HAL open science

Sequential Monte Carlo and Applications in Molecular Dynamics

Qiming Du

▶ To cite this version:

Qiming Du. Sequential Monte Carlo and Applications in Molecular Dynamics. Probability [math.PR]. Sorbonne Université, 2020. English. NNT : 2020SORUS040 . tel-02969115v2

HAL Id: tel-02969115

https://theses.hal.science/tel-02969115v2

Submitted on 16 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Sequential Monte Carlo and Applications in Molecular Dynamics

Qiming Du

Laboratoire de Probabilités, Statistique et Modélisation - UMR 8001
Sorbonne Université

Thèse pour l'obtention du grade de:

Docteur de l'université Sorbonne Université

Dirigée par : Arnaud Guyader et Tony Lelièvre

Rapportée par : Josselin Garnier (École Polytechnique)
Ajay Jasra (King Abdullah University of Science and Technology)

Présentée devant un jury composé de :
Gérard Biau (Sorbonne Université)
Pierre Del Moral (INRIA Bordeaux)
Gersende Fort (IMT Toulouse)
Josselin Garnier (École Polytechnique)
Arnaud Guyader (Sorbonne Université)
Tony Lelièvre (École des Ponts ParisTech)
Mathias Rousset (INRIA Rennes)

École Doctorale de Sciences Mathématiques de
Paris-Centre
Section Mathématiques appliquées

17 juin 2020

To my parents
To Mengying
To Siyan

青玉案·元夕

东风夜放花千树，更吹落，星如雨。
宝马雕车香满路。凤箫声动，玉壶光转，一夜鱼龙舞。
蛾儿雪柳黄金缕，笑语盈盈暗香去。
众里寻他千百度，蓦然回首，那人却在，灯火阑珊处。

——辛弃疾

Acknowledgements

First and foremost I would like to thank my supervisor Arnaud Guyader. Throughout the past three years, his helpful advices and encouragements are invaluable to me. His strict supervision will guide me in the rest of my life, encouraging me to always pursue a higher standard. I thank him for letting me know the beautiful world of Interacting Particle System, and I feel lucky to have had the chance to dive into it for the past three years. Apart from research, I am also very grateful for his help on my daily life. At the same time, I want to thank my supervisor Tony Lelièvre. His kindness and encouragement have always been so inspiring, especially during the most difficult time of my Ph.D. career. As a former statistics student, I thank him for taking so much time to explain the ideas in Molecular Dynamics to me, and in a extremely clear and insightful way. In particular, I want to thank him for letting me be brave enough to study the asymmetric resampling scheme, which is one of the main contributions of this thesis. I do not know if I could have another chance to take such a long journey in mathematics in the future, and this is definitely a precious experience of my life. For both of my supervisors, I appreciate that they have helped me in so many ways, and I learned so much from them.

I want to thank Gérard and Raphaël for helping me to finish this thesis, especially during the difficult time of COVID-19. I feel so lucky to be able to start a new journey in my career with both of them, and I am sure that there will be a lot of fruitful collaborations in the future.

I consider myself very fortunate to have so many lovely colleagues and friends in Laboratoirede Probabilités, Statistique et Modélisation (LPSM). First, I want to thank Aude, Cyril, Emilie, Eric, Gloria, Simon and Thibault for having shared the same office with me. I thank Tabea and Vincent for letting me be the teaching assistant in the courses "Statistique (4M015)" and "Probabilités Numériques et Statistiques Computationnelles (4M074)" for the past three years, which was an extremely pleasant experience. In addition, I want to thank Adeline, Anna, Anna, Charlotte, Claire, Lucie, Maud, Marie, Maxime, Nazih, Nicolas, Pierre, Sébastien, Taieb and Yohann. The names will go on and on. Every coffee break with you was relaxing and inspiring. I learned a lot from the discussions with them. In particular, I want to mention that I have also learned a lot of "extremely difficult French vocabularies" from the crossword games in the resting room on the second floor of 15-25. I thank the excellent administration team of LPSM, especially I thank Corinne and Louise for their generous help and support. Special thanks to my Chinese friends Chenguang, Shuo, Wenqian, Yating and Zhuchao, for all the lunch discussions and encouragements during the past years.

From the deepest of my heart, I want to thank my family in China, for everything I can imagine. Being the single child of my parents, I owe them a lot for not being able to stay by their side in the past years. Without their support and sacrifice, I can not be free to pursue my dream on the other side of the world.

Most importantly, I want to thank my wife Mengying for her love and patience. Without her, this thesis would not have been possible. Finally, I want to thank the sweetest rare event that I have failed to estimate, Siyan, my daughter. Without her smiles, I could never achieve so many, including this thesis.

Abstract

This thesis consists in three parts, all connecting to the Sequential Monte Carlo (SMC) framework. The primary motivation is to understand the generalized Adaptive Multilevel Splitting (gAMS, [BGG+16]), an algorithm that aims at estimating the rare-event transition probability between metastable states in the context of Molecular Dynamics. In the first part, we deal with the Adaptive SMC framework, introduced by [BJKT16]. We prove that the variance estimator proposed by Lee and Whiteley [LW18] in the non-adaptive setting, is still consistent in this adaptive setting, under a slightly weaker assumption than in [BJKT16]. In the theoretical perspective, We propose a new strategy that deals with the adaptiveness and genealogy of the particle system separately, based on coalescent tree-based expansions [CDMG11]. In the second part, we propose Asymmetric SMC framework, a generalization of the classical SMC framework. The motivation is to reduce the computational burden brought by the mutation kernels. We provide Central Limite Theorem for the assoticated Feynman-Kac measures, along with consistent asymptotic variance estimators. We remark that in some specific setting, the gAMS algorithm enters into the Asymmetric SMC framework, which leads to a consistent variance estimator and asymptotic normality for the gAMS algorithm. However, this result does not cover the general setting of gAMS algorithm. Our analysis is based on generalized coalescent tree-based expansions, which may provide an universal strategy that can be used to derive consistent variance estimators in the general SMC context. In the third part, we propose some strategies that combine the gAMS algorithm and modern statistical/machine learning. We investigate the coupling of gAMS algorithm and a nonparametric regressor Mondrian Forests [LRT14], to improve the performance of gAMS algorithm. The proposed iterative updating strategy may be helpful in developing automated and efficient rare-event estimation strategy in a high-dimensional and low temperature setting.

Keywords: Sequential Monte Carlo, Interacting Particle System, Variance estimation, Genealogy, Central Limite Theorem, Rare-event simulation, Random Forests, Online learning.

Contents

1 Introduction **1**

1.1 General background on Sequential Monte Carlo . . . 1

1.1.1 Particle Filters and Hidden Markov Models . . . 1

1.1.2 Subset Simulation and rare-event simulation . . . 4

1.1.3 Feynman-Kac particle models . . . 9

1.2 Rare-event simulation in Molecular Dynamics . . . 14

1.2.1 Overdamped Langevin dynamics and metastability . . . 14

1.2.2 Reaction coordinate . . . 15

1.2.3 A brief introduction to gAMS . . . 16

1.2.4 Level-indexed process . . . 19

1.3 Summary of the Chapters . . . 20

1.3.1 Chapter 2: Variance estimation in Adaptive Sequential Monte Carlo . . . 20

1.3.2 Chapter 3: Asymmetric Sequential Monte Carlo . . . 22

1.3.3 Chapter 4: Estimating committor function with Mondrian Forests . . . 23

2 Variance Estimation in Adaptive Sequential Monte Carlo **25**

2.1 Introduction . . . 26

2.2 Adaptive Sequential Monte Carlo . . . 29

2.2.1 Framework . . . 29

2.2.2 ASMC algorithm . . . 30

2.2.3 Assumptions . . . 31

2.2.4 Central limit theorems . . . 32

2.3 Variance estimations . . . 33

2.3.1 Coalescent tree-based variance expansion . . . 33

2.3.2 Term by term estimator . . . 36

2.3.3 Disjoint ancestral lines estimator . . . 38

2.4 Proofs . . . 39

2.4.1 Almost sure convergence . . . 39

2.4.2 Proof of Theorem 2.2.1 . . . 41

2.4.3 Proof of Theorem 2.3.1 . . . 44

2.4.4 Technical results . . . 45

2.4.5 Connection between the estimators . . . 55

2.5 Many-body Feynman-Kac models . . . 56

2.5.1 Duality formula . . . 57

2.5.2 Some intuition . . . 62

2.5.3 Connection with SMC . . . 63

3 Asymmetric Sequential Monte Carlo **65**
3.1 Introduction . . . 66
3.1.1 Main results . . . 68
3.1.2 Notation . . . 68
3.2 SMC framework . . . 71
3.2.1 Setting . . . 71
3.2.2 Interacting particle system . . . 74
3.2.3 Asymptotic results . . . 74
3.2.4 Genealogy and survival history . . . 75
3.3 Variance estimations . . . 76
3.3.1 Asymptotic variance expansion . . . 76
3.3.2 Term by term asymptotic variance estimators . . . 77
3.3.3 Unbiased non-asymptotic variance estimator . . . 80
3.3.4 Connection between the estimators . . . 80
3.3.5 Efficient asymptotic variance estimators . . . 81
3.4 Discussions . . . 82
3.4.1 Unbiasedness condition . . . 82
3.4.2 Other comments . . . 82
3.5 Algorithms to compute variance estimators . . . 84
3.6 Coalescent tree-based expansions . . . 90
3.6.1 Original coalescent tree-based measures . . . 90
3.6.2 Coalescent Feynman-Kac kernels . . . 91
3.6.3 Binary decompositions . . . 94
3.6.4 Coalescent tree occupation measures . . . 96
3.6.5 Feynman-Kac measures flow in a random environment. . . . 99
3.6.6 Efficient estimator of $\widehat{\Gamma}_n^{\dagger,(p)}$ . . . 101
3.7 Proofs . . . 102
3.7.1 Martingales . . . 103
3.7.2 Verification of asymptotic variance expansion . . . 106
3.7.3 Proof of Theorem 3.2.1 . . . 108
3.7.4 Proof of Theorem 3.2.2 . . . 109
3.7.5 Proof of Proposition 3.3.1 . . . 112
3.7.6 Proof of Theorem 3.6.1 . . . 113
3.7.7 Proof of Proposition 3.6.5 . . . 114
3.7.8 Technical results . . . 115

4 Estimating Committor Function with Mondrian Forests **153**
4.1 Introduction . . . 154
4.2 Setting . . . 155
4.2.1 Overdampled Langevin dynamics . . . 155
4.2.2 Ground truth . . . 156
4.2.3 On the choice of regressors . . . 158
4.3 A brief introduction to Mondrian Forests . . . 158
4.3.1 The mechanism of a Mondrian tree . . . 159
4.3.2 Theoretical aspects of MF . . . 161
4.4 Numerical illustrations . . . 163
4.4.1 Learning with perfect data . . . 163

4.4.2 Learning with noisy data 163
4.4.3 Capability of recovering from a ruined model 165
4.4.4 Conclusion of the numerical tests 165
4.5 Iterative updating strategy with gAMS and MF 166
4.5.1 Crude interaction between gAMS and MF 166
4.5.2 Tempering 166
4.6 Discussions 168

Appendices **171**

A Generalized AMS under Asymmetric SMC framework **173**
A.1 Setting 173
A.2 Uncertainty control with a single simulation 174
A.3 VoV behaviors 174

List of Figures

1.1 An illustration of Hidden Markov Models. 2
1.2 The big picture of Subset Simulation. 5
1.3 The construction of a $\mathbf{P}_{A_n}$-invariant kernel. 6
1.4 Feynman-Kac measures flow for Asymmetric SMC. 13
1.5 Schematic picture of a 2-dimensional toy example with the level sets of V (dashed lines). 15
1.6 An example of reaction coordinate $\xi : (x, y) \mapsto x$. 16
1.7 An illustration of M_L that enters into gAMS framework. 17

2.1 A representation of the coalescent tree-based measure $\Gamma_n^{(p)}$. 34
2.2 An IPS with $n + 1 = 7$ levels and $N = 5$ particles at each level. 35
2.3 The first family of $\ell_{0:6}^{[2]}$ such that (2.7) is verified. 35
2.4 The second family of $\ell_{0:6}^{[2]}$ such that (2.7) is verified. 36
2.5 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 $\ell_5^{[2]} = (1, 2)$. 62
2.6 The coupled particle block tracked by the event defined by (2.28). 62
2.7 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 $\ell_0^{[2]} = (3, 5)$, $\ell_1^{[2]} = (3, 4)$, $\ell_2^{[2]} = (2, 5)$, $\ell_3^{[2]} = (1, 2)$, $\ell_4^{[2]} = (2, 5)$ and $\ell_5^{[2]} = (1, 3)$. 63

3.1 Illustration of the Feynman-Kac measures flow. 72
3.2 A representation of the original coalescent tree-based measure $\Gamma_n^{(p)}$. 91

4.1 Schematic picture of transition paths. 154
4.2 The illustration of iterative updating strategy. 155
4.3 Representation of the three-hole energy landscape. 157
4.4 The numerical solution of (4.2) with inverse temperature $\beta = 1.67$. 157

4.5 First split in the construction of MT. 159
4.6 Decision tree corresponding to the first split. 160
4.7 Second split in the construction of MT. 160
4.8 Second and third splits in the construction of MT. 160
4.9 The decision tree with three splits. 161
4.10 Learning with the ground truth, where the training data is uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$. 164
4.11 Learning with slightly perturbed data, where (X_i) are uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$. 164
4.12 Learning with largely perturbed data, where (X_i) are uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$. 165
4.13 The figure on the left represents the prediction made by a MF model trained by 100 largely perturbed data. The one on the right represents the model trained with 1000 more perfect data. (X_i) are uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$. 166
4.14 Learning with the training data generated by gAMS algorithm. The reaction coordinate is set to be the distance to the state A (i.e., ξ_1). The inverse temperature is $\beta = 1.67$. 167
4.15 Learning with the data re-generated by gAMS algorithm. The reaction coordinate is the trained MF model. The inverse temperature is $\beta = 1.67$. 167
4.16 Illustrations of committor functions for different temperatures. 168

A.1 An illustration for the estimation of AMS method along with 95% confidential interval. The number of particles N varies from 100 to 50000 and $\iota = 0.01$. . . . 174
A.2 An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 10$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators. . . 175
A.3 An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 50$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators. . . 175
A.4 An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 100$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators. . . 176
A.5 An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 200$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators. . . 176
A.6 An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 500$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators. . . 177

List of Tables

3.1 Time and space complexity under multinomial resampling scheme. 89
3.2 Time and space complexity under asymmetric resampling scheme. 89

List of Algorithms

1 Simulation of an IPS with genealogy and survival history. 84
2 Computation of $\gamma_n^N(f)$ and $\eta_n^N(f)$. 85
3 Generate backward genealogy tracing matrix: Θ. 85
4 Compute star inner product in $\mathbf{R}^3$: starProduct(X, Y). 85
5 Consistent asymptotic variance estimator for $\gamma_n^N(f)$. 86
6 Consistent asymptotic variance estimator for $\eta_n^N(f)$. 87
7 Unbiased non-asymptotic variance estimator for $\gamma_n^N(f)$. 88

Chapter 1

Introduction

1.1 General background on Sequential Monte Carlo

Generally speaking, Sequential Monte Carlo methods (SMC) are a set of simulation-based methods aiming at sampling a sequence of probability or finite measures, usually connected by some nonlinear operators in high dimensional state spaces. They are among the most widely used tools in computational statistics, chemistry, finance and many other disciplines. It seems that the term "Sequential Monte Carlo" was first coined by Liu and Chen in [LC98], where two typical problems are discussed: Bayesian missing-data problem and state-space models. In the latter case, SMC methods are also called Particles Filters, and this term was first proposed by Del Moral [DM96]. The most basic, and also the most famous Particle Filter was proposed a little bit earlier by Gordon, Salmond and Smith in [Gor93], which is often referred to as Bootstrap Filter. Beyond Filtering and Bayesian Inference, SMC methods also provide powerful tools in rare-event simulation problems. A typical family of algorithms are called Subset Simulation, or Multilevel Splitting methods. The original idea was brought in the 1950s by Kahn and Harris [KH51] and Rosenbluth and Rosenbluth [RR55] to study particle transmission energies and molecular polymer conformations. More recently, a more detailed study was conducted by Au and Beck [AB01, AB03] for estimation of small failure probabilities in high dimensional setting. A more refined version of this algorithm, called Adaptive Multilevel Splitting (AMS), was proposed by Cérou and Guyader in [CG07]. Generalized Adaptive Multilevel Splitting methods (gAMS) were introduced later in [BGG+16] by Bréhier, Gazeau, Goudenège, Lelièvre and Rousset, which are designed to solve some typical rare-event simulation problems in Molecular Dynamics. These are the models at the core of this thesis.

In this section, we provide some of the most important motivating examples of the SMC methods mentioned above, as well as the basic mathematical language used in this thesis. The reader is referred, for example, to [DdFG01] and [DM04] for a wider list of applications and theoretical analysis. In particular, a good survey for the recent development of AMS methods can be found in [CGR19b].

1.1.1 Particle Filters and Hidden Markov Models

The filtering problem consists in estimating the internal states in some unobserved dynamical systems when partial observations are made, and random perturbations are non-negligible in the observers as well as in the underlying dynamical system. The main objective is to compute and/or sample from the posterior distributions of the states of some Markov chain, given some

noisy and partial observations. In this section, we present a basic setting of Hidden Markov Models (HMM, cf. Figure 1.1) and the original algorithm of Bootstrap Filter to illustrate the basic idea of the implementation of Particle Filters.

$$\begin{tikzcd}
X_0 \arrow[r] \arrow[d] & X_1 \arrow[r] \arrow[d] & X_2 \arrow[d] \arrow[r, dotted, no head] & X_n \arrow[d] \arrow[r, dotted, no head] & \text{Signal} \\
Y_0 & Y_1 & Y_2 \arrow[r, dotted, no head] & Y_n \arrow[r, dotted, no head] & \text{Observation}
\end{tikzcd}$$

Figure 1.1: An illustration of Hidden Markov Models.

We assume that $(X_n; n \geq 0)$ is a time-homogeneous Markov chain taking values in $\mathbf{R}^{d_x}$, with initial distribution $p_0(x)dx$, where dx denotes the Lebesgue measure on $\mathbf{R}^{d_x}$ and $p_0(x)$ is a probability density function. At each time step $k \geq 1$, we suppose that

$$\mathcal{L}\left(X_k \mid X_{k-1} = x_{k-1}\right) = p\left(x_k \mid x_{k-1}\right) dx_k,$$

where $p\left(x_k \mid x_{k-1}\right)$ denotes the probability transition density function of the underlying Markov chain. The observations $(Y_n; n \geq 0)$ take values in $\mathbf{R}^{d_y}$ and d_y can be different from d_x. We suppose that $(Y_0, Y_1, \ldots, Y_n)$ are conditionally independent given $(X_0, X_1, \ldots, X_n)$. Moreover, we assume that

$$\mathcal{L}\left(Y_k \mid X_k = x_k\right) = p(y_k \mid x_k) dy_k.$$

A standard model that allows this structure can be reformulated as follows:

$$\begin{cases} X_n = g(X_{n-1}) + W_n \\ Y_n = h(X_n) + V_n. \end{cases}$$

We assume that g and h are some known functions and $(W_n; n \geq 0)$ and $(V_n; n \geq 0)$ are independent sequences with known probability density functions. We remark that when g and h are linear functions, and when W_n and V_n are Gaussian random variables, the problem can be analytically solved by the famous Kalman Filter, also known as linear quadratic estimation. However, when g and h are nonlinear, the analytical solutions are intractable in general. This is one of the main motivations of the Monte Carlo-based numerical methods such as Particle Filters. A typical goal of HMM is to estimate recursively in time the posterior density $p(x_n \mid y_0, y_1, \ldots, y_{n-1})$ given $Y_0 = y_0$, $Y_1 = y_1, \ldots, Y_{n-1} = y_{n-1}$. In other words, for any test function f_n, we are interested in estimating the integral

$$I_n(f_n) := \int f_n(x_n) p(x_n \mid y_0, y_1, \ldots, y_{n-1}) dx_n.$$

Thanks to Bayes formula, we have

$$p(x_0, x_1, \ldots, x_n \mid y_0, y_1 \ldots, y_{n-1}) = \frac{p(y_0, y_1, \ldots, y_{n-1} \mid x_0, x_1, \ldots, x_n) p(x_0, x_1, \ldots, x_n)}{p(y_0, y_1, \ldots, y_{n-1})}.$$

By the conditional independence in the HMM construction, it is easily checked that

$$p(y_0, y_1, \ldots, y_{n-1} \mid x_0, x_1, \ldots, x_n) = \prod_{k=0}^{n-1} p(y_k \mid x_k),$$

which also yields

$$p(y_0, y_1, \ldots, y_{n-1}) = \int p(x_0, x_1, \ldots, x_n) \prod_{k=0}^{n-1} p(y_k \mid x_k) dx_0 dx_1 \ldots dx_n.$$

Therefore, the integral of interest becomes

$$I_n(f_n) = \frac{\int f_n(x_n) \prod_{k=0}^{n-1} p(y_k \mid x_k) p(x_0) \prod_{k=0}^{n} p(x_k \mid x_{k-1}) dx_0 dx_1 \ldots dx_n}{\int \prod_{k=0}^{n-1} p(y_k \mid x_k) p(x_0) \prod_{k=0}^{n} p(x_k \mid x_{k-1}) dx_0 dx_1 \ldots dx_n}.$$

The idea of Particle Filters is to simulate N particles at each iteration of the algorithm. Then, by applying the importance sampling and bootstrap resampling steps, the quantity of interest $I_n(f_n)$ can be estimated recursively in time by exploiting the corresponding empirical measures. More concretely, the mechanism of the Bootstrap Filter is the following:

(i) Initialization:

Set $k = 0$;

For $i = 1, 2, \ldots, N$, sample $x_0^{(i)} \sim p_0(x_0)$.

(ii) Importance weights calculations:

For $i = 1, 2, \ldots, N$, calculate importance weights $\omega_k^{(i)} = p\left(y_k \,\middle|\, x_k^{(i)}\right)$.

(iii) Resampling step:

Resample with replacement N particles

$$\left(\widetilde{x}_k^{(1)}, \widetilde{x}_k^{(2)}, \ldots, \widetilde{x}_k^{(N)}\right) \text{ from the set } \left(x_k^{(1)}, x_k^{(2)}, \ldots, x_k^{(N)}\right)$$

according to the importance weights $\left(\omega_k^{(1)}, \omega_k^{(2)}, \ldots, \omega_k^{(N)}\right)$.

(iv) Importance sampling step:

Set $k = k + 1$;

For $i = 1, 2, \ldots, N$, sample $x_k^{(i)} \sim p\left(x_k \,\middle|\, \widetilde{x}_{k-1}^{(i)}\right)$;

Go to step (ii).

By replacing the posterior distribution with its empirical version, the integral of interest can therefore be estimated by

$$\hat{I}_n(f_n) := \frac{1}{N} \sum_{i=1}^{N} f_n(x_n^{(i)}).$$

As shown above, the implementation of Particle Filter is straightforward and can be applied to a very large class of state-space models. The price to pay for this simplicity is the relatively expensive computational cost. For this topic, the reader is referred to [DdFG01] for more detailed analysis and possible variance reduction techniques that can be applied to increase the performance in real-world applications.

1.1.2 Subset Simulation and rare-event simulation

A rare event is an event with non-zero, but extremely small probability. To give an idea, this probability can be lower than 10^{-20}. Generally speaking, when the probability of interest is extremely small, it is nearly impossible to have a single sample such that the event of interest occurs within a reasonable simulation time through a crude Monte Carlo approach, let alone to collect a handful of realizations to conduct any meaningful estimation. One typical example in real-world applications is the risk management in insurance, in which case it is important to estimate accurately the probability of some kind of catastrophe. Another example is when we are interested in simulating some certain transition paths between metastable states of a stochastic dynamical system. It can be a transition procedure that is not rare at the macroscopic scale. However, when the associated dynamic can only be simulated with a very small time step, the event of interest then becomes rare in the simulation timescale.

A classic solution for rare-event simulation problems is to apply Important Sampling (IS) techniques. If the underlying probability distribution is η, we draw samples according to another auxiliary distribution π, and weight each observation $X = x$ by the Radon-Nikodym derivative $w(x) = d\eta(x)/d\pi(x)$. When the auxiliary distribution is well chosen, Importance Sampling methods may dramatically decrease the variance of the estimated probability. As a consequence, sample sizes needed to conduct reliable estimation may also be dramatically reduced. The readers are referred to [RC10] for a discussion on Importance Sampling techniques in general, and to [Buc10] for the context of rare-event estimation. However, the efficiency of IS methods depends highly on the choice of the auxiliary distribution, and it may completely fail for bad choices even compared to naive Monte Carlo. A good choice of auxiliary distribution π requires strong knowledge and careful analysis on the underlying probability measure η and the rare event of interest, which is not always tractable.

In this thesis, we mainly discuss another possible approach, that is referred to as Subset Simulation, or Multilevel Splitting. Compared to IS methods, fewer information is needed to conduct reliable estimation and inference, at the cost of a heavier computational cost. Some comments on the comparison of these two approaches can be found in the survey [CGR19b], which will not be discussed in this thesis. In this section, we present a static rare-event simulation problem in order to have a big picture in mind. A more detailed setting in Molecular Dynamics is provided in the following sections. Let E be some Polish space, namely, a separable completely metrizable topological space. As is shown in Figure 1.2, we assume that there exists a decreasing sequence of events

$$E = A_{-1} \supset A_0 \supset A_1 \supset A_2 \supset \cdots \supset A_{n-1} \supset A_n = A_*,$$

where A_* denotes the rare event of interest. Let $\eta(dx)$ be a probability measure on E, and the goal is to estimate the probability $p_* := \eta(A_*)$. Let X be a random variable with distribution $\eta(dx)$. Thanks to Bayes formula, we have

$$\mathbf{P}\left(X \in A_*\right) = \prod_{p=0}^{n} \mathbf{P}\left(X \in A_p \,\middle|\, X \in A_{p-1}\right). \tag{1.1}$$

The basic idea of Subset Simulation is to estimate the probabilities

$$\mathbf{P}\left(X \in A_p \,\middle|\, X \in A_{p-1}\right)$$

term by term, and the final estimator is therefore constructed as the product of these estimators. The intuition behind this formulation is that each term of the conditional probability is not so

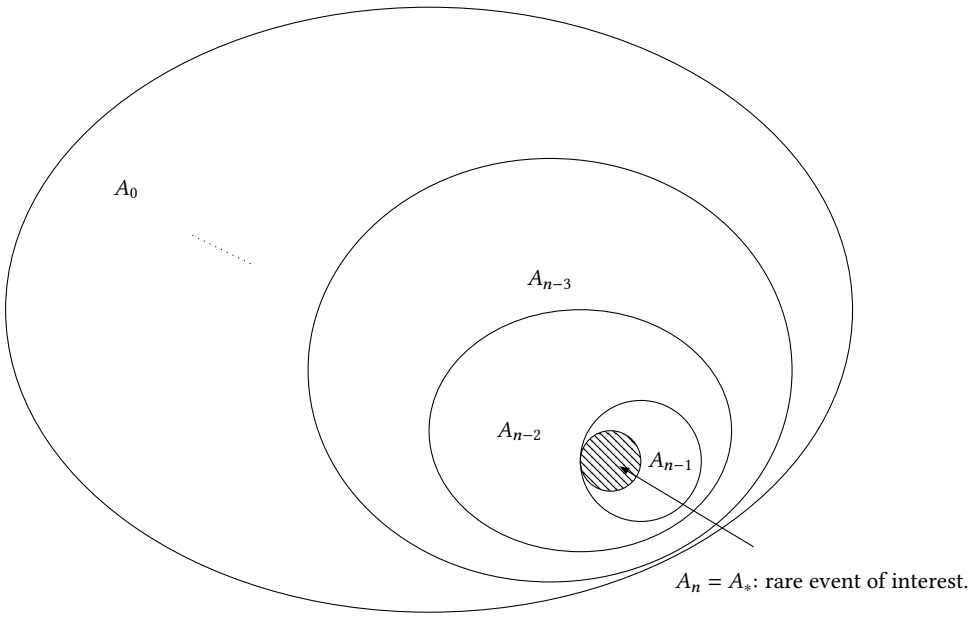

Figure 1.2: The big picture of Subset Simulation.

small compared to the probability of the rare event of interest, and one expects that when they are estimated separately, the variance of the estimation can thus be reduced. In order to achieve this, it is required to be able to simulate according to some nontrivial transition kernel $M_n(x, dy)$ at each time step $n \geq 0$ such that

$$\mathbf{P}_{A_n} M_n = \mathbf{P}_{A_n},$$

where $\mathbf{P}_{A_n}$ denotes the law of X given that $X \in A_n$, namely,

$$\mathbf{P}_{A_n}(dx) = \mathbf{1}_{x \in A_n} \eta(dx)/\eta(A_n).$$

A standard construction is given by Metropolis-Hastings techniques (cf. Figure 1.3). More precisely, we assume that $\eta(dx)$ is absolutely continuous w.r.t. some reference probability measure dx on E, and there exists an η-reversible Markov kernel K on the state space E, that is

$$\forall (x, y) \in E \times E, \quad \eta(dx)K(x, dy) = \eta(dy)K(y, dx).$$

A Metropolis-Hastings kernel can then be constructed:

$$M_n(x, dy) := K(x, dy)\mathbf{1}_{A_n}(x)\mathbf{1}_{A_n}(y) + \delta_x(dy)\left(\mathbf{1}_{E \setminus A_n}(x) + \mathbf{1}_{A_n}(x)K(x, E \setminus A_n)\right).$$

Since K is assumed to be η-reversible, standard calculations ensure that M_n is indeed $\mathbf{P}_{A_n}$-invariant. Roughly speaking, the Markov kernel M_n only accepts the transitions within the event A_n; all the other kinds of transitions are rejected. In practice, this Markov transition kernel may be applied several times to ensure a certain level of acceptance rate in order to ensure decorrelation between the initial condition and the new sample. Namely, one may replace M_n by $M_n^{(k)}$ defined by

$$M_n^{(k)} := \underbrace{M_n \cdot M_n \cdots M_n}_{k \text{ times}}.$$

Now, let us introduce the dynamic of Subset Simulation.

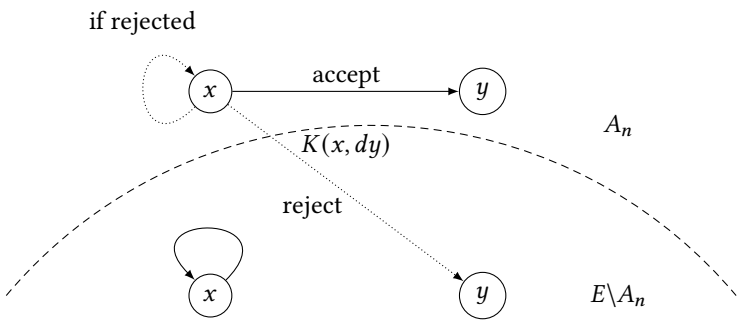

Figure 1.3: The construction of a $\mathbf{P}_{A_n}$-invariant kernel.

(i) Initialization:

Set $\tau_N = p = 0$;

For $i = 1, 2, \ldots, N$, draw $X_0^i \sim \eta(dx)$.

(ii) Absorbing time:

Denote by I_p the set of the indices of the surviving particles at level p, i.e.,

$$I_p := \left\{ i \in [N] \;\middle|\; X_p^i \in A_p \right\},$$

with $[N] := \{1, 2, \ldots, N\}$.

Stop the algorithm if $I_p = \varnothing$ and then set $\tau_N = (p-1) \vee 0$.

(iii) Selection:

For $i \in I_p$, let $\widetilde{X}_p^i = X_p^i$.

For $i \notin I_p$, uniformly draw a random index α_p^i in I_p and let $\widetilde{X}_p^i = X_p^{\alpha_p^i}$.

(iv) Mutation:

Set $\tau_N = p + 1$ and $p = p + 1$;

For $i = 1, 2, \ldots, N$, sample $X_p^i \sim M_p(\widetilde{X}_{p-1}^i, dx)$;

Stop the algorithm if $p = n$;

Otherwise, go to step (ii).

At each iteration of the algorithm, each term in (1.1) is estimated separately by

$$\mathbf{P}\left(X \in A_p \;\middle|\; X \in A_{p-1}\right) \approx \frac{\#I_p}{N} \times \mathbf{1}_{\tau_N \geq p}.$$

Therefore, the estimator of p_* is defined by

$$p_*^N := \frac{\#\left\{ i \;\middle|\; X_n^i \in A_* \right\}}{N} \prod_{p=0}^{n-1} \frac{\#I_p}{N} \times \mathbf{1}_{\tau_N \geq n}.$$

Note that it is possible that the particle system degenerates before the desired time horizon n, namely $\tau_N < n$, when the sequence of subsets $(A_p; 1 \leq p \leq n)$ or the transition kernels $(M_p; 1 \leq$

$p \leq n$) are poorly designed. Theoretically speaking, one may encounter a situation where Subset Simulation is worse than naive Monte Carlo. Hence, in practice, it is of crucial importance to construct properly and wisely the sequence of subsets. This is one of the key motivation of AMS methods, where these subsets are generated "on the fly" and in an optimal way. Let us provide a slightly modified version of the AMS algorithm introduced in [CDMFG12, CG16]. The setting will be referred to as *static* AMS in the following sections.

Adaptive Multilevel Splitting

We suppose that the rare event of interest is defined by

$$A_* := \{S(X) > L^*\},$$

where X is a random variable on E with distribution $\eta(dx)$ and $S : E \mapsto \mathbf{R}$ is a function that can be computed for each input value x, but is explicitly inaccessible or too complex to be studied analytically. The goal is to estimate the probability

$$p_* := \mathbf{P}\left(S(X) > L^*\right).$$

We call L^* the level of interest and this problem can be reformulated in the Subset Simulation framework by defining

$$A_p := \left\{S(X) > L_p\right\}$$

for a fixed sequence of levels

$$-\infty = L_{-1} < L_0 < L_1 < \cdots < L_{n-1} < L_n = L^*.$$

The idea of static AMS is to estimate a sequence of *adaptive* levels when generating the particle system at each iteration of the algorithm, by fixing a minimum number of particles K_* (or $K_*(N)$ which depends on the number of particles N) to kill at each resampling step. Accordingly, the mutation kernel M_n in the previous case is replaced by a sequence of Markov kernels M_L indexed by the level L. More precisely,

$$M_L(x, dy) := K(x, dy)\mathbf{1}_{\{S(x)>L\}}\mathbf{1}_{\{S(y)>L\}} + \delta_x(dy)\left(\mathbf{1}_{\{S(x)\leq L\}} + \mathbf{1}_{\{S(x)>L\}}K(x, S^{-1}\left((-\infty, L]\right)\right).$$

The construction is then identical to the previous case. Now, let us present the dynamic of static AMS as follows.

(i) Initialization:

Set $\tau_N = p = 0$;

For $i = 1, 2, \ldots, N$, draw $X_0^i \sim \eta(dx)$.

(ii) Level calculation:

Order the particles at iteration p such that

$$S(X_p^{\sigma(1)}) \leq S(X_p^{\sigma(2)}) \leq \cdots \leq S(X_p^{\sigma(N-1)}) \leq S(X_p^{\sigma(N)}),$$

where σ denotes a permutation on $[N]$;

Set $L_p^N = S(X_p^{\sigma(K^*)})$;

Stop the algorithm if $L_p^N > L^*$.

(iii) Absorbing time:

Denote by I_p the set of the indices of the surviving particles at iteration p,

$$I_p = \left\{ i \in [N] \;\middle|\; S(X_p^i) > L_p^N \right\}.$$

Stop the algorithm if $I_p = \varnothing$ and then set $\tau_N = (p-1) \vee 0$.

(iv) Selection:

For $i \in I_p$, let $X_{p+1}^i = X_p^i$;

For $i \notin I_p$, uniformly draw a random index $\alpha_p^i \in I_p$ and let $\widetilde{X}_p^i = X_p^{\alpha_p^i}$.

(v) Mutation:

Set $\tau_N = p+1$ and $p = p+1$;

For $i \in [N] \backslash I_p$, sample $X_p^i \sim M_{L_{p-1}^N}(\widetilde{X}_{p-1}^i, dx)$;

Go to step (ii).

Similarly, the estimator of the rare-event probability p_* is defined by

$$p_*^N := \frac{\#\left\{ i \;\middle|\; S(X_{\tau_N}^i) > L^* \right\}}{N} \prod_{p=0}^{\tau_N - 1} \frac{\# I_p}{N} \times \mathbf{1}_{L_{\tau_N}^N > L^*}.$$

We remark that the algorithm above is indeed different from the algorithm discussed in [CG16] due to some major difference in the resampling scheme. In the AMS framework discussed in [CG16], the particles on the same level are killed according to a new sequence of uniform random variables, in order to ensure that there are exactly K^* particles to be killed at each iteration. To the best of our knowledge, the adaptive method presented above has not been rigorously studied in theory. Partial results will be given later under the Asymmetric SMC framework in some special cases, that is, when the image of S is a finite set and K^* is set to be 1. Roughly speaking, this algorithm can be regarded as a static version of the gAMS algorithm introduced in [BGG+16]: at each iteration p, the resampled particles perform a mutation step according to the adaptive kernel $M_{L_p^N}$ and the surviving particles stay at the same site. Only the construction of the Markov kernel M_L is different. This resampling scheme is different from the mechanism of Subset Simulation introduced earlier in this section: in Subset Simulation, all the particles mutate after each resampling step, while in the AMS, only the killed particles mutate. The intuition of this modification is to reduce the computational cost brought by the mutation kernel $M_{L_p^N}$ at each resampling step. Imagine that K^* is set to be very small compared to N, say, $K^* = 1$. Most of the particles will survive at each resampling step. If we only execute the mutation to the resampled particles, the computational cost will be dramatically reduced. This resampling scheme will be referred to as *asymmetric* resampling scheme, namely, the surviving particles and the resampled particles mutate w.r.t. different Markov kernels. This difference will be addressed several times in the following sections as it is one of the core problems discussed in this thesis.

In the same spirit as for the static AMS framework studied in [CG16], we expect that with the same regularity assumptions, when the number of particles to kill $K_*(N)$ is set to be

$$K_*(N) := \lfloor (1-\alpha) N \rfloor$$

for a prefixed ratio $\alpha \in (0,1)$, the asymptotic behavior of the algorithm presented above is identical to a "limit" fixed-level Multilevel Splitting models with

$$L_0^* = F_{S(X)}^{-1}(1-\alpha) \le L_1^* = F_{S(X)}^{-1}(1-\alpha^2) \le \cdots \le L_{n-1}^* = F_{S(X)}^{-1}(1-\alpha^n) \le L^*,$$

where $F_{S(X)}^{-1}(1-\alpha)$ denotes the $(1-\alpha)$-quantile of $S(X)$. In this limit case, we have equal transition probabilities at each level, namely,

$$\mathbf{P}\left(S(X) > L_p^* \,\middle|\, S(X) > L_{p-1}^*\right) = \alpha.$$

It is well understood that this balanced case is the optimal one in terms of asymptotic variance for the fixed-level models. The interested reader is referred to Section 3.5 of [CGR19b] and reference therein for details.

1.1.3 Feynman-Kac particle models

For the theoretical analysis of the algorithms presented above, we use the language of Feynman-Kac particle models mainly developed by Del Moral and his co-authors (see, e.g., [DM04]), which provides a powerful and elegant approach to reformulate different models under the SMC framework. To give an idea, we start with a brief introduction on the connection between Feynman-Kac particle models and the Feynman-Kac semigroup. Let $(X_t; t \ge 0)$ be a Markov process taking values in $\mathbf{R}^d$ with initial distribution X_0 and transition function $P_{s,t}$. For a Borelian function $V : \mathbf{R}^d \mapsto \mathbf{R}$, the Feynman-Kac semigroup is defined by

$$\forall A \in \mathcal{B}(\mathbf{R}^d), \quad P_{s,t}^V(x,A) := \mathbf{E}\left[\mathbf{1}_A(X_t)\exp\left(\int_s^t V(X_s)ds\right) \,\middle|\, X_s = x\right].$$

Revealed by Feynman-Kac formula, the Feynman-Kac semigroup is strongly connected to the solutions of parabolic Partial Differential Equations. If one is interested in simulating according to this semigroup numerically, it is then necessary to consider the discretization in time. Taking $s = 0$, for any $t > 0$, we consider a time discretization

$$0 = t_0 < t_1 < t_2 < \cdots < t_{n-1} < t_n = t.$$

Under proper regularity assumptions, the Feynman-Kac kernel $P_{0,t}^V$ can be approximated by

$$P_{0,t}^V(x,A) \approx \mathbf{E}\left[\mathbf{1}_A(X_{t_n})\prod_{p=0}^{n-1}\exp\left\{V(X_{t_p})(t_{p+1}-t_p)\right\} \,\middle|\, X_0 = x\right],$$

where the Markov chain $(X_{t_n}; n \ge 0)$ is a discrete-time approximation of the continuous-time Markov process $(X_t; t \ge 0)$. With a slight abuse of notation, we denote $X_n := X_{t_n}$. The Markov kernel from X_{n-1} to X_n is denoted by M_n. In addition, we also denote

$$G_p(x) := \exp\left\{V(x)(t_{p+1}-t_p)\right\}.$$

As a consequence, we have

$$P_{0,t}^V(x,A) \approx Q_{0,n}(x,A) := \mathbf{E}\left[\mathbf{1}_A(X_n)\prod_{p=0}^{n-1}G_p(X_p) \,\middle|\, X_0 = x\right].$$

Similarly, for each $p < n$, $P^V_{t_p,t}(x, A)$ can be approximated by

$$P^V_{t_p,t}(x, A) \approx Q_{p,n}(x, A) := \mathbf{E}\left[\mathbf{1}_A(X_n)\prod_{k=p}^{n-1} G_k(X_k) \,\middle|\, X_p = x\right]. \tag{1.2}$$

The semigroup $Q_{p,n}$ defined above is called a discrete-time Feynman-Kac semigroup. More generally, given a Markov chain $(X_n; n \geq 0)$ with initial distribution η_0 and transition kernels $(M_n; n \geq 1)$, as well as a sequence of bounded nonnegative potential functions $(G_n; n \geq 0)$, we call $Q_{p,n}$ defined in (1.2) an element of the Feynman-Kac semigroup and we call $Q_n(x, dy) := Q_{n-1,n}(x, dy) = G_{n-1}(x)M_n(x, dy)$ a Feynman-Kac kernel. The measure γ_n defined by

$$\gamma_n(A) := \eta_0 Q_{0,n}(A) = \mathbf{E}\left[\mathbf{1}_A(X_n)\prod_{p=0}^{n-1} G_p(X_p)\right]$$

is called the Feynman-Kac terminal measure at time n. We also define the corresponding normalized version η_n by

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

Remark that η_n is well-defined if and only if

$$\forall n \geq 0, \quad \mathbf{E}\left[\prod_{p=0}^{n} G_p(X_p)\right] > 0,$$

which is assumed in the following. The elegant part of this formulation is that even for some a priori different applications, such as HMM and rare-event simulation, the essentials of the problems are identical. More precisely, in HMM, if we take $G_n(x) := p(y_n \mid x)$, the prediction problem of estimating $I_n(f_n)$ can be reformulated as to estimate $\eta_n(f_n)$. Similarly, for rare-event simulation, if we consider $G_n(x) := \mathbf{1}_{A_n}(x)$, the estimation of p_* can be reformulated as the estimation of $\gamma_n(\mathbf{1}_{A_*})$. Hence, in order to study the properties of the stochastic algorithms, which are essentially the same as explained in the next section, there is no need to return to the details of the models.

McKean interpretation and Interacting Particle System

Now, we give a brief introduction on how we can simulate according to the Feynman-Kac measures γ_n and η_n. First, we remark that

$$\gamma_n = \eta_n \prod_{p=0}^{n-1} \eta_p(G_p). \tag{1.3}$$

By definition, it is obvious that the unnormalized measures γ_{n-1} and γ_n are connected by the Feynman-Kac kernel Q_n, i.e., $\gamma_n = \gamma_{n-1}Q_n$. A natural question to ask is whether the sequence of probability measures $(\eta_n; n \geq 0)$ can also be connected by some transition kernels. We consider the McKean-type kernel $K_{n,\eta}(x, dy)$ indexed by some entry probability measure η such that $\eta(G_{n-1}) > 0$ and defined by

$$K_{n,\eta}(x, dy) := \epsilon G_{n-1}(x)M_n(x, dy) + (1 - \epsilon G_{n-1}(x))\frac{\eta Q_n(dy)}{\eta(G_{n-1})},$$

where ϵ is a constant in $[0, 1/\|G_{n-1}\|_\infty]$. Standard calculations give

$$\eta_{n-1}K_{n,\eta_{n-1}} = \eta_n.$$

By "McKean kernel" we mean that the definition of such kernel depends on the probability measure of the previous step, and a discrete-time stochastic process that is connected by McKean kernels is referred to as a McKean chain. Unlike the Markov case, the transition of the normalized measures from η_{n-1} to η_n depends on the probability measure η_{n-1} itself. Therefore, it is not possible to simulate directly according to $K_{n,\eta_{n-1}}$. A natural solution is to construct an Interacting Particle System (IPS), such that at step $n-1$, N particles $\mathbf{X_{n-1}} := (X^1_{n-1}, X^2_{n-1}, \ldots, X^N_{n-1})$ samples an approximation of η_{n-1}. Then, by exploiting the empirical measure

$$\eta^N_{n-1} := \frac{1}{N}\sum_{i=1}^{N} \delta_{X^i_{n-1}}, \tag{1.4}$$

we are able to sample from the approximated kernel $K_{n,\eta^N_{n-1}}$. Similarly, by replacing each normalized measure η_p by its particle approximation η^N_p in (1.3), we define

$$\gamma^N_n := \eta^N_n \prod_{p=0}^{n-1} \eta^N_p(G_p).$$

In this way, an IPS is therefore constructed, along with the particle approximations of $(\eta_n; n \geq 0)$ and $(\gamma_n; n \geq 0)$. Now, let us provide the dynamic of the Feynman-Kac IPS.

(i) Initialization:

Set $p = 0$ and $\tau_N = 0$;

Sample $\mathbf{X_0} \sim \eta_0^{\otimes N}$.

(ii) Absorbing time:

Stop if $\eta^N_p(G_p) = 0$ and then set $\tau_N = (p-1) \vee 0$.

(iii) Markov transition:

Set $\tau_N = p+1$ and $p = p+1$;

Sample $\mathbf{X_p} \sim \bigotimes_{i=1}^N K_{p,\eta^N_{p-1}}(X^i_{p-1}, \cdot)$;

Go to step (ii).

By definition, it is obvious that $(\mathbf{X}_n; n \geq 0)$ is a Markov chain with absorption taking values in E^N. Roughly speaking, the IPS can be seen as a numerical technique that uses a Markov chain defined in a high dimensional space to approximate a McKean chain in low dimensional space. Another remark is that all the details in the mechanism of resampling are encoded in the design of the McKean kernel $K_{n,\eta}$. For example, in HMM discussed in Section 1.1.1, the Bootstrap Filter corresponds to the case where $\epsilon = 0$. This resampling scheme is often referred to as *multinomial* resampling scheme. The setting is thus referred to as *Multinomial SMC*. More concretely, when $\epsilon = 0$, we have

$$K_{n,\eta}(x, dy) = \frac{\eta Q_n(dy)}{\eta(G_{n-1})}.$$

Let η_{n-1}^N denote the empirical measure defined in (1.4). It is then easy to verify that

$$K_{n,\eta_{n-1}^N}(x, dy) = \sum_{k=1}^N \frac{G_{n-1}(X_{n-1}^k)}{\sum_{\ell=1}^N G_{n-1}(X_{n-1}^\ell)} M_n(X_{n-1}^k, dy).$$

In practice, in order to sample $X_{n-1}^i \rightsquigarrow X_n^i$ according to this kernel, we first sample randomly a parent index based on the values of the potential functions

$$A_{n-1}^i \sim \sum_{k=1}^N \frac{G_{n-1}(X_{n-1}^k)}{\sum_{\ell=1}^N G_{n-1}(X_{n-1}^\ell)} \delta_k,$$

which corresponds to a multinomial *selection* mechanism. Then, based on the parent index, a Markov transition step is executed and often referred to as a *mutation* step:

$$X_n^i \sim M_n(X_{n-1}^{A_{n-1}^i}, \cdot).$$

For the Subset Simulation discussed in Section 1.1.2, the mechanism of the algorithm corresponds to the case $\epsilon = 1$. In this case, $\|G_n\|_\infty = 1$ since $G_n = \mathbf{1}_{A_n}$. Different from the previous case where $\epsilon = 0$, each particle performs a Bernoulli survival test based on its potential:

$$B_{n-1}^i \sim G_{n-1}(X_{n-1}^i)\delta_1 + \left(1 - G_{n-1}(X_{n-1}^i)\right)\delta_0.$$

Then, given the results of the survival tests $(B_{n-1}^i, 1 \le i \le N)$, the killed particles, namely, the particle such that $B_{n-1}^i = 0$, execute a multinomial resampling. Finally, all the particles, i.e., the surviving ones and the resampled ones, mutate according to the Markov kernel M_n.

In this way, by executing the mechanism encoded in the McKean kernel $K_{n,\eta}$, we are able to construct an Interacting Particle System (IPS), along with the particle approximations $(\eta_n^N; n \ge 0)$ and $(\gamma_n^N; n \ge 0)$ of the Feynman-Kac measures. Both of these cases are intensely studied in the nonadaptive context and many theoretical results, such as consistency and asymptotic normality of γ_n^N and η_n^N, can be found in the literature such as [DM04, DM13].

Genealogy, survival history and variance estimation

We call the parent indices $(A_n^i; 1 \le i \le N, n \ge 0)$ the *genealogy* of an IPS and we call the results of the Bernoulli survival tests $(B_n^i; 1 \le i \le N, n \ge 0)$ the *survival history* of an IPS. These structures provide additional information on the fluctuation of the particle system. In particular, we are interested by the variance estimation problem of the Feynman-Kac particle models.

Chan and Lai [CL13] have shown that one can derive consistent asymptotic variance estimators for η_n^N using the genealogy of the particle system. Combined with the asymptotic normality, one is able to conduct statistical inference with a single run of the algorithm. Then, Lee and Whiteley [LW18] provide a more refined analysis on the genealogy of the particle system, and an unbiased non-asymptotic variance estimator for γ_n^N is proposed. They also provide new asymptotic variance estimators for both η_n^N and γ_n^N, for which each term found in the decomposition of the asymptotic variance can be estimated separately. All of these studies are done under the multinomial resampling scheme, namely, for the case $\epsilon = 0$. In this thesis, we provide new variance estimators for the case $\epsilon = 1$, and it turns out that the information encoded in the survival history has to be taken into consideration. The study on these additional structures of the IPS is still new in the domain and very few theoretical tools are available.

Symmetric and asymmetric resampling schemes

Now, let us return to the static AMS methods introduced in Section 1.1.2. Besides the difference in the choice of levels, we have mentioned that the mutation mechanism is also different: in the static AMS setting, we only apply the mutation procedure to the resampled particles, while the surviving particles stay at the same place. For a given level L_p, let us denote

$$G_p(x) := \mathbf{1}_{\{S(x)>L_p\}}.$$

It is easy to check that this resampling scheme can be reformulated by the following McKean kernel:

$$K_{n,\eta}(x, dy) := G_{n-1}(x)\delta_x(dy) + (1 - G_{n-1}(x))\frac{\eta Q_n(dy)}{\eta(G_{n-1})}.$$

To the best of our knowledge, this setting is not covered in the literature of classic Feynman-Kac particle models. To make it clearer, we consider a slightly generalized setting. We assume that the sequence of potential functions $(G_n; n \geq 0)$ is $[0, 1]$-valued and two sequences of Markov kernels $(\mathring{M}_n; n \geq 1)$ and $(\dot{M}_n; n \geq 1)$ exist, such that

$$\gamma_p \mathring{Q}_{p,n} = \gamma_p \dot{Q}_{p,n} = \gamma_n,$$

where $\mathring{Q}_{p,n}$ and $\dot{Q}_{p,n}$ denote respectively the Feynman-Kac semigroups defined by

$$\mathring{Q}_n(x, dy) := G_{n-1}(x)\mathring{M}_n(x, dy),$$

and

$$\dot{Q}_n(x, dy) := G_{n-1}(x)\dot{M}_n(x, dy).$$

To give an idea, an illustration of the measures flow is presented in Figure 1.4.

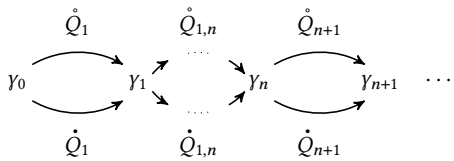

Figure 1.4: Feynman-Kac measures flow for Asymmetric SMC.

Now, we consider the McKean kernel $K_{n,\eta}$ defined by

$$K_{n,\eta}(x, dy) := \dot{Q}_n(x, dy) + (1 - G_{n-1}(x))\frac{\eta \mathring{Q}_n(dy)}{\eta(G_{n-1})}.$$

This setting will be referred to as *Asymmetric SMC* in this thesis, to emphasize that the surviving particles and the resampled particles mutate according to different Markov kernels. In particular, when $\mathring{Q}_n \equiv \dot{Q}_n$ for all $n \geq 1$, that is, all the particles mutate according to the same mutation kernel at each step, we say that the resampling scheme is *symmetric*. The setting thus returns to the classic Feynman-Kac particle models for the case $\epsilon = 1$ when the potential functions $(G_n; n \geq 0)$ are $[0, 1]$-valued. With a slight abuse of language, this setting is referred to as *Symmetric SMC*, to emphasize that the surviving particles and the resampled particles mutate according to the same mutation kernel. It is well-known that symmetric SMC admits smaller asymptotic variance than Multinomial SMC (see, e.g., Section 3.6.1). In general, the only additional requirement to

implement Symmetric SMC compared to Multinomial SMC is that an upper-bound of the potential function G_n needs to be tractable, in order to define "normalized" potential functions that are $[0,1]$-valued.

We also want to mention that in the static AMS algorithm, if S is a function that only takes finite values, and in addition, the number of particles to kill K^* is set to be 1, the static AMS algorithm enters the Asymmetric SMC framework. This important and nontrivial observation is inspired by the equivalence between SMC and gAMS revealed by Mathias Rousset in personal communications and the reader is referred to [CGR19a] for discussion on this topic. This is the original motivation for the introduction of Asymmetric SMC, which is one of the main topics discussed in this thesis.

It is well-known that the estimation given by the unnormalized measure γ_n^N is unbiased in the classic Feynman-Kac particle models, namely, when all the particles mutate according to the same Markov kernel at each iteration. However, if the underlying resampling scheme is asymmetric, this lack-of-bias property will be lost in general. At the same time, the consistency and the asymptotic normality are still valid. Considering the diversity of applications that can be reformulated as the estimation of Feynman-Kac measures, we hope that Asymmetric SMC framework may also provide new insights in other disciplines.

1.2 Rare-event simulation in Molecular Dynamics

In this section, we present a concrete example in Molecular Dynamics, namely, rare-event estimation in the overdamped Langevin dynamics with two metastable states. This example is the prime motivation for the gAMS framework widely used to sample reactive trajectories in Molecular Dynamics. The interested reader is referred to [BGG+16] for more details and a very rich list of numerical experiments.

Besides the original gAMS algorithm proposed in [BGG+16], we also provide a variant with multinomial resampling scheme, which is referred to as *Multinomial gAMS*. The reference mutation kernel (denoted by π_z in [BGG+16]) is set to be the most popular one in the AMS context (cf. Figure 1.7). Since the regularity assumptions on the mutation kernel, especially Assumption 2 in [BGG+16], are fundamental for the unbiasedness of gAMS, and quite abstract in their original form, we also provide some intuitive interpretations. The connection between gAMS and Asymmetric SMC framework is revealed through recent developments on the level-indexed process, which provide novel approaches to study the properties of gAMS thanks to Fleming-Viot particle systems. The reader is referred to [CDGR18a] for a more profound and detailed introduction. We also present the gap that exists in theoretical analysis, where the Asymmetric SMC framework failed to cover, and why its continuous-time generalization is an interesting topic to explore. Finally, we discuss some particular generalizations of gAMS algorithm that enter into Asymmetric SMC framework or its continuous-time generalization, and we explain heuristically why they are promising in solving the multi-channel problems in Molecular Dynamics.

1.2.1 Overdamped Langevin dynamics and metastability

The Markov chain $\mathbf{X} = (X_t, t \in \mathbf{N})$ is defined by the discretization of the overdamped Langevin dynamics:

$$\forall t \in \mathbf{N},\ X_{t+1} - X_t = -\nabla V(X_t)h + \sqrt{2\beta^{-1}}(W_{(t+1)h} - W_{th}). \tag{1.5}$$

Particularly, we let $X_0 = x_0 \in E := \mathbf{R}^{3d}$ and $X_t \in \mathbf{R}^{3d}$ is a random variable giving the positions of d particles in $\mathbf{R}^3$ at time th with $h > 0$ being the time step size. The function $V : E \mapsto \mathbf{R}$ is the potential energy function of the system. The constant $\beta = (k_B T)^{-1}$ denotes the inverse temperature and $(W_t; t \in \mathbf{R}_+)$ is a $3d$-dimensional Wiener process on E. We denote $P(x, dy)$ the transition kernel of $\mathbf{X}$.

A metastable state for $\mathbf{X}$ is an open subset of the state space, such that when the stochastic process $\mathbf{X}$ is trapped in it, it takes an extremely long time to escape. In the overdamped Langevin dynamics, a metastable state is typically a well in the energy landscape. In Transition Path Theory (see, e.g. [VE14]), it is important to study the reactive trajectories that transit between such metastable states, which is essentially a rare-event simulation problem. Let us denote A, B two metastable states in $\mathbf{R}^{3d}$. We assume that $x_0 \notin A \cup B$ but x_0 is close to A. As is illustrated in Figure 1.5, our goal is to estimate the probability of $\mathbf{X}$ reaching B before A, i.e.,

$$p_* := \mathbf{P}_{x_0}(\tau_B < \tau_A) = \mathbf{P}(\tau_B < \tau_A \mid X_0 = x_0), \tag{1.6}$$

where the stopping times are defined respectively by

$$\tau_A := \inf\{t \in \mathbf{N} : X_t \in A\} \quad \text{and} \quad \tau_B := \inf\{t \in \mathbf{N} : X_t \in B\}.$$

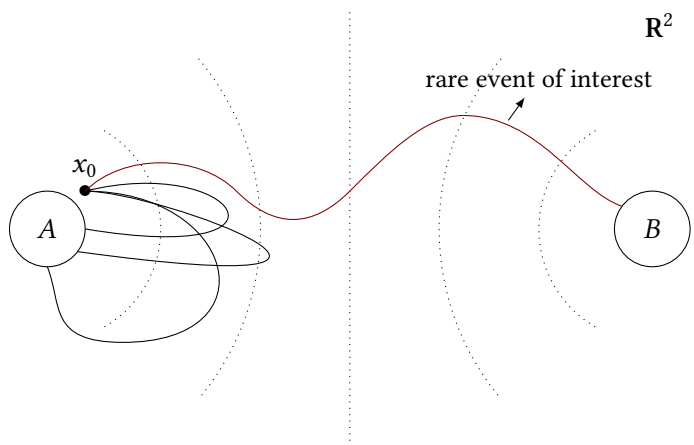

Figure 1.5: Schematic picture of a 2-dimensional toy example with the level sets of V (dashed lines).

1.2.2 Reaction coordinate

A reaction coordinate $\xi : E \backslash (A \cup B) \mapsto \mathbf{R}$ is a function designed to measure the advance of a reactive trajectory towards the metastable state B (cf. Figure 1.6). Generally speaking, the introduction of reaction coordinate can be regarded as a model reduction technique that uses a one dimensional summary statistics to collect information on the original Markov process or Markov chain. We suppose that there exists $L^* \in \mathbf{R}$ such that

$$B \subset \{x \in E : \xi(x) \in]L^*, \infty[\}.$$

For a given level $L \in \mathbf{R}$, we denote $S_L : E^{\mathbf{N}} \mapsto \mathbf{N}$ defined by

$$\forall \boldsymbol{x} = (x_t; t \in \mathbf{N}) \in E^{\mathbf{N}}, \quad S_L(\boldsymbol{x}) := \inf\{t \in \mathbf{N} : \xi(x_t) > L\}.$$

Similarly, for an open subset $F \subset E$, we denote

$$S_F(\boldsymbol{x}) := \inf\{t \in \mathbf{N} : x_t \in F\}.$$

We define the maximum level of a trajectory $\boldsymbol{x} \in E^{\mathbf{N}}$ by

$$\Xi(\boldsymbol{x}) := \sup_{t \in \mathbf{N}}\{\xi(x_{t \wedge S_{A \cup B}(\boldsymbol{x})})\}. \tag{1.7}$$

Remark that Ξ can be seen as a reaction coordinate in the path space $E^{\mathbf{N}}$.

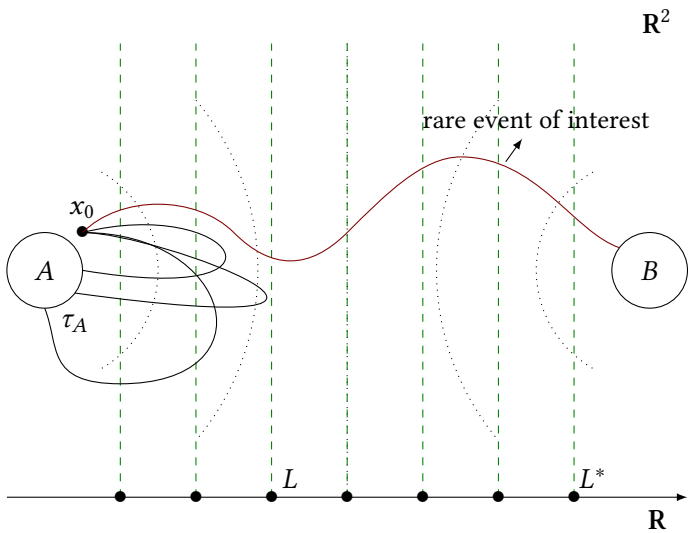

Figure 1.6: An example of reaction coordinate $\xi : (x, y) \mapsto x$.

We know that the choice of the reaction coordinate is crucial for the performance of gAMS, as well as for other rare-event simulation methods in this context. The optimal choice ξ^* in terms of asymptotic variance of the gAMS estimator of the probability p_* is the *committor function* (see, e.g., [BLR15, CGR19b]), which is defined by

$$\xi^*(x) = \mathbf{P}\left(\tau_B < \tau_A \mid X_0 = x\right). \tag{1.8}$$

It is known that for the overdamped Langevin dynamics, ξ^* is the solution of the following elliptic Partial Differential Equation:

$$-\nabla V \cdot \nabla u + \beta^{-1}\Delta u = 0 \quad \text{on } E \backslash (A \cup B),$$

with the boundary conditions

$$\begin{cases} u = 0 & \text{on } \partial A; \\ u = 1 & \text{on } \partial B. \end{cases}$$

1.2.3 A brief introduction to gAMS

Roughly speaking, gAMS is a generalized version of the static AMS algorithm introduced in Section 1.1.2, dedicated to sampling reactive trajectories between metastable states. Before proceeding further, we remark that the resampling in gAMS takes place in the path space $\boldsymbol{E} := E^{\mathbf{N}}$,

which is indeed a Polish space. Before giving the precise description of the gAMS algorithm, we want to mention that gAMS and static AMS as presented in Section 1.1.2 share the same mechanism when regarded as adaptive Feynman-Kac particle models. More concretely, if we replace the statistics S in the static AMS models by Ξ defined in (1.7), the only difference between these two algorithms is the construction of the adaptive mutation kernel M_L. In the static AMS models, M_L is built using a Metropolis-Hastings kernel (see Figure 1.3), while in the gAMS framework, the construction of M_L is guided by two regularity assumptions (Assumptions 1 and 2 in [BGG+16]). A popular choice of M_L that enters into gAMS framework is defined as follows, and an intuitive illustration is provided in Figure 1.7:

$$\mathbf{Y} \sim M_L(\boldsymbol{x}, \cdot) \Leftrightarrow \begin{cases} \widetilde{Y}_t = x_t & \text{if } t \leq S_L(\mathbf{X}); \\ \widetilde{Y}_t \sim P(\widetilde{Y}_{t-1}, \cdot) & \text{if } t > S_L(\mathbf{X}). \end{cases} \quad \text{and} \quad \mathbf{Y} = \left(\widetilde{Y}_{t \wedge S_{A \cup B}(\widetilde{Y})}\right)_{t \in \mathbf{N}}. \tag{1.9}$$

Let us consider $\mathbf{X}_{AB} := (X_{t \wedge S_{A \cup B}(\mathbf{X})}; t \in \mathbf{N})$ and denote by $\mathbf{P}_{\Xi > L}$ the law of $\mathbf{X}_{AB}$ given that $\Xi(\mathbf{X}_{AB}) > L$. It is easy to verify that M_L is indeed $\mathbf{P}_{\Xi > L}$-invariant. Notice that this resampling kernel heavily relies on the fact that we are in a dynamical setting, i.e., the random object of interest is a Markov process, whereas the static case discussed in Section 1.1.2 is more general, where the small probability is defined for any random variable.

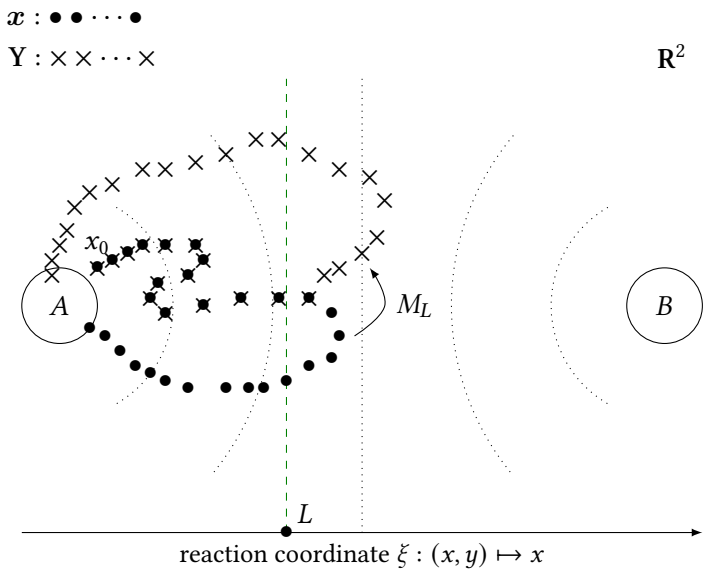

Figure 1.7: An illustration of M_L that enters into gAMS framework.

Now, let us present the mechanism of the gAMS algorithm. This time, we also track the genealogy and the survival history when generating the particle system. Recall that K^* denotes the minimum number of particles to kill at each resampling step.

(i) Initialization:

Set $\tau_N = p = 0$;

For $i = 1, 2, \ldots, N$, sample $\mathbf{X}^{(i,0)}$ independently starting from $X_0^{(i,0)} = x_0 \in E \backslash (A \cup B)$ with transition kernel P described by the time discretization of the overdamped Langevin dynamics (1.5), stopped when reaching A or B.

(ii) Level calculation:

Order the particles at iteration p such that

$$\Xi(\mathbf{X}^{(\sigma(1),p)}) \leq \Xi(\mathbf{X}^{(\sigma(2),p)}) \leq \cdots \leq \Xi(\mathbf{X}^{(\sigma(N-1),p)}) \leq \Xi(\mathbf{X}^{(\sigma(N),p)}),$$

where σ denotes a permutation on $[N]$;

Set $L_p^N = \Xi(\mathbf{X}^{(\sigma(K^*),p)})$;

Stop the algorithm if $L_p^N > L^*$.

(iii) Absorbing time:

Denote by I_p the set of the indices of the surviving particles at iteration p,

$$I_p = \left\{ i \in [N] \;\middle|\; \Xi(\mathbf{X}^{(i,p)}) > L_p^N \right\}.$$

Stop the algorithm if $I_p = \varnothing$ and then set $\tau_N = (p-1) \vee 0$.

(iv) Selection:

For $i \in I_p$, set $\mathbf{X}^{(i,p+1)} = \mathbf{X}^{(i,p)}$, $A_p^i = i$ and $B_p^i = 1$;

For $i \in [N] \backslash I_p$, uniformly draw a random index $\alpha_p^i \in I_p$ and set $\widetilde{\mathbf{X}}^{(i,p)} = \mathbf{X}^{(\alpha_p^i,p)}$, $A_p^i = \alpha_p^i$ and $B_p^i = 0$.

(v) Mutation:

Set $\tau_N = p+1$ and $p = p+1$;

For $i \in [N] \backslash I_p$, sample $\mathbf{X}^{(i,p)} \sim M_{L_{p-1}^N}(\widetilde{\mathbf{X}}^{(i,p)}, \cdot)$, where $M_{L_{p-1}^N}$ is defined in (1.9);

Go to step (ii).

An unbiased estimator p_*^N of the rare-event probability p_* is defined by

$$p_*^N := \frac{\#\left\{ i \;\middle|\; S_B(\mathbf{X}^{(i,\tau_N)}) < S_A(\mathbf{X}^{(i,\tau_N)}) \right\}}{N} \prod_{p=0}^{\tau_N - 1} \frac{\#I_p}{N} \times \mathbf{1}_{L_{\tau_N}^N > L^*}.$$

In particular, if the selection step (iv) and the mutation step (v) are replaced by the following variants, the corresponding algorithm will be referred to as *Multinomial gAMS*, for the fact that it can thus be reformulated as an adaptive Feynman-Kac particle model with multinomial resampling scheme. The Multinomial gAMS serves as a benchmark to be compared with the original gAMS algorithm, but is of no interest in practice.

- Selection:

  For $i \in [N]$, uniformly draw a random index $\alpha_p^i \in I_p$ and set $\widetilde{\mathbf{X}}^{(i,p)} = \mathbf{X}^{(\alpha_p^i,p)}$, $E_{p+1}^i = E_p^{\alpha_p^i}$ and $B_p^i = 0$.

- Mutation:

  Set $\tau_N = p+1$ and $p = p+1$;

  For $i \in [N]$, sample $\mathbf{X}^{(i,p)} \sim M_{L_{p-1}^N}(\widetilde{\mathbf{X}}^{(i,p)}, \cdot)$;

  Go to step (ii).

Remark that in Multinomial gAMS, there is no need to track the survival history, as all the particles are resampled at each iteration of the algorithm.

1.2.4 Level-indexed process

In static AMS, there is no unbiasedness property in general. However, by the nontrivial assumptions studied in [BGG+16], the gAMS turns out to provide unbiased estimations. This in practice allows the implementation of parallel computing. The consistency is therefore ensured by running multiple gAMS algorithms and using the average as the final estimator. To the best of our knowledge, the consistency and fluctuation analysis of gAMS w.r.t. the particle number N are still open problems. Recent results can be found in [CDGR18a], where the underlying dynamic is a continuous-time Markov process. The key ingredient is a nontrivial construction called *level-indexed process*. Without loss of generality, we suppose that $\xi(x_0) = 0$. For the Markov chain $\mathbf{X} = (X_n; n \in \mathbf{N})$, we define the level-indexed process $\mathbf{Y} = (Y_t; 0 \leq t \leq L^*)$ by

$$X_0 = Y_0 \quad \text{and} \quad \begin{cases} Y_t = X_{S_t \wedge L^*} & \text{if } S_t(\mathbf{X}) < S_A(\mathbf{X}); \\ Y_t = \boldsymbol{\partial} & \text{if } S_t(\mathbf{X}) \geq S_A(\mathbf{X}), \end{cases} \tag{1.10}$$

where $\boldsymbol{\partial}$ denotes the cemetery point. To simplify the notation, we assume that $B = \{\xi > L^*\}$. In this case, we consider the potential function $G_t(y) := \mathbf{1}_{y \neq \boldsymbol{\partial}}$. First, we provide an intuitive interpretation of Assumption 2 of [BGG+16]. Let M_L be a $\mathbf{P}_{\Xi > L}$-invariant kernel. When a reference trajectory $\boldsymbol{x}$ is resampled according to M_L, the part $(x_0, x_1, \ldots, x_{S_L(\boldsymbol{x})-1})$ can not be changed. Otherwise, the mutation kernel M_L does not enter into gAMS framework. Roughly speaking the assumptions of [BGG+16] can be regarded as the condition such that gAMS w.r.t. $\mathbf{X}$ can be reformulated as some continuous-time Symmetrical SMC w.r.t. $\mathbf{Y}$, namely, the Fleming-Viot particle systems. Therefore, the unbiasedness property is provided by the structure of the continuous-time Feynman-Kac particle models. In particular, it is easily checked that in the case where ξ can only take finite values, say $0, 1, 2, \ldots, n^*$, the estimation of p^* can be reformulated by Symmetric SMC framework w.r.t. $(Y_m; 0 \leq m \leq n^*)$. However, when these assumptions are not verified, it is then impossible to construct the level-indexed process in general. Fortunately, we still can reformulate the gAMS algorithm as Asymmetric SMC models, expressed in its continuous-time generalizations on the path space. In such scenario, our work on the discrete-time Asymmetric SMC failed to provide theoretical guarantee. So, it would be interesting to explore the continuous-time Asymmetric SMC framework, which is left for future research.

In our analysis, we will also discuss generalizations of gAMS to new resampling kernels. The motivation for introducing the resampling kernels that do not enter into gAMS framework is the multi-channel problem, which is also presented in [BGG+16]. Roughly speaking, starting from the neighborhood of A, the reactive trajectories may reach B from different channels that are extremely unbalanced in terms of probability. This induces large variances on the estimators. The multi-channel problem is partially due to the dependence between the trajectories, especially by the fact that a resampled trajectory and its parent share the same piece before the time when they first pass the reference level of the mutation kernel. To counter this, an elementary approach is to conduct rejection sampling from a lower level, which is not always practical due to the uncontrollable computational costs. A better choice is the Particle Markov Chain Monte Carlo (PMCMC) [ADH10] sampler since one can easily construct a $\mathbf{P}_{\Xi > L}$-invariant kernel by implementing another gAMS using the freezing technique. The computational costs can therefore be controlled by the size of the IPS. In both cases, the mutation kernels do not enter into gAMS framework, even though the invariant property w.r.t. $\mathbf{P}_{\Xi > L}$ is verified.

1.3 Summary of the Chapters

In this section, we summarize the main contributions of the present thesis. We also briefly present some intuition or heuristic knowledge we have collected when studying these subjects, which may help future developments on these topics.

1.3.1 Chapter 2: Variance estimation in Adaptive Sequential Monte Carlo

A generalized version of the *Adaptive SMC via summary statistics* introduced in Section 2 of [BJKT16] is investigated. We provide the conditions under which the asymptotic normality is available, such that the asymptotic variance is identical to a "limiting" non-adaptive SMC model. At the same time, we prove that the variance estimators proposed by Lee and Whiteley in a non-adaptive context [LW18] are still consistent in the adaptive context.

Main results

Let $(E_n; n \geq 0)$ be a sequence of Polish spaces. For each level $n \geq 1$, we consider a family of potential functions $G_{n-1,z} : E_{n-1} \mapsto \mathbf{R}_+$ and Markov kernels $M_{n,z} : (E_{n-1}, \mathcal{B}(E_n)) \mapsto [0,1]$, parametrized by $z \in \mathbf{R}^d$. Accordingly, we define the family of 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

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

By exploiting some summary statistics $\zeta_n : E_n \mapsto \mathbf{R}^d$, such that, for all $n \geq 0$, $\eta_n(\zeta_n) = z_n^*$, we consider the adaptive Feynman-Kac particle models identified by the McKean kernel

$$K_{n,\eta}(x, dy) := \frac{\eta Q_{n,\eta(\zeta_{n-1})}(dy)}{\eta(G_{n-1,\eta(\zeta_{n-1})})}.$$

This is an adaptive Feynman-Kac particle model under multinomial resampling scheme. Now, we introduce the regularity assumptions required to establish asymptotic normality and to conduct variance estimation. They are a slight generalization of the ones presented in [BJKT16] to obtain consistency and asymptotic normality in the same context.

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

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

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

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

The function h_n is assumed to satisfy the following properties:

- The Euclidean norm $|h_n|$ is bounded over $E_n \times \mathbf{R}^d$ 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 $\left|z_n - z_n^*\right| < \delta$ implies

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

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

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 $k \in [d]$, ζ_n^k belongs to $\mathcal{B}_b(E_n)$.

Under these assumptions, we have, for any test function $f \in \mathcal{B}_n(E_n)$,

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

and

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

Moreover, convergent variance estimators of the asymptotic variances are provided respectively by

$$N\gamma_n^N(1)^2 V_n^N(f) \xrightarrow[N\to\infty]{\mathbf{P}} \sigma_{\gamma_n}^2(f),$$

and

$$N V_n^N(f - \eta_n^N(f)) \xrightarrow[N\to\infty]{\mathbf{P}} \sigma_{\eta_n}^2(f - \eta_n(f)),$$

with, if E_n^i denotes the index of the ancestor of the particle X_n^i at step 0,

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

Note that these estimators are the same as the ones proposed by Lee and Whiteley in [LW18] in a non-adaptive context.

Other comments

In Chapter 2, we provide a methodology such that the adaptiveness of the Feynman-Kac particle models and the variance estimation problem can be treated separately. The theoretical tools we developed are mainly inspired by the pioneering work [LW18] of Lee and Whiteley, and we made some modifications such that the randomness brought by the adaptiveness is easier to deal with. As a consequence, the framework is also more consistent with previous works such as [CDMG11] and [DMPR09]. We expect that the adaptive models that are not covered by our regularity assumptions, such as [CG16], can also be studied in a similar way. Heuristically speaking, the regularity assumptions needed to conduct variance estimation are in general weaker than the ones required to establish asymptotic normality. Therefore, the take-home message is simple: it is expected that when an adaptive model has the same asymptotic behaviors as a "limiting" non-adaptive model, the asymptotic variance estimators are also available and are identical to the ones that can be derived in the corresponding non-adaptive context.

1.3.2 Chapter 3: Asymmetric Sequential Monte Carlo

As briefly discussed in Section 1.1.3, Asymmetric SMC is proposed as a generalization of the classic Feynman-Kac particle models, aiming at reducing the computational costs brought by the mutation kernels at each iteration of the algorithm. In a very general setting, we provide consistency and asymptotic normality for the Asymmetric SMC framework, which, in particular, also covers the Symmetric SMC framework. In addition, we construct consistent and efficient asymptotic variance estimators, that are available as a by-product of the simulation of a single run of the particle system. Moreover, we also propose unbiased variance estimators for the unnormalized measure γ_n^N under Symmetric SMC framework. Since the computations of the variance estimators are highly nontrivial, we provide detailed and efficient algorithms w.r.t. time and space complexity in Section 3.5.

Main results

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

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

Assuming that $\gamma_n(1) > 0$ for any $n \geq 0$, we also define $\eta_n(f) := \gamma_n(f)/\gamma_n(1)$. For any test function $f \in \mathcal{B}_b(E_n)$, when the number of particle N tends to infinity, the estimators given by Algorithm 2 in Section 3.5, denoted respectively by $\gamma_n^N(f)$ and $\eta_n^N(f)$, converge almost surely to $\gamma_n(f)$ and $\eta_n(f)$ if for any $n \geq 1$, we have

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

Moreover, we also have

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

as well as

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

where the computations of $\hat{\sigma}_{\gamma_n^N}(f)$ and $\hat{\sigma}_{\eta_n^N}(f - \eta_n^N(f))$ are respectively provided in Algorithm 5 and Algorithm 6 in Section 3.5.

Moreover, under the condition discussed in Section 3.4.1, which at least contains the case where $\dot{M}_n \equiv \overset{\circ}{M}_n$ for any $n \geq 1$, the estimator $\gamma_n^N(f)$ is an unbiased estimator for $\gamma_n(f)$. Moreover, the estimator provided by Algorithm 7 is an unbiased estimator for the non-asymptotic variance of $\gamma_n^N(f)$.

Other comments

The theoretical tools introduced in Chapter 3 can be regarded as a development of the ones introduced in Chapter 2, in order to deal with the challenge brought by the asymmetry in the associated particle system. Although no rigorous mathematical foundation of the adaptive Asymmetric SMC is provided in this chapter, we expect that in the same spirit as for the Adaptive SMC

framework studied in Chapter 2, the variance estimators would still be valid if the underlying resampling scheme is also changed accordingly. Many by-products of the analysis on the variance estimators can be found in Section 3.7.8, which also help to understand the difference between Symmetric and Asymmetric SMC frameworks. We also expect that a similar methodology can be applied to derive variance estimators for Fleming-Viot particle systems [DCGR17], which enters the continuous-time generalization of Symmetric SMC. In this respect, since we have mentioned many times that gAMS enters into Asymmetric SMC framework when the reaction coordinate ξ only takes finite values and the number of particles to kill is set to be $K^* = 1$, it would also be interesting to explore the continuous-time generalization of Asymmetric SMC framework, i.e., when ξ can be chosen as a continuous function.

1.3.3 Chapter 4: Estimating committor function with Mondrian Forests

As briefly mentioned in Section 1.2.2 and illustrated in the rich list of numerical experiments in [BGG+16], the choice of the reaction coordinate ξ is crucial to the performance of gAMS and many other rare-event estimation algorithms in Molecular Dynamics. In many cases, the optimal choice is the so-called committor function ξ^* defined in (1.8). Interestingly, gAMS can also be used to estimate the values of the committor function. Hence, if we provide a regressor that uses the estimations of gAMS, with some conventional reaction coordinate, as the training data, and we use the trained model as the reaction coordinate to rerun gAMS algorithm to generate new training data, it is expected that such iterative updating strategy may help to improve the accuracy of both gAMS and the regressor. This is the crucial idea of Chapter 4.

In this chapter, we provide some preliminary analysis on the principles of the choice of the regressor, and our choice is a new member in the family of Random Forests called Mondrian Forests, recently proposed in [LRT14]. Roughly speaking, Mondrian Forests are a Random Forests-based model such that online learning is available, and implemented in an elegant way. Several interaction strategies between gAMS and Mondrian Forests are presented and we also discuss some possible improvements that can be applied to make Mondrian Forests even more suitable.

Chapter 2

Variance Estimation in Adaptive Sequential Monte Carlo

abstract: 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 chapter 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.* [BJKT16]. We prove that, under natural assumptions, the estimator introduced by Lee and Whiteley [LW18] 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.

2.1 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 [DdFG01] 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 [DM04, DM13]).

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., [DM04, Cho04, DM08])

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

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 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 [CL13], by using the ancestral information encoded in the genealogy of the associated IPS. Then, Lee and Whiteley [LW18] proposed an unbiased variance estimator for the unnormalized measures γ_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 chapter, 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 [BJKT16] and [CG16]. The framework discussed in the present paper is just a slightly generalized version of the one presented in Section 2 in [BJKT16] 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 [AB01, AB03, CG07, CDMFG12, CG16]. 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 [CG16] and [BJKT16] 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 [LW18], 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 Γ_n^b that appear in the expansion of the variance given in [CDMG11]. 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 [DMKP16] to study propagation of chaos properties of Conditional Particle Markov Chain Monte Carlo methods [ADH10]. Above the specific context of the present chapter, 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 $(E, \mathcal{B}(E))$ while $\mathcal{B}_b(E)$ denotes the collection of the bounded measurable functions from $(E, \mathcal{B}(E))$ to $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ 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

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

  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

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

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 $H : E \times E \mapsto \mathbf{R}$, we have

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

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

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

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

$$m^{\otimes 2}(\boldsymbol{x}) = \frac{N-1}{N} m^{\odot 2}(\boldsymbol{x}) C_0 + \frac{1}{N} m^{\odot 2}(\boldsymbol{x}) C_1. \tag{2.1}$$

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

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

2.2 Adaptive Sequential Monte Carlo

This section presents the formal definition and the regularity assumptions of the ASMC framework studied in this chapter. The motivation is mainly from ASMC via summary statistics introduced in Section 2 of [BJKT16]. We refer the reader to the latter for details on motivating examples such as filtering or sequential Bayesian parameter inference.

2.2.1 Framework

The notations that are adopted are essentially those in the pair of books [DM04, DM13]. 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 $G_{n-1,z} : E_{n-1} \mapsto \mathbf{R}_+$ and Markov kernels $M_{n,z} : (E_{n-1}, \mathcal{B}(E_n)) \mapsto [0,1]$ parametrized by $z \in \mathbf{R}^d$. Accordingly, we define the family of nonnegative 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

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

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

$$\gamma_n(1) = \eta_n(f_n) \prod_{p=0}^{n-1} \eta_p(G_p). \tag{2.2}$$

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

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

2.2.2 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 $(\mathbf{X_n})_{n\geq 0}$ taking values in $(E_n^N, \mathcal{B}(E_n)^{\otimes N})_{n\geq 0}$ with genealogy $(\mathbf{A_n})_{n\geq 0}$ 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 η_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 (2.2), 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 $\mathbf{X_0} \sim \eta_0^{\otimes N}$.

(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 $\mathbf{X_p} = \boldsymbol{x_p}$, we make an independent multinomial selection of the parent of each particle by

$$S_{p,N}(\boldsymbol{x_p}, da_p^i) = \sum_{k=1}^N \frac{G_{p,N}(x_p^k)}{\sum_{j=1}^N G_{p,N}(x_p^j)}\delta_k(da_p^i). \tag{2.3}$$

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

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

- Mutation: given the parent indices $\mathbf{A_p} = \boldsymbol{a_p}$, 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 $\mathbf{X_p}$ and $\mathbf{A_p}$, we have

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

2.2.3 Assumptions

Our assumptions are introduced in a similar way as in [BJKT16], 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 [BJKT16] explains why this case is particularly interesting in practice. In the following sections, we use $\mathcal{A}$ as a shorthand for *Assumption*.

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

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

Notice that, under $\mathcal{A}1$, Equation (2.3) 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 τ_N defined by

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

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 thesis. Let us mention that the strict positivity and boundedness of the potential functions is also required in [BJKT16] (see page 1116 and Assumption 1 page 1118). In our second assumption, "$\langle \cdot, \cdot \rangle$" stands for the Euclidean scalar product in $\mathbf{R}^d$ 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 $h_n : (E_n \times \mathbf{R}^d, \mathcal{B}(E_n) \otimes \mathcal{B}(\mathbf{R}^d)) \to (\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ such that, for all $(x, z_n) \in E_n \times \mathbf{R}^d$,*

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

The function h_n is assumed to satisfy the following properties:

- *The Euclidean norm $|h_n|$ is bounded over $E_n \times \mathbf{R}^d$ 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 $\left|z_n - z_n^*\right| < \delta$ implies*

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

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

Moreover, the summary statistics $\zeta_n = (\zeta_n^1, \dots, \zeta_n^d)$ satisfies $z_n^* = \eta_n(\zeta_n)$ and is such that, for all $k \in [d]$, ζ_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 [BJKT16]. Indeed, our function h_n coincides with the function ω defined in (2.17) of [BJKT16], 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 [BJKT16] 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

$$\forall \epsilon > 0,\ \exists g_n \in \mathcal{B}_b(E_n),\ \exists \delta > 0,\ s.t.$$
$$\left|z_n - z_n^*\right| < \delta \implies \forall x \in E_n,\ \left|h_n(x, z_n) - h_n(x, z_n^*)\right| < g_n(x)\epsilon.$$

We expect that 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 [BJKT16] and the AMS framework studied in [CG16].

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 2.2.1 below and, more specifically, in the estimation of the asymptotic variance. Nevertheless, let us briefly mention that to establish the consistency of γ_n^N and η_n^N, one just needs

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

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

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

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

2.2.4 Central limit theorems

As explained before, the present chapter 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 [BJKT16] under slightly weaker assumptions. This is why, in Section 2.4.2, we propose a different strategy for the proof.

Theorem 2.2.1. *Assume $\mathcal{A}1$-$\mathcal{A}2$. For any test function $f \in \mathcal{B}_n(E_n)$, we have*

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

and

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

where

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

One can notice that the CLT for η_n^N is just a consequence of the CLT for γ_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 chapter is to estimate the asymptotic variances $\sigma^2_{\gamma_n}(f)$ and $\sigma^2_{\eta_n}(f - \eta_n(f))$ by a single simulation of the particle system, exactly as is done by Lee and Whiteley in [LW18] in a nonadaptive context.

2.3 Variance estimations

In this section, we recall the coalescent tree-based expansion of the variance firstly introduced in [CDMG11] from which we deduce a new variance estimator. We also recall the variance estimator proposed by Lee and Whiteley in [LW18] and explain the connection between both estimators.

2.3.1 Coalescent tree-based variance expansion

We call $b := (b_0, \dots, b_n) \in \{0,1\}^{n+1}$ a coalescence indicator where $b_p = 1$ indicates that there is a coalescence at level p.

Definition 2.3.1. *We associate with any coalescence indicator $b \in \{0,1\}^{n+1}$ the nonnegative measures Γ_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 2.1). When there is no coalescence at all, that is $b = (0, \dots, 0)$, we have

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

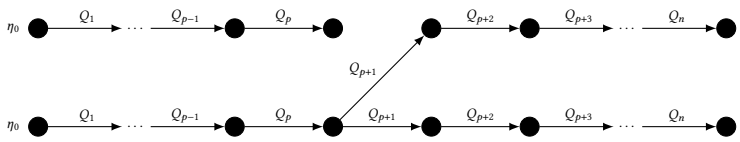

Figure 2.1: A representation of the coalescent tree-based measure $\Gamma_n^{(p)}$.

By definition, it is easy to verify that

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

As noticed in [CDMG11], the latter yields alternative representations for the asymptotic variances of Theorem 2.2.1, namely

$$\sigma^2_{\gamma_n}(f) = \sum_{p=0}^{n} \left(\Gamma_n^{(p)}(f^{\otimes 2}) - \Gamma_n^{(\varnothing)}(f^{\otimes 2})\right), \tag{2.4}$$

and

$$\sigma^2_{\eta_n}(f) = \sum_{p=0}^{n} \left(\bar{\Gamma}_n^{(p)}(f^{\otimes 2}) - \bar{\Gamma}_n^{(\varnothing)}(f^{\otimes 2})\right). \tag{2.5}$$

As a consequence, if for any coalescence indicator $b := (b_0, \ldots, b_n) \in \{0,1\}^{n+1}$, we can construct a consistent estimator $\bar{\Gamma}^b_{n,N}$ of $\bar{\Gamma}^b_n$, then we automatically deduce consistent estimators for the asymptotic variances of Theorem 2.2.1. 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 $X_n^{\ell_n^{[2]}} = (X_n^{\ell_n^1}, X_n^{\ell_n^2})$ to shorten the writings.

Definition 2.3.2. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator b, we define the estimator $\bar{\Gamma}^b_{n,N}$ of the measure $\bar{\Gamma}^b_n$ by*

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

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

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

The estimator of Γ_n^b is defined by

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

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

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

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

$$\bar{\Gamma}^b_{0,N}(F) := \frac{1}{N(N-1)} \sum_{\ell_0^{[2]} \in (N)^2} C_{b_0}(F)(X_0^{\ell_0^{[2]}}) = \frac{1}{N(N-1)} \sum_{i \neq j} C_{b_0}(F)(X_0^i, X_0^j). \tag{2.6}$$

We also adopt the convention

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

A toy example As the definition of the estimator $\Gamma^b_{n,N}$ is not completely straightforward, we illustrate the idea on a simple example. For this, we consider the IPS of Figure 2.2.

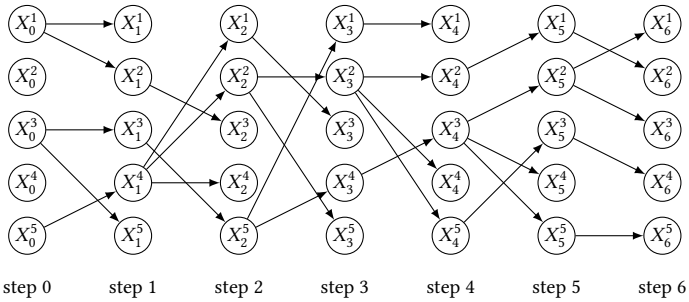

Figure 2.2: An IPS with $n + 1 = 7$ levels and $N = 5$ particles at each level.

Suppose we want to estimate $\bar{\Gamma}^{(3)}_6(F)$ by $\bar{\Gamma}^{(3)}_{6,5}(F)$. 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 $\ell^{[2]}_{0:6}$ such that

$$\prod_{p=0}^{5} \lambda_p^{b^*}(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) = 1. \tag{2.7}$$

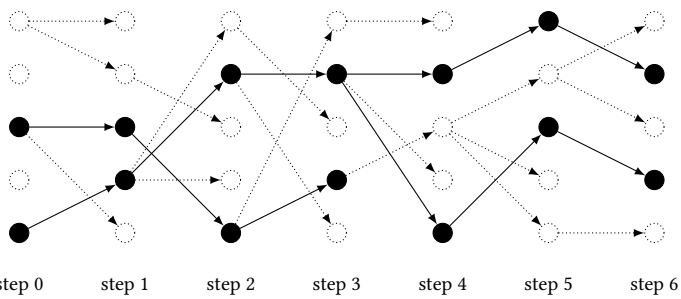

Figure 2.3: The first family of $\ell^{[2]}_{0:6}$ such that (2.7) is verified.

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 2.3):

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

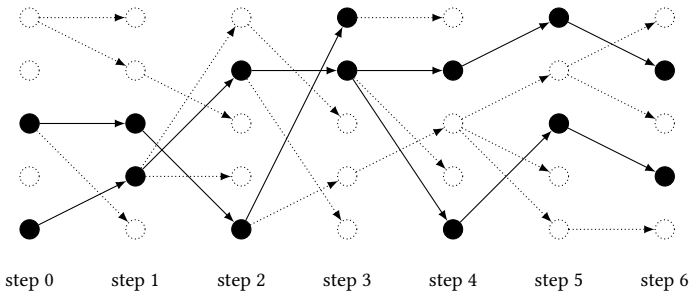

Figure 2.4: The second family of $\ell_{0:6}^{[2]}$ such that (2.7) is verified.

The second couple of ancestral lines is (see Figure 2.4):

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

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 2.3.1 (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*

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

The proof is given in Section 2.4.3.

2.3.2 Term by term estimator

Considering (2.4), (2.5), and Theorem 2.3.1, we are now in a position to provide term by term variance estimators for $\sigma_{Y_n}^2(f)$ and $\sigma_{\eta_n}^2(f)$.

Definition 2.3.3 (Estimators of the asymptotic variances). *Given a test function $f \in \mathcal{B}_b(E_n)$, we let*

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

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 2.3.1 ensures the consistency of both $\Gamma_{n,N}^{(p)}(f^{\otimes 2})$ and $\Gamma_{n,N}^{(\varnothing)}(f^{\otimes 2})$. Returning to the equation (2.4), this amounts to say that

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

Similarly, for the consistency of $\sigma_{\eta_{n,N}}^2(f-\eta_n^N(f))$, since by (2.5) 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,

$$\bar{\Gamma}_{n,N}^{b}\left(\left[f-\eta_n^N(f)\right]^{\otimes 2}\right) \xrightarrow[N\to\infty]{\mathbf{P}} \bar{\Gamma}_{n}^{b}\left(\left[f-\eta_n^N(f)\right]^{\otimes 2}\right). \tag{2.8}$$

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

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

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 2.2.1 is that

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

the verification of (2.8) is just a consequence of Theorem 2.3.1 and Slutsky's Lemma. Hence, we have obtained the following result.

Theorem 2.3.2 (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*

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

as well as

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

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 [LW18]. The interested reader can find more details on this point in Appendix 2.5.3.

2.3.3 Disjoint ancestral lines estimator

Let us now recall the variance estimator proposed in [LW18], which can be seen as a disjoint ancestral lines estimator. Namely, given a test function $f \in \mathcal{B}_b(E_n)$, consider

$$V_n^N(f) := \eta_n^N(f)^2 - \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{E_n^i \neq E_n^j} f(X_n^i) f(X_n^j), \tag{2.9}$$

where E_n^i is the ancestor index of X_n^i at level 0. Returning to the toy example of Section 2.3.1, 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 [LW18] when the number N of particles is the same at each step. The reader is referred to [LW18] for an efficient algorithm to compute this estimator.

According to our notation, since $E_n^i \neq E_n^j$ corresponds to the case $b = (0, \dots, 0) = (\varnothing)$ 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 2.4.5.

Proposition 2.3.1. *Assume $\mathcal{A}1$. For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$N V_n^N(f) - \sigma^2_{\eta_{n,N}}(f) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{N}\right),$$

and

$$N V_n^N(f - \eta_n^N(f)) - \sigma^2_{\eta_{n,N}}(f - \eta_n^N(f)) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{N}\right).$$

By combining Theorem 2.3.2 and Proposition 2.3.1, we finally obtain the main result of the present chapter.

Theorem 2.3.3. *Assume $\mathcal{A}1$-$\mathcal{A}2$. For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$N \gamma_n^N(1)^2 V_n^N(f) - \sigma^2_{\gamma_n}(f) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right),$$

and

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

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 in 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 2.5.3.

As mentioned 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.

2.4 Proofs

2.4.1 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.

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

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

  where, given $\mathbf{X_{p-1}}$, K_{p,η_{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 2.2.2. 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 $k \in [d]$, ζ_n^k belongs to $\mathcal{B}_b(E_n)$.

Theorem 2.4.1. *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 γ_n^N implies the convergence of η_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

$$\begin{aligned}&\left|\gamma_n^N(f) - \gamma_n(f)\right| \\ &\quad \leq \underbrace{\left|\gamma_n^N(f) - \gamma_{n-1}^N Q_{n,N}(f)\right|}_{P_1(N)} + \underbrace{\left|\gamma_{n-1}^N Q_{n,N}(f) - \gamma_{n-1}^N Q_n(f)\right|}_{P_2(N)} + \underbrace{\left|\gamma_{n-1}^N Q_n(f) - \gamma_{n-1} Q_n(f)\right|}_{P_3(N)}. \end{aligned} \tag{2.10}$$

- 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

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

We consider the filtration $(\mathcal{E}_n^i)_{0\leq i\leq N}$ defined by

$$\mathcal{E}_n^i := \mathcal{F}_{n-1}^N \vee \sigma(X_n^1, \ldots, X_n^i)$$

with $\mathcal{E}_n^0 = \mathcal{F}_{n-1}^N := \sigma(\mathbf{X_0}, \ldots, \mathbf{X_{n-1}})$. By definition, we have

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

Thus, $(U_{n,N}^i)_{0\leq i\leq N}$ is an $(\mathcal{E}_n^i)_{0\leq i\leq N}$-martingale difference array. Under $\mathcal{A}1$, we have

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

Therefore, for any $\epsilon > 0$, Hoeffding-Azuma inequality gives

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

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 that

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

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 (2.10), 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. □

2.4.2 Proof of Theorem 2.2.1

We prove by induction that

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

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

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

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

$$\forall 0 \leq p \leq n-1, \quad \sqrt{N}\left|Z_{p-1}^N - z_{p-1}^*\right| = \mathcal{O}_{\mathbf{p}}(1). \tag{2.11}$$

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

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

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

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

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

$$\sqrt{N} \left(\gamma_n^N(f) - \gamma_n(f) \right) = \sum_{k=1}^{(n+1)N} \left(U_k^N + \frac{1}{N} D_{p_k}^N \right) = \sum_{k=1}^{(n+1)N} U_k^N + \sum_{p=0}^{n} D_p^N. \tag{2.12}$$

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

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

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

$$\begin{aligned} &\left| \sqrt{N} \left\langle \gamma_{p-1}^N \left(h_{p-1}(\cdot, Z_{p-1}^N) - h_{p-1}(\cdot, z_{p-1}^*) \right), Z_{p-1}^N - z_{p-1}^* \right\rangle \right| \\ \leq& \sqrt{N}\, \gamma_{p-1}^N(1) \left| Z_{p-1}^N - z_{p-1}^* \right| \sup_{x \in E_{p-1}} \left| h_{p-1}(x, Z_{p-1}^N) - h_{p-1}(x, z_{p-1}^*) \right|. \end{aligned}$$

  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 2.4.1, Ω_{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 (2.11) that

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

- For the second part, since Theorem 2.4.1 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 (2.11) that

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

Hence we have proved that

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

which leads to

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

Next, it is easy to check that $(U_k^N)_{1\le k\le (n+1)N}$ is an $(\mathcal{E}_k^N)_{1\le k\le (n+1)N}$-martingale difference array. In order to apply Theorem 2.3 in [McL74], we just have to check that

- By $\mathcal{A}1$,

$$\max_{1\le k\le (n+1)N} \left|U_k^N\right| \le \frac{2}{\sqrt{N}} \|f\|_\infty \max_{1\le p\le n} \prod_{q=0}^{p-1} \left\|G_{q,\cdot}\right\|_\infty \le \frac{2}{\sqrt{N}} \|f\|_\infty \sum_{p=1}^{n} \prod_{q=0}^{p-1} \left\|G_{q,\cdot}\right\|_\infty, \tag{2.13}$$

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

- From (2.13), we also get that

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

- Standard calculation gives

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

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

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

  Then, by applying Theorem 2.4.1, we obtain

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

Therefore, we have the following central limit theorem

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

Returning to (2.12), the conclusion follows from Slutsky's Lemma.

2.4.3 Proof of Theorem 2.3.1

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

$$\Gamma^b_{n,N}(\phi \otimes \psi) - \Gamma^b_n(\phi \otimes \psi) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

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

$$\Gamma^b_p := \Gamma^{(b_0,\dots,b_p)}_p \qquad \text{and} \qquad \Gamma^b_{p,N} := \Gamma^{(b_0,\dots,b_p)}_{p,N}$$

with the convention

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

Note that, with this convention, we have

$$\Gamma^b_p = \Gamma^b_{p-1} Q_p^{\otimes 2} C_{b_p}.$$

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

$$Q_n^{\otimes 2} C_{b_n}(\phi \otimes \psi) = f \otimes g. \tag{2.14}$$

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 $b_0 = 1$, (2.6) and Definition 2.3.1 give

$$\mathbf{E}\left[\Gamma^b_{0,N}(\phi \otimes \psi)\right] = \mathbf{E}\left[\frac{1}{N}\sum_{i=1}^N \phi(X_0^i)\psi(X_0^i)\right] = \eta_0(\phi\psi) = \Gamma^b_0(\phi \otimes \psi).$$

  Hence, the central limit theorem yields

$$\sqrt{N}\left(\Gamma^b_{0,N}(\phi \otimes \psi) - \Gamma^b_0(\phi \otimes \psi)\right) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \eta_0(\phi^2\psi^2) - \eta_0(\phi\psi)^2\right),$$

  so that

$$\Gamma^b_{0,N}(\phi \otimes \psi) - \Gamma^b_0(\phi \otimes \psi) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

- 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

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

Thanks to (2.1), one has

$$\Gamma^b_{0,N}(\phi\otimes\psi) = \frac{N}{N-1}\left(\eta^N_0(\phi)\eta^N_0(\psi) - \frac{1}{N^2}\sum_{i=1}^N \phi(X^i_0)\psi(X^i_0)\right).$$

Combined with Definition 2.3.1 and the law of large numbers, one deduces that

$$\Gamma^b_{0,N}(\phi\otimes\psi) - \Gamma^b_0(\phi\otimes\psi) = \Gamma^b_{0,N}(\phi\otimes\psi) - \eta_0(\phi)\eta_0(\psi) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$

- 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

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

Next, we consider the following decomposition

$$\begin{aligned}\Gamma^b_{n,N}(\phi\otimes\psi) - \Gamma^b_n(\phi\otimes\psi) =& \underbrace{\Gamma^b_{n,N}(\phi\otimes\psi) - \Gamma^b_{n-1,N}Q^{\otimes 2}_{n,N}C_{b_n}(\phi\otimes\psi)}_{R_1(N)}\\ &+\underbrace{\Gamma^b_{n-1,N}Q^{\otimes 2}_{n,N}C_{b_n}(\phi\otimes\psi) - \Gamma^b_{n-1,N}Q^{\otimes 2}_{n}C_{b_n}(\phi\otimes\psi)}_{R_2(N)}\\ &+\underbrace{\Gamma^b_{n-1,N}Q^{\otimes 2}_{n}C_{b_n}(\phi\otimes\psi) - \Gamma^b_{n-1}Q^{\otimes 2}_{n}C_{b_n}(\phi\otimes\psi).}_{R_3(N)}\end{aligned} \tag{2.15}$$

The tools to terminate the proof are the following ones:

- Lemma 2.4.1 shows that
$$R_1(N) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$
- Lemma 2.4.2 and the fact that one may write $C_{b_n}(\phi\otimes\psi)$ as $f\otimes g$ for any b_n ensure that
$$R_2(N) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$
- Finally, the convergence rate
$$R_3(N) = \mathcal{O}_{\mathrm{p}}\left(\frac{1}{\sqrt{N}}\right).$$
is a direct consequence of (2.14) and the induction hypothesis.

2.4.4 Technical results

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

$$\Gamma^b_{n,N}(1) := \gamma^N_n(1)^2\frac{N^{n-1}}{(N-1)^{n+1}}\sum_{\ell^{[2]}_{0:n}\in((N)^2)^{\times(n+1)}}\left\{\prod_{p=0}^{n-1}\lambda^b_p(A^{\ell^{[2]}_{p+1}}_p, \ell^{[2]}_p)\right\}.$$

If we set

$$\Lambda_n^{\ell_n^{[2]}} := \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{ \prod_{p=0}^{n-1} \lambda_p^b (A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\} \tag{2.16}$$

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

$$\Gamma_{n,N}^b(1) := \gamma_n^N(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell_n^{[2]} \in (N)^2} \Lambda_n^{\ell_n^{[2]}}, \tag{2.17}$$

so that

$$\Gamma_{n,N}^b(1)^2 = \gamma_n^N(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}} \right)^2 \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2}} \Lambda_n^{\ell_n^{[2]}} \Lambda_n^{\ell_n'^{[2]}}. \tag{2.18}$$

Note that (2.17) is still true when $n = 0$. Then, for $n \geq 1$, we have by definition

$$\Lambda_n^{\ell_n^{[2]}} = \sum_{\ell_{n-1}^{[2]} \in (N)^2} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \lambda_{n-1}^b (A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}). \tag{2.19}$$

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

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

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

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

As defined in Section 2.2.2, the IPS associated with ASMC is a Markov chain $(\mathbf{X_n})_{n\geq 0}$ with genealogy $(\mathbf{A_n})_{n\geq 0}$ 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 $(\mathcal{G}_n^N)_{n\geq 0}$ with the genealogy of the IPS is defined by

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

and, for $n \geq 1$,

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

By combining (2.18) and (2.19), and taking into account that

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

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

$$\mathbf{E}\left[\Gamma_{n,N}^b(1)^2 \,\middle|\, \mathcal{G}_{n-1}^N\right] = \gamma_{n-1}^N(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{'[2]})\in((N)^2)^{\times 2}} \Lambda_{n-1}^{\ell_{n-1}^{[2]}}\Lambda_{n-1}^{\ell_{n-1}^{'[2]}}$$
$$\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}} m(\mathbf{X_{n-1}})(G_{n-1,N})^4\, \mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}},\ell_{n-1}^{'[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]. \quad (2.20)$$

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

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

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 (2.20) with respect to the three terms that appear in the right-hand side of the latter. Recall from (2.3) that, given $\mathbf{X_{n-1}}$, we make an independent multinomial selection of the parent of each particle at step n according to the discrete probability measure

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

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

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

We also recall that

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

- 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

$$\begin{aligned}
&\mathbf{E}\left[\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell'^{[2]}_n}_{n-1}, \ell'^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\leq \mathbf{E}\left[\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\quad = \mathbf{1}_{\{b_{n-1}=0\}}\mathbf{P}\left(A^{\ell^1_n}_{n-1} = \ell^1_{n-1}, A^{\ell^2_n}_{n-1} = \ell^2_{n-1} \,\middle|\, \mathcal{G}^N_{n-1}\right) \\
&\qquad + \mathbf{1}_{\{b_{n-1}=1\}}\mathbf{P}\left(A^{\ell^1_n}_{n-1} = \ell^1_{n-1} = A^{\ell^2_n}_{n-1} \,\middle|\, \mathcal{G}^N_{n-1}\right) \\
&\leq \left(\frac{\left\|G_{n-1,\cdot}\right\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^2.
\end{aligned}$$

The analysis for the case where

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

is analogue. Hence, we conclude that

$$\mathbf{E}\left[\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell'^{[2]}_n}_{n-1}, \ell'^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \leq \left(\frac{\left\|G_{n-1,\cdot}\right\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^2. \tag{2.21}$$

Meanwhile, we notice that

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

Putting all things together yields

$$\begin{aligned}
\sum_{(\ell^{[2]}_n, \ell'^{[2]}_n) \in ((N)^2)^{\times 2} \cap [N]^4_2} m(\mathbf{X_{n-1}})(G_{n-1,N})^4\, \mathbf{E}&\left[\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell'^{[2]}_n}_{n-1}, \ell'^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\leq \frac{2(N-1)}{N}\left\|G_{n-1,\cdot}\right\|^4_\infty.
\end{aligned}$$

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

  As noticed in the previous case, the number of different indices within $(\ell^{[2]}_n, \ell'^{[2]}_n)$ is the only thing that matters for the upper-bound in (2.21). Accordingly, the same reasoning gives this time

$$\mathbf{E}\left[\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1}, \ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell'^{[2]}_n}_{n-1}, \ell'^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \leq \left(\frac{\left\|G_{n-1,\cdot}\right\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^3.$$

  Since the total number of choices is

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

it comes

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

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

  This time, we get

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

  and

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

  so that

$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in(N)^4} m(\mathbf{X_{n-1}})(G_{n-1,N})^4\,\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\mathcal{G}_{n-1}^N\right]$$
$$\leq\frac{(N-1)(N-2)(N-3)}{N^3}\left\|G_{n-1,\cdot}\right\|_\infty^4.$$

As a consequence, since

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

an upper-bound for (2.20) is

$$\mathbf{E}\left[\Gamma_{n,N}^b(1)^2\,\middle|\,\mathcal{G}_{n-1}^N\right]$$
$$\leq 7\left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2\frac{N-1}{N}\left\|G_{n-1,\cdot}\right\|_\infty^4\Upsilon_{n-1}^N(1)^4\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}}\Lambda_{n-1}^{\ell_{n-1}^{[2]}}\Lambda_{n-1}^{\ell_{n-1}'^{[2]}}.$$

Replacing n with $(n-1)$ in (2.18) allows us to conclude that

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

Finally, by applying the induction hypothesis, we have

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

which ends the proof of Proposition 2.4.1. □

Lemma 2.4.1. *Under $\mathcal{A}1$, for any test functions $f,g\in\mathcal{B}_b(E_n)$, we have, for all $n\geq 1$,*

$$\mathbf{E}\left[\Gamma_{n,N}^b(f\otimes g)\,\middle|\,\mathcal{G}_{n-1}^N\right]=\Gamma_{n-1,N}^b Q_{n,N}^{\otimes 2}C_{b_n}(f\otimes g),\tag{2.22}$$

as well as

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

Proof. First, by exploiting the notation defined in (2.16), we have

$$\Gamma^b_{n,N}(f\otimes g) := \gamma^N_n(1)^2 \frac{N^{n-1}}{(N-1)^{n+1}} \sum_{\ell^{[2]}_n \in (N)^2} \Lambda^{\ell^{[2]}_n}_n C_{b_n}(f\otimes g)(X^{\ell^{[2]}_n}_n),$$

and (2.22) is then a direct consequence of Proposition 2.5.1 since for any $\ell^{[2]}_n \in (N)^2$

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

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

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

For this, by (2.22), we just have to show that

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

Then, recall that, by definition,

$$\begin{aligned}
&\Gamma^b_{n,N}(f\otimes g)^2 \\
=&\gamma^N_n(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell^{[2]}_n, \ell'^{[2]}_n) \in ((N)^2)^{\times 2}} \Lambda^{\ell^{[2]}_n}_n \Lambda^{\ell'^{[2]}_n}_n \left(C_{b_n}(f\otimes g)\right)^{\otimes 2} (X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n) \\
=&\underbrace{\gamma^N_n(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell^{[2]}_n, \ell'^{[2]}_n) \in (N)^4} \Lambda^{\ell^{[2]}_n}_n \Lambda^{\ell'^{[2]}_n}_n \left(C_{b_n}(f\otimes g)\right)^{\otimes 2} (X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n)}_{R_1(N)} \\
&+\underbrace{\gamma^N_n(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell^{[2]}_n, \ell'^{[2]}_n) \in ((N)^2)^{\times 2} \setminus (N)^4} \Lambda^{\ell^{[2]}_n}_n \Lambda^{\ell'^{[2]}_n}_n \left(C_{b_n}(f\otimes g)\right)^{\otimes 2} (X^{\ell^{[2]}_n}_n, X^{\ell'^{[2]}_n}_n)}_{R_2(N)}.
\end{aligned}$$

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

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

  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 2.5.1 respectively for $\ell^{[2]}_n$ and for $\ell'^{[2]}_n$, we have

$$\begin{aligned}
&\mathbf{E}\left[\gamma^N_n(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \Lambda^{\ell^{[2]}_n}_n \Lambda^{\ell'^{[2]}_n}_n C_{b_n}(f\otimes g)(X^{\ell^{[2]}_n}_n) C_{b_n}(f\otimes g)(X^{\ell'^{[2]}_n}_n) \,\middle|\, \mathcal{G}^N_{n-1}\right] \\
&\qquad = \frac{1}{N^2(N-1)^2} (\Gamma^N_{n-1} Q^{\otimes 2}_{n,N} C_{b_n}(f\otimes g))^2.
\end{aligned}$$

Then, since

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

we deduce that

$$\mathbf{E}\left[R_1(N) - (\Gamma^b_{n-1,N} Q^{\otimes 2}_{n,N} C_{b_n}(f\otimes g))^2 \,\middle|\, \mathcal{G}^N_{n-1}\right] = \left(\frac{N(N-1)(N-2)(N-3)}{N^2(N-1)^2} - 1\right)(\Gamma^N_{n-1} Q^{\otimes 2}_{n,N} C_{b_n}(f\otimes g))^2,$$

and

$$\mathbf{E}\left[R_1(N) - (\Gamma^b_{n-1,N} Q^{\otimes 2}_{n,N} C_{b_n}(f\otimes g))^2\right] = \left(\frac{N(N-1)(N-2)(N-3)}{N^2(N-1)^2} - 1\right)\mathbf{E}\left[(\Gamma^N_{n-1} Q^{\otimes 2}_{n,N} C_{b_n}(f\otimes g))^2\right] = \mathcal{O}\left(\frac{1}{N}\right),$$

where the final equality is due to Proposition 2.4.1, 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 $\Lambda_n^{\ell_n^{[2]}}$ implies

$$\mathbf{E}[R_2(N)] \le \mathbf{E}\left[\gamma_n^N(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_n^{\ell_n^{[2]}}\Lambda_n^{\ell_n'^{[2]}}\right] \|f\|_\infty^2 \|g\|_\infty^2 .$$

So the proof will be finished once we have shown that

$$\mathbf{E}\left[\gamma_n^N(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_n^{\ell_n^{[2]}}\Lambda_n^{\ell_n'^{[2]}}\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

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

$$\frac{1}{N^2(N-1)^2} \sum_{(\ell_0^{[2]},\ell_0'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} 1 = 1 - \frac{N(N-1)(N-2)(N-3)}{N^2(N-1)^2} = \mathcal{O}\left(\frac{1}{N}\right).$$

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

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

The adaptation of (2.20) to the present context gives

$$\mathbf{E}\left[\gamma_n^N(1)^4\left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_n^{\ell_n^{[2]}}\Lambda_n^{\ell_n'^{[2]}} \,\middle|\, \mathcal{G}_{n-1}^N\right]$$
$$= \gamma_{n-1}^N(1)^4\left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}} \Lambda_{n-1}^{\ell_{n-1}^{[2]}}\Lambda_{n-1}^{\ell_{n-1}'^{[2]}}$$
$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} m(\mathbf{X_{n-1}})(G_{n-1,N})^4$$
$$\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right].$$

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

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

- 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 (2.3),

$$\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]$$
$$\leq \left(\frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^2 \left(\mathbf{1}_{\{b_{n-1}=1,\ell_{n-1}^1=\ell_{n-1}'^1\}} + \mathbf{1}_{\{b_{n-1}=0,\ell_{n-1}^1=\ell_{n-1}'^1,\ell_{n-1}^2=\ell_{n-1}'^2\}}\right)$$
$$\leq \left(\frac{\|G_{n-1,\cdot}\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^2 \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}'^1,\ell_{n-1}^2,\ell_{n-1}'^2\}<4\}},$$

  and since

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

  it comes

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

- 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

$$\begin{aligned}
&\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\leq \left(\frac{\left\|G_{n-1,\cdot}\right\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^3 \left(\mathbf{1}_{\{b_{n-1}=1, \ell_{n-1}^1 = \ell_{n-1}'^1\}} + \mathbf{1}_{\{b_{n-1}=0, \ell_{n-1}^1 = \ell_{n-1}'^1\}}\right) \\
&\leq \left(\frac{\left\|G_{n-1,\cdot}\right\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^3 \mathbf{1}_{\left\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\}<4\right\}}.
\end{aligned}$$

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

$$\begin{aligned}
&\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\leq \left(\frac{\left\|G_{n-1,\cdot}\right\|_\infty}{N\, m(\mathbf{X_{n-1}})(G_{n-1,N})}\right)^3 \mathbf{1}_{\left\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\}<4\right\}}.
\end{aligned}$$

In addition, since

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

we get this time

$$\begin{aligned}
&\sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \cap [N]_3^4} m(\mathbf{X_{n-1}})(G_{n-1,N})^4 \\
&\qquad \mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\qquad\qquad \leq \frac{4N(N-1)(N-2)}{N^3} \left\|G_{n-1,\cdot}\right\|_\infty^4 \mathbf{1}_{\left\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\}<4\right\}}.
\end{aligned}$$

By gathering both cases, we have

$$\begin{aligned}
&\mathbf{E}\left[\gamma_n^N(1)^4 \left(\frac{N^{n-1}}{(N-1)^{n+1}}\right)^2 \sum_{(\ell_n^{[2]}, \ell_n'^{[2]}) \in ((N)^2)^{\times 2} \setminus (N)^4} \Lambda_n^{\ell_n^{[2]}} \Lambda_n^{\ell_n'^{[2]}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\leq \gamma_{n-1}^N(1)^4 \left(\frac{N^{n-2}}{(N-1)^n}\right)^2 \frac{N^2}{(N-1)^2} \left(\frac{2N(N-1)}{N^2} + \frac{4N(N-1)(N-2)}{N^3}\right) \left\|G_{n-1,\cdot}\right\|_\infty^4 \\
&\qquad \sum_{(\ell_{n-1}^{[2]}, \ell_{n-1}'^{[2]}) \in ((N)^2)^{\times 2}} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \Lambda_{n-1}^{\ell_{n-1}'^{[2]}} \mathbf{1}_{\left\{\#\{\ell_{n-1}^1, \ell_{n-1}'^1, \ell_{n-1}^2, \ell_{n-1}'^2\}<4\right\}} \\
&\leq 6\frac{N^2}{(N-1)^2} \left\|G_{n-1,\cdot}\right\|_\infty^4 \gamma_{n-1}^N(1)^4 \left(\frac{N^{n-2}}{(N-1)^n}\right)^2 \sum_{(\ell_{n-1}^{[2]}, \ell_{n-1}'^{[2]}) \in ((N)^2)^{\times 2} \setminus (N)^4} \Lambda_{n-1}^{\ell_{n-1}^{[2]}} \Lambda_{n-1}^{\ell_{n-1}'^{[2]}}.
\end{aligned}$$

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

This ends the proof of Lemma 2.4.1. □

Lemma 2.4.2. *Assume $\mathcal{A}1$-$\mathcal{A}2$, then for any test functions $f, g \in \mathcal{B}_b(E_n)$ and for all $n \geq 1$,*

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

Proof. The verification shares some resemblance with the convergence of $P_2(N)$ in the proof of Theorem 2.4.1. Specifically, we start with the following decomposition:

$$\begin{aligned}
&\left|\Gamma^b_{n-1,N} Q^{\otimes 2}_{n,N}(f \otimes g) - \Gamma^b_{n-1,N} Q^{\otimes 2}_{n}(f \otimes g)\right| \\
&\qquad \leq \underbrace{\left|\Gamma^b_{n-1,N} Q^{\otimes 2}_{n,N}(f \otimes g) - \Gamma^b_{n-1,N}(Q_{n,N} \otimes Q_n)(f \otimes g)\right|}_{D_1(N)} \\
&\qquad\qquad + \underbrace{\left|\Gamma^b_{n-1,N}(Q_{n,N} \otimes Q_n)(f \otimes g) - \Gamma^b_{n-1,N} Q^{\otimes 2}_{n}(f \otimes g)\right|}_{D_2(N)}.
\end{aligned}$$

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

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

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 2.4.1, one has

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

In addition, a by-product (2.11) of Theorem 2.2.1 is that

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

Hence, one concludes that

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

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

2.4.5 Connection between the estimators

In this section, we give some combinatorial results on the coalescent tree-based measures $\Gamma^b_{n,N}$. In particular, these results allow us to connect the variance estimator (2.9) of Lee & Whiteley and our term by term estimators. As mentioned before, these results do not depend on $\mathcal{A}2$: they are provided by the structure of the IPS and the underlying multinomial selection scheme.

Proposition 2.4.2. *Under $\mathcal{A}1$, for any test function $F \in \mathcal{B}_b(E_n^2)$, we have the decompositions:*

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

and

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

Proof. Since

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

we have

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

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

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

To conclude, one just has to observe that

$$\sum_{\ell^{[2]}_0\in(N)^2} \cdots \sum_{\ell^{[2]}_{n-1}\in(N)^2} \left\{ \prod_{p=0}^{n-1} \left(\frac{1}{N} \mathbf{1}_{\{A_p^{\ell^1_{p+1}} = A_p^{\ell^2_{p+1}} = \ell^1_p \neq \ell^2_p\}} + \frac{N-1}{N} \mathbf{1}_{\{A_p^{\ell^1_{p+1}} = \ell^1_p \neq A_p^{\ell^2_{p+1}} = \ell^2_p\}} \right) \right\} = \left(\frac{N-1}{N} \right)^n,$$

while, by (2.1),

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

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

We can now proceed with the proof of Proposition 2.3.1. Recall that the goal is to show that

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

and

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

By construction, we have

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

An implication of Proposition 2.4.1 is that, for any test function f and any coalescence indicator b,

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

Thus, a consequence of Proposition 2.4.2 is

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

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

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

the same algebraic manipulation yields

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

This closes the proof of Proposition 2.3.1.

2.5 Many-body Feynman-Kac models

The many-body Feynman-Kac model was proposed in [DMKP16] to study the propagation of chaos property of the Conditional Particle Markov Chain Monte Carlo introduced in [ADH10]. 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.

2.5.1 Duality formula

At each layer, the particles in the original IPS are denoted by $\mathbf{X_p}$, with its genealogy $\mathbf{A_{p-1}}$. 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 $\boldsymbol{x_p}$, $\boldsymbol{a_{p-1}}$, $\tilde{x}_p^{[q]}$ 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:

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

As for the basic idea of Particle Markov Chain Monte Carlo method [ADH10], we study respectively the distributions of

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

and

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

Thanks to the specific construction, as well as the relatively simple multinomial resampling scheme of Feynman-Kac IPS, Lemma 2.5.1 provides a duality formula to connect both distributions and leads in particular to Proposition 2.5.1. This latter result is crucial to prove the consistency of our term by term estimator in Theorem 2.3.1.

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 2.2.2 can be expressed as

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

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

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

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

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

Then we have the following pivotal duality formula, which is simply a generalization of Lemma 4.1 in [DMKP16]. We will apply it in the proof of Proposition 2.5.1 with $q = 2$.

Lemma 2.5.1. *For $p \geq 1$, $q \in [N]$ and $\ell_p^{[q]} \in (N)^q$, we have the following duality formula between integral operators*

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

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

$$\mathcal{G}_{p-1}^{(q)}(\boldsymbol{x_{p-1}}) \left\{\sum_{k=1}^n \frac{G_{p-1,N}(x_{p-1}^k)}{N\, m(\boldsymbol{x_{p-1}})(G_{p-1,N})}\delta_k\right\}^{\otimes q}(da_{p-1}^{\ell_p^{[q]}})\, M_{p,N}^{\otimes q}(x_{p-1}^{a_{p-1}^{\ell_p^{[q]}}}, dx_p^{\ell_p^{[q]}})$$
$$\mathbf{M}_p^{\ell_p^{[q]}}\left((\boldsymbol{a_{p-1}}, \boldsymbol{x_p}), d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]})\right) = \mathbb{Q}_p^{(q)}(\boldsymbol{x_{p-1}}, d(\tilde{a}_{p-1}^{[q]}, \tilde{x}_p^{[q]}))\, \delta_{\tilde{x}_p^{[q]}}(dx_p^{\ell_p^{[q]}})\, \delta_{\tilde{a}_{p-1}^{[q]}}(da_{p-1}^{\ell_p^{[q]}}).$$

Fixing $\ell_p^{[q]} \in (N)^q$ and $\boldsymbol{x_{p-1}} \in E_{p-1}^N$, consider a function $\mathbf{F} \in \mathcal{B}_b([N]^q \times [N]^q \times E_p^q \times E_p^q)$. 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

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

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

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

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

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

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

This ends the proof of the duality formula. □

Let us recall (2.16) and (2.19):

$$\Lambda_n^{\ell_n^{[2]}} = \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{ \prod_{p=0}^{n-1} \lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\} = \sum_{\ell_{n-1}^{[2]} \in (N)^2} \Lambda_{n-1}^{\ell_{n-1}^{[2]}}\ \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}). \tag{2.23}$$

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

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

The upcoming result is useful in the proof of Lemma 2.4.1.

Proposition 2.5.1. *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*

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

Proof. By applying (2.23), 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

$$\begin{aligned}&\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1,N})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\\&=\frac{1}{N^2}C_{b_{n-1}}Q_{n,N}^{\otimes 2}C_{b_n}(f\otimes g)(X_{n-1}^{\ell_{n-1}^{[2]}}).\end{aligned}\tag{2.24}$$

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 $(\mathbf{A_{n-1}},\mathbf{X_n})$, 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 $\mathbf{Q}_p^{(2)}(\boldsymbol{x_{p-1}},d(\boldsymbol{a_{p-1}},\boldsymbol{x_p}))$,

$$\begin{aligned}&\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1,N})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(X_n^{\ell_n^{[2]}})F(\mathbf{A_{0:n-2}},\mathbf{X_{0:n-1}})\right]\\&=\int\mathbf{Q}_n^{(2)}(\boldsymbol{x_{n-1}},d(\boldsymbol{a_{n-1}},\boldsymbol{x_n}))\lambda_{n-1}(a_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(x_n^{\ell_n^{[2]}})\\&\qquad F(\boldsymbol{a_{0:n-2}},\boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}},d\boldsymbol{x_{0:n-1}}),\end{aligned}\tag{2.25}$$

where μ_{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

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

is a Markov kernel, we can introduce it in the right-hand side of (2.25) to obtain

$$\begin{aligned}&\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1,N})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(X_n^{\ell_n^{[2]}})F(\boldsymbol{a_{0:n-2}},\boldsymbol{x_{0:n-1}})\right]\\&=\int\mathbf{Q}_n^{(2)}(\boldsymbol{x_{n-1}},d(\boldsymbol{a_{n-1}},\boldsymbol{x_n}))\mathbf{M}_n^{\ell_n^{[2]}}\left((\boldsymbol{a_{n-1}},\boldsymbol{x_n}),d(\tilde{a}_{n-1}^{[2]},\tilde{x}_n^{[2]})\right)\\&\qquad\lambda_{n-1}(a_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(x_n^{\ell_n^{[2]}})F(\boldsymbol{a_{0:n-2}},\boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}},d\boldsymbol{x_{0:n-1}}).\end{aligned}$$

The design of many-body Feynman-Kac models allows us to replace $(a_{n-1}^{\ell_n^{[2]}},x_n^{\ell_n^{[2]}})$ 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:

$$\begin{aligned}&\int\mathbf{Q}_n^{(2)}(\boldsymbol{x_{n-1}},d(\boldsymbol{a_{n-1}},\boldsymbol{x_n}))\mathbf{M}_n^{\ell_n^{[2]}}\left((\boldsymbol{a_{n-1}},\boldsymbol{x_n}),d(\tilde{a}_{n-1}^{[2]},\tilde{x}_n^{[2]})\right)\\&\qquad\lambda_{n-1}(a_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(x_n^{\ell_n^{[2]}})F(\boldsymbol{a_{0:n-2}},\boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}},d\boldsymbol{x_{0:n-1}})\\&=\int\mathbf{Q}_n^{(2)}(\boldsymbol{x_{n-1}},d(\boldsymbol{a_{n-1}},\boldsymbol{x_n}))\mathbf{M}_n^{\ell_n^{[2]}}\left((\boldsymbol{a_{n-1}},\boldsymbol{x_n}),d(\tilde{a}_{n-1}^{[2]},\tilde{x}_n^{[2]})\right)\\&\qquad\lambda_{n-1}(\tilde{a}_{n-1}^{[2]},\ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]})F(\boldsymbol{a_{0:n-2}},\boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}},d\boldsymbol{x_{0:n-1}}).\end{aligned}$$

Now, the duality formula given in Lemma 2.5.1 yields

$$\begin{aligned}
&\int \mathbf{Q}_n^{(2)}(\boldsymbol{x_{n-1}}, d(\boldsymbol{a_{n-1}}, \boldsymbol{x_n}))\mathbf{M}_n^{\ell_n^{[2]}}\left((\boldsymbol{a_{n-1}}, \boldsymbol{x_n}), d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]})\right) \\
&\quad \lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]})F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}) \\
&= \int \mathbb{Q}_n^{(2)}(\boldsymbol{x_{n-1}}, d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]}))\ \mathbb{M}_n^{\tilde{a}_{n-1}^{[2]}, \ell_n^{[2]}, \tilde{x}_n^{[2]}}(\boldsymbol{x_{n-1}}, d(\boldsymbol{a_{n-1}}, \boldsymbol{x_n})) \\
&\quad \lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]})F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}).
\end{aligned}$$

In addition, since

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

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

$$\begin{aligned}
&\int \mathbb{Q}_n^{(2)}(\boldsymbol{x_{n-1}}, d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]}))\ \mathbb{M}_n^{\tilde{a}_{n-1}^{[2]}, \ell_n^{[2]}, \tilde{x}_n^{[2]}}(\boldsymbol{x_{n-1}}, d(\boldsymbol{a_{n-1}}, \boldsymbol{x_n})) \\
&\quad \lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]})F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}) \\
&= \int \mathbb{Q}_n^{(2)}(\boldsymbol{x_{n-1}}, d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]}))\ \lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]}) \\
&\quad F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}).
\end{aligned}$$

Next, let us recall that

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

and

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

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

$$\begin{aligned}
&\int \mathbb{Q}_n^{(2)}(\boldsymbol{x_{n-1}}, d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]}))\ \lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]}) \\
&\quad F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}) \qquad (2.26) \\
&= \int \frac{1}{N^2}Q_{n,N}^{\otimes 2}C_{b_n}(f\otimes g)(x_{n-1}^{\ell_{n-1}^{[2]}})F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}).
\end{aligned}$$

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

$$\begin{aligned}
&\int \mathbb{Q}_n^{(2)}(\boldsymbol{x_{n-1}}, d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]}))\ \lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]}) \\
&\quad F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}) \qquad (2.27) \\
&= \int \frac{1}{N^2}Q_{n,N}^{\otimes 2}C_{b_n}(f\otimes g)(x_{n-1}^{\ell_{n-1}^{(1,1)}})F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}).
\end{aligned}$$

Combining (2.26) and (2.27), we safely deduce that

$$\begin{aligned}
&\int \mathbb{Q}_n^{(2)}(\boldsymbol{x_{n-1}}, d(\tilde{a}_{n-1}^{[2]}, \tilde{x}_n^{[2]}))\ \lambda_{n-1}(\tilde{a}_{n-1}^{[2]}, \ell_{n-1}^{[2]})C_{b_n}(f\otimes g)(\tilde{x}_n^{[2]}) \\
&\quad F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}) \\
&= \int \frac{1}{N^2}C_{b_{n-1}}Q_{n,N}^{\otimes 2}C_{b_n}(f\otimes g)(x_{n-1}^{\ell_{n-1}^{[2]}})F(\boldsymbol{a_{0:n-2}}, \boldsymbol{x_{0:n-1}})\mu_{n-1}(d\boldsymbol{a_{0:n-2}}, d\boldsymbol{x_{0:n-1}}) \\
&= \mathbf{E}\left[\frac{1}{N^2}C_{b_{n-1}}Q_{n,N}^{\otimes 2}C_{b_n}(f\otimes g)(X_{n-1}^{\ell_{n-1}^{[2]}})F(\mathbf{A_{0:n-2}}, \mathbf{X_{0:n-1}})\right].
\end{aligned}$$

In conclusion, we have established that

$$\begin{aligned}&\mathbf{E}\left[m(\mathbf{X_{n-1}})(G_{n-1,N})^2\lambda^b_{n-1}(A^{\ell^{[2]}_n}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(f\otimes g)(X^{\ell^{[2]}_n}_n)F(\mathbf{A_{0:n-2}},\mathbf{X_{0:n-1}})\right]\\ =&\mathbf{E}\left[\frac{1}{N^2}C_{b_{n-1}}Q^{\otimes 2}_{n,N}C_{b_n}(f\otimes g)(X^{\ell^{[2]}_{n-1}}_{n-1})F(\mathbf{A_{0:n-2}},\mathbf{X_{0:n-1}})\right],\end{aligned}$$

which terminates the verification of (2.24) and the proof of Proposition 2.5.1.

□

2.5.2 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 2.5.1 recursively applied in a mini IPS from level 0 to level 5 to some randomly chosen indices $\ell^{[2]}_{0:5}$ (see Figure 2.5).

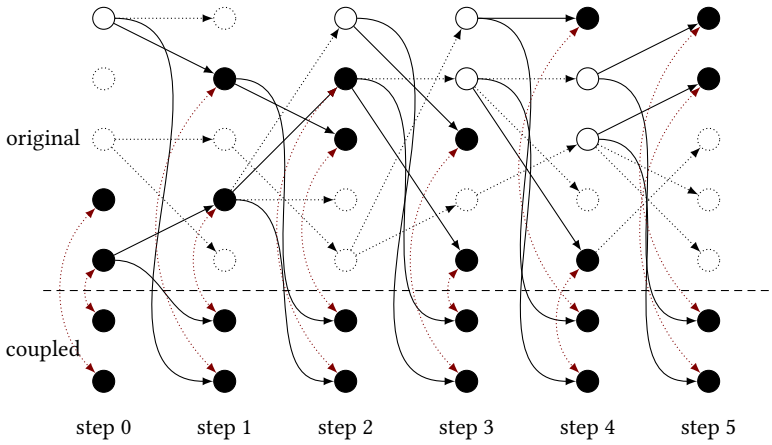

Figure 2.5: An illustration of the duality formula recursively applied to a mini IPS of $n+1=6$ levels with 5 particles at each level. Every straight black or dotted arrow within the original IPS represents a Markov transition $M_{p,N}$ and the black twisted ones pointing to the particles in the coupled 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 $\ell^{[2]}_0=(4,5)$, $\ell^{[2]}_1=(2,4)$, $\ell^{[2]}_2=(2,3)$, $\ell^{[2]}_3=(3,5)$, $\ell^{[2]}_4=(1,5)$ and $\ell^{[2]}_5=(1,2)$.

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

$$\left\{\ell^{[2]}_{p-2}=\tilde{A}^{[2]}_{p-2},\ell^1_{p-1}=\tilde{A}^1_{p-1}=\tilde{A}^2_{p-1}\neq\ell^2_{p-1},\ell^{[2]}_p=\tilde{A}^{[2]}_p\right\}. \tag{2.28}$$

On this event, we are able to track the coalescent tree-based form as in Figure 2.6.

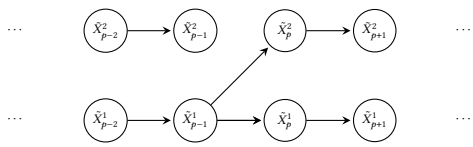

Figure 2.6: The coupled particle block tracked by the event defined by (2.28).

The coupled 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 2.7).

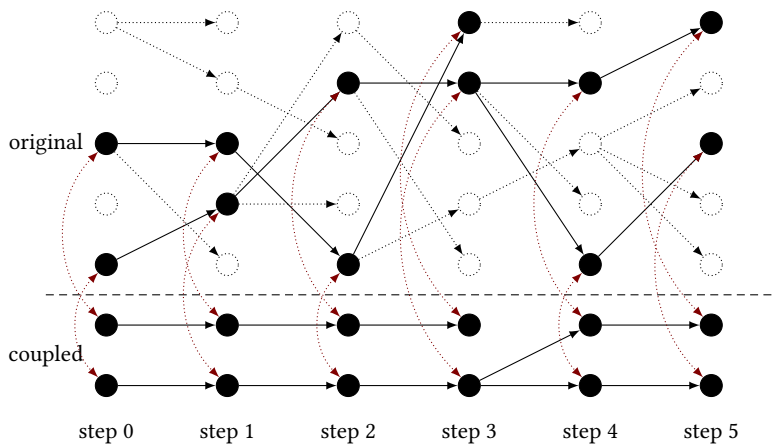

Figure 2.7: An illustration of the duality formula recursively applied to a mini IPS of $n+1=6$ levels with 5 particles at each level. Every straight black or dotted arrow within the original IPS represents a Markov transition $M_{p,N}$ and the black ones within the coupled particle block represent the 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 $\ell_0^{[2]}=(3,5)$, $\ell_1^{[2]}=(3,4)$, $\ell_2^{[2]}=(2,5)$, $\ell_3^{[2]}=(1,2)$, $\ell_4^{[2]}=(2,5)$ and $\ell_5^{[2]}=(1,3)$.

The duality formula provides a way to touch the adaptive versions of the coalescent tree-based measures Γ_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 $\Gamma_{n,N}^b$.

2.5.3 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 2.4.2. On the opposite, in a nonadaptive case (SMC), the estimation is unbiased, exactly as γ_n^N is an unbiased estimation of γ_n (see for example [DM04] 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 2.4.1 and (2.15) give the following proposition.

Proposition 2.5.2. *Assume $\mathcal{A}1$-$\mathcal{A}2$ and suppose that $h_n \equiv 0$ for all $n \geq 0$. Then, for all test functions $f, g \in \mathcal{B}_b(E_n)$,*

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

In particular, we also have

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

In fact, the essential technical results in Section 2.4.4 and Section 2.4.5 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 chapter.

Another remark is about the difference between $\Gamma^b_{n,N}$ and μ_b as defined in Section 3.2. of [LW18] 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 2.3 in Section 2.3.1, our estimator is not the most "precise" that one could propose. Let us look at the case where

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

For the terminal point X^4_3, the conditional distribution of A^4_2 is simply the categorical distribution since X^4_3 is a terminal point. Roughly speaking, once all the genealogy of the terminal points is calculated, one can deduce μ_b. Hence, the take-home message is simple: if one is interested in estimating Γ^b_n numerically, then the estimator μ_b proposed in [LW18] 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 μ_b. This also simplifies the analysis in an adaptive context where there is already more randomness than in a nonadaptive context.

Chapter 3

Asymmetric Sequential Monte Carlo

abstract: Sequential Monte Carlo is a general framework aiming at sampling a sequence of measures $(\eta_n; n \geq 0)$ connected by some nonlinear operators. In the classical setting, the simulation consists in a multinomial resampling selection step and a Markov mutation step at each iteration of the algorithm. When the potential functions are $[0, 1]$-valued, a well-known variant is to conduct a Bernoulli *survival test* before the multinomial resampling step: the surviving particles will not be resampled, whilst the non-survived particles perform a multinomial resampling. We go one step further, that is, we suppose that the surviving particles and non-surviving particles will mutate according to different Markov kernels. This setting is referred to as "asymmetric resampling" in this chapter. The idea is natural in rare-event simulation, where the Markov kernel at step n is η_n-invariant. In this scenario, since the surviving particles have already "well placed", there is no need to implement another mutation. The idea is to reduce the computational burden induced by the mutation kernels. We provide a CLT-type result as well as consistent variance estimators, which allows to conduct statistical inference with a single run of the simulation. We also provide an unbiased variance estimator for the unnormalized measures under certain conditions. To do this, we introduce generalized coalescent tree-based measures and their particle approximations as a complement of the ones introduced respectively in [CDMG11] and [DG19](Chapter 2). We firmly believe that they represent an important and natural family of mathematical objects in the general framework of SMC. We expect the same methodology may also inspire further analysis for the models in a continuous-time setting, such as Fleming-Viot particle systems (see, e.g., [DCGR17]).

3.1 Introduction

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

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

In contrast to the resampling strategy discussed above, we are interested in the case where there are two different Markov kernels, denoted respectively by $\overset{\circ}{M}_n$ and $\overset{\bullet}{M}_n$, at each iteration of the algorithm. In the rest of this chapter, they are referred to as *mutation kernels*. Roughly speaking, at each iteration, say, from level $n-1$ to level n, each particle performs a Bernoulli *survival test* w.r.t. the $[0,1]$-valued potential function G_{n-1}: the survived particle mutates according to $\overset{\bullet}{M}_n$ while the non-surviving ones execute a multinomial resampling w.r.t. the potential function G_{n-1}, after which a mutation according to the kernel $\overset{\circ}{M}_n$ will be executed. The precise mathematical definition will be given in Section 3.2.4.

The main motivating example is the generalized Adaptive Multilevel Splitting (gAMS) methods introduced in [BGG+16], where the kernel $\overset{\bullet}{M}_n$ is set to be the identity $\delta_x(dy)$. It is a natural choice in the applications such as Particle Tempering and other Subset Simulation methods, where $\overset{\circ}{M}_n$ is designed to be an η_n-invariant kernel. The primal motivation of this implementation is to reduce the unnecessary computational costs brought by the mutation kernel $\overset{\circ}{M}_n$. Moreover, since the invention of Particle Markov Chain Monte Carlo methods (PMCMC, cf. [ADH10]), the design of mutation kernels $\overset{\circ}{M}_n$ becomes much easier and more computationally demanding at the same time. One typical example is SMC^2 methods (see, e.g., [CJP13, CRGP15]). The basic idea is to use another SMC-based IPS and freezing techniques to construct an η_n-invariant kernel at each level. One can imagine that the computational costs are mainly from the implementation of the mutation kernels $(\overset{\circ}{M}_n; n \geq 1)$, which can be dramatically reduced by applying asymmetric resampling scheme if we choose $\overset{\bullet}{M}_n$ to be the identity or some other "cheaper" kernels.

Another motivation is the variance estimation in the context of symmetric sampling, namely, the case where all the particles mutate according to the same Markov kernel at each step. More precisely, it means that $\overset{\bullet}{M}_n \equiv \overset{\circ}{M}_n$ for each $n \geq 1$ and this setting enters into the classic Feynman-Kac particle models (e.g. [DM04]). It is well-known that the asymptotic variance is smaller than the one under classical multinomial resampling scheme. The exact difference at each step is presented in (3.20). As the symmetric resampling can also slightly reduce the computational costs required by the multinomial resampling, there is no practical reason to implement multinomial resampling scheme if an upper bound for G_n is available.

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

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

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

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

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

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

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

3.1.1 Main results

On one hand, in a very general setting, we provide consistent estimations for the target measures in the SMC context with controllable asymptotic uncertainty under our specific asymmetric resampling scheme. Since the computations of the variance estimators are highly nontrivial, we provide detailed and efficient algorithms with time and space complexity analysis in Section 3.5. If there is any ambiguity w.r.t. the notation, the reader is referred to Section 3.1.2.

Theorem 3.1.1. *Let $(E_n; n \leq 0)$ be a sequence of Polish state spaces. Given a sequence of $[0,1]$-valued potential functions $(G_n; n \geq 0)$ and a canonical Markov chain $(X_n; n \geq 0)$ taking values in $(E_n; n \geq 0)$, with initial distribution η_0 and transition kernels $(\mathring{M}_n; n \geq 1)$, we define the family of measures $(\gamma_n; n \geq 0)$ by*

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

Assuming that $\gamma_n(1) > 0$ for any $n \geq 0$, we also define $\eta_n(f) := \gamma_n(f)/\gamma_n(1)$. For any test function $f \in \mathcal{B}_b(E_n)$, when the number of particle N tends to infinity, the estimators given by Algorithm 2 in Section 3.5, denoted respectively by $\gamma_n^N(f)$ and $\eta_n^N(f)$, converge almost surely to $\gamma_n(f)$ and $\eta_n(f)$ if for any $n \geq 1$, we have

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

Moreover, we also have

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

as well as

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

where the computation of $\hat{\sigma}^2_{\gamma_n^N}(f)$ and $\hat{\sigma}^2_{\eta_n^N}(f - \eta_n^N(f))$ are respectively provided in Algorithm 5 and Algorithm 6 in Section 3.5.

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

Theorem 3.1.2. *Assume the same setting as in Theorem 3.1.1. Under the condition discussed in Section 3.4.1, which at least contains the case where $\dot{M}_n \equiv \mathring{M}_n$ for any $n \geq 1$, the estimator $\gamma_n^N(f)$ is an unbiased estimator for $\gamma_n(f)$. Moreover, the estimator provided by Algorithm 7 in Section 3.5 is an unbiased estimator for the non-asymptotic variance of $\gamma_n^N(f)$.*

3.1.2 Notation

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

- The underlying probability space is denoted by $(\Omega, \mathcal{F}, \mathbf{P})$. For σ-fields $\mathcal{E}, \mathcal{G} \subset \mathcal{F}$, $\mathcal{E} \vee \mathcal{G}$ denotes the smallest σ-field on Ω containing $\mathcal{E}$ and $\mathcal{G}$. For any $x, y \in \mathbf{R}$, we denote $x \wedge y := \min\{x, y\}$ and $x \vee y := \max\{x, y\}$. We also adopt the standard convention $\inf \varnothing = \infty$.

- Let X be a number, a function or a random variable. We adopt the following convention:

$$\frac{1}{\mathrm{X}}\mathbf{1}_{\mathrm{X}\neq 0} := \begin{cases} \frac{1}{\mathrm{X}} & \text{if} \quad \mathrm{X} \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

  Therefore, we admit the calculation

$$\mathrm{X} \times \frac{1}{\mathrm{X}}\mathbf{1}_{\mathrm{X}\neq 0} = \mathbf{1}_{\mathrm{X}\neq 0}.$$

- Random variables take values in Polish spaces, i.e., a topological space E which is metrizable, separable and complete for some distance d_E. It is endowed with the Borel σ-algebra generated by d_E, denoted by $\mathcal{B}(E)$.

- We denote respectively by $\mathcal{M}(E)$, $\mathcal{M}_+(E)$ and $\mathcal{P}(E)$ the set of all signed finite measures, the subset of all nonnegative finite measures and the subset of all probability measures on $(E, \mathcal{B}(E))$. The set $\mathcal{P}(E)$ is endowed with the Prohorov-Lévy metric, i.e., the weak convergence "$\xrightarrow{\mathrm{d}}$" is tested with continuous bounded functions.

- $\mathcal{B}_b(E)$ denotes the collection of all the bounded measurable functions from $(E, \mathcal{B}(E))$ to $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ equipped with uniform norm $\|\cdot\|_\infty$, among which the constant function will be denoted by 1 with a slight abuse of notation. Given a probability measure η in $\mathcal{P}(E)$ and for all test functions in $\mathcal{B}_b(E)$, we denote η-ess $\sup(f)$ the essential supremum of f. It is defined by

$$\eta\text{-ess}\sup(f) := \inf\left\{a \in \mathbf{R} : \eta(x \in E : f(x) > a) = 0\right\}.$$

- For all $\mu \in \mathcal{M}(E)$ and for all test functions $f \in \mathcal{B}_b(E)$, $\mu(f)$ denotes the integration

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

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

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

  More precisely, for all $x \in E$, $Q(x, \cdot)$ is a finite nonnegative measure in $\mathcal{M}_+(F)$ and for all $A \in \mathcal{B}(F)$, $x \mapsto Q(x, A)$ is a $\mathcal{B}(E)$-measurable function. We say that Q is a Markov transition kernel if Q is a finite transition kernel and for all $x \in E$, $Q(x, \cdot)$ is a probability measure in $\mathcal{P}(F)$. For a signed measure $\mu \in \mathcal{M}(E)$ and a test function $f \in \mathcal{B}_b(F)$, we denote respectively $\mu Q \in \mathcal{M}(E)$ and $Qf \in \mathcal{B}_b(E)$ are respectively defined as follows:

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

  and

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

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

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

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

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

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

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

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

In particular, we denote

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

Accordingly, we denote

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

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

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

Similarly, we also denote

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

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

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

and

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

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

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

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

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

We denote

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

and

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

A straightforward computation shows that

$$m^{\otimes 2}(\boldsymbol{x}) = \frac{N-1}{N} m^{\odot 2}(\boldsymbol{x}) C_0 + \frac{1}{N} m^{\odot 2}(\boldsymbol{x}) C_1. \tag{3.2}$$

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

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

as well as

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

3.2 SMC framework

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

3.2.1 Setting

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

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

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

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

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

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

We adopt the convention

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

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

$$\forall n \geq 0, \quad \gamma_n = \left\{ \prod_{p=0}^{n-1} \eta_p(G_p) \right\} \eta_n. \tag{3.3}$$

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

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

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

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

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

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

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

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

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

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

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

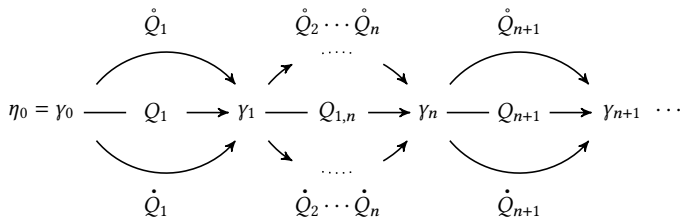

Figure 3.1: Illustration of the Feynman-Kac measures flow.

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

Definition 3.2.1. *We introduce the asymmetric McKean kernel $K_{n,\mu}$ from E_{n-1} to E_n, parameterized by some positive finite measure $\mu \in \mathcal{M}_+(E_{n-1})$ such that $\mu(G_{n-1}) > 0$, defined as follows*

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

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

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

with the convention

$$\forall x \in E_0, \quad Q_{0,\mu}(x, A) := \eta_0(A). \tag{3.4}$$

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

$$\eta_{n-1}K_{n,\eta_{n-1}} = \eta_n \quad \text{and} \quad \gamma_{n-1}Q_{n,\eta_{n-1}} = \gamma_n. \tag{3.5}$$

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

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

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

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

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

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

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

3.2.2 Interacting particle system

The Interacting Particle System (IPS) in this chapter refers to a Markov chain $(\mathbf{X_n}; n \geq 0)$ with absorption in the product spaces $(E_n^N, \mathcal{B}(E_n)^{\otimes N}; n \geq 0)$. As we have seen in the previous section (3.5), the normalized Feynman-Kac measures η_n and η_{n+1} are connected by K_{n+1,η_n}, which depends on the measure of the previous step η_n. Hence, it is not possible to simulate directly according to the kernel K_{n+1,η_n}. The idea of the IPS is to simulate N particles $\mathbf{X_n} = (X_n^1, X_n^2, \ldots, X_n^N)$ step by step. Therefore, by exploiting the empirical measure $m(\mathbf{X_n})$ to approximate its "limiting" measure η_n, we are able to simulate the next layer of particles $\mathbf{X_{n+1}}$ with the approximated kernel $K_{n+1,m(\mathbf{X_n})}$. In this section, we deal with the version without the genealogy (i.e., the indices of the parent of each particle) and the survival history of IPS. The mechanism of the IPS is defined as follows:

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

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

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

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

A more detailed explanation on the algorithm can be found in Section 3.2.4. The particle approximation of the normalized measure η_n is defined by

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

According to (3.3), the unnormalized version γ_n^N is defined by

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

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

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

3.2.3 Asymptotic results

In this section, we establish some basic convergence results such as law of large numbers and central limit theorem for the empirical Feynman-Kac measures. These results are standard in the case where $\dot{Q}_n \equiv \mathring{Q}_n$ (see, e.g., Chapter 7 of [DM04]) and the proofs are housed respectively in Section 3.7.3 and Section 3.7.4. The goal is to understand the consequences of the introduction of $\dot{Q}_n$, especially on the form of the asymptotic variances.

Theorem 3.2.1. *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

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

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

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

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

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

Theorem 3.2.2. *For any test function* $f \in \mathcal{B}_b(E_n)$, *we have*

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

with the asymptotic variance defined as follows:

$$\sigma^2_{\gamma_n}(f) := \sum_{p=0}^{n}\left(\gamma_p^{\otimes 2} C_1 Q_{p,n}^{\otimes 2}(f^{\otimes 2}) - \gamma_{p-1}^{\otimes 2} C_1 Q_{p,\eta_{p-1}}^{\otimes 2} Q_{p,n}^{\otimes 2}(f^{\otimes 2})\right). \tag{3.6}$$

Similarly, we also have

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

with $\sigma^2_{\eta_n}$ *defined by*

$$\forall \varphi \in \mathcal{B}_b(E_n), \quad \sigma^2_{\eta_n}(\varphi) := \sigma^2_{\gamma_n}(\varphi)/\gamma_n(1)^2. \tag{3.7}$$

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

3.2.4 Genealogy and survival history

In this section, we give a more detailed version of the IPS defined in Section 3.2.2, namely, the actual simulation algorithm we execute in practice. Specifically, we trace two kinds of information: the genealogy $\mathbf{A_n} = (A_n^1, \ldots, A_n^N) \in [N]^N$ and the so-called survival history $\mathbf{B_n} = (B_n^1, \ldots, B_n^N) \in \{0,1\}^N$. They are both intermediate random variables introduced in the real-world algorithm, so that one can simulate according to an approximated kernel $K_{n+1,m(\mathbf{X_n})}$. Note that

$$A_n^i = j$$

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

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

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

(i) Initial distribution:

At step 0, we let $\mathbf{X_0} \sim \eta_0^{\otimes N}$.

(ii) Stopping criterion:

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

(iii) Transition kernels:

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

- Survival test: Let B_n^i be a Bernoulli random variable with probability $G_n(X_n^i)$, that is

$$B_n^i \sim G_n(X_n^i)\delta_1 + (1 - G_n(X_n^i))\delta_0.$$

- Selection: If $B_n^i = 1$, we let $A_n^i = i$. Otherwise, the parent index is selected by the following multinomial selection

$$A_n^i \sim \sum_{k=1}^{N} \frac{G_n(X_n^k)}{\sum_{j=1}^{N} G_n(X_n^j)} \delta_k.$$

Therefore, given $B_n^i = \beta_n^i$, we have

$$A_n^i \sim \beta_n^i \delta_i + (1 - \beta_n^i) \sum_{k=1}^{N} \frac{G_n(X_n^k)}{\sum_{j=1}^{N} G_n(X_n^j)} \delta_k.$$

- Mutation: Given $B_n^i = \beta_n^i$ and $A_n^i = a_n^i$, each particle X_n^i evolves independently from level n to level $n+1$ according to the following transition kernel:

$$X_{n+1}^i \sim \beta_n^i \dot{M}_{n+1}(X_n^i, \cdot) + (1 - \beta_n^i) \mathring{M}_{n+1}(X_n^{a_n^i}, \cdot).$$

3.3 Variance estimations

In this section, we provide estimators for the asymptotic variances $\sigma^2_{\gamma_n}(f)$ and $\sigma^2_{\eta_n}(f)$: we provide a term by term asymptotic variance estimator, an unbiased variance estimator under symmetric resampling scheme, and finally, an efficient asymptotic variance estimator. The strategy is almost identical as in [DG19](Chapter 2). First, we give an alternative representation of the asymptotic variance $\sigma^2_{\gamma_n}(f)$ using some generalized coalescent tree-based measures. Next, we provide convergence results of the particle approximations of these generalized coalescent tree-based measures, which gives naturally a term by term variance estimator. Finally, we connect this term by term estimator to the non-asymptotic variance using a nontrivial combinatorial property of the IPS given in Theorem 3.6.1, from which we derive an efficient variance estimator that can be computed with the optimal $\mathcal{O}(nN)$ time complexity.

3.3.1 Asymptotic variance expansion

In this section, we revisit the asymptotic variance $\sigma^2_{\gamma_n}(f)$ of Theorem 3.2.2 using some novel coalescent tree-based measures. More precisely, unlike the multinomial case, the form of the asymptotic variance $\sigma^2_{\gamma_n}(f)$ is relatively complex under asymmetric resampling and there is no

free coalescent tree-based expansion as in [DG19](Chapter 2). Hence, we need to introduce some generalized coalescent tree-based measures as a supplement of the one introduced in [CDMG11]. The goal is plain and simple: we want to establish an alternative representation of the asymptotic variance based on some coalescent tree-based measures. To begin, let us define the so-called *coalescent Feynman-Kac kernels*:

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

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

The full description of this new family of kernels can be found in Section 3.6.2. Using the partial semigroup property of the coalescent Feynman-Kac kernels defined above, we introduce some generalized coalescent tree-based measures. They will be referred to as *coalescent Feynman-Kac measures* in this chapter. In the next definition, we call $b := (b_0, \ldots, b_n) \in \{0,1\}^{n+1}$ a coalescence indicator where $b_p = 1$ indicates that there is a coalescence at level p.

Definition 3.3.1. *For any $n \geq 1$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, we define the signed finite measures $\Gamma_n^{\dagger,b}$ and $\widetilde{\Gamma}_n^{\dagger,b}$ by*

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

and

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

with the convention

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

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

The connection of the original coalescent tree-based measures proposed in [CDMG11] and the generalized version defined above will be discussed in Section 3.6.1 and Section 3.6.2. By exploiting this novel pair of coalescent Feynman-Kac measures, we have the following alternative representation of the asymptotic variance $\sigma_{\gamma_n}^2(f)$. The rigorous verification is housed in Section 3.7.2.

$$\sigma_{\gamma_n}^2(f) := \sum_{p=0}^{n} \left(\Gamma_n^{\dagger,(p)}(f^{\otimes 2}) - \Gamma_n^{\dagger,(\varnothing)}(f^{\otimes 2}) \right) + \sum_{p=0}^{n-1} \widetilde{\Gamma}_n^{\dagger,(p)}(f^{\otimes 2}). \tag{3.8}$$

3.3.2 Term by term asymptotic variance estimators

Thanks to the alternative representation (3.8) given in the last section, the variance estimation problem is reformulated as how we can estimate the corresponding coalescent Feynman-Kac measures. Using the same idea as in [DG19](Chapter 2), we construct the particle approximation of $\Gamma_n^{\dagger,b}$ and $\widetilde{\Gamma}_n^{\dagger,b}$. They will be referred to as *coalescent tree occupation measures* in this chapter.

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

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

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

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

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

Definition 3.3.2. *For any $n \geq 0$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, the estimator $\Gamma_{n,N}^{\ddagger,b}$ of $\Gamma_n^{\dagger,b}$ is defined by*

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

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

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

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

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

Notice that, by standard convention, we get

$$\Gamma_{0,N}^{\ddagger,b} := \frac{1}{N(N-1)} \sum_{\ell_0^{[2]} \in (N)^2} C_{b_0}(F)(X_0^{\ell_0^{[2]}}) = \frac{1}{N(N-1)} \sum_{i \neq j} C_{b_0}(F)(X_0^i, X_0^j). \tag{3.9}$$

Remark. In fact, $\Gamma_{n,N}^{\ddagger,b}$ defined above is exactly the same estimator as $\Gamma_{n,N}^b$ defined in Definition 3.2 of [DG19](Chapter 2). The change of notation is due to the change of resampling scheme, and the exact reason lies in a technical result (cf. Proposition 3.7.9). The inhomogeneity of the notation w.r.t. "‡" and "†" is due to some nontrivial combinatorial structure of the asymmetric SMC framework. The detailed explanation can be found in Proposition 3.6.3 and other remarks in Section 3.6.2.

Definition 3.3.3. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator b, the estimator $\widetilde{\Gamma}_{n,N}^{\dagger,b}$ of $\widetilde{\Gamma}_n^{\dagger,b}$ is defined by*

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

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

$$\forall (\ell_{p:p+1}^{[2]}, \boldsymbol{\beta_p}, \boldsymbol{x_p}) \in \left((N)^2\right)^{\times 2} \times \{0,1\}^N \times E_p^N,$$

we let

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

and

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

We also define

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

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

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

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

$$\widetilde{\mathbf{G}}_{\mathbf{p}}^{\dagger,0}(\ell_{p:p+1}^{[2]},\boldsymbol{\beta}_{\boldsymbol{p}},\boldsymbol{x}_{\boldsymbol{p}}) := \beta_p^{\ell_{p+1}^1}\beta_p^{\ell_{p+1}^2} m(\boldsymbol{x}_{\boldsymbol{p}})^{\odot 2}(G_p) + \frac{N}{N-1}(1-\beta_p^{\ell_{p+1}^1}\beta_p^{\ell_{p+1}^2})m^{\otimes 2}(\boldsymbol{x}_{\boldsymbol{p}})(G_p^{\otimes 2}). \tag{3.10}$$

Returning to the coalescent tree-based expansion given in (3.8), it is natural to define the term by term estimators $\sigma^2_{\gamma_n^N}(f)$ as follows:

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

Then, by (3.7), it is natural to consider

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

Therefore, thanks to Theorem 3.6.3 and Corollary 3.6.3.1 , we have the consistency of these term by term variance estimators.

Theorem 3.3.1 (Consistency of $\sigma^2_{\gamma_n^N}$ and $\sigma^2_{\eta_n^N}$)**.** *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

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

as well as

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

Remark. We do not provide the algorithm to compute these estimators, since, to the best of our knowledge, they can only be computed with time complexity $\mathcal{O}(nN^2)$. Therefore, they mainly serve as theoretical handy tools prove the consistency of the efficient estimator given in Algorithm 5 and Algorithm 6. However, with the same techniques as in these two Algorithms, one should be able to design an algorithm such that each term in the asymptotic variance can be evaluated separately, with time complexity $\mathcal{O}(nN)$. The details are given in Section 3.6.6.

3.3.3 Unbiased non-asymptotic variance estimator

In this section, we provide an unbiased non-asymptotic variance estimator which is only valid under symmetric resampling scheme, i.e., $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 0$. However, we prove that in the general case, this estimator also yields a consistent asymptotic variance estimator. In fact, in order to provide an unbiased non-asymptotic variance estimator, the idea is much more straightforward: thanks to Theorem 3.2.1, we know that under symmetric sampling scheme, the estimation $\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}$ is unbiased. As a consequence, one has

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

It is then clear that constructing an unbiased non-asymptotic variance estimator is equivalent to constructing an unbiased estimator for the measure $\gamma_n^{\otimes 2}$. Therefore, the following proposition is a direct consequence of Proposition 3.7.3. The detailed computation is provided in Algorithm 7 in Section 3.5.

Theorem 3.3.2. *Assume symmetric resampling, that is, $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 0$. For any test function $f \in \mathcal{B}_b(E_n)$, the estimator $V_n^N(f)$ defined below is an unbiased variance estimator of $\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}$:*

$$V_n^N(f) := \left(\gamma_n^N(f)^2 - \Gamma_{n,N}^{(\varnothing)}(f^{\otimes 2})\right)\mathbf{1}_{\tau_N \geq n}. \tag{3.11}$$

Remark. In fact, this unbiased estimator can also be used by AMS methods if the image of the reaction coordinate is a finite set, and under some regularity assumption on the resampling kernel (e.g. Assumption 1, 2 of [BGG+16]) is satisfied. Although the associated IPS is not simulated by symmetric resampling, recent results (cf. [CDGR18b]) show that one can construct an IPS with particles defined by some level-indexed processes that are "mathematically symmetrically resampled". It can be regarded as an almost sure equivalence between an artificial asymmetric IPS, i.e. the real-world algorithm, and a symmetric IPS constructed by some abstract mathematical objects. More generally, when the reaction coordinate is finite-valued, the AMS method enters the asymmetric SMC framework. More discussions on this topic can be found in Section 3.4.1. As a consequence, for all the unbiasedness results in this chapter, we only use the condition *under symmetric resampling* or *assume symmetric resampling* in order to simplify the writings.

3.3.4 Connection between the estimators

The connection between the term by term estimators and the non-asymptotic variance estimator is based on Theorem 3.6.1. Unfortunately, to the best of our knowledge, both of these estimators can only be computed with $\mathcal{O}(nN^2)$ time complexity. However, this connection inspired the construction of the efficient consistent estimator provided in the next section. The proof is provided in Section 3.7.5.

Proposition 3.3.1. *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

$$\sup_{N>1} N\mathbf{E}\left[\left|NV_n^N(f) - \sigma^2_{\eta_n^N}(f)\right|\right] < +\infty, \tag{3.12}$$

and

$$NV_n^N(f - \eta_n^N(f)) - \sigma^2_{\eta_n^N}(f - \eta_n^N(f)) = \mathcal{O}_{\mathbf{P}}\left(\frac{1}{N}\right). \tag{3.13}$$

3.3.5 Efficient asymptotic variance estimators

The efficient variance estimator can be regarded as a "mix" of the term by term estimator and the non-asymptotic estimator provided respectively in Section 3.3.2 and Section 3.3.3. Inspired by a by-product (3.41) in the proof of Proposition 3.3.1, an intermediate estimator can be proposed as follows

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

By the same technique as the one proposed in [LW18], the former term can be approximated by

$$N\left(\gamma_n^N(f)^2 - \Gamma_{n,N}^{\ddagger,(\varnothing)}(f^{\otimes 2})\right),$$

which can be computed with $\mathcal{O}(nN)$ time complexity. This term corresponds to $\mathrm{V}^{\ddagger}$ founded in Algorithm 5 in Section 3.5. Hence, the design of an efficient variance estimator amounts to constructing an efficient estimator for $\widetilde{\Gamma}_n^{\dagger,(p)}$. Since we failed to provide an efficient algorithm to compute $\widetilde{\Gamma}_{n,N}^{\dagger,(p)}$, we consider some particle approximation $\widetilde{\Gamma}_{n,N}^{\ddagger,(p)}$ of $\widetilde{\Gamma}_n^{\dagger,(p)}$, such that $\widetilde{\Gamma}_{n,N}^{\ddagger,(p)}$ is "close" enough to the original one $\widetilde{\Gamma}_{n,N}^{\dagger,(p)}$, and which, at the same time, can be computed with $\mathcal{O}(nN)$ time complexity.

Definition 3.3.4. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator b, the efficient estimator $\widetilde{\Gamma}_{n,N}^{\ddagger,b}$ of $\widetilde{\Gamma}_n^{\dagger,b}$ is defined by*

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

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

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

Proposition 3.3.2. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and any coalescence indicator $b \in \{0,1\}^{n+1}$, we have*

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

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

$$\hat{\sigma}_{\gamma_n^N}^2(f) := N\left(\gamma_n^N(f)^2 - \Gamma_{n,N}^{\ddagger,(\varnothing)}\right) + \sum_{p=0}^{n-1} \widetilde{\Gamma}_{n,N}^{\ddagger,(p)}(f^{\otimes 2}), \tag{3.14}$$

and

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

Thanks to Proposition 3.6.4, Proposition 3.3.2, Proposition 3.6.4, Theorem 3.3.1 and Theorem 3.2.1, we have the following consistency result.

Theorem 3.3.3 (Consistency of $\hat{\sigma}^2_{\gamma_n^N}$ and $\hat{\sigma}^2_{\eta_n^N}$). *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

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

as well as

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

3.4 Discussions

3.4.1 Unbiasedness condition

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

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

The major motivation is to give an intuitive interpretation of Assumption 2 of [BGG+16] in our fixed level setting. Recent works (cf. [DCGR17] and [CDGR18b]) show that Assumption 2 of [BGG+16] can eventually be regarded as a requirement to reformulate AMS methods as Fleming-Viot particle systems, which enters the continuous-time generalization of the current framework, with symmetric resampling, namely, the case where $\dot{Q}_n \equiv \mathring{Q}_n$. More concretely, if the real-world simulation is under asymmetric resampling scheme, but we are somehow able to find out an underlying mathematical structure that is symmetrically resampled, the unbiasedness will be recovered. In addition, as discussed in Section 3.6.1, it also gives a smaller asymptotic variance than the multinomial resampling. Unfortunately, the only setting we could provide that allows this property is the family of AMS methods in the dynamical setting. The readers are referred to [DCGR17], [CDGR18b] and [BGG+16] for more details. In this respect, since the real mathematical object is simply a symmetric SMC model, we did not emphasize the more complex condition discussed above, which is also due to the fact that the variance estimators are the same in the symmetric and asymmetric resampling schemes. Moreover, it would be interesting to see if we can implement the same technique in Particle Tempering methods, at least in some specific situations.

3.4.2 Other comments

On the adaptive SMC models First, we want to mention that the current setting covers the Adaptive Multilevel Splitting methods when the image of the reaction coordinate is a finite set. This is due to the fact that under asymmetric resampling scheme, when the potential of a particle is 1, there is no additional computational cost required to evolve the IPS. In practice, it corresponds to the case where the reaction coordinate function is calculated by some prefixed grid. Next, since the ingredients we need to implement asymmetric SMC are nearly the same as for the classical multinomial SMC, we can therefore consider the corresponding adaptive methods as in our previous work in [DG19](Chapter 2). Since all the technical results are done in a similar style, we expect the adaptive version with Assumption 2 of [DG19](Chapter 2) to be a simple generalization from a mathematical point of view, at the price of some notational complications. Therefore, the variance estimators may be used as a reference if the underlying resampling scheme is changed to the asymmetric one. We also expect the asymptotic variance estimators to

apply to another family of adaptive SMC models, which contain an online adaptive resampling strategy, such as [GDM17, DMDJ12]. Roughly speaking, the resampling is executed when some summary statistics, such as the popular *Effective Sample Size*, attains some prefixed threshold. In this scenario, as the adaptive model and the fixed reference model are connected by a coupling argument, one is also encouraged to use our estimators in the real-world applications as a reference. However, there are situations where we are sure that the estimators provided in this chapter will fail. The first example is in Section 2 of [BJKT16]. If the stability property given in Theorem 2.3 is not verified: namely, if the "limit" model has different asymptotic variance, then, it is not possible to conduct variance estimations with our estimators. The same argument also applies to the *Adaptive Tempering* introduced in section 3 of [BJKT16]: it is not possible to use our variance estimator as a reference even if we change the underlying resampling scheme. In all cases, the rigorous analysis of the adaptive context requires more attention on the regularity of the adaptive parametrizations.

On the non-asymptotic variance expansion and long-term behaviors There is an angle that we did not cover in the present work: using the finer analysis given in Lemma 3.7.10 and the decomposition in Theorem 3.6.1, a very sharp upper bound of the non-asymptotic variance can be obtained, w.r.t. both n and N. Since it is well understood that the non-asymptotic variance may also contribute to the bound encountered in the propagation of chaos property of the particle system (cf. [DMPR09]), we expect that one can also derive the sharp propagation of chaos bound for the non-asymptotic empirical measures. In particular, this kind of analysis can provide information on the bias associated to the estimation $\eta_n^N(f)\mathbf{1}_{\tau_N \geq n}$. The same idea can also be applied to obtain sharp $\mathbb{L}^p$-bound estimates. This is a relatively large topic, the rigorous analysis is thus left for future investigations. Returning to the variance estimation problems, we remark that all the consistent variance estimators provided in this chapter are essentially for "short-term" models, namely, n is set to be finite and N tends to infinity. When more stability properties of the Feynman-Kac kernels are available, it would be interesting to investigate the fixed-lag variance estimator such as the one introduced in [OD19]. The same kind of estimators, i.e., by considering only part of the genealogy, as well as the survival history, is expected to be more numerically stable in the long term. We expect that the regularity requirements will be the same as in the multinomial case.

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

3.5 Algorithms to compute variance estimators

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

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

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

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

```
Initialization:
Allocate memory for IPS, GENE, SH and EVE;
T = 0;
SumG = 0;
for i ∈ {1, 2, ..., N} do
    IPS[0, i] ~ η_0;
    SumG = SumG+G_0(IPS[0, i]);
    EVE[0, i] = i;
end
Iteration:
while SumG > 0 and T < n do
    SumG = 0;
    for i ∈ {1, 2, ..., N} do
        U ~ Uniform[0, 1];
        if U ≤ G_T(IPS[T, i]) then
            ParentIndex = i;
            IPS[T + 1, i] ~ Ṁ_{T+1}(IPS[T, ParentIndex], ·);
            SH[T, i] = 1;
        else
            ParentIndex ~ Categorical(G_T(X_T^1), G_T(X_T^2), ..., G_T(X_T^N));
            IPS[T + 1, i] ~ M̊_{T+1}(IPS[T, ParentIndex], ·);
            SH[T, i] = 0;
        end
        EVE[T + 1, i] = EVE[T, ParentIndex];
        GENE[T, i] = ParentIndex;
        SumG = SumG+G_{T+1}(IPS[T + 1, i]);
    end
    T = T + 1;
end
T = max{0, T − 1}1_{T<n} + n1_{T=n}.
```

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

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

```
if T < n then
    η_n^N(f) = 0;
    γ_n^N(f) = 0;
else
    Normalizer = 1;
    for p ∈ {0, 1, ..., n − 1} do
        Normalizer = Normalizer × (1/N) Σ_{i=1}^N G_p(IPS[p, i]);
    end
    η_n^N(f) = (1/N) Σ_{i=i}^N f(IPS[n, i]);
    γ_n^N(f) = Normalizer × η_n^N(f);
end
return γ_n^N(f) and η_n^N(f).
```

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

Algorithm 3: Generate backward genealogy tracing matrix: Θ.

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

```
Initialization:
Allocate memory for Θ ∈ M_{T×N}([N]);
Iteration:
for i ∈ {1, 2, ..., N} do
    CurrentIndex = i;
    for p ∈ {1, 2, ..., T} do
        ParentIndex = GENE[T − p, CurrentIndex];
        Θ[T − p, i] = ParentIndex;
        CurrentIndex = ParentIndex;
    end
end
```

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

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

```
return starProduct(X, Y) = X[1] × Y[1] + X[2] × Y[3] + X[3] × Y[2].
```

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

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

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

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

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

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

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

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

```
if T < n then
    return 0;
else
    Allocate memory for MeanG, MeanG2 and MeanGdot2 ∈ M_{n×1}(R);
    for p ∈ {0, 1, ..., n − 1} do
        MeanG[p] = (1/N) Σ_{i=1}^{N} G_p(IPS[p, i]);
        MeanG2[p] = (1/N) Σ_{i=1}^{N} G_p(IPS[p, i])^2;
        MeanGdot2[p] = (MeanG[p] − MeanG2[p]/N) × N/(N−1);
    end
    V^(∅) = 0;
    for i ∈ {1, 2, ..., N − 1} do
        for j ∈ {i + 1, ..., N} do
            if EVE[n, i] ≠ EVE[n, j] then
                ProdCouple = f(IPS[n, i]) × f(IPS[n, j]);
                Index1 = i, Index2 = j;
                for p ∈ {0, 1, ..., n − 1} do
                    ParentIndex1 = GENE[n − p − 1, Index1];
                    ParentIndex2 = GENE[n − p − 1, Index2];
                    if SH[n − p − 1, Index1] = 1 & SH[n − p − 1, Index2] = 1 then
                        ProdCouple = ProdCouple × MeanGdot2[n − p − 1];
                    else if SH[n − p − 1, Index1] = 1 & SH[n − p − 1, Index2] = 0 then
                        ProdCouple = ProdCouple × MeanG[n − p − 1] × {MeanG[n − p − 1] × N/(N−1)
                          − N × (G_{n−p−1}(IPS[n−p−1, ParentIndex1]) × MeanG[n−p−1] − MeanG2[n−p−1])
                            / ((N−1) × (N−1−MeanG[n−p−1] + G_{n−p−1}(IPS[n−p−1, ParentIndex1]))) };
                    else if SH[n − p − 1, Index1] = 0 & SH[n − p − 1, Index2] = 1 then
                        ProdCouple = ProdCouple × MeanG[n − p − 1] × {MeanG[n − p − 1] × N/(N−1)
                          − N × (G_{n−p−1}(IPS[n−p−1, ParentIndex2]) × MeanG[n−p−1] − MeanG2[n−p−1])
                            / ((N−1) × (N−1−MeanG[n−p−1] + G_{n−p−1}(IPS[n−p−1, ParentIndex2]))) };
                    else
                        ProdCouple = ProdCouple × N/(N−1) × MeanG[n − p − 1]^2;
                    end
                    Index1 = ParentIndex1;
                    Index2 = ParentIndex2;
                end
                V^(∅) = V^(∅) + ProdCouple;
            end
        end
    end
end
V^(∅) = 2 × V^(∅)/(N(N − 1));
return γ_n^N(f)^2 − V^(∅).
```

Above, we also provide Algorithm 7 to compute unbiased non-asymptotic variance estimator for $\gamma_n^N(f)$. A tremendous amount of effort has been spent in order to find an $\mathcal{O}(nN)$ time complexity algorithm, which, unfortunately, does not pay back. The Algorithm 7 is of $\mathcal{O}(nN^2)$ time complexity, which means that even with a very little n, the computation will be intractable when $N > 10^6$. In fact, the construction for the consistent asymptotic variance estimators, i.e., Algorithm 5 and Algorithm 6, of $\mathcal{O}(nN)$ time complexity is highly nontrivial, and we failed to apply the same technique to reduce the time complexity for the unbiased non-asymptotic variance estimator. Still, this estimator may be useful for rare-event simulation problems or other applications whose target measure is γ_n. One can thus take advantage of the parallel computing for relatively small N. Though, the crude estimator is a by-product in this case, the average of this estimator may still be typically more accurate and the statistical inference for the variance of the variance estimator is also available if an unbiased variance estimator is provided. Note that the *lack-of-bias* is not free in general, and we provide a relatively general condition in Section 3.4.1. In fact, even without unbiased condition, this estimator multiplied by N is also a consistent asymptotic variance estimator for $\gamma_n^N(f)$. Another remark is that $\mathrm{V}^{\ddagger}$ found in Algorithm 5 and Algorithm 6 represent respectively the estimators $NV_n^N(f)\gamma_n^N(1)^2$ and $NV_n^N(f - \eta_n^N(f))$ provided in [LW18]. Due to the change of resampling scheme, some modifications, namely, $\widetilde{\mathrm{V}}^{\dagger}$ have to be taken into consideration. Now, we provide the time complexity and space complexity of the algorithms in the SMC context. The multinomial resampling scheme will be set as benchmark, with the variance estimators provided in [LW18].

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

Table 3.1: Time and space complexity under multinomial resampling scheme.

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

Table 3.2: Time and space complexity under asymmetric resampling scheme.

We remark that we did not provide the algorithm to compute $\gamma_n^N(f)$ and $\eta_n^N(f)$ with $\mathcal{O}(N)$ space complexity, which is readily obtained with some modification of Algorithm 1, since we mainly focus on the variance estimation problems. We can see from Table 3.1 and Table 3.2 that the main drawback of the presented setting is the space complexity and the unbiased variance estimator for the non-asymptotic variance of $\gamma_n^N(f)$. However, the main computational consumption, in general, is brought by the resampling kernels $\mathring{M}_n$ at each iteration of the algorithm. If a cheap kernel $\dot{M}_n$ is available, it is expected that the asymmetric resampling scheme would dramatically reduce the computational cost of the simulation of IPS. In fact, in real-world

applications, the computational consumption of the asymptotic variance estimators is negligible compared to the simulation of IPS. In addition, the asymmetric setting also gives smaller variance in some specific situations, such as AMS methods.

3.6 Coalescent tree-based expansions

As we have seen in Theorem 3.2.2, the asymptotic variance becomes sophisticated when the asymmetric resampling is implemented. Therefore, we need to develop a novel mathematical language in order to conduct calculations and eventually, to understand the structures behind. In this section, we give a detailed development of the coalescent tree-based expansions encountered in the asymmetric SMC framework. Before going further, let us list some definitions and properties associated to McKean-type Feynman-Kac kernel $Q_{n,\mu}$ within reach by some straightforward algebraic calculations in the following proposition. For the sake of simplification, these properties will be of constant use in the following sections and may be applied without reference.

Proposition 3.6.1. *For any probability measure $\mu \in \mathcal{P}(E_{n-1})$ and test function $f \in \mathcal{B}_b(E_n)$, we have the following properties:*

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

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

$$\forall (x,A) \in E_{n-1}\times\mathcal{B}(E_n), \quad R_{n,\mu}(x,A) := \mu(G_{n-1})\dot{Q}_n(x,A) - G_{n-1}(x)\mu\mathring{Q}_n(A), \tag{3.15}$$

with the convention

$$\forall x \in E_0, \quad R_{0,\mu}(x,A) := \eta_0(A). \tag{3.16}$$

Then, we have

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

and

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

(iii) *We have*

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

as well as

$$\eta_{n-1}\left(Q_{n,\eta_{n-1}}(f)^2\right) = \eta_{n-1}\mathring{Q}_n(f)^2 + \eta_{n-1}\left(R_{n,\eta_{n-1}}(f)^2\right). \tag{3.17}$$

3.6.1 Original coalescent tree-based measures

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

Definition 3.6.1. *For any $n' \geq n$, we associate with any coalescence indicator $b \in \{0,1\}^{n'+1}$ the nonnegative measures $\Gamma_n^b \in \mathcal{M}_+(E_n^2)$ defined for any $F \in \mathcal{B}_b(E_n^2)$ by*

$$\Gamma_n^b(F) := \eta_0^{\otimes 2}C_{b_0}\mathring{Q}_1^{\otimes 2}C_{b_1}\cdots\mathring{Q}_n^{\otimes 2}C_{b_n}(F).$$

When there is only one coalescence at, say, level p, we write $\Gamma_n^{(p)}(F)$ instead of $\Gamma_n^b(F)$ (see Figure 3.2). When there is no coalescence at all, that is $b = (\varnothing)$, we have

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

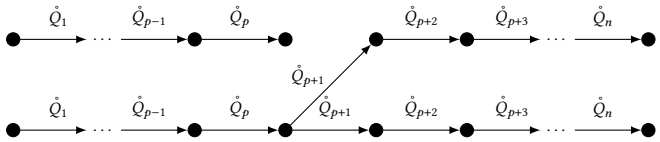

Figure 3.2: A representation of the original coalescent tree-based measure $\Gamma_n^{(p)}$.

Comparison of asymptotic variance. Now, we suppose that $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 1$. Let us go back to the form of the asymptotic variance $\sigma^2_{\gamma_n}(f)$ defined in (3.6). It is easy to verify that

$$\Gamma_n^{(p)}(f^{\otimes 2}) = \gamma_p(1)\, \gamma_p\left(Q_{p,n}(f)^2\right). \tag{3.18}$$

By applying (3.17) with some standard algebraic manipulations, we have

$$\sigma^2_{\gamma_n}(f) = \sum_{p=0}^{n}\left(\Gamma_n^{(p)}(f^{\otimes 2}) - \Gamma_n^{(\varnothing)}(f^{\otimes 2})\right) - \sum_{p=1}^{n}\gamma_{p-1}(1)\gamma_{p-1}\left(R_{p,\eta_{p-1}}Q_{p,n}(f)^2\right). \tag{3.19}$$

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

$$\gamma_{p-1}(1)\gamma_{p-1}\left(R_{p,\eta_{p-1}}Q_{p,n}(f)^2\right) \tag{3.20}$$

is nonnegative, we deduce that the choice $\dot{Q}_n \equiv \mathring{Q}_n$ is always better than the multinomial resampling scheme in terms of asymptotic variance. Moreover, we notice that the original coalescent tree-based measures introduced in [CDMG11] and [DG19](Chapter 2) failed to provide a full description of the asymptotic variance, even in this simple symmetric case. This is the main difficulty compared to the multinomial resampling scheme, where the alternative representation is free. Therefore, we need to develop some new tools to understand the term given in (3.20).

3.6.2 Coalescent Feynman-Kac kernels

As we have seen in the last section, the original coalescent tree-based measures fail to provide insights on the asymptotic variance $\sigma^2_{\gamma_n}(f)$. In order to go one step further, let us go back to Definition 3.6.1. We consider the following alternative writing

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

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

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

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

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

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

$$\begin{cases} \mathbf{Q}_n^0 := \left(\mathring{Q}_n + \eta_{n-1}^{\otimes 2}(1 \otimes G_{n-1})(\dot{Q}_n - \mathring{Q}_n)\right)^{\otimes 2}; \\ \mathbf{Q}_n^1 := C_1 \mathbf{Q}_n^0. \end{cases} \tag{3.21}$$

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

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

Next, we need to introduce several important coalescent Feynman-Kac kernels, which, at the moment, is not as intuitive as the one introduced above. Their introduction is motivated by an observation in the proof of a technical result (cf. Proposition 3.7.9):

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

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

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

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

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

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

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

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

all of the other kernels conserve the coalescence structure. This observation may be the intrinsic reason why they play a particularly important role in variance related problems. No matter how anecdotal it seems, we claim that these kernels are at the core of the analysis of the variance related problems. Although we are not able to clarify the exact purpose of the construction of these coalescent Feynman-Kac kernels at the moment, we can explain, however, the logic of our notation: the number of daggers "†" indicates the number of kernels between $b_{n-1} = 0$ and $b_{n-1} = 1$, that is changed from the original definition (3.21). At the same time, the kernel for $b_{n-1} = 1$ is always replaced before the kernel for $b_{n-1} = 0$. This is why all the kernels that have only one dagger share the same $\mathbf{Q}_n^0$ for the case $b_{n-1} = 0$. Since the number of particles N is

also involved in the definition, we add parenthesis "(N)" to specify the number of particles N, in order to differentiate from the coalescent tree occupation measures. With a slight abuse of notation, when there is no ambiguity, we omit the part "(N)" for simplicity. For example, we may use $\mathbf{Q}_n^{\ddagger,b}$ to denote $\mathbf{Q}_{n,(N)}^{\ddagger,b}$. All the kernels defined from point (i) to point (v) are introduced to describe the binary decomposition w.r.t. "+" in the definition of point (vi). We say that the kernels defined above are in the class $\mathcal{Q}_n^{(2)}$, or the kernels are of $\mathcal{Q}_n^{(2)}$-class.

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

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

with the convention

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

Definition 3.6.2. *For any $n \geq 1$, $N \geq 2$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, we define the signed finite measures $\Gamma_{n,(N)}^{\ddagger,b}$ by*

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

with the convention

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

Similar as in Definition 3.6.1, when there is only one coalescence at level p, we write $\Gamma_{n,(N)}^{\ddagger,(p)}(F)$ instead. When there is no coalescence , we denote $\Gamma_{n,(N)}^{\ddagger,(\varnothing)}(F)$.

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

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

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

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

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

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

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

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

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

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

We remark that the kernels of $\mathcal{Q}^{(2)}_{\hat{n}}$-class will not be used to define the coalescent tree-based measures, the introduction is purely for technical reasons (cf. Proposition 3.7.9, Proposition 3.7.10, Lemma 3.7.10 and Lemma 3.7.11). They are eventually proved to be very "close" to the $\mathcal{Q}_n$-class kernels (cf. Lemma 3.7.12) by the propagation of chaos property of the IPS (cf. Proposition 3.7.5).

3.6.3 Binary decompositions

Next, we define some auxiliary coalescent Feynman-Kac kernels using the same idea. Before that, we need some new notation to describe the coalescence structure that is a little bit more complicated than the basic binary structure illustrated in Figure 3.2. For one coalescence indicator b, we use $|b|$ to denote the number of 1 in b, namely

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

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

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

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

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

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

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

$$\begin{cases} \mathbf{Q}^{0|0}_n := \mathbf{Q}^{\dagger,0}_n \\ \mathbf{Q}^{1|0}_n := \widetilde{\mathbf{Q}}^{\dagger,1}_n \\ \mathbf{Q}^{0|1}_n := \check{\mathbf{Q}}^{\dagger,1}_n \\ \mathbf{Q}^{1|1}_n := \mathbf{Q}^{\dagger,1}_n \end{cases} \quad \text{and} \quad \begin{cases} \mathbf{Q}^{0|0}_{n,(N)} := \mathbf{Q}^{\dagger,0}_n \\ \mathbf{Q}^{1|0}_{n,(N)} := \widetilde{\mathbf{Q}}^{\dagger,1}_{n,(N)} \\ \mathbf{Q}^{0|1}_{n,(N)} := \check{\mathbf{Q}}^{\dagger,1}_{n,(N)} \\ \mathbf{Q}^{1|1}_{n,(N)} := \mathbf{Q}^{\dagger,1}_n \end{cases} \tag{3.22}$$

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

$$\underbrace{0 : \mathbf{Q}^{\ddagger,0}_{n,(N)}}_{0|0:\mathbf{Q}^{\dagger,0}_n + \widetilde{\mathbf{Q}}^{\dagger,1}_{n,(N)}:1|0} \quad \text{and} \quad \underbrace{1 : \mathbf{Q}^{\ddagger,1}_{n,(N)}}_{1|1:\mathbf{Q}^{\dagger,1}_n + \check{\mathbf{Q}}^{\dagger,1}_{n,(N)}:0|1}, \tag{3.23}$$

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

Definition 3.6.3. *For any $n' \geq n \geq 1$, $N \geq 2$ and for any coalescence indicators $b, b' \in \{0,1\}^{n'+1}$, we define the signed finite measures $\Gamma_n^{b'|b}$ and $\Gamma_{n,N}^{b'|b}$ respectively by*

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma_n^{b'|b}(F) := \eta_0^{\otimes 2}\, \mathbb{Q}_1^{b'_0|b_0}\, \mathbb{Q}_2^{b'_1|b_1} \cdots \mathbb{Q}_n^{b'_{n-1}|b_{n-1}} C_{b_n}(F),$$

and

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma_{n,(N)}^{b'|b}(F) := \eta_0^{\otimes 2}\, \mathbb{Q}_{1,(N)}^{b'_0|b_0}\, \mathbb{Q}_{2,(N)}^{b'_1|b_1} \cdots \mathbb{Q}_{n,(N)}^{b'_{n-1}|b_{n-1}} C_{b_n}(F),$$

with the convention

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

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

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

Proposition 3.6.2. *For any $n' \geq n \geq 1$, $N \geq 2$, for any coalescence indicators $b, b' \in \{0,1\}^{n'+1}$ and for any test function $F \in \mathcal{B}_b(E_n^2)$, we have the following equalities:*

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

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

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

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

$$\begin{aligned}
\Gamma_{n,(N)}^{\ddagger,b}(F) &= \sum_{b' \in \mathcal{S}_n(b)} \Gamma_{n,(N)}^{b'|b}(F) \\
&= \Gamma_{n,(N)}^{b|b}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n(b)} \Gamma_{n,(N)}^{b'|b}(F) \\
&= \Gamma_n^{b|b}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n(b)} \left(\frac{1}{N-1}\right)^{|b'-b|} \Gamma_n^{b'|b}(F) \\
&= \Gamma_n^{\dagger,b}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n(b)} \left(\frac{1}{N-1}\right)^{|b'-b|} \Gamma_n^{b'|b}(F).
\end{aligned} \tag{3.24}$$

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

$$\begin{aligned}
\Gamma^{\ddagger,(\varnothing)}_{n,(N)}(F) =& \sum_{b' \in \mathcal{S}_n(b)} \Gamma^{b'|b}_{n,(N)}(F) \\
=& \Gamma^{(\varnothing)|(\varnothing)}_{n,(N)}(F) + \sum_{b' \in \mathring{\mathcal{S}}_n((\varnothing))} \Gamma^{b'|(\varnothing)}_{n,(N)}(F) \\
=& \Gamma^{\dagger,(\varnothing)}_{n}(F) + \sum_{b' \in \mathcal{S}^{>}_n((\varnothing))} \left(\frac{1}{N-1}\right)^{|b'|} \widetilde{\Gamma}^{\dagger,b'}_{n}(F) \\
=& \Gamma^{(\varnothing)}_{n}(F) + \frac{1}{N-1} \sum_{p=0}^{n-1} \widetilde{\Gamma}^{\dagger,(p)}_{n}(F) + \sum_{b' \in \mathcal{S}^{>}_n(b), |b'| \geq 2} \left(\frac{1}{N-1}\right)^{|b'|} \widetilde{\Gamma}^{\dagger,b'}_{n}(F).
\end{aligned} \tag{3.25}$$

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

Proposition 3.6.3. *For any $n' \geq n \geq 1$ and for any coalescent indicator $b \in \{0,1\}^{n'+1}$, we have*

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

In particular, we have

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

Remark. Above lies part of the reason why we have inhomogeneity in the notation w.r.t. "‡" and "†" in Definition 3.3.2. In fact, the strategy to prove the consistency given in Theorem 3.6.3 is divided into two steps, and the latter is done by Proposition 3.6.3 above:

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

3.6.4 Coalescent tree occupation measures

In this section, we introduce the particle approximations of the coalescent tree-based measures discussed in the last section. To do this, we need to adopt some notation from [DG19](Chapter 2). The readers are also referred to Appendix A.2 of [DG19](Chapter 2) to find more intuitions and the connection between the construction of these particle approximations and the Particle Markov Chain Monte Carlo methods. Due to the fact that the underlying resampling scheme is changed, the analysis also becomes more challenging. Fortunately, the basic idea remains the same: we exploit the information encoded in the genealogy, and in addition, the information encoded in the survival history, to approximate the coalescent Feynman-Kac measures. The key idea is to collect all the corresponding coalescent tree-type forms illustrated in Figure 3.2, and the coalescent tree occupation measures are constructed as weighted empirical terminal measures of these particle blocks. The intuition of this procedure remains identical to the previous work in [DG19](Chapter 2). More precisely, the major difference is about the weights mentioned

above, which correspond to the potential function of the original IPS in many-body Feynman-Kac models (cf. $\mathcal{G}_p^{(q)}(\boldsymbol{x}_p)$ defined in A.1 of [DG19](Section 2.5.1)). Under asymmetric resampling scheme, it is necessary to consider the influence of the survival history. Thus, it is expected that the constructions become more sophisticated. Another remark is on the measure $\widetilde{\Gamma}_n^{\dagger,b}$: since the related coalescent Feynman-Kac kernels do not conserve the coalescence structure, its particle approximation also turns out to be a little bit different. Recall that $\lambda_p^b(\tilde{a}_p^{[2]}, \ell_p^{[2]}) \in \{0,1\}$ is an indicator function defined by

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

Definition 3.6.4 (Coalescent tree occupation measures)**.** *For any $n' \geq n \geq 0$ and for any coalescence indicator $b \in \{0,1\}^{n'+1}$, the random measure $\bar{\Gamma}_{n,N}^b$ is defined by*

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

The next theorem is brought from Proposition 4.2 of [DG19](Chapter 2), and the proof is a slightly modified version of the one given in Section 4.5 of [DG19] (Section 2.4.5). It is one of the most important ingredients that connect the coalescent tree occupation measures and non-asymptotic variance of Feynman-Kac IPS. It provides information on the combinatorial structure of the IPS in regard to the coalescent tree occupation measures $\bar{\Gamma}_{n,N}^b$, which does not depend on the resampling scheme and regularity assumptions whilst the IPS is well-defined. Namely, it reveals the essential combinatorial properties that apply to all genealogy tree-based particle systems when each particle has only one parent. This combinatorial property, in particular, is also valid in continuous-time settings and/or in the frameworks with even more complex resampling schemes. In this chapter, it is the bridge between the asymptotic variance, non-asymptotic variance, and eventually, the construction of our variance estimators. The proof is given in Section 3.7.6.

Theorem 3.6.1. *For any test function $F \in \mathcal{B}_b(E_n^2)$, we have the following decompositions.*

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

and

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

Below we list the most important lack-of-bias and convergence results of the coalecent tree occupation measures, serving as the particle approximations of the coalescent Feynman-Kac measures. The proofs are divided into several technical results: they are direct consequences of the combination of Lemma 3.7.1, Lemma 3.7.2, Proposition 3.7.2, Proposition 3.7.3 and Proposition 3.6.3, with some standard manipulations of bounded i.i.d. random variables for the case $n = 0$, which is provided in Lemma 3.7.4. The only remark is that for any coalescent indicator b and for any coalescent Feynman-Kac kernel, e.g., $\widetilde{\mathbf{Q}}_n^{\dagger,b}$, we have, by definition,

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

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

Theorem 3.6.2 (Unbiasedness). *Assume symmetric resampling, that is $\dot{Q}_n \equiv \mathring{Q}_n$ for all $n \geq 1$, then, we have*

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

Theorem 3.6.3 (Consistency). *For any coalescence indicator $b \in \{0,1\}^{n+1}$ and any test function $F \in \mathcal{B}_b(E_n)^{\otimes 2}$, we have*

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

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

Remark. The notation

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

means that

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

The reader is referred to the beginning of Section 3.7 for details.

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

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

as well as

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

Then, Theorem 3.2.1 gives

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

and

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

Finally, combined to Proposition 3.7.7 and Proposition 3.7.8, we have the following corollary.

Corollary 3.6.3.1. *For any coalescence indicator $b \in \{0,1\}^{n+1}$ and any test function $f \in \mathcal{B}_b(E_n)$, we have*

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

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

3.6.5 Feynman-Kac measures flow in a random environment.

In this section, we gives some intuition on the construction of the coalescent tree occupation measures from another point of view. One of the main message of [DG19](Chapter 2) is to provide some intuition on the construction of the coalescent tree occupation measures using the many-body Feynman-Kac models introduced in [DMKP16], by considering a Gibbs sampler w.r.t. the *original IPS* and *coupled particle block* on a sophisticated path space. Then, we define the event that traps the desired coalescent particle block and eventually, we construct the estimator given in Definition 3.6.4. This methodology gives the foundation of the present work: Definition 3.3.3 is also obtained by this procedure, though it is not discussed in detail as in [DG19](Chapter 2). Now, let us look at this family of random measures from a different angle. We begin with some basic observations. To facilitate the writings, let us fix a time horizon $T \in \mathbf{N}^*$. The following discussion is valid on the event $\{\tau_N \geq T\}$ and $0 \leq n \leq T$. Given $\mathcal{W}_T^N$ (cf. Section 3.7.1) and fixing $\ell_{n-1}^{[2]} \in (N)^2$, we have

$$\forall b \in \{0,1\}^{T+1},\ \forall n \in [T], \qquad \sum_{\ell_n^{[2]} \in (N)^2} \lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) = 1.$$

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

$$\left(\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})\right)_{\left(\ell_{n-1}^{[2]}, \ell_n^{[2]}\right) \in ((N)^2)^{\times 2}}, \tag{3.26}$$

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

$$\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})$$

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

$$\mathbf{H}_n^{\ddagger,b}[\ell_{n-1}^{[2]}, \ell_n^{[2]}] := \mathbf{G}_{\mathbf{n}}^{\ddagger,b}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}),$$

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

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

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

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

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

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

In particular, we denote

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

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

$$\Lambda_n^{\ddagger,b}[\ell_n^{[2]}] := m^{\odot 2}([N]) \cdot \mathbf{H}_1^{\ddagger,b} \cdot \mathbf{H}_2^{\ddagger,b} \cdots \mathbf{H}_n^{\ddagger,b}[\ell_n^{[2]}], \tag{3.27}$$

with the convention $\Lambda_0^{\ddagger,b} := m^{\odot 2}([N])$. Accordingly, the random test function $\mathbf{F}_n^b$ is defined by

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

Consequently, we have

$$\Lambda_n^{\ddagger,b}(\mathbf{F}_n^b) = \Gamma_{n,N}^{\ddagger,b}(F). \tag{3.28}$$

The random measure notation above is frequently used in the proof of technical results, an explicit form can be found later in (3.56). Similarly, we denote

$$\widetilde{\mathbf{H}}_n^{\dagger,b}[\ell_{n-1}^{[2]}, \ell_n^{[2]}] := \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}),$$

and

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

with the convention $\widetilde{\Lambda}_0^{\dagger,b} := m^{\odot 2}([N])$. Apart from the fact that this writing guided and simplified some of the proofs of the technical results, the main motivation is to provide a decomposition result similart to the one in (3.25), which is essentially due to the partial $\mathcal{R}$-algebra homomorphism. We define $\mathbf{H}_n^0 := \widetilde{\mathbf{H}}_n^{\dagger,0}$. By definition, we have

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

which yields, for the associated random matrix,

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

Therefore, we have the following decomposition:

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

which is equivalent to

$$\begin{aligned}
\Gamma_{n,N}^{\ddagger,(\varnothing)}(F) =& \Gamma_{n,N}^{(\varnothing)}(F) + \sum_{b' \in \mathcal{S}_n^{>}((\varnothing))} \left(\frac{1}{N-1}\right)^{|b'|} \widetilde{\Gamma}_{n,N}^{\dagger,b'}(F) \\
=& \Gamma_{n,N}^{(\varnothing)}(F) + \frac{1}{N-1}\sum_{p=0}^{n-1} \widetilde{\Gamma}_{n,N}^{\dagger,(p)}(F) + \sum_{b' \in \mathcal{S}_n^{>}(b), |b'| \geq 2} \left(\frac{1}{N-1}\right)^{|b'|} \widetilde{\Gamma}_{n,N}^{\dagger,b'}(F).
\end{aligned} \tag{3.29}$$

Thanks to Proposition 3.7.8 and the decomposition (3.29) above, we have the following proposition.

Proposition 3.6.4. *For any $n' \geq n \geq 1$ and for any coalescent indicator $b \in \{0,1\}^{n'+1}$, we have*

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

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

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

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

Next, using the same semigroup property of these random matrix and random test function as in (3.27) and (3.28), by consider the initial distribution $m^{\odot 2}([N])$, we define the new coalescent tree occupation measure $\widetilde{\Gamma}_{n,N}^{\ddagger,b}$ for each $b \in \{0,1\}^{n+1}$ by

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

with

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

The theoretical property of the approximation discussed above is provided in the following proposition. The proof is housed in Section 3.7.7.

Proposition 3.6.5. *For any test function $F \in \mathcal{B}_b(E_n^2)$ and for any coalescence indicator b, we have*

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

In particular, for any $p \in \{0,1,\ldots,n-1\}$, we have

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

Finally, without loss of generality, we explain why the estimator given in (3.14) can be computed with $\mathcal{O}(nN)$ time complexity. In fact, the first part of the estimator

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

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

Lemma 3.6.1. *Let $(\mathcal{R}, +, \star)$ be a ring, and $(\mathcal{E}_i)_{i\in[k]}$ be a disjoint partition of $[N]$ for some $k \geq 1$, then for any sequence $(a_i)_{i\in[N]}$ composed by elements of $\mathcal{R}$, we have the following equality:*

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

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

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

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

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

This is useful in calculating the term

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

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

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

Hence, by applying Lemma 3.6.1, one can therefore compute $\widetilde{\Gamma}_{n,N}^{\ddagger,(p)}(f^{\otimes 2})$ with $\mathcal{O}(nN)$ time complexity. The details can be found in Algorithm 5, and the design of the Algorithm 6 is similar. We remark that we are not able to construct the ring homomorphism discussed above for the measure $\widetilde{\Gamma}_{n,N}^{\dagger,b}$. This is the intrinsic reason why we failed to apply the same technique to reduce the time complexity of the non-asymptotic variance estimator $V_n^N(f)$. In the Algorithm 7 of Section 3.5, the corresponding term

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

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

3.7 Proofs

In this section, we list all the proofs in the present work. Some notation are gathered in Section 3.7.1, such as the formal definitions of the filtrations frequently used in the proofs, along with the most important martingale decompositions. A little plan on the organization of the technical results is also provided. In order to facilitate the writing, the stochastic bounds introduced in [Jan11] are intensely involved in our technical results. More precisely, we use frequently the notation $\mathcal{O}_{\mathbf{p}}$, $\mathcal{O}_{\mathbb{L}^p}$ and $\mathcal{O}_{a.s.}$. Let $(a_N; N \in \mathbf{N})$ be a sequence of natural numbers, where N represents the number of particles in the IPS. The notation

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

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

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

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

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

means that the sequence $(X_N/a_N; N \in \mathbf{N})$ is uniformly bounded in $\mathbb{L}^{\mathbf{p}}$-norm w.r.t. N. The notation

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

indicates that

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

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

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

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

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

one also has

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

3.7.1 Martingales

We present some important martingales encountered in the analysis of SMC framework. They are crucial to some of the technical results in this chapter. We also hope that the similar construction may inspire the future work in different context. Before proceeding further, let us define some filtrations associated to the Feynman-Kac IPS.

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

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

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

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

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

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

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

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

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

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

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

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

Proposition 3.7.1. *For any test function $f \in \mathcal{B}_b(n)$, we define*

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

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

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

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

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

$$\left|U_k^N(f)\right| \leq 3\,\|f\|_\infty, \quad a.s. \tag{3.30}$$

which gives the integrability. Then, by the fact that

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

one writes

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

This ends the verification of Proposition 3.7.1. □

Recall that, by definition, we have

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

Hence,

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

which yields

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

We denote

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

with

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

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

$$\begin{aligned}
&\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f) \\
=&\sum_{p=0}^{n}\left(\gamma_p^N(f_{p,n})\mathbf{1}_{\tau_N \geq p} - \gamma_{p-1}^N Q_{p,\eta_{p-1}^N}(f_{p,n})\mathbf{1}_{\tau_N \geq p-1}\right) \\
=&\frac{1}{N}\sum_{k=1}^{(n+1)N} U_k^N(f) + \sum_{p=1}^{n} D_{p,n}^N(f),
\end{aligned} \tag{3.31}$$

taking into account the convention

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

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

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

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

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

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

First, the integrability is guaranteed by Proposition 3.7.7. Then, thanks to Proposition 3.7.9, we get a almost sure equality which is very "close" to a martingale structure:

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

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

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

as well as

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

Thanks to Lemma 3.7.10, Lemma 3.7.12, the Minkowski's inequality and conservation of coalescence structure, we deduce the following Lemma 3.7.1. Then, the Proposition 3.7.2 is a direct application of Doob decomposition theorem.

Lemma 3.7.1. *For any test function $F \in \mathcal{B}_b(E_T^2)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, we have*

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

and

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

Proposition 3.7.2. *For any test function $F \in \mathcal{B}_b(E_T^2)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, the integrable process $(\mathbb{X}_n^{\ddagger,b}(F); 0 \le n \le T)$ can be decomposed to a $(\mathcal{G}_n^N; 0 \le n \le T)$-martingale $(\mathbb{M}_n^{\ddagger,b}(F); 0 \le n \le T)$, and a integrable predictable process $(\mathbb{A}_n^{\ddagger,b}(F); 0 \le n \le T)$, respectively defined by*

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

and

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

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

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

as well as

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

and

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

Thanks to Proposition 3.7.10, Lemma 3.7.11 and Lemma 3.7.12, we have the following results. The unbiasedness given in Proposition 3.7.3 is a direct consequence of the definition of $\widetilde{\mathbf{Q}}^{\dagger,(\varnothing)}_{\hat{n}}$ and $\widetilde{\mathbf{Q}}^{\dagger,(\varnothing)}_n$.

Lemma 3.7.2. *For any test function $F \in \mathcal{B}_b(E^2_T)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, we have*

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

and

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

Proposition 3.7.3. *For any test function $F \in \mathcal{B}_b(E^2_T)$ and any coalescence indicator $b \in \{0,1\}^{T+1}$, the integrable process $(\widetilde{\mathbb{X}}^{\dagger,b}_n(F); 0 \leq n \leq T)$ can be decomposed to a $(\mathcal{W}^N_n; 0 \leq n \leq T)$-martingale $(\widetilde{\mathbb{M}}^{\dagger,b}_n(F); 0 \leq n \leq T)$, and a integrable predictable process $(\widetilde{\mathbb{A}}^{\dagger,b}_n(F); 0 \leq n \leq T)$, respectively defined by*

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

and

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

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

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

which yields

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

is a $(\mathcal{W}^N_n; 0 \leq n \leq T)$-martingale.

3.7.2 Verification of asymptotic variance expansion

In this section, we verify the asymptotic variance expansion (3.8) given in Section 3.3.1. Recall that, with the introduction of coalescent Feynman-Kac kernels, we have

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

By definition, since

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

we have

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

Next, for the latter term, since

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

by applying (3.17), we deduce that

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

Note that

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

whence

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

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

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

Taking into account that

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

and

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

as well as the convention (3.4) for the case $p = 0$, that writes

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

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

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

3.7.3 Proof of Theorem 3.2.1

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

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

for the case $\dot{Q}_n \equiv \mathring{Q}_n$ is a direct consequence of the martingale decomposition (3.31). For the almost sure convergence, the proof is done by induction. For the step 0, the almost sure convergence of η_0^N is a direct consequence of law of large numbers for i.i.d. random variables. For step $n \geq 1$, we suppose that for each $0 \leq p \leq n-1$, we have

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

We first check that

$$\frac{1}{N} \sum_{k=1}^{(n+1)N} U_k^N(f) \xrightarrow[N\to\infty]{a.s.} 0. \tag{3.32}$$

Taking into account that

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

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

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

Hence, the almost sure convergence (3.32) is then ensured by Borel-Cantelli lemma. On the other hand, the induction hypothesis gives

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

which yields

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

The verification of the almost sure convergence for $\gamma_n^N(f)\mathbf{1}_{\tau_N \geq n}$ is then complete. The almost sure convergence of η_n^N is then trivial since for any test function $f \in \mathcal{B}_b(E_n)$, the convention (3.1) allows the writing

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

3.7.4 Proof of Theorem 3.2.2

Lemma 3.7.3. *Let μ^N be an empirical on E_{n-1}, we suppose that there exists a probability measure μ on E_{n-1}, such that for any test function $\phi \in \mathcal{B}_b(E_{n-1})$, one has*

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

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

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

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

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

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

with $R_{n,\mu}$ defined in (3.15). Basic algebraic mulipulation gives

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

Recall that, by definition, we have

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

and

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

whence we deduce that

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

Since $G_{n-1}, \mathring{Q}_n(f) \in \mathcal{B}_b(E_{n-1})$, Theorem 3.2.1 gives that

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

On the other hand, as

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

we safely deduce that

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

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

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

whence

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

Finally, as $\mathring{Q}_n(f), \dot{Q}_n(f) \in \mathcal{B}_b(E_{n-1})$, point (i) and Theorem 3.2.1 combined with the fact that

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

ensure the desired convergence in point (ii). This closes the proof of Lemma 3.7.3. □

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

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

Again, let us go back to the decomposition (3.31). First, we prove that

$$\frac{1}{\sqrt{N}} \sum_{k=1}^{(n+1)N} U_k^N(f) \xrightarrow[N\to\infty]{\mathrm{d}} \mathcal{N}\left(0, \sigma^2_{\gamma_n}\right). \tag{3.33}$$

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

- The boundness of G_n gives that

$$\max_{1\leq k\leq (n+1)N} \left|\frac{1}{\sqrt{N}} U_k^N(f)\right| \leq \frac{3}{\sqrt{N}} \|f\|_\infty, \tag{3.34}$$

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

- From (3.34), one also gets that

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

- For the asymptotic variance, we deduce that

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

First, let us prove that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N} P_1^N(k) \xrightarrow[N\to\infty]{a.s.} \sum_{p=0}^{n} \gamma_p(1)\gamma_p(f_{p,n}^2). \tag{3.35}$$

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

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

Hence, Theorem 3.2.1 gives the desired convergence (3.35). Second, for the term concerning $P_2(k)$, we would like to show that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N} P_2^N(k) \xrightarrow[N\to\infty]{a.s.} \sum_{p=0}^{n} \gamma_{p-1}(1)\gamma_{p-1}\left(Q_{p,\eta_{p-1}}(f_{p,n})^2\right). \tag{3.36}$$

Similar to the previous case, we deduce that

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

The convergence (3.36) is then obtained by combining Theorem 3.2.1 and the point (ii) of Lemma 3.7.3. Then, for the term concerning $P_3(k)$, we prove that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N} P_3^N(k) \xrightarrow[N\to\infty]{a.s.} \sum_{p=0}^{n} \gamma_{p-1}(1)\gamma_{p-1}\left(Q_{p,\eta_{p-1}}(f_{p,n})^2\right). \tag{3.37}$$

Notice that

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

Hence, by exploiting the already proved convergence (3.36), it is sufficient to verify that

$$\frac{1}{N}\sum_{k=1}^{(n+1)N}\left(P_3^N(k)-P_2^N(k)\right) \xrightarrow[N\to\infty]{a.s.} 0. \tag{3.38}$$

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

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

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

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

Thanks to Azuma-Hoeffding inequality, one obtains

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

The almost sure convergence (3.38) is then followed from Borel-Cantelli lemma. In conclusion, by combining (3.35),(3.36) and (3.37), one gets

$$\frac{1}{N}\sum_{k=1}^{(n+1)N}\left(U_k^N(f)\right)^2 \xrightarrow[N\to\infty]{a.s./\mathbf{P}} \sigma_{\gamma_n}^2(f). \tag{3.39}$$

Next, the induction hypothesis, Lemma 3.7.5 and Theorem 3.2.1 ensure that

$$D_{p,n}^N(f) = \underbrace{\left[\eta_{p-1}^N(G_{p-1}) - \eta_{p-1}(G_{p-1})\right]\mathbf{1}_{\tau_N \geq p-1}}_{\mathcal{O}_{\mathbf{P}}\left(\frac{1}{\sqrt{N}}\right)} \underbrace{\gamma_{p-1}^N(\dot{Q}_p - Q_p)(f_{p,n})\mathbf{1}_{\tau_N \geq p-1}}_{o_{\mathbf{P}}(1)}, \tag{3.40}$$

whence

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

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

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

This ends the proof of Theorem 3.2.2.

3.7.5 Proof of Proposition 3.3.1

First, we notice that

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

We start by study the first term on the right-hand side of the equality above. Thanks to Theorem 3.6.1, and by considering the stochastic bound given in Proposition 3.7.7, we have, on the event $\{\tau_N \geq n\}$,

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

Notice that

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

which yields

$$N\left(\gamma_n^N(1)^2 - \Gamma_{n,N}^{\ddagger,(\varnothing)}\right)\mathbf{1}_{\tau\geq n} = \sum_{p=0}^{n}\left(\Gamma_{n,N}^{\ddagger,(p)}(f^{\otimes 2}) - \Gamma_{n,N}^{(\varnothing)}(f^{\otimes 2})\right) + \mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{N}\right). \tag{3.41}$$

Next, by applying the decomposition given in Proposition 3.6.4, we deduce that

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

Combining the two parts, we finally obtain the desired stochastic bound in (3.12). For (3.13), the reasoning is similar by the same algebraic manipulations. The only remark is that due to the "normalization" procedure, the stochastic bound w.r.t. $\mathbb{L}^2$-norm given by Proposition 3.7.7 and Proposition 3.7.8 will be replaced by a weaker version, namely,

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

This is ensured by Theorem 3.2.1.

3.7.6 Proof of Theorem 3.6.1

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

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

we have

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

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

$$\begin{aligned}
&\sum_{b\in\{0,1\}^{n+1}}\left\{\prod_{p=0}^{n}\frac{(N-1)^{1-b_p}}{N}\right\}\bar{\Gamma}_{n,N}^{b}(F)\\
&= \sum_{\ell_0^{[2]}\in(N)^2}\cdots\sum_{\ell_{n-1}^{[2]}\in(N)^2}\left\{\prod_{p=0}^{n-1}\left(\frac{1}{N}\mathbf{1}_{\{A_p^{\ell_{p+1}^1}=A_p^{\ell_{p+1}^2}=\ell_p^1\neq\ell_p^2\}} + \frac{N-1}{N}\mathbf{1}_{\{A_p^{\ell_{p+1}^1}=\ell_p^1\neq A_p^{\ell_{p+1}^2}=\ell_p^2\}}\right)\right\}\\
&\left(\frac{N}{N-1}\right)^n\left\{\frac{N-1}{N}m^{\odot 2}(\mathbf{X_n})C_0(F) + \frac{1}{N}m^{\odot 2}(\mathbf{X_n})C_1(F)\right\}.
\end{aligned}$$

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

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

while, by (3.2),

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

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

3.7.7 Proof of Proposition 3.6.5

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

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

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

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

with

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

Similarly, we also define

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \widetilde{\Lambda}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n^b) = m^{\odot 2}([N]) \cdot \widetilde{\mathbf{H}}_{1,(N)}^{\ddagger,b'|b} \cdot \widetilde{\mathbf{H}}_{2,(N)}^{\ddagger,b'|b} \cdots \widetilde{\mathbf{H}}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n).$$

Remark that $0 = 0 \times \frac{1}{N-1}$. Hence, by definition, we have

$$\widetilde{\Lambda}_{n,(N)}^{\ddagger,b'|b}(\mathbf{F}_n^b) = \left(\frac{1}{N-1} \right)^{|b'-b|} \widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n^b). \quad a.s.$$

Next, we consider the binary decomposition w.r.t. a coalescence indicator b. More precisely,

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

Therefore, it suffices to verify that for any coalescence indicator b' and b, we have

$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) = \mathcal{O}_{\mathbb{L}^2}(1).$$

By definition, if there exists $n_0 \geq 0$ such that

$$b'_{n_0} = 0 \quad \text{and} \quad b_{n_0} = 1,$$

we have

$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) \equiv 0.$$

If not, let us consider the mapping $\phi : \{0,1\}^2 \mapsto \{0,1\}$ defined by

$$\phi(0,0) = 0, \ \phi(1,0) = 1 \quad \text{and} \quad \phi(1,1) = 1.$$

We also denote $b_\phi := \big(\phi(b_0, b'_0), \phi(b_1, b'_1), \ldots, \phi(b_n, b'_n)\big)$. It is then easily checked that

$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) = \widetilde{\Lambda}_n^{\dagger,b_\phi}(\mathbf{F}_n) = \widetilde{\Gamma}_n^{\dagger,b_\phi}(F).$$

As a consequence, thanks to Proposition 3.7.8, we have

$$\widetilde{\Lambda}_n^{\ddagger,b'|b}(\mathbf{F}_n) = \mathcal{O}_{\mathbb{L}^2}(1).$$

This is sufficient to end the proof of Proposition 3.6.5.

3.7.8 Technical results

In this section, we list some technical results in support of the proofs given in the following sections. We remark that Lemma 3.7.5 serves as a technical lemma, designed to prove Proposition 3.7.4 by induction. Since the latter one is proved to be true, the hypothesis in Lemma 3.7.5 can thus be removed. This is why in the Proposition 3.7.5, it can be used without induction argument. In the proof of Proposition 3.7.7, we do not give the finest analysis, which is done later in the proof of Lemma 3.7.10. This organization is due to the complication of the notation in the present work. Since the rougher analysis in the proof of Proposition 3.7.7 is more straightforward than the finer version in Lemma 3.7.10, we consider it to be a good warm-up to the techniques involved in this section, which are highly repetitive in regard of the application of the pivotal decomposition (3.63).

Lemma 3.7.4. *For any test function $f, g \in \mathcal{B}_b(E_0)$ and for both $b_0 = 0$ and $b_0 = 1$, we have*

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

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

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

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

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

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

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

which guarantees

$$\mathbf{E}\left[\eta_0^N(f)\mathbf{1}_{\tau_N \geq 0} - \eta_0(f)\right] = \mathcal{O}\left(\frac{1}{N}\right), \tag{3.42}$$

whence, by Cauchy-Schwarz inequality,

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

For the case $b_0 = 0$, since

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

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

Lemma 3.7.5. *If for all $0 \leq p \leq n-1$, we have*

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

then, we also have

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

Proof. By definition, we have

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

For $n \geq 1$, thanks to the bias-martingale decomposition (3.31), the almost sure boundness (3.30) and Azuma-Hoeffding inequality, we have

$$\mathbf{P}\left(\gamma_n^N(1)\mathbf{1}_{\tau_N \geq n} < \frac{\gamma_n(1)}{2}\right) \leq \underbrace{\mathbf{P}\left(\left|\frac{1}{N}\sum_{k=1}^{(n+1)N} U_k^N(1)\right| > \frac{\gamma_n(1)}{4}\right)}_{\text{exponential decay rate w.r.t. } N} + \mathbf{P}\left(\left|\sum_{p=1}^{n} D_{p,n}^N(1)\right| > \frac{\gamma_n(1)}{4}\right) \tag{3.43}$$

Then, we verify that

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

By Markov's inequality, one has

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

Thanks to Cauchy-Schwarz inequality, one derives

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

In addition, we also have

$$\begin{aligned}
&\left\|\gamma^N_{p-1}(1)\eta^N_{p-1}(\dot{Q}_p - \mathring{Q}_p)\left(Q_{p,n}(1)\right)\mathbf{1}_{\tau_N \geq p-1}\right\|_{\mathbb{L}^2} \\
\leq&\left\|\gamma^N_{p-1}(1)\left(\eta^N_{p-1}\dot{Q}_p\left(Q_{p,n}(1)\right)\mathbf{1}_{\tau_N \geq p-1} - \eta_{p-1}\dot{Q}_p\left(Q_{p,n}(1)\right)\right)\mathbf{1}_{\tau_N \geq p-1}\right\|_{\mathbb{L}^2} \\
&+\left\|\gamma^N_{p-1}(1)\left(\eta^N_{p-1}\mathring{Q}_p\left(Q_{p,n}(1)\right)\mathbf{1}_{\tau_N \geq p-1} - \eta_{p-1}\mathring{Q}_p\left(Q_{p,n}(1)\right)\right)\mathbf{1}_{\tau_N \geq p-1}\right\|_{\mathbb{L}^2}.
\end{aligned}$$

Therefore, consider the hypothesis, and the fact that

$$\gamma^N_{n-1}(1) \leq 1, \qquad a.s.$$

we get

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

which yields

$$\mathbf{P}\left(\gamma^N_n(1)\mathbf{1}_{\tau_N \geq n} < \frac{\gamma_n(1)}{2}\right) = \mathcal{O}\left(\frac{1}{N}\right). \tag{3.44}$$

Next, since

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

one derives that

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

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

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

□

Proposition 3.7.4. *For any test function $f \in \mathcal{B}_b(E_n)$, we have*

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

In particular, one also has

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

Proof. The proof is done by induction. For $n = 0$, the stochastic bound is clear for the bounded i.i.d. random variables. which is guaranteed by Lemma 3.7.4. For step $n \geq 1$, we suppose that

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

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

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

By the definition of the absorbing time τ_N, one has

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

whence

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

Then, by the fact that

$$\begin{aligned}\mathbf{1}_{\tau_N \geq n} = \mathbf{1}_{\tau_N \geq n} - \mathbf{1}_{\Omega_n^N} + \mathbf{1}_{\Omega_n^N} &\leq \left|\mathbf{1}_{\tau_N \geq n} - 1 + 1 - \mathbf{1}_{\Omega_n^N}\right| + \mathbf{1}_{\Omega_n^N} \\ &\leq \left|1 - \mathbf{1}_{\tau_N \geq n}\right| + \left|1 - \mathbf{1}_{\Omega_n^N}\right| + \mathbf{1}_{\Omega_n^N} \leq 2\mathbf{1}_{(\Omega_n^N)^c} + \mathbf{1}_{\Omega_n^N},\end{aligned} \tag{3.45}$$

we obtain

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

Then, by applying the induction hypothesis, and thanks to a by-product (3.44) of Lemma 3.7.5, the inequality above leads to

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

It is thus sufficient to verify that

$$\left\|\left(\eta_n^N(f) - \eta_n(f)\right)\mathbf{1}_{\Omega_n^N}\right\|_{\mathbb{L}^2} = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{3.46}$$

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

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

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

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

As a consequence, to prove that

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

and

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

One only needs to verify that

$$\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n}-\gamma_n(f)=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right). \tag{3.47}$$

According to the bias-martingale decomposition (3.31), we have

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

whence

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

By definition, we have

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

By applying the induction hypothesis, one obtains

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

which gives

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

and, by Cauchy-Schwarz inequality,

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

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

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

where the convergence is a by-product (3.39) of the proof of Theorem 3.2.2 and dominated convergence theorem. Hence, we obtain

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

In summary, we have

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

which, thanks to Cauchy-Schwarz inequality, leads to

$$\left\|\gamma_n^N(f)\mathbf{1}_{\tau_N\geq n} - \gamma_n(f)\right\|_{\mathbb{L}^2} = \mathcal{O}\left(\frac{1}{\sqrt{N}}\right). \tag{3.48}$$

This ends the verification of (3.47) and the proof of this proposition.

□

Proposition 3.7.5 ($\mathbb{L}^2$-propagation of chaos)**.** *For any test function $f, g \in \mathcal{B}_b(E_n)$, we have*

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

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

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

Notice that

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

Thanks to Proposition 3.7.4 and Lemma 3.7.5, one derives

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

which implies that

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

Next, considering the decomposition (3.2), we deduce that

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

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

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

In addition, since

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

it is then sufficient to verify that

$$\forall f\in\mathcal{B}_b(E_n),\quad\left(\eta_n^N(f)-\eta_n(f)\right)\mathbf{1}_{\tau_N\geq n}=\mathcal{O}_{\mathbb{L}^2}\left(\frac{1}{\sqrt{N}}\right).\tag{3.49}$$

which is guaranteed by Proposition 3.7.4 since

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

The proof is then finished. □

Proposition 3.7.6 (Biasedness). *For any test function $f\in\mathcal{B}_b(E_n)$, we have*

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

In particular, we also have

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

Remark. Different from the other technical results, the order of the bias given in this proposition will not be used to prove the consistency of the variance estimator. They are put in this section simply because we think the order of bias is important but not as relevant in the present work, where the most results we discussed are "short term" asymptotic properties of the IPS. In addition, we want to mention that by the same strategy, one can obtain an explicit bound w.r.t. both n and N for the bias. The main difference from the classic Feynman-Kac particle models discussed in [DM04] is the "lack-of-martingale" or, said differently, the bias-martingale structure (cf. (3.31)). As a consequence, the decay rate of the absorbing time is not exponential w.r.t. N any more. Instead, it is replace by $\mathcal{O}(1/N)$, as stated in Lemma 3.7.5. This is why the order of bias w.r.t. N is not affected. The "lack-of-martingale" structure also requires an induction in order to deal with the bias term encountered in the bias-martingale decomposition (3.31). This technique is frequently used in the adaptive SMC context (cf. [DG19](Chapter 2)).

Proof. The proof is done by induction. Thanks to a by-product (3.42) of Lemma 3.7.4, we have

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

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

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

By the bias-martingale decomposition (3.31), it gives

$$\forall\psi\in\mathcal{B}_b(E_n),\quad \mathbf{E}\left[\gamma_n^N(\varphi)\mathbf{1}_{\tau_N\geq n-1}-\gamma_n(f)\right]=\mathcal{O}\left(\frac{1}{N}\right). \tag{3.50}$$

Next, standard calculations give

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

with

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

Remark that, by definition,

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

Then, by applying Lemma 3.7.5, we noticed that

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

Mutatis mutandis, one also has

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

Therefore, considering the induction hypothesis (3.50), we only have to show that

$$\begin{aligned}&\left(\frac{\gamma_n(1)}{\gamma_n^N(1)}-1\right)\left(\gamma_n^N\left(f_n\right)-\gamma_n\left(f_n\right)\right)\mathbf{1}_{\tau_N\geq n}\\=&-\left(\frac{\gamma_n^N(1)\mathbf{1}_{\tau_N\geq n}-\gamma_n(1)}{\gamma_n^N(1)}\right)\left(\gamma_n^N\left(f_n\right)\mathbf{1}_{\tau_N\geq n}-\gamma_n\left(f_n\right)\right)\mathbf{1}_{\tau_N\geq n}=\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right).\end{aligned} \tag{3.51}$$

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

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

and we have $\mathbf{1}_{\Omega_n^N} \leq \mathbf{1}_{\tau_N \geq n} \leq 2\mathbf{1}_{(\Omega_n^N)^c} + \mathbf{1}_{\Omega_n^N}$. Notice that, by Lemma 3.7.5 and the definition of f_n, one has

$$\begin{aligned}
&\left(\frac{\gamma_n(1)}{\gamma_n^N(1)} - 1\right)\left(\gamma_n^N(f_n) - \gamma_n(f_n)\right)\mathbf{1}_{(\Omega_n^N)^c} \\
\leq &\left|\eta_n^N(f)\mathbf{1}_{\tau_N \geq n} - \eta_n(f)\right|\mathbf{1}_{(\Omega_n^N)^c} + \left|\gamma_n^N(f_n)\mathbf{1}_{\tau_N \geq n} - \gamma_n(f_n)\right|\mathbf{1}_{(\Omega_n^N)^c} \\
\leq &4\|f\|_\infty \mathbf{1}_{(\Omega_n^N)^c} = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right).
\end{aligned} \tag{3.52}$$

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

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

Thus, thanks to a by-product (3.48) in the proof of Proposition 3.7.4, we obtain

$$\left(\frac{\gamma_n(1)}{\gamma_n^N(1)} - 1\right)\left(\gamma_n^N(f_n) - \gamma_n(f_n)\right)\mathbf{1}_{\Omega_n^N} = \mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{N}\right). \tag{3.53}$$

Finally, combining both (3.52) and (3.53) terminates the verification of (3.51), which also ends the proof of Proposition 3.7.6.

□

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

$$[N]_p^q := \left\{(i_1, \ldots, i_q) \in [N]^q : \mathrm{Card}\{i_1, \ldots, i_q\} = p\right\}. \tag{3.54}$$

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

$$\left((N)^2\right)^{\times q} := \underbrace{(N)^2 \times (N)^2 \times \cdots \times (N)^2}_{q \text{ times}}. \tag{3.55}$$

With a slight abuse of notation, we admit that

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

We also adopt the notation introduced in Section 3.6.5. Fixing some $b \in \{0,1\}^{n+1}$, we denote

$$\Lambda_n^{\ddagger,b}[\ell_n^{[2]}] := \frac{1}{N(N-1)} \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{\prod_{p=0}^{n-1} \mathbf{G}_{\mathbf{p}}^{\ddagger}(\mathbf{X_p})\lambda_p^b(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]})\right\}, \tag{3.56}$$

with the convention

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

It is readily checked that

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

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

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \Gamma^{\ddagger,b}_{n,N}(F) = \sum_{\ell_n^{[2]} \in (N)^2} \Lambda^{\ddagger}_n[\ell_n^{[2]}] C_{b_n}(F)(X_n^{\ell_n^{[2]}}), \tag{3.57}$$

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

$$\widetilde{\Lambda}^{\dagger,b}_n[\ell_n^{[2]}] := \frac{1}{N(N-1)} \sum_{\ell_{0:n-1}^{[2]} \in ((N)^2)^{\times n}} \left\{ \prod_{p=0}^{n-1} \widetilde{\mathbf{G}}^{\dagger,b_p}_{\mathbf{p}}(\ell_{p:p+1}^{[2]}, \mathbf{B_p}, \mathbf{X_p}) \lambda_p^{(\varnothing)}(A_p^{\ell_{p+1}^{[2]}}, \ell_p^{[2]}) \right\}, \tag{3.58}$$

with the convention

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

We also have the decomposition

$$\widetilde{\Lambda}^{\dagger,b}_n[\ell_n^{[2]}] = \sum_{\ell_{n-1}^{[2]} \in (N)^2} \widetilde{\Lambda}^{\dagger,b}_{n-1}[\ell_{n-1}^{[2]}] \widetilde{\mathbf{G}}^{\dagger,b_{n-1}}_{\mathbf{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}}) \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}). \tag{3.59}$$

As is shown in the previous case (3.57), ${}^{\dagger}\widetilde{\Gamma}^{b}_{n,N}$ admits the following alternative representation:

$$\forall F \in \mathcal{B}_b(E_n^2), \quad \widetilde{\Gamma}^{\dagger,b}_{n,N}(F) = \sum_{\ell_n^{[2]} \in (N)^2} \widetilde{\Lambda}^{\dagger,b}_n[\ell_n^{[2]}] F(X_n^{\ell_n^{[2]}}), \tag{3.60}$$

which covers the case $n = 0$.

Proposition 3.7.7. *For any coalescence indicator $b \in \{0,1\}^{n+1}$, we have*

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

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

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

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

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

By the alternative representation (3.57), for all $N \geq 4$, we have

$$\begin{aligned} &\mathbf{E}\left[\Gamma^{\ddagger,b}_{n,N}(1)^2 \mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{G}^N_{n-1} \right] \\ &= \sum_{(\ell_{n-1}^{[2]}, \ell_{n-1}^{\prime[2]}) \in ((N)^2)^{\times 2}} \Lambda^{\ddagger,b}_{n-1}[\ell_{n-1}^{[2]}] \Lambda^{\ddagger,b}_{n-1}[\ell_{n-1}^{\prime[2]}] \mathbf{1}_{\tau_N \geq n-1} \\ &\quad \mathbf{E}\left[\sum_{(\ell_n^{[2]}, \ell_n^{\prime[2]}) \in ((N)^2)^{\times 2}} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2 \lambda^b_{n-1}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \lambda^b_{n-1}(A_{n-1}^{\ell_n^{\prime[2]}}, \ell_{n-1}^{\prime[2]}) \,\middle|\, \mathcal{G}^N_{n-1} \right], \end{aligned} \tag{3.61}$$

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

$$\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N))^{\times 2}} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell_n'^{[2]}}_{n-1},\ell'^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right]\mathbf{1}_{\tau_N=n-1} = 0.$$

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

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

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

$$\begin{aligned}
&m(\mathbf{X_{n-1}})(G_{n-1})\sum_{i=1}^N\mathbf{P}\left(A^i_{n-1}=j\mid\mathcal{G}^N_{n-1}\right)\\
=&m(\mathbf{X_{n-1}})(G_{n-1})G_{n-1}(X^j_{n-1})+\sum_{i=1}^N(1-G_{n-1}(X^i_{n-1}))\frac{G_{n-1}(X^j_{n-1})}{N}\\
=&G_{n-1}(X^j_{n-1})\leq 1.\quad a.s.
\end{aligned}\tag{3.62}$$

With the notation introduced in (3.54) and (3.55), for $N\geq 4$, we have the decomposition

$$((N)^2)^{\times 2}=\left(((N)^2)^{\times 2}\cap[N]^4_2\right)\cup\left(((N)^2)^{\times 2}\cap[N]^4_3\right)\cup(N)^4.\tag{3.63}$$

The rest of the proof consists in studying the term

$$\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N))^{\times 2}} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})\lambda^b_{n-1}(A^{\ell_n'^{[2]}}_{n-1},\ell'^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right]$$

with respect to the decomposition above and the bound given in (3.62).

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

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

$$\left(A^{\ell^1_n}_{n-1},A^{\ell^2_n}_{n-1},A^{\ell'^1_n}_{n-1},A^{\ell'^2_n}_{n-1}\right).$$

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

$$\underbrace{\mathrm{C}^2_4/2}_{\substack{\text{possible choices to assign}\\ \text{two distinct couples in }(4)^4.}}-\underbrace{2/2}_{\text{limitation by }((N)^2)^{\times 2}.}=2.\tag{3.64}$$

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

$$A^{\ell^1_n}_{n-1}=A^{\ell'^1_n}_{n-1},\ A^{\ell^2_n}_{n-1}=A^{\ell'^2_n}_{n-1}\quad\text{and}\quad A^{\ell^1_n}_{n-1}=A^{\ell'^2_n}_{n-1},\ A^{\ell^2_n}_{n-1}=A^{\ell'^1_n}_{n-1}.$$

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

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

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

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

Now, let us deduce that

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

By the conditional independence between $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n^2}$, one deduces

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} m(\mathbf{X_{n-1}})(G_n)^2 \mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1\right\}} \mathbf{1}_{\left\{A_{n-1}^{\ell_n^2}=b_{n-1}\ell_{n-1}^1+(1-b_{n-1})\ell_{n-1}^2\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
= &\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\qquad \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A_{n-1}^{\ell_n^2}=b_{n-1}\ell_{n-1}^1+(1-b_{n-1})\ell_{n-1}^2\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \qquad (3.65) \\
\leq &\sum_{\ell_n^1\in[N]} \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\sum_{\ell_n^2\in[N]} \mathbf{E}\left[m(\mathbf{X_{n-1}})(G_n)\mathbf{1}_{\left\{A_{n-1}^{\ell_n^2}=b_{n-1}\ell_{n-1}^1+(1-b_{n-1})\ell_{n-1}^2\right\}} \,\middle|\, \mathcal{G}_{n-1}^N\right].
\end{aligned}$$

Combined to (3.62), one gets

$$\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G_{n-1}^{\ddagger}}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq 1^2 \times \frac{N}{N-1}. \quad a.s. \qquad (3.66)$$

which gives

$$\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1}) \,\middle|\, \mathcal{G}^N_{n-1}\right] \leq \left(\frac{N}{N-1}\right)^2. \quad a.s. \tag{3.67}$$

Considering the choices of assignments mentioned above, one finally gets

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

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

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

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

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

$$\underbrace{\mathrm{C}^3_4}_{\substack{\text{possible choices to divise}\\ (4)^4 \text{ into three distinct parts.}}} - \underbrace{0}_{\text{limitation by } ((N)^2)^{\times 2}.} = 4. \tag{3.68}$$

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

between $A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}$ and $A_{n-1}^{\ell_n'^1}$, one derives

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

Again, combined with (3.62), one deduces

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

Considering the number of assignments, we obtain

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

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

In this case, all the random variables

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

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

(3.62), one gets

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

Combining the three cases discussed above, we safely deduce that

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

Now, let us go back to (3.61), by taking expectation on both sides, we have, for $N\geq 4$,

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

This closes the proof of Proposition 3.7.7. □

Proposition 3.7.8. *For any coalescence indicator $b\in\{0,1\}^{n+1}$, we have*

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

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

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

Remark. Before starting the proof of Proposition 3.7.8, we would like to mention that the techniques involved are similar but a little bit different from the ones in the proof of Proposition 3.7.7. First, $\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}$ is not nonnegative in general except for the case where G_n is indicator function for all $n \geq 0$. In addition, it is not obvious that

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

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

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

is not a.s. bounded in general, which is the main difficult part in the following technical results. Hence, unlike the previous case shown in the proof of Proposition 3.7.7, one should be extremely careful where dealing with the bound associated to the term

$$\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]}, \mathbf{B_n}, \mathbf{X_n})\right| \lambda_n^{(\varnothing)}(A_n^{\ell_{n+1}^{[2]}}, \ell_n^{[2]})\mathbf{1}_{\tau_N \geq n}.$$

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

Lemma 3.7.6. *For any nonnegative real numbers $a, b, c \in \mathbf{R}$, if $a \leq b$, $b > 0$ and $c \geq 0$, then, we have*

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

Proof. Direct calculation gives

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

The conclusion follows. □

Lemma 3.7.7. *For any $\ell_n^{[2]}$ and $\ell_{n+1}^{[2]} \in (N)^2$, we have*

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

and

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

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

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

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

is always well-defined on the event $\{\tau_N \geq n\}$ with the convention (3.1). In fact, when

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

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

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

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

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

The conclusion follows. □

Lemma 3.7.8. *For any $\ell_n^{[2]} \in (N)^2$, and for any coalescence indicator $b' \in \{0,1\}^{n+1}$ and $b_n \in \{0,1\}$, we have almost surely*

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

In particular, we have

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

Remark. We remark that in the definition of $\widetilde{\Gamma}_{n,N}^{\dagger,b}$, we do not need to investigate $\lambda_n^{b'}$ for a different coalescence indicator b' in general. The reason that b' is not set to be $(\varnothing)$ lies in the fact that Lemma 3.7.8 is applied in the proof of Lemma 3.7.9.

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

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

and given $\overline{\mathcal{W}}_n^N$, one has

$$\forall \ell\in[N],\quad A_n^\ell \sim B_n^\ell\delta_\ell(\cdot) + (1-B_n^\ell)\sum_{k=1}^N\frac{G_n(X_n^k)}{Nm(\mathbf{X_n})(G_n)}\delta_k(\cdot).$$

By definition, one has

$$\begin{aligned}
\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| \leq& B_n^{\ell^1_{n+1}}B_n^{\ell^2_{n+1}}m(\mathbf{X_n})(G_n^2)\\
&+B_n^{\ell^1_{n+1}}(1-B_n^{\ell^2_{n+1}})m(\mathbf{X_n})(G_n)\frac{\left|G_n(X_n^{\ell^1_n})m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))\right|}{\sum_{k\neq\ell^1_n}\left(1-G_n(X_n^k)\right)/N}\\
&+B_n^{\ell^2_{n+1}}(1-B_n^{\ell^1_{n+1}})m(\mathbf{X_n})(G_n)\frac{\left|G_n(X_n^{\ell^2_n})m(\mathbf{X_n})(G_n) - m(\mathbf{X_n}(G_n^2))\right|}{\sum_{k\neq\ell^2_n}\left(1-G_n(X_n^k)\right)/N},
\end{aligned}$$

whence

$$\begin{aligned}
&\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| \\
\leq & B_n^{\ell_{n+1}^1}B_n^{\ell_{n+1}^2}m(\mathbf{X_n})(G_n^2)+\frac{2B_n^{\ell_{n+1}^1}(1-B_n^{\ell_{n+1}^2})m(\mathbf{X_n})(G_n)}{\sum_{k\neq\ell_n^1}\left(1-G_n(X_n^k)\right)/N}+\frac{2B_n^{\ell_{n+1}^2}(1-B_n^{\ell_{n+1}^1})m(\mathbf{X_n})(G_n)}{\sum_{k\neq\ell_n^2}\left(1-G_n(X_n^k)\right)/N}.
\end{aligned}$$

Since $B_n^{\ell_n^1}$ and $B_n^{\ell_n^2}$ are both $\{0,1\}$-valued, we deduce that

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_{n+1}^{[2]}\in[N]^2}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|\lambda_n^{b'}(A_n^{\ell_{n+1}^{[2]}},\ell_n^{[2]})\mathbf{1}_{\tau_N\geq n}\,\middle|\,\overline{\mathcal{W}}_n^N\right] \\
\leq & B_n^{\ell_n^1}B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2}m(\mathbf{X_n})(G_n^2)\mathbf{1}_{\tau_N\geq n} \\
&+B_n^{\ell_n^1}\sum_{\ell_{n+1}^2=1}^N(1-B_n^{\ell_{n+1}^2})\frac{2G_n(X_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2})}{\sum_{k\neq\ell_n^1}1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n} \\
&+B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2}\sum_{\ell_{n+1}^1=1}^N(1-B_n^{\ell_{n+1}^1})\frac{2G_n(X_n^{\ell_n^1})}{\sum_{k\neq\ell_n^2}1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n}.
\end{aligned}$$

By applying Lemma 3.7.6 and considering the convention (3.1), we have

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_{n+1}^{[2]}\in[N]^2}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|\lambda_n^{b'}(A_n^{\ell_{n+1}^{[2]}},\ell_n^{[2]})\mathbf{1}_{\tau_N\geq n}\,\middle|\,\overline{\mathcal{W}}_n^N\right] \\
\leq & B_n^{\ell_n^1}B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2}m(\mathbf{X_n})(G_n^2)\mathbf{1}_{\tau_N\geq n} \\
&+B_n^{\ell_n^1}\sum_{\ell_{n+1}^2=1}^N(1-B_n^{\ell_{n+1}^2})\frac{2G_n(X_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2})+1-G_n(X_n^{\ell_n^1})}{\sum_{k=1}^N1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n} \\
&+B_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2}\sum_{\ell_{n+1}^1=1}^N(1-B_n^{\ell_{n+1}^1})\frac{2G_n(X_n^{\ell_n^1})+1-G_n(X_n^{\ell_n^2})}{\sum_{k=1}^N1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n}.
\end{aligned}$$

The simple fact $0=B_n^k(1-B_n^k)\mathbf{1}_{\tau_N\geq n}\leq G_n(X_n^k)(1-G_n(X_n^k)\mathbf{1}_{\tau_N\geq n})$ for any $k\in[N]$ yields

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_{n+1}^{[2]}\in[N]^2}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|\lambda_n^{b'}(A_n^{\ell_{n+1}^{[2]}},\ell_n^{[2]})\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_n^N\right] \\
\leq & \left(G_n(X_n^{\ell_n^1})G_n(X_n^{\ell_n^2})\vee G_n(X_n^{\ell_n^1})\right)m(\mathbf{X_n})(G_n^2)\mathbf{1}_{\tau_N\geq n} \\
&+G_n(X_n^{\ell_n^1})\sum_{\ell_{n+1}^2=1}^N(1-G_n(X_n^{\ell_{n+1}^2}))\frac{2G_n(X_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2})+1-G_n(X_n^{\ell_n^1})}{\sum_{k=1}^N1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n} \\
&+G_n(X_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2})\sum_{\ell_{n+1}^1=1}^N(1-G_n(X_n^{\ell_{n+1}^1}))\frac{2G_n(X_n^{\ell_n^1})+1-G_n(X_n^{\ell_n^2})}{\sum_{k=1}^N1-G_n(X_n^k)}\mathbf{1}_{\tau_N\geq n} \\
= & \left(G_n(X_n^{\ell_n^1})m(\mathbf{X_n})(G_n^2)+4G_n(X_n^{\ell_n^1})G_n(X_n^{b'_n\ell_n^1+(1-b'_n)\ell_n^2})+2\right)\mathbf{1}_{\tau_N\geq n}\leq 7.\quad a.s.
\end{aligned}$$

By definition, since

$$\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,0}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| = \mathbf{G}_{\mathbf{n}}^{\ddagger}(\mathbf{X_n}) + \frac{1}{N-1}\left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right|,$$

the analysis for the case $b_n = 0$ is the combination of the case $b_n = 1$ and the similar reasoning in (3.65), namely, a direct consequence of (3.62). This terminates the proof of Lemma 3.7.8.

□

Lemma 3.7.9. *For any $\ell_n^{[2]} \in (N)^2$ and for any coalescence indicator $b \in \{0,1\}^{n+1}$, we have almost surely*

$$\forall \ell_{n+1}^2 \in [N], \quad \sup_{N>1} \mathbf{E}\left[\sum_{\ell_{n+1}^1 \in [N]} \left|\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})\right| \mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1} \mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{W}_n^N\right] < +\infty, \tag{3.69}$$

as well as

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

Proof. By symmetry of the definition of $\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,b_n}(\ell_{n:n+1}^{[2]},\mathbf{B_n},\mathbf{X_n})$, it suffices to verify (3.69). In fact, by simple observation, one has

$$\mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1} = \mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1} \mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1} \leq \mathbf{1}_{A_n^{\ell_{n+1}^1}=\ell_n^1} \sum_{\ell_{n+1}^2=1}^{N} \mathbf{1}_{A_n^{\ell_{n+1}^2}=\ell_n^1}. \quad a.s.$$

Therefore, one only needs to show that

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

which, by taking $b' = (n)$, is guaranteed by Lemma 3.7.8. This terminates the proof of Lemma 3.7.9. □

Proof of Proposition 3.7.8. Now, we start the proof by induction. It is trivial for the step $n = 0$ as $\widetilde{\Gamma}_{0,N}^{\dagger,b}(1) = 1$. For step $n \geq 1$, we suppose that

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

By the alternative representation (3.60), for all $N \geq 4$, we have

$$\begin{aligned}
&\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(1)^2 \mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{W}_{n-1}^N\right] \\
&= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]}) \in ((N))^{\times 2}} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}] \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}] \mathbf{1}_{\tau_N \geq n-1} \\
&\quad \mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]}) \in ((N)^2)^{\times 2}} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right. \\
&\qquad\qquad \left. \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{W}_{n-1}^N\right],
\end{aligned} \tag{3.70}$$

since by definition of $\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}$, one has

$$\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\ \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\Big|\, \mathcal{W}_{n-1}^N\Bigg]\mathbf{1}_{\tau_N=n-1}=0.$$

As we have mentioned several times, in the following part of the proof, we omit the notation $\mathbf{1}_{\tau_N\geq n-1}$. Similar as in the proof of Proposition 3.7.7, the rest of the reasoning relies on the decomposition (3.63).

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

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

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

As we have already mentioned in (3.64), there are 2 possible assignments such that we can fix two distinct random variables within the tuple above. Without loss of generality, we suppose that $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n'^1}$ are two distinct random variables, and we fix one of these two assignments. Then, when $(\ell_n^1,\ell_n'^1)$ varies freely in $(N)^2$, the values of $A_{n-1}^{\ell_n^2}$ and $A_{n-1}^{\ell_n'^2}$ will be a.s. determined by the chosen assignment and the values of $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n'^1}$.

$$\begin{aligned}&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\ &\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\Big|\, \overline{\mathcal{W}}_{n-1}^N\Bigg] \\ =&\mathbf{E}\Bigg[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\ &\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\Big|\, \overline{\mathcal{W}}_{n-1}^N\Bigg]\end{aligned} \tag{3.71}$$

Given $\overline{\mathcal{W}}_{n-1}^N$, by the conditional independence of $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n'^1}$ under the chosen assignment, we have

$$\begin{aligned}&\mathbf{E}\Bigg[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\ &\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\Big|\, \overline{\mathcal{W}}_{n-1}^N\Bigg] \\ \leq&\mathbf{E}\Bigg[\sum_{\ell_n^1\in[N]} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right|\mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1\right\}} \,\Big|\, \overline{\mathcal{W}}_{n-1}^N\Bigg] \\ &\mathbf{E}\Bigg[\sum_{\ell_n'^1\in[N]} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right|\mathbf{1}_{\left\{A_{n-1}^{\ell_n'^1}=\ell_{n-1}'^1\right\}} \,\Big|\, \overline{\mathcal{W}}_{n-1}^N\Bigg].\end{aligned} \tag{3.72}$$

Then, by applying Lemma 3.7.9, one gets

$$\mathbf{E}\left[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \right.$$
$$\left. \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{W}_{n-1}^N\right] = \mathcal{O}_{a.s.}(1).$$

Since the number of different assignments 2 does not depend on N, one deduces that

$$\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \right.$$
$$\left. \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{W}_{n-1}^N\right] = \mathcal{O}_{a.s.}(1).$$

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

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

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

As is calculated in (3.68), the number of assignment is 4 at this time. Let us fix one assignment. We suppose that $A_{n-1}^{\ell_n^{[2]}}$ and $A_{n-1}^{\ell_n'^1}$ are three distinct random variables. Then, whilst $(\ell_n^{[2]},\ell_n'^1)$ varies freely in $(N)^3$, the value of $A_{n-1}^{\ell_n'^2}$ is a.s. determined by the chosen assignment and the values of $(A_{n-1}^{\ell_n^{[2]}}, A_{n-1}^{\ell_n'^1})$. Given $\overline{\mathcal{W}}_{n-1}^N$, by the conditional independence between $A_{n-1}^{\ell_n^1}, A_{n-1}^{\ell_n^2}$ and $A_{n-1}^{\ell_n'^1}$, one derives

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \right.\\
&\qquad\qquad \left. \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \overline{\mathcal{W}}_{n-1}^N\right]\\
=&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{1})\in(N)^3} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \right.\\
&\qquad\qquad \left. \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \overline{\mathcal{W}}_{n-1}^N\right]\\
\le&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right| \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \,\middle|\, \overline{\mathcal{W}}_{n-1}^N\right]\\
&\mathbf{E}\left[\sum_{\ell_n'^1\in[N]} \left|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right| \mathbf{1}_{\{A_{n-1}^{\ell_n'^1}=\ell_{n-1}'^1\}} \,\middle|\, \overline{\mathcal{W}}_{n-1}^N\right].
\end{aligned} \tag{3.73}$$

Thanks to Lemma 3.7.8 and Lemma 3.7.9, we get

$$\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{1})\in(N)^3} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\ \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\Big|\,\mathcal{W}_{n-1}^N\Bigg] = \mathcal{O}_{a.s.}(1).$$

Again, since the number of the different assignments 4 does not depend on N, we obtain

$$\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \Big|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\Big| \\ \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\Big|\,\mathcal{W}_{n-1}^N\Bigg] = \mathcal{O}_{a.s.}(1). \tag{3.74}$$

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

In this case, all the random variables

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

are distinct. This time, Lemma 3.7.8 gives directly

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in(N)^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\ &\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\Big|\,\mathcal{W}_{n-1}^N\Bigg] \\
\le&\mathbf{E}\Bigg[\sum_{\ell_n^{[2]}\in[N]^2} \Big|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\Big|\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\,\Big|\,\mathcal{W}_{n-1}^N\Bigg] \\
&\mathbf{E}\Bigg[\sum_{\ell_n'^{[2]}\in[N]^2} \Big|\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\Big|\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\Big|\,\mathcal{W}_{n-1}^N\Bigg] = \mathcal{O}_{a.s.}(1).
\end{aligned} \tag{3.75}$$

Combining the three cases above, we finally obtain

$$\sup_{N>0}\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\ \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\Big|\,\mathcal{W}_{n-1}^N\Bigg] < +\infty. \quad a.s.$$

Returning to (3.70) and induction hypothesis, the verification of step n is then finished, so as the proof of Proposition 3.7.8.

□

Proposition 3.7.9. *For any test function $F \in \mathcal{B}_b(E_n^2)$, we have*

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

Proof. First, by the alternative representation we have introduced in (3.57), we have

$$\begin{aligned}
&\mathbf{E}\left[\Gamma_{n,N}^{\ddagger,b}(F)\mathbf{1}_{\tau_N \geq n} \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
=&\sum_{(\ell_{n-1}^{[2]})\in(N)^2} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\mathbf{1}_{\tau_N \geq n-1} \\
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right],
\end{aligned} \tag{3.76}$$

Thus, it suffices to show that

$$\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N \geq n-1} \\
=&\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}})\mathbf{1}_{\tau_N \geq n-1}.
\end{aligned} \tag{3.77}$$

Alghough it may seem unnecessary, we recall that we have almost surely

$$\begin{aligned}
\frac{N-1}{N}\mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}}) =& B_{n-1}^{\ell_n^1}B_{n-1}^{\ell_n^2}m(\mathbf{X_{n-1}})(G_{n-1})^2 \\
&+ B_{n-1}^{\ell_n^1}(1-B_{n-1}^{\ell_n^2})m(\mathbf{X_{n-1}})(G_{n-1})^2 \\
&+ B_{n-1}^{\ell_n^2}(1-B_{n-1}^{\ell_n^1})m(\mathbf{X_{n-1}})(G_{n-1})^2 \\
&+ (1-B_{n-1}^{\ell_n^1})(1-B_{n-1}^{\ell_n^2})m(\mathbf{X_{n-1}})(G_{n-1})^2.
\end{aligned}$$

Given $\overline{\mathcal{W}}_{n-1}^N$, for any $\ell \in [N]$, the definition of the Feynman-Kac IPS gives

$$(A_{n-1}^\ell, X_n^\ell) \sim B_{n-1}^\ell \delta_\ell(dA_{n-1}^\ell)\dot{M}_n(X_{n-1}^\ell, dX_n^\ell) + (1-B_{n-1}^\ell)\sum_{k=1}^N \frac{\delta_k(dA_{n-1}^\ell)\mathring{Q}_n(X_{n-1}^k, dX_n^\ell)}{Nm(\mathbf{X_{n-1}})(G_{n-1})}. \tag{3.78}$$

(i) Case $b_{n-1} = 0$:

Notice that

$$\begin{aligned}
&\frac{N-1}{N}\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_{\mathbf{n}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \overline{\mathcal{W}}_{n-1}^N\right] \\
=&B_{n-1}^{\ell_{n-1}^1}B_{n-1}^{\ell_{n-1}^2}m^{\otimes 2}(\mathbf{X_{n-1}})(G_{n-1}^{\otimes 2})\dot{M}_n^{\otimes 2}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}}) \\
&+\frac{1}{N}B_{n-1}^{\ell_{n-1}^1}\sum_{\ell_n^2\neq\ell_{n-1}^1}(1-B_{n-1}^{\ell_n^2})m(\mathbf{X_{n-1}})(G_{n-1})(\dot{M}_n\otimes\mathring{Q}_n)C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}}) \\
&+\frac{1}{N}B_{n-1}^{\ell_{n-1}^2}\sum_{\ell_n^1\neq\ell_{n-1}^2}(1-B_{n-1}^{\ell_n^1})m(\mathbf{X_{n-1}})(G_{n-1})(\mathring{Q}_n\otimes\dot{M}_n)C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}}) \\
&+\frac{1}{N^2}\sum_{\ell_n^{[2]}\in(N)^2}(1-B_{n-1}^{\ell_n^1})(1-B_{n-1}^{\ell_n^2})\mathring{Q}_n^{\otimes 2}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}}),
\end{aligned}$$

which yields

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\\
=&\frac{N}{N-1}m^{\otimes 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})\dot{Q}^{\otimes 2}_n C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\
&+\frac{1}{N-1}\sum_{\ell^2_n\neq\ell^1_{n-1}}(1-G_{n-1}(X^{\ell^2_n}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})(\dot{Q}_n\otimes\mathring{Q}_n)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\
&+\frac{1}{N-1}\sum_{\ell^1_n\neq\ell^2_{n-1}}(1-G_{n-1}(X^{\ell^1_n}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})(\mathring{Q}_n\otimes\dot{Q}_n)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\
&+\frac{1}{N(N-1)}\sum_{\ell_n^{[2]}\in(N)^2}(1-G_{n-1}(X^{\ell^1_n}_{n-1}))(1-G_{n-1}(X^{\ell^2_n}_{n-1}))\mathring{Q}^{\otimes 2}_n C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1}),
\end{aligned}
\tag{3.79}
$$

First, by decomposition (3.2), we noticed that

$$
\frac{N}{N-1}m^{\otimes 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1}) = \left(m^{\odot 2}(\mathbf{X_n})(G^{\otimes 2}_{n-1}) + \frac{1}{N-1}m(\mathbf{X_{n-1}})(G^2_{n-1}).\right)
\tag{3.80}
$$

Then, we deduce that

$$
\begin{aligned}
&\frac{1}{N-1}\sum_{\ell^2_n\neq\ell^1_{n-1}}(1-G_{n-1}(X^{\ell^2_n}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})\\
=&\frac{N}{N-1}m^{\otimes 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&-\frac{1}{N-1}(1-G_{n-1}(X^{\ell^1_{n-1}}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})\\
=&m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&+\frac{1}{N(N-1)}\sum_{k=1}^N\left(G_{n-1}(X^{\ell^1_{n-1}}_{n-1})-G_{n-1}(X^k_{n-1})\right)G_{n-1}(X^k_{n-1})\\
=&m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&+\frac{1}{N-1}\left(G_{n-1}(X^{\ell^1_{n-1}}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G^2_{n-1})\right).
\end{aligned}
\tag{3.81}
$$

The exactly same manipulations also give

$$
\begin{aligned}
&\frac{1}{N-1}\sum_{\ell^1_n\neq\ell^2_{n-1}}(1-G_{n-1}(X^{\ell^1_n}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})\\
=&m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\\
&+\frac{1}{N-1}\left(G_{n-1}(X^{\ell^2_{n-1}}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G^2_{n-1})\right).
\end{aligned}
\tag{3.82}
$$

Now, let us put (3.80), (3.81) and (3.82) back into (3.79). One derives

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\\
=&\Bigg\{\Bigg[m^{\odot 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})\dot{Q}^{\otimes 2}_n + m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\dot{Q}_n\otimes\mathring{Q}_n\\
&+ m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))\mathring{Q}_n\otimes\dot{Q}_n + m^{\odot 2}(\mathbf{X_{n-1}})((1-G_{n-1})^{\otimes 2})\mathring{Q}^{\otimes 2}_n\Bigg]\\
&+\frac{1}{N-1}\Bigg[m(\mathbf{X_{n-1}})(G_{n-1})\left((G_{n-1}\times\dot{Q}_n)\otimes\mathring{Q}_n+\mathring{Q}_n\otimes(G_{n-1}\times\dot{Q}_n)\right)\\
&+ m(\mathbf{X_{n-1}})(G^2_{n-1})\left(\dot{Q}^{\otimes 2}_n-\dot{Q}_n\otimes\mathring{Q}_n-\mathring{Q}_n\otimes\dot{Q}_n\right)\Bigg]\Bigg\}(C_{b_n}(F))(X^{\ell_{n-1}^{[2]}}_{n-1}),
\end{aligned}
\tag{3.83}
$$

which, by definition, turns out to be the following equality:

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N\geq n-1}\\
=&\left(\mathbf{Q}^0_{\hat{n}}+\frac{1}{N-1}\widetilde{\mathbf{Q}}^{\dagger,0}_{\hat{n}}\right)\left(C_{b_n}(F)\right)\left(X^{\ell_{n-1}^{[2]}}_{n-1}\right).
\end{aligned}
\tag{3.84}
$$

(ii) Case $b_{n-1}=1$:

By the similar calculations done in the previous case, one has

$$
\begin{aligned}
&\frac{N-1}{N}\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n}}(\mathbf{X_{n-1}})\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\overline{\mathcal{W}}^N_{n-1}\right]\\
=&\ 0\\
&+\frac{1}{N}B^{\ell^1_{n-1}}_{n-1}\sum_{\ell^2_n\neq\ell^1_{n-1}}(1-B^{\ell^2_n}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})(\dot{M}_n\otimes\mathring{Q}_n)C_{b_n}(F)(X^{\ell^{(1,1)}_{n-1}}_{n-1})\\
&+\frac{1}{N}B^{\ell^1_{n-1}}_{n-1}\sum_{\ell^1_n\neq\ell^1_{n-1}}(1-B^{\ell^1_n}_{n-1})m(\mathbf{X_{n-1}})(G_{n-1})(\mathring{Q}_n\otimes\dot{M}_n)C_{b_n}(F)(X^{\ell^{(1,1)}_{n-1}}_{n-1})\\
&+\frac{1}{N^2}\sum_{\ell_n^{[2]}\in(N)^2}(1-B^{\ell^1_n}_{n-1})(1-B^{\ell^2_n}_{n-1})\mathring{Q}^{\otimes 2}_n C_{b_n}(F)(X^{\ell^{(1,1)}_{n-1}}_{n-1}),
\end{aligned}
$$

which yields

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\\
&=0\\
&\quad+\frac{1}{N-1}\sum_{\ell_n^2\neq\ell^1_{n-1}}(1-G_{n-1}(X^{\ell_n^2}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})C_1\left(\dot{Q}_n\otimes\mathring{Q}_n\right)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\
&\quad+\frac{1}{N-1}\sum_{\ell_n^1\neq\ell^1_{n-1}}(1-G_{n-1}(X^{\ell_n^1}_{n-1}))m(\mathbf{X_{n-1}})(G_{n-1})C_1\left(\mathring{Q}_n\otimes\dot{Q}_n\right)C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1})\\
&\quad+\frac{1}{N(N-1)}\sum_{\ell_n^{[2]}\in(N)^2}(1-G_{n-1}(X^{\ell_n^1}_{n-1}))(1-G_{n-1}(X^{\ell_n^2}_{n-1}))C_1\mathring{Q}_n^{\otimes 2}C_{b_n}(F)(X^{\ell^{[2]}_{n-1}}_{n-1}).
\end{aligned}
\tag{3.85}
$$

Then, taking into account the equality (3.81), one obtains a similar equation as (3.83):

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N\geq n-1}\\
=&\Bigg\{\Bigg[m^{\odot 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})C_1\dot{Q}_n^{\otimes 2}+m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))C_1\left(\dot{Q}_n\otimes\mathring{Q}_n\right)\\
&+m^{\odot 2}(\mathbf{X_{n-1}})(G_{n-1}\otimes(1-G_{n-1}))C_1\left(\mathring{Q}_n\otimes\dot{Q}_n\right)+m^{\odot 2}(\mathbf{X_{n-1}})((1-G_{n-1})^{\otimes 2})C_1\mathring{Q}_n^{\otimes 2}\Bigg]\\
&-m^{\odot 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})C_1\dot{Q}_n^{\otimes 2}\\
&+\frac{1}{N-1}\Bigg[m(\mathbf{X_{n-1}})(G_{n-1})C_1\left((G_{n-1}\times\dot{Q}_n)\otimes\mathring{Q}_n+\mathring{Q}_n\otimes(G_{n-1}\times\dot{Q}_n)\right)\\
&-m(\mathbf{X_{n-1}})(G^2_{n-1})C_1\left(\dot{Q}_n\otimes\mathring{Q}_n+\mathring{Q}_n\otimes\dot{Q}_n\right)\Bigg]\Bigg\}(C_{b_n}(F))(X^{\ell^{[2]}_{n-1}}_{n-1})\mathbf{1}_{\tau_N\geq n-1}.
\end{aligned}
\tag{3.86}
$$

By definition, we finally obtain that on the event $\{\tau_N\geq n-1\}$,

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\\
&=\left(\mathbf{Q}^1_{\hat{n}}-m^{\odot 2}(\mathbf{X_{n-1}})(G^{\otimes 2}_{n-1})\mathbf{1}_{\tau_N\geq n-1}C_1\dot{Q}_n^{\otimes}+\frac{1}{N-1}\check{\mathbf{Q}}^{\dagger,1}_{\hat{n}}\right)\left(C_{b_n}(F)\right)(X^{\ell^{[2]}_{n-1}}_{n-1})\\
&=\left(\mathbf{Q}^{\dagger,1}_{\hat{n}}+\frac{1}{N-1}\check{\mathbf{Q}}^{\dagger,1}_{\hat{n}}\right)\left(C_{b_n}(F)\right)(X^{\ell^{[2]}_{n-1}}_{n-1}).
\end{aligned}
\tag{3.87}
$$

Combining the two cases, we conclude that we have proved that

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}^{\ddagger}_{\mathbf{n-1}}(\mathbf{X_{n-1}})^2\lambda^b_{n-1}(A^{\ell_n^{[2]}}_{n-1},\ell^{[2]}_{n-1})C_{b_n}(F)(X^{\ell_n^{[2]}}_n)\,\middle|\,\mathcal{W}^N_{n-1}\right]\mathbf{1}_{\tau_N\geq n-1}\\
=&\mathbf{Q}^{\ddagger,b_{n-1}}_{\hat{n}}\left(C_{b_n}(F)\right)(X^{\ell^{[2]}_{n-1}}_{n-1})\mathbf{1}_{\tau_N\geq n-1}.\quad a.s.
\end{aligned}
\tag{3.88}
$$

In addition, since

$$\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}\left(C_{b_n}(F)\right)(X_{n-1}^{\ell_{n-1}^{[2]}})\mathbf{1}_{\tau_N\geq n-1}$$

is $\mathcal{G}_{n-1}^N$-measurable, the verification of (3.77) is then finished, so as the proof of Proposition 3.7.9. □

Proposition 3.7.10. *For any test function $F\in\mathcal{B}_b(E_n^2)$, we have*

$$\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_{n-1}^N\right]=\widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}C_{b_n}(F)\mathbf{1}_{\tau_N\geq n-1}.$$

Proof. By combining Proposition 3.7.9 and the fact that

$$\begin{aligned}&\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,0}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\\&=\left(\mathbf{G}_{\mathbf{n}}^{\ddagger}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})-\frac{1}{N-1}\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\right)\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}),\ a.s.\end{aligned}$$

it suffices to check the case $b_{n-1}=1$. Similar as in the proof of Proposition 3.7.9, we consider the alternative representation (3.60), which gives

$$\begin{aligned}&\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_{n-1}^N\right]\\=&\sum_{(\ell_{n-1}^{[2]})\in(N)^2}\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\mathbf{1}_{\tau_N\geq n-1}\\&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,1}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{W}_{n-1}^N\right],\end{aligned}\tag{3.89}$$

Thus, it suffices to show that

$$\begin{aligned}&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,1}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{W}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n-1}\\=&\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,1}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}})\mathbf{1}_{\tau_N\geq n-1}.\end{aligned}\tag{3.90}$$

We omit the notation $\mathbf{1}_{\tau_N\geq n-1}$ in the rest of the proof. Recall that

$$\begin{aligned}&\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,1}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\\=&B_{n-1}^{\ell_n^1}B_{n-1}^{\ell_n^2}m(\mathbf{X_{n-1}})(G_{n-1}^2)\\&+B_{n-1}^{\ell_n^1}(1-B_{n-1}^{\ell_n^2})m(\mathbf{X_{n-1}})(G_{n-1})\frac{G_{n-1}(X_{n-1}^{\ell_{n-1}^1})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)}{\sum_{k\neq\ell_{n-1}^1}G_{n-1}(X_{n-1}^k)}\\&+B_{n-1}^{\ell_n^2}(1-B_{n-1}^{\ell_n^1})m(\mathbf{X_{n-1}})(G_{n-1})\frac{G_{n-1}(X_{n-1}^{\ell_{n-1}^2})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)}{\sum_{k\neq\ell_{n-1}^2}G_{n-1}(X_{n-1}^k)}.\end{aligned}$$

Hence, with the definition (3.78), standard calculation gives

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\,\middle|\,\overline{\mathcal{W}}_{n-1}^N\right]\\
=&\Bigg\{B_{n-1}^{\ell_{n-1}^1}B_{n-1}^{\ell_{n-1}^2}m(\mathbf{X_{n-1}})(G_{n-1}^2)\dot{M}_n^{\otimes 2}\\
&+B_{n-1}^{\ell_{n-1}^1}\sum_{\ell_n^2\neq\ell_{n-1}^1}(1-B_{n-1}^{\ell_n^2})\frac{G_{n-1}(X_{n-1}^{\ell_{n-1}^1})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)}{\sum_{k\neq\ell_{n-1}^1}(1-G_{n-1}(X_{n-1}^k))}\dot{M}_n\otimes\mathring{M}_n\\
&+B_{n-1}^{\ell_{n-1}^2}\sum_{\ell_n^1\neq\ell_{n-1}^2}(1-B_{n-1}^{\ell_n^1})\frac{G_{n-1}(X_{n-1}^{\ell_{n-1}^2})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)}{\sum_{k\neq\ell_{n-1}^2}(1-G_{n-1}(X_{n-1}^k))}\mathring{M}_n\otimes\dot{M}_n\\
&\Bigg\}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}}),
\end{aligned}
$$

whence

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]\\
=&\Bigg\{m(\mathbf{X_{n-1}})(G_{n-1}^2)\dot{Q}_n^{\otimes 2}\\
&+\left(G_{n-1}(X_{n-1}^{\ell_{n-1}^1})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)\right)\dot{Q}_n\otimes\mathring{Q}_n\\
&+\left(G_{n-1}(X_{n-1}^{\ell_{n-1}^2})m(\mathbf{X_{n-1}})(G_{n-1})-m(\mathbf{X_{n-1}})(G_{n-1}^2)\right)\mathring{Q}_n\otimes\dot{Q}_n\\
&\Bigg\}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}}),
\end{aligned}
$$

which yields

$$
\begin{aligned}
&\mathbf{E}\left[\sum_{\ell_n^{[2]}\in(N)^2}\widetilde{\mathbf{G}}_{\mathbf{n}}^{\dagger,1}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n-1}\\
=&\Bigg\{m(\mathbf{X_{n-1}})(G_{n-1})\left((G_{n-1}\times\dot{Q}_n)\otimes\mathring{Q}_n+\mathring{Q}_n\otimes(G_{n-1}\times\dot{Q}_n)\right)\\
&+m(\mathbf{X_{n-1}})(G_{n-1}^2)\left(\dot{Q}_n^{\otimes 2}-\dot{Q}_n\otimes\mathring{Q}_n-\mathring{Q}_n\otimes\dot{Q}_n\right)\Bigg\}C_{b_n}(F)(X_{n-1}^{\ell_{n-1}^{[2]}})\mathbf{1}_{\tau_N\geq n-1}.
\end{aligned}
\tag{3.91}
$$

By definition, this gives the desired equality (3.90).

□

Lemma 3.7.10. *For any test function $F\in\mathcal{B}_b(E_n^2)$ and any coalescent indicator $b\in\{0,1\}^{n+1}$, we have*

$$
\Gamma_{n,N}^{\ddagger,b}(F)\mathbf{1}_{\tau_N\geq n}-\Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}C_{b_n}(F)\mathbf{1}_{\tau_N\geq n-1}=\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right).
$$

Proof. Thanks to Cauchy-Schwarz inequality and Proposition 3.7.9, it suffices to verify that

$$
\mathbf{E}\left[\Gamma_{n,N}^{\ddagger,b}(F)^2\mathbf{1}_{\tau_N\geq n}-\Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}C_{b_n}(F)^2\mathbf{1}_{\tau_N\geq n-1}\right]=\mathcal{O}\left(\frac{1}{N}\right).
$$

Next, thanks to the alternative representation (3.57), one derives that

$$
\begin{aligned}
&\mathbf{E}\left[\Gamma_{n,N}^{\ddagger,b}(F)^2\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{G}_{n-1}^N\right]\\
=&\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{'[2]})\in((N)^2)^{\times 2}}\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{'[2]}]\mathbf{1}_{\tau_N\geq n-1}\\
&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}}\mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{'[2]}},\ell_{n-1}^{'[2]})C_{b_n}(F)^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n^{'[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right].
\end{aligned}
$$

To simplify the notation, we denote

$$
R_1(N):=\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}\setminus(N)^4}\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{[2]}]\Lambda_n^{\ddagger,b}[\ell_n^{'[2]}]C_{b_n}(F)^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n^{'[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n},
$$

and

$$
\begin{aligned}
R_2(N):=\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}\setminus(N)^4}&\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{[2]}]C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\\
&\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{'[2]}]C_{b_n}(F)(X_n^{\ell_n^{'[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n},
\end{aligned}
$$

as well as

$$
\begin{aligned}
R_3(N):=\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in((N)^2)^{\times 2}}&\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{[2]}]C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\\
&\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{'[2]}]C_{b_n}(F)(X_n^{\ell_n^{'[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n}
\end{aligned}
$$

Thanks to the conditional independence between $A_{n-1}^{\ell_n^{[2]}}$ and $A_{n-1}^{\ell_n^{'[2]}}$ given $\mathcal{G}_{n-1}^N$, we have

$$
\begin{aligned}
&R_3(N)-R_2(N)\\
=&\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in(N)^4}\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{[2]}]C_{b_n}(F)(X_n^{\ell_n^{[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{'[2]}]C_{b_n}(F)(X_n^{\ell_n^{'[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n}\\
=&\sum_{(\ell_n^{[2]},\ell_n^{'[2]})\in(N)^4}\mathbf{E}\left[\Lambda_n^{\ddagger,b}[\ell_n^{[2]}]\Lambda_n^{\ddagger,b}[\ell_n^{'[2]}]C_{b_n}(F)^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n^{'[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n},
\end{aligned}
$$

whence we deduce that

$$
\mathbf{E}\left[\Gamma_{n,N}^{\ddagger,b}(F)^2\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{G}_{n-1}^N\right]=R_1(N)-R_2(N)+R_3(N).\quad a.s.
$$

In addition, since one of the by-product (3.88) in the proof of Proposition 3.7.9 yields

$$\begin{aligned}
&\left(\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}} C_{b_n}(F)\right)^{\otimes 2}(X_n^{\ell_n^{[2]}}, X_n^{\ell_n'^{[2]}}) \\
&= \sum_{(\ell_n^{[2]})\in((N)^2)^{\times 2}} \mathbf{E}\left[\mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\quad \sum_{(\ell_n'^{[2]})\in((N)^2)^{\times 2}} \mathbf{E}\left[\mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^{b}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]})C_{b_n}(F)(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&= \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}} \mathbf{E}\left[\mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^{b}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \\
&\qquad \mathbf{E}\left[\mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^{b}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]})C_{b_n}(F)(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{G}_{n-1}^N\right],
\end{aligned}$$

one has

$$R_3(N) = \Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}} C_{b_n}(F)^2 \mathbf{1}_{\tau_N \geq n-1}, \quad a.s.$$

which guarantees that

$$\mathbf{E}\left[R_3(N)\right] = \mathbf{E}\left[\Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}} C_{b_n}(F)^2 \mathbf{1}_{\tau_N \geq n-1}\right].$$

Therefore, it suffices to verify that

$$\mathbf{E}\left[R_1(N)\right] = \mathcal{O}\left(\frac{1}{N}\right) \quad \text{and} \quad \mathbf{E}\left[R_2(N)\right] = \mathcal{O}\left(\frac{1}{N}\right). \tag{3.92}$$

Together, we prove both of the two convergence above by induction, as the proofs share the same mechanism. Without loss of generality, we suppose that $F \equiv 1$. For $n = 0$, standard calculations give

$$R_1(N) = R_2(N) = \frac{4N-6}{N(N-1)} = \mathcal{O}\left(\frac{1}{N}\right).$$

For $n \geq 1$, we suppose that

$$\mathbf{E}\left[\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N \geq n-1}\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

Now, it is time to go back to the decomposition (3.63). As is mentioned for many times, we may omit the notation $\mathbf{1}_{\tau_N \geq n-1}$.

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

As we have seen in the proof of Proposition 3.7.7, there are 2 different assignments

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

such that two distinct random variables can be found in

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

Let us fix one assignment. This time, in order to execute a finer analysis, we suppose that ℓ_n^1 and $\ell_n'^1$ vary freely in $(N)^2$ and the values of $A_{n-1}^{\ell_n^2}$ and $A_{n-1}^{\ell_n'^2}$ will be almost surely determined by the values of $A_{n-1}^{\ell_n^1}$ and $A_{n-1}^{\ell_n'^1}$. Since the potential function $\mathbf{G}_{\mathbf{n}}^{\ddagger}$ and indicator functions are both nonnegative, we extend $(N)^2$ to $[N]^2$, which gives the following inequality:

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\\
\leq&\mathbf{E}\left[\sum_{\ell_n^1\in[N]} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\\
&\mathbf{E}\left[\sum_{\ell_n'^1\in[N]} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}.
\end{aligned} \tag{3.93}$$

Now, we explain why there is an indicator function

$$\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}$$

at the r.h.s. above. For the case $b_{n-1}=0$, one has

$$\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) = \mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1\neq\ell_{n-1}^2=A_{n-1}^{\ell_n^2}\right\}}\mathbf{1}_{\left\{A_{n-1}^{\ell_n'^1}=\ell_{n-1}'^1\neq\ell_{n-1}'^2=A_{n-1}^{\ell_n'^2}\right\}}.$$

Hence, since $(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\backslash(N)^4$, we have

$$\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2\} = \#\{A_{n-1}^{\ell_{n-1}^1},A_{n-1}^{\ell_{n-1}^2},A_{n-1}^{\ell_{n-1}'^1},A_{n-1}^{\ell_{n-1}'^2}\} \leq \#\{\ell_n^1,\ell_n^2,\ell_n'^1,\ell_n'^2\} < 4.$$

At the same time, when $b_{n-1}=1$, one has

$$\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) = \mathbf{1}_{\left\{A_{n-1}^{\ell_n^1}=\ell_{n-1}^1=A_{n-1}^{\ell_n^2}\neq\ell_{n-1}^2\right\}}\mathbf{1}_{\left\{A_{n-1}^{\ell_n'^1}=\ell_{n-1}'^1=A_{n-1}^{\ell_n'^2}\neq\ell_{n-1}'^2\right\}}.$$

Since $\ell_n^1\neq\ell_n^2$ and $\ell_n'^1\neq\ell_n'^2$, we have

$$\exists i,j\in\{1,2\},\quad s.t.\quad \ell_n^i=\ell_n'^j.$$

In no matter which case mentioned above, it is necessary that $\ell_{n-1}^1=\ell_{n-1}'^1$. Therefore, one also gets

$$\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2\} < 4.$$

The arguments above will be applied repeatedly in the rest of the proof. Next, by the same procedure in the proof of Proposition 3.7.7, we obtain

$$\begin{aligned}
&\mathbf{E}\left[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right]\\
&\qquad\qquad\leq C_1\times\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},
\end{aligned}$$

where C_1 denotes positive constant which does not depend on N. Meanwhile, by the same procedure, we also have

$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2$$
$$\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq C_1' \times \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

where C_1' is also a positive constant which does not depend on N.

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

Again, similart in the proof of Proposition 3.7.7, there are 4 different assignments such that there are 3 distinct random variables within

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

Let us fix one assignment. As is done in the previous case, to conduct a finer study, we suppose that $(\ell_n^{[2]},\ell_n'^1)$ vary freely in $(N)^3$ and the values of $A_{n-1}^{\ell_n'^2}$ will be almost surely determined by the choice of assignment and the values of $A_{n-1}^{\ell_n^{[2]}}$ and $A_{n-1}^{\ell_n'^1}$. This time, by extending $(N)^3$ to $(N)^2\cup[N]$, we get

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell_n^1,\ell_n'^1)\in(N)^3} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \\
&\qquad\qquad \lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})C_{b_n}(F)^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n'^{[2]}}) \,\Bigg|\, \mathcal{G}_{n-1}^N\Bigg] \\
\leq&\mathbf{E}\Bigg[\sum_{\ell_n^{[2]}\in(N)^2} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})C_{b_n}(F)(X_n^{\ell_n^{[2]}}) \,\Bigg|\, \mathcal{G}_{n-1}^N\Bigg] \\
&\mathbf{E}\Bigg[\sum_{\ell_n'^1\in[N]} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})C_{b_n}(F)(X_n^{\ell_n'^{[2]}}) \,\Bigg|\, \mathcal{G}_{n-1}^N\Bigg] \\
&\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}.
\end{aligned} \tag{3.94}$$

Again, by similar argument given in the proof of Proposition 3.7.7, we get

$$\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\Bigg|\, \mathcal{G}_{n-1}^N\Bigg]$$
$$\leq C_2 \times \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

and

$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \mathbf{G}_{\mathbf{n-1}}^{\ddagger}(\mathbf{X_{n-1}})^2$$
$$\mathbf{E}\left[\lambda_{n-1}^b(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]}) \,\middle|\, \mathcal{G}_{n-1}^N\right] \leq C_2' \times \mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

where C_2 and C_2' are positive constant which does not depend on N.

Combining both cases, one gets

$$\begin{aligned}
&R_1(N)\\
&\leq \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{\prime[2]})\in((N)^2)^{\times 2}} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{\prime[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}^{\prime 1},\ell_{n-1}^{\prime 2}<4\}\}}\\
&= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{\prime[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{\prime[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1'), \quad a.s.
\end{aligned}$$

and

$$\begin{aligned}
&R_2(N)\\
&\leq \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{\prime[2]})\in((N)^2)^{\times 2}} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{\prime[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}^{\prime 1},\ell_{n-1}^{\prime 2}<4\}\}}\\
&= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{\prime[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{[2]}]\Lambda_{n-1}^{\ddagger,b}[\ell_{n-1}^{\prime[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2'). \quad a.s.
\end{aligned}$$

The desired convergence (3.92) are therefore guaranteed by the induction hypothesis by taking the expectation on both sides of the inequalities above. This is the end of the proof of Lemma 3.7.10.

□

Lemma 3.7.11. *For any test function $F\in\mathcal{B}_b(E_n^2)$ and any coalescent indicator $b\in\{0,1\}^{n+1}$, we have*

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

Proof. Before starting, we mention that this proof bears a resemblance to the one of Lemma 3.7.10. Thanks to Cauchy-Schwarz inequality, it is sufficient to verify that

$$\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)^2\mathbf{1}_{\tau_N\geq n}-\widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)^2\mathbf{1}_{\tau_N\geq n-1}\right]=\mathcal{O}\left(\frac{1}{N}\right).$$

Again, similar to the equation (3.70), by the alternative representation (3.60) and decomposition (3.59), we deduce that

$$\begin{aligned}
&\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)^2\mathbf{1}_{\tau_N\geq n}\,\middle|\,\mathcal{W}_{n-1}^N\right]\\
&=\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{\prime[2]})\in((N)^2)^{\times 2}} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{\prime[2]}]\mathbf{1}_{\tau_N\geq n-1}\\
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n^{\prime[2]})\in((N)^2)^{\times 2}} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{\prime[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\\
&\qquad\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{\prime[2]}},\ell_{n-1}^{\prime[2]})F^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n^{\prime[2]}})\,\Bigg|\,\mathcal{W}_{n-1}^N\Bigg].
\end{aligned}$$

Similar to the previous case, we denote

$$R_1(N):=\sum_{(\ell_n^{[2]},\ell_n^{\prime[2]})\in((N)^2)^{\times 2}\setminus(N)^4}\mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}]\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{\prime[2]}]F^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n^{\prime[2]}})\,\middle|\,\mathcal{G}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n},$$

and

$$R_2(N) := \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}]F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n'^{[2]}]F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \mathbf{1}_{\tau_N\geq n},$$

as well as

$$R_3(N) := \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}]F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \mathbf{E}\left[\Lambda_n^{\dagger,b}[\ell_n'^{[2]}]F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \mathbf{1}_{\tau_N\geq n}$$

Thanks to the conditional independence between $A_{n-1}^{\ell_n^{[2]}}$, $B_{n-1}^{\ell_n^{[2]}}$ and $A_{n-1}^{\ell_n'^{[2]}}$, $B_{n-1}^{\ell_n'^{[2]}}$ given $\mathcal{W}_{n-1}^N$, we have

$$\begin{aligned}
&R_3(N) - R_2(N)\\
&= \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in(N)^4} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}]F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n'^{[2]}]F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \mathbf{1}_{\tau_N\geq n}\\
&= \sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in(N)^4} \mathbf{E}\left[\widetilde{\Lambda}_n^{\dagger,b}[\ell_n^{[2]}]\widetilde{\Lambda}_n^{\dagger,b}[\ell_n'^{[2]}]F^{\otimes 2}(X_n^{\ell_n^{[2]}}, X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \mathbf{1}_{\tau_N\geq n},
\end{aligned}$$

from which we get

$$\mathbf{E}\left[\widetilde{\Gamma}_{n,N}^{\dagger,b}(F)^2 \mathbf{1}_{\tau_N\geq n} \,\middle|\, \mathcal{W}_{n-1}^N\right] = R_1(N) - R_2(N) + R_3(N). \quad a.s.$$

Notice that Proposition 3.7.10 gives, on the event $\{\tau_N \geq n-1\}$,

$$\begin{aligned}
&\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}} \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}}, \ell_{n-1}^{[2]})F(X_n^{\ell_n^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right]\\
&\qquad \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]}, \mathbf{B_{n-1}}, \mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}}, \ell_{n-1}'^{[2]})F(X_n^{\ell_n'^{[2]}}) \,\middle|\, \mathcal{W}_{n-1}^N\right]\\
&= \left(\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)\right)^{\otimes 2}(X_{n-1}^{\ell_{n-1}^{[2]}}, X_{n-1}^{\ell_{n-1}'^{[2]}}).
\end{aligned}$$

Hence, one has

$$R_3(N) = \widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)^2 \mathbf{1}_{\tau_N\geq n-1}, \quad a.s.$$

which guarantees that

$$\mathbf{E}\left[R_3(N)\right] = \mathbf{E}\left[\widetilde{\Gamma}_{n-1,N}^{\dagger,b}\widetilde{\mathbf{Q}}_{\hat{n}}^{\dagger,b_{n-1}}(F)^2 \mathbf{1}_{\tau_N\geq n-1}\right].$$

Therefore, it suffices to verify that

$$\mathbf{E}\left[R_1(N)\right] = \mathcal{O}\left(\frac{1}{N}\right) \quad \text{and} \quad \mathbf{E}\left[R_2(N)\right] = \mathcal{O}\left(\frac{1}{N}\right). \tag{3.95}$$

Without loss of generality, we suppose that $F \equiv 1$. The rest of the proof is done by induction. For $n = 0$, by definition, we have

$$R_1(N) = R_2(N) = \frac{4N-6}{N(N-1)} = \mathcal{O}\left(\frac{1}{N}\right).$$

For $n \geq 1$, we suppose that

$$\mathbf{E}\left[\sum_{(\ell_{n-1}^{[2]},\ell_{n-1}^{\prime[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{\prime[2]}]\mathbf{1}_{\tau_N\geq n-1}\right] = \mathcal{O}\left(\frac{1}{N}\right).$$

As is stated many times, we may omit the notation $\mathbf{1}_{\tau_N\geq n-1}$ since it is $\mathcal{W}_{n-1}^N$-measurable. Once again, let us return to the decomposition (3.63). Since the essential idea is highly repetitive w.r.t. the reasoning in the proof of Proposition 3.7.10, we skip some of the unnecessary details in the rest of the proof.

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

By the same procedure given in the proof of Proposition 3.7.8 and Lemma 3.7.10, we obtain

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n^{\prime[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{\prime[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\
&\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{\prime[2]}},\ell_{n-1}^{\prime[2]})F^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n^{\prime[2]}}) \,\Bigg|\, \mathcal{W}_{n-1}^N\Bigg]\mathbf{1}_{\tau_N\geq n-1} \\
&\leq C_1\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}^{\prime 1},\ell_{n-1}^{\prime 2}<4\}\}},
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{(\ell_n^{[2]},\ell_n^{\prime[2]})\in((N)^2)^{\times 2}\cap[N]_2^4} \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]}) \,\middle|\, \mathcal{W}_{n-1}^N\right] \\
&\qquad \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{\prime[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{\prime[2]}},\ell_{n-1}^{\prime[2]}) \,\middle|\, \mathcal{W}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n-1} \\
&\leq C_1'\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}^{\prime 1},\ell_{n-1}^{\prime 2}<4\}\}},
\end{aligned}$$

where C_1 and C_1' are some constant that does not depend on N.

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

This time, we get

$$\begin{aligned}
&\mathbf{E}\Bigg[\sum_{(\ell_n^{[2]},\ell_n^{\prime[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{\prime[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}}) \\
&\qquad \lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{\prime[2]}},\ell_{n-1}^{\prime[2]})F^{\otimes 2}(X_n^{\ell_n^{[2]}},X_n^{\ell_n^{\prime[2]}}) \,\Bigg|\, \mathcal{W}_{n-1}^N\Bigg]\mathbf{1}_{\tau_N\geq n-1} \\
&\leq C_2\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}^{\prime 1},\ell_{n-1}^{\prime 2}<4\}\}},
\end{aligned}$$

and

$$\sum_{(\ell_n^{[2]},\ell_n'^{[2]})\in((N)^2)^{\times 2}\cap[N]_3^4} \mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n^{[2]}},\ell_{n-1}^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]$$
$$\mathbf{E}\left[\widetilde{\mathbf{G}}_{\mathbf{n-1}}^{\dagger,b_{n-1}}(\ell_{n-1:n}'^{[2]},\mathbf{B_{n-1}},\mathbf{X_{n-1}})\lambda_{n-1}^{(\varnothing)}(A_{n-1}^{\ell_n'^{[2]}},\ell_{n-1}'^{[2]})\,\middle|\,\mathcal{W}_{n-1}^N\right]\mathbf{1}_{\tau_N\geq n-1}$$
$$\leq C_2'\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}},$$

where C_2 and C_2' are some constant that does not depend on N.

By combining the both cases, we establish that

$$\begin{aligned}R_1(N) &\leq \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}\\ &= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_1+C_1'),\quad a.s.\end{aligned}$$

and

$$\begin{aligned}R_2(N) &\leq \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2')\mathbf{1}_{\{\#\{\ell_{n-1}^1,\ell_{n-1}^2,\ell_{n-1}'^1,\ell_{n-1}'^2<4\}\}}\\ &= \sum_{(\ell_{n-1}^{[2]},\ell_{n-1}'^{[2]})\in((N)^2)^{\times 2}\setminus(N)^4} \widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}^{[2]}]\widetilde{\Lambda}_{n-1}^{\dagger,b}[\ell_{n-1}'^{[2]}]\mathbf{1}_{\tau_N\geq n-1}(C_2+C_2').\quad a.s.\end{aligned}$$

By taking the expectation on both sides of the inequalities above, the desired convergence (3.92) are then verified thanks to the induction hypothesis. The conclusion follows. □

Lemma 3.7.12. *For any test function $F\in\mathcal{B}_b(E_n)^{\otimes 2}$ and any coalescent indicator $b\in\{0,1\}^{n+1}$, we have*

$$\Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{\hat{n}}^{\ddagger,b_{n-1}}C_{b_n}(F)-\Gamma_{n-1,N}^{\ddagger,b}\mathbf{Q}_{n}^{\ddagger,b_{n-1}}C_{b_n}(F)=\mathcal{O}_{\mathbb{L}^1}\left(\frac{1}{\sqrt{N}}\right),$$

as well as

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

Proof. First, we noticed that for any test functions $F_1,F_2\in\mathcal{B}_b(E_n)^{\otimes 2}$, Minkowski's inequality gives

$$\begin{aligned}&\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_1+F_2)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1+F_2)\right|_{\mathbb{L}^1}\right\|\\ \leq\ &\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1)\right|\right\|_{\mathbb{L}^1}\\ &+\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_2)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_2)\right|\right\|_{\mathbb{L}^1}.\end{aligned}\tag{3.96}$$

Second, thanks to Cauchy-Schwarz inequality, Proposition 3.7.5 and Proposition 3.7.7, we deduce that

$$\begin{aligned}&\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta_{n-1}^N)^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1)\right|\right\|_{\mathbb{L}^1}\\ \leq&\left\|\Gamma_{n-1,N}^{\ddagger,b}(1)\mathbf{1}_{\tau_N\geq n-1}\right\|_{\mathbb{L}^2}\left\|(\eta_{n-1}^N)^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta_{n-1}^{\otimes 2}(F_1)\right\|_{\mathbb{L}^2}\\ =&\mathcal{O}\left(\frac{1}{\sqrt{N}}\right).\end{aligned}\tag{3.97}$$

Similarly, thanks to Proposition 3.7.8, we also have

$$\begin{aligned}
&\left\|\widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta^N_{n-1})^{\odot 2}(F_1+F_2)\mathbf{1}_{\tau_N\geq n-1}-\eta^{\otimes 2}_{n-1}(F_1+F_2)\right|\right\|_{\mathbb{L}^1}\\
\leq\quad &\left\|\widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta^N_{n-1})^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta^{\otimes 2}_{n-1}(F_1)\right|\right\|_{\mathbb{L}^1}\\
&+\left\|\widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta^N_{n-1})^{\odot 2}(F_2)\mathbf{1}_{\tau_N\geq n-1}-\eta^{\otimes 2}_{n-1}(F_2)\right|\right\|_{\mathbb{L}^1}.
\end{aligned}\tag{3.98}$$

and

$$\begin{aligned}
&\left\|\widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N\geq n-1}\left|(\eta^N_{n-1})^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta^{\otimes 2}_{n-1}(F_1)\right|\right\|_{\mathbb{L}^1}\\
\leq&\left\|\widetilde{\Gamma}^{\dagger,b}_{n-1,N}(1)\mathbf{1}_{\tau_N\geq n-1}\right\|_{\mathbb{L}^2}\left\|(\eta^N_{n-1})^{\odot 2}(F_1)\mathbf{1}_{\tau_N\geq n-1}-\eta^{\otimes 2}_{n-1}(F_1)\right\|_{\mathbb{L}^2}\\
=&\mathcal{O}\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}\tag{3.99}$$

Finally, returning to two pairs of decompositions (3.76), (3.86) and (3.89), (3.91), the boundness of G_{n-1} and the homogeneous structure in these two decompositions allow us to apply respectively (3.96), (3.97) and (3.98), (3.99). Note that

$$m(\mathbf{X_{n-1}})(G_{n-1}) = m^{\odot 2}(\mathbf{X_{n-1}})(1\otimes G_{n-1}),$$

and

$$m(\mathbf{X_{n-1}})(G^2_{n-1}) = m^{\odot 2}(\mathbf{X_{n-1}})(1\otimes G^2_{n-1}).$$

The desired $\mathbb{L}^1$-bound can therefore be obtained with some standard algebraic manipulations.

□

Chapter 4

Estimating Committor Function with Mondrian Forests

abstract: In Molecular Dynamics, the rare event of interest can be modeled by the transition of some underlying Markov process between metastable states. In Transition Path Theory, a very important model reduction technique is to design a *reaction coordinate*, which is a function that measures the advance of a reactive trajectory towards a metastable state. Let A and B be two metastable states. A *committor function* is the perfect choice of reaction coordinate in the sense that it measures exactly the probability of reaching B before A. The committor function is also, in some sense, the perfect reaction coordinate for generalized Adaptive Multilevel Splitting (gAMS, see [BGG+16]) and crucial to the performances of many other rare-event estimation algorithms. We investigate the performance of Mondrian Forests (MF) [LRT14] to estimate the committor function, and we provide strategies to couple gAMS iteratively: first, since gAMS can also provide information on the committor function, we can use this algorithm to generate training data for MF, and, conversely, to update gAMS by using the trained MF model as its reaction coordinate. As iterations go, the updated-gAMS to generate the training data of better quality, with with, MF is expected to be able to yield a good approximation of the committor function.

4.1 Introduction

In Transition Path Theory (TPT, see, e.g., [EVE10]), a typical problem is to sample the transition paths between a metastable state A and another metastable state B (see Figure 4.1). More precisely, let E denote a state space in which the underlying dynamics is modeled by a Markov process $\mathbf{X} := (X_t; t \geq 0)$. A metastable state of $\mathbf{X}$ is an open subset of E such that when $\mathbf{X}$ is trapped in such set, it takes an extremely long time for $\mathbf{X}$ to escape (see Figure 4.1). Let A, B be two metastable sets in E, then the committor function at point x is the probability that $\mathbf{X}$ starting from x, reaches B before A.

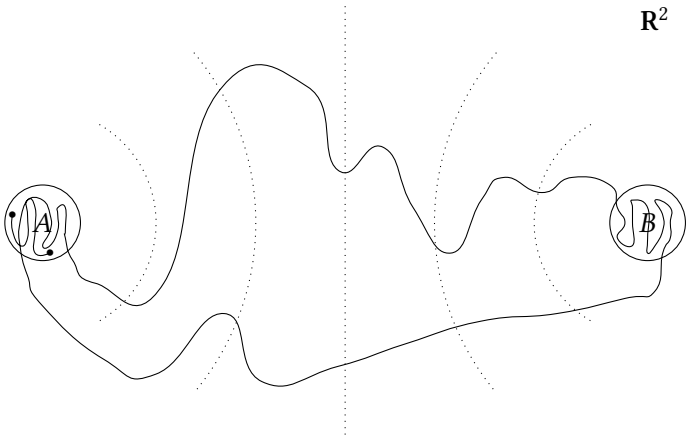

Figure 4.1: Schematic picture of transition paths.

By the metastability of A and B, when evaluating the values of the committor function close to A, crude Monte Carlo typically fails because the process is very much likely to go back to A rather than to go to B: the probability to be evaluated is thus very small. This metastability problem is related to a time scale problem: although the molecular transitions are not so rare at the macroscopic timescale, the dynamics encoded in the Markov process can only be simulated w.r.t. femtosecond timescale. Therefore, in order to ensure a certain level of accuracy, the wall-clock time of simulation is typically intractable.

One popular numerical approach to counter this is the gAMS framework [BGG+16]. The basic idea is to generate an Interacting Particle System (IPS) based on an adaptive level updating strategy, where the trajectories that advance more survive. A level is calculated w.r.t. a reaction coordinate, which is of crucial importance to the performance of the algorithm. It is well-known (see, e.g., [BLR15, CGR19b]) that the committor function is the optimal reaction coordinate for gAMS. In addition, since gAMS is able to evaluate efficiently the values of the committor function close to A, the idea of designing a regressor to estimate the committor function is therefore natural.

An elementary approach is to split the state space E or some compact subset based on a regular grid cells, and a natural approximation of the committor function can then be derived in each cell thanks to ensemble of paths generated by the gAMS algorithm. The reader is referred to [LL19] for a rich list of numerical experiments and a concrete application on alanine dipeptide. However, this natural construction will only work in low dimensional settings, since it is not possible to create a uniform mesh when the dimension of the state space is large. Intuitively, a possible generalization of such a method is to find an intelligent way to create an adaptive mesh,

such that it may also work in a high-dimensional setting.

Mondrian Forests (MF) [LRT14], a variant of Random Forests (RF), i.e., an ensemble of randomized decision trees, are proposed as the regressor of committor function. The construction of MF are based on a stochastic process called Mondrian process (MP, see, e.g., [RT09]), taking values in guillotine partitions of an axis-aligned box. Roughly speaking, a Mondrian process is a high-dimensional generalization of the partition on an interval that is split by a Poisson process. Said differently, MF provide an intelligent way to create "randomized uniform mesh" that is tractable in a high-dimensional setting. From a different angle, MF can be regarded as an RF-based regressor, such that *online learning* is available. More concretely, when new training data come, there is no need to retrain the model in order to have a better quality regressor. This property is crucial for our iterative updating strategy (cf. Figure 4.2).

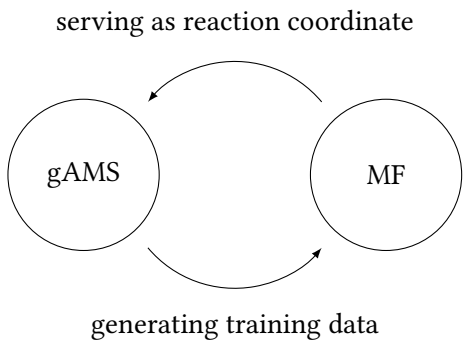

Figure 4.2: The illustration of iterative updating strategy.

Unfortunately, the theoretical understanding of MF are still in its infancy. To the interested readers, we refer to [MGS17] and [MGS18] for recent theoretical developments on MF, where a Purely Random Forests version of MF are proposed, along with a min-max convergence rate analysis. Therefore, we mainly provide numerical illustrations and ideas on how to use MF to design efficient strategy to estimate the committor function.

4.2 Setting

4.2.1 Overdampled Langevin dynamics

We consider an overdamped Langevin process $\mathbf{X} = (X_t; t \geq 0)$ taking values in the state space $E = \mathbf{R}^d$ defined by

$$dX_t = -\nabla V(X_t)dt + \sqrt{2\beta^{-1}}dW_t, \tag{4.1}$$

where $(W_t; t \geq 0)$ denotes a d-dimensional Wiener process, V denotes the associated energy and the inverse temperature $(\kappa_B T)^{-1}$ is denoted by β.

A metastable state is an open subset of E such that when $\mathbf{X}$ is trapped in such set, it takes an extremely long time for $\mathbf{X}$ to escape. Let us denote A and B two metastable states in a state space E. For the Markov process $\mathbf{X} := (X_t, t \geq 0)$, let τ_A and τ_B denote respectively the stopping times

$$\tau_A := \inf\{t \geq 0 \mid X_t \in A\},$$

and

$$\tau_B := \inf\{t \geq 0 \mid X_t \in B\}.$$

The *committor function* $\xi^* : E\backslash(A \cup B) \mapsto [0,1]$ is defined by

$$\xi^*(x) := \mathbf{P}\left(\tau_B < \tau_A \mid X_0 = x\right).$$

Our goal is to design an efficient strategy to estimate this function on some compact subset of the state space E. In particular, we are interested in exploiting this estimation to improve the performance of gAMS and vice versa.

4.2.2 Ground truth

In this section, we show how one can obtain the reference values of a committor function in overdamped Langevin dynamics for some low-dimensional toy examples, which will be referred to as *ground truth*, namely the "real" values of the committor function. This method only works in low-dimensional setting. It is well known (see, e.g. [BLR15]) that ξ^* is the solution of the following elliptic Partial Differential Equation (PDE):

$$-\nabla V \cdot \nabla u + \beta^{-1}\Delta u = 0 \quad \text{on} \quad E\backslash(A \cup B), \tag{4.2}$$

with the boundary conditions

$$\begin{cases} u = 0 & \text{on } \partial A; \\ u = 1 & \text{on } \partial B. \end{cases} \tag{4.2'}$$

For sufficiently fine grids, the numerical solution of (4.2) using Finite Difference method is precise enough so that it can be regarded as the real committor function.

Three-hole potential The main example studied in this chapter is the following toy exmaple (see Figure 4.3):

$$\begin{aligned} V(x,y) :=& 3\exp\left(-x^2 - (y-1/3)^2\right) - 3\exp\left(-x^2 - (y-5/3)^2\right) \\ & - 5\exp\left(-(x-1)^2 - y^2\right) - 5\exp\left(-(x+1)^2 - y^2\right) + x^4/5 + (y-1/3)^2/5. \end{aligned}$$

We consider the following two metastable states respectively defined by

$$A := \{(x,y) \in \mathbf{R}^d : (x+1)^2 + y^2 < 0.1\},$$

and

$$B := \{(x,y) \in \mathbf{R}^d : (x-1)^2 + y^2 < 0.1\}.$$

The reader is referred to [MSVE06] for a wider list of toy examples. In order to solve the elliptic PDE (4.2) by Finite Difference method, we consider a rectangular domain $\Omega = [-2,2] \times [-2,2]$ and a uniform mesh with stepsize 0.01. We also add a Neumann boundary condition on the boundary $\partial\Omega$ of the rectangular domain, i.e., $\partial_{\vec{n}} u = 0$ where $\vec{n}$ denotes the unit normal vector on the boundary. These fictitious boundary conditions do not affect too much the quality of the result since when starting far from A and B, the difference of the values of the committor function for two close points is negligible. This ensures that the Finite Difference result yields an accurate approximation of the solution to (4.2)-(4.2'). The numerical solution of (4.2)-(4.2') for the inverse temperature $\beta = 1.67$ is illustrated in Figure 4.4.

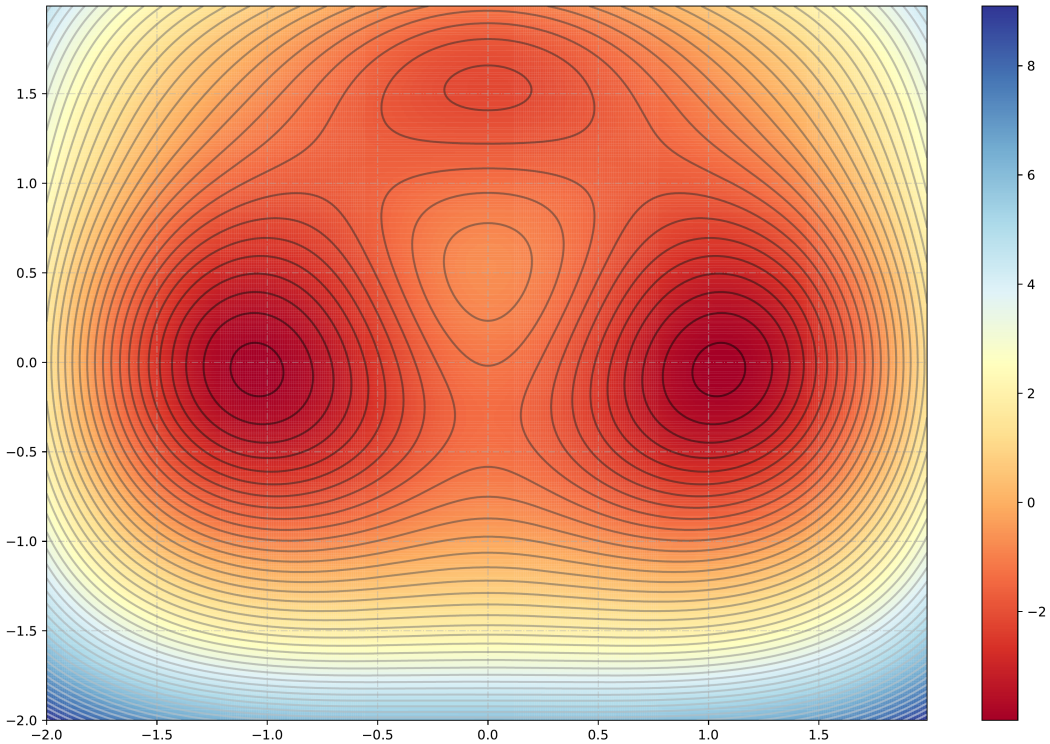

Figure 4.3: Representation of the three-hole energy landscape.

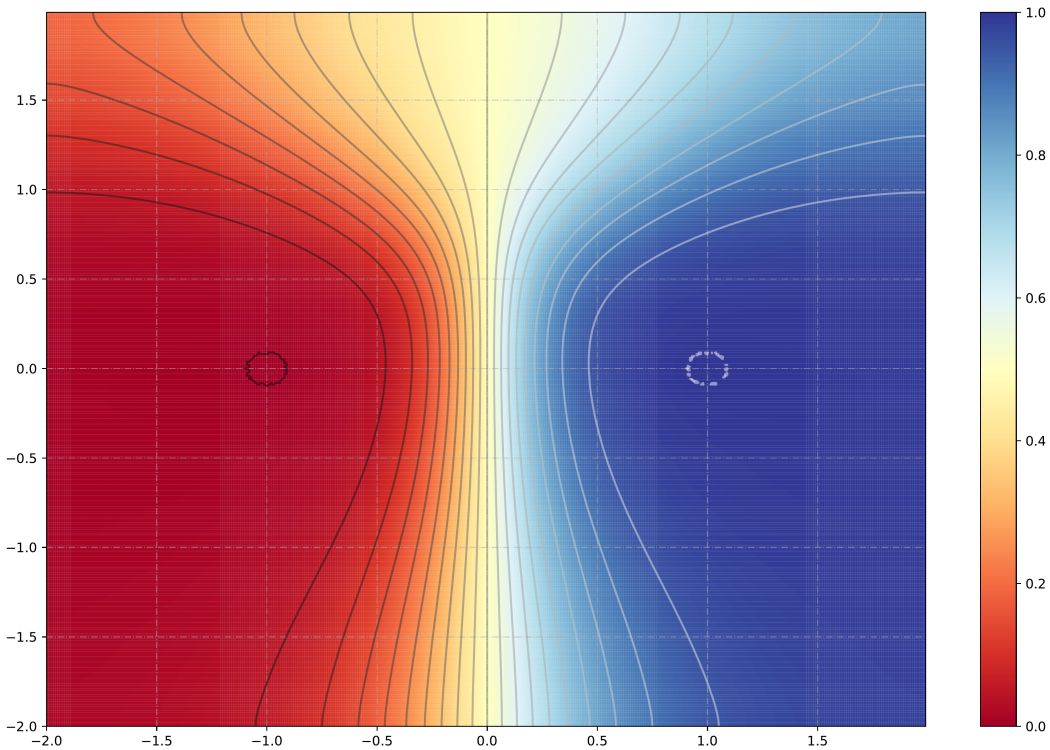

Figure 4.4: The numerical solution of (4.2) with inverse temperature $\beta = 1.67$.

4.2.3 On the choice of regressors

The committor function is a smooth function, taking values in $[0, 1]$. In this sense, any *state-of-the-art* machine learning model may serve as a regressor the committor function. However, when interacting with gAMS methods, certain practical aspects have to be taken into account when choosing the learning method:

- First, the regressor has to be fast in terms of prediction speed. In fact, gAMS algorithm requires that at each state of each trajectory, the reaction coordinate is evaluated. This is a *huge* amount of computation costs if the prediction of the regressor is complicated, and most of them would not be useful. In this sense, a sophisticated fine-tuned Deep Learning model may not be a relevant candidate.
- Second, the regressor must be able to execute online learning or incremental learning. When new data come, the regressor should be able to update in order to have better accuracy. In this sense, classic Random Forests may not be a good choice.
- Finally, we expect it to be as adaptive and robust as possible in high dimension, since we do not have much insights on how to tune a specific model to learn a high-dimensional function.

As such, we may consider two families of regressors in order to estimate the committor function efficiently. The advantages and disadvantages are listed as follows.

(i) Gradient-based learning algorithms, such as logistic regression, XGBoosting, or some shallow neural network:

  (a) Online learning is available by a Stochastic Gradient Descent-based optimization algorithm;

  (b) It is easy to add regularization modules, such as the "smoothness" of the prediction, etc;

  (c) The parameters of the algorithm are hard to tune in general;

  (d) Sometimes, the results are hard to interpret.

(ii) Ensemble-based learning algorithms, mainly variants of Random Forests:

  (a) The ensemble methods are super robust in high-dimensional settings by the design of Monte Carlo-typed structure;

  (b) Nearly no tuning is needed, which also means that the regularization modules are not easy to implement in general;

  (c) Online learning is generally difficult to design;

  (d) It is straightforward to understand and control possible dangerous situations.

4.3 A brief introduction to Mondrian Forests

Mondrian Forests (MF) were introduced in [LRT14], named after the famous Dutch painter Piet Mondrian, as the partitions created by each Mondrian tree (MT) and Mondrian's paintings have similar style. The fundamental idea of the construction is based on a guillotine-partition-valued stochastic process called the Mondrian Process (MP). For details, the readers are referred to

[RT09] and [BW15]. In general, MF are a variant of RF such that online learning is available, and designed in a smart way. In this section, we present the basic mechanism of a Mondrian tree and we explain why MF are a competitive candidate as a regressor for estimating the committor function.

4.3.1 The mechanism of a Mondrian tree

In this section, we provide an intuitive interpretation on the construction of a Mondrian tree on a 2-dimensional toy example. For the generic algorithm, the reader is referred to [LRT14]. Let us fix a parameter $\lambda \in \mathbf{R} \cup \{+\infty\}$, which defines the *lifetime* of a Mondrian tree. Let us denote by $(X_i, Y_i; 1 \leq i \leq n)$ the training data. Before starting, we remark that the partition defined by a MT only depends on the inputs $(X_i; 1 \leq i \leq n)$ of the training data.

Construction of MT in dimension 2 At step 0, we denote $ABCD$ the minimum rectangle that covers $X_1, X_2, \ldots, X_n$ (see Figure 4.5). Then, we sample an exponential random variable E_0 with rate $(|AB| + |BC|)$. If $E_0 < \lambda$, a split on the side AB or BC will be executed. More precisely, we sample a uniform random variable U on the interval $[0, |AB| + |BC|]$. If $U \leq |AB|$, a splitting point will be uniformly sampled on the side AB; otherwise, the split will be done on the side BC. After the splitting point is determined, a split will be executed orthogonally to the chosen side. For short, this amounts to say that an orthogonal split is performed uniformly on ABC. In Figure 4.5, the splitting point is on the side AB and its abscissa is denoted by x_0. Therefore, the training data is vertically divided into two subgroups.

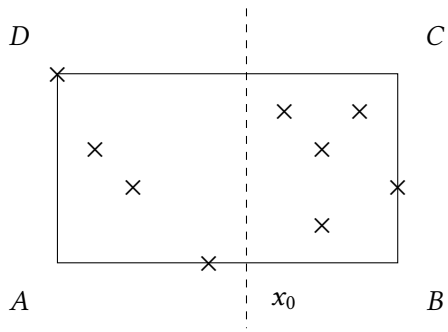

Figure 4.5: First split in the construction of MT.

At the same time, a node in a decision tree according to the split is therefore constructed (see Figure 4.6). By definition, the condition node $x > x_0$ illustrated in Figure 4.6 can be determined by the pair of sets $(\{x > x_0\}, \{x \leq x_0\})$. We say that the lifetime of the node $(\{x > x_0\}, \{x \leq x_0\})$ is E_0. Note that a decision tree uniquely determines a partition on the whole state space $\mathbf{R}^2$.

Next, we perform the splitting for these two subgroups of data recursively. Let us start by the subgroup of data on the left-hand side (see Figure 4.7). Denote again by $ABCD$ the minimum rectangle that covers all the data in this subgroup of data. We sample E_1^1 w.r.t. an exponential distribution with rate $(|AB| + |BC|)$. Now, if $E_0 + E_1^1 < \lambda$, we perform a uniform split on BCD, in the same way as presented at step 0. The same mechanism is applied mutatis mutandis to the right-hand side (see Figure 4.8). The associated decision tree is also updated accordingly, and one tracks the lifetime of each node (see Figure 4.9).

The splitting procedure stops when the lifetime of the proposed condition node surpasses the prefixed lifetime λ. We also remark that, when a subset contains only 1 data, the splitting stops

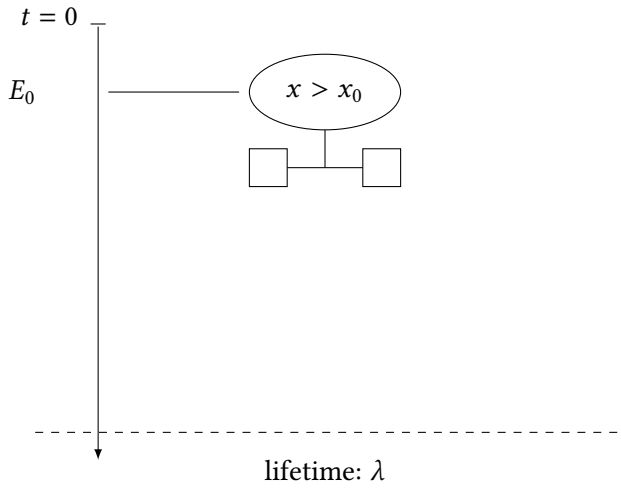

Figure 4.6: Decision tree corresponding to the first split.

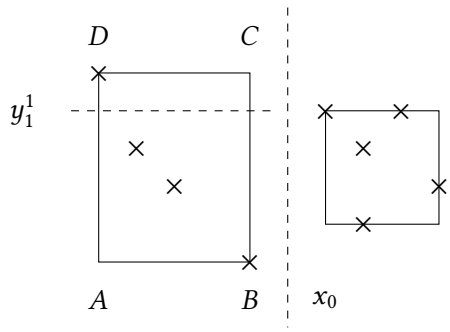

Figure 4.7: Second split in the construction of MT.

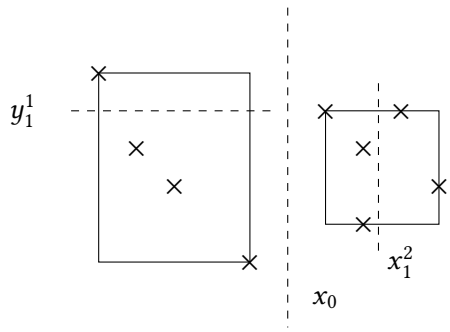

Figure 4.8: Second and third splits in the construction of MT.

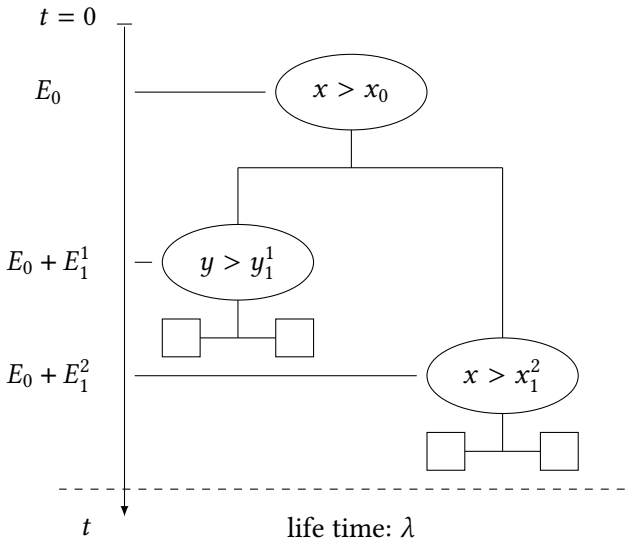

Figure 4.9: The decision tree with three splits.

automatically since the minimum rectangle that contains one point degenerates to a point. In the final stage, a randomized decision tree is therefore constructed, along with a random partition of the state space $\mathbf{R}^2$. In a high-dimensional setting, the orthogonal lines used to execute the splitting procedures are replaced by hyperplanes, and the rectangles are replaced accordingly by hypercubes. The basic mechanism remains the same. The prediction of MT is then provided following the decision tree. More precisely, the prediction on the point x^* is defined as the average of the outputs of the training data that are in the same hypercube as x^*. The online learning of MT exploits the memoryless property of exponential distribution. Indeed, when new data come, one regenerates part of the decision tree such that the existing structure is not changed. It turns out that the online training of a Mondrian tree does not change its posterior distribution given the same data (cf. (4.4)), and the reader is referred to Section 5 of [LRT14] for detailed algorithms.

4.3.2 Theoretical aspects of MF

In this section, we provide some theoretical properties of MF, highlighting the missing part in the theory.

Properties of Ensemble methods MF are ensemble methods, meaning that many weak regressors are constructed independently and we use the average of their predictions as the final estimation. Here, these weak regressors are Mondrian trees: they are variants of classic decision trees used in Random Forests. Denote by $D_{[n]} = (D_i; 1 \leq i \leq n) = (X_i, Y_i; 1 \leq i \leq n)$ a sequence of training data. We also denote $\mathbb{T}_1(D_{[n]}), \cdots, \mathbb{T}_M(D_{[n]})$ a collection of conditional i.i.d. Mondrian trees. We denote $\mathbf{Pred}_{\mathbb{T}_m}\left(x \mid D_{[n]}\right)$ the prediction at point x made by $\mathbb{T}_m(D_{[n]})$. Then, the final prediction is

$$\mathbf{Pred}_{\mathbb{T}_{[M]}}\left(x \mid D_{[n]}\right) := \frac{1}{M}\sum_{m=1}^{M} \mathbf{Pred}_{\mathbb{T}_m}\left(x \mid D_{[n]}\right)$$

Hoeffding inequality and Borel-Cantelli lemma give the following law of large numbers as the prediction is uniformly bounded by 1 for the committor function estimation problem:

$$\mathbf{Pred}_{\mathbb{T}_{[M]}}\left(x \mid D_{[n]}\right) \xrightarrow[M\to\infty]{a.s.} \mathbf{E}\left[\mathbf{Pred}_{\mathbb{T}_1}(x) \mid D_{[n]}\right].$$

At the same time, we have

$$\mathrm{Var}\left[\mathbf{Pred}_{\mathbb{T}_{[M]}}\left(x \mid D_{[n]}\right)\right] = \frac{1}{M}\mathrm{Var}\left[\mathbf{Pred}_{\mathbb{T}_m}\left(x \mid D_{[n]}\right)\right].$$

The final prediction is then more consistent than the prediction made by each regressor. Hence, M can be set as large as possible in practice such that we only have to deal with the randomness introduced by a single MT.

Missing parts in the convergence analysis As is shown above, the variance brought by ensemble methods can be reduced simply by adding more MT in MF. However, for the expectation of the prediction provided by MT given data, the consistency is not guaranteed in general. Let us give a more detailed setting. Assume that the distribution of D_i writes,

$$\begin{cases} X_i \sim \mathrm{Unif}(\Omega); \\ \epsilon_i \sim \mathcal{N}(0, \sigma_i^2); \\ Y_i = \xi^*(X_i) + \epsilon_i, \end{cases}$$

there is no theoretical guarantee such that

$$\mathbf{E}\left[\mathbf{Pred}_{\mathbb{T}_1}(x) \mid D_{[n]}\right] \xrightarrow[n\to\infty]{a.s. \text{ or } \mathbf{P} \text{ or } L^p} \xi^*(x). \tag{4.3}$$

The assumption that ϵ_i is Gaussian is quite natural since both crude Monte Carlo and gAMS provide estimation with normal limit distribution. Recent results [MGS17] show that when $\lambda < +\infty$, there is no consistency in general. A more refined analysis and a variant of Purely Random Forests version of MF are proposed in [MGS18], with a min-max rate for α-Hölder functions. However, we did not use this variant since the construction of the purely random Mondrian tree allows empty cell, and 0 is set to be the estimation when the evaluation is needed to be conducted in such cells. This induces problems in the implementation of gAMS algorithm since $\xi = 0$ means a sudden death of a transition path, which makes the implementation of gAMS numerically unstable.

Since providing a huge amount of training data is not possible due to the computational cost of gAMS methods, we are more interested by the performance when relatively few training data is provided. Hence, asymptotic properties such as consistency are not the priority for the applications in order to improve the performance of gAMS, since the final estimation of rare-event simulation is eventually estimated by gAMS, and even with non-converged estimation of committor function, we still have a theoretical-guaranteed unbiased estimator. In addition, since the construction of decision tree-type regressor ensures that the estimator can only take finite values, gAMS enters into Asymmetric SMC framework introduced in Chapter 3 and can also provide theoretical guaranteed consistent estimations.

Self-consistency of MT The online training of each MT does not depend on the arrival orders of $D_i = (X_i, Y_i)$. More precisely, we have

$$\mathbb{T}_m(D_{[n+1]}) \sim \mathcal{T}_\lambda(D_{[n+1]}) \Leftarrow \begin{cases} \mathbb{T}_m(D_{[n]}) \sim \mathcal{T}_\lambda(D_{[n]}); \\ \mathbb{T}_m(D_{[n+1]}) \mid D_{n+1} \sim \mathcal{M}_{D_{n+1}}\left(\mathbb{T}_m(D_{[n]}), \cdot\right), \end{cases} \tag{4.4}$$

where $\mathcal{M}_{D_{n+1}}$ represents the online updating strategy of Mondrian tree given the data D_{n+1} and $\mathcal{T}_\lambda(D_{[n]})$ denotes the distribution of Mondrian tree with lifetime λ given the data $D_{[n]}$. According to the author of [LRT14], this is the only construction of decision tree available such that the property above is verified. However, this beautiful property comes with a cost: the splits of the MF do not use the value of the Y_i. Therefore, they are not the "optimal" splits given the training data used in classical decision tree's construction. Hence, the robustness over the "extremely bad quality data" will be affected. This will be discussed in the numerical illustration in the next section.

4.4 Numerical illustrations

In this section, we provide some numerical illustrations and interpretations on the estimation of committor function. The energy is set to be the three-hole energy introduced in Section 4.2.2. In the following sections, the sample points are uniformly sampled in the rectangle $\Omega = [-2, 2] \times [-2, 2]$, and by "perfect training data" we mean the numerical solution of (4.2) given by Finite Difference methods.

4.4.1 Learning with perfect data

Unlike the typical estimation problems in statistics and machine learning context, the quality of the training data is indeed controllable in the committor function estimation problem. Although it is of no practical interest to do this kind of trade-off, we investigate the performance of training MF with perfect data (ground truth) to test its adequacy and to have an idea on how many data is needed to provide reasonable approximations. The results are given in Figure 4.10.

4.4.2 Learning with noisy data

Now, we consider adding some artificial Gaussian noise to the training data. For a gAMS algorithm with $N = 100$ and $K^* = 20$, the typical variance over the rectangle Ω is between 10^{-6} to 10^{-4}. Therefore, we consider adding a slightly larger centered normal noise with variance 4×10^{-4}. The results are shown in Figure 4.11. Although it seems that MF cannot handle the situation perfectly, the approximation is still quite impressive. Since the added noises are i.i.d. normal random variables, the relative variance on the left ($x \leq 0$) is noticeably larger than on the right ($x \geq 0$). This is why the estimation quality on the left is worse than on the right in general. The situation where the noise is relatively large is presented in Figure 4.12. We use the possibly largest variance, 1, that can be made by crude Monte Carlo or gAMS. This time, unsurprisingly, MF failed to provide reasonable predictions.

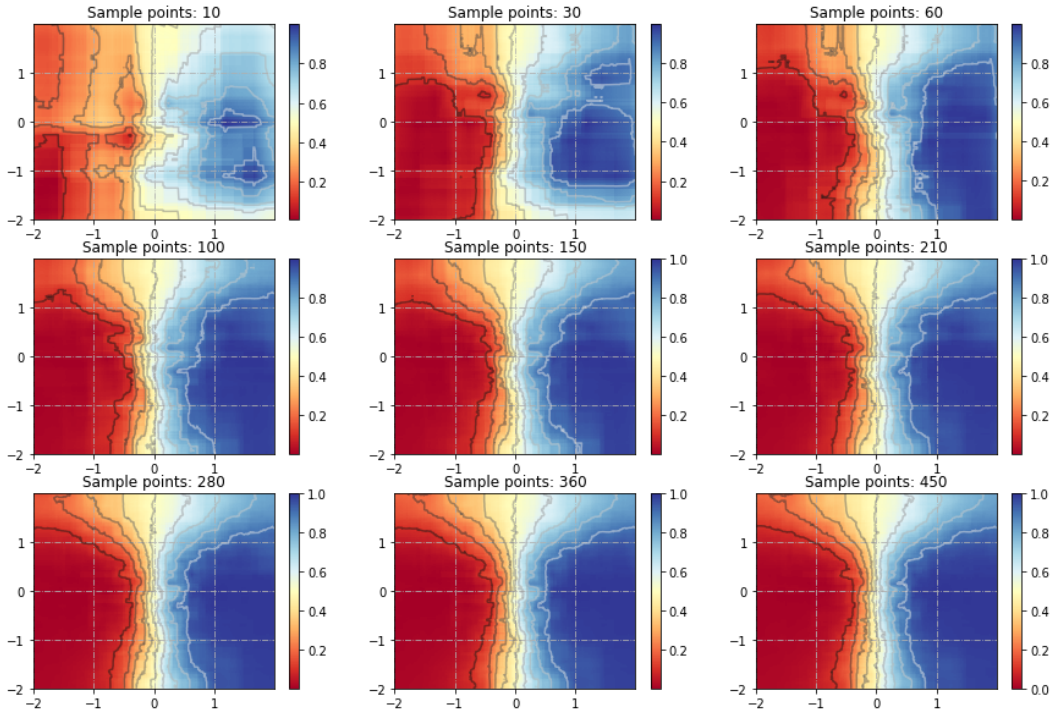

Figure 4.10: Learning with the ground truth, where the training data is uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$.

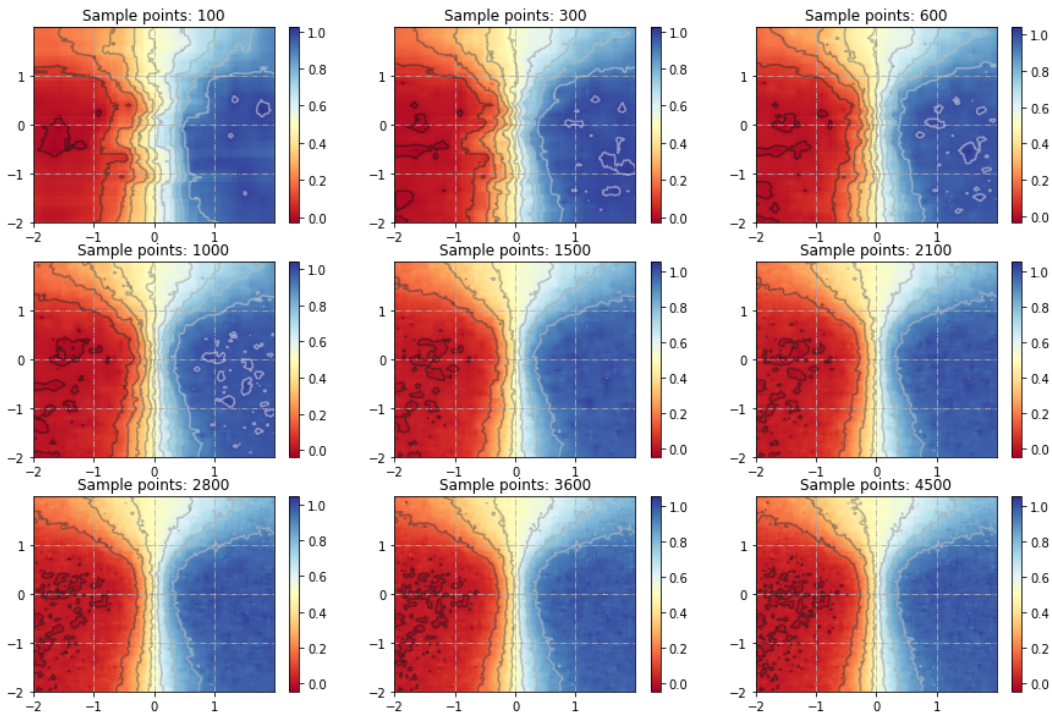

Figure 4.11: Learning with slightly perturbed data, where (X_i) are uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$.

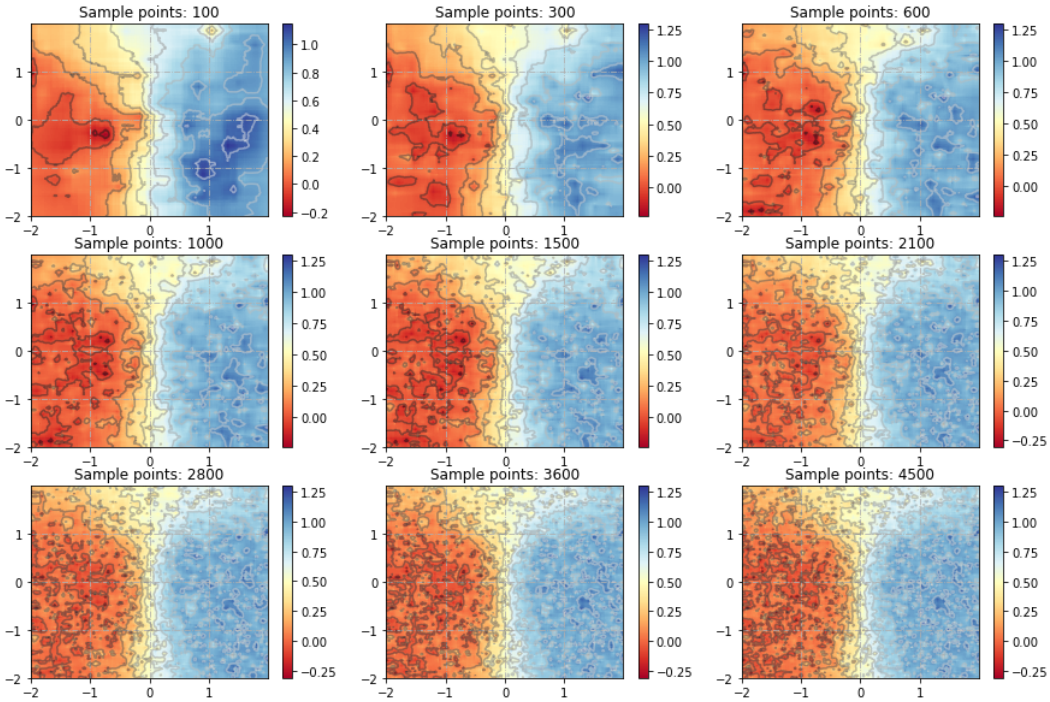

Figure 4.12: Learning with largely perturbed data, where (X_i) are uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$.

4.4.3 Capability of recovering from a ruined model

By a "ruined model" we mean the model generated using extremely low quality training data. We start by training a MF model with largely perturbed data, and then, we continue by providing perfect data to the MF model, to see if it could recover from the ruined model. As presented in Figure 4.13, it may be difficult for an MF model to recover from a ruined one, meaning that a huge amount of high quality training data is required. According to Figure 4.10, a high quality training data of size 1000 is enough to provide an accurate approximation.

4.4.4 Conclusion of the numerical tests

In this section, we summarize the empirical knowledge on MF that we have collected in the numerical tests:

(i) MF do not need a huge sample size to give a reasonable approximation of committor functions;

(ii) MF prefer having less high quality data rather than more low quality data;

(iii) MF are not robust to largely perturbed data;

(iv) It is difficult for a ruined MF model to recover by updating through online learning with high quality training data.

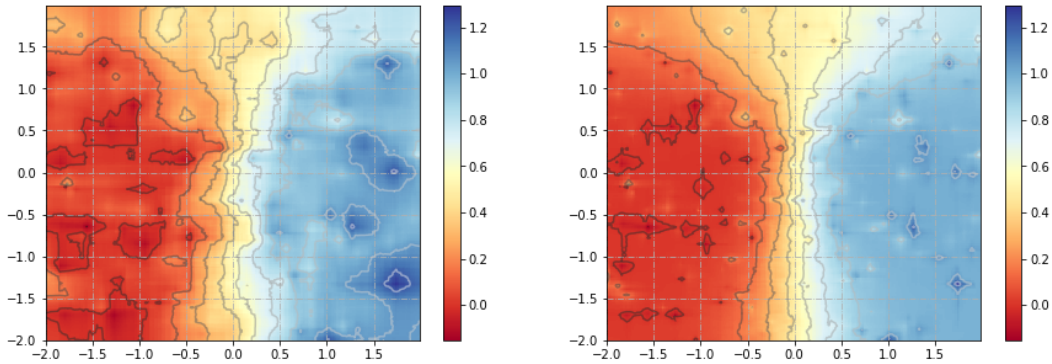

Figure 4.13: The figure on the left represents the prediction made by a MF model trained by 100 largely perturbed data. The one on the right represents the model trained with 1000 more perfect data. (X_i) are uniformly sampled over $\Omega = [-2, 2] \times [-2, 2]$. For the parameters of MF, we set $\lambda = \infty$ and $M = 50$. The inverse temperature is $\beta = 1.67$.

As a consequence, in the design of the iterative updating strategy mentioned in the previous sections, it is crucial to ensure the quality of the training data for MF. This intuition is contrary to the one in typical Machine Learning applications, where the size of the data is usually of top priority.

4.5 Iterative updating strategy with gAMS and MF

In this section, we discuss some possible combinations of MF and gAMS algorithm, and explain how they can help each other in order to improve accuracy. Some numerical illustrations are also provided.

4.5.1 Crude interaction between gAMS and MF

Let us start by using gAMS to estimate committor function in a crude way: we use the estimation provided by gAMS as the training data of MF, with some prefixed reaction coordinate such as the Euclidean distance to the state A. The number of replicas is $N = 100$, and the minimum number of replicas to be killed at each iteration is $K^* = 20$. The reaction coordinate ξ_1 is the Euclidean distance to the point $(-1, 0)$ with the threshold $L^* = 1.8$. The starting point X_0 is sampled uniformly over the rectangle Ω. In particular, when the sampled point turns out to be in A (resp. B), we provide directly 0 (resp. 1) as the output of gAMS. At each point, we simulate independently $n_{sim} = 50$ runs of gAMS, and the final estimation is the average. The prediction given by the MF model is presented in Figure 4.14. Then, we use the trained MF model as the new reaction coordinate and the updated gAMS is therefore implemented to generate new training data for MF. The final performance of the MF model is provided in Figure 4.15.

Clearly, the quality of approximation is improved. This procedure can be done repeatedly and we expect the approximation of the committor function to be more and more accurate.

4.5.2 Tempering

In the previous section, we need to have a reasonable reaction coordinate in order to run gAMS in the first place. This is not always tractable, especially in high dimensional situations. Here,

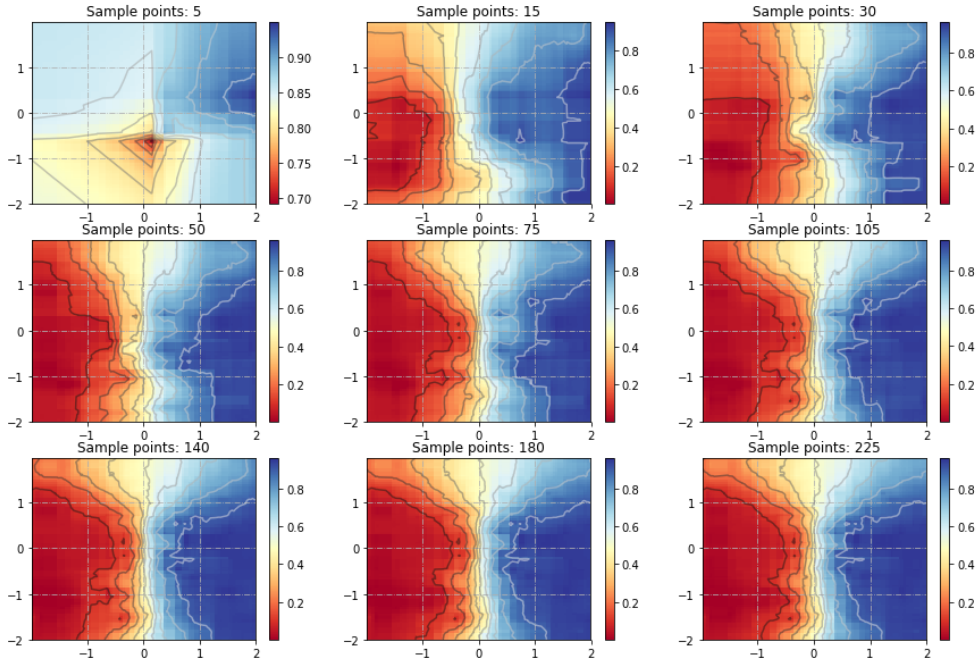

Figure 4.14: Learning with the training data generated by gAMS algorithm. The reaction coordinate is set to be the distance to the state A (i.e., ξ_1). The inverse temperature is $\beta = 1.67$.

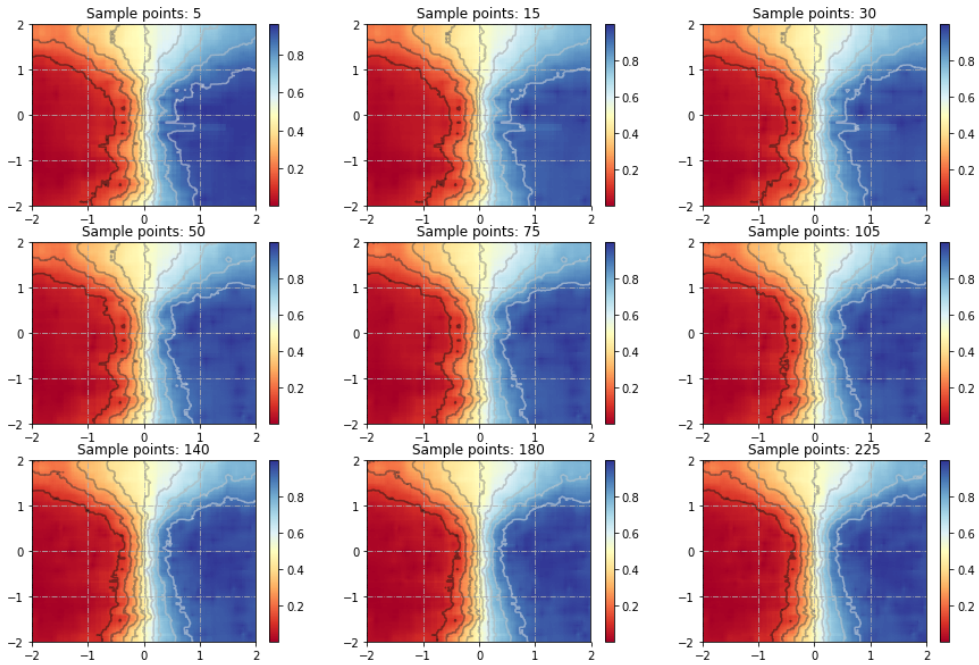

Figure 4.15: Learning with the data re-generated by gAMS algorithm. The reaction coordinate is the trained MF model. The inverse temperature is $\beta = 1.67$.

we provide a possible strategy that allows to "start out of nowhere". The idea is simple: when the temperature is high, the rare events associated to the energy barriers become less rare. We start with a small inverse temperature β_0, such that crude Monte Carlo provides reasonable estimations. Therefore, we estimate the associated committor function ξ_{β_0} by crude Monte Carlo. Next, we consider a tempering sequence (see Figure 4.16) $\beta_0 < \beta_1, < \cdots < \beta_{q^*-1} < \beta_{q^*} = \beta^*$, where β^* denotes the target inverse temperature. For each $q \in [n]$, we estimate ξ_{β_q} by using $\xi_{\beta_{q-1}}$ as reaction coordinate, until we get the target inverse temperature. Although this strategy is computationally intensive, it allows to design gAMS algorithm in a much more adaptive and optimal way.

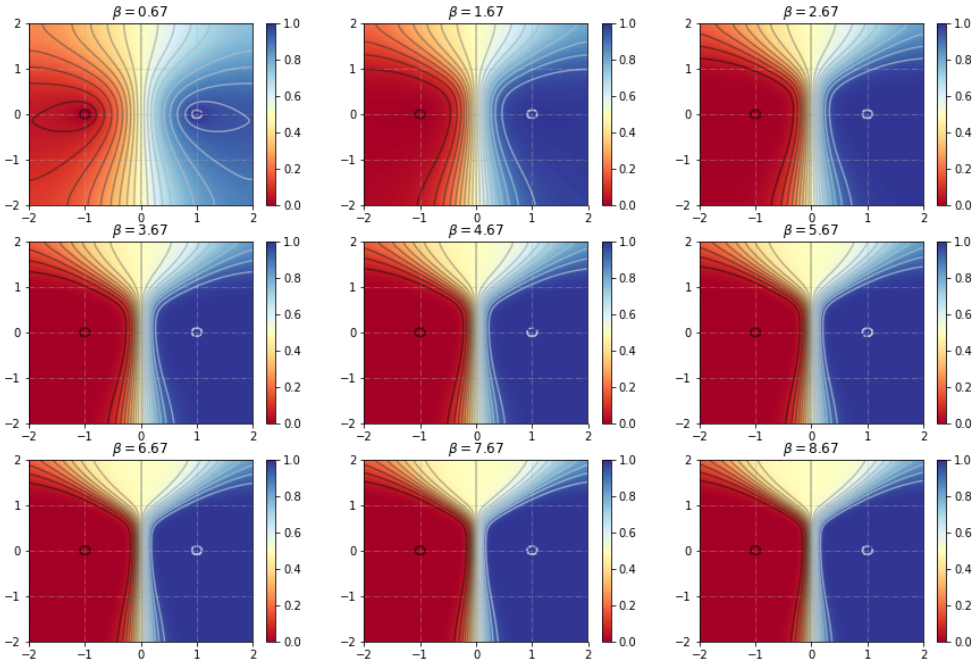

Figure 4.16: Illustrations of committor functions for different temperatures.

4.6 Discussions

In this chapter, we mainly discussed the possibility of using MF to estimate committor functions. This paves the way to a wide range of possible interacting strategies of gAMS and MF. However, a more refined study still needs to be conducted.

The first remark is on the sample points of the training data. In high-dimensional settings, the uniform sampling strategy is typically intractable. In fact, one does not need to estimate the whole picture of the committor function if one only seeks to improve the performance of gAMS. A possible approach is to sample along the reactive trajectories generated by gAMS. In this way, it is expected that the prediction of MF to be more accurate along these transition paths, which yields a more accurate gAMS estimator. However, to develop a proper sampling strategy for the sample points is not trivial. For example, for the overdamped Langevin dynamics, the reactive trajectories stay longer in the well of the energy landscape, which means that by uniformly sampling the points along the reactive trajectories is not efficient in general. Hence, the details still need to be investigated.

The second remark is on hybrid approaches that combines gAMS and crude Monte Carlo. In fact, the implementation of gAMS is mainly due to the behaviors of committor function close to the metastable states A and B. We remark that when close to B the rare event is of form $1 - p_*$, namely the values of the committor function is very close to 1. This means that in a lot of places where the values of the committor function is around 0.5, crude Monte Carlo is already able to provide reasonable estimations. Therefore, the design of an adaptive hybrid approach may greatly increase the efficiency of the algorithm.

The final remark is on Mondrian Forests. Since it is proved that finite lifetime parameter λ yields non-consistent estimator, it would be interesting to explore another alternative stopping criterion for the growth of MT, in order to improve its robustness against largely perturbed data. One possible choice is standard in a Random Forests context, that is to fix a threshold to control the minimum number of data in each subset of the partition of the decision tree. Said differently, we fix a number $N_{\min}$ in $\mathbf{N}$ such that when a subset in the partition contains less data than $N_{\min}$, the split is rejected. The ideal case is that we can somehow design a strategy such that this number $N_{\min}$ can be evaluated by the variance of the output of the data. Since variance estimation is available for both crude Monte Carlo and gAMS with MF as reaction coordinate, it would then be possible to develop more automated and advanced algorithms.

Appendices

Appendix A

Generalized AMS under Asymmetric SMC framework

This appendix sets out to stress the implementation and numerical performance of the generalized AMS method ([BGG+16]) with discrete reaction coordinate under Asymmetric SMC framework. The equivalence of AMS and Asymmetrical SMC is not straightforward to implement in practice—several adaptations should be taken into consideration, especially for the variance estimators. This implementation will simply be referred to as AMS for short in the following. An optimized and structured (readable) `python` package (powered by `Cython` [BBC+11]) is provided in the following `github` repository:

https://github.com/MGIMM/aAMS

Several numerical experiments on the performance of different variance estimators will also be provided. In particular, we are interested in the behaviors of the Variance of different Variance estimators. The Variance of Variance estimators (VoV) will be used to assess the quality of different variance estimators (through the length of the 95% confidential interval). Finally, we provide some general suggestions on the application of AMS methods under Asymmetric SMC framework.

A.1 Setting

Before proceeding further, we remark that all the numerical experiments are based on the Overdamped Langevin dynamic with three-hole potential (see Section 4.2.2). The discretization step is fixed to $\mathtt{dt} = 0.05$. In order to make sure the reaction coordinate only take finite values, we introduce a new parameter ι (`iota`) that indicates the discretization step of the reaction coordinate. More precisely, we consider the reaction coordinate

$$\xi(x) = \sqrt{(x+1)^2 + y^2}.$$

The corresponding discretized version ξ_ι is defined by

$$\forall x \in \mathbf{R}^2, \quad \xi_\iota(x) = \xi\left(\lfloor x/\iota \rfloor \iota\right).$$

The metastable states are defined as follows:

$$A := \{(x,y) \in \mathbf{R}^d : (x+1)^2 + y^2 < 0.2^2\} \text{ and } B := \{(x,y) \in \mathbf{R}^d : (x-1)^2 + y^2 < 0.2^2\}.$$

Our goal is to estimate the rare-event probability

$$p_* = \mathbf{P}\left(\tau_B < \tau_A \mid X_0 = (-0.75, 0)\right),$$

where $\tau_A := \inf\{t \geq 0 \mid X_t \in A\}$ and $\tau_B := \inf\{t \geq 0 \mid X_t \in B\}$. If not mentioned otherwise, the inverse temperature $\beta = 4.1$ and the final level $L^* = 1.75$. In this case, the reference value of p_* based on Naive Monte Carlo is 3.368×10^{-5}, and the associated naive asymptotic variance estimator is 3.374×10^{-5}.

A.2 Uncertainty control with a single simulation

Now, let us implement AMS with the efficient variance estimator provided in Section 3.3.5. The illustration is shown in Figure A.1. Note that the asymptotic variance for the AMS with $N = 50000$ is around 1.844×10^{-7}, which is nearly 200 times better than the Naive Monte Carlo with the same particle size.

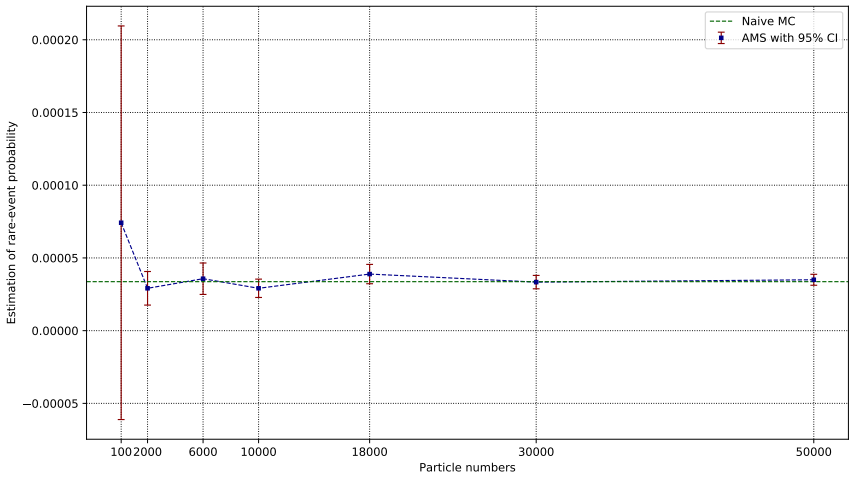

Figure A.1: An illustration for the estimation of AMS method along with 95% confidential interval. The number of particles N varies from 100 to 50000 and $\iota = 0.01$.

A.3 VoV behaviors

Let us continue to see the comportment of the VoV with different particle size N varies from 10 to 500. We sample the IPS independently for 200 to 10000 times, and the confidential interval is obtained from the crude variance estimation of our variance estimators. The illustration can be found in Figure A.2-A.6. First, it is clear that when N is small, the efficient asymptotic variance estimator is heavily biased (see Figure A.2-A.3). Then, it also shows that the average of the unbiased variance estimators over the independent IPS depends strongly on the crude variance estimator. Finally, we remark that from a relatively small particle size ($N = 500$, Figure A.6), the VoV of efficient variance estimator becomes noticeably smaller than the unbiased variance estimator. Hence, when using one particle system of large size for the rare-event estimation, the efficient variance estimator is expected to be more accurate and more efficient at the same time.

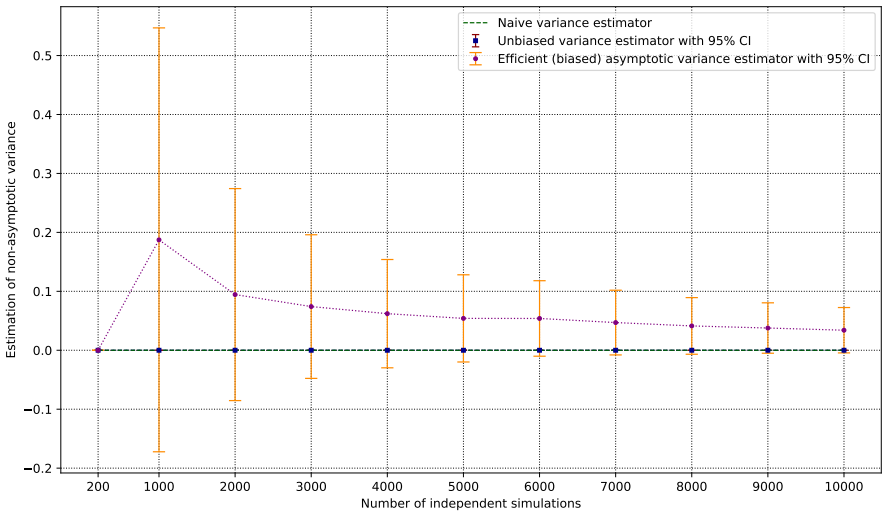

Figure A.2: An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 10$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators.

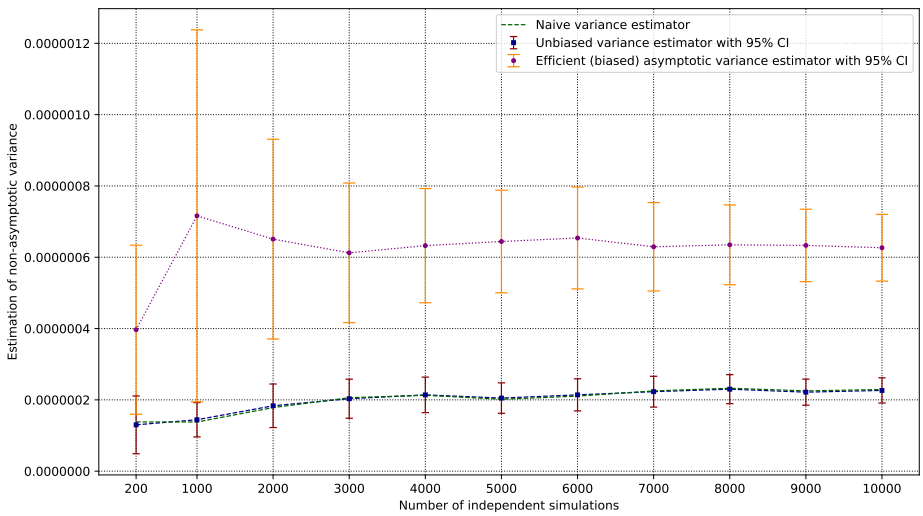

Figure A.3: An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 50$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators.

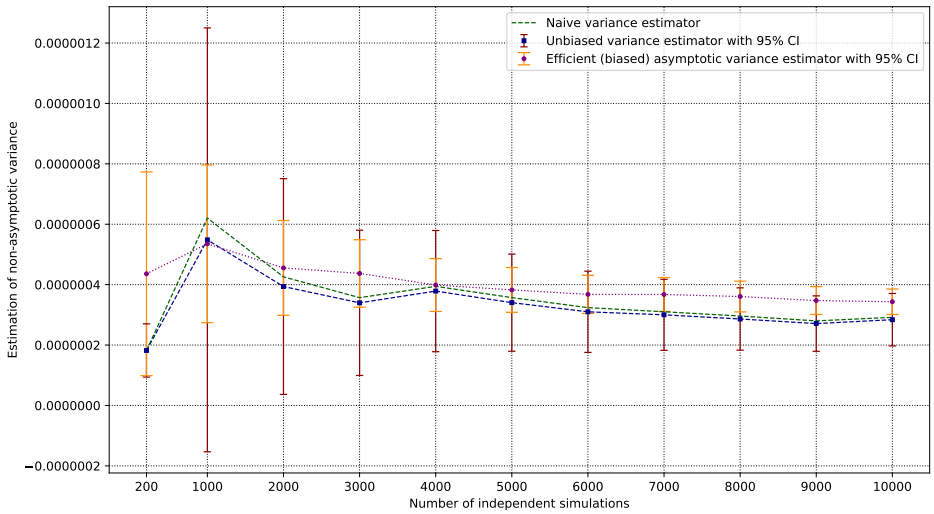

Figure A.4: An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 100$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators.

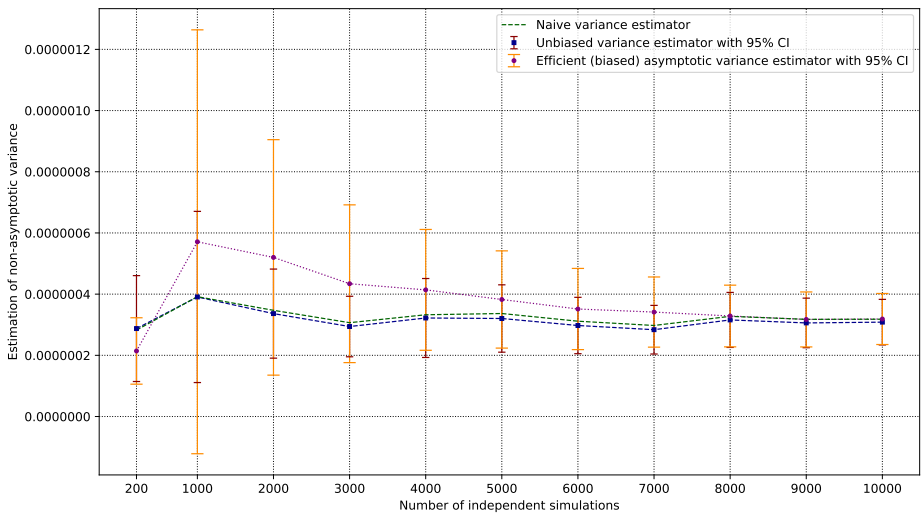

Figure A.5: An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 200$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators.

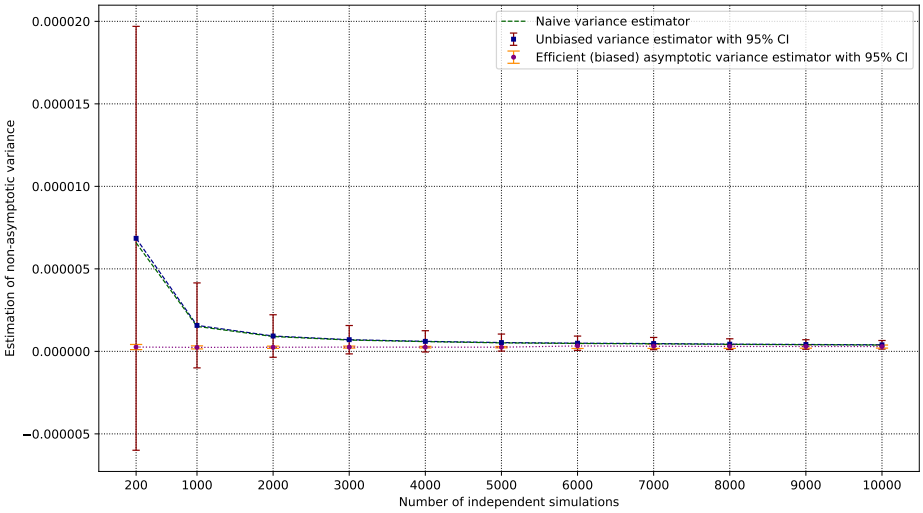

Figure A.6: An illustration for the estimation of asymptotic variance with 95% confidential interval. The number of particles $N = 500$ and $\iota = 0.01$. The naive variance estimator is computed on the same sample of unbiased variance estimators.

Bibliography

[AB01] S.K. Au and J.L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4):263–277, 2001.

[AB03] S.K. Au and J.L. Beck. Subset simulation and its application to seismic risk based on dynamic analysis. *Journal of Engineering Mechanics*, 129(8):901–917, 2003.

[ADH10] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 72(3):269–342, 2010.

[ALV18] Christophe Andrieu, Anthony Lee, and Matti Vihola. Uniform ergodicity of the iterated conditional smc and geometric ergodicity of particle gibbs samplers. *Bernoulli*, 24(2):842–872, 05 2018.

[BBC+11] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D.S. Seljebotn, and K. Smith. Cython: The best of both worlds. *Computing in Science Engineering*, 13(2):31 –39, 2011.

[BGG+16] Charles-Edouard Bréhier, Maxime Gazeau, Ludovic Goudenège, Tony Lelièvre, and Mathias Rousset. Unbiasedness of some generalized adaptive multilevel splitting algorithms. *Ann. Appl. Probab.*, 26(6):3559–3601, 2016.

[BJKT16] A. Beskos, A. Jasra, N. Kantas, and A. Thiery. On the convergence of adaptive sequential Monte Carlo methods. *Ann. Appl. Probab.*, 26(2):1111–1146, 2016.

[BLR15] Bréhier, Charles-Edouard, Lelièvre, Tony, and Rousset, Mathias. Analysis of adaptive multilevel splitting algorithms in an idealized case. *ESAIM: PS*, 19:361–394, 2015.

[Buc10] James Bucklew. *Introduction to Rare Event Simulation.* Springer Publishing Company, Incorporated, 1st edition, 2010.

[BW15] Matej Balog and Yee Whye Teh. The Mondrian Process for Machine Learning. *arXiv e-prints*, page arXiv:1507.05181, Jul 2015.

[CDGR18a] Frédéric Cérou, Bernard Delyon, Arnaud Guyader, and Mathias Rousset. On the Asymptotic Normality of Adaptive Multilevel Splitting. *arXiv e-prints*, page arXiv:1804.08494, Apr 2018.

[CDGR18b] Frédéric Cérou, Bernard Delyon, Arnaud Guyader, and Mathias Rousset. On the asymptotic normality of adaptive multilevel splitting. *SIAM/ASA Journal on Uncertainty Quantification*, 7, 04 2018.

[CDMFG12] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.

[CDMG11] F. Cérou, P. Del Moral, and A. Guyader. A nonasymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(3):629–649, 2011.

[CG07] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.*, 25(2):417–443, 2007.

[CG16] F. Cérou and A. Guyader. Fluctuation analysis of adaptive multilevel splitting. *Ann. Appl. Probab.*, 26(6):3319–3380, 2016.

[CGR19a] Frédéric Cérou, Arnaud Guyader, and Mathias Rousset. On Synchronized Fleming-Viot Particle Systems. *arXiv e-prints*, page arXiv:1911.05366, Nov 2019.

[CGR19b] Frédéric Cérou, Arnaud Guyader, and Mathias Rousset. Adaptive multilevel splitting: Historical perspective and recent results. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(4):043108, 2019.

[Cho04] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6):2385–2411, 2004.

[CJP13] N. Chopin, P. E. Jacob, and O. Papaspiliopoulos. SMC^2: an efficient algorithm for sequential analysis of state space models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 75(3):397–426, 2013.

[CL13] H. P. Chan and T. L. Lai. A general theory of particle filters in hidden Markov models and some applications. *Ann. Statist.*, 41(6):2877–2904, 2013.

[Cog80] Robert Cogburn. Markov chains in random environments: The case of markovian environments. *Ann. Probab.*, 8(5):908–916, 10 1980.

[CRGP15] Nicolas Chopin, James Ridgway, Mathieu Gerber, and Omiros Papaspiliopoulos. Towards automatic calibration of the number of state particles within the SMC^2 algorithm. *ArXiv preprint 1506.00570*, Jun 2015.

[DCGR17] Bernard Delyon, Frédéric Cérou, Arnaud Guyader, and Mathias Rousset. A Central Limit Theorem for Fleming-Viot Particle Systems with Hard Killing. *arXiv e-prints*, page arXiv:1709.06771, Sep 2017.

[DdFG01] A. Doucet, N. de Freitas, and N. Gordon, editors. *Sequential Monte Carlo methods in practice*. Statistics for Engineering and Information Science. Springer-Verlag, New York, 2001.

[DG19] Qiming Du and Arnaud Guyader. Variance estimation in adaptive sequential monte carlo. *arXiv*, 2019.

[DM96] P. Del Moral. Nonlinear filtering: interacting particle solution. *Markov Process. Related Fields*, 2(4):555–579, 1996.

[DM04] P. Del Moral. *Feynman-Kac formulae: Genealogical and interacting particle systems with applications*. Probability and its Applications. Springer-Verlag, New York, 2004.

[DM08] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, 36(5):2344–2376, 2008.

[DM13] P. Del Moral. *Mean field simulation for Monte Carlo integration*, volume 126 of *Monographs on Statistics and Applied Probability*. CRC Press, Boca Raton, FL, 2013.

[DMDJ12] Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. On adaptive resampling strategies for sequential monte carlo methods. *Bernoulli*, 18(1):252–278, 02 2012.

[DMKP16] P. Del Moral, R. Kohn, and F. Patras. On particle Gibbs samplers. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1687–1733, 2016.

[DMPR09] P. Del Moral, F. Patras, and S. Rubenthaler. Tree based functional expansions for Feynman-Kac particle models. *Ann. Appl. Probab.*, 19(2):778–825, 2009.

[EVE10] Weinan E and Eric Vanden-Eijnden. Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, 61(1):391–420, 2010. PMID: 18999998.

[GCW17] Mathieu Gerber, Nicolas Chopin, and Nick Whiteley. Negative association, ordering and convergence of resampling methods. *arXiv e-prints*, page arXiv:1707.01845, Jul 2017.

[GDM17] François Giraud and Pierre Del Moral. Nonasymptotic analysis of adaptive and annealed feynman–kac particle models. *Bernoulli*, 23(1):670–709, 02 2017.

[Gor93] N.J. Gordon. Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140:107–113(6), April 1993.

[HSG06] J. D. Hol, T. B. Schon, and F. Gustafsson. On resampling algorithms for particle filters. In *2006 IEEE Nonlinear Statistical Signal Processing Workshop*, pages 79–82, Sep. 2006.

[Jan11] Svante Janson. Probability asymptotics: notes on notation. *arXiv e-prints*, page arXiv:1108.3924, Aug 2011.

[KH51] H. Kahn and T. E. Harris. Estimation of particle transmission by random sampling. *National Bureau of Standards Applied Mathematical Series*, 12:27–30, 1951.

[LC98] Jun S. Liu and Rong Chen. Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.*, 93(443):1032–1044, 1998.

[LL19] Laura J. S. Lopes and Tony Lelièvre. Analysis of the adaptive multilevel splitting method on the isomerization of alanine dipeptide. *Journal of Computational Chemistry*, 40(11):1198–1208, 2019.

[LRT14] Balaji Lakshminarayanan, Daniel M Roy, and Yee Whye Teh. Mondrian forests: Efficient online random forests. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3140–3148. Curran Associates, Inc., 2014.

[LW18] A. Lee and N. Whiteley. Variance estimation in the particle filter. *Biometrika*, 105(3):609–625, 2018.

[McL74] D. L. McLeish. Dependent central limit theorems and invariance principles. *Ann. Probability*, 2:620–628, 1974.

[MGS17] Jaouad Mourtada, Stéphane Gaïffas, and Erwan Scornet. Universal consistency and minimax rates for online mondrian forests. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3758–3767. Curran Associates, Inc., 2017.

[MGS18] Jaouad Mourtada, Stéphane Gaïffas, and Erwan Scornet. Minimax optimal rates for mondrian trees and forests. *arXiv preprint arXiv:1803.05784*, 2018.

[MSVE06] Philipp Metzner, Christof Schütte, and Eric Vanden-Eijnden. Illustration of transition path theory on a collection of simple examples. *The Journal of chemical physics*, 125:084110, 09 2006.

[OD19] Jimmy Olsson and Randal Douc. Numerically stable online estimation of variance in particle filters. *Bernoulli*, 25(2):1504–1535, 05 2019.

[RC10] Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer Publishing Company, Incorporated, 2010.

[RR55] Marshall N. Rosenbluth and Arianna W. Rosenbluth. Monte carlo calculation of the average extension of molecular chains. *The Journal of Chemical Physics*, 23(2):356–359, 1955.

[RT09] Daniel M Roy and Yee W. Teh. The mondrian process. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1377–1384. Curran Associates, Inc., 2009.

[VE14] Eric Vanden-Eijnden. *Transition Path Theory*, pages 91–100. Springer Netherlands, Dordrecht, 2014.