\forall F\in\mathcal{B}_b(E_n^2),\quad \widetilde{\Gamma}^{\ddagger,b'|b}_{n,N}(F) := \widetilde{\Lambda}^{\ddagger,b'|b}_n(\mathbf{F}^b_n) = m^{\odot 2}([N])\cdot\widetilde{\mathbf{H}}_1^{\ddagger,b'|b}\cdot\widetilde{\mathbf{H}}_2^{\ddagger,b'|b}\cdots\widetilde{\mathbf{H}}_n^{\ddagger,b'|b}(\mathbf{F}_n),$$

with
$$\mathbf{F}_n[\ell_n^{[2]}] := F(X_n^{\ell_n^{[2]}}).$$
Similarly, we also define
$$\forall F \in \mathcal{B}_b(E_n^2), \quad \widetilde{\Lambda}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n^b) = m^{\odot 2}([N]) \cdot \widetilde{\mathbf{H}}_{1,(N)}^{\ddagger,b'|b} \cdot \widetilde{\mathbf{H}}_{2,(N)}^{\ddagger,b'|b} \cdots \widetilde{\mathbf{H}}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n).$$
Remark that $0 = 0 \times \frac{1}{N-1}$. Hence, by definition, we have
$$\widetilde{\Lambda}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n^b) = \left(\frac{1}{N-1}\right)^{|b'-b|} \widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n^b). \quad a.s.$$
Next, we consider the binary decomposition w.r.t. a coalescence indicator b. More precisely,
$$\begin{aligned}\widetilde{\Lambda}_n^{\ddagger,b}(\mathbf{F}_n) &= \sum_{b' \in \mathcal{S}(b)} \widetilde{\Lambda}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n) \\ &= \widetilde{\Lambda}_{n,(N)}^{\ddagger,b|b}(\mathbf{F}_n) + \sum_{b' \in \mathring{\mathcal{S}}(b)} \widetilde{\Lambda}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n) \\ &= \widetilde{\Lambda}_n^{\ddagger,b|b}(\mathbf{F}_n) + \sum_{b' \in \mathring{\mathcal{S}}(b)} \widetilde{\Lambda}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n) \\ &= \widetilde{\Lambda}_n^{\dagger,b}(\mathbf{F}_n) + \sum_{b' \in \mathring{\mathcal{S}}(b)} \left(\frac{1}{N-1}\right)^{|b'-b|} \widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n).\end{aligned}$$
Therefore, it suffices to verify that for any coalescence indicator b' and b, we have
$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) = \mathcal{O}_{\mathbb{L}^2}(1).$$
By definition, if there exists $n_0 \geq 0$ such that
$$b'_{n_0} = 0 \quad \text{and} \quad b_{n_0} = 1,$$
we have
$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) \equiv 0.$$
If not, let us consider the mapping $\phi : \{0,1\}^2 \mapsto \{0,1\}$ defined by
$$\phi(0,0) = 0, \ \phi(1,0) = 1 \quad \text{and} \quad \phi(1,1) = 1.$$
We also denote $b_\phi := \big(\phi(b_0, b'_0), \phi(b_1, b'_1), \ldots, \phi(b_n, b'_n)\big)$. It is then easily checked that
$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) = \widetilde{\Lambda}_n^{\dagger,b_\phi}(\mathbf{F}_n) = \widetilde{\Gamma}_n^{\dagger,b_\phi}(F).$$
As a consequence, thanks to Proposition C.8, we have
$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) = \mathcal{O}_{\mathbb{L}^2}(1).$$
This is sufficient to end the proof of Proposition B.5.

C.8 Technical results

In this section, we list some technical results in support of the proofs given in the following sections. We remark that Lemma C.5 serves as a technical lemma, designed to prove Proposition C.4 by induction. Since the latter one is proved to be true, the hypothesis in Lemma C.5 can thus be removed. This is why in the Proposition C.5, it can be used without induction argument. In the proof of Proposition C.7, we do not give the finest analysis, which is done later in the proof of Lemma C.10. This organization is due to the complication of the notation in the present work. Since the rougher analysis in the proof of Proposition C.7 is more straightforward than the finer version in Lemma C.10, we consider it to be a good warm-up to the techniques involved in this section, which are highly repetitive in regard of the application of the pivotal decomposition (63).

Lemma C.4. *For any test function $f, g \in \mathcal{B}_b(E_0)$ and for both $b_0 = 0$ and $b_0 = 1$, we have*

$$(\eta_0^N)^{\otimes 2} C_{b_0}(f \otimes g) \mathbf{1}_{\tau_N \geq 0} - \eta_0^{\otimes 2} C_{b_0}(f \otimes g) = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

Proof. For the case $b_0 = 1$, it is sufficient to verify that

$$\forall \varphi \in \mathcal{B}_b(E_0), \quad \eta_0^N(\varphi) \mathbf{1}_{\tau_N \geq 0} - \eta_n(\varphi) = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right),$$

which is clear for the bounded i.i.d. random variables. More precisely, we have

$$\begin{aligned} \mathbf{E}\left[\eta_0^N(f)(\mathbf{1}_{\tau_N \geq 0} + 1 - 1) - \eta_0(f)\right] =& \mathbf{E}\left[\eta_0^N(f)(\mathbf{1}_{\tau_N \geq 0} - 1)\right] \\ &\leq \|f\|_\infty \mathbf{P}\left(\eta_0^N(G_0) = 0\right). \end{aligned}$$

Since $\mathbf{E}[\eta_0^N(G_0)] = \eta_0(G_0)$, we have, thanks to Hoeffding's inequality for bounded i.i.d. random variables,

$$\mathbf{P}\left(\eta_0^N(G_0) = 0\right) \leq \underbrace{\mathbf{P}\left(\eta_0^N(G_0) < \frac{\eta_0(G_0)}{2}\right)}_{\text{exponential decay rate.}},$$

which guarantees

$$\mathbf{E}\left[\eta_0^N(f) \mathbf{1}_{\tau_N \geq 0} - \eta_0(f)\right] = \mathcal{O}\left(\frac{1}{N}\right), \tag{42}$$

whence, by Cauchy-Schwartz inequality,

$$\left\|\eta_0^N(f) \mathbf{1}_{\tau_N \geq 0} - \eta_0(f)\right\|_{\mathbb{L}^2} = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right).$$

For the case $b_0 = 0$, since

$$\begin{aligned} &(\eta_0^N)^{\otimes 2} C_{b_0}(f \otimes g) \mathbf{1}_{\tau_N \geq 0} - \eta_0^{\otimes 2} C_{b_0}(f \otimes g) \\ =& \eta_0^N(f)\left(\eta_0^N(g) - \eta_0(g)\right) \mathbf{1}_{\tau_N \geq 0} + \eta_0(g)\left(\eta_0^N(f) - \eta_0(f)\right) \mathbf{1}_{\tau_N \geq 0}, \end{aligned}$$

The conclusion is also straightforward by considering the case $b_0 = 1$. $\square$

Lemma C.5. *If for all* $0 \le p \le n-1$, *we have*

$$\forall \varphi_p \in \mathcal{B}_b(E_p), \quad \eta_p^N(\varphi_p)\mathbf{1}_{\tau_N \ge p} - \eta_p(\varphi_p) = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right),$$

then, we also have

$$\mathbf{P}\left(\tau_N < n\right) = \mathcal{O}\left(\frac{1}{N}\right).$$

Proof. By definition, we have

$$\mathbf{P}\left(\tau_N < 0\right) = 0.$$

For $n \ge 1$, thanks to the bias-martingale decomposition (31), the almost sure boundness (30) and Azuma-Hoeffiding inequality, we have

$$\mathbf{P}\left(\gamma_n^N(1)\mathbf{1}_{\tau_N \ge n} < \frac{\gamma_n(1)}{2}\right) \le \underbrace{\mathbf{P}\left(\left|\frac{1}{N}\sum_{k=1}^{(n+1)N} U_k^N(1)\right| > \frac{\gamma_n(1)}{4}\right)}_{\text{exponential decay rate w.r.t. } N} + \mathbf{P}\left(\left|\sum_{p=1}^{n} D_{p,n}^N(1)\right| > \frac{\gamma_n(1)}{4}\right) \tag{43}$$

Then, we verify that

$$\mathbf{P}\left(\left|\sum_{p=1}^{n} D_{p,n}^N(1)\right| > \frac{\gamma_n(1)}{4}\right) = \mathcal{O}\left(\frac{1}{N}\right).$$

By Markov's inequality, one has

$$\mathbf{P}\left(\left|\sum_{p=1}^{n} D_{p,n}^N(1)\right| > \frac{\gamma_n(1)}{4}\right) \le \frac{4\left\|\sum_{p=1}^n D_{p,n}^N(1)\right\|_{\mathbb{L}^1}}{\gamma_n(1)} \le \frac{4\sum_{p=1}^n \left\|D_{p,n}^N(1)\right\|_{\mathbb{L}^1}}{\gamma_n(1)}$$

Thanks to Cauchy-Schwartz inequality, one derives

$$\begin{aligned}
&\left\|D_{p,n}^N(1)\right\|_{\mathbb{L}^1} \\
&= \left\|\gamma_{p-1}^N(1)\eta_{p-1}^N(\dot{Q}_p - \mathring{Q}_p)\left(Q_{p,n}(1)\right)\left[\eta_{p-1}^N(G_{p-1}) - \eta_{p-1}(G_{p-1})\right]\mathbf{1}_{\tau_N \ge p-1}\right\|_{\mathbb{L}^1} \\
&\le \left\|\gamma_{p-1}^N(1)\eta_{p-1}^N(\dot{Q}_p - \mathring{Q}_p)\left(Q_{p,n}(1)\right)\mathbf{1}_{\tau_N \ge p-1}\right\|_{\mathbb{L}^2}\left\|\left(\eta_{p-1}^N(G_{p-1}) - \eta_{p-1}(G_{p-1})\right)\mathbf{1}_{\tau_N \ge p-1}\right\|_{\mathbb{L}^2}.
\end{aligned}$$

In addition, we also have

$$\begin{aligned}
&\left\|\gamma_{p-1}^N(1)\eta_{p-1}^N(\dot{Q}_p - \mathring{Q}_p)\left(Q_{p,n}(1)\right)\mathbf{1}_{\tau_N \ge p-1}\right\|_{\mathbb{L}^2} \\
&\le \left\|\gamma_{p-1}^N(1)\left(\eta_{p-1}^N\dot{Q}_p\left(Q_{p,n}(1)\right)\mathbf{1}_{\tau_N \ge p-1} - \eta_{p-1}\dot{Q}_p\left(Q_{p,n}(1)\right)\right)\mathbf{1}_{\tau_N \ge p-1}\right\|_{\mathbb{L}^2} \\
&\quad + \left\|\gamma_{p-1}^N(1)\left(\eta_{p-1}^N\mathring{Q}_p\left(Q_{p,n}(1)\right)\mathbf{1}_{\tau_N \ge p-1} - \eta_{p-1}\mathring{Q}_p\left(Q_{p,n}(1)\right)\right)\mathbf{1}_{\tau_N \ge p-1}\right\|_{\mathbb{L}^2}.
\end{aligned}$$

Therefore, consider the hypothesis, and the fact that

$$\gamma_{n-1}^N(1) \le 1, \qquad a.s.$$

we get

$$\left\|D_{p,n}^N(1)\right\|_{\mathbb{L}^1} = \mathcal{O}\left(\frac{1}{N}\right),$$

which yields

$$\mathbf{P}\left(\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} < \frac{\gamma_n(1)}{2}\right) = \mathcal{O}\left(\frac{1}{N}\right). \tag{44}$$

Next, since

$$\begin{aligned}\{\tau_N < n\} \subset \{\tau_N \leq n\} =& \{\tau_N \leq n-1\} \cup \left\{\gamma_n^N(1) = 0\right\} \\ &\subset \{\tau_N \leq n-1\} \cup \left\{\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} < \frac{\gamma_n(1)}{2}\right\},\end{aligned}$$

one derives that

$$\mathbf{P}(\tau_N \leq n) \leq \mathbf{P}(\tau_N \leq n-1) + \mathbf{P}\left(\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} < \frac{\gamma_n(1)}{2}\right).$$

By applying the inequality above recursively from n to 0, one finally obtains

$$\mathbf{P}(\tau_N \leq n) = \mathcal{O}\left(\frac{1}{N}\right).$$

□

Proposition C.4. *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f) = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

In particular, one also has

$$\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f) = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

Proof. The proof is done by induction. For $n = 0$, the stochastic bound is clear for the bounded i.i.d. random variables. which is guaranteed by Lemma C.4. For step $n \geq 1$, we suppose that

$$\forall \varphi_p \in \mathcal{B}_b(E_p), \quad \eta_p^N(\varphi_p)\mathbf{1}_{\tau_N \geq p} - \eta_p(\varphi_p) = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

We consider the event $\Omega_n^N \subset \Omega$ defined by

$$\Omega_n^N := \left\{\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} \geq \frac{\gamma_n(1)}{2}\right\}.$$

By the definition of the absorbing time τ_N, one has

$$\Omega_n^N = \left\{\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} \geq \frac{\gamma_n(1)}{2} \text{ and } \tau_N \geq n\right\} \subset \{\tau_N \geq n\},$$

whence

$$\mathbf{1}_{\Omega_n^N} \leq \mathbf{1}_{\tau_N \geq n}.$$

Then, by the fact that

$$\begin{aligned}\mathbf{1}_{\tau_N \geq n} = \mathbf{1}_{\tau_N \geq n} - \mathbf{1}_{\Omega_n^N} + \mathbf{1}_{\Omega_n^N} &\leq \left|\mathbf{1}_{\tau_N \geq n} - 1 + 1 - \mathbf{1}_{\Omega_n^N}\right| + \mathbf{1}_{\Omega_n^N} \\ &\leq \left|1 - \mathbf{1}_{\tau_N \geq n}\right| + \left|1 - \mathbf{1}_{\Omega_n^N}\right| + \mathbf{1}_{\Omega_n^N} \leq 2\mathbf{1}_{(\Omega_n^N)^c} + \mathbf{1}_{\Omega_n^N},\end{aligned} \tag{45}$$

we obtain

$$\left|\eta_n^N(f) - \eta_n(f)\right| \mathbf{1}_{\tau_N \geq n} \leq \left|\eta_n^N(f) - \eta_n(f)\right| \mathbf{1}_{\Omega_n^N} + 4\left\|f\right\|_\infty \mathbf{1}_{(\Omega_n^N)^c}. \quad a.s.$$

Then, by applying the induction hypothesis, and thanks to a by-product (44) of Lemma C.5, the inequality above leads to

$$\left\|\left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\tau_N \geq n}\right\|_{\mathbb{L}^2} \leq \left\|\left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\Omega_n^N}\right\|_{\mathbb{L}^2} + \underbrace{4\left\|f\right\|_\infty \sqrt{\mathbf{P}\left((\Omega_n^N)^c\right)}}_{\mathcal{O}\left(\frac{1}{\sqrt{N}}\right)}.$$

It is thus sufficient to verify that

$$\left\|\left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\Omega_n^N}\right\|_{\mathbb{L}^2} = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{46}$$

By the fact that $\mathbf{1}_{\Omega_n^N}\mathbf{1}_{\tau_N \geq n} = \mathbf{1}_{\Omega_n^N}$, we have the following equality:

$$\begin{aligned}\left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\Omega_n^N} =& \frac{1}{\gamma_n^N(1)}\left(\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\right)\mathbf{1}_{\Omega_n^N} \\ &- \frac{\eta_n(f)}{\gamma_n^N(1)}\left(\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} - \gamma_n(1)\right)\mathbf{1}_{\Omega_n^N}.\end{aligned}$$

Then, by the definition of the event Ω_n^N, we have

$$\frac{1}{\gamma_n^N(1)}\left(\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\right)\mathbf{1}_{\Omega_n^N} \leq \frac{2}{\gamma_n(1)}\left(\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\right)\mathbf{1}_{\Omega_n^N}. \quad a.s.$$

As a consequence, to prove that

$$\frac{1}{\gamma_n^N(1)}\left(\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\right)\mathbf{1}_{\Omega_n^N} = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right),$$

and

$$\frac{\eta_n(f)}{\gamma_n^N(1)}\left(\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} - \gamma_n(1)\right)\mathbf{1}_{\Omega_n^N} = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

One only needs to verify that

$$\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f) = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right). \tag{47}$$

According to the bias-martingale decomposition (31), we have

$$\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f) = \frac{1}{N}\sum_{k=1}^{(n+1)N} U_k^N(f) + \sum_{p=1}^{n} D_{p,n}^N(f),$$

whence

$$\begin{aligned}
&N\left(\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}-\gamma_n(f)\right)^2\\
=&\frac{1}{N}\left(\sum_{k=1}^{(n+1)N}U_k^N(f)\right)^2+N\left(\sum_{p=1}^{n}D_{p,n}^N(f)\right)^2+2\sum_{k=1}^{(n+1)N}U_k^N(f)\sum_{p=1}^{n}D_{p,n}^N(f).
\end{aligned}$$

By definition, we have

$$\begin{aligned}
\left\|D_{p,n}^N(f)\right\|_{\mathbb{L}^2}=&\left\|\gamma_{p-1}^N(1)\eta_{p-1}^N(\dot{Q}_p-\mathring{Q}_p)\left(Q_{p,n}(f)\right)\left[\eta_{p-1}^N(G_{p-1})-\eta_{p-1}(G_{p-1})\right]\mathbf{1}_{\tau_N\geq p-1}\right\|_{\mathbb{L}^2}\\
\leq&2\,\|f\|_\infty\left\|\left(\eta_{p-1}^N(G_{p-1})\mathbf{1}_{\tau_N\geq p-1}-\eta_{p-1}(G_{p-1})\right)\mathbf{1}_{\tau_N\geq p-1}\right\|_{\mathbb{L}^2}.
\end{aligned}$$

By applying the induction hypothesis, one obtains

$$D_{p,n}^N(f)=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right),$$

which gives

$$\sum_{p=1}^{n}D_{p,n}^N(f)=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right),$$

and, by Cauchy-Schwartz inequality,

$$\left(\sum_{p=1}^{n}D_{p,n}^N(f)\right)^2=\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right).$$

Meanwhile, since $\left(U_k^N(f)\right)_{1\leq k\leq(n+1)N}$ is a martingale difference array, we have

$$\frac{1}{N}\mathbf{E}\left[\left(\sum_{k=1}^{(n+1)N}U_k^N(f)\right)^2\right]=\frac{1}{N}\sum_{k=0}^{(n+1)N}\mathbf{E}\left[U_k^N(f)^2\right]\xrightarrow[N\to\infty]{}\sigma_{\gamma_n}^2(f)<+\infty,$$

where the convergence is a by-product (39) of the proof of Theorem 2.2 and dominated convergence theorem. Hence, we obtain

$$\sum_{k=1}^{(n+1)N}U_k^N(f)=\mathcal{O}_{\mathbb{L}^2}\left(\sqrt{N}\right).$$

In summary, we have

$$\begin{aligned}
&N\mathbf{E}\left[\left|\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}-\gamma_n(f)\right|^2\right]\\
=&\frac{1}{N}\sum_{k=0}^{(n+1)N}\mathbf{E}\left[U_k^N(f)^2\right]+2\mathbf{E}\left[\underbrace{\left|\sum_{k=0}^{(n+1)N}U_k^N(f)\right|}_{\mathcal{O}_{\mathbb{L}^2}(\sqrt{N})}\underbrace{\left|\sum_{p=1}^{n}D_{p,n}^N(f)\right|}_{\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right)}\right]+2N\mathbf{E}\left[\underbrace{\left(\sum_{p=1}^{n}D_{p,n}^N(f)\right)^2}_{\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right)}\right],
\end{aligned}$$

which, thanks to Cauchy-Schwartz inequality, leads to

$$\left\|\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}-\gamma_n(f)\right\|_{\mathbb{L}^2}=\mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{48}$$

This ends the verification of (47) and the proof of this proposition.

□

Proposition C.5 ($\mathbb{L}^2$-propagation of chaos). *For any test function $f, g \in \mathcal{B}_b(E_n)$, we have*

$$(\eta_n^N)^{\odot 2}(f\otimes g)\mathbf{1}_{\tau_N\geq n}-\eta_n^{\otimes 2}(f\otimes g)=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

Proof. Before starting the proof, let us mention that by Minkowski's inequality, for two random variables X and Y, one has

$$X=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right)\quad\text{and}\quad Y=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right)\implies X+Y=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

Notice that

$$(\eta_n^N)^{\odot 2}(f\otimes g)\mathbf{1}_{\tau_N\geq n}-\eta_n^{\otimes 2}(f\otimes g)=\left((\eta_n^N)^{\odot 2}(f\otimes g)-\eta_n^{\otimes 2}(f\otimes g)\right)\mathbf{1}_{\tau_N\geq n}+\eta_n^{\otimes 2}(f\otimes g)(1-\mathbf{1}_{\tau_N\geq n}).$$

Thanks to Proposition C.4 and Lemma C.5, one derives

$$\mathbf{E}\left[\left|\eta_n^{\otimes 2}(f\otimes g)(1-\mathbf{1}_{\tau_N\geq n})\right|^2\right]\leq\|f\|_\infty\|g\|_\infty\,\mathbf{E}\left[1-\mathbf{1}_{\tau_N\geq n}\right]\leq\|f\|_\infty\|g\|_\infty\underbrace{\mathbf{P}\left(\tau<n\right)}_{\mathcal{O}(\frac{1}{N})},$$

which implies that

$$\eta_n^{\otimes 2}(f\otimes g)(1-\mathbf{1}_{\tau_N\geq n})=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).$$

Next, considering the decomposition (2), we deduce that

$$\begin{aligned}\left((\eta_n^N)^{\odot 2}(f\otimes g)-\eta_n^{\otimes 2}(f\otimes g)\right)\mathbf{1}_{\tau_N\geq n}=&\frac{N}{N-1}\left((\eta_n^N)^{\otimes 2}(f\otimes g)-\eta_n^{\otimes 2}(f\otimes g)\right)\mathbf{1}_{\tau_N\geq n}\\&+\frac{1}{N-1}\left(\eta_n^N(fg)+\eta_n(f)\eta_n(g)\right)\mathbf{1}_{\tau_N\geq n}.\end{aligned}$$

Concerning the term at the right-hand side of the equality above, we noticed that

$$\left\|\frac{1}{N-1}\left(\eta_n^N(fg)+\eta_n(f)\eta_n(g)\right)\mathbf{1}_{\tau_N\geq n}\right\|_{\mathbb{L}^2}\leq\frac{2}{N-1}\|f\|_\infty\|g\|_\infty.$$

In addition, since

$$\begin{aligned}&\left((\eta_n^N)^{\otimes 2}(f\otimes g)-\eta_n^{\otimes 2}(f\otimes g)\right)\mathbf{1}_{\tau_N\geq n}\\=&\left(\eta_n^N(f)\left[\eta_n^N(g)-\eta_n(g)\right]+\eta_n(g)\left[\eta_n^N(f)-\eta_n(f)\right]\right)\mathbf{1}_{\tau_N\geq n}\\\leq&2\left(\|f\|_\infty\vee\|g\|_\infty\right)\left(\left[\eta_n^N(f)-\eta_n(f)\right]\vee\left[\eta_n^N(g)-\eta_n(g)\right]\right)\mathbf{1}_{\tau_N\geq n},\quad a.s.\end{aligned}$$

it is then sufficient to verify that

$$\forall f \in \mathcal{B}_b(E_n), \quad \left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\tau_N \geq n} = \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right). \tag{49}$$

which is guaranteed by Proposition C.4 since

$$\left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\tau_N \geq n} = \left(\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right) \mathbf{1}_{\tau_N \geq n}. \quad a.s.$$

The proof is then finished. □

Proposition C.6 (Biasedness)**.** *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\mathbf{E}\left[\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

In particular, we also have

$$\mathbf{E}\left[\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

Remark. Different from the other technical results, the order of the bias given in this proposition will not be used to prove the consistency of the variance estimator. They are put in this section simply because we think the order of bias is important but not as relevant in the present work, where the most results we discussed are "short term" asymptotic properties of the IPS. In addition, we want to mention that by the same strategy, one can obtain an explicit bound w.r.t. both n and N for the bias. The main difference from the classic Feynman-Kac particle models discussed in [DM04] is the "lack-of-martingale" or, said differently, the bias-martingale structure (cf. (31)). As a consequence, the decay rate of the absorbing time is not exponential w.r.t. N any more. Instead, it is replace by $\mathcal{O}(1/N)$, as stated in Lemma C.5. This is why the order of bias w.r.t. N is not affected. The "lack-of-martingale" structure also requires an induction in order to deal with the bias term encountered in the bias-martingale decomposition (31). This technique is frequently used in the adaptive SMC context (cf. [DG19](Chapter 2)).

Proof. The proof is done by induction. Thanks to a by-product (42) of Lemma C.4, we have

$$\mathbf{E}\left[\eta_0^N(f)\mathbf{1}_{\tau_N \geq 0} - \eta_0(f)\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

For step $n \geq 1$, we suppose that

$$\forall \varphi \in \mathcal{B}_b(E_{n-1}), \quad \mathbf{E}\left[\eta_{n-1}^N(\varphi)\mathbf{1}_{\tau_N \geq n-1} - \eta_{n-1}(f)\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

By the bias-martingale decomposition (31), it gives

$$\forall \psi \in \mathcal{B}_b(E_n), \quad \mathbf{E}\left[\gamma_n^N(\varphi)\mathbf{1}_{\tau_N \geq n-1} - \gamma_n(f)\right] = \mathcal{O}\left(\frac{1}{N}\right). \tag{50}$$

Next, standard calculations give

$$\left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\tau_N \geq n} = \left(\frac{\gamma_n^N(f)}{\gamma_n^N(1)} - \frac{\gamma_n(f)}{\gamma_n(1)}\right) \mathbf{1}_{\tau_N \geq n}$$
$$= \frac{\gamma_n(1)}{\gamma_n^N(1)} \left(\gamma_n^N(f_n) - \gamma_n(f_n)\right) \mathbf{1}_{\tau_N \geq n},$$

with

$$f_n := \frac{1}{\gamma_n(1)}(f - \eta_n(f)).$$

Remark that, by definition,

$$\gamma_n(f_n) = 0.$$

Then, by applying Lemma C.5, we noticed that

$$\begin{aligned}
&\mathbf{E}\left[\left(\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right) - \left(\eta_n^N(f) - \eta_n(f)\right) \mathbf{1}_{\tau_N \geq n}\right] \\
=&\mathbf{E}\left[\left(\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right)(1 - \mathbf{1}_{\tau_N \geq n})\right] \\
\leq& 2\,\|f\|_\infty \mathbf{P}(\tau_N < n) = \mathcal{O}\left(\frac{1}{N}\right).
\end{aligned}$$

Mutatis mutandis, one also has

$$\mathbf{E}\left[\left(\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f)\right) - \left(\gamma_n^N(f) - \gamma_n(f)\right) \mathbf{1}_{\tau_N \geq n}\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

Therefore, considering the induction hypothesis (50), we only have to show that

$$\begin{aligned}
&\left(\frac{\gamma_n(1)}{\gamma_n^N(1)} - 1\right)\left(\gamma_n^N(f_n) - \gamma_n(f_n)\right) \mathbf{1}_{\tau_N \geq n} \\
=& -\left(\frac{\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} - \gamma_n(1)}{\gamma_n^N(1)}\right)\left(\gamma_n^N(f_n)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f_n)\right) \mathbf{1}_{\tau_N \geq n} = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right).
\end{aligned} \tag{51}$$

Recall that the event $\Omega_n^N \subset \Omega$ is defined by

$$\Omega_n^N := \left\{\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} \geq \frac{\gamma_n(1)}{2}\right\},$$

and we have $\mathbf{1}_{\Omega_n^N} \leq \mathbf{1}_{\tau_N \geq n} \leq 2\mathbf{1}_{(\Omega_n^N)^c} + \mathbf{1}_{\Omega_n^N}$. Notice that, by Lemma C.5 and the definition of f_n, one has

$$\begin{aligned}
&\left(\frac{\gamma_n(1)}{\gamma_n^N(1)} - 1\right)\left(\gamma_n^N(f_n) - \gamma_n(f_n)\right) \mathbf{1}_{(\Omega_n^N)^c} \\
\leq& \left|\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right| \mathbf{1}_{(\Omega_n^N)^c} + \left|\gamma_n^N(f_n)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f_n)\right| \mathbf{1}_{(\Omega_n^N)^c} \\
\leq& 4\,\|f\|_\infty \mathbf{1}_{(\Omega_n^N)^c} = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right).
\end{aligned} \tag{52}$$

In addition, by definition of Ω_n^N, one gets

$$\begin{aligned}
&-\left(\frac{\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} - \gamma_n(1)}{\gamma_n^N(1)}\right)\left(\gamma_n^N(f_n)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f_n)\right) \mathbf{1}_{\Omega_n^N} \\
\leq& \frac{2}{\gamma_n(1)} \left|\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} - \gamma_n(1)\right| \left|\gamma_n^N(f_n)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f_n)\right|. \quad a.s.
\end{aligned}$$

Thus, thanks to a by-product (48) in the proof of Proposition C.4, we obtain

$$\left(\frac{\gamma_n(1)}{\gamma_n^N(1)} - 1\right)\left(\gamma_n^N(f_n) - \gamma_n(f_n)\right)\mathbf{1}_{\Omega_n^N} = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right). \tag{53}$$

Finally, combining both (52) and (53) terminates the verification of (51), which also ends the proof of Proposition C.6.

□

Before proceeding further, we recall and introduce some notation that is used frequently in the following technical results. For $N \in \mathbf{N}^*$, we denote

$$[N]_p^q := \left\{(i_1, \ldots, i_q) \in [N]^q : \mathrm{Card}\{i_1, \ldots, i_q\} = p\right\}. \tag{54}$$

In particular, we denote $(N)^q := [N]_q^q$. We also write

$$\left((N)^2\right)^{\times q} := \underbrace{(N)^2 \times (N)^2 \times \cdots \times (N)^2}_{q \text{ times}}. \tag{55}$$

With a slight abuse of notation, we admit that

$$((i,j),k) = (i,j,k). \quad \text{and} \quad ((i,j),(k,l)) = (i,j,k,l).$$

We also adopt the notation introduced in Section B.5. Fixing some $b \in \{0,1\}^{n+1}$, we denote

$$\Lambda_n^{\ddagger,b}[\ell_n^{[2]}] := \frac{1}{N(N-1)} \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{\prod_{p=0}^{n-1} \mathbf{G}_\mathbf{p}^\ddagger(\mathbf{X}_\mathbf{p}) \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]})\right\}, \tag{56}$$

with the convention

$$\Lambda_0^{\ddagger,b}[\ell_0^{[2]}] := \frac{1}{N(N-1)}.$$

It is readily checked that

$$\Lambda_n^{\ddagger,b}[\ell_n^{[2]}] = \sum_{\ell_{n-1}^{[2]} \in (N)^2} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}] \mathbf{G}_{\mathbf{n-1}}^\ddagger(\mathbf{X}_{\mathbf{n-1}}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}).$$

This allows an alternative representation of $\Gamma_{n,N}^{\ddagger,b}$:

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma_{n,N}^{\ddagger,b}(F) = \sum_{\ell_n^{[2]} \in (N)^2} \Lambda_n^\ddagger[\ell_n^{[2]}] C_{b_n}(F)(X_n^{\ell_n^{[2]}}), \tag{57}$$

which covers the case $n = 0$. Similarly, we also denote

$$\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}] := \frac{1}{N(N-1)} \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{\prod_{p=0}^{n-1} \widetilde{\mathbf{G}}_\mathbf{p}^{\dagger,b_p}(\ell_{p:p+1}^{[2]}, \mathbf{B}_\mathbf{p}, \mathbf{X}_\mathbf{p}) \lambda_p^{(\varnothing)}(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]})\right\}, \tag{58}$$

with the convention

$$\widetilde{\Lambda}_0^{\dagger,b}[\ell_0^{[2]}] := \frac{1}{N(N-1)}.$$

We also have the decomposition

$$\widetilde{\Lambda}_{n}^{\dagger,b}[\ell_{n}^{[2]}] = \sum_{\ell_{n-1}^{[2]}\in(N)^{2}} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\emptyset)}(A_{n-1}^{\ell_{n}^{[2]}},\ell_{n-1}^{[2]}). \tag{59}$$

As is shown in the previous case (57), ${}^{\dagger}\widetilde{\Gamma}_{n,N}^{b}$ admits the following alternative representation:

$$\forall F \in \mathcal{B}_b(E_n^2),\quad \widetilde{\Gamma}_{n,N}^{\dagger,b}(F) = \sum_{\ell_{n}^{[2]}\in(N)^{2}} \widetilde{\Lambda}_{n}^{\dagger,b}[\ell_{n}^{[2]}]F(X_{n}^{\ell_{n}^{[2]}}), \tag{60}$$

which covers the case $n = 0$.

Proposition C.7. *For any coalescence indicator $b \in \{0,1\}^{n+1}$, we have*

$$\Gamma_{n,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n} = \mathcal{O}_{\mathbb{L}^2}(1).$$

In particular, for any test function $F \in \mathcal{B}_b(E_n^2)$, we also have

$$\Gamma_{n,N}^{\ddagger,b}(F)\mathbf{1}_{\tau_N\geq n} = \mathcal{O}_{\mathbb{L}^1}(1).$$

Proof. The proof is done by induction. For the step $n = 0$, it is clear since ${}^{\ddagger}\Gamma_{0,N}^{b}(1) = 1$. For step $n \geq 1$, we suppose that

$$\sup_{N>0}\mathbf{E}\left[\Gamma_{n-1,N}^{\ddagger,b}(1)^2\mathbf{1}_{\tau_N\geq n-1}\right] < +\infty.$$

By the alternative representation (57), for all $N \geq 4$, we have

$$\begin{aligned}
&\mathbf{E}\left[\Gamma_{n,N}^{\ddagger,b}(1)^2\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{G}_{n-1}^{N}\right]\\
&= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{\prime[2]})\in((N)^2)^{\times 2}} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{\prime[2]}]\mathbf{1}_{\tau_N\geq n-1}\\
&\quad\mathbf{E}\left[\sum_{(\ell_{n}^{[2]},\ell_{n}^{\prime[2]})\in((N)^2)^{\times 2}} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^{b}(A_{n-1}^{\ell_{n}^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{b}(A_{n-1}^{\ell_{n}^{\prime[2]}},\ell_{n-1}^{\prime[2]})\,\middle|\,\mathcal{G}_{n-1}^{N}\right],
\end{aligned} \tag{61}$$

since the definition of $\mathbf{G}_{\mathbf{n-1}}^{\ddagger}$ gives

$$\mathbf{E}\left[\sum_{(\ell_{n}^{[2]},\ell_{n}^{\prime[2]})\in((N))^{\times 2}} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^{b}(A_{n-1}^{\ell_{n}^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{b}(A_{n-1}^{\ell_{n}^{\prime[2]}},\ell_{n-1}^{\prime[2]})\,\middle|\,\mathcal{G}_{n-1}^{N}\right]\mathbf{1}_{\tau_N=n-1} = 0.$$

In order to simplify the notation, we may omit $\mathbf{1}_{\tau_N\geq n-1}$, which is $\mathcal{G}_{n-1}^{N}$-measurable, in the rest of the proof. Before proceeding, we recall the conditional distribution of the selection step. Given $\mathcal{G}_{n-1}^{N}$, we have

$$\forall i \in [N],\quad A_{n-1}^{i} \sim G_{n-1}(X_{n-1}^{i})\delta_i(\cdot) + (1 - G_{n-1}(X_{n-1}^{i}))\sum_{k=1}^{N}\frac{G_{n-1}(X_{n-1}^{k})}{Nm(\mathbf{X_{n-1}})(G_{n-1})}\delta_k(\cdot).$$

Hence, for any $j \in [N]$, we have

$$\begin{aligned}
&m(\mathbf{X_{n-1}})(G_{n-1}) \sum_{i=1}^{N} \mathbf{P}\left(A_{n-1}^{i} = j \mid \mathcal{G}_{n-1}^{N}\right) \\
=&m(\mathbf{X_{n-1}})(G_{n-1})G_{n-1}(X_{n-1}^{j}) + \sum_{i=1}^{N}(1 - G_{n-1}(X_{n-1}^{i}))\frac{G_{n-1}(X_{n-1}^{j})}{N} \\
=&G_{n-1}(X_{n-1}^{j}) \leq 1. \quad a.s.
\end{aligned} \tag{62}$$

With the notation introduced in (54) and (55), for $N \geq 4$, we have the decomposition

$$((N)^2)^{\times 2} = \left(((N)^2)^{\times 2} \cap [N]_2^4\right) \cup \left(((N)^2)^{\times 2} \cap [N]_3^4\right) \cup (N)^4. \tag{63}$$

The rest of the proof consists in studying the term

$$\mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N))^{\times 2}} \mathbf{G_{n-1}^{\ddagger}}(\mathbf{X_{n-1}})^2 \lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^{b}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^{N}\right]$$

with respect to the decomposition above and the bound given in (62).

(i) Case: $(\ell_n^{[2]}, \ell_n'^{[2]}) \in \left((N)^2\right)^{\times 2} \cap [N]_2^4$:

In this case, there are only two distinct random variables in the tuple

$$\left(A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n'^1}, A_{n-1}^{\ell_n'^2}\right).$$

Due to the symmetry of the particles, we first calculate the number of choices to assign two different random variables in the tuple above.

$$\underbrace{C_4^2/2}_{\substack{\text{possible choices to assign} \\ \text{two distinct couples in } (4)^4.}} \quad - \quad \underbrace{2/2}_{\text{limitation by } ((N)^2)^{\times 2}.} \quad = 2. \tag{64}$$

More precisely, in this case, the two possible assignments are

$$A_{n-1}^{\ell_n^1} = A_{n-1}^{\ell_n'^1},\ A_{n-1}^{\ell_n^2} = A_{n-1}^{\ell_n'^2} \quad \text{and} \quad A_{n-1}^{\ell_n^1} = A_{n-1}^{\ell_n'^2},\ A_{n-1}^{\ell_n^2} = A_{n-1}^{\ell_n'^1}.$$

Without loss of generality, we suppose that $A_{n-1}^{\ell_n^{[2]}}$ are two distinct random variables, with one of the two assignments above. Then, when $\ell_n^{[2]}$ varies freely in $(N)^2$, the values of $A_{n-1}^{\ell_n'^{[2]}}$ will be a.s. determined by the chosen assignment and the value of $A_{n-1}^{\ell_n^{[2]}}$. By the fact that λ_{n-1}^{b} is indicator function, we have

$$0 \leq \lambda_{n-1}^{b} \leq 1.$$

In addition, since G_n varies on the interval $[0, 1]$, we have

$$0 \leq m(\mathbf{X_n})(G_n) \leq 1. \quad a.s.$$

Now, let us deduce that

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell_n^{\prime[2]}}_{n-1},\ell^{\prime[2]}_{n-1})\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
\leq&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
\leq&\left(\frac{N}{N-1}\right)^2\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}m(\mathbf{X_{n-1}})(G_n)^2\mathbf{1}_{\left\{A^{\ell^1_n}_{n-1}=\ell^1_{n-1}\right\}}\mathbf{1}_{\left\{A^{\ell^2_n}_{n-1}=b_{n-1}\ell^1_{n-1}+(1-b_{n-1})\ell^2_{n-1}\right\}}\,\middle|\,\mathcal{G}^N_{n-1}\right].
\end{aligned}
$$

By the conditional independence between $A^{\ell^1_n}_{n-1}$ and $A^{\ell^2_n}_{n-1}$, one deduces

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}m(\mathbf{X_{n-1}})(G_n)^2\mathbf{1}_{\left\{A^{\ell^1_n}_{n-1}=\ell^1_{n-1}\right\}}\mathbf{1}_{\left\{A^{\ell^2_n}_{n-1}=b_{n-1}\ell^1_{n-1}+(1-b_{n-1})\ell^2_{n-1}\right\}}\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
&=\sum_{\ell_n^{[2]}\in(N)^2}\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A^{\ell^1_n}_{n-1}=\ell^1_{n-1}\right\}}\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
&\qquad\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A^{\ell^2_n}_{n-1}=b_{n-1}\ell^1_{n-1}+(1-b_{n-1})\ell^2_{n-1}\right\}}\,\middle|\,\mathcal{G}^N_{n-1}\right] \qquad (65)\\
&\leq\sum_{\ell^1_n\in[N]}\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A^{\ell^1_n}_{n-1}=\ell^1_{n-1}\right\}}\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
&\qquad\sum_{\ell^2_n\in[N]}\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A^{\ell^2_n}_{n-1}=b_{n-1}\ell^1_{n-1}+(1-b_{n-1})\ell^2_{n-1}\right\}}\,\middle|\,\mathcal{G}^N_{n-1}\right].
\end{aligned}
$$

Combined to (62), one gets

$$
\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})\,\middle|\,\mathcal{G}^N_{n-1}\right]\leq 1^2\times\frac{N}{N-1}.\quad a.s. \qquad (66)
$$

which gives

$$
\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})\,\middle|\,\mathcal{G}^N_{n-1}\right]\leq\left(\frac{N}{N-1}\right)^2.\quad a.s. \qquad (67)
$$

Considering the choices of assignments mentioned above, one finally gets

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n^{\prime[2]})\in((N))^{\times 2}\cap[N]^4_2}\mathbf{G}^{\ddagger}_{\mathbf{n}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell_n^{\prime[2]}}_{n-1},\ell^{\prime[2]}_{n-1})\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
&\qquad\qquad\leq 2\times\left(\frac{N}{N-1}\right)^2.\quad a.s.
\end{aligned}
$$

(ii) Case: $(\ell_n^{[2]}, \ell_n^{\prime[2]}) \in \left((N)^2\right)^{\times 2} \cap [N]_3^4$:

In this case, there are three distinct random variables in the tuple

$$\left(A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n^{\prime 1}}, A_{n-1}^{\ell_n^{\prime 2}}\right).$$

Similar to the previous case, we calculate the number of choices to assign three different random variables in the tuple above.

$$\underbrace{\mathrm{C}_4^3}_{\substack{\text{possible choices to divise}\\ (4)^4 \text{ into three distinct parts.}}} - \underbrace{0}_{\text{limitation by } ((N)^2)^{\times 2}.} = 4. \tag{68}$$

Let us fix one assignment. We suppose that $\ell_n^{[2]}$ and $\ell_n^{\prime 1}$ are three distinct numbers. Then, whilst $(\ell_n^{[2]}, \ell_n^{\prime 1})$ varies freely in $(N)^3$, the value of $A_{n-1}^{\ell_n^{\prime 2}}$ is a.s. determined by the chosen assignment and the values of $(A_{n-1}^{\ell_n^{[2]}}, A_{n-1}^{\ell_n^{\prime 1}})$. Given $\mathcal{G}_{n-1}^N$, by the conditional independence between $A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}$ and $A_{n-1}^{\ell_n^{\prime 1}}$, one derives

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n^{\prime 1}) \in (N)^3} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X}_{\mathbf{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{\prime[2]}}, \ell_{n-1}^{\prime[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
\leq &\mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n^{\prime 1}) \in [N]^3} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X}_{\mathbf{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n^{\prime 1}} = \ell_{n-1}^{\prime 1}\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
\leq &\left(\frac{N}{N-1}\right)^2 \mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n^{\prime 1}) \in [N]^3} m(\mathbf{X}_{\mathbf{n-1}})(G_{n-1})^3 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n^{\prime 1}} = \ell_{n-1}^{\prime 1}\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
= &\left(\frac{N}{N-1}\right)^2 \sum_{(\ell_n^{[2]}, \ell_n^{\prime 1}) \in (N)^3} \mathbf{E}\left[m(\mathbf{X}_{\mathbf{n-1}})(G_{n-1}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n^1} = \ell_{n-1}^1\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\qquad \mathbf{E}\left[m(\mathbf{X}_{\mathbf{n-1}})(G_{n-1}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n^2} = b_{n-1}\ell_{n-1}^1 + (1-b_{n-1})\ell_{n-1}^2\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\qquad \mathbf{E}\left[m(\mathbf{X}_{\mathbf{n-1}})(G_{n-1}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n^{\prime 1}} = \ell_{n-1}^{\prime 1}\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right]
\end{aligned}$$

Again, combined with (62), one deduces

$$\mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n^{\prime 1}) \in (N)^3} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X}_{\mathbf{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{\prime[2]}}, \ell_{n-1}^{\prime[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq 1^3 \times \left(\frac{N}{N-1}\right)^2.$$

Considering the number of assignments, we obtain

$$\mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n^{\prime[2]}) \in ((N))^{\times 2} \cap [N]_3^4} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X}_{\mathbf{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{\prime[2]}}, \ell_{n-1}^{\prime[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq 4 \times \left(\frac{N}{N-1}\right)^2.$$

(iii) Case: $(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4$:

In this case, all the random variables

$$\left(A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n'^1}, A_{n-1}^{\ell_n'^2}\right)$$

are distinct. Similarly, by the conditional independence of $\mathbf{A_{n-1}}$ and the bound given in (62), one gets

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2 \lambda^b_{n-1}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda^b_{n-1}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
\leq &\left(\frac{N}{N-1}\right)^2 \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4} \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)^4 \lambda^b_{n-1}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
= &\left(\frac{N}{N-1}\right)^2 \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4} \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n^1} = \ell_{n-1}^1\right\}} \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\qquad \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n^2} = b_{n-1}\ell_{n-1}^1 + (1-b_{n-1})\ell_{n-1}^2\right\}} \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\qquad \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n'^1} = \ell_{n-1}'^1\right\}} \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\qquad \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1}) \mathbf{1}_{\left\{A_{n-1}^{\ell_n'^2} = b_{n-1}\ell_{n-1}'^1 + (1-b_{n-1})\ell_{n-1}'^2\right\}} \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
\leq &1^4 \times \left(\frac{N}{N-1}\right)^2
\end{aligned}$$

Combining the three cases discussed above, we safely deduce that

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N))^{\times 2}} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2 \lambda^b_{n-1}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda^b_{n-1}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\qquad \leq \underbrace{(2+4+1)}_{=7} \times \left(\frac{N}{N-1}\right)^2. \quad a.s.
\end{aligned}$$

Now, let us go back to (61), by taking expectation on both sides, we have, for $N \geq 4$,

$$\begin{aligned}
\mathbf{E}\left[\Gamma^{\ddagger,b}_{n,N}(1)^2 \mathbf{1}_{\tau_N \geq n}\right] &\leq 7\left(\frac{N}{N-1}\right)^2 \mathbf{E}\left[\Gamma^{\ddagger,b}_{n-1,N}(1)^2 \mathbf{1}_{\tau_N \geq n-1}\right] \\
&\leq \frac{112}{9} \mathbf{E}\left[\Gamma^{\ddagger,b}_{n-1,N}(1)^2 \mathbf{1}_{\tau_N \geq n-1}\right] < +\infty.
\end{aligned}$$

This closes the proof of Proposition C.7. $\square$

Proposition C.8. *For any coalescence indicator $b \in \{0,1\}^{n+1}$, we have*

$$\widetilde{\Gamma}^{\dagger,b}_{n,N}(1)\mathbf{1}_{\tau_N \geq n} = \mathcal{O}_{\mathbb{L}^2}(1).$$

In particular, for any test function $F \in \mathcal{B}_b(E_n^2)$, we also have

$$\widetilde{\Gamma}^{\dagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} = \mathcal{O}_{\mathbb{L}^1}(1).$$

Remark. Before starting the proof of Proposition C.8, we would like to mention that the techniques involved are similar but a little bit different from the ones in the proof of Proposition C.7. First, $\widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}$ is not nonnegative in general except for the case where G_n is indicator function for all $n \geq 0$. In addition, it is not obvious that

$$\widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}(\ell^{[2]}_{n:n+1}, \mathbf{B_n}, \mathbf{X_n})\mathbf{1}_{\tau_N \geq n}$$

is bounded almost surely. In fact, it is easy prove that it is a.s. upper bounded by 3. However, there is no obvious reason that this term is a.s. lower bounded. This leads to the fact that

$$\left|\widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}(\ell^{[2]}_{n:n+1}, \mathbf{B_n}, \mathbf{X_n})\right| \mathbf{1}_{\tau_N \geq n}$$

is not a.s. bounded in general, which is the main difficult part in the following technical results. Hence, unlike the previous case shown in the proof of Proposition C.7, one should be extremely careful where dealing with the bound associated to the term

$$\left|\widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}(\ell^{[2]}_{n:n+1}, \mathbf{B_n}, \mathbf{X_n})\right| \lambda^{(\emptyset)}_n(A_n^{\ell^{[2]}_{n+1}}, \ell^{[2]}_n)\mathbf{1}_{\tau_N \geq n}.$$

Therefore, we introduce the following Lemma in order to facilitate the proof.

Lemma C.6. *For any nonnegative real numbers $a, b, c \in \mathbf{R}$, if $a \leq b$, $b > 0$ and $c \geq 0$, then, we have*

$$\frac{a}{b} \leq \frac{a+c}{b+c}.$$

Proof. Direct calculation gives

$$\frac{a+c}{b+c} - \frac{a}{b} = \frac{ab + bc - ab - ac}{b(b+c)} = \frac{(b-a)c}{b(b+c)} \geq 0.$$

The conclusion follows. □

Lemma C.7. *For any $\ell^{[2]}_n$ and $\ell^{[2]}_{n+1} \in (N)^2$, we have*

$$B_n^{\ell^1_{n+1}}(1 - B_n^{\ell^2_{n+1}})m(\mathbf{X_n})(G_n)\frac{G_n(X_n^{\ell^1_n})m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))}{\sum_{k \neq \ell^1_n}\left(1 - G_n(X_n^k)\right)/N},$$

and

$$B_n^{\ell^2_{n+1}}(1 - B_n^{\ell^1_{n+1}})m(\mathbf{X_n})(G_n)\frac{G_n(X_n^{\ell^2_n})m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))}{\sum_{k \neq \ell^2_n}\left(1 - G_n(X_n^k)\right)/N}$$

are both well-defined on the event $\{\tau_N \geq n\}$.

Proof. By symmetry of the definition, we only show that

$$B_n^{\ell^1_{n+1}}(1 - B_n^{\ell^2_{n+1}})m(\mathbf{X_n})(G_n)\frac{G_n(X_n^{\ell^1_n})m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))}{\sum_{k\neq\ell^1_n}\left(1 - G_n(X_n^k)\right)/N}$$

is always well-defined on the event $\{\tau_N \geq n\}$ with the convention (1). In fact, when

$$\sum_{k\neq\ell^1_n}\left(1 - G_n(X_n^k)\right) = 0,$$

we have $1 - G_n(X_n^{\ell^2_{n+1}}) = 0$ on the event $\{\tau_N \geq n\}$. By definition of $B_n^{\ell^2_{n+1}}$, we have

$$1 - B_n^{\ell^2_{n+1}} = 0.$$

This implies that, on the event $\{\tau_N \geq n\}$, we have

$$\begin{aligned}
&B_n^{\ell^1_{n+1}}(1 - B_n^{\ell^2_{n+1}})m(\mathbf{X_n})(G_n)\frac{G_n(X_n^{\ell^1_n})m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))}{\sum_{k\neq\ell^1_n}\left(1 - G_n(X_n^k)\right)/N}\\
=&B_n^{\ell^1_{n+1}}(1 - B_n^{\ell^2_{n+1}})m(\mathbf{X_n})(G_n)\frac{G_n(X_n^{\ell^1_n})m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))}{\sum_{k\neq\ell^1_n}\left(1 - G_n(X_n^k)\right)/N}\mathbf{1}_{\left\{\sum_{k\neq\ell^1_n}\left(1-G_n(X_n^k)\right)>0\right\}}.
\end{aligned}$$

The conclusion follows. □

Lemma C.8. *For any $\ell_n^{[2]} \in (N)^2$, and for any coalescence indicator $b' \in \{0,1\}^{n+1}$ and $b_n \in \{0,1\}$, we have almost surely*

$$\sup_{N>1}\mathbf{E}\left[\sum_{\ell^{[2]}_{n+1}\in[N]^2}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell^{[2]}_{n:n+1},\mathbf{B_n},\mathbf{X_n})\right|\lambda_n^{b'}(A_n^{\ell^{[2]}_{n+1}},\ell_n^{[2]})\mathbf{1}_{\tau_N\geq n}\;\middle|\;\mathcal{W}_n^N\right] < +\infty.$$

In particular, we have

$$\sup_{N>1}\mathbf{E}\left[\sum_{\ell^{[2]}_{n+1}\in(N)^2}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell^{[2]}_{n:n+1},\mathbf{B_n},\mathbf{X_n})\right|\lambda_n^{(\emptyset)}(A_n^{\ell^{[2]}_{n+1}},\ell_n^{[2]})\mathbf{1}_{\tau_N\geq n}\;\middle|\;\mathcal{W}_n^N\right] < +\infty.$$

Remark. We remark that in the definition of $\widetilde{\Gamma}^{\dagger,b}_{n,N}$, we do not need to investigate $\lambda_n^{b'}$ for a different coalescence indicator b' in general. The reason that b' is not set to be $(\emptyset)$ lies in the fact that Lemma C.8 is applied in the proof of Lemma C.9.

Proof. When not mentioned, the calculations of the random variables is only valid on the event $\{\tau_N \geq n\}$. Recall that, given $\ell_n^{[2]} \in (N)^2$,

$$\lambda_n^{b'}(A_n^{\ell^{[2]}_{n+1}},\ell_n^{[2]}) = \mathbf{1}_{\left\{A_n^{\ell^1_{n+1}}=\ell^1_n\right\}}\mathbf{1}_{\left\{A_n^{\ell^2_{n+1}}=b'_n\ell^1_n+(1-b'_n)\ell^2_n\right\}},$$

and given $\overline{\mathcal{W}}_n^N$, one has

$$\forall \ell\in[N],\quad A_n^\ell \sim B_n^\ell\delta_\ell(\cdot) + (1 - B_n^\ell)\sum_{k=1}^N\frac{G_n(X_n^k)}{Nm(\mathbf{X_n})(G_n)}\delta_k(\cdot).$$

By definition, one has

$$
\begin{aligned}
\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| \leq & B_n^{\ell_{n+1}^1} B_n^{\ell_{n+1}^2} m(\mathbf{X_n})(G_n^2) \\
&+ B_n^{\ell_{n+1}^1}(1-B_n^{\ell_{n+1}^2}) m(\mathbf{X_n})(G_n) \frac{\left|G_n(X_n^{\ell_n^1}) m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))\right|}{\sum_{k\neq\ell_n^1}\left(1-G_n(X_n^k)\right)/N} \\
&+ B_n^{\ell_{n+1}^2}(1-B_n^{\ell_{n+1}^1}) m(\mathbf{X_n})(G_n) \frac{\left|G_n(X_n^{\ell_n^2}) m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))\right|}{\sum_{k\neq\ell_n^2}\left(1-G_n(X_n^k)\right)/N},
\end{aligned}
$$

whence

$$
\begin{aligned}
&\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| \\
\leq & B_n^{\ell_{n+1}^1} B_n^{\ell_{n+1}^2} m(\mathbf{X_n})(G_n^2) + \frac{2B_n^{\ell_{n+1}^1}(1-B_n^{\ell_{n+1}^2}) m(\mathbf{X_n})(G_n)}{\sum_{k\neq\ell_n^1}\left(1-G_n(X_n^k)\right)/N} + \frac{2B_n^{\ell_{n+1}^2}(1-B_n^{\ell_{n+1}^1}) m(\mathbf{X_n})(G_n)}{\sum_{k\neq\ell_n^2}\left(1-G_n(X_n^k)\right)/N}.
\end{aligned}
$$

Since $B_n^{\ell_n^1}$ and $B_n^{\ell_n^2}$ are both $\{0,1\}$-valued, we deduce that

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_{n+1}^{[2]}\in[N]^2} \left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| \lambda_n^{b'}(A_n^{\ell_{n+1}^{[2]}},\ell_n^{[2]}) \mathbf{1}_{\tau_N\geq n} \,\middle|\, \overline{\mathcal{W}}_n^N\right] \\
\leq & B_n^{\ell_n^1} B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2} m(\mathbf{X_n})(G_n^2)\mathbf{1}_{\tau_N\geq n} \\
&+ B_n^{\ell_n^1}\sum_{\ell_{n+1}^2=1}^N (1-B_n^{\ell_{n+1}^2}) \frac{2G_n(X_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2})}{\sum_{k\neq\ell_n^1} 1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n} \\
&+ B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2}\sum_{\ell_{n+1}^1=1}^N (1-B_n^{\ell_{n+1}^1}) \frac{2G_n(X_n^{\ell_n^1})}{\sum_{k\neq\ell_n^2} 1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n}.
\end{aligned}
$$

By applying Lemma C.6 and considering the convention (1), we have

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_{n+1}^{[2]}\in[N]^2} \left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| \lambda_n^{b'}(A_n^{\ell_{n+1}^{[2]}},\ell_n^{[2]}) \mathbf{1}_{\tau_N\geq n} \,\middle|\, \overline{\mathcal{W}}_n^N\right] \\
\leq & B_n^{\ell_n^1} B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2} m(\mathbf{X_n})(G_n^2)\mathbf{1}_{\tau_N\geq n} \\
&+ B_n^{\ell_n^1}\sum_{\ell_{n+1}^2=1}^N (1-B_n^{\ell_{n+1}^2}) \frac{2G_n(X_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2}) + 1 - G_n(X_n^{\ell_n^1})}{\sum_{k=1}^N 1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n} \\
&+ B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2}\sum_{\ell_{n+1}^1=1}^N (1-B_n^{\ell_{n+1}^1}) \frac{2G_n(X_n^{\ell_n^1}) + 1 - G_n(X_n^{\ell_n^2})}{\sum_{k=1}^N 1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n}.
\end{aligned}
$$

The simple fact $0 = B_n^k(1-B_n^k)\mathbf{1}_{\tau_N\geq n} \leq G_n(X_n^k)(1-G_n(X_n^k)\mathbf{1}_{\tau_N\geq n})$ for any $k\in[N]$ yields

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_{n+1}^{[2]}\in[N]^2}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|\lambda_n^{b'}(A_n^{\ell_{n+1}^{[2]}},\ell_n^{[2]})\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_n^N\right]\\
&\leq\left(G_n(X_n^{\ell_n^1})G_n(X_n^{\ell_n^2})\vee G_n(X_n^{\ell_n^1})\right)m(\mathbf{X_n})(G_n^2)\mathbf{1}_{\tau_N\geq n}\\
&\quad+G_n(X_n^{\ell_n^1})\sum_{\ell_{n+1}^2=1}^{N}(1-G_n(X_n^{\ell_{n+1}^2}))\frac{2G_n(X_n^{b_n'\ell_n^1+(1-b_n')\ell_n^2})+1-G_n(X_n^{\ell_n^1})}{\sum_{k=1}^N 1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n}\\
&\quad+G_n(X_n^{b_n'\ell_n^1+(1-b_n')\ell_n^2})\sum_{\ell_{n+1}^1=1}^{N}(1-G_n(X_n^{\ell_{n+1}^1}))\frac{2G_n(X_n^{\ell_n^1})+1-G_n(X_n^{\ell_n^2})}{\sum_{k=1}^N 1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n}\\
&=\left(G_n(X_n^{\ell_n^1})m(\mathbf{X_n})(G_n^2)+4G_n(X_n^{\ell_n^1})G_n(X_n^{b_n'\ell_n^1+(1-b_n')\ell_n^2})+2\right)\mathbf{1}_{\tau_N\geq n}\leq 7.\quad a.s.
\end{aligned}$$

By definition, since

$$\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,0}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| = \mathbf{G}_{\mathbf{n}}^{\ddagger}(\mathbf{X_n}) + \frac{1}{N-1}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|,$$

the analysis for the case $b_n = 0$ is the combination of the case $b_n = 1$ and the similar reasoning in (65), namely, a direct consequence of (62). This terminates the proof of Lemma C.8.

□

Lemma C.9. *For any $\ell_n^{[2]}\in(N)^2$ and for any coalescence indicator $b\in\{0,1\}^{n+1}$, we have almost surely*

$$\forall\ell_{n+1}^2\in[N],\quad\sup_{N>1}\mathbf{E}\left[\sum_{\ell_{n+1}^1\in[N]}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|\mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1}\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_n^N\right]<+\infty,\quad(69)$$

as well as

$$\forall\ell_{n+1}^1\in[N],\quad\sup_{N>1}\mathbf{E}\left[\sum_{\ell_{n+1}^2\in[N]}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|\mathbf{1}_{A_n^{\ell_{n+1}^2}=\ell_n^2}\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_n^N\right]<+\infty.$$

Proof. By symmetry of the definition of $\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})$, it suffices to verify (69). In fact, by simple observation, one has

$$\mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1}=\mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1}\mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1}\leq\mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1}\sum_{\ell_{n+1}^2=1}^{N}\mathbf{1}_{A_n^{\ell_{n+1}^2}=\ell_n^1}.\quad a.s.$$

Therefore, one only needs to show that

$$\sup_{N>1}\mathbf{E}\left[\sum_{\ell_{n+1}^{[2]}\in[N]^2}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|\lambda_n^{(n)}(A_n^{\ell_{n+1}^{[2]}},\ell_n^{[2]})\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_n^N\right]<+\infty,$$

which, by taking $b' = (n)$, is guaranteed by Lemma C.8. This terminates the proof of Lemma C.9. □

Proof of Proposition C.8. Now, we start the proof by induction. It is trivial for the step $n = 0$ as $\widetilde{\Gamma}^{\dagger,b}_{0,N}(1) = 1$. For step $n \geq 1$, we suppose that

$$\sup_{N>0} \mathbf{E}\left[\widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)^2 \mathbf{1}_{\tau_N \geq n-1}\right] < +\infty.$$

By the alternative representation (60), for all $N \geq 4$, we have

$$\begin{aligned}
&\mathbf{E}\left[\widetilde{\Gamma}^{\dagger,b}_{n,N}(1)^2 \mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{W}^N_{n-1}\right] \\
&= \sum_{(\ell^{[2]}_{n-1},\ell'^{[2]}_{n-1}) \in ((N))^{\times 2}} \widetilde{\Lambda}^{\dagger,b}_{n-1}[\ell^{[2]}_{n-1}]\widetilde{\Lambda}^{\dagger,b}_{n-1}[\ell'^{[2]}_{n-1}]\mathbf{1}_{\tau_N \geq n-1} \\
&\quad \mathbf{E}\Bigg[\sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n}) \in ((N)^2)^{\times 2}} \widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell'^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \\
&\qquad\qquad \lambda^{(\varnothing)}_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})\lambda^{(\varnothing)}_{n-1}(A^{\ell'^{[2]}_n}_{n-1}, \ell'^{[2]}_{n-1}) \,\Bigg|\, \mathcal{W}^N_{n-1}\Bigg],
\end{aligned} \tag{70}$$

since by definition of $\widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}$, one has

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n}) \in ((N)^2)^{\times 2}} \widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell'^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \\
&\qquad\qquad \lambda^{(\varnothing)}_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})\lambda^{(\varnothing)}_{n-1}(A^{\ell'^{[2]}_n}_{n-1}, \ell'^{[2]}_{n-1}) \,\Bigg|\, \mathcal{W}^N_{n-1}\Bigg]\mathbf{1}_{\tau_N = n-1} = 0.
\end{aligned}$$

As we have mentioned several times, in the following part of the proof, we omit the notation $\mathbf{1}_{\tau_N \geq n-1}$. Similar as in the proof of Proposition C.7, the rest of the reasoning relies on the decomposition (63).

(i) Case: $(\ell^{[2]}_n, \ell'^{[2]}_n) \in \left((N)^2\right)^{\times 2} \cap [N]^4_2$:
In this case, there are only two distinct random variables in the tuple

$$\left(A^{\ell^1_n}_{n-1}, A^{\ell^2_n}_{n-1}, A^{\ell'^1_n}_{n-1}, A^{\ell'^2_n}_{n-1}\right).$$

As we have already mentioned in (64), there are 2 possible assignments such that we can fix two distinct random variables within the tuple above. Without loss of generality, we suppose that $A^{\ell^1_n}_{n-1}$ and $A^{\ell'^1_n}_{n-1}$ are two distinct random variables, and we fix one of these two assignments. Then, when (ℓ^1_n, ℓ'^1_n) varies freely in $(N)^2$, the values of $A^{\ell^2_n}_{n-1}$ and $A^{\ell'^2_n}_{n-1}$ will be a.s. determined by the chosen assignment and the values of $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n'^1}$.

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right.\\
&\qquad\left.\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\overline{\mathcal{W}}_{n-1}^N\right]\\
=&\mathbf{E}\left[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right.\\
&\qquad\left.\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\overline{\mathcal{W}}_{n-1}^N\right]
\end{aligned}\tag{71}$$

Given $\overline{\mathcal{W}}_{n-1}^N$, by the conditional independence of $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n'^1}$ under the chosen assignment, we have

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right.\\
&\qquad\left.\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\overline{\mathcal{W}}_{n-1}^N\right]\\
\leq&\mathbf{E}\left[\sum_{\ell_n^1\in[N]}\left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right|\mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1\right\}}\,\middle|\,\overline{\mathcal{W}}_{n-1}^N\right]\\
&\mathbf{E}\left[\sum_{\ell_n'^1\in[N]}\left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right|\mathbf{1}_{\left\{A_{n-1}^{\ell_n'^1}=\ell_{n-1}'^1\right\}}\,\middle|\,\overline{\mathcal{W}}_{n-1}^N\right].
\end{aligned}\tag{72}$$

Then, by applying Lemma C.9, one gets

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right.\\
&\qquad\left.\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]=\mathcal{O}_{a.s.}(1).
\end{aligned}$$

Since the number of different assignments 2 does not depend on N, one deduces that

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right.\\
&\qquad\left.\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]=\mathcal{O}_{a.s.}(1).
\end{aligned}$$

(ii) Case: $(\ell_n^{[2]},\ell_n'^{[2]})\in\left((N)^2\right)^{\times 2}\cap[N]_3^4$:

In this case, there are three distinct random variables in the tuple

$$\left(A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n'^1}, A_{n-1}^{\ell_n'^2}\right).$$

As is calculated in (68), the number of assignment is 4 at this time. Let us fix one assignment. We suppose that $A_{n-1}^{\ell_n^{[2]}}$ and $A_{n-1}^{\ell_n'^1}$ are three distinct random variables. Then, whilst $(\ell_n^{[2]}, \ell_n'^1)$ varies freely in $(N)^3$, the value of $A_{n-1}^{\ell_n'^2}$ is a.s. determined by the chosen assignment and the values of $(A_{n-1}^{\ell_n^{[2]}}, A_{n-1}^{\ell_n'^1})$. Given $\overline{\mathcal{W}}_{n-1}^N$, by the conditional independence between $A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}$ and $A_{n-1}^{\ell_n'^1}$, one derives

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell'^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \\
&\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell'^{[2]}_{n-1}) \,\Bigg|\, \overline{\mathcal{W}}_{n-1}^N\Bigg] \\
=&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^1)\in(N)^3} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell'^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \\
&\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell'^{[2]}_{n-1}) \,\Bigg|\, \overline{\mathcal{W}}_{n-1}^N\Bigg] \\
\leq&\mathbf{E}\Bigg[\sum_{\ell_n^{[2]}\in(N)^2} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\right| \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \,\Bigg|\, \overline{\mathcal{W}}_{n-1}^N\Bigg] \\
&\mathbf{E}\Bigg[\sum_{\ell_n'^1\in[N]} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell'^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\right| \mathbf{1}_{\{A_{n-1}^{\ell_n'^1}=\ell'^1_{n-1}\}} \,\Bigg|\, \overline{\mathcal{W}}_{n-1}^N\Bigg].
\end{aligned} \tag{73}$$

Thanks to Lemma C.8 and Lemma C.9, we get

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^1)\in(N)^3} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell'^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \\
&\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell'^{[2]}_{n-1}) \,\Bigg|\, \mathcal{W}_{n-1}^N\Bigg] = \mathcal{O}_{a.s.}(1).
\end{aligned}$$

Again, since the number of the different assignments 4 does not depend on N, we obtain

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell'^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\right| \\
&\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell'^{[2]}_{n-1}) \,\Bigg|\, \mathcal{W}_{n-1}^N\Bigg] = \mathcal{O}_{a.s.}(1).
\end{aligned} \tag{74}$$

(iii) Case: $(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4$:

In this case, all the random variables

$$\left(A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n'^1}, A_{n-1}^{\ell_n'^2}\right)$$

are distinct. This time, Lemma C.8 gives directly

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in(N)^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \right.\\
&\qquad\qquad\qquad\qquad \left. \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \;\middle|\; \mathcal{W}_{n-1}^N\right] \\
\leq &\mathbf{E}\left[\sum_{\ell_n^{[2]}\in[N]^2} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\right| \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \;\middle|\; \mathcal{W}_{n-1}^N\right] \\
&\mathbf{E}\left[\sum_{\ell_n'^{[2]}\in[N]^2} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\right| \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \;\middle|\; \mathcal{W}_{n-1}^N\right] = \mathcal{O}_{a.s.}(1).
\end{aligned} \tag{75}$$

Combining the three cases above, we finally obtain

$$\begin{aligned}
\sup_{N>0} \mathbf{E}&\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \right.\\
&\qquad\qquad \left. \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \;\middle|\; \mathcal{W}_{n-1}^N\right] < +\infty. \quad a.s.
\end{aligned}$$

Returning to (70) and induction hypothesis, the verification of step n is then finished, so as the proof of Proposition C.8.

□

Proposition C.9. *For any test function $F \in \mathcal{B}_b(E_n^2)$, we have*

$$\mathbf{E}\left[\Gamma_{n,N}^{\ddagger,b}(F)\mathbf{1}_{\tau_N\geq n} \;\middle|\; \mathcal{G}_{n-1}^N\right] = \Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}C_{b_n}(F)\mathbf{1}_{\tau_N\geq n-1}.$$

Proof. First, by the alternative representation we have introduced in (57), we have

$$\begin{aligned}
&\mathbf{E}\left[\Gamma_{n,N}^{\ddagger,b}(F)\mathbf{1}_{\tau_N\geq n} \;\middle|\; \mathcal{G}_{n-1}^N\right] \\
=&\sum_{(\ell_{n-1}^{[2]})\in(N)^2} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\mathbf{1}_{\tau_N\geq n-1} \\
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \;\middle|\; \mathcal{G}_{n-1}^N\right],
\end{aligned} \tag{76}$$

Thus, it suffices to show that

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \;\middle|\; \mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n-1} \\
=&\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}})\mathbf{1}_{\tau_N\geq n-1}.
\end{aligned} \tag{77}$$

Alghough it may seem unnecessary, we recall that we have almost surely

$$\begin{aligned}\frac{N-1}{N}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}}) =& B^{\ell^1_n}_{n-1}B^{\ell^2_n}_{n-1}m(\mathbf{X_{n-1}})(G_{n-1})^2\\ &+ B^{\ell^1_n}_{n-1}(1-B^{\ell^2_n}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})^2\\ &+ B^{\ell^2_n}_{n-1}(1-B^{\ell^1_n}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})^2\\ &+ (1-B^{\ell^1_n}_{n-1})(1-B^{\ell^2_n}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})^2.\end{aligned}$$

Given $\overline{\mathcal{W}}^N_{n-1}$, for any $\ell \in [N]$, the definition of the Feynman-Kac IPS gives

$$(A^\ell_{n-1}, X^\ell_n) \sim B^\ell_{n-1}\delta_\ell(dA^\ell_{n-1})\dot{M}_n(X^\ell_{n-1}, dX^\ell_n)+(1-B^\ell_{n-1})\sum_{k=1}^N \frac{\delta_k(dA^\ell_{n-1})\mathring{Q}_n(X^k_{n-1}, dX^\ell_n)}{Nm(\mathbf{X_{n-1}})(G_{n-1})}. \quad (78)$$

(i) Case $b_{n-1}=0$:
Notice that

$$\begin{aligned}&\frac{N-1}{N}\mathbf{E}\left[\sum_{\ell^{[2]}_n\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell^{[2]}_n}_n)\,\middle|\,\overline{\mathcal{W}}^N_{n-1}\right]\\ =& B^{\ell^1_{n-1}}_{n-1}B^{\ell^2_{n-1}}_{n-1}m^{\otimes 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})\dot{M}^{\otimes 2}_n C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\ &+\frac{1}{N}B^{\ell^1_{n-1}}_{n-1}\sum_{\ell^2_n\neq\ell^1_{n-1}}(1-B^{\ell^2_n}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})(\dot{M}_n\otimes\mathring{Q}_n)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\ &+\frac{1}{N}B^{\ell^2_{n-1}}_{n-1}\sum_{\ell^1_n\neq\ell^2_{n-1}}(1-B^{\ell^1_n}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})(\mathring{Q}_n\otimes\dot{M}_n)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\ &+\frac{1}{N^2}\sum_{\ell^{[2]}_n\in(N)^2}(1-B^{\ell^1_n}_{n-1})(1-B^{\ell^2_n}_{n-1})\mathring{Q}^{\otimes 2}_n C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1}),\end{aligned}$$

which yields

$$\begin{aligned}&\mathbf{E}\left[\sum_{\ell^{[2]}_n\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell^{[2]}_n}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\\ =&\frac{N}{N-1}m^{\otimes 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})\dot{Q}^{\otimes 2}_n C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\ &+\frac{1}{N-1}\sum_{\ell^2_n\neq\ell^1_{n-1}}(1-G_{n-1}(X^{\ell^2_n}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})(\dot{Q}_n\otimes\mathring{Q}_n)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\ &+\frac{1}{N-1}\sum_{\ell^1_n\neq\ell^2_{n-1}}(1-G_{n-1}(X^{\ell^1_n}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})(\mathring{Q}_n\otimes\dot{Q}_n)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\ &+\frac{1}{N(N-1)}\sum_{\ell^{[2]}_n\in(N)^2}(1-G_{n-1}(X^{\ell^1_n}_{n-1}))(1-G_{n-1}(X^{\ell^2_n}_{n-1}))\mathring{Q}^{\otimes 2}_n C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1}),\end{aligned} \quad (79)$$

First, by decomposition (2), we noticed that

$$\frac{N}{N-1}m^{\otimes 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1}) = \left(m^{\odot 2}(\mathbf{X_n})(G^{\otimes 2}_{n-1}) + \frac{1}{N-1}m(\mathbf{X_{n-1}})(G^2_{n-1}).\right) \quad (80)$$

Then, we deduce that

$$
\begin{aligned}
&\frac{1}{N-1}\sum_{\ell_n^2\neq \ell_{n-1}^1}(1-G_{n-1}(X_{n-1}^{\ell_n^2}))m(\mathbf{X_{n-1}})(G_{n-1})\\
=&\frac{N}{N-1}m^{\otimes 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&-\frac{1}{N-1}(1-G_{n-1}(X_{n-1}^{\ell_{n-1}^1}))m(\mathbf{X_{n-1}})(G_{n-1})\\
=&m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&+\frac{1}{N(N-1)}\sum_{k=1}^{N}\left(G_{n-1}(X_{n-1}^{\ell_{n-1}^1})-G_{n-1}(X_{n-1}^k)\right)G_{n-1}(X_{n-1}^k)\\
=&m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&+\frac{1}{N-1}\left(G_{n-1}(X_{n-1}^{\ell_{n-1}^1})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)\right).
\end{aligned}
\tag{81}
$$

The exactly same manipulations also give

$$
\begin{aligned}
&\frac{1}{N-1}\sum_{\ell_n^1\neq \ell_{n-1}^2}(1-G_{n-1}(X_{n-1}^{\ell_n^1}))m(\mathbf{X_{n-1}})(G_{n-1})\\
=&m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&+\frac{1}{N-1}\left(G_{n-1}(X_{n-1}^{\ell_{n-1}^2})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)\right).
\end{aligned}
\tag{82}
$$

Now, let us put (80), (81) and (82) back into (79). One derives

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{W}_{n-1}^N\right]\\
=&\Bigg\{\Big[m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}^{\otimes 2})\dot{Q}_n^{\otimes 2}+m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\dot{Q}_n\otimes\mathring{Q}_n\\
&+m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\mathring{Q}_n\otimes\dot{Q}_n+m^{\odot 2}(\mathbf{X_{n-1}})((1-G_{n-1})^{\otimes 2})\mathring{Q}_n^{\otimes 2}\Big]\\
&+\frac{1}{N-1}\Big[m(\mathbf{X_{n-1}})(G_{n-1})\left((G_{n-1}\times\dot{Q}_n)\otimes\mathring{Q}_n+\mathring{Q}_n\otimes(G_{n-1}\times\dot{Q}_n)\right)\\
&+m(\mathbf{X_{n-1}})(G_{n-1}^2)\left(\dot{Q}_n^{\otimes 2}-\dot{Q}_n\otimes\mathring{Q}_n-\mathring{Q}_n\otimes\dot{Q}_n\right)\Big]\Bigg\}(C_{b_n}(F))(X_{n-1}^{\ell_{n-1}^{[2]}}),
\end{aligned}
\tag{83}
$$

which, by definition, turns out to be the following equality:

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{W}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n-1}\\
=&\left(\mathbf{Q}_{\hat{n}}^0+\frac{1}{N-1}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,0}\right)(C_{b_n}(F))\,(X_{n-1}^{\ell_{n-1}^{[2]}}).
\end{aligned}
\tag{84}
$$

(ii) Case $b_{n-1}=1$:

By the similar calculations done in the previous case, one has

$$
\begin{aligned}
&\frac{N-1}{N}\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_\mathbf{n}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \overline{\mathcal{W}}_{n-1}^{N}\right]\\
&= 0\\
&\quad + \frac{1}{N}B_{n-1}^{\ell_{n-1}^1}\sum_{\ell_n^2\neq\ell_{n-1}^1}(1-B_{n-1}^{\ell_n^2})m(\mathbf{X_{n-1}})(G_{n-1})(\dot{M}_n\otimes\mathring{Q}_n)C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{(1,1)}})\\
&\quad + \frac{1}{N}B_{n-1}^{\ell_{n-1}^1}\sum_{\ell_n^1\neq\ell_{n-1}^1}(1-B_{n-1}^{\ell_n^1})m(\mathbf{X_{n-1}})(G_{n-1})(\mathring{Q}_n\otimes\dot{M}_n)C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{(1,1)}})\\
&\quad + \frac{1}{N^2}\sum_{\ell_n^{[2]}\in(N)^2}(1-B_{n-1}^{\ell_n^1})(1-B_{n-1}^{\ell_n^2})\mathring{Q}_n^{\otimes 2}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{(1,1)}}),
\end{aligned}
$$

which yields

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_\mathbf{n-1}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^{N}\right]\\
&= 0\\
&\quad + \frac{1}{N-1}\sum_{\ell_n^2\neq\ell_{n-1}^1}(1-G_{n-1}(X_{n-1}^{\ell_n^2}))m(\mathbf{X_{n-1}})(G_{n-1})C_1\left(\dot{Q}_n\otimes\mathring{Q}_n\right)C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}})\\
&\quad + \frac{1}{N-1}\sum_{\ell_n^1\neq\ell_{n-1}^1}(1-G_{n-1}(X_{n-1}^{\ell_n^1}))m(\mathbf{X_{n-1}})(G_{n-1})C_1\left(\mathring{Q}_n\otimes\dot{Q}_n\right)C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}})\\
&\quad + \frac{1}{N(N-1)}\sum_{\ell_n^{[2]}\in(N)^2}(1-G_{n-1}(X_{n-1}^{\ell_n^1}))(1-G_{n-1}(X_{n-1}^{\ell_n^2}))C_1\mathring{Q}_n^{\otimes 2}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}}).
\end{aligned}
\tag{85}
$$

Then, taking into account the equality (81), one obtains a similar equation as (83):

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_\mathbf{n-1}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^{N}\right]\mathbf{1}_{\tau_N\geq n-1}\\
&=\Bigg\{\Bigg[m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}^{\otimes 2})C_1\dot{Q}_n^{\otimes 2} + m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))C_1\left(\dot{Q}_n\otimes\mathring{Q}_n\right)\\
&\quad + m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))C_1\left(\mathring{Q}_n\otimes\dot{Q}_n\right) + m^{\odot 2}(\mathbf{X_{n-1}})((1-G_{n-1})^{\otimes 2})C_1\mathring{Q}_n^{\otimes 2}\Bigg]\\
&\quad - m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}^{\otimes 2})C_1\dot{Q}_n^{\otimes 2}\\
&\quad + \frac{1}{N-1}\Bigg[m(\mathbf{X_{n-1}})(G_{n-1})C_1\left((G_{n-1}\times\dot{Q}_n)\otimes\mathring{Q}_n + \mathring{Q}_n\otimes(G_{n-1}\times\dot{Q}_n)\right)\\
&\quad - m(\mathbf{X_{n-1}})(G_{n-1}^2)C_1\left(\dot{Q}_n\otimes\mathring{Q}_n + \mathring{Q}_n\otimes\dot{Q}_n\right)\Bigg]\Bigg\}(C_{b_n}(F))(X_{n-1}^{\ell_{n-1}^{[2]}})\mathbf{1}_{\tau_N\geq n-1}.
\end{aligned}
\tag{86}
$$

By definition, we finally obtain that on the event $\{\tau_N \geq n-1\}$,

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2 \lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}^N_{n-1}\right] \\
&= \left(\mathbf{Q}^1_{\hat{n}} - m^{\odot 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})\mathbf{1}_{\tau_N \geq n-1} C_1 \dot{\mathbf{Q}}^{\otimes}_n + \frac{1}{N-1}\check{\mathbf{Q}}^{\dagger,1}_{\hat{n}}\right)\left(C_{b_n}(F)\right)(X^{\ell^{[2]}_{n-1}}_{n-1}) \qquad (87)\\
&= \left(\mathbf{Q}^{\dagger,1}_{\hat{n}} + \frac{1}{N-1}\check{\mathbf{Q}}^{\dagger,1}_{\hat{n}}\right)\left(C_{b_n}(F)\right)(X^{\ell^{[2]}_{n-1}}_{n-1}).
\end{aligned}$$

Combining the two cases, we conclude that we have proved that

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2 \lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N \geq n-1} \qquad (88)\\
&= \mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}}\left(C_{b_n}(F)\right)(X^{\ell^{[2]}_{n-1}}_{n-1})\mathbf{1}_{\tau_N \geq n-1}. \quad a.s.
\end{aligned}$$

In addition, since

$$\mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}}\left(C_{b_n}(F)\right)(X^{\ell^{[2]}_{n-1}}_{n-1})\mathbf{1}_{\tau_N \geq n-1}$$

is $\mathcal{G}^N_{n-1}$-measurable, the verification of (77) is then finished, so as the proof of Proposition C.9. □

Proposition C.10. *For any test function $F \in \mathcal{B}_b(E_n^2)$, we have*

$$\mathbf{E}\left[\widetilde{\Gamma}^{\dagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{W}^N_{n-1}\right] = \widetilde{\Gamma}^{\dagger,b}_{n-1,N}\widetilde{\mathbf{Q}}^{\dagger,b_{n-1}}_{\hat{n}} C_{b_n}(F)\mathbf{1}_{\tau_N \geq n-1}.$$

Proof. By combining Proposition C.9 and the fact that

$$\begin{aligned}
&\widetilde{\mathbf{G}}^{\dagger,0}_{\mathbf{n}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\lambda^{(\emptyset)}_{n-1}(A^{\ell_n^{[2]}}_{n-1}, \ell^{[2]}_{n-1}) \\
&= \left(\mathbf{G}^{\ddagger}_{\mathbf{n}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) - \frac{1}{N-1}\widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\right)\lambda^{(\emptyset)}_{n-1}(A^{\ell_n^{[2]}}_{n-1}, \ell^{[2]}_{n-1}), \; a.s.
\end{aligned}$$

it suffices to check the case $b_{n-1} = 1$. Similar as in the proof of Proposition C.9, we consider the alternative representation (60), which gives

$$\begin{aligned}
&\mathbf{E}\left[\widetilde{\Gamma}^{\dagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{W}^N_{n-1}\right] \\
&= \sum_{(\ell^{[2]}_{n-1})\in(N)^2} \widetilde{\Lambda}^{\dagger,b}_{n-1}[\ell^{[2]}_{n-1}]\mathbf{1}_{\tau_N \geq n-1} \qquad (89)\\
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\lambda^{(\emptyset)}_{n-1}(A^{\ell_n^{[2]}}_{n-1}, \ell^{[2]}_{n-1})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}^N_{n-1}\right],
\end{aligned}$$

Thus, it suffices to show that

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\lambda^{(\emptyset)}_{n-1}(A^{\ell_n^{[2]}}_{n-1}, \ell^{[2]}_{n-1})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N \geq n-1} \qquad (90)\\
&= \widetilde{\mathbf{Q}}^{\dagger,1}_{\hat{n}} C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\mathbf{1}_{\tau_N \geq n-1}.
\end{aligned}$$

We omit the notation $\mathbf{1}_{\tau_N \geq n-1}$ in the rest of the proof. Recall that

$$
\begin{aligned}
&\widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \\
=&B^{\ell^1_n}_{n-1} B^{\ell^2_n}_{n-1} m(\mathbf{X_{n-1}})(G^2_{n-1}) \\
&+ B^{\ell^1_n}_{n-1}(1 - B^{\ell^2_n}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) \frac{G_{n-1}(X^{\ell^1_{n-1}}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) - m(\mathbf{X_{n-1}})(G^2_{n-1})}{\sum_{k \neq \ell^1_{n-1}} G_{n-1}(X^k_{n-1})} \\
&+ B^{\ell^2_n}_{n-1}(1 - B^{\ell^1_n}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) \frac{G_{n-1}(X^{\ell^2_{n-1}}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) - m(\mathbf{X_{n-1}})(G^2_{n-1})}{\sum_{k \neq \ell^2_{n-1}} G_{n-1}(X^k_{n-1})}.
\end{aligned}
$$

Hence, with the definition (78), standard calculation gives

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell^{[2]}_n \in (N)^2} \widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \lambda^{(\emptyset)}_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1}) \,\middle|\, \overline{\mathcal{W}}^N_{n-1} \right] \\
=&\Bigg\{ B^{\ell^1_{n-1}}_{n-1} B^{\ell^2_{n-1}}_{n-1} m(\mathbf{X_{n-1}})(G^2_{n-1}) \dot{M}^{\otimes 2}_n \\
&+ B^{\ell^1_{n-1}}_{n-1} \sum_{\ell^2_n \neq \ell^1_{n-1}} (1 - B^{\ell^2_n}_{n-1}) \frac{G_{n-1}(X^{\ell^1_{n-1}}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) - m(\mathbf{X_{n-1}})(G^2_{n-1})}{\sum_{k \neq \ell^1_{n-1}} (1 - G_{n-1}(X^k_{n-1}))} \dot{M}_n \otimes \mathring{M}_n \\
&+ B^{\ell^2_{n-1}}_{n-1} \sum_{\ell^1_n \neq \ell^2_{n-1}} (1 - B^{\ell^1_n}_{n-1}) \frac{G_{n-1}(X^{\ell^2_{n-1}}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) - m(\mathbf{X_{n-1}})(G^2_{n-1})}{\sum_{k \neq \ell^2_{n-1}} (1 - G_{n-1}(X^k_{n-1}))} \mathring{M}_n \otimes \dot{M}_n \\
&\Bigg\} C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1}),
\end{aligned}
$$

whence

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell^{[2]}_n \in (N)^2} \widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \lambda^{(\emptyset)}_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1}) \,\middle|\, \mathcal{W}^N_{n-1} \right] \\
=&\Bigg\{ m(\mathbf{X_{n-1}})(G^2_{n-1}) \dot{Q}^{\otimes 2}_n \\
&+ \left(G_{n-1}(X^{\ell^1_{n-1}}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) - m(\mathbf{X_{n-1}})(G^2_{n-1}) \right) \dot{Q}_n \otimes \mathring{Q}_n \\
&+ \left(G_{n-1}(X^{\ell^2_{n-1}}_{n-1}) m(\mathbf{X_{n-1}})(G_{n-1}) - m(\mathbf{X_{n-1}})(G^2_{n-1}) \right) \mathring{Q}_n \otimes \dot{Q}_n \\
&\Bigg\} C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1}),
\end{aligned}
$$

which yields

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell^{[2]}_n \in (N)^2} \widetilde{\mathbf{G}}^{\dagger,1}_{\mathbf{n}}(\ell^{[2]}_{n-1:n}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \lambda^{(\emptyset)}_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1}) \,\middle|\, \mathcal{W}^N_{n-1} \right] \mathbf{1}_{\tau_N \geq n-1} \\
=&\Bigg\{ m(\mathbf{X_{n-1}})(G_{n-1}) \left((G_{n-1} \times \dot{Q}_n) \otimes \mathring{Q}_n + \mathring{Q}_n \otimes (G_{n-1} \times \dot{Q}_n) \right) \\
&+ m(\mathbf{X_{n-1}})(G^2_{n-1}) \left(\dot{Q}^{\otimes 2}_n - \dot{Q}_n \otimes \mathring{Q}_n - \mathring{Q}_n \otimes \dot{Q}_n \right) \Bigg\} C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1}) \mathbf{1}_{\tau_N \geq n-1}.
\end{aligned} \tag{91}
$$

By definition, this gives the desired equality (90).

$\square$

Lemma C.10. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescent indicator $b \in \{0,1\}^{n+1}$, we have*

$$\Gamma^{\ddagger,b}_{n,N}(F)\mathbf{1}_{\tau_N \geq n} - \Gamma^{\ddagger,b}_{n-1,N}\mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}} C_{b_n}(F)\mathbf{1}_{\tau_N \geq n-1} = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right).$$

Proof. Thanks to Cauchy-Schwartz inequality and Proposition C.9, it suffices to verify that

$$\mathbf{E}\left[\Gamma^{\ddagger,b}_{n,N}(F)^2\mathbf{1}_{\tau_N \geq n} - \Gamma^{\ddagger,b}_{n-1,N}\mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}} C_{b_n}(F)^2\mathbf{1}_{\tau_N \geq n-1}\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

Next, thanks to the alternative representation (57), one derives that

$$\begin{aligned}
&\mathbf{E}\left[\Gamma^{\ddagger,b}_{n,N}(F)^2\mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&= \sum_{(\ell^{[2]}_{n-1},\ell'^{[2]}_{n-1})\in((N)^2)^{\times 2}} \Lambda^{\ddagger,b}_{n-1}[\ell^{[2]}_{n-1}]\Lambda^{\ddagger,b}_{n-1}[\ell'^{[2]}_{n-1}]\mathbf{1}_{\tau_N \geq n-1} \\
&\mathbf{E}\left[\sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n})\in((N)^2)^{\times 2}} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X}_{\mathbf{n-1}})^2\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell'^{[2]}_n}_{n-1},\ell'^{[2]}_{n-1})C_{b_n}(F)^{\otimes 2}(X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right].
\end{aligned}$$

To simplify the notation, we denote

$$R_1(N) := \sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n})\in((N)^2)^{\times 2}\setminus(N)^4} \mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell^{[2]}_{n}]\Lambda^{\ddagger,b}_{n}[\ell'^{[2]}_{n}]C_{b_n}(F)^{\otimes 2}(X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right]\mathbf{1}_{\tau_N \geq n},$$

and

$$\begin{aligned}
R_2(N) := \sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n})\in((N)^2)^{\times 2}\setminus(N)^4} &\mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell^{[2]}_{n}]C_{b_n}(F)(X^{\ell^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell'^{[2]}_{n}]C_{b_n}(F)(X^{\ell'^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right]\mathbf{1}_{\tau_N \geq n},
\end{aligned}$$

as well as

$$\begin{aligned}
R_3(N) := \sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n})\in((N)^2)^{\times 2}} &\mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell^{[2]}_{n}]C_{b_n}(F)(X^{\ell^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell'^{[2]}_{n}]C_{b_n}(F)(X^{\ell'^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right]\mathbf{1}_{\tau_N \geq n}
\end{aligned}$$

Thanks to the conditional independence between $A^{\ell^{[2]}_n}_{n-1}$ and $A^{\ell'^{[2]}_n}_{n-1}$ given $\mathcal{G}^N_{n-1}$, we have

$$\begin{aligned}
&R_3(N) - R_2(N) \\
&= \sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n})\in(N)^4} \mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell^{[2]}_{n}]C_{b_n}(F)(X^{\ell^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right]\mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell'^{[2]}_{n}]C_{b_n}(F)(X^{\ell'^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right]\mathbf{1}_{\tau_N \geq n} \\
&= \sum_{(\ell^{[2]}_{n},\ell'^{[2]}_{n})\in(N)^4} \mathbf{E}\left[\Lambda^{\ddagger,b}_{n}[\ell^{[2]}_{n}]\Lambda^{\ddagger,b}_{n}[\ell'^{[2]}_{n}]C_{b_n}(F)^{\otimes 2}(X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right]\mathbf{1}_{\tau_N \geq n},
\end{aligned}$$

whence we deduce that

$$\mathbf{E}\left[\Gamma^{\ddagger,b}_{n,N}(F)^2\mathbf{1}_{\tau_N\ge n}\,\middle|\,\mathcal{G}^N_{n-1}\right]=R_1(N)-R_2(N)+R_3(N).\quad a.s.$$

In addition, since one of the by-product (88) in the proof of Proposition C.9 yields

$$\begin{aligned}
&\left(\mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}}C_{b_n}(F)\right)^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n'^{[2]}})\\
&=\sum_{(\ell_n^{[2]})\in((N)^2)^{\times 2}}\mathbf{E}\left[\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
&\qquad\sum_{(\ell_n'^{[2]})\in((N)^2)^{\times 2}}\mathbf{E}\left[\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n'^{[2]}}_{n-1},\ell'^{[2]}_{n-1})C_{b_n}(F)(X_n^{\ell_n'^{[2]}})\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
&=\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}}\mathbf{E}\left[\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{G}^N_{n-1}\right]\\
&\qquad\qquad\mathbf{E}\left[\mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n'^{[2]}}_{n-1},\ell'^{[2]}_{n-1})C_{b_n}(F)(X_n^{\ell_n'^{[2]}})\,\middle|\,\mathcal{G}^N_{n-1}\right],
\end{aligned}$$

one has

$$R_3(N)=\Gamma^{\ddagger,b}_{n-1,N}\mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}}C_{b_n}(F)^2\mathbf{1}_{\tau_N\ge n-1},\quad a.s.$$

which guarantees that

$$\mathbf{E}\left[R_3(N)\right]=\mathbf{E}\left[\Gamma^{\ddagger,b}_{n-1,N}\mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}}C_{b_n}(F)^2\mathbf{1}_{\tau_N\ge n-1}\right].$$

Therefore, it suffices to verify that

$$\mathbf{E}\left[R_1(N)\right]=\mathcal{O}\left(\frac{1}{N}\right)\quad\text{and}\quad\mathbf{E}\left[R_2(N)\right]=\mathcal{O}\left(\frac{1}{N}\right).\tag{92}$$

Together, we prove both of the two convergence above by induction, as the proofs share the same mechanism. Without loss of generality, we suppose that $F\equiv 1$. For $n=0$, standard calculations give

$$R_1(N)=R_2(N)=\frac{4N-6}{N(N-1)}=\mathcal{O}\left(\frac{1}{N}\right).$$

For $n\ge 1$, we suppose that

$$\mathbf{E}\left[\sum_{(\ell^{[2]}_{n-1},\ell'^{[2]}_{n-1})\in((N)^2)^{\times 2}\setminus(N)^4}\Lambda^{\ddagger,b}_{n-1}[\ell^{[2]}_{n-1}]\Lambda^{\ddagger,b}_{n-1}[\ell'^{[2]}_{n-1}]\mathbf{1}_{\tau_N\ge n-1}\right]=\mathcal{O}\left(\frac{1}{N}\right).$$

Now, it is time to go back to the decomposition (63). As is mentioned for many times, we may omit the notation $\mathbf{1}_{\tau_N\ge n-1}$.

(i) Case: $(\ell_n^{[2]},\ell_n'^{[2]})\in\left((N)^2\right)^{\times 2}\cap[N]^4_2$:

As we have seen in the proof of Proposition C.7, there are 2 different assignments

$$A^{\ell_n^1}_{n-1}=A^{\ell_n'^1}_{n-1},\ A^{\ell_n^2}_{n-1}=A^{\ell_n'^2}_{n-1}\quad\text{and}\quad A^{\ell_n^1}_{n-1}=A^{\ell_n'^2}_{n-1},\ A^{\ell_n^2}_{n-1}=A^{\ell_n'^1}_{n-1},$$

such that two distinct random variables can be found in

$$\left(A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n'^1}, A_{n-1}^{\ell_n'^2}\right).$$

Let us fix one assignment. This time, in order to execute a finer analysis, we suppose that ℓ_n^1 and $\ell_n'^1$ vary freely in $(N)^2$ and the values of $A_{n-1}^{\ell_n^2}$ and $A_{n-1}^{\ell_n'^2}$ will be almost surely determined by the values of $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n'^1}$. Since the potential function $\mathbf{G}_\mathbf{n}^\ddagger$ and indicator functions are both nonnegative, we extend $(N)^2$ to $[N]^2$, which gives the following inequality:

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^\ddagger(\mathbf{X}_{\mathbf{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
\leq &\mathbf{E}\left[\sum_{\ell_n^1\in[N]} \mathbf{G}_{\mathbf{n-1}}^\ddagger(\mathbf{X}_{\mathbf{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\mathbf{E}\left[\sum_{\ell_n'^1\in[N]} \mathbf{G}_{\mathbf{n-1}}^\ddagger(\mathbf{X}_{\mathbf{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}.
\end{aligned} \tag{93}$$

Now, we explain why there is an indicator function

$$\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}$$

at the r.h.s. above. For the case $b_{n-1} = 0$, one has

$$\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) = \mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1\neq\ell_{n-1}^2=A_{n-1}^{\ell_n^2}\right\}}\mathbf{1}_{\left\{A_{n-1}^{\ell_n'^1}=\ell_{n-1}'^1\neq\ell_{n-1}'^2=A_{n-1}^{\ell_n'^2}\right\}}.$$

Hence, since $(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2}\backslash(N)^4$, we have

$$\#\{\ell_{n-1}^1, \ell_{n-1}^2, \ell_{n-1}'^1, \ell_{n-1}'^2\} = \#\{A_{n-1}^{\ell_{n-1}^1}, A_{n-1}^{\ell_{n-1}^2}, A_{n-1}^{\ell_{n-1}'^1}, A_{n-1}^{\ell_{n-1}'^2}\} \leq \#\{\ell_n^1, \ell_n^2, \ell_n'^1, \ell_n'^2\} < 4.$$

At the same time, when $b_{n-1} = 1$, one has

$$\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) = \mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1=A_{n-1}^{\ell_n^2}\neq\ell_{n-1}^2\right\}}\mathbf{1}_{\left\{A_{n-1}^{\ell_n'^1}=\ell_{n-1}'^1=A_{n-1}^{\ell_n'^2}\neq\ell_{n-1}'^2\right\}}.$$

Since $\ell_n^1 \neq \ell_n^2$ and $\ell_n'^1 \neq \ell_n'^2$, we have

$$\exists i, j \in \{1, 2\}, \quad s.t. \quad \ell_n^i = \ell_n'^j.$$

In no matter which case mentioned above, it is necessary that $\ell_{n-1}^1 = \ell_{n-1}'^1$. Therefore, one also gets

$$\#\{\ell_{n-1}^1, \ell_{n-1}^2, \ell_{n-1}'^1, \ell_{n-1}'^2\} < 4.$$

The arguments above will be applied repeatedly in the rest of the proof. Next, by the same procedure in the proof of Proposition C.7, we obtain

$$\begin{aligned}
\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \mathbf{G}_{\mathbf{n-1}}^\ddagger(\mathbf{X}_{\mathbf{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]& \\
\leq C_1 \times \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}&,
\end{aligned}$$

where C_1 denotes positive constant which does not depend on N. Meanwhile, by the same procedure, we also have

$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2$$
$$\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq C_1' \times \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

where C_1' is also a positive constant which does not depend on N.

(ii) Case: $(\ell_n^{[2]},\ell_n'^{[2]}) \in \left((N)^2\right)^{\times 2} \cap [N]_3^4$:
Again, similart in the proof of Proposition C.7, there are 4 different assignments such that there are 3 distinct random variables within

$$\left(A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n'^1}, A_{n-1}^{\ell_n'^2}\right).$$

Let us fix one assignment. As is done in the previous case, to conduct a finer study, we suppose that $(\ell_n^{[2]},\ell_n'^1)$ vary freely in $(N)^3$ and the values of $A_{n-1}^{\ell_n'^2}$ will be almost surely determined by the choice of assignment and the values of $A_{n-1}^{\ell_n^{[2]}}$ and $A_{n-1}^{\ell_n'^1}$. This time, by extending $(N)^3$ to $(N)^2 \cup [N]$, we get

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^3} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \right.\\
&\qquad\left. \lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) C_{b_n}(F)^{\otimes 2}(X_n^{\ell_n^{[2]}}, X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\\
\leq& \mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\\
&\mathbf{E}\left[\sum_{\ell_n'^1\in[N]} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}}) \lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) C_{b_n}(F)(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\\
&\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}.
\end{aligned} \tag{94}$$

Again, by similar argument given in the proof of Proposition C.7, we get

$$\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]$$
$$\leq C_2 \times \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

and

$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2$$
$$\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq C_2' \times \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

where C_2 and C_2' are positive constant which does not depend on N.

Combining both cases, one gets

$$
\begin{aligned}
&R_1(N)\\
&\leq \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}\\
&= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1'), \quad a.s.
\end{aligned}
$$

and

$$
\begin{aligned}
&R_2(N)\\
&\leq \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}\\
&= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2'). \quad a.s.
\end{aligned}
$$

The desired convergence (92) are therefore guaranteed by the induction hypothesis by taking the expectation on both sides of the inequalities above. This is the end of the proof of Lemma C.10.

□

Lemma C.11. *For any test function $F\in\mathcal{B}_b(E_n^2)$ and any coalescent indicator $b\in\{0,1\}^{n+1}$, we have*

$$
\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)\mathbf{1}_{\tau_N\geq n}-\widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)\mathbf{1}_{\tau_N\geq n-1}=\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right).
$$

Proof. Before starting, we mention that this proof bears a resemblance to the one of Lemma C.10. Thanks to Cauchy-Schwartz inequality, it is sufficient to verify that

$$
\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)^2\mathbf{1}_{\tau_N\geq n}-\widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)^2\mathbf{1}_{\tau_N\geq n-1}\right]=\mathcal{O}\left(\frac{1}{N}\right).
$$

Again, similar to the equation (70), by the alternative representation (60) and decomposition (59), we deduce that

$$
\begin{aligned}
&\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)^2\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_{n-1}^N\right]\\
&=\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}\\
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\\
&\qquad\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})F^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n'^{[2]}})\,\Bigg|\,\mathcal{W}_{n-1}^N\Bigg].
\end{aligned}
$$

Similar to the previous case, we denote

$$
R_1(N):=\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4}\mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}]\widetilde{\Lambda}_n^{\dagger,b}[\ell_n'^{[2]}]F^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n'^{[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n},
$$

and

$$R_2(N) := \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \setminus (N)^4} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}] F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n'^{[2]}] F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \mathbf{1}_{\tau_N \geq n},$$

as well as

$$R_3(N) := \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2}} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}] F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n'^{[2]}] F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \mathbf{1}_{\tau_N \geq n}$$

Thanks to the conditional independence between $A_{n-1}^{\ell_n^{[2]}}$, $B_{n-1}^{\ell_n^{[2]}}$ and $A_{n-1}^{\ell_n'^{[2]}}$, $B_{n-1}^{\ell_n'^{[2]}}$ given $\mathcal{W}_{n-1}^N$, we have

$$\begin{aligned}
&R_3(N) - R_2(N) \\
&= \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}] F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n'^{[2]}] F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \mathbf{1}_{\tau_N \geq n} \\
&= \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}] \widetilde{\Lambda}_n^{\dagger,b}[\ell_n'^{[2]}] F^{\otimes 2}(X_n^{\ell_n^{[2]}}, X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \mathbf{1}_{\tau_N \geq n},
\end{aligned}$$

from which we get

$$\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)^2 \mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{W}_{n-1}^N \right] = R_1(N) - R_2(N) + R_3(N). \quad a.s.$$

Notice that Proposition C.10 gives, on the event $\{\tau_N \geq n-1\}$,

$$\begin{aligned}
&\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2}} \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \\
&\qquad \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N \right] \\
&= \left(\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F) \right)^{\otimes 2} (X_{n-1}^{\ell_{n-1}^{[2]}}, X_{n-1}^{\ell_{n-1}'^{[2]}}).
\end{aligned}$$

Hence, one has

$$R_3(N) = \widetilde{\Gamma}_{n-1,N}^{\dagger,b} \widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)^2 \mathbf{1}_{\tau_N \geq n-1}, \quad a.s.$$

which guarantees that

$$\mathbf{E}[R_3(N)] = \mathbf{E}\left[\widetilde{\Gamma}_{n-1,N}^{\dagger,b} \widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)^2 \mathbf{1}_{\tau_N \geq n-1} \right].$$

Therefore, it suffices to verify that

$$\mathbf{E}[R_1(N)] = \mathcal{O}\left(\frac{1}{N}\right) \quad \text{and} \quad \mathbf{E}[R_2(N)] = \mathcal{O}\left(\frac{1}{N}\right). \tag{95}$$

Without loss of generality, we suppose that $F \equiv 1$. The rest of the proof is done by induction. For $n = 0$, by definition, we have

$$R_1(N) = R_2(N) = \frac{4N-6}{N(N-1)} = \mathcal{O}\left(\frac{1}{N}\right).$$

For $n \geq 1$, we suppose that

$$\mathbf{E}\left[\sum_{(\ell^{[2]}_{n-1},\ell'^{[2]}_{n-1})\in((N)^2)^{\times 2}\setminus(N)^4} \widetilde{\Lambda}^{\dagger,b}_{n-1}[\ell^{[2]}_{n-1}]\widetilde{\Lambda}^{\dagger,b}_{n-1}[\ell'^{[2]}_{n-1}]\mathbf{1}_{\tau_N \geq n-1}\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

As is stated many times, we may omit the notation $\mathbf{1}_{\tau_N\geq n-1}$ since it is $\mathcal{W}^N_{n-1}$-measurable. Once again, let us return to the decomposition (63). Since the essential idea is highly repetitive w.r.t. the reasoning in the proof of Proposition C.10, we skip some of the unnecessary details in the rest of the proof.

(i) Case: $(\ell^{[2]}_n, \ell'^{[2]}_n) \in \left((N)^2\right)^{\times 2} \cap [N]^4_2$:
By the same procedure given in the proof of Proposition C.8 and Lemma C.10, we obtain

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell^{[2]}_n,\ell'^{[2]}_n)\in((N)^2)^{\times 2}\cap[N]^4_2} \widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell'^{[2]}_{n-1:n},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right.\\
&\qquad\left.\lambda^{(\emptyset)}_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})\lambda^{(\emptyset)}_{n-1}(A^{\ell'^{[2]}_n}_{n-1},\ell'^{[2]}_{n-1})F^{\otimes 2}(X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n)\,\middle|\, \mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N\geq n-1}\\
&\leq C_1\mathbf{1}_{\{\#\{\ell^1_{n-1},\ell^2_{n-1},\ell'^1_{n-1},\ell'^2_{n-1}<4\}\}},
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{(\ell^{[2]}_n,\ell'^{[2]}_n)\in((N)^2)^{\times 2}\cap[N]^4_2} \mathbf{E}\left[\widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda^{(\emptyset)}_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})\,\middle|\,\mathcal{W}^N_{n-1}\right]\\
&\qquad\mathbf{E}\left[\widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell'^{[2]}_{n-1:n},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda^{(\emptyset)}_{n-1}(A^{\ell'^{[2]}_n}_{n-1},\ell'^{[2]}_{n-1})\,\middle|\,\mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N\geq n-1}\\
&\leq C'_1\mathbf{1}_{\{\#\{\ell^1_{n-1},\ell^2_{n-1},\ell'^1_{n-1},\ell'^2_{n-1}<4\}\}},
\end{aligned}$$

where C_1 and C'_1 are some constant that does not depend on N.

(ii) Case: $(\ell^{[2]}_n, \ell'^{[2]}_n) \in \left((N)^2\right)^{\times 2} \cap [N]^4_3$:
This time, we get

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell^{[2]}_n,\ell'^{[2]}_n)\in((N)^2)^{\times 2}\cap[N]^4_3} \widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell^{[2]}_{n-1:n},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell'^{[2]}_{n-1:n},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right.\\
&\qquad\left.\lambda^{(\emptyset)}_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})\lambda^{(\emptyset)}_{n-1}(A^{\ell'^{[2]}_n}_{n-1},\ell'^{[2]}_{n-1})F^{\otimes 2}(X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n)\,\middle|\, \mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N\geq n-1}\\
&\leq C_2\mathbf{1}_{\{\#\{\ell^1_{n-1},\ell^2_{n-1},\ell'^1_{n-1},\ell'^2_{n-1}<4\}\}},
\end{aligned}$$

and

$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\emptyset)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]$$
$$\mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\emptyset)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n-1}$$
$$\leq C_2'\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

where C_2 and C_2' are some constant that does not depend on N.

By combining the both cases, we establish that

$$\begin{aligned}R_1(N)\leq&\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}}\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}\\ =&\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4}\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1'),\quad a.s.\end{aligned}$$

and

$$\begin{aligned}R_2(N)\leq&\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}}\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}\\ =&\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4}\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2').\quad a.s.\end{aligned}$$

By taking the expectation on both sides of the inequalities above, the desired convergence (92) are then verified thanks to the induction hypothesis. The conclusion follows. □

Lemma C.12. *For any test function $F\in\mathcal{B}_b(E_n)^{\otimes 2}$ and any coalescent indicator $b\in\{0,1\}^{n+1}$, we have*

$$\Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}C_{b_n}(F)-\Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{n}^{\ddagger,b_{n-1}}C_{b_n}(F)=\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right),$$

as well as

$$\widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)-\widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{n}^{\dagger,b_{n-1}}(F)=\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right).$$

Proof. First, we noticed that for any test functions $F_1,F_2\in\mathcal{B}_b(E_n)^{\otimes 2}$, Minkowski's inequality gives

$$\begin{aligned}&\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_1+F_2)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1+F_2)\right|_{\mathbb{L}^1}\right\|\\ \leq&\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1)\right|\right\|_{\mathbb{L}^1}\\ &+\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_2)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_2)\right|\right\|_{\mathbb{L}^1}.\end{aligned}\tag{96}$$

Second, thanks to Cauchy-Schwartz inequality, Proposition C.5 and Proposition C.7, we deduce that

$$\begin{aligned}&\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1)\right|\right\|_{\mathbb{L}^1}\\ \leq&\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\right\|_{\mathbb{L}^2}\left\|(\eta_{n-1}^N)^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1)\right\|_{\mathbb{L}^2}\\ =&\mathcal{O}\left(\frac{1}{\sqrt{N}}\right).\end{aligned}\tag{97}$$

Similarly, thanks to Proposition C.8, we also have

$$
\begin{aligned}
&\left\| \widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N \geq n-1} \left| (\eta^N_{n-1})^{\odot 2}(F_1+F_2)\mathbf{1}_{\tau_N \geq n-1} - \eta^{\otimes 2}_{n-1}(F_1+F_2) \right| \right\|_{\mathbb{L}^1} \\
\leq \quad & \left\| \widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N \geq n-1} \left| (\eta^N_{n-1})^{\odot 2}(F_1)\mathbf{1}_{\tau_N \geq n-1} - \eta^{\otimes 2}_{n-1}(F_1) \right| \right\|_{\mathbb{L}^1} \\
+ & \left\| \widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N \geq n-1} \left| (\eta^N_{n-1})^{\odot 2}(F_2)\mathbf{1}_{\tau_N \geq n-1} - \eta^{\otimes 2}_{n-1}(F_2) \right| \right\|_{\mathbb{L}^1} .
\end{aligned}
\tag{98}
$$

and

$$
\begin{aligned}
&\left\| \widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N \geq n-1} \left| (\eta^N_{n-1})^{\odot 2}(F_1)\mathbf{1}_{\tau_N \geq n-1} - \eta^{\otimes 2}_{n-1}(F_1) \right| \right\|_{\mathbb{L}^1} \\
\leq & \left\| \widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N \geq n-1} \right\|_{\mathbb{L}^2} \left\| (\eta^N_{n-1})^{\odot 2}(F_1)\mathbf{1}_{\tau_N \geq n-1} - \eta^{\otimes 2}_{n-1}(F_1) \right\|_{\mathbb{L}^2} \\
= & \mathcal{O}\left(\frac{1}{\sqrt{N}} \right).
\end{aligned}
\tag{99}
$$

Finally, let us go back to two pairs of decompositions (76), (86) and (89), (91), the boundness of G_{n-1} and the homogeneous structure in these two decompositions allow us to apply respectively (96), (97) and (98), (99). Note that

$$
m(\mathbf{X_{n-1}})(G_{n-1}) = m^{\odot 2}(\mathbf{X_{n-1}})(1 \otimes G_{n-1}),
$$

and

$$
m(\mathbf{X_{n-1}})(G^2_{n-1}) = m^{\odot 2}(\mathbf{X_{n-1}})(1 \otimes G^2_{n-1}).
$$

The desired $\mathbb{L}^1$-bound can therefore be obtained with some standard algebraic manipulations.

□

D Partial $\mathcal{R}$-algebra structure for transition operators

In the development of SMC framework, the Feynman-Kac semigroup plays a crucial role in the theoretic analysis: it provides the natural martingale or bias-martingale structure constructed by focusing on the local sampling errors at each level and/or by each particle. In the present work, it is also at the core of the proofs of technical results. However, in asymmetric SMC framework, we need to consider a new type of structure, such as the decompositions (24), (25) and (29). Alghough the coalescent Feynman-Kac measures and coalescent tree occupation measures is very different in terms of construction, they share the same algebraic structure. In the classic literature in the domain, the $\mathcal{R}$-algebra structure is regarded trivial: it is extremely easy to verify and no meaningful calculations are conducted according to this structure. However, in the variance related problem of asymmetric SMC framework, it is crucial to the construction of the variance estimators. Therefore, we think it is important to summary and justify this particular calculations, which is novel to the general framework in the SMC context. We start by the case where all the particles share a same state space, in order to illustrate the intuition of our construction.

D.1 $\mathcal{R}$-algebra structure on homogeneous state spaces

Let us assume that $E = E_0 = E_1 = \cdots = E_n$. First, we notice that $(\mathcal{B}_b(E), +, \times)$ admits a ring structure, where "+" and "×" use the usual convention, that is,

$$\forall f, g \in \mathcal{B}_b(E), \quad f + g : E \ni x \mapsto f(x) + g(x) \in \mathbf{R},$$

and

$$\forall f, g \in \mathcal{B}_b(E), \quad f \times g : E \ni x \mapsto f(x) \times g(x) \in \mathbf{R}.$$

Given a sequence of transition kernel $(M_n, n \geq 1)$, we consider the uniform finite transition kernels at step $n \geq 1$ defined by

$$\mathcal{Q}_n := \{g_{n-1}(x)M_n(x, dy) \mid g_{n-1} \in \mathcal{B}_b(E)\}.$$

It is readily checked that $\mathcal{Q}_n$ can be regarded a ring with $(+, \cdot)$. More precisely, the plus "+" is defined by

$$\forall \mu \in \mathcal{M}(E), \ \forall f \in \mathcal{B}_b(E), \quad \mu(Q_1 + Q_2)(f) := \mu Q_1(f) + \mu Q_2(f).$$

Fixing a time horizon $T \in \mathbf{N}^*$, we denote $\mathcal{Q}$ the ring generated by $\cup_{n=0}^T \mathcal{Q}_n$. Then, $(\mathcal{Q}, +, \cdot)$ allows a ring structure. Let us consider the ring $\mathcal{R} := \mathcal{B}_b(E)$, it is then natural to construct an $\mathcal{R}$-algebra (cf. Figure 3).

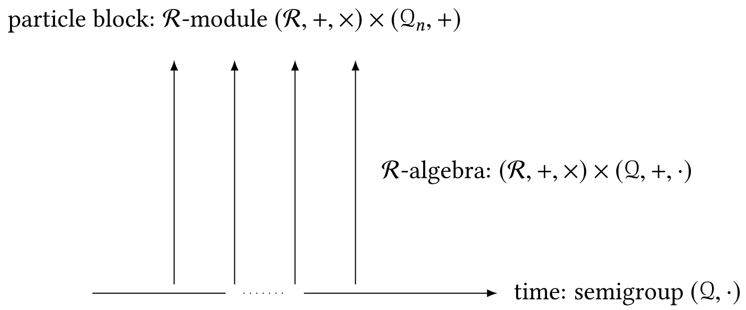

Figure 3: $\mathcal{R}$-algebra structure of Feynman-Kac kernels

In this article, the first family of decompositions (24) and (25) is obtained by taking $\mathcal{R} := \mathcal{B}_b(E)^{\otimes 2}$. We consider the $\mathcal{R}$-algebra structure on $\mathcal{Q}^{(2)}$, defined by the ring generated by $\cup_{n=0}^T \mathcal{Q}_n^{(2)}$, with

$$\mathcal{Q}_n^{(2)} := \mathcal{Q}_n^{\otimes 2} \cup C_1 \mathcal{Q}_n^{\otimes 2},$$

where

$$C_1 \mathcal{Q}_n^{\otimes 2} := \left\{ C_1 \mathbf{Q}_n \,\middle|\, \mathbf{Q}_n \in \mathcal{Q}_n^{\otimes 2} \right\}.$$

The similar argument can also be used to define, $\mathbf{P}$-almost surely, the $\mathcal{R}$-algebra for the random matrix with which we define the coalescent tree occupation measures for each $1 < N < +\infty$. The similar decomposition (29), which is valid $\mathbf{P}$-almost surely, can thus be regarded as the direct consequence of the $\mathcal{R}$-algebra homomorphism.

D.2 Partial $\mathcal{R}$-algebra structure

More generally, since the state spaces may vary w.r.t. time horizon n, the semigroup structure is then not intrinsic: one should consider the partial semigroup, in which the associated composition law is associative when it is compatible. Moreover, using the same idea, one may define the partial $\mathcal{R}$-algebra, which, roughly speaking, is an algebraic structure such that all the operations in an $\mathcal{R}$-algebra is valid when it is compatible.

References

[ADH10] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):269–342, 2010.

[ALV18] Christophe Andrieu, Anthony Lee, and Matti Vihola. Uniform ergodicity of the iterated conditional smc and geometric ergodicity of particle gibbs samplers. *Bernoulli*, 24(2):842–872, 05 2018.

[BGG$^+$16] Charles-Edouard Bréhier, Maxime Gazeau, Ludovic Goudenège, Tony Lelièvre, and Mathias Rousset. Unbiasedness of some generalized adaptive multilevel splitting algorithms. *Ann. Appl. Probab.*, 26(6):3559–3601, 2016.

[BJKT16] A. Beskos, A. Jasra, N. Kantas, and A. Thiery. On the convergence of adaptive sequential Monte Carlo methods. *Ann. Appl. Probab.*, 26(2):1111–1146, 2016.

[CDGR18] Frédéric Cérou, Bernard Delyon, Arnaud Guyader, and Mathias Rousset. On the asymptotic normality of adaptive multilevel splitting. *SIAM/ASA Journal on Uncertainty Quantification*, 7, 04 2018.

[CDMG11] F. Cérou, P. Del Moral, and A. Guyader. A nonasymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(3):629–649, 2011.

[Cho04] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6):2385–2411, 2004.

[CJP13] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. SMC^2: an efficient algorithm for sequential analysis of state space models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(3):397–426, 2013.

[CL13] H. P. Chan and T. L. Lai. A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.*, 41(6):2877–2904, 2013.

[Cog80] Robert Cogburn. Markov chains in random environments: The case of markovian environments. *Ann. Probab.*, 8(5):908–916, 10 1980.

[CRGP15] Nicolas Chopin, James Ridgway, Mathieu Gerber, and Omiros Papaspiliopoulos. Towards automatic calibration of the number of state particles within the SMC^2 algorithm. *ArXiv preprint 1506.00570*, Jun 2015.

[DCGR17] Bernard Delyon, Frédéric Cérou, Arnaud Guyader, and Mathias Rousset. A Central Limit Theorem for Fleming-Viot Particle Systems with Hard Killing. *arXiv e-prints*, page arXiv:1709.06771, Sep 2017.

[DdFG01] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice.* Statistics for Engineering and Information Science. Springer-Verlag, New York, 2001.

[DG19] Qiming Du and Arnaud Guyader. Variance estimation in adaptive sequential monte carlo. *arXiv*, 2019.

[DM04] P. Del Moral. *Feynman-Kac formulae: Genealogical and interacting particle systems with applications*. Probability and its Applications. Springer-Verlag, New York, 2004.

[DM08] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, 36(5):2344–2376, 2008.

[DM13] P. Del Moral. *Mean field simulation for Monte Carlo integration*, volume 126 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2013.

[DMDJ12] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. On adaptive resampling strategies for sequential monte carlo methods. *Bernoulli*, 18(1):252–278, 02 2012.

[DMKP16] P. Del Moral, R. Kohn, and F. Patras. On particle Gibbs samplers. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1687–1733, 2016.

[DMPR09] P. Del Moral, F. Patras, and S. Rubenthaler. Tree based functional expansions for Feynman-Kac particle models. *Ann. Appl. Probab.*, 19(2):778–825, 2009.

[GCW17] Mathieu Gerber, Nicolas Chopin, and Nick Whiteley. Negative association, ordering and convergence of resampling methods. *arXiv e-prints*, page arXiv:1707.01845, Jul 2017.

[GDM17] François Giraud and Pierre Del Moral. Nonasymptotic analysis of adaptive and annealed feynman–kac particle models. *Bernoulli*, 23(1):670–709, 02 2017.

[HSG06] J. D. Hol, T. B. Schon, and F. Gustafsson. On resampling algorithms for particle filters. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 79–82, Sep. 2006.

[Jan11] Svante Janson. Probability asymptotics: notes on notation. *arXiv e-prints*, page arXiv:1108.3924, Aug 2011.

[LW18] A. Lee and N. Whiteley. Variance estimation in the particle filter. *Biometrika*, 105(3):609–625, 2018.

[McL74] D. L. McLeish. Dependent central limit theorems and invariance principles. *Ann. Probability*, 2:620–628, 1974.

[OD19] Jimmy Olsson and Randal Douc. Numerically stable online estimation of variance in particle filters. *Bernoulli*, 25(2):1504–1535, 05 2019.