# VARIANCE ESTIMATION IN ADAPTIVE SEQUENTIAL MONTE CARLO

BY QIMING DU[1], AND ARNAUD GUYADER[2]

[1]*Sorbonne Université, LPSM & CERMICS, France, qiming.du@upmc.fr*

[2]*Sorbonne Université, LPSM & CERMICS, France, arnaud.guyader@upmc.fr*

Sequential Monte Carlo (SMC) methods represent a classical set of techniques to simulate a sequence of probability measures through a simple selection/mutation mechanism. However, the associated selection functions and mutation kernels usually depend on tuning parameters that are of first importance for the efficiency of the algorithm. A standard way to address this problem is to apply Adaptive Sequential Monte Carlo (ASMC) methods, which consist in exploiting the information given by the history of the sample to tune the parameters. This article is concerned with variance estimation in such ASMC methods. Specifically, we focus on the case where the asymptotic variance coincides with the one of the "limiting" Sequential Monte Carlo algorithm as defined by Beskos *et al.* [4]. We prove that, under natural assumptions, the estimator introduced by Lee and Whiteley [16] in the nonadaptive case (i.e., SMC) is also a consistent estimator of the asymptotic variance for ASMC methods. To do this, we introduce a new estimator that is expressed in terms of coalescent tree-based measures, and explain its connection with the previous one. Our estimator is constructed by tracing the genealogy of the associated Interacting Particle System. The tools we use connect the study of Particle Markov Chain Monte Carlo methods and the variance estimation problem in SMC methods. As such, they may give some new insights when dealing with complex genealogy-involved problems of Interacting Particle Systems in more general scenarios.

**1. Introduction.** Sequential Monte Carlo (SMC) methods are classical Monte Carlo techniques widely used in Bayesian inference, filtering, rare events simulations and many other fields (see for example [15] and references therein). The principle is to approximate a sequence of probability measures $(\eta_n)_{n \geq 0}$ by simulating an Interacting Particle System (IPS) via an importance sampling and resampling mechanism. The flow of measures is then approximated by the empirical version $(\eta_n^N)_{n \geq 0}$. A lot of convergence results when the sample size $N$ goes to infinity can be found in the literature (see for example [11, 12]).

In practice, when applying these SMC methods, it is also very important to have a control on the constructed estimators, such as confidence intervals. For this, if one has a CLT type theorem for the test function $f$ such as (see, e.g., [11, 10, 14])

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

it suffices to provide a consistent estimator $\sigma_n^N(f)$ of $\sigma_n(f)$ since Slutsky's lemma then ensures that

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

A natural way to achieve this aim is by resimulating the IPS independently many times and by estimating $\sigma_n(f)^2$ with the crude variance estimator. However, since a single run of the

1

algorithm may take a lot of time, this is usually intractable. In addition, as the estimator $\eta_n^N(f)$ of $\eta_n(f)$ provided by SMC is typically biased, it is also nontrivial to implement parallel computing for a large number of 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 first consistent estimator of this type was proposed by Chan and Lai [9], by using the ancestral information encoded in the genealogy of the associated IPS. Then, Lee and Whiteley [16] proposed an unbiased variance estimator for the unnormalized measures $\gamma_n^N$ and a term by term estimator, with insights on the genealogy of the IPS. Both estimators are studied in the classical SMC framework, meaning in a nonadaptive setting where the weight functions and the Markov proposal kernels are fixed a priori.

In this article, we deal with *adaptive* SMC methods. At each resampling step, the weight functions and/or Markov proposal kernels depend upon the history of the simulated process. The idea is to approximate an ideal "limiting" SMC algorithm, which is usually out of reach, by exploiting the induced information tracked by some summary statistics. Such approaches are expected to be more efficient and more automated than the nonadaptive ones since they require less user-specified tuning parameters.

Specifically, we are interested in the case where the adaptive SMC algorithm is asymptotically identical to a "limiting" SMC algorithm. More precisely, we expect the asymptotic variance of the adaptive SMC algorithm to be identical to the "ideal" nonadaptive one. This kind of stability property is at the core of the pair of articles [4] and [8]. The framework discussed in the present paper is just a slightly generalized version of the one presented in Section 2 in [4] but still ensures the stability property of their Theorem 2.3.

Another remark is about Adaptive Multilevel Splitting (AMS), also known as Subset Simulation, see for example [2, 3, 7, 5, 8]. This is a class of ASMC algorithms dedicated to rare event estimation and simulation. Despite the fact that our assumptions are not verified in the AMS framework, we expect that the variance estimator would also work in this context. Nonetheless, we believe that this case requires a specific analysis as well as different assumptions. To account for this, one can notice that the proofs in [8] and [4] differ in many points, although the take-home message is the same. In a nutshell, the main difficulty in the AMS framework comes from the indicator functions in the potential functions as well as in the transitions kernels, leading to severe regularity issues when dealing with CLT type results and asymptotic variances.

From a theoretical viewpoint, to prove the consistency of the variance estimator proposed in [16], we were not able to adapt their technical tools. This is due to the additional randomness brought by the weight functions and Markov kernels in the adaptive case. As a consequence, we propose to develop new techniques in order to estimate the terms $\Gamma_n^b$ that appear in the expansion of the variance given in [6]. The mains ideas are: first, our term by term estimator is consistent and, second, the difference between our estimator and the one of Lee and Whiteley goes to 0 in probability when the sample size $N$ goes to infinity. However, in practice, one uses the estimator proposed by Lee and Whiteley, which is computationally very simple, while the one we introduce here may be seen as a handy tool to prove the consistency of the former.

The construction of our estimators $\Gamma_{n,N}^b$ uses the idea of many-body Feynman-Kac models, which were designed in [13] to study propagation of chaos properties of Conditional Particle Markov Chain Monte Carlo methods [1]. Above the specific context of the present article, these connections may give some insights on how to deal with complex genealogy-involved problems in more general settings.

*Notation.* Before proceeding, let us provide some notation that will be of constant use in the following.

- For any Polish space $E$, we denote respectively by $\mathcal{M}(E)$, $\mathcal{M}_+(E)$ and $\mathcal{P}(E)$ the sets of signed finite measures, nonnegative finite measures, and probability measures on $E$ endowed with Borel $\sigma$-algebra $\mathcal{B}(E)$, while $\mathcal{B}_b(E)$ denotes the collection of the bounded measurable functions from $(E, \mathcal{B}(E))$ to $(\mathrm{R}, \mathcal{B}(\mathrm{R}))$ equipped with uniform norm $\|\cdot\|_\infty$.

- For any $\mu \in \mathcal{M}(E)$ and any test function $f \in \mathcal{B}_b(E)$, we write

$$\mu(f) := \int_E f(x)\mu(dx).$$

A finite nonnegative kernel $Q$ from $(E, \mathcal{B}(E))$ to $(F, \mathcal{B}(F))$ is a function

$$Q : E \times \mathcal{B}(F) \mapsto \mathrm{R}_+$$

such that, for all $x \in E$, $Q(x, \cdot) \in \mathcal{M}_+(F)$ and, for all $A \in \mathcal{B}(F)$, $Q(x, A)$ is a $\mathcal{B}(E)$-measurable function. We say that $Q$ is a Markov transition kernel if, moreover, 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 by $\mu Q \in \mathcal{M}(E)$ and $Q(f) \in \mathcal{B}_b(E)$ the measure and function respectively defined by

$$\mu Q(A) := \int_E \mu(dx)Q(x, A) \qquad \forall A \in \mathcal{B}(F),$$

and

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

Given two finite nonnegative kernels $Q_1$ and $Q_2$ respectively from $E_0$ to $E_1$ and $E_1$ to $E_2$, $Q_1Q_2$ is the nonnegative kernel from $E_0$ to $E_2$ defined by

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

- For two functions $f, g \in \mathcal{B}(E)$, their tensor product is the function

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

and, in particular, we denote $f^{\otimes 2} := f \otimes f$. For two finite nonnegative kernels $Q$ and $H$ from $(E, \mathcal{B}(E))$ to $(F, \mathcal{B}(F))$, we denote

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

for all $(x, y) \in E \times E$ and all $(A, B) \in \mathcal{B}(F) \otimes \mathcal{B}(F)$. Accordingly, we write $Q^{\otimes 2} := Q \otimes Q$.

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

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

and

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

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

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

We also denote

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

and

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

A straightforward computation shows that

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

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

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

**2. Adaptive Sequential Monte Carlo.** This section presents the formal definition and the regularity assumptions of the ASMC framework studied in this article. The motivation is mainly from ASMC via summary statistics introduced in Section 2 of [4]. We refer the reader to the latter for details on motivating examples such as filtering or sequential Bayesian parameter inference.

2.1. *Framework.* The notations that are adopted are essentially those in the pair of books [11, 12]. Let $(E_n, \mathcal{B}(E_n))_{n \geq 0}$ be a sequence of Polish spaces. For each level $n \geq 1$, we consider a family of potential functions $G_{n-1,z} : E_{n-1} \mapsto \mathbb{R}_+$ and Markov kernels $M_{n,z} : (E_{n-1}, \mathcal{B}(E_n)) \mapsto [0,1]$ parametrized by $z \in \mathbb{R}^d$. Accordingly, we define the family of non-negative Feynman-Kac kernels $Q_{n,z}$ by

$$Q_{n,z}(x, A) := G_{n-1,z}(x) M_{n,z}(x, A).$$

We suppose that there exists a sequence of reference parameters $(z_n^*)_{n \geq 0}$ and, for each $n \geq 1$, we denote

$$G_{n-1} := G_{n-1,z_{n-1}^*}, \qquad M_n := M_{n,z_{n-1}^*} \quad \text{and} \quad Q_n := Q_{n,z_{n-1}^*}.$$

Starting with a known probability measure $\gamma_0 := \eta_0 \in \mathcal{P}(E_0)$, we define the unnormalized Feynman-Kac measures $\gamma_n$ by

$$\gamma_n := \gamma_0 Q_1 \cdots Q_n,$$

along with the normalized measures

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

Assumption 1 below ensures that, for all $n \geq 0$, $G_n$ is strictly positive so that

$$\gamma_n(1) = \prod_{p=0}^{n-1} \eta_p(G_p) > 0.$$

Another formulation of the connection between normalized and unnormalized measures is thus given by

(2) $$\gamma_n(f_n) = \eta_n(f_n) \prod_{p=0}^{n-1} \eta_p(G_p).$$

For $p < n$, we define the Feynman-Kac semigroup

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

and $Q_{n,n}(x, A) := \delta_x(A)$. In this context, ASMC algorithms aim at approximating the sequences of measures $(\gamma_n)_{n \geq 0}$ and $(\eta_n)_{n \geq 0}$ by exploiting some summary statistics

$$\zeta_n : E_n \mapsto \mathbf{R}^d$$

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

$$\eta_n(\zeta_n) = z_n^*.$$

2.2. *ASMC algorithm.*  In practice, ASMC and SMC algorithms share the same selection/mutation mechanisms. However, since in most situations of interest the parameters $(z_n^*)_{n \geq 0}$ are not analytically tractable, the potential functions $(G_n)_{n \geq 0}$ and transition kernels $(M_n)_{n \geq 1}$ are estimated on the fly through the design of an adaptive algorithm.

Let $N \in \mathbf{N}^*$ be the number of particles (or samples). The Interacting Particle System (IPS) associated to the ASMC algorithm is a Markov chain $(X_n)_{n \geq 0}$ taking values in $(E_n^N, \mathcal{B}(E_n)^{\otimes N})_{n \geq 0}$ with genealogy $(A_n)_{n \geq 0}$ tracking the indice of the parent of each particle at each level. Specifically, $A_{p-1}^i = j$ means that the parent of the particle $X_p^i$ at layer $p$ is $X_{p-1}^j$ at layer $p - 1$. The estimation of the normalized measure $\eta_n$ is given by the empirical measure

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

At each level $n \geq 0$, the estimated parameters are defined by $Z_n^N := \eta_n^N(\zeta_n)$. In order to lighten the notation, we denote

$$G_{n-1,N} := G_{n-1,Z_{n-1}^N}, \qquad M_{n,N} := M_{n,Z_{n-1}^N}, \qquad \text{and} \qquad Q_{n,N} := Q_{n,Z_{n-1}^N}.$$

Then, considering (2), the unnormalized Feynman-Kac measures are estimated by

$$\gamma_n^N(f_n) := \eta_n^N(f_n) \prod_{p=0}^{n-1} \eta_p^N(G_{p,N}).$$

In the following sections, we use the convention

$$\eta_{-1}^N = \gamma_{-1}^N := \eta_0.$$

Let us give the formal definition of the IPS associated with the ASMC algorithm:

(i) Initial distribution:
   At step 0, let $X_0 \sim \eta_0^{\otimes N}$.
(ii) Transition kernels:
   For all $p \geq 0$, set $Z_p^N = \eta_p^N(\zeta_p)$. The transition $X_p^i \rightsquigarrow X_{p+1}^i$ is decomposed into two steps:
   • Selection: given $X_p = x_p$, we make an independent multinomial selection of the parent of each particle by

   (3) $$S_{p,N}(x_p, da_p^i) = \sum_{k=1}^{N} \frac{G_{p,N}(x_p^k)}{\sum_{j=1}^{N} G_{p,N}(x_p^j)} \delta_k(da_p^i).$$

   Thus, the genealogy of level $p$ to level $p + 1$ is tracked by

   $$A_p \sim \bigotimes_{i=1}^{N} S_{p,N}(X_p, \cdot)$$

- Mutation: given the parent indices $A_p = a_p$, each particle at level $p$ evolves independently according to the transition kernel $M_{p+1,N}$, meaning that for $i \in [N]$,

$$X_{p+1}^i \sim M_{p+1,N}(X_p^{a_p^i}, \cdot).$$

Said differently, given $X_p$ and $A_p$, we have

$$X_{p+1} \sim \bigotimes_{i=1}^N M_{p+1,N}(X_p^{A_p^i}, \cdot).$$

2.3. *Assumptions.* Our assumptions are introduced in a similar way as in [4], but just slightly weaker. The reason why we can relax their assumptions is because we are only interested in the specific situation where the asymptotic variance of the ASMC estimator is identical to the "limiting" SMC algorithm which uses ideal potential functions and proposal kernels, namely $G_p = G_{p,z_{p-1}^*}$ and $M_p = M_{p,z_{p-1}^*}$. Considering stability properties, Section 2.7 in [4] explains why this case is particularly interesting in practice. In the following sections, we use $\mathcal{A}$ as a short-hand for *Assumption*.

ASSUMPTION 1. For each $n \geq 0$, we assume that $G_{n,z}$ is strictly positive and bounded uniformly over $z \in R^d$, i.e.,

$$\|G_{n,\cdot}\|_\infty := \sup_{(x,z) \in E_n \times R^d} G_{n,z}(x) < +\infty.$$

Notice that, under $\mathcal{A}1$, Equation (3) above is always well-defined for the denominator is always strictly positive. In the case where $G_{p,z}$ is only assumed to be nonnegative, as in the AMS framework, one may consider the stopping time $\tau_N$ defined by

$$\tau_N := \inf \left\{ p \in N : \eta_p^N(G_{p,N}) = 0 \right\}.$$

We believe that similar techniques can be applied to obtain results of the same taste as in the present paper, but at the cost of considerable technical complications which are out of the scope of this article. Let us mention that the strict positivity and boundedness of the potential functions is also required in [4] (see page 1116 and Assumption 1 page 1118). In our second assumption, "$\langle \cdot, \cdot \rangle$" stands for the Euclidean scalar product in $R^d$ and $|\cdot|$ for the associated norm.

ASSUMPTION 2. For any test function $f_{n+1} \in \mathcal{B}_b(E_{n+1})$, there exists a measurable function $h_n : (E_n \times R^d, \mathcal{B}(E_n) \otimes \mathcal{B}(R^d)) \to (R^d, \mathcal{B}(R^d))$ such that, for all $(x, z_n) \in E_n \times R^d$,

$$Q_{n+1,z_n}(f_{n+1})(x) - Q_{n+1}(f_{n+1})(x) = \langle h_n(x, z_n), z_n - z_n^* \rangle.$$

The function $h_n$ is assumed to satisfy the following properties:

- The Euclidean norm $|h_n|$ is bounded over $E_n \times R^d$ by $\|h_n\|_\infty$.
- The application $z \mapsto h_n(x, z)$ is continuous at $z_n^*$ uniformly over $x \in E_n$. More precisely, for any $\epsilon > 0$, there exists $\delta > 0$, such that $|z_n - z_n^*| < \delta$ implies

$$\sup_{x \in E_n} |h_n(x, z_n) - h_n(x, z_n^*)| < \epsilon.$$

- $h_n$ satisfies the equality $\eta_n(h_n(\cdot, z_n^*)) = 0$.

Moreover, the summary statistics $\zeta_n = (\zeta_n^1, \ldots, \zeta_n^d)$ satisfies $z_n^* = \eta_n(\zeta_n)$ and is such that, for all $k \in [d]$, $\zeta_n^k$ belongs to $\mathcal{B}_b(E_n)$.

$\mathcal{A}2$ guarantees some regularity properties of the transition kernels $Q_{n,z}$ with respect to the parameter $z$ and is just a slight generalization of the framework studied in Section 2 of [4]. Indeed, our function $h_n$ coincides with the function $\omega$ defined in (2.17) of [4], that is

$$h_n(x, z_n) = \int_0^1 \partial_z Q_{n+1,z}(f_{n+1})(x)\big|_{z=z_n^*+\lambda(z_n-z_n^*)} \, d\lambda.$$

As such, the first two conditions on $h_n$ are satisfied as soon as Assumption 2 in [4] is verified. In this respect, our third condition on $h_n$ corresponds to their condition (2.19) in Theorem 2.3, which is precisely the "limiting" case mentioned above. Finally, the hypothesis that the summary statistics are bounded is also required in their Assumption 1, while the relation $z_n^* = \eta_n(\zeta_n)$ corresponds in their notation to $\bar{\xi}_n = \eta_{n-1}(\xi_n)$.

We also want to mention that the second point is equivalent to

$$\forall \epsilon > 0, \ \exists g_n \in \mathcal{B}_b(E_n), \ \exists \delta > 0, \ s.t.$$

$$|z_n - z_n^*| < \delta \implies \forall x \in E_n, \ |h_n(x, z_n) - h_n(x, z_n^*)| < g_n(x)\epsilon \le \|g_n\|_\infty \epsilon.$$

We expect that, in this alternative formulation, the functions $g_n$ and $h_n$ can be relaxed to some unbounded functions, belonging for example to $\mathbb{L}^2(\eta_n)$, along with stronger conditions on the test function $f_{n+1}$. We believe that this is one of the main differences between the ASMC framework studied in [4] and the AMS framework studied in [8].

In general, it is not easy to verify the existence of such $h_n$. However, we have, at least, a direction to explore in the case where $Q_{n,z}(f)$ is not globally differentiable with respect to $z$. We also remark that we do not study the consistency of $\gamma_n^N(f)$ and $\eta_n^N(f)$ with weaker assumptions, as we are only interested in the CLT type result of Theorem 2.1 below and, more specifically, in the estimation of the asymptotic variance. Nevertheless, let us briefly mention that to establish the consistency of $\gamma_n^N$ and $\eta_n^N$, one just needs

$$\gamma_{n-1}^N Q_{n,N}(f_n) - \gamma_{n-1}^N Q_n(f_n) = o_p(1)$$

for any test function $f_n \in \mathcal{B}_b(E_n)$. This does not require such a strong assumption as $\mathcal{A}2$. However, for CLT type results with the "stable" asymptotic variance, it is necessary that

$$\gamma_{n-1}^N Q_{n,N}(f_n) - \gamma_{n-1}^N Q_n(f_n) = o_p\left(\frac{1}{\sqrt{N}}\right).$$

A stronger regularity assumption like $\mathcal{A}2$ over the parametrization is therefore required.

2.4. *Central limit theorems.* As explained before, the present article only deals with the case where the asymptotic variance is identical to the "limiting" one, which is only a special case of the Central Limit Theorem 2.2 given in [4] under slightly weaker assumptions. This is why, in Section 4.2, we propose a different strategy for the proof.

THEOREM 2.1. *Assume $\mathcal{A}1$-$\mathcal{A}2$. For any test function $f \in \mathcal{B}_n(E_n)$, we have*

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

*and*

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

*where*

$$\sigma_{\gamma_n}^2(f) := \sum_{p=0}^n \left(\gamma_p(1)\gamma_p(Q_{p,n}(f)^2) - \gamma_n(f)^2\right) \qquad and \qquad \sigma_{\eta_n}^2(f) := \sigma_{\gamma_n}^2(f)/\gamma_n(1)^2.$$

One can notice that the CLT for $\eta_n^N$ is just a consequence of the CLT for $\gamma_n^N$, Slutsky's Lemma, and the decomposition

$$\sqrt{N}\left(\eta_n^N(f) - \eta_n(f)\right) = \frac{1}{\gamma_n^N(1)}\sqrt{N}\left(\gamma_n^N(f - \eta_n(f)) - \gamma_n(f - \eta_n(f))\right).$$

The main goal of this paper is to estimate the asymptotic variances $\sigma_{\gamma_n}^2(f)$ and $\sigma_{\eta_n}^2(f - \eta_n(f))$ by a single simulation of the particle system, exactly as is done by Lee and Whiteley in [16] in a nonadaptive context.

**3. Variance estimations.** In this section, we recall the coalescent tree-based expansion of the variance firstly introduced in [6] from which we deduce a new variance estimator. We also recall the variance estimator proposed by Lee and Whiteley in [16] and explain the connection between both estimators.

3.1. *Coalescent tree-based variance expansion.* 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. We associate with any coalescence indicator $b \in \{0, 1\}^{n+1}$ the nonnegative measures $\Gamma_n^b$ and $\bar{\Gamma}_n^b \in \mathcal{M}_+(E_n^2)$ defined for any $F \in \mathcal{B}_b(E_n^2)$ by

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

and

$$\bar{\Gamma}_n^b(F) := \frac{1}{\gamma_n(1)^2}\Gamma_n^b(F).$$

When there is only one coalescence at, say, level $p$, we write $\Gamma_n^{(p)}(F)$ and $\bar{\Gamma}_n^{(p)}(F)$ instead of $\Gamma_n^b(F)$ and $\bar{\Gamma}_n^b(F)$ (see Figure 1). When there is no coalescence at all, that is $b = (0, \ldots, 0)$, we have

$$\Gamma_n^{(\varnothing)}(F) = \gamma_n^{\otimes 2}(F) \qquad \text{and} \qquad \bar{\Gamma}_n^{(\varnothing)}(F) = \eta_n^{\otimes 2}(F).$$
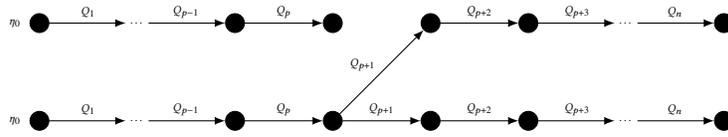


FIG 1. *A representation of the coalescent tree-based measure* $\Gamma_n^{(p)}$.

It is easy to verify from the definition that

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

As noticed in [6], the latter yields alternative representations for the asymptotic variances of Theorem 2.1, namely

$$(4) \qquad \sigma_{\gamma_n}^2(f) = \sum_{p=0}^{n}\left(\Gamma_n^{(p)}(f^{\otimes 2}) - \Gamma_n^{(\varnothing)}(f^{\otimes 2})\right),$$

and

$$\sigma_{\eta_n}^2(f) = \sum_{p=0}^{n} \left( \bar{\Gamma}_n^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_n^{(\varnothing)}(f^{\otimes 2}) \right).$$

(5)

As a consequence, if for any coalescence indicator $b := (b_0, \dots, b_n) \in \{0, 1\}^{n+1}$, we can construct a consistent estimator $\bar{\Gamma}_{n,N}^b$ of $\bar{\Gamma}_n^b$, then we automatically deduce consistent estimators for the asymptotic variances of Theorem 2.1. This is the idea behind our next definition.

In this definition, $\tilde{a}_p^{[2]} = (\tilde{a}_p^1, \tilde{a}_p^2)$ 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

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

Additionally, we use the notation $X_n^{\ell_n^{[2]}}$ as a short-hand for $(X_n^{\ell_n^1}, X_n^{\ell_n^2})$.

DEFINITION 3.2.    For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator $b$, we define the estimator $\bar{\Gamma}_{n,N}^b$ of the measure $\bar{\Gamma}_n^b$ by

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

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

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

The estimator of $\Gamma_n^b$ is defined by

$$\Gamma_{n,N}^b(F) = \gamma_n^N(1)^2 \, \bar{\Gamma}_{n,N}^b(F).$$

Since $\ell_p^1 \neq \ell_p^2$, we also have

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

Notice that, for $n = 0$, we get

(6)       $$\bar{\Gamma}_{0,N}^b(F) := \frac{1}{N(N-1)} \sum_{\ell_0^{[2]} \in (N)^2} C_{b_0}(F)(X_0^{\ell_0^{[2]}}) = \frac{1}{N(N-1)} \sum_{i \neq j} C_{b_0}(F)(X_0^i, X_0^j).$$

We also adopt the convention

$$\bar{\Gamma}_{-1,N}^b(F) = \Gamma_{-1,N}^b(F) := \eta_0^{\otimes 2} C_{b_0}(F).$$

3.1.1. *A toy example.*    As the definition of the estimator $\Gamma_{n,N}^b$ is not completely straightforward, we illustrate the idea on a simple example. For this, we consider the IPS of Figure 2.

Suppose we want to estimate $\bar{\Gamma}_6^{(3)}(F)$ by $\bar{\Gamma}_{6,5}^{(3)}(F)$. We denote $b^* = (0, 0, 0, 1, 0, 0, 0)$ the corresponding coalescence indicator. In the associated IPS, we have to find the choices of $\ell_{0:6}^{[2]}$ such that

(7)                              $$\prod_{p=0}^{5} \lambda_p^{b^*}(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) = 1.$$
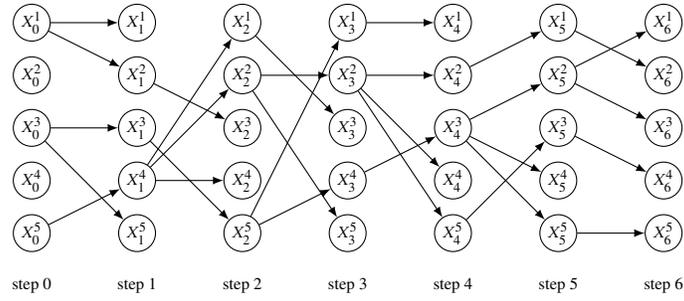
FIG 2. *An IPS with n + 1 = 7 levels and N = 5 particles at each level.*
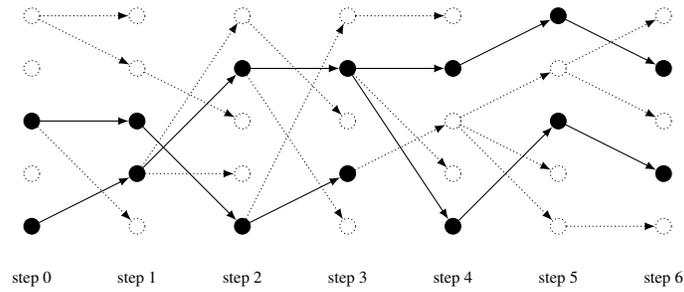


FIG 3. *The first family of $\ell_{0:6}^{[2]}$ such that (7) is verified.*

It turns out that there are 4 possible choices, taking into account that $F(x, x')$ is not necessarily symmetric in its variables. Namely, the first couple of ancestral lines is (see Figure 3):

- $\ell_{0:6}^{[2]} = ((5, 3), (4, 3), (2, 5), (2, 4), (2, 5), (1, 3), (2, 4))$;
- $\ell_{0:6}^{[2]} = ((5, 3), (4, 3), (2, 5), (2, 4), (5, 2), (3, 1), (4, 2))$.
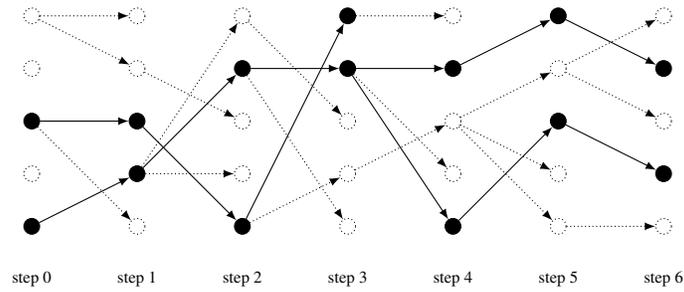


FIG 4. *The second family of $\ell_{0:6}^{[2]}$ such that (7) is verified.*

The second couple of ancestral lines is (see Figure 4):

- $\ell_{0:6}^{[2]} = ((5, 3), (4, 3), (2, 5), (2, 1), (2, 5), (1, 3), (2, 4))$;
- $\ell_{0:6}^{[2]} = ((5, 3), (4, 3), (2, 5), (2, 1), (5, 2), (3, 1), (4, 2))$.

Hence, the number of choices of $\ell_{0:6}^{[2]}$ where $\ell_6^{[2]} = (2,4)$ is 2, and the number of choices of $\ell_{0:6}^{[2]}$ where $\ell_6^{[2]} = (4,2)$ is also 2. As a consequence, we have

$$\bar{\Gamma}_{6,5}^{(3)}(F) = 2 \times \left\{ \frac{5^5}{4^7} \left( F(X_6^2, X_6^4) + F(X_6^4, X_6^2) \right) \right\}.$$

Our next result ensures the convergence of our estimators.

THEOREM 3.1 (Convergence of $\Gamma_{n,N}^b$). *Assume* $\mathcal{A}1$-$\mathcal{A}2$. *For any test functions* $\phi, \psi \in \mathcal{B}_b(E_n)$ *and for any coalescence indicator* $b \in \{0,1\}^{n+1}$, *we have*

$$\Gamma_{n,N}^b(\phi \otimes \psi) - \Gamma_n^b(\phi \otimes \psi) = \mathcal{O}_p\left( \frac{1}{\sqrt{N}} \right).$$

The proof is given in Section 4.3.

3.2. *Term by term estimator.* Considering (4), (5), and Theorem 3.1, we are now in a position to provide term by term variance estimators for $\sigma_{\gamma_n}^2(f)$ and $\sigma_{\eta_n}^2(f)$.

DEFINITION 3.3 (Estimators of the asymptotic variances). Given a test function $f \in \mathcal{B}_b(E_n)$, we let

$$\sigma_{\gamma_{n,N}}^2(f) := \sum_{p=0}^{n} \left( \Gamma_{n,N}^{(p)}(f^{\otimes 2}) - \Gamma_{n,N}^{(\varnothing)}(f^{\otimes 2}) \right),$$

and

$$\sigma_{\eta_{n,N}}^2(f) := \sum_{p=0}^{n} \left( \bar{\Gamma}_{n,N}^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}) \right).$$

Theorem 3.1 ensures the consistency of both $\Gamma_{n,N}^{(p)}(f^{\otimes 2})$ and $\Gamma_{n,N}^{(\varnothing)}(f^{\otimes 2})$. Returning to (4), this amounts to saying that

$$\sigma_{\gamma_{n,N}}^2(f) = \sum_{p=0}^{n} \left( \Gamma_{n,N}^{(p)}(f^{\otimes 2}) - \Gamma_{n,N}^{(\varnothing)}(f^{\otimes 2}) \right) \xrightarrow[N\to\infty]{\mathrm{P}} \sum_{p=0}^{n} \left( \Gamma_n^{(p)}(f^{\otimes 2}) - \Gamma_n^{(\varnothing)}(f^{\otimes 2}) \right) = \sigma_{\gamma_n}^2(f).$$

Similarly, for the consistency of $\sigma_{\eta_{n,N}}^2(f - \eta_n^N(f))$, since by (5) we know that

$$\sigma_{\eta_n}^2(f) = \sum_{p=0}^{n} \left( \bar{\Gamma}_n^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_n^{(\varnothing)}(f^{\otimes 2}) \right),$$

it suffices to verify that, for any coalescent indicator $b$,

$$(8) \qquad \bar{\Gamma}_{n,N}^b\left( [f - \eta_n^N(f)]^{\otimes 2} \right) \xrightarrow[N\to\infty]{\mathrm{P}} \bar{\Gamma}_n^b\left( [f - \eta_n^N(f)]^{\otimes 2} \right).$$

Clearly, the linearity of $\bar{\Gamma}_{n,N}^b$ yields

$$\bar{\Gamma}_{n,N}^b\left( [f - \eta_n^N(f)]^{\otimes 2} \right)$$

$$= \bar{\Gamma}_{n,N}^b(f^{\otimes 2}) - \eta_n^N(f) \left( \bar{\Gamma}_{n,N}^b(1 \otimes f) + \bar{\Gamma}_{n,N}^b(f \otimes 1) \right) + \eta_n^N(f)^2 \bar{\Gamma}_{n,N}^b\left( 1^{\otimes 2} \right).$$

Mutatis mutandis, the same relation holds for $\bar{\Gamma}_n^b \left( \left[ f - \eta_n^N(f) \right]^{\otimes 2} \right)$. Since a by-product of Theorem 2.1 is that

$$\eta_n^N(f) - \eta_n(f) = \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right),$$

the verification of (8) is just a consequence of Theorem 3.1 and Slutsky's Lemma. Hence, we have obtained the following result.

THEOREM 3.2 (Consistency of $\sigma_{\gamma_{n,N}}^2$ and $\sigma_{\eta_{n,N}}^2$). *Assume $\mathcal{A}1$-$\mathcal{A}2$. For $f \in \mathcal{B}_b(E_n)$, we have*

$$\sigma_{\gamma_{n,N}}^2(f) - \sigma_{\gamma_n}^2(f) = \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right),$$

*as well as*

$$\sigma_{\eta_{n,N}}^2(f - \eta_n^N(f)) - \sigma_{\eta_n}^2(f - \eta_n(f)) = \mathcal{O}_p \left( \frac{1}{\sqrt{N}} \right).$$

Even if the term by term estimator is very natural in theory, the computational cost is quite heavy in practice since one has to trace the whole genealogy of a particle system and calculate all the corresponding terms one by one. Therefore, we do not provide an efficient algorithm to calculate this estimator. Instead, we show in the next section that this estimator can be connected to the one given by Lee & Whiteley in a nonadaptive context (SMC), which is very simple and fast to calculate. Let us also mention that our term by term estimator is different from the one introduced in Section 4.1 of [16]. The interested reader can find more details on this point in Appendix A.3.

3.3. *Disjoint ancestral lines estimator.* Let us now recall the variance estimator proposed in [16], which can be seen as a disjoint ancestral lines estimator. Namely, given a test function $f \in \mathcal{B}_b(E_n)$, consider

$$(9) \qquad V_n^N(f) := \eta_n^N(f)^2 - \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{E_n^i \neq E_n^j} f(X_n^i) f(X_n^j),$$

where $E_n^i$ is the ancestor index of $X_n^i$ at level 0. Returning to the toy example of Section 3.1, the couples $(i, j)$ such that $i < j$ and $E_n^i \neq E_n^j$ are: $(1,2), (1,4), (2,3), (2,5), (3,4), (4,5)$.

In a nonadaptive context (SMC), this is the variance estimator introduced in [16] when the number $N$ of particles is the same at each step. The reader is referred to [16] for an efficient algorithm to compute this estimator.

According to our notation, since $E_n^i \neq E_n^j$ corresponds to the case $b = (0, \ldots, 0) = (\varnothing)$ of disjoint ancestral lines, we may also write

$$V_n^N(f) = \eta_n^N(f)^2 - \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}).$$

The following proposition makes a connection between $V_n^N(f)$ and our estimators. Notice that this result does not depend on $\mathcal{A}2$, but is provided by the structure of the IPS and the underlying multinomial selection scheme. The proof is housed in Section 4.5.

PROPOSITION 3.1. *Assume $\mathcal{A}1$. For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$NV_n^N(f) - \sigma_{\eta_{n,N}}^2(f) = \mathcal{O}_p \left( \frac{1}{N} \right),$$

and

$$NV_n^N(f - \eta_n^N(f)) - \sigma_{\eta_{n,N}}^2(f - \eta_n^N(f)) = \mathcal{O}_p\left(\frac{1}{N}\right).$$

By combining Theorem 3.2 and Proposition 3.1, we finally obtain the main result of the present article.

THEOREM 3.3. *Assume $\mathcal{A}1$-$\mathcal{A}2$. For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$N\gamma_n^N(1)^2 V_n^N(f) - \sigma_{\gamma_n}^2(f) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right),$$

*and*

$$NV_n^N(f - \eta_n^N(f)) - \sigma_{\eta_n}^2(f - \eta_n(f)) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

Hence, the main message of the present work is that the computationally very simple estimator proposed by Lee and Whiteley in a nonadaptive framework (SMC) is still consistent in an adaptive one (ASMC). However, since we could not adapt easily their proof to our adaptive context, we propose a new approach to show this consistency result. More details on the connection between both estimators are given in Appendix A.3.

As emphasized before, among other ingredients, the tools we use connect the study of Particle Markov Chain Monte Carlo methods and the variance estimation problem in SMC methods. As such, more generally, they may give some new insights when dealing with complex genealogy-involved problems of Interacting Particle Systems.

Before going into the details of the proofs, let us mention that a numerical experiment on a toy example is proposed in Appendix B to illustrate the consistency of the Lee and Whiteley variance estimators in the adaptive and nonadaptive cases. Finally, Appendix C makes a connection between our term by term estimators and the truncated variance estimators recently proposed by Olsson and Douc in [18] to address the issue of degeneracy in the ancestral lines.

## 4. Proofs.

4.1. *Almost sure convergence.* In this section, we provide classical almost sure convergence results on SMC framework under our specific parameterization, namely with adaptive potential functions and transition kernels. We focus on the properties that do not use the additional information given by the genealogy of the associated IPS. Therefore, in order to simplify the story, we give a "rougher" definition of the associated IPS without considering the genealogy.

- $X_0 \sim \eta_0^{\otimes N}$
- For $p \geq 1$, we let

$$X_p \sim \bigotimes_{i=1}^N K_{p,\eta_{p-1}^N}(X_{p-1}^i, \cdot)$$

where, given $X_{p-1}$, $K_{p,\eta_{p-1}^N}$ is the Markov kernel defined by

$$\forall (x, A) \in E_{p-1} \times \mathcal{B}(E_p), \qquad K_{p,\eta_{p-1}^N}(x, A) := \frac{\eta_{p-1}^N Q_{p,N}(x, A)}{\eta_{p-1}^N\left(G_{p-1,N}\right)}.$$

It is easy to check that the distributions of the particles are identical to the ones defined in Section 2.2. Let us begin with the consistency of the corresponding adaptive estimators. Recall that, by $\mathcal{A}2$, the summary statistics $\zeta_n = (\zeta_n^1, \cdots, \zeta_n^d)$ satisfies $\eta_n(\zeta_n) = z_n^*$ and, for all $k \in [d]$, $\zeta_n^k$ belongs to $\mathcal{B}_b(E_n)$.

THEOREM 4.1. *Assume $\mathcal{A}1$-$\mathcal{A}2$. For any $f \in \mathcal{B}_b(E_n)$, we have*

$$\gamma_n^N(f) \xrightarrow[N\to\infty]{a.s.} \gamma_n(f),$$

*and*

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

*In particular, we also have*

$$Z_n^N = \eta_n^N(\zeta_n) \xrightarrow[N\to\infty]{a.s.} \eta_n(\zeta_n) = z_n^*.$$

PROOF. By definition, it is clear that the convergence of $\gamma_n^N$ implies the convergence of $\eta_n^N$. Therefore, it is sufficient to establish the first one. We prove by induction that

$$\forall f \in \mathcal{B}_b(E_n), \quad \gamma_n^N(f) \xrightarrow[N\to\infty]{a.s.} \gamma_n(f).$$

Step 0:

The almost sure convergence of $\gamma_0^N = \eta_0^N$ to $\gamma_0 = \eta_0$ with respect to a test function in $\mathcal{B}_b(E_0)$ is given by the strong law of large numbers.

Step $n \geq 1$:

We assume that

$$Z_{n-1}^N \xrightarrow[N\to\infty]{a.s.} z_{n-1}^*$$

and, for any $\phi \in \mathcal{B}_b(E_{n-1})$,

$$\gamma_{n-1}^N(\phi) \xrightarrow[N\to\infty]{a.s.} \gamma_{n-1}(\phi).$$

For any $f \in \mathcal{B}_b(E_n)$, the triangular inequality yields

(10) $\left| \gamma_n^N(f) - \gamma_n(f) \right|$

$$\leq \underbrace{\left| \gamma_n^N(f) - \gamma_{n-1}^N Q_{n,N}(f) \right|}_{P_1(N)} + \underbrace{\left| \gamma_{n-1}^N Q_{n,N}(f) - \gamma_{n-1}^N Q_n(f) \right|}_{P_2(N)} + \underbrace{\left| \gamma_{n-1}^N Q_n(f) - \gamma_{n-1} Q_n(f) \right|}_{P_3(N)}.$$

• For $P_1(N)$, we denote

$$U_{n,N}^i := \eta_{n-1}^N(G_{n-1,N}) f(X_n^i) - \eta_{n-1}^N Q_{n,N}(f).$$

It is readily seen that

$$P_1(N) = \gamma_{n-1}^N(1) \frac{1}{N} \sum_{i=1}^N U_{n,N}^i.$$

Given $\mathcal{F}_{n-1}^N := \sigma(X_0, \ldots, X_{n-1})$, the random variables $(U_{n,N}^i)_{1 \le i \le N}$ are i.i.d. and such that

$$\mathrm{E}\left[U_{n,N}^i \,\middle|\, \mathcal{F}_{n-1}^N\right] = \eta_{n-1}^N(G_{n-1,N})\frac{\eta_{n-1}^N Q_{n,N}(f)}{\eta_{n-1}^N(G_{n-1,N})} - \eta_{n-1}^N Q_{n,N}(f) = 0.$$

Under $\mathcal{A}1$, we can also see that

$$\left|U_{n,N}^i\right| \le C_n := 2\left\|G_{n-1,\cdot}\right\|_\infty \|f\|_\infty.$$

Therefore, for any $\epsilon > 0$, Hoeffding's inequality gives

$$\mathrm{P}\left(\left|\sum_{i=1}^N U_{n,N}^i\right| \ge N\epsilon \,\middle|\, \mathcal{F}_{n-1}^N\right) \le 2\exp\left(\frac{-\epsilon^2 N}{2C_n^2}\right).$$

Since this upper-bound is deterministic, this amounts to saying that

$$\mathrm{P}\left(\left|\sum_{i=1}^N U_{n,N}^i\right| \ge N\epsilon\right) \le 2\exp\left(\frac{-\epsilon^2 N}{2C_n^2}\right).$$

Consequently, Borel-Cantelli Lemma ensures that

$$\frac{1}{N}\sum_{i=1}^N U_{n,N}^i \xrightarrow[N\to\infty]{a.s.} 0.$$

Combined with the induction hypothesis, we get

$$P_1(N) = \gamma_{n-1}^N(1)\frac{1}{N}\sum_{i=1}^N U_{n,N}^i \xrightarrow[N\to\infty]{a.s.} 0.$$

- For $P_2(N)$, $\mathcal{A}2$ implies that there exists a function $h_{n-1}$ such that

$$Q_{n,N}(f)(x) - Q_n(f)(x) = \left\langle h_{n-1}(x, Z_{n-1}^N), Z_{n-1}^N - z_{n-1}^*\right\rangle.$$

Hence, since $|h_{n-1}|$ and the potential functions $G_{n,z}$ are bounded, Cauchy-Schwarz inequality gives

$$P_2(N) \le \gamma_{n-1}^N(1)\|h_{n-1}\|_\infty\left|Z_{n-1}^N - z_{n-1}^*\right| \le \left\{\prod_{p=0}^{n-2}\|G_{p,\cdot}\|_\infty\right\}\|h_{n-1}\|_\infty\left|Z_{n-1}^N - z_{n-1}^*\right|.$$

By induction hypothesis, we conclude that

$$P_2(N) \xrightarrow[N\to\infty]{a.s.} 0.$$

- For $P_3(N)$, under $\mathcal{A}1$, we have that $Q_n(f) \in \mathcal{B}_b(E_{n-1})$. Thus, the induction hypothesis gives

$$P_3(N) \xrightarrow[N\to\infty]{a.s.} 0.$$

Considering (10), the verification of the convergence

$$\forall f \in \mathcal{B}_b(E_n), \quad \gamma_n^N(f) \xrightarrow[N\to\infty]{a.s.} \gamma_n(f)$$

is then complete. □

4.2. *Proof of Theorem 2.1.* We prove by induction that

$$\sqrt{N}\left(\gamma_n^N(f) - \gamma_n(f)\right) \xrightarrow[N\to\infty]{d} \mathcal{N}\left(0, \sigma_{\gamma_n}^2(f)\right).$$

The verification of step 0 comes from the CLT for i.i.d. random variables. For step $n \geq 1$, we suppose that

$$\forall 0 \leq p \leq n-1, \quad \sqrt{N}\left(\gamma_p^N(f) - \gamma_p(f)\right) \xrightarrow[N\to\infty]{d} \mathcal{N}\left(0, \sigma_{\gamma_p}^2(f)\right).$$

Notice that, by $\mathcal{A}2$, this implies that

(11) $$\forall 0 \leq p \leq n-1, \quad \sqrt{N}\left|Z_{p-1}^N - z_{p-1}^*\right| = \mathcal{O}_p(1).$$

For any test function $f \in \mathcal{B}_b(E_n)$, we denote $f_p := Q_{p,n}(f) \in \mathcal{B}_b(E_p)$. For any $(x, A) \in E_0 \times \mathcal{B}(E_0)$ we set $Q_0(x, A) = Q_{0,N}(x, A) = \delta_x(A)$. Taking into account the convention $\gamma_{-1}^N = \gamma_0 = \eta_0$ and the fact that $\gamma_n = \gamma_0 Q_{0,n}$, we have the telescoping decomposition

$$\gamma_n^N(f) - \gamma_n(f)$$

$$= \sum_{p=0}^{n}\left(\gamma_p^N(f_p) - \gamma_{p-1}^N Q_p(f_p)\right)$$

$$= \frac{1}{N}\sum_{p=0}^{n}\sum_{i=1}^{N}\left\{\left(\gamma_p^N(1)f_p(X_p^i) - \gamma_{p-1}^N Q_{p,N}(f_p)\right) + \left(\gamma_{p-1}^N Q_{p,N}(f_p) - \gamma_{p-1}^N Q_p(f_p)\right)\right\}.$$

For $k \in [(n+1)N]$, we denote

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

We define the filtration

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

Then, we set

$$U_k^N := \frac{1}{\sqrt{N}}\left(\gamma_{p_k}^N(1)f_{p_k}(X_{p_k}^{i_k}) - \gamma_{p_k-1}^N Q_{p_k,N}(f_{p_k})\right),$$

and

$$D_p^N := \sqrt{N}\left(\gamma_{p-1}^N Q_{p,N}(f_p) - \gamma_{p-1}^N Q_p(f_p)\right),$$

so that

(12) $$\sqrt{N}\left(\gamma_n^N(f) - \gamma_n(f)\right) = \sum_{k=1}^{(n+1)N}\left(U_k^N + \frac{1}{N}D_{p_k}^N\right) = \sum_{k=1}^{(n+1)N} U_k^N + \sum_{p=0}^{n} D_p^N.$$

From $\mathcal{A}2$, we know that there exists a function $h_{p-1}$ such that

$$D_p^N = \sqrt{N}\left\langle \gamma_{p-1}^N\left(h_{p-1}(\cdot, Z_{p-1}^N)\right), Z_{p-1}^N - z_{p-1}^*\right\rangle$$

$$= \sqrt{N}\left\langle \gamma_{p-1}^N\left(h_{p-1}(\cdot, Z_{p-1}^N) - h_{p-1}(\cdot, z_{p-1}^*)\right), Z_{p-1}^N - z_{p-1}^*\right\rangle$$

$$+ \sqrt{N}\left\langle \gamma_{p-1}^N\left(h_{p-1}(\cdot, z_{p-1}^*)\right), Z_{p-1}^N - z_{p-1}^*\right\rangle.$$

- For the first part, we have by Cauchy-Schwarz inequality

$$\left| \sqrt{N} \left\langle \gamma_{p-1}^N \left( h_{p-1}(\cdot, Z_{p-1}^N) - h_{p-1}(\cdot, z_{p-1}^*) \right), Z_{p-1}^N - z_{p-1}^* \right\rangle \right|$$

$$\leq \sqrt{N} \, \gamma_{p-1}^N(1) \left| Z_{p-1}^N - z_{p-1}^* \right| \sup_{x \in E_{p-1}} \left| h_{p-1}(x, Z_{p-1}^N) - h_{p-1}(x, z_{p-1}^*) \right|.$$

Then, let us consider

$$\Omega_{p-1} = \left\{ \omega \in \Omega : Z_{p-1}^N(\omega) \xrightarrow[N \to \infty]{} z_{p-1}^* \right\}.$$

By Theorem 4.1, $\Omega_{p-1}$ has probability one. Therefore, by $\mathcal{A}2$, for all $\omega \in \Omega_{p-1}$ and all $\epsilon > 0$, there exists $N(\omega, \epsilon) > 0$ such that, for all $N > N(\omega, \epsilon)$,

$$\sup_{x \in E_{p-1}} \left| h_{p-1}(x, Z_{p-1}^N(\omega)) - h_{p-1}(x, z_{p-1}^*) \right| < \epsilon.$$

This means that

$$\sup_{x \in E_{p-1}} \left| h_{p-1}(x, Z_{p-1}^N) - h_{p-1}(x, z_{p-1}^*) \right| \xrightarrow[N \to \infty]{a.s.} 0.$$

Thus, we deduce from (11) that

$$\sqrt{N} \left\langle \gamma_{p-1}^N \left( h_{p-1}(\cdot, Z_{p-1}^N) - h_{p-1}(\cdot, z_{p-1}^*) \right), Z_{p-1}^N - z_{p-1}^* \right\rangle \xrightarrow[N \to \infty]{P} 0.$$

- For the second part, since Theorem 4.1 and $\mathcal{A}2$ imply that

$$\gamma_{p-1}^N (h_{p-1}(\cdot, z_{p-1}^*)) \xrightarrow[N \to \infty]{a.s.} \gamma_{p-1}(h_{p-1}(\cdot, z_{p-1}^*)) = 0,$$

we conclude by (11) that

$$\sqrt{N} \left\langle \gamma_{p-1}^N \left( h_{p-1}(\cdot, z_{p-1}^*) \right), Z_{p-1}^N - z_{p-1}^* \right\rangle \xrightarrow[N \to \infty]{P} 0.$$

Hence we have proved that

$$D_p^N \xrightarrow[N \to \infty]{P} 0,$$

which leads to

$$\sum_{p=0}^{n} D_p^N \xrightarrow[N \to \infty]{P} 0.$$

Next, it is easy to check that $(U_k^N)_{1 \leq k \leq (n+1)N}$ is an $(\mathcal{E}_k^N)_{1 \leq k \leq (n+1)N}$-martingale difference array. In order to apply Theorem 2.3 in [17], we just have to check that

- By $\mathcal{A}1$,

$$(13) \qquad \max_{1 \leq k \leq (n+1)N} \left| U_k^N \right| \leq \frac{2}{\sqrt{N}} \|f\|_\infty \max_{1 \leq p \leq n} \prod_{q=0}^{p-1} \left\| G_{q,\cdot} \right\|_\infty \leq \frac{2}{\sqrt{N}} \|f\|_\infty \sum_{p=1}^{n} \prod_{q=0}^{p-1} \left\| G_{q,\cdot} \right\|_\infty,$$

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

- From (13), we also get that

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

- Standard calculation gives

$$\sum_{k=1}^{(n+1)N} \left(U_k^N\right)^2$$

$$= \sum_{p=0}^{n} \left(\gamma_p^N(1)^2 \eta_p^N(f_p^2) + (\gamma_{p-1}^N Q_{p,N}(f_p))^2 - 2\gamma_p^N(1)\eta_p^N(f_p)\gamma_{p-1}^N Q_{p,N}(f_p)\right).$$

As shown above, the convergence of $D_p^N$ indicates that

$$\gamma_{p-1}^N Q_{p,N}(f_p) - \gamma_{p-1}^N Q_p(f_p) \xrightarrow[N\to\infty]{\text{P}} 0.$$

Then, by applying Theorem 4.1, we obtain

$$\sum_{k=1}^{(n+1)N} \left(U_k^N\right)^2 \xrightarrow[N\to\infty]{\text{P}} \sigma_{\gamma_n}^2(f),$$

Therefore, we have the following central limit theorem

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

Returning to (12), the conclusion follows from Slutsky's Lemma.

4.3. *Proof of Theorem 3.1.* We want to show that, under $\mathcal{A}1$-$\mathcal{A}2$, for any test functions $\phi, \psi \in \mathcal{B}_b(E_n)$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, we have

$$\Gamma_{n,N}^b(\phi \otimes \psi) - \Gamma_n^b(\phi \otimes \psi) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

Before proceeding, let us introduce some additional notation. With a slight abuse of notation, for a coalescence indicator $b = (b_0, \ldots, b_n) \in \{0,1\}^{n+1}$, we denote, for all $0 \le p \le n$,

$$\Gamma_p^b := \Gamma_p^{(b_0,\ldots,b_p)} \qquad \text{and} \qquad \Gamma_{p,N}^b := \Gamma_{p,N}^{(b_0,\ldots,b_p)}$$

with the convention

$$\Gamma_{-1,N}^b = \Gamma_{-1}^b := \eta_0^{\otimes 2} C_{b_0}.$$

Note that, with this convention, we have

$$\Gamma_p^b = \Gamma_{p-1}^b Q_p^{\otimes 2} C_{b_p}.$$

We also remark that, for any $b_n \in \{0,1\}$ and any $\phi, \psi \in \mathcal{B}_b(E_n)$, there exists $f$ and $g$ in $\mathcal{B}_b(E_{n-1})$ such that

(14) $$Q_n^{\otimes 2} C_{b_n}(\phi \otimes \psi) = f \otimes g.$$

Specifically, for $b_n = 0$, it suffices to consider $f = Q_n(\phi)$ and $g = Q_n(\psi)$, while for $b_n = 1$ one can take $f = Q_n(\phi\psi)$ and $g = Q_n(1) = G_{n-1}$. As usual, the proof is done by induction.

- Step 0:

- If $b_0 = 1$, (6) and Definition 3.1 give

$$\mathrm{E}\left[\Gamma_{0,N}^{b}(\phi \otimes \psi)\right] = \mathrm{E}\left[\frac{1}{N}\sum_{i=1}^{N}\phi(X_0^i)\psi(X_0^i)\right] = \eta_0(\phi\psi) = \Gamma_0^b(\phi \otimes \psi).$$

Hence, the central limit theorem yields

$$\sqrt{N}\left(\Gamma_{0,N}^{b}(\phi \otimes \psi) - \Gamma_0^b(\phi \otimes \psi)\right) \xrightarrow[N \to \infty]{\mathrm{d}} \mathcal{N}\left(0, \eta_0(\phi^2\psi^2) - \eta_0(\phi\psi)^2\right),$$

so that

$$\Gamma_{0,N}^{b}(\phi \otimes \psi) - \Gamma_0^b(\phi \otimes \psi) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

- If $b_0 = 0$, the central limit theorem ensures that

$$\eta_0^N(\phi) - \eta_0(\phi) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right) \quad \text{and} \quad \eta_0^N(\psi) - \eta_0(\psi) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

Therefore, we have

$$\eta_0^N(\phi)\eta_0^N(\psi) - \eta_0(\phi)\eta_0(\psi)$$
$$= \left(\eta_0^N(\phi) - \eta_0(\phi)\right)\eta_0^N(\psi) + \eta_0(\phi)\left(\eta_0^N(\psi) - \eta_0(\psi)\right)$$
$$= \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

Thanks to (1), one has

$$\Gamma_{0,N}^{b}(\phi \otimes \psi) = \frac{N}{N-1}\left(\eta_0^N(\phi)\eta_0^N(\psi) - \frac{1}{N^2}\sum_{i=1}^{N}\phi(X_0^i)\psi(X_0^i)\right).$$

Combined with Definition 3.1 and the law of large numbers, one deduces that

$$\Gamma_{0,N}^{b}(\phi \otimes \psi) - \Gamma_0^b(\phi \otimes \psi) = \Gamma_{0,N}^{b}(\phi \otimes \psi) - \eta_0(\phi)\eta_0(\psi) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

- Step $n \geq 1$:

We suppose that for any test functions $f, g \in \mathcal{B}_b(E_{n-1})$ and coalescence indicator $b$, we have

$$\Gamma_{n-1,N}^{b}(f \otimes g) - \Gamma_{n-1}^b(f \otimes g) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

Next, we consider the following decomposition

(15)
$$\Gamma_{n,N}^{b}(\phi \otimes \psi) - \Gamma_n^b(\phi \otimes \psi) = \underbrace{\Gamma_{n,N}^{b}(\phi \otimes \psi) - \Gamma_{n-1,N}^{b}Q_{n,N}^{\otimes 2}C_{b_n}(\phi \otimes \psi)}_{R_1(N)}$$
$$+ \underbrace{\Gamma_{n-1,N}^{b}Q_{n,N}^{\otimes 2}C_{b_n}(\phi \otimes \psi) - \Gamma_{n-1,N}^{b}Q_n^{\otimes 2}C_{b_n}(\phi \otimes \psi)}_{R_2(N)}$$
$$+ \underbrace{\Gamma_{n-1,N}^{b}Q_n^{\otimes 2}C_{b_n}(\phi \otimes \psi) - \Gamma_{n-1}^{b}Q_n^{\otimes 2}C_{b_n}(\phi \otimes \psi)}_{R_3(N)}.$$

The tools to terminate the proof are the following ones:

- Lemma 4.1 shows that

$$R_1(N) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

- Lemma 4.2 and the fact that one may write $C_{b_n}(\phi \otimes \psi)$ as $f \otimes g$ for any $b_n$ ensure that

$$R_2(N) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

- Finally, the convergence rate

$$R_3(N) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

is a direct consequence of (14) and the induction hypothesis.

4.4. *Technical results.* This section presents some useful technical results. Before going further, remind that

$$\Gamma_{n,N}^b(1) := \gamma_n^N(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_{0:n}^{[2]} \in ((N)^2)^{\times(n+1)}} \left\{ \prod_{p=0}^{n-1} \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\}.$$

If we set

$$(16) \qquad \Lambda_n^{\ell_n^{[2]}} := \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{ \prod_{p=0}^{n-1} \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\}$$

together with the convention $\Lambda_0^{\ell_0^{[2]}} := 1$, we may write

$$(17) \qquad \Gamma_{n,N}^b(1) := \gamma_n^N(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_n^{[2]} \in (N)^2} \Lambda_n^{\ell_n^{[2]}},$$

so that

$$(18) \qquad \Gamma_{n,N}^b(1)^2 = \gamma_n^N(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2}} \Lambda_n^{\ell_n^{[2]}} \Lambda_n^{\ell_n'^{[2]}}.$$

Note that (17) is still true when $n = 0$. Then, for $n \geq 1$, we have by definition

$$(19) \qquad \Lambda_n^{\ell_n^{[2]}} = \sum_{\ell_{n-1}^{[2]} \in (N)^2} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}).$$

This decomposition will appear several times in the sequel for it is a keystone to study the behavior of the coalescent tree-based measures.

PROPOSITION 4.1. Assume $\mathcal{A}1$. For any coalescence indicator $b$, we have

$$\sup_{N > 1} \mathrm{E}\left[\Gamma_{n,N}^b(1)^2\right] < +\infty.$$

In particular, the sequence $(\Gamma_{n,N}^b(1); N \geq 1)$ is uniformly tight.

PROOF. We give a proof by induction. The verification for step 0 is trivial as $\Gamma_{0,N}^b(1) = 1$. For $n \geq 1$, we suppose that

$$\sup_{N>1} \mathrm{E}\left[\Gamma_{n-1,N}^b(1)^2\right] < +\infty.$$

As defined in Section 2.2, the IPS associated with ASMC is a Markov chain $(X_n)_{n\geq0}$ with genealogy $(A_n)_{n\geq0}$ tracking the indice of the parent of each particle at each level. More precisely, $A_{p-1}^i = j$ means that the parent of the particle $X_p^i$ is $X_{p-1}^j$. Accordingly, the filtration $(\mathcal{G}_n^N)_{n\geq0}$ with the genealogy of the IPS is defined by

$$\mathcal{G}_0^N := \sigma(X_0)$$

and, for $n \geq 1$,

$$\mathcal{G}_n^N := \sigma(A_0, \ldots, A_{n-1}, X_0, \ldots, X_n).$$

By combining (18) and (19), and taking into account that

$$\gamma_n^N(1) = \prod_{p=0}^{n-1} \eta_p^N(G_{p,N}) = \gamma_{n-1}^N(1)\eta_{n-1}^N(G_{n-1,N}) = \gamma_{n-1}^N(1)m(X_{n-1})(G_{n-1,N})$$

is $\mathcal{G}_{n-1}^N$-measurable, we have

$$(20) \quad \mathrm{E}\left[\Gamma_{n,N}^b(1)^2 \,\big|\, \mathcal{G}_{n-1}^N\right] = \gamma_{n-1}^N(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_{n-1}^{[2]}, \ell_{n-1}'^{[2]}) \in ((N)^2)^{\times 2}} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \Lambda_{n-1}^{\ell_{n-1}'^{[2]}}$$

$$\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2}} m(X_{n-1})(G_{n-1,N})^4 \, \mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\bigg|\, \mathcal{G}_{n-1}^N\right].$$

For the notation concerning the indices in the IPS, we use

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

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

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

With a slight abuse of notation, we admit that

$$((i, j), (k, l)) = (i, j, k, l).$$

With this notation, for $N \geq 4$, we have the decomposition

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

The idea of the proof consists in analyzing (20) with respect to the three terms that appear in the right-hand side of the latter. Recall from (3) that, given $X_{n-1}$, we make an independent multinomial selection of the parent of each particle at step $n$ according to the discrete probability measure

$$S_{n-1,N}(X_{n-1}, \cdot) = \sum_{k=1}^N \frac{G_{n-1,N}(X_{n-1}^k)}{\sum_{j=1}^N G_{n-1,N}(X_{n-1}^j)} \delta_k = \sum_{k=1}^N \frac{G_{n-1,N}(X_{n-1}^k)}{N \, m(X_{n-1})(G_{n-1,N})} \delta_k,$$

with, for all $k \in [N]$,

$$0 < \frac{G_{n-1,N}(X_{n-1}^k)}{N\, m(X_{n-1})(G_{n-1,N})} \leq \frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(X_{n-1})(G_{n-1,N})}.$$

We also recall that

$$\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) = 1_{\{b_{n-1}=0\}} 1_{\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1 \neq A_{n-1}^{\ell_n^2}=\ell_{n-1}^2\}} + 1_{\{b_{n-1}=1\}} 1_{\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1=A_{n-1}^{\ell_n^2} \neq \ell_{n-1}^2\}}.$$

- Case 1: $(\ell_n^{[2]}, \ell_n^{'[2]}) \in ((N)^2)^{\times 2} \cap [N]_2^4$.

  In this case, there are only two distinct random variables among $A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}, A_{n-1}^{\ell_n^{'1}}, A_{n-1}^{\ell_n^{'2}}$. Recall that $\ell_n^1 \neq \ell_n^2$ by construction. Let us first suppose that

$$\ell_n^1 = \ell_n^{'1} \quad \text{and} \quad \ell_n^2 = \ell_n^{'2}.$$

Thus, we deduce that

$$\mathrm{E}\left[ \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}}, \ell_{n-1}^{'[2]}) \,\middle|\, \mathcal{G}_{n-1}^N \right]$$

$$\leq \mathrm{E}\left[ \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N \right]$$

$$= 1_{\{b_{n-1}=0\}} \mathrm{P}\left( A_{n-1}^{\ell_n^1} = \ell_{n-1}^1, A_{n-1}^{\ell_n^2} = \ell_{n-1}^2 \,\middle|\, \mathcal{G}_{n-1}^N \right)$$

$$+ 1_{\{b_{n-1}=1\}} \mathrm{P}\left( A_{n-1}^{\ell_n^1} = \ell_{n-1}^1 = A_{n-1}^{\ell_n^2} \,\middle|\, \mathcal{G}_{n-1}^N \right)$$

$$\leq \left( \frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(X_{n-1})(G_{n-1,N})} \right)^2.$$

The analysis for the case where

$$\ell_n^1 = \ell_n^{'2} \quad \text{and} \quad \ell_n^2 = \ell_n^{'1}$$

is analogue. Hence, we conclude that

$$(21) \qquad \mathrm{E}\left[ \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}}, \ell_{n-1}^{'[2]}) \,\middle|\, \mathcal{G}_{n-1}^N \right] \leq \left( \frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(X_{n-1})(G_{n-1,N})} \right)^2.$$

Meanwhile, we notice that

$$\#\left( ((N)^2)^{\times 2} \cap [N]_2^4 \right) = 2N(N-1).$$

Putting all things together yields

$$\sum_{(\ell_n^{[2]}, \ell_n^{'[2]}) \in ((N)^2)^{\times 2} \cap [N]_2^4} m(X_{n-1})(G_{n-1,N})^4 \, \mathrm{E}\left[ \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}}, \ell_{n-1}^{'[2]}) \,\middle|\, \mathcal{G}_{n-1}^N \right]$$

$$\leq \frac{2(N-1)}{N} \|G_{n-1,\cdot}\|_\infty^4.$$

- Case 2: $(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \cap [N]_3^4$.

  As noticed in the previous case, the number of different indices within $(\ell_n^{[2]}, \ell_n'^{[2]})$ is the only thing that matters for the upper-bound in (21). Accordingly, the same reasoning gives this time

$$\mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq \left(\frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(X_{n-1})(G_{n-1,N})}\right)^3.$$

Since the total number of choices is

$$\#\left(((N)^2)^{\times 2} \cap [N]_3^4\right) = 4N(N-1)(N-2),$$

it comes

$$\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \cap [N]_3^4} m(X_{n-1})(G_{n-1,N})^4\, \mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]$$

$$\leq \frac{4(N-1)(N-2)}{N^2}\|G_{n-1,\cdot}\|_\infty^4.$$

- Case 3: $(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4$.

  This time, we get

$$\mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq \left(\frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(X_{n-1})(G_{n-1,N})}\right)^4,$$

and

$$\#\left((N)^4\right) = N(N-1)(N-2)(N-3),$$

so that

$$\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in (N)^4} m(X_{n-1})(G_{n-1,N})^4\, \mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]$$

$$\leq \frac{(N-1)(N-2)(N-3)}{N^3}\|G_{n-1,\cdot}\|_\infty^4.$$

As a consequence, since

$$2 + \frac{4(N-2)}{N} + \frac{(N-2)(N-3)}{N^2} \leq 7,$$

an upper-bound for (20) is

$$\mathrm{E}\left[\Gamma_{n,N}^b(1)^2 \,\middle|\, \mathcal{G}_{n-1}^N\right]$$

$$\leq 7\left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \frac{N-1}{N}\|G_{n-1,\cdot}\|_\infty^4\, \gamma_{n-1}^N(1)^4 \sum_{(\ell_{n-1}^{[2]}, \ell_{n-1}'^{[2]}) \in ((N)^2)^{\times 2}} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \Lambda_{n-1}^{\ell_{n-1}'^{[2]}}.$$

Replacing $n$ with $(n-1)$ in (18) allows us to conclude that

$$\mathrm{E}\left[\Gamma_{n,N}^b(1)^2 \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq 7\frac{N}{N-1}\|G_{n-1,\cdot}\|_\infty^4\, \Gamma_{n-1,N}^b(1)^2.$$

Finally, by applying the induction hypothesis, we have

$$\sup_{N \geq 4} \mathrm{E}\left[\Gamma_{n,N}^{b}(1)^2\right] = \sup_{N \geq 4} \mathrm{E}\left[\mathrm{E}\left[\Gamma_{n,N}^{b}(1)^2 \mid \mathcal{G}_{n-1}^{N}\right]\right] \leq \frac{28}{3} \left\|G_{n-1,\cdot}\right\|_{\infty}^{4} \sup_{N \geq 4} \mathrm{E}\left[\Gamma_{n-1,N}^{b}(1)^2\right] < +\infty,$$

which ends the proof of Proposition 4.1. $\qquad \square$

LEMMA 4.1. Under $\mathcal{A}1$, for any test functions $f, g \in \mathcal{B}_b(E_n)$, we have, for all $n \geq 1$,

(22)
$$\mathrm{E}\left[\Gamma_{n,N}^{b}(f \otimes g) \mid \mathcal{G}_{n-1}^{N}\right] = \Gamma_{n-1,N}^{b} Q_{n,N}^{\otimes 2} C_{b_n}(f \otimes g),$$

as well as

$$\Gamma_{n,N}^{b}(f \otimes g) - \Gamma_{n-1,N}^{b} Q_{n,N}^{\otimes 2} C_{b_n}(f \otimes g) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

PROOF. First, by exploiting the notation defined in (16), we have

$$\Gamma_{n,N}^{b}(f \otimes g) := \gamma_{n}^{N}(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_{n}^{[2]} \in (N)^2} \Lambda_{n}^{\ell_{n}^{[2]}} C_{b_n}(f \otimes g)(X_{n}^{\ell_{n}^{[2]}}),$$

and (22) is then a direct consequence of Proposition A.1 since for any $\ell_{n}^{[2]} \in (N)^2$

$$\mathrm{E}\left[\gamma_{n}^{N}(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \Lambda_{n}^{\ell_{n}^{[2]}} C_{b_n}(f \otimes g)(X_{n}^{\ell_{n}^{[2]}}) \mid \mathcal{G}_{n-1}^{N}\right] = \frac{1}{N(N-1)} \Gamma_{n-1,N}^{b} Q_{n,N}^{\otimes 2} C_{b_n}(f \otimes g),$$

where the right-hand side does not depend on $\ell_{n}^{[2]}$. Second, thanks to Chebyshev's inequality, it suffices to verify that

$$\mathrm{Var}\left[\Gamma_{n,N}^{b}(f \otimes g) - \Gamma_{n-1,N}^{b} Q_{n,N}^{\otimes 2} C_{b_n}(f \otimes g)\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

For this, by (22), we just have to show that

$$\mathrm{E}\left[\Gamma_{n,N}^{b}(f \otimes g)^2 - (\Gamma_{n-1,N}^{b} Q_{n,N}^{\otimes 2} C_{b_n}(f \otimes g))^2\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

Then, recall that, by definition,

$$\Gamma_{n,N}^{b}(f \otimes g)^2$$

$$= \gamma_{n}^{N}(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_{n}^{[2]}, \ell_{n}^{'[2]}) \in ((N)^2)^{\times 2}} \Lambda_{n}^{\ell_{n}^{[2]}} \Lambda_{n}^{\ell_{n}^{'[2]}} \left(C_{b_n}(f \otimes g)\right)^{\otimes 2} (X_{n}^{\ell_{n}^{[2]}}, X_{n}^{\ell_{n}^{'[2]}})$$

$$= \underbrace{\gamma_{n}^{N}(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_{n}^{[2]}, \ell_{n}^{'[2]}) \in (N)^4} \Lambda_{n}^{\ell_{n}^{[2]}} \Lambda_{n}^{\ell_{n}^{'[2]}} \left(C_{b_n}(f \otimes g)\right)^{\otimes 2} (X_{n}^{\ell_{n}^{[2]}}, X_{n}^{\ell_{n}^{'[2]}})}_{R_1(N)}$$

$$+ \underbrace{\gamma_{n}^{N}(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_{n}^{[2]}, \ell_{n}^{'[2]}) \in ((N)^2)^{\times 2} \backslash (N)^4} \Lambda_{n}^{\ell_{n}^{[2]}} \Lambda_{n}^{\ell_{n}^{'[2]}} \left(C_{b_n}(f \otimes g)\right)^{\otimes 2} (X_{n}^{\ell_{n}^{[2]}}, X_{n}^{\ell_{n}^{'[2]}})}_{R_2(N)}.$$

- For $R_1(N)$, our goal is to establish that

$$E\left[R_1(N) - (\Gamma^b_{n-1,N}Q^{\otimes 2}_{n,N}C_{b_n}(f \otimes g))^2\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

In fact, for any $(\ell^{[2]}_n, \ell^{'[2]}_n) \in (N)^4$,

$$(A^{\ell^1_n}_n, X^{\ell^1_n}_n, A^{\ell^2_n}_n, X^{\ell^2_n}_n) \qquad \text{and} \qquad (A^{\ell'^1_n}_n, X^{\ell'^1_n}_n, A^{\ell'^2_n}_n, X^{\ell'^2_n}_n)$$

are conditionally independent given $\mathcal{G}^N_{n-1}$ by construction of the IPS. Hence, by applying Proposition A.1 respectively for $\ell^{[2]}_n$ and for $\ell^{'[2]}_n$, we have

$$E\left[\gamma^N_n(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \Lambda^{\ell^{[2]}_n}_n \Lambda^{\ell^{'[2]}_n}_n C_{b_n}(f \otimes g)(X^{\ell^{[2]}_n}_n)C_{b_n}(f \otimes g)(X^{\ell^{'[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right]$$

$$= \frac{1}{N^2(N-1)^2}(\Gamma^N_{n-1}Q^{\otimes 2}_{n,N}C_{b_n}(f \otimes g))^2.$$

Then, since

$$\#\left((N)^4\right) = N(N-1)(N-2)(N-3),$$

we deduce that

$$E\left[R_1(N) - (\Gamma^b_{n-1,N}Q^{\otimes 2}_{n,N}C_{b_n}(f \otimes g))^2 \,\middle|\, \mathcal{G}^N_{n-1}\right]$$

$$= \left(\frac{N(N-1)(N-2)(N-3)}{N^2(N-1)^2} - 1\right)(\Gamma^N_{n-1}Q^{\otimes 2}_{n,N}C_{b_n}(f \otimes g))^2,$$

and

$$E\left[R_1(N) - (\Gamma^b_{n-1,N}Q^{\otimes 2}_{n,N}C_{b_n}(f \otimes g))^2\right]$$

$$= \left(\frac{N(N-1)(N-2)(N-3)}{N^2(N-1)^2} - 1\right)E\left[(\Gamma^N_{n-1}Q^{\otimes 2}_{n,N}C_{b_n}(f \otimes g))^2\right] = \mathcal{O}\left(\frac{1}{N}\right),$$

where the final equality is due to Proposition 4.1, taking into account that $f$ and $g$ are bounded, and so is $G_{n-1,N}$ uniformly with respect to $N$ by $\mathcal{A}1$.

- For $R_2(N)$, the nonnegativity of $\Lambda^{\ell^{[2]}_n}_n$ implies

$$E[R_2(N)] \le E\left[\gamma^N_n(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell^{[2]}_n, \ell^{'[2]}_n) \in ((N)^2)^{\times 2}\setminus(N)^4} \Lambda^{\ell^{[2]}_n}_n \Lambda^{\ell^{'[2]}_n}_n\right]\|f\|^2_\infty \|g\|^2_\infty.$$

So the proof will be finished once we have shown that

$$E\left[\gamma^N_n(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell^{[2]}_n, \ell^{'[2]}_n) \in ((N)^2)^{\times 2}\setminus(N)^4} \Lambda^{\ell^{[2]}_n}_n \Lambda^{\ell^{'[2]}_n}_n\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

Once again, we proceed by induction. At step 0, we have

$$\frac{1}{N^2(N-1)^2} \sum_{(\ell^{[2]}_0, \ell^{'[2]}_0) \in ((N)^2)^{\times 2}\setminus(N)^4} 1 = 1 - \frac{N(N-1)(N-2)(N-3)}{N^2(N-1)^2} = \mathcal{O}\left(\frac{1}{N}\right).$$

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

$$\mathrm{E}\left[\gamma_{n-1}^N(1)^4\left(\frac{N^{n-2}}{(N-1)^n}\right)^2\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{'[2]})\in((N)^2)^{\times 2}\backslash(N)^4}\Lambda_{n-1}^{\ell_{n-1}^{[2]}}\Lambda_{n-1}^{\ell_{n-1}^{'[2]}}\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

The adaptation of (20) to the present context gives

$$\mathrm{E}\left[\gamma_n^N(1)^4\left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}\backslash(N)^4}\Lambda_n^{\ell_n^{[2]}}\Lambda_n^{\ell_n^{'[2]}}\,\middle|\,\mathcal{G}_{n-1}^N\right]$$

$$= \gamma_{n-1}^N(1)^4\left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{'[2]})\in((N)^2)^{\times 2}}\Lambda_{n-1}^{\ell_{n-1}^{[2]}}\Lambda_{n-1}^{\ell_{n-1}^{'[2]}}$$

$$\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}\backslash(N)^4}m(X_{n-1})(G_{n-1,N})^4$$

$$\mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}},\ell_{n-1}^{'[2]})\,\middle|\,\mathcal{G}_{n-1}^N\right].$$

Now, for $N \geq 4$, it is clear that

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

– Case 1: $(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}\cap[N]_2^4$.
  By definition of $S_{n-1,N}$ in (3),

$$\mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}},\ell_{n-1}^{'[2]})\,\middle|\,\mathcal{G}_{n-1}^N\right]$$

$$\leq \left(\frac{\|G_{n-1,\cdot}\|_\infty}{N\,m(X_{n-1})(G_{n-1,N})}\right)^2\left(1_{\{b_{n-1}=1,\ell_{n-1}^1=\ell_{n-1}^{'1}\}}+1_{\{b_{n-1}=0,\ell_{n-1}^1=\ell_{n-1}^{'1},\ell_{n-1}^2=\ell_{n-1}^{'2}\}}\right)$$

$$\leq \left(\frac{\|G_{n-1,\cdot}\|_\infty}{N\,m(X_{n-1})(G_{n-1,N})}\right)^2 1_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^{'1},\ell_{n-1}^2,\ell_{n-1}^{'2}\}<4\}},$$

and since

$$\#\left(((N)^2)^{\times 2}\cap[N]_2^4\right) = 2N(N-1),$$

it comes

$$\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}\cap[N]_2^4}m(X_{n-1})(G_{n-1,N})^4$$

$$\mathrm{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}},\ell_{n-1}^{'[2]})\,\middle|\,\mathcal{G}_{n-1}^N\right]$$

$$\leq \frac{2N(N-1)}{N^2}\|G_{n-1,\cdot}\|_\infty^4 1_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^{'1},\ell_{n-1}^2,\ell_{n-1}^{'2}\}<4\}}.$$

- Case 2: $(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \cap [N]_3^4$.

  First, we suppose that $\ell_n^1 = \ell_n'^1$. As for the previous case, we have

$$
\mathrm{E}\left[ \lambda_{n-1}^b (A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b (A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\bigg|\, \mathcal{G}_{n-1}^N \right]
$$

$$
\leq \left( \frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(\mathrm{X}_{n-1})(G_{n-1,N})} \right)^3 \left( 1_{\{b_{n-1}=1, \ell_{n-1}^1 = \ell_{n-1}'^1\}} + 1_{\{b_{n-1}=0, \ell_{n-1}^1 = \ell_{n-1}'^1\}} \right)
$$

$$
\leq \left( \frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(\mathrm{X}_{n-1})(G_{n-1,N})} \right)^3 1_{\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\} < 4\}}.
$$

By the same reasoning, for $\ell_n^1 = \ell_n'^2$, $\ell_n^2 = \ell_n'^1$ and $\ell_n^2 = \ell_n'^2$, we also have

$$
\mathrm{E}\left[ \lambda_{n-1}^b (A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b (A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\bigg|\, \mathcal{G}_{n-1}^N \right]
$$

$$
\leq \left( \frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(\mathrm{X}_{n-1})(G_{n-1,N})} \right)^3 1_{\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\} < 4\}}.
$$

In addition, since

$$
\# \left( ((N)^2)^{\times 2} \cap [N]_3^4 \right) = 4N(N-1)(N-2),
$$

we get this time

$$
\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \cap [N]_3^4} m(\mathrm{X}_{n-1})(G_{n-1,N})^4
$$

$$
\mathrm{E}\left[ \lambda_{n-1}^b (A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda_{n-1}^b (A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\bigg|\, \mathcal{G}_{n-1}^N \right]
$$

$$
\leq \frac{4N(N-1)(N-2)}{N^3} \|G_{n-1,\cdot}\|_\infty^4 1_{\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\} < 4\}}.
$$

By gathering both cases, we have

$$
\mathrm{E}\left[ \gamma_n^N(1)^4 \left( \frac{N^{n-1}}{(N-1)^{n+1}} \right)^2 \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \backslash (N)^4} \Lambda_n^{\ell_n^{[2]}} \Lambda_n^{\ell_n'^{[2]}} \,\bigg|\, \mathcal{G}_{n-1}^N \right]
$$

$$
\leq \gamma_{n-1}^N(1)^4 \left( \frac{N^{n-2}}{(N-1)^n} \right)^2 \frac{N^2}{(N-1)^2} \left( \frac{2N(N-1)}{N^2} + \frac{4N(N-1)(N-2)}{N^3} \right) \|G_{n-1,\cdot}\|_\infty^4
$$

$$
\sum_{(\ell_{n-1}^{[2]}, \ell_{n-1}'^{[2]}) \in ((N)^2)^{\times 2}} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \Lambda_{n-1}^{\ell_{n-1}'^{[2]}} 1_{\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\} < 4\}}
$$

$$
\leq 6 \frac{N^2}{(N-1)^2} \|G_{n-1,\cdot}\|_\infty^4 \gamma_{n-1}^N(1)^4 \left( \frac{N^{n-2}}{(N-1)^n} \right)^2 \sum_{(\ell_{n-1}^{[2]}, \ell_{n-1}'^{[2]}) \in ((N)^2)^{\times 2} \backslash (N)^4} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \Lambda_{n-1}^{\ell_{n-1}'^{[2]}}.
$$

The conclusion follows from the induction hypothesis by taking the expectation on both sides.

This ends the proof of Lemma 4.1. □

LEMMA 4.2. Assume $\mathcal{A}1$-$\mathcal{A}2$, then for any test functions $f, g \in \mathcal{B}_b(E_n)$ and for all $n \geq 1$,

$$\Gamma^b_{n-1,N} Q^{\otimes 2}_{n,N}(f \otimes g) - \Gamma^b_{n-1,N} Q^{\otimes 2}_n(f \otimes g) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

PROOF. The verification shares some resemblance with the convergence of $P_2(N)$ in the proof of Theorem 4.1. Specifically, we start with the following decomposition:

$$\left|\Gamma^b_{n-1,N} Q^{\otimes 2}_{n,N}(f \otimes g) - \Gamma^b_{n-1,N} Q^{\otimes 2}_n(f \otimes g)\right|$$

$$\leq \underbrace{\left|\Gamma^b_{n-1,N} Q^{\otimes 2}_{n,N}(f \otimes g) - \Gamma^b_{n-1,N}(Q_{n,N} \otimes Q_n)(f \otimes g)\right|}_{D_1(N)}$$

$$+ \underbrace{\left|\Gamma^b_{n-1,N}(Q_{n,N} \otimes Q_n)(f \otimes g) - \Gamma^b_{n-1,N} Q^{\otimes 2}_n(f \otimes g)\right|}_{D_2(N)}.$$

For $D_1(N)$, we may write

$$(Q^{\otimes 2}_{n,N}(f \otimes g) - (Q_{n,N} \otimes Q_n)(f \otimes g))(x, y) = Q_{n,N}(f)(x)(Q_{n,N}(g)(y) - Q_n(g)(y)).$$

By $\mathcal{A}2$, for any $g \in \mathcal{B}_b(E_n)$, there exists a bounded function $h_{n-1}$ such that

$$\left|Q_{n,N}(g)(y) - Q_n(g)(y)\right| = \left|\langle h_{n-1}(y, Z^N_{n-1}), Z^N_{n-1} - z^*_{n-1}\rangle\right| \leq \|h_{n-1}\|_\infty \left|Z^N_{n-1} - z^*_{n-1}\right|.$$

Since, in addition,

$$|Q_{n,N}(f)(x)| \leq \left\|G_{n-1,\cdot}\right\|_\infty \|f\|_\infty,$$

it comes

$$D_1(N) \leq \Gamma^b_{n-1,N}(1) \left\|G_{n-1,\cdot}\right\|_\infty \|f\|_\infty \|h_{n-1}\|_\infty \left|Z^N_{n-1} - z^*_{n-1}\right|.$$

By Proposition 4.1, one has

$$\Gamma^b_{n-1,N}(1) = \mathcal{O}_p(1).$$

In addition, a by-product (11) of Theorem 2.1 is that

$$\left|Z^N_{n-1} - z^*_{n-1}\right| = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

Hence, one concludes that

$$D_1(N) = \mathcal{O}_p\left(\frac{1}{\sqrt{N}}\right).$$

The reasoning for $D_2(N)$ is the same. □

4.5. *Connection between the estimators.* In this section, we give some combinatorial results on the coalescent tree-based measures $\Gamma^b_{n,N}$. In particular, they allow us to connect the variance estimator (9) of Lee & Whiteley and our term by term estimators. As mentioned before, these relations do not depend on $\mathcal{A}2$: they are provided by the structure of the IPS and the underlying multinomial selection scheme. In this respect, recall that, under $\mathcal{A}1$, Equation (3) is always well-defined for the denominator is always strictly positive, and the same holds true for the IPS itself. This is in fact the only condition required here.

PROPOSITION 4.2. Provided that the IPS is well-defined, then for any test function $F \in \mathcal{B}_b(E_n^2)$, we have the decompositions:

$$(\gamma_n^N)^{\otimes 2}(F) = \sum_{b \in \{0,1\}^{n+1}} \left\{ \prod_{p=0}^n \frac{(N-1)^{1-b_p}}{N} \right\} \Gamma_{n,N}^b(F),$$

and

$$(\eta_n^N)^{\otimes 2}(F) = \sum_{b \in \{0,1\}^{n+1}} \left\{ \prod_{p=0}^n \frac{(N-1)^{1-b_p}}{N} \right\} \bar{\Gamma}_{n,N}^b(F).$$

PROOF. Since

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

we have, for any $b = (b_0, b_1, \ldots, b_n) \in \{0,1\}^{n+1}$,

$$\left\{ \prod_{p=0}^n \frac{(N-1)^{1-b_p}}{N} \right\} \bar{\Gamma}_{n,N}^b(F)$$

$$= \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_{0:n}^{[2]} \in ((N)^2)^{\times(n+1)}} \left\{ \prod_{p=0}^n \frac{(N-1)^{1-b_p}}{N} \right\} \left\{ \prod_{p=0}^{n-1} \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\} C_{b_n}(F)(X_n^{\ell_n^{[2]}}).$$

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

$$\sum_{b \in \{0,1\}^{n+1}} \left\{ \prod_{p=0}^n \frac{(N-1)^{1-b_p}}{N} \right\} \bar{\Gamma}_{n,N}^b(F)$$

$$= \sum_{\ell_0^{[2]} \in (N)^2} \cdots \sum_{\ell_{n-1}^{[2]} \in (N)^2} \left\{ \prod_{p=0}^{n-1} \left( \frac{1}{N} 1_{\{A_p^{\ell_{p+1}^1} = A_p^{\ell_{p+1}^2} = \ell_p^1 \neq \ell_p^2\}} + \frac{N-1}{N} 1_{\{A_p^{\ell_{p+1}^1} = \ell_p^1 \neq A_p^{\ell_{p+1}^2} = \ell_p^2\}} \right) \right\}$$

$$\left( \frac{N}{N-1} \right)^n \left\{ \frac{N-1}{N} m^{\odot 2}(X_n) C_0(F) + \frac{1}{N} m^{\odot 2}(X_n) C_1(F) \right\}.$$

To conclude, one just has to observe that

$$\sum_{\ell_0^{[2]} \in (N)^2} \cdots \sum_{\ell_{n-1}^{[2]} \in (N)^2} \left\{ \prod_{p=0}^{n-1} \left( \frac{1}{N} 1_{\{A_p^{\ell_{p+1}^1} = A_p^{\ell_{p+1}^2} = \ell_p^1 \neq \ell_p^2\}} + \frac{N-1}{N} 1_{\{A_p^{\ell_{p+1}^1} = \ell_p^1 \neq A_p^{\ell_{p+1}^2} = \ell_p^2\}} \right) \right\} = \left( \frac{N-1}{N} \right)^n$$

while, by (1),

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

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

We can now proceed with the proof of Proposition 3.1. Recall that the goal is to show that

$$NV_n^N(f) - \sigma_{\eta_{n,N}}^2(f) = NV_n^N(f) - \sum_{p=0}^{n} \left\{ \bar{\Gamma}_{n,N}^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}) \right\} = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{N}\right),$$

and

$$NV_n^N(f - \eta_n^N(f)) - \sum_{p=0}^{n} \left\{ \bar{\Gamma}_{n,N}^{(p)}\left( \left[ f - \eta_n^N(f) \right]^{\otimes 2} \right) - \bar{\Gamma}_{n,N}^{(\varnothing)}\left( \left[ f - \eta_n^N(f) \right]^{\otimes 2} \right) \right\} = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{N}\right).$$

By construction, we have

$$V_n^N(f) = \eta_n^N(f)^2 - \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}) = (\eta_n^N)^{\otimes 2}(f^{\otimes 2}) - \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}).$$

An implication of Proposition 4.1 is that, for any test function $f$ and any coalescence indicator $b$,

$$\bar{\Gamma}_{n,N}^{b}(f^{\otimes 2}) = \mathcal{O}_{\mathrm{p}}\left( 1 \right).$$

Thus, a consequence of Proposition 4.2 is

$$(\eta_n^N)^{\otimes 2}(f^{\otimes 2}) = \left(\frac{N-1}{N}\right)^{n+1} \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}) + \frac{1}{N}\left(\frac{N-1}{N}\right)^{n} \sum_{p=0}^{n} \bar{\Gamma}_{n,N}^{(p)}(f^{\otimes 2}) + \mathcal{O}_{\mathrm{p}}\left(\frac{1}{N^2}\right).$$

The desired formula is then obtained by remarking that

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

Similarly, since

$$\bar{\Gamma}_{n,N}^{b}\left( \left[ f - \eta_n^N(f) \right]^{\otimes 2} \right) = \mathcal{O}_{\mathrm{p}}\left( 1 \right),$$

the same algebraic manipulation yields

$$NV_n^N(f - \eta_n^N(f)) - \sum_{p=0}^{n} \left\{ \bar{\Gamma}_{n,N}^{(p)}\left( \left[ f - \eta_n^N(f) \right]^{\otimes 2} \right) - \bar{\Gamma}_{n,N}^{(\varnothing)}\left( \left[ f - \eta_n^N(f) \right]^{\otimes 2} \right) \right\} = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{N}\right).$$

This closes the proof of Proposition 3.1.

## APPENDIX A: MANY-BODY FEYNMAN-KAC MODELS

The many-body Feynman-Kac model was proposed in [13] to study the propagation of chaos property of the Conditional Particle Markov Chain Monte Carlo introduced in [1]. The basic idea is to trace the information of all particles in the IPS along with its genealogy, and to construct an instrumental particle block which is heavily dependent (identical) to some specific particles. We call these instrumental particles the *coupled particle block* of the IPS.

**A.1. Duality formula.** At each layer, the particles in the original IPS are denoted by $X_p$, with its genealogy $A_{p-1}$. The coupled particle block of $q$ particles is denoted by $\tilde{X}_p^{[q]}$, with its genealogy denoted by $\tilde{A}_{p-1}^{[q]}$. The corresponding variables in the integral operators will be denoted by $x_p$, $a_{p-1}$, $\tilde{x}_p^{[q]}$ and $\tilde{a}_{p-1}^{[q]}$ respectively.

Before giving specific definitions, we want to mention that the mathematical object we would like to look into is the *whole particle system*, namely the *original IPS* and the *coupled particle block* with genealogy. At each layer $p$, we are interested by the tuple:

$$(X_p, A_{p-1}, \tilde{X}_p^{[q]}, \tilde{A}_{p-1}^{[q]}).$$

As for the basic idea of Particle Markov Chain Monte Carlo method [1], we study respectively the distributions of

$$X_p, A_{p-1} \mid \tilde{X}_p^{[q]}, \tilde{A}_{p-1}^{[q]}$$

and

$$\tilde{X}_p^{[q]}, \tilde{A}_{p-1}^{[q]} \mid X_p, A_{p-1}.$$

Thanks to the specific construction, as well as the relatively simple multinomial resampling scheme of Feynman-Kac IPS, Lemma A.1 provides a duality formula to connect both distributions and leads in particular to Proposition A.1. This latter result is crucial to prove the consistency of our term by term estimator in Theorem 3.1.

In this section, a transition kernel denoted by the letter $Q$ is a Feynman-Kac kernel, meaning that its total mass is not necessarily 1, and it can be expressed by the product of a positive potential function and a Markov kernel. All transition kernels denoted by the letter $M$ are Markov kernels.

Notice that the transition from level $p-1$ to level $p$ of the IPS with its genealogy defined in Section 2.2 can be expressed as

$$(A_{p-1}, X_p) \sim \bigotimes_{i=1}^{N} \Phi_{p,N}(X_{p-1}, d(A_{p-1}^i, X_p^i))$$

with $\Phi_{p,N}$ defined by

$$\Phi_{p,N}(\boldsymbol{x_{p-1}}, d(a_{p-1}^i, x_p^i)) = S_{p-1,N}(\boldsymbol{x_{p-1}}, da_{p-1}^i) \times M_{p,N}(x_{p-1}^{a_{p-1}^i}, dx_p^i).$$

We define the transition of the original IPS with its genealogy by

$$\mathcal{M}_p(\boldsymbol{x_{p-1}}, d(\boldsymbol{a_{p-1}}, \boldsymbol{x_p})) := \prod_{i=1}^{N} \Phi_{p,N}(\boldsymbol{x_{p-1}}, d(a_{p-1}^i, x_p^i))$$

and the potential function of the particle block of size $q$ by

$$\mathcal{G}_{p-1}^{(q)}(\boldsymbol{x_{p-1}}) := m(\boldsymbol{x_{p-1}})(G_{p-1,N})^q.$$

We denote the associated Feynman-Kac transition kernel

$$Q_p^{(q)}(\boldsymbol{x_{p-1}}, d(\boldsymbol{a_{p-1}}, \boldsymbol{x_p})) := \mathcal{G}_{p-1}^{(q)}(\boldsymbol{x_{p-1}}) \times \mathcal{M}_p(\boldsymbol{x_{p-1}}, d(\boldsymbol{a_{p-1}}, \boldsymbol{x_p})).$$

Given $\ell_p^{[q]} \in (N)^q$, $\tilde{a}_{p-1}^{[q]} \in [N]^q$ and $\tilde{x}_p^{[q]} \in E_p^q$, we define

$$\mathbb{M}_p^{\tilde{a}_{p-1}^{[q]}, \ell_p^{[q]}, \tilde{x}_p^{[q]}}(\boldsymbol{x_{p-1}}, d(\boldsymbol{a_{p-1}}, \boldsymbol{x_p}))$$

$$:= \prod_{i \in [N] \setminus \{\ell_p^1, \dots, \ell_p^q\}} \left\{ \Phi_{p,N}(\boldsymbol{x_{p-1}}, d(a_{p-1}^i, x_p^i)) \right\} \times \delta_{\tilde{x}_p^{[q]}}(dx_p^{\ell_p^{[q]}}) \times \delta_{\tilde{a}_{p-1}^{[q]}}(da_{p-1}^{\ell_p^{[q]}})$$

the conditional transition for the original particle system given the coupled particle block $\tilde{X}_p^{[q]} = \tilde{x}_p^{[q]}$ at position $\ell_p^{[q]}$ with frozen genealogy $\tilde{A}_{p-1}^{[q]} = \tilde{a}_{p-1}^{[q]}$. In particular, we denote

$$\mathbb{M}_0^{\ell_0^{[q]}, \tilde{x}_0^{[q]}}(d\boldsymbol{x_0}) := \left\{ \prod_{i \in [N] \setminus \{\ell_0^1, \dots, \ell_0^q\}} \eta_0(dx_0^i) \right\} \times \delta_{\tilde{x}_0^{[q]}}(dx_0^{\ell_0^{[q]}}).$$

We also define

$$\mathbb{Q}_p^{(q)}(\boldsymbol{x_{p-1}}, d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})) := m([N])^{\otimes q}(d\tilde{a}_{p-1}^{[q]}) \, Q_{p,N}^{\otimes q}(x_{p-1}^{\tilde{a}_{p-1}^{[q]}}, d\tilde{x}_p^{[q]}),$$

and

$$\mathrm{M}_p^{\ell_p^{[q]}}\Big((\boldsymbol{a_{p-1}}, \boldsymbol{x_p}), d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})\Big) := \delta_{a_{p-1}^{\ell_p^{[q]}}}(d\tilde{a}_{p-1}^{[q]}) \, \delta_{x_p^{\ell_p^{[q]}}}(d\tilde{x}_p^{[q]}).$$

Then we have the following pivotal duality formula, which is simply a generalization of Lemma 4.1 in [13]. We will apply it in the proof of Proposition A.1 with $q = 2$.

LEMMA A.1. For $p \geq 1$, $q \in [N]$ and $\ell_p^{[q]} \in (N)^q$, we have the following duality formula between integral operators

$$Q_p^{(q)}(\boldsymbol{x_{p-1}}, d(\boldsymbol{a_{p-1}}, \boldsymbol{x_p})) \, \mathrm{M}_p^{\ell_p^{[q]}}\Big((\boldsymbol{a_{p-1}}, \boldsymbol{x_p}), d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})\Big)$$

$$= \mathbb{Q}_p^{(q)}(\boldsymbol{x_{p-1}}, d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})) \, \mathbb{M}_p^{\tilde{a}_{p-1}^{[q]}, \ell_p^{[q]}, \tilde{x}_p^{[q]}}(\boldsymbol{x_{p-1}}, d(\boldsymbol{a_{p-1}}, \boldsymbol{x_p})),$$

and

$$\eta_0^{\otimes N}(d\boldsymbol{x_0}) \, \delta_{x_0^{\ell_0^{[q]}}}(d\tilde{x}_0^{[q]}) = \eta_0^{\otimes q}(d\tilde{x}_0^{[q]}) \, \mathbb{M}_0^{\ell_0^{[q]}, \tilde{x}_0^{[q]}}(d\boldsymbol{x_0}).$$

PROOF. Step 0 is clear. For $p \geq 1$, it suffices to check that the nonidentical parts are equal, namely

$$\mathcal{G}_{p-1}^{(q)}(\boldsymbol{x_{p-1}}) \left\{ \sum_{k=1}^n \frac{G_{p-1,N}(x_{p-1}^k)}{N \, m(\boldsymbol{x_{p-1}})(G_{p-1,N})} \delta_k \right\}^{\otimes q} (da_p^{\ell_p^{[q]}}) \, M_{p,N}^{\otimes q}(x_{p-1}^{a_{p-1}^{\ell_p^{[q]}}}, dx_p^{\ell_p^{[q]}})$$

$$\mathrm{M}_p^{\ell_p^{[q]}}\Big((\boldsymbol{a_{p-1}}, \boldsymbol{x_p}), d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})\Big) = \mathbb{Q}_p^{(q)}(\boldsymbol{x_{p-1}}, d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})) \, \delta_{\tilde{x}_p^{[q]}}(dx_p^{\ell_p^{[q]}}) \, \delta_{\tilde{a}_{p-1}^{[q]}}(da_{p-1}^{\ell_p^{[q]}}).$$

Fixing $\ell_p^{[q]} \in (N)^q$ and $\boldsymbol{x_{p-1}} \in E_{p-1}^N$, consider a function $\mathrm{F} \in \mathcal{B}_b([N]^q \times [N]^q \times E_p^q \times E_p^q)$. Moreover, let $a^{[q]} = (a_1, \dots, a_q)$ and $x^{[q]} = (x_1, \dots, x_q)$ denote generic variables belonging respectively to $[N]^q$ and $E_p^q$. Then, we may write

$$\int m(\boldsymbol{x_{p-1}})(G_{p-1,N})^q \left\{ \sum_{k=1}^N \frac{G_{p-1,N}(x_{p-1}^k)}{N \, m(\boldsymbol{x_{p-1}})(G_{p-1,N})} \delta_k \right\}^{\otimes q} (da_p^{\ell_p^{[q]}})$$

$$M_{p,N}^{\otimes q}(x_{p-1}^{a_{p-1}^{\ell_p^{[q]}}}, dx_p^{\ell_p^{[q]}}) \, \delta_{x_p^{\ell_p^{[q]}}}(d\tilde{x}_p^{[q]}) \, \delta_{a_{p-1}^{\ell_p^{[q]}}}(d\tilde{a}_{p-1}^{[q]}) \, \mathrm{F}(\tilde{a}_{p-1}^{[q]}, a_{p-1}^{\ell_p^{[q]}}, x_p^{\ell_p^{[q]}}, \tilde{x}_p^{[q]})$$

$$= \int m([N])^{\otimes q}(da^{[q]}) \, G_{p-1,N}^{\otimes q}(x_{p-1}^{a^{[q]}}) \, M_{p,N}^{\otimes q}(x_{p-1}^{a^{[q]}}, dx^{[q]}) \, \mathrm{F}(a^{[q]}, a^{[q]}, x^{[q]}, x^{[q]})$$

$$= \int m([N])^{\otimes q}(da^{[q]}) \, Q_{p,N}^{\otimes q}(x_{p-1}^{a^{[q]}}, dx^{[q]}) \, \mathrm{F}(a^{[q]}, a^{[q]}, x^{[q]}, x^{[q]})$$

$$= \int m([N])^{\otimes q}(d\tilde{a}_{p-1}^{[q]}) \, Q_{p,N}^{\otimes q}(x_{p-1}^{\tilde{a}_{p-1}^{[q]}}, d\tilde{x}_p^{[q]}) \, \delta_{\tilde{x}_p^{[q]}}(dx_p^{\ell_p^{[q]}}) \, \delta_{\tilde{a}_{p-1}^{[q]}}(da_p^{\ell_p^{[q]}}) \, \mathrm{F}(\tilde{a}_{p-1}^{[q]}, a_p^{\ell_p^{[q]}}, x_p^{\ell_p^{[q]}}, \tilde{x}_p^{[q]})$$

$$= \int \mathbb{Q}_p^{(q)}(\boldsymbol{x_{p-1}}, d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})) \, \delta_{\tilde{x}_p^{[q]}}(dx_p^{\ell_p^{[q]}}) \, \delta_{\tilde{a}_{p-1}^{[q]}}(da_p^{\ell_p^{[q]}}) \, \mathrm{F}(\tilde{a}_{p-1}^{[q]}, a_p^{\ell_p^{[q]}}, x_p^{\ell_p^{[q]}}, \tilde{x}_p^{[q]}).$$

This ends the proof of the duality formula.                                    □

Let us recall (16) and (19):

$$(23) \qquad \Lambda_n^{\ell_n^{[2]}} = \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{ \prod_{p=0}^{n-1} \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\} = \sum_{\ell_{n-1}^{[2]} \in (N)^2} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}).$$

with the convention $\Lambda_0^{\ell_0^{[2]}} = 1$. In fact, this gives another representation of the approximation of the coalescent tree-based measures:

$$\Gamma_{n,N}^b(f \otimes g) = \gamma_n^N(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_n^{[2]} \in (N)^2} \Lambda_n^{\ell_n^{[2]}} C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}}).$$

Recall that

$$\mathcal{G}_{n-1}^N := \sigma(X_0, \ldots, X_{n-1}, A_0, \ldots, A_{n-2}).$$

The upcoming result is useful in the proof of Lemma 4.1.

PROPOSITION A.1.    Under $\mathcal{A}1$, for any $\ell_n^{[2]} \in (N)^2$, any coalescence indicator $b$, and any test functions $f$ and $g$ in $\mathcal{B}_b(E_n)$, we have, for all $n \geq 1$, that

$$E\left[ \gamma_n^N(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \Lambda_n^{\ell_n^{[2]}} C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N \right] = \frac{1}{N(N-1)} \Gamma_{n-1,N}^b Q_{n,N}^{\otimes 2} C_{b_n}(f \otimes g).$$

PROOF.  By applying (23), we obtain

$$\gamma_n^N(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \Lambda_n^{\ell_n^{[2]}} C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}})$$

$$= \gamma_n^N(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_{n-1}^{[2]} \in (N)^2} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}}).$$

Since $\Lambda_{n-1}^{\ell_{n-1}^{[2]}}$ is $\mathcal{G}_{n-1}^N$-measurable, it is sufficient to show that for each $\ell_{n-1}^{[2]} \in (N)^2$, we have

$$(24)$$
$$E\left[ m(X_{n-1})(G_{n-1,N})^2 \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N \right]$$

$$= \frac{1}{N^2} C_{b_{n-1}} Q_{n,N}^{\otimes 2} C_{b_n}(f \otimes g)(X_{n-1}^{\ell_{n-1}^{[2]}}).$$

Before starting our reasoning, for the sake of simplification, we remark that

$$\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}})$$

can be seen as a bounded measurable function of $(A_{n-1}, X_n)$, rather than a measurable function of $(X_n^{\ell_n^{[2]}}, A_{n-1}^{\ell_n^{[2]}})$. With this in mind, for any test function

$$F \in \mathcal{B}_b(\underbrace{[N]^N \times \cdots \times [N]^N}_{(n-1) \text{ times}} \times E_0^N \times \cdots \times E_{n-1}^N),$$

we have, by definition of $Q_p^{(2)}(x_{p-1}, d(a_{p-1}, x_p))$,

$$\mathrm{E}\left[m(X_{n-1})(G_{n-1,N})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}})F(A_{0:n-2}, X_{0:n-1})\right]$$

(25)
$$= \int Q_n^{(2)}(x_{n-1}, d(a_{n-1}, x_n))\lambda_{n-1}(a_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(x_n^{\ell_n^{[2]}})$$

$$F(a_{0:n-2}, x_{0:n-1})\mu_{n-1}(da_{0:n-2}, dx_{0:n-1}),$$

where $\mu_{n-1}$ denotes the measure corresponding to the underlying joint distribution of the IPS from step 0 to step $n-1$ with genealogy. Taking into account that

$$\mathrm{M}_n^{\ell_n^{[2]}}\left((a_{n-1}, x_n), d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]})\right) = \delta_{a_{n-1}^{\ell_n^{[2]}}}(d\tilde{a}_{n-1}^{[2]})\,\delta_{x_n^{\ell_n^{[2]}}}(d\tilde{x}_n^{[2]})$$

is a Markov kernel, we can introduce it in the right-hand side of (25) to obtain

$$\mathrm{E}\left[m(X_{n-1})(G_{n-1,N})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(X_n^{\ell_n^{[2]}})F(a_{0:n-2}, x_{0:n-1})\right]$$

$$= \int Q_n^{(2)}(x_{n-1}, d(a_{n-1}, x_n))\mathrm{M}_n^{\ell_n^{[2]}}\left((a_{n-1}, x_n), d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]})\right)$$

$$\lambda_{n-1}(a_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(x_n^{\ell_n^{[2]}})F(a_{0:n-2}, x_{0:n-1})\mu_{n-1}(da_{0:n-2}, dx_{0:n-1}).$$

The construction of many-body Feynman-Kac models allows replacing $(a_{n-1}^{\ell_n^{[2]}}, x_n^{\ell_n^{[2]}})$ with $(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]})$ in the observation functions, as they are equal by definition. Hence, one has the following equality:

$$\int Q_n^{(2)}(x_{n-1}, d(a_{n-1}, x_n))\mathrm{M}_n^{\ell_n^{[2]}}\left((a_{n-1}, x_n), d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]})\right)$$

$$\lambda_{n-1}(a_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(x_n^{\ell_n^{[2]}})F(a_{0:n-2}, x_{0:n-1})\mu_{n-1}(da_{0:n-2}, dx_{0:n-1})$$

$$= \int Q_n^{(2)}(x_{n-1}, d(a_{n-1}, x_n))\mathrm{M}_n^{\ell_n^{[2]}}\left((a_{n-1}, x_n), d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]})\right)$$

$$\lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(\tilde{x}_n^{[2]})F(a_{0:n-2}, x_{0:n-1})\mu_{n-1}(da_{0:n-2}, dx_{0:n-1}).$$

Now, the duality formula given in Lemma A.1 yields

$$\int Q_n^{(2)}(x_{n-1}, d(a_{n-1}, x_n))\mathrm{M}_n^{\ell_n^{[2]}}\left((a_{n-1}, x_n), d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]})\right)$$

$$\lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(\tilde{x}_n^{[2]})F(a_{0:n-2}, x_{0:n-1})\mu_{n-1}(da_{0:n-2}, dx_{0:n-1})$$

$$= \int \mathbb{Q}_n^{(2)}(x_{n-1}, d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]}))\,\mathbb{M}_n^{\tilde{a}_{n-1}^{[2]}, \ell_n^{[2]}, \tilde{x}_n^{[2]}}(x_{n-1}, d(a_{n-1}, x_n))$$

$$\lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f \otimes g)(\tilde{x}_n^{[2]})F(a_{0:n-2}, x_{0:n-1})\mu_{n-1}(da_{0:n-2}, dx_{0:n-1}).$$

In addition, since

$$\mathbb{M}_n^{\tilde{a}_{n-1}^{[2]}, \ell_n^{[2]}, \tilde{x}_n^{[2]}}(x_{n-1}, d(a_{n-1}, x_n))$$

is a Markov kernel for any choice of $(\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_n, \tilde{x}^{[2]}_n)$, we deduce that

$$\int \mathbb{Q}^{(2)}_n(\boldsymbol{x_{n-1}}, d(\tilde{a}^{[2]}_{n-1}, \tilde{x}^{[2]}_n)) \, \mathbb{M}^{\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_n, \tilde{x}^{[2]}_n}_n(\boldsymbol{x_{n-1}}, d(\boldsymbol{a_{n-1}}, \boldsymbol{x_n}))$$

$$\lambda_{n-1}(\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(f \otimes g)(\tilde{x}^{[2]}_n) F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}})$$

$$= \int \mathbb{Q}^{(2)}_n(\boldsymbol{x_{n-1}}, d(\tilde{a}^{[2]}_{n-1}, \tilde{x}^{[2]}_n)) \, \lambda_{n-1}(\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(f \otimes g)(\tilde{x}^{[2]}_n)$$

$$F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}).$$

Next, let us recall that

$$\mathbb{Q}^{(2)}_n(\boldsymbol{x_{n-1}}, d(\tilde{a}^{[2]}_{n-1}, \tilde{x}^{[2]}_n)) := m([N])^{\otimes 2}(d\tilde{a}^{[2]}_{n-1}) \, Q^{\otimes 2}_{n,N}(x^{\tilde{a}^{[2]}_{n-1}}_{n-1}, d\tilde{x}^{[2]}_n)$$

and

$$\lambda_{n-1}(\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_{n-1}) := \mathbb{1}_{\{b_{n-1}=1, \tilde{a}^1_{n-1} = \tilde{a}^2_{n-1} = \ell^1_{n-1} \neq \ell^2_{n-1}\}} + \mathbb{1}_{\{b_{n-1}=0, \tilde{a}^1_{n-1} = \ell^1_{n-1} \neq \tilde{a}^2_{n-1} = \ell^2_{n-1}\}},$$

whence we get the equality concerning the operator $C_{b_{n-1}}$. More precisely, if $b_{n-1} = 0$, we have

$$\int \mathbb{Q}^{(2)}_n(\boldsymbol{x_{n-1}}, d(\tilde{a}^{[2]}_{n-1}, \tilde{x}^{[2]}_n)) \, \lambda_{n-1}(\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(f \otimes g)(\tilde{x}^{[2]}_n)$$

(26)
$$F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}})$$

$$= \int \frac{1}{N^2} Q^{\otimes 2}_{n,N} C_{b_n}(f \otimes g)(x^{\ell^{[2]}_{n-1}}_{n-1}) F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}).$$

Otherwise, if $b_{n-1} = 1$, we get, with the convention $x^{\ell^{(1,1)}_{n-1}}_{n-1} = (x^{\ell^1_{n-1}}_{n-1}, x^{\ell^1_{n-1}}_{n-1})$,

$$\int \mathbb{Q}^{(2)}_n(\boldsymbol{x_{n-1}}, d(\tilde{a}^{[2]}_{n-1}, \tilde{x}^{[2]}_n)) \, \lambda_{n-1}(\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(f \otimes g)(\tilde{x}^{[2]}_n)$$

(27)
$$F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}})$$

$$= \int \frac{1}{N^2} Q^{\otimes 2}_{n,N} C_{b_n}(f \otimes g)(x^{\ell^{(1,1)}_{n-1}}_{n-1}) F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}).$$

Combining (26) and (27), we safely deduce that

$$\int \mathbb{Q}^{(2)}_n(\boldsymbol{x_{n-1}}, d(\tilde{a}^{[2]}_{n-1}, \tilde{x}^{[2]}_n)) \, \lambda_{n-1}(\tilde{a}^{[2]}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(f \otimes g)(\tilde{x}^{[2]}_n)$$

$$F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}})$$

$$= \int \frac{1}{N^2} C_{b_{n-1}} Q^{\otimes 2}_{n,N} C_{b_n}(f \otimes g)(x^{\ell^{[2]}_{n-1}}_{n-1}) F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}}) \mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}})$$

$$= \mathrm{E}\left[\frac{1}{N^2} C_{b_{n-1}} Q^{\otimes 2}_{n,N} C_{b_n}(f \otimes g)(X^{\ell^{[2]}_{n-1}}_{n-1}) F(\mathrm{A}_{0:n-2}, \mathrm{X}_{0:n-1})\right].$$

In conclusion, we have established that

$$\mathrm{E}\left[m(\mathrm{X_{n-1}})(G_{n-1,N})^2 \lambda^b_{n-1}(\mathrm{A}^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1}) C_{b_n}(f \otimes g)(\mathrm{X}^{\ell^{[2]}_n}_n) F(\mathrm{A}_{0:n-2}, \mathrm{X}_{0:n-1})\right]$$

$$= \mathrm{E}\left[\frac{1}{N^2} C_{b_{n-1}} Q^{\otimes 2}_{n,N} C_{b_n}(f \otimes g)(\mathrm{X}^{\ell^{[2]}_{n-1}}_{n-1}) F(\mathrm{A}_{0:n-2}, \mathrm{X}_{0:n-1})\right],$$

which terminates the verification of (24) and the proof of Proposition A.1.

$\square$

**A.2. Some intuition.** In general, the coupled particle block does not necessarily have the parents-children relations. Let us see a representation of the duality formula given in Lemma A.1 recursively applied in a mini IPS from level 0 to level 5 to some randomly chosen indices $\ell_{0:5}^{[2]}$ (see Figure 5).
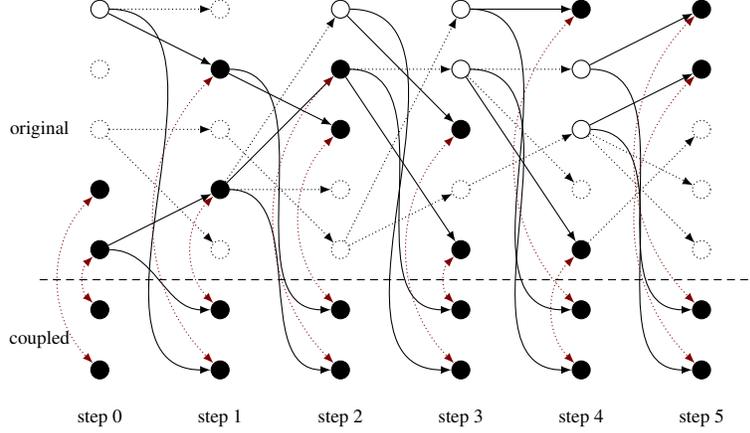


FIG 5. *An illustration of the duality formula recursively applied to a mini IPS of $n+1 = 6$ levels with 5 particles at each level. Every straight black or dotted arrow within the original IPS represents a Markov transition $M_{p,N}$ and the black twisted ones pointing to the particles in the coupled particle block represent the Feynman-Kac transition kernels $Q_{p,N}$. The red dotted bending arrows are identities. The indices of the original particles in the coupled particle block are $\ell_0^{[2]} = (4,5)$, $\ell_1^{[2]} = (2,4)$, $\ell_2^{[2]} = (2,3)$, $\ell_3^{[2]} = (3,5)$, $\ell_4^{[2]} = (1,5)$ and $\ell_5^{[2]} = (1,2)$.*

However, we can get any ancestral relations or coalescent tree-based form by manipulating the genealogical information encoded in the coupled genealogy. This is the essential idea we used by introducing many-body Feynman-Kac models. To make it clearer, we consider an event defined by

$$(28) \qquad \left\{ \ell_{p-2}^{[2]} = \tilde{A}_{p-2}^{[2]}, \ell_{p-1}^1 = \tilde{A}_{p-1}^1 = \tilde{A}_{p-1}^2 \neq \ell_{p-1}^2, \ell_p^{[2]} = \tilde{A}_p^{[2]} \right\}.$$

On this event, we are able to track the coalescent tree-based form as in Figure 6. The coupled
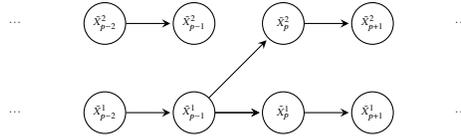


FIG 6. *The coupled particle block tracked by the event defined by (28).*

particle block and its genealogy are defined as the copies of certain particles and parents indices in the associated original IPS. On one hand, we select certain events such that the desired structure is trapped in the coupled particle block. On the other hand, we define the estimator based on the information reflected in the original IPS as no additional randomness are added by introducing the coupled particle block. Since their distributions are connected by the duality formula, we can use the information coded in the original IPS to estimate the measures corresponding to these coalescent tree-based particle blocks (see Figure 7).
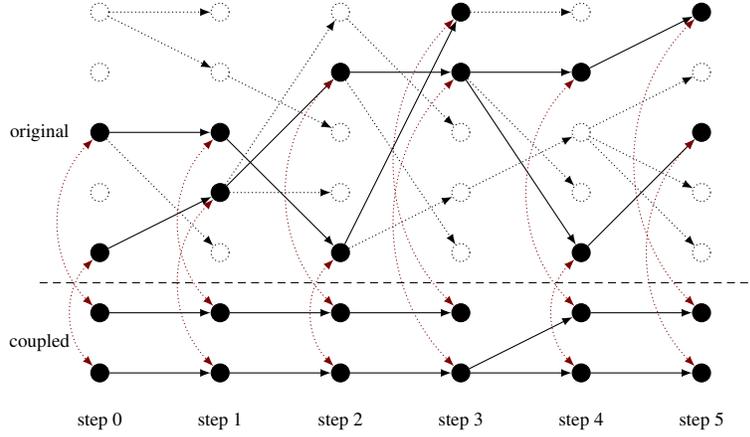
FIG 7. *An illustration of the duality formula recursively applied to a mini IPS of $n + 1 = 6$ levels with 5 particles at each level. Every straight black or dotted arrow within the original IPS represents a Markov transition $M_{p,N}$ and the black ones within the coupled particle block represent the Feynman-Kac transition kernels $Q_{p,N}$. The red dotted bending arrows are identities. The indices of the original particles in the coupled particle block are $t_0^{[2]} = (3,5)$, $t_1^{[2]} = (3,4)$, $t_2^{[2]} = (2,5)$, $t_3^{[2]} = (1,2)$, $t_4^{[2]} = (2,5)$ and $t_5^{[2]} = (1,3)$.*

The duality formula provides a way to touch the adaptive versions of the coalescent tree-based measures $\Gamma_n^b$, i.e., all the Feynman-Kac transition kernels $Q_p$ in the definition are replaced with the adaptive version $Q_{p,N}$. This is the idea underlying the construction of the estimators $\Gamma_{n,N}^b$.

**A.3. Connection with SMC.** To conclude, let us say a few words about the behavior of $\Gamma_{n,N}^b$. One remark is that, in general, this estimator is not unbiased in the ASMC framework. This is a consequence of the adaptive parametrization, as witnessed by Lemma 4.2. On the opposite, in a nonadaptive case (SMC), the estimation is unbiased, exactly as $\gamma_n^N$ is an unbiased estimation of $\gamma_n$ (see for example [11] Section 3.5.1). It turns out that the classical SMC framework corresponds to the case where the function $h_n$ in $\mathcal{A}2$ is equal to zero, meaning that $Q_{n,N} = Q_n$ for all $n$. Thus, Lemma 4.1 and (15) give the following proposition.

PROPOSITION A.2. Assume $\mathcal{A}1$-$\mathcal{A}2$ and suppose that $h_n \equiv 0$ for all $n \geq 0$. Then, for all test functions $f, g \in \mathcal{B}_b(E_n)$,

$$\mathrm{E}\left[\Gamma_{n,N}^b(f \otimes g)\right] = \Gamma_n^b(f \otimes g).$$

In particular, we also have

$$\mathrm{E}\left[\gamma_n^N(1)^2 V_n^N(f)\right] = \mathrm{Var}\left[\gamma_n^N(f)\right].$$

In fact, the essential technical results in Section 4.4 and Section 4.5 only require $\mathcal{A}1$. In other words, $\mathcal{A}2$ can be studied separately in order to adapt to applications not covered in this article.

Another remark is about the difference between $\Gamma_{n,N}^b$ and $\mu_b$ as defined in Section 3.2. of [16] in the nonadaptive context. However, since it is not straightforward to compare these estimators that are extremely notation-heavy, we would just like to briefly and heuristically mention that the main difference comes from the step where there is a coalescence, namely

$b_p = 1$. If we consider Figure 3 in Section 3.1, our estimator is not the most "precise" that one could propose. Let us look at the case where

$$\ell_{0:6}^{[2]} = ((5,3),(4,3),(2,5),(2,4),(2,5),(1,3),(2,4)).$$

For the terminal point $X_3^4$, the conditional distribution of $A_2^4$ is simply the categorical distribution since $X_3^4$ is a terminal point. Roughly speaking, once all the genealogy of the terminal points is calculated, one can deduce $\mu_b$. Hence, the take-home message is simple: if one is interested in estimating $\Gamma_n^b$ numerically, then the estimator $\mu_b$ proposed in [16] is expected to be more accurate, meaning that the variance should be smaller in general.

Nevertheless, as a theoretical tool, our estimator is easier to deal with in the adaptive framework. Indeed, induction is highly involved in our proof of consistency, so estimators that are stepwise easy to manipulate are required. Another difference is that we do not use instrumental random variables such as $K^1$ and $K^2$ in the definition of $\mu_b$. This also simplifies the analysis in an adaptive context where there is already more randomness than in a nonadaptive context.

## APPENDIX B: NUMERICAL EXPERIMENT

We provide in this section a numerical experiment based on the same toy example as the one presented in Section 4.1 of [4]. In particular, this ensures that assumptions $\mathcal{A}1$-$\mathcal{A}2$ are satisfied. Namely, consider a sequence of centered Gaussian target distributions $(\eta_n; 0 \le n \le 50)$ on $\mathrm{R}^{10}$ given by

$$\eta_n(x) \propto \exp\left(-\frac{1}{2}\left\langle x, \Sigma_n^{-1} x\right\rangle\right).$$

Denote by Id the identity matrix on $\mathrm{R}^{10}$ and J the lower triangular matrix such that $\mathrm{J}_{ij} = 1$ for $1 \le j \le i - 1 \le 9$. The covariance matrices are defined by

$$\Sigma_n = \mathrm{L}_n \mathrm{L}_n^{\mathrm{T}}, \text{ with } \mathrm{L}_n = \left(10\left(1 - \frac{n}{99}\right) + \frac{1}{10}\frac{n}{99}\right)\mathrm{Id} + \frac{1}{2}\frac{n}{99}\mathrm{J}.$$

Thus, the initial distribution $\eta_0$ consists in 10 centered and independent Gaussian components with variance 10. As $n$ grows, the covariance structure becomes more complicated. We consider an implementation of SMC with (nonadaptive) potential functions

$$G_n(x) := \exp\left(-\frac{1}{2}\left\langle x, \left(\Sigma_{n+1}^{-1} - \Sigma_n^{-1}\right) x\right\rangle\right),$$

and some random walk Metropolis kernels $M_n$ such that, at each step, $M_n$ is reversible with respect to $\eta_n$. In this scenario, a popular choice for $M_n$ is based on the Gaussian proposal with covariance matrix $\Sigma_n$. This is the "limiting" (nonadaptive) scenario that we will consider in the sequel. When one does not know the covariance matrices $\Sigma_n$, a natural choice is to use the estimated covariance matrix $\Sigma_n^N$. Our goal is to compare the respective behaviors of adaptive SMC and nonadaptive SMC. In particular, we want to show that the Lee and Whiteley variance estimator, in an adaptive context, goes to the asymptotic variance of the "limiting" (nonadaptive) SMC when $N$ grows.

For this, we consider the test function $f : \mathrm{R}^{10} \ni (x^{(1)}, x^{(2)}, \ldots, x^{(10)}) \mapsto x^{(1)} \in \mathrm{R}$. Keeping the notation of the previous sections, we illustrate the asymptotic variance estimators $NV_n^N\left(f - \eta_n^N(f)\right)$, which estimate the asymptotic variances of $\eta_n^N(f)$ respectively for the adaptive and nonadaptive SMC algorithms, see Figure 8. On the latter, the so-called reference value is the estimation of the theoretical asymptotic variance $\sigma_{\eta_n}^2(f - \eta_n(f))$. This value is estimated via Crude Monte Carlo through $2 \times 10^3$ independent runs of nonadaptive SMC
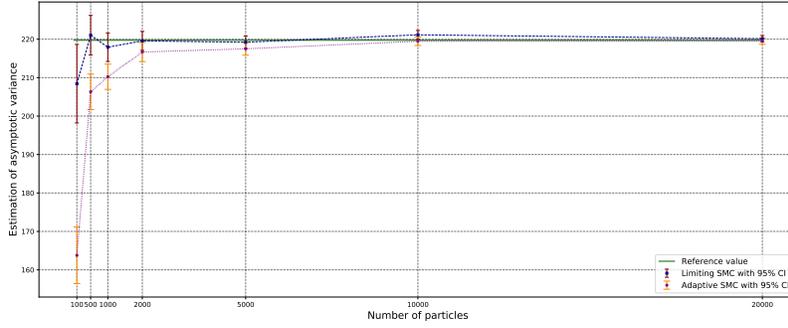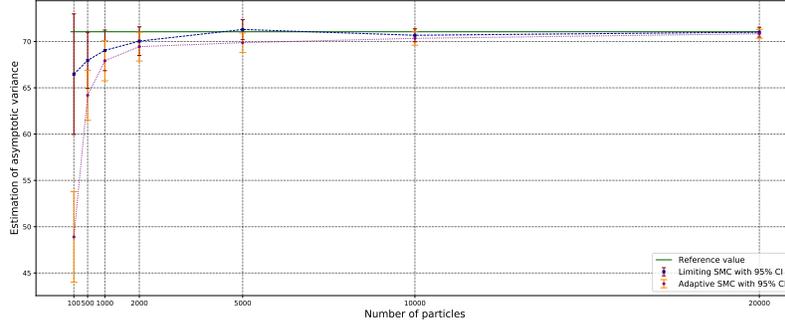
(a) Variance estimation of $\eta_n^N(f)$ with $n = 10$.



(b) Variance estimation of $\eta_n^N(f)$ with $n = 50$.

FIG 8. *Variance estimators for adaptive and nonadaptive "limiting" SMC for $n = 10$ and $n = 50$ with $N$ varying from $100$ to $2 \times 10^4$. We trace the variance estimators and its $95\%$ confidential intervals based on $500$ independent runs of each algorithm. The reference values $\sigma_{\eta_n}^2(f - \eta_n(f))$ are obtained through $2 \times 10^3$ independent runs of nonadaptive "limiting" SMC with $N = 5 \times 10^3$.*

with $N = 5 \times 10^3$ (notice that, stricto sensu, it does of course not depend on N). At each iteration of the algorithm, the random walk Metropolis kernel is applied 4 times in order to ensure a certain level of acceptance.

It is clear that when $N$ is relatively small, the variance estimations are biased. However, as the number $N$ of particles grows, we see that the behaviors of the adaptive and nonadaptive algorithms are similar in terms of asymptotic variance estimations. More precisely, the overlaps of the $95\%$ confidential intervals indicate that the adaptive SMC algorithm is indeed very "close" to its nonadaptive "limiting" counterpart. As expected, both variance estimators converge to the reference value $\sigma_{\eta_n}^2(f - \eta_n(f))$ as $N$ grows.

## APPENDIX C: TRUNCATED VARIANCE ESTIMATORS

As mentioned in [16], their variance estimators degenerate when $n$ is very large compared to the number $N$ of particles. Typically, no disjoint ancestral lines exist in such a particle system. In this case, we recommend to use the same kind of fixed-lag variance estimators as the ones proposed in [18]. More precisely, we only use part of the genealogy of the particle system (e.g., truncated at time $n - H$ for a relatively small lag $H \in \mathbb{N}^*$) to construct the variance estimators. Hereafter, we provide a heuristic in order to justify the relevance of such estimators in practice.

In general, the application of SMC sampling on a large time scale requires some "forgetting" (mixing) properties of the underlying model. For example, in the toy example presented in Appendix B, if the Metropolis kernel is implemented a large amount of times at each iteration, the particles will somehow "forget" the dependence caused by the genealogy. Basically, in terms of coalescent tree-based measures, $\bar{\Gamma}_n^{(p)}$ would be very "close" to the disjoint ancestral lines based measure $\bar{\Gamma}_n^{(\varnothing)}$. More concretely, if we look at the asymptotic variance $\sigma_{\eta_n}^2$, we have

$$\sigma_{\eta_n}^2(f) = \underbrace{\sum_{p=0}^{n-H-1} \left( \bar{\Gamma}_n^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_n^{(\varnothing)}(f^{\otimes 2}) \right)}_{\text{small by the "forgetting" properties of the model}} + \sum_{p=n-H}^{n} \left( \bar{\Gamma}_n^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_n^{(\varnothing)}(f^{\otimes 2}) \right).$$

Accordingly, a natural idea is to estimate only the second part of the right hand side in order to approximate the asymptotic variance. A truncated term by term estimator can therefore be defined as

$$\sum_{p=n-H}^{n} \left( \bar{\Gamma}_{n,N}^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}) \right).$$

Unfortunately, this estimator is not always numerically stable, as it requires that disjoint ancestral lines exist in the particle system from time 0 to time $n - H$. Following the same mechanism as in the proofs of Proposition 3.1 and Proposition 4.2, we can show that

$$N \left( \eta_n^N(f)^2 - \bar{\Gamma}_{n,N}^{(\varnothing,H)}(f^{\otimes 2}) \right) \approx \sum_{p=n-H}^{n} \left( \bar{\Gamma}_{n,N}^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_{n,N}^{(\varnothing)}(f^{\otimes 2}) \right),$$

where, if $E_n^i(H)$ denotes the index of the ancestor of $X_n^i$ at step $n - H$,

$$\bar{\Gamma}_{n,N}^{(\varnothing,H)}(f^{\otimes 2}) := \frac{N^{H-1}}{(N-1)^{H+1}} \sum_{E_n^i(H) \neq E_n^j(H)} f(X_n^i)f(X_n^j).$$

The estimator $N(\eta_n^N(f)^2 - \bar{\Gamma}_{n,N}^{(\varnothing,H)}(f^{\otimes 2}))$ is more or less the one proposed in [18] and is indeed a truncated version of $NV_n^N(f)$ proposed in [16]. We refer the interested reader to [18] for theoretical results as well as numerical illustrations. When $H$ is properly chosen, the fixed-lag variance estimator is expected to be able to balance the memory and the degeneracy of the genealogy of the particle system. However, finding a suitable $H$ in a specific application is highly nontrivial. As explained in [18], it is then natural to consider adaptive mechanisms to determine $H$. Nevertheless, to the best of our knowledge, this is still an open problem, which is beyond the scope of the present paper.

## REFERENCES

[1] ANDRIEU, C., DOUCET, A. and HOLENSTEIN, R. (2010). Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 269–342. MR2758115

[2] AU, S. K. and BECK, J. L. (2001). Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics* **16** 263-277.

[3] AU, S. K. and BECK, J. L. (2003). Subset Simulation and its Application to Seismic Risk Based on Dynamic Analysis. *Journal of Engineering Mechanics* **129** 901-917.

[4] BESKOS, A., JASRA, A., KANTAS, N. and THIERY, A. (2016). On the convergence of adaptive sequential Monte Carlo methods. *Ann. Appl. Probab.* **26** 1111–1146. MR3476634

[5] CÉROU, F., DEL MORAL, P., FURON, T. and GUYADER, A. (2012). Sequential Monte Carlo for rare event estimation. *Stat. Comput.* **22** 795–808. MR2909622

[6] CÉROU, F., DEL MORAL, P. and GUYADER, A. (2011). A nonasymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.* **47** 629–649. MR2841068

[7] CÉROU, F. and GUYADER, A. (2007). Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.* **25** 417–443. MR2303095 (2008k:60049)

[8] CÉROU, F. and GUYADER, A. (2016). Fluctuation analysis of adaptive multilevel splitting. *Ann. Appl. Probab.* **26** 3319–3380. MR3582805

[9] CHAN, H. P. and LAI, T. L. (2013). A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.* **41** 2877–2904. MR3161451

[10] CHOPIN, N. (2004). Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.* **32** 2385–2411. MR2153989

[11] DEL MORAL, P. (2004). *Feynman-Kac formulae: Genealogical and interacting particle systems with applications. Probability and its Applications*. Springer-Verlag, New York. MR2044973

[12] DEL MORAL, P. (2013). *Mean field simulation for Monte Carlo integration. Monographs on Statistics and Applied Probability* **126**. CRC Press, Boca Raton, FL. MR3060209

[13] DEL MORAL, P., KOHN, R. and PATRAS, F. (2016). On particle Gibbs samplers. *Ann. Inst. Henri Poincaré Probab. Stat.* **52** 1687–1733. MR3573292

[14] DOUC, R. and MOULINES, E. (2008). Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.* **36** 2344–2376. MR2458190

[15] DOUCET, A., DE FREITAS, N. and GORDON, N., eds. (2001). *Sequential Monte Carlo methods in practice. Statistics for Engineering and Information Science*. Springer-Verlag, New York. MR1847783

[16] LEE, A. and WHITELEY, N. (2018). Variance estimation in the particle filter. *Biometrika* **105** 609–625. MR3842888

[17] MCLEISH, D. L. (1974). Dependent central limit theorems and invariance principles. *Ann. Probability* **2** 620–628. MR0358933

[18] OLSSON, J. and DOUC, R. (2019). Numerically stable online estimation of variance in particle filters. *Bernoulli* **25** 1504–1535.