Fluctuation Analysis
of Adaptive Multilevel Splitting

Frédéric Cérou[1]
INRIA Rennes & IRMAR, France
frederic.cerou@inria.fr

Arnaud Guyader
Université Pierre et Marie Curie, France
arnaud.guyader@upmc.fr

Abstract

Multilevel Splitting, also called Subset Simulation, is a Sequential Monte Carlo method to simulate realisations of a rare event as well as to estimate its probability. This article is concerned with the convergence and the fluctuation analysis of Adaptive Multilevel Splitting techniques. In contrast to their fixed level version, adaptive techniques estimate the sequence of levels on the fly and in an optimal way, with only a low additional computational cost. However, very few convergence results are available for this class of adaptive branching models, mainly because the sequence of levels depends on the occupation measures of the particle systems. This article proves the consistency of these methods as well as a central limit theorem. In particular, we show that the precision of the adaptive version is the same as the one of the fixed-levels version where the levels would have been placed in an optimal manner.

Index Terms — Sequential Monte Carlo, Rare events, Interacting particle systems, Feynman-Kac semigroups.

2010 Mathematics Subject Classification: 47D08, 65C35, 60J80, 65C05.

1   Introduction

Multilevel Splitting techniques were introduced as natural heuristics in the 1950s' by Kahn and Harris [27] and Rosenbluth and Rosenbluth [32] to analyze particle transmission energies and molecular polymer conformations. In their basic form, these methods can be interpreted as a genetic mutation-selection algorithm. The mutation transition reflects the free evolution of the

---

[1]Corresponding author.

physical model at hand, while the selection is an acceptance-rejection transition equipped with a recycling mechanism. The samples entering a critical level are more likely to be selected and duplicated. The genealogy associated with these genetic type particles model represents the statistical behavior of the system passing through a cascade of critical rare events.

Interestingly, these models can also be seen as a mean field particle approximation of Feynman-Kac measures. This interpretation depends on the application area of interest. In scientific computing and mathematical biology, these stochastic techniques are often termed genetic algorithms. In machine learning and advanced signal processing, they are referred as Sequential Monte Carlo or Particle Filters. In computational and quantum physics, they belong to the class of Diffusion Monte Carlo methods. The analysis of this class of branching and mean field type particle methods is now well understood (see, for instance, [9, 10, 14, 17] and references therein).

The present article is concerned with the convergence analysis of a more sophisticated class of adaptive particle methods where both the selection functions and the mutation transitions depend on the occupation of the system. The selection functions are chosen to sequentially achieve a prescribed proportion of samples in an higher critical level set, while the mutation transitions are dictated by some Markov Chain Monte Carlo (MCMC) methods on the adaptive critical level sets. This adaptive multilevel technique is a natural and popular approach amongst practitioners, but there are very few convergence results for this class of models.

The first analysis of this class of models has been developed by Del Moral, Doucet and Jasra in [16], in terms of adaptive resampling times associated with some criteria such as the effective sample size. Nonetheless, their result only applies to adaptive models associated with parametric level sets and equipped with sequential resampling times related to some fixed threshold. Thus, this does not correspond precisely to the purpose of Adaptive Multilevel Splitting methods that we are presently interested in.

In a slightly different framework, the recent article by Beskos, Jasra, Kantas and Thiéry [3] is also related to the present paper. Specifically, the authors present a detailed analysis for a class of adaptive Sequential Monte Carlo models under regularity properties on the dependency of the mutation transitions and the selection functions with respect to the occupation measures of the system. The proofs in [3] reveal that these regularity properties are essential to develop a first order perturbation analysis between the adaptive particle models and their limiting measures. Unfortunately, this framework does not apply to indicator selection functions arising in classical

multilevel splitting methodologies and developed in the present article. As a consequence, even if the global goal here is roughly the same as in [16, 3], the techniques developed for establishing our convergence results are quite different. Note also that in the context of adaptive tempering (a context considered in [3]), Giraud and Del Moral give non-asymptotic bounds on the error in [22].

Let us first specify our framework and notation. In all the paper, we suppose that X is a random vector in $\mathbb{R}^d$ with law η that we can simulate, and S is a mapping from $\mathbb{R}^d$ to $\mathbb{R}$, also called a score function. Then, given a threshold $L^\star$ which lies far out in the right hand tail of the distribution of $S(X)$, our goal is to estimate the rare event probability $P = \mathbb{P}(S(X) > L^\star)$. This very general context includes applications in queuing networks, insurance risks, random graphs (as found in social networks, or epidemiology), etc., see e.g. [24] for some of them, and a discussion on practical implementations.

In this context, a crude Monte Carlo uses an i.i.d. N-sample $X_1, \ldots, X_N$ to estimate P by the fraction $\hat{P}_{mc} = \#\{i : S(X_i) > L^\star\}/N$. However, in order to obtain a reasonable precision of the estimate given by the relative variance $\mathbb{V}(\hat{P}_{mc})/P^2 = (1 - P)/(NP)$, one needs a sample size N of order at least P^{-1}. Obviously, this becomes unrealistic when P is very small, hence the use of variance reduction techniques.

Importance Sampling, which draws samples according to π and weights each observation $X = x$ by $w(x) = d\eta(x)/d\pi(x)$, may decrease the variance of the estimated probability dramatically, which in turn reduces the need for such large sample sizes. We refer to Robert and Casella [31] for a discussion on Importance Sampling techniques in general, and to Bucklew [7] and L'Ecuyer, Mandjes and Tuffin [33, Chapter 2] for the application in the context of rare event estimation. Notice that, in rare event estimation, it is customary to design an importance sampling scheme using a large deviation principle. Although it often gives an efficient method, this approach may fail dramatically, even compared to crude Monte Carlo, when the rare event has two or more most likely occurrences. As explained by Glasserman and Wang in the introduction of [23], "Simply put, an analysis of a first moment cannot be expected to carry a guarantee about the behavior of a second moment."

Multilevel Splitting represents another powerful algorithm for rare event estimation. The basic idea of Multilevel Splitting, adapted to our problem, is to fix a set of increasing levels $-\infty = L_{-1} < L_0 < \cdots < L_{n-1} < L_n = L^\star$, and to decompose the tail probability thanks to Bayes formula, that is

$$\mathbb{P}(S(X) > L^\star) = \prod_{p=0}^{n} \mathbb{P}(S(X) > L_p | S(X) > L_{p-1}).$$

3

Each conditional probability $\mathbb{P}(S(X) > L_p | S(X) > L_{p-1})$ is then estimated separately. We refer the reader to L'Ecuyer, Le Gland, Lezaud and Tuffin [33, Chapter 3] for an in-depth review of the Multilevel Splitting method and a detailed list of references. Two practical issues associated with the implementation of Multilevel Splitting are: first, the need for computationally efficient algorithms for estimating the successive conditional probabilities; second, the optimal selection of the sequence of levels.

The first question can be addressed thanks to the introduction of Markov Chain Monte Carlo procedures at each step of the algorithm. This trick was proposed in different contexts and through slightly different variants by Au and Beck [1, 2], Del Moral, Doucet and Jasra [15], Botev and Kroese [5], Rubinstein [34].

The second question is straightforward in the idealized situation where one could estimate the successive quantities $\mathbb{P}(S(X) > L_p | S(X) > L_{p-1})$ independently at each step. Indeed, considering the variance of the estimator, it is readily seen that the best thing to do is to place the levels as evenly as possible in terms of the intermediate probabilities, that is to take, for all p,

$$\mathbb{P}(S(X) > L_p | S(X) > L_{p-1}) = \mathbb{P}(S(X) > L^\star)^{\frac{1}{n+1}}.$$

But, since little might be known about the mapping S, the only way to achieve this goal is to do it on the fly by taking advantage of the information of the current sample at each step. This method is called Subset Simulation (see Au and Beck [1, 2]) or Adaptive Multilevel Splitting (see Cérou and Guyader [11]), and may be seen as an adaptive Sequential Monte Carlo method specifically dedicated to rare event estimation.

However, except in the idealized situation where one considers a new independent sample at each step (see Cérou, Del Moral, Furon and Guyader [8], Guyader, Hengartner and Matzner-Løber [25], Bréhier, Lelièvre and Rousset [6], and Simonnet [36]), there are only very few results about the theoretical properties of this efficient algorithm. From a broader point of view, as duly noticed in [16, 3], this disparity between theory and practice holds true for adaptive Sequential Monte Carlo methods in general. As such, the present article is in the same vein as [16, 3] and might be seen as a new step towards a better understanding of the statistical properties of adaptive Sequential Monte Carlo methods.

In particular, the take-home message here is the same as in [16, 3], namely that the asymptotic variance of the adaptive version is the same as the one of the fixed-levels version where the levels would have been placed in an optimal

manner. However, there are substantial differences between [16, 3] and the present contribution.

In [16], the adaptive parameter is the time at which one needs to resample. This approach can be used for rare event if we choose a possibly long sequence of deterministic levels $L_1, \ldots, L_n$, and resample only when the current level sees a given proportion of particles to be already killed. The authors provide convergence results, including a CLT, when the number N of particles goes to ∞, but for fixed levels $L_1, \ldots, L_n$. To get the kind of results of the present contribution, one would need to let also n go to ∞, and this cannot be achieved by the coupling technique used in [16] due to the inherent jittering of the adaptive levels, which is typically of order $1/\sqrt{N}$. If the granularity of the levels goes to 0 as N goes to ∞, then there is little hope that the adaptive particle system coincides with the optimal one with large probability as in their Theorem 2.3.

In [3], the authors consider different scenarios, including adaptive proposal and adaptive tempering, where they can make a Taylor expansion of the adaptive selection function, and the adaptive kernel, in the vicinity of the optimal parameter. This leads to additional terms in the asymptotic variance that may cancel in some cases (adaptive proposal), giving the same variance as in the non adaptive optimal case. Yet, let us emphasize again that the inherent unsmoothness of the selection functions of interest here (going abruptly from 0 to 1 when crossing a level set for S) leads to different proofs, meaning that their results and even techniques, although very interesting in and by themselves, can definitely not be applied in our context.

The paper is organized as follows. In Section 2, we introduce some notation and describe the Multilevel Splitting algorithms. The asymptotic results (laws of large numbers and central limit theorems) are presented in Section 3. Section 4 comes back on the assumption required for our CLT type result to be valid. Section 5 is devoted to the proofs of the theorems, while technical results are postponed to Section 6.

2 Multilevel splitting techniques

2.1 Framework and notation

We consider an $\mathbb{R}^d$-valued random variable X with distribution η, for some $d \geq 1$. We assume that η has a density with respect to Lebesgue's measure dx on $\mathbb{R}^d$ and, by a slight abuse of notation, we denote $\eta(x)$ this density. We also consider a mapping S from $\mathbb{R}^d$ to $\mathbb{R}$. If S is Lipschitz with $|DS| > 0$

almost everywhere, where $|DS|$ stands for the Euclidean norm of the gradient of S, then the coarea formula (see for example [21], page 118, Proposition 3) ensures that the random variable $Y = S(X)$ is absolutely continuous with respect to Lebesgue's measure on $\mathbb{R}$, and its density is given by the formula

$$f_Y(s) = \int_{S(x)=s} \eta(x) \frac{dx}{|DS(x)|}, \qquad (2.1)$$

where dx stands for the Hausdorff measure on the level set $S^{-1}(s) = \{x \in \mathbb{R}^d, S(x) = s\}$. In this notation, given $\alpha \in (0,1)$, the $(1-\alpha)$ quantile of Y is simply $F_Y^{-1}(1-\alpha)$, where F_Y stands for the cumulative distribution function (cdf for short) of Y.

Consider a real number (or level) $L^\star$ lying far away in the right hand tail of $S(X)$ so that the probability $P = \mathbb{P}(Y \geq L^\star)$ is very small. For any bounded and measurable function $f : \mathbb{R}^d \to \mathbb{R}$ (denoted $f \in \mathcal{B}(\mathbb{R}^d)$ in all the paper) which is null below $L^\star$ (implicitly: with respect to S), our goal is to estimate its expectation with respect to η, that is the quantity

$$E = \mathbb{E}[f(X)] = \mathbb{E}[f(X)\mathbf{1}_{S(X) \geq L^\star}]. \qquad (2.2)$$

To this end, we fix an $\alpha \in (0,1)$ (in practice one may typically choose $\alpha = 3/4$), and consider the decomposition

$$P = \mathbb{P}(Y \geq L^\star) = r \times \alpha^n \quad \text{with} \quad n = \left\lfloor \frac{\log \mathbb{P}(Y \geq L^\star)}{\log \alpha} \right\rfloor, \qquad (2.3)$$

so that $r \in (\alpha, 1]$. For the sake of simplicity and since this is always the case in practice, we assume that r belongs to the open interval $(\alpha, 1)$. With the convention $L_{-1} = -\infty$, we define the increasing sequence of levels $(L_p)_{p \geq -1}$ as follows

$$L_0 = F_Y^{-1}(1-\alpha) < \cdots < L_{n-1} = F_Y^{-1}(1-\alpha^n) < L^\star < L_n = F_Y^{-1}(1-\alpha^{n+1}).$$

Once and for all, we assume that the density f_Y, as defined in equation (2.1), is **continuous** and **strictly positive** at each L_p, for $p \in \{0, \ldots, n\}$. This will guarantee that the quantiles are well defined and that the empirical ones have good convergence properties.

Following the notations of [13, 14], we associate to these successive levels the potential functions

$$\forall -1 \leq p < n, \qquad G_p = \mathbf{1}_{\mathcal{A}_p} \quad \text{with} \quad \mathcal{A}_p = \{x \in \mathbb{R}^d : S(x) \geq L_p\}.$$

6

The restriction of η to $\mathcal{A}_{p-1}$ is then denoted η_p. More formally, we have

$$\eta_p(dx) = \alpha^{-p} \mathbf{1}_{\mathcal{A}_{p-1}}(x)\eta(x)dx = \alpha^{-p} G_{p-1}(x)\eta(x)dx.$$

By construction, we have

$$\eta_p(G_p) = \eta_p(\mathbf{1}_{\mathcal{A}_p}) = \mathbb{P}\left(S(X) \geq L_p | S(X) \geq L_{p-1}\right) = \alpha.$$

We also notice that the interpolating measures η_p are connected by the Boltzmann-Gibbs transformation

$$\eta_{p+1}(dx) = \Psi_{G_p}(\eta_p)(dx) = \frac{1}{\eta_p(G_p)}G_p(x)\eta_p(dx) = \alpha^{-1}G_p(x)\eta_p(dx).$$

Moreover, we consider a collection of Markov transitions from $\mathcal{A}_{p-1}$ into itself defined for any $x \in \mathcal{A}_{p-1}$ by

$$M_p(x, dx') = K_p(x, dx')\mathbf{1}_{\mathcal{A}_{p-1}}(x') + K_p(x, \bar{\mathcal{A}}_{p-1})\ \delta_x(dx'),$$

where $\bar{\mathcal{A}}_{p-1} = \mathbb{R}^d - \mathcal{A}_{p-1}$, and K_p stands for a collection of η-reversible Markov transitions on $\mathbb{R}^d$, meaning that for all p and all couple (x, x'), we have the detailed balance equation

$$\eta(dx)K_p(x, dx') = \eta(dx')K_p(x', dx). \tag{2.4}$$

We extend M_p into a transition kernel on $\mathbb{R}^d$ by setting $M_p(x, dx') = \delta_x(dx')$ whenever $x \notin \mathcal{A}_{p-1}$. Under the assumption that K_p is η-symmetric, it is easy to check that M_p is η_p-invariant, meaning that $\eta_p M_p = \eta_p$ for all $p \geq 1$. In addition, we have the recursion

$$\eta_p(dx') = \alpha^{-1}(\eta_{p-1}Q_p)(dx') = \alpha^{-1}\int \eta_{p-1}(dx)Q_p(x, dx'),$$

with the integral operators

$$Q_p(x, dx') = G_{p-1}(x)M_p(x, dx').$$

Next, let us denote $(X_p)_{p \geq 0}$ a non homogeneous Markov chain with initial distribution $\eta_0 = \eta$ and elementary transitions M_{p+1}. In this situation, it is readily seen that

$$\alpha^n\ \eta_n(f) = \mathbb{E}\left[f(X_n)\prod_{q=0}^{n-1}G_q(X_q)\right] \quad \Longleftrightarrow \quad \alpha^n\ \eta_n = \eta_0 Q_{0,n}. \tag{2.5}$$

with the Feynman-Kac semigroup $Q_{0,n}$ associated with the integral operators Q_p defined by

$$\forall 0 \leq p \leq n \qquad Q_{p,n} = Q_{p+1} Q_{p+1,n}$$

In this notation, we have

$$
\begin{aligned}
E &= \mathbb{E}[f(X)] = \mathbb{E}[f(X)\mathbf{1}_{S(X)\geq L^\star}] = \alpha^n \times \eta_n(f \times \mathbf{1}_{S(\cdot)\geq L^\star}) \\
P &= \mathbb{P}(Y \geq L^\star) = \mathbb{P}(S(X) \geq L^\star) = \alpha^n \times \eta_n(\mathbf{1}_{S(\cdot)\geq L}) = \alpha^n \times r
\end{aligned}
$$

and

$$
f = f \times \mathbf{1}_{S(\cdot)\geq L^\star} \implies C = \mathbb{E}[f(X)|S(X) \geq L^\star] = \frac{\eta_n(f)}{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})} = \frac{\eta_n(f)}{r}.
$$

$$\tag{2.6}$$

We will now describe two multilevel splitting techniques in order to estimate these quantities. The optimal Feynman-Kac particle approximation of the flow (2.5) corresponds to the fixed-levels method that we describe in Section 2.2. As this approximation is not possible in practice, we detail in Section 2.3 the corresponding adaptive Feynman-Kac particle approximation, known as Adaptive Multilevel Splitting or Subset Simulation.

2.2 The fixed-levels method

Following the notation of [13], the fixed-levels approximation of the flow (2.5) works as follows. Let $(X_p^1, \ldots, X_p^N)_{0 \leq p \leq n}$ be an $(\mathbb{R}^d)^N$-valued Markov chain with initial distribution $\eta_0^{\otimes N}$ and for which each elementary transition $X_p^i \rightsquigarrow X_{p+1}^i$ is decomposed into the following separate mechanisms:

1. Selection step: compute $\check{\eta}_p^N(G_p)$, which is the proportion of the sample $(X_p^1, \ldots, X_p^N)$ such that $S(X_p^i) \geq L_p$.

2. Multinomial step: from the $\check{\eta}_p^N(G_p)N$-sample with distribution η_{p+1}, draw an N-sample $(X_{p+1/2}^1, \ldots, X_{p+1/2}^N)$ with the same distribution.

3. Transition step: each $X_{p+1/2}^i$ evolves independently to a new site X_{p+1}^i randomly chosen with distribution $M_{p+1}(X_{p+1/2}^i, dx')$.

4. Incrementation step: $p = p + 1$. If $p = n$, then stop the algorithm, else go to step 1 (selection step).

Let us denote $\check{\gamma}_n^N(1)$ the normalizing constant defined by

$$\check{\gamma}_n^N(1) = \prod_{p=0}^{n-1} \check{\eta}_p^N(G_p).$$

In our framework, its deterministic counterpart is simply

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

For any $f \in \mathcal{B}(\mathbb{R}^d)$, the normalized and unnormalized measures $\check{\eta}_n^N(f)$ and $\check{\gamma}_n^N(f)$ are respectively defined by

$$\check{\eta}_n^N(f) = \frac{1}{N}\sum_{i=1}^{N} f(X_n^i) \quad \text{and} \quad \check{\gamma}_n^N(f) = \check{\gamma}_n^N(1) \times \check{\eta}_n^N(f).$$

The fixed-levels algorithm provides the following estimates:

(i) The estimate of the expectation $E = \mathbb{E}[f(X)\mathbf{1}_{S(X)\geq L^\star}] = \gamma_n(f \times \mathbf{1}_{S(\cdot)\geq L^\star})$ is given by $\check{E} = \check{\gamma}_n^N(f \times \mathbf{1}_{S(\cdot)\geq L^\star})$.

(ii) The rare event probability $P = \mathbb{P}(S(X) \geq L^\star)$ is estimated by the quantity $\check{P} = \check{\gamma}_n^N(\mathbf{1}_{S(\cdot)\geq L^\star})$.

(iii) The estimate of the conditional expectation $C = \mathbb{E}[f(X)|S(X) \geq L^\star]$ is
$$\check{C} = \frac{\check{\eta}_n^N(f \times \mathbf{1}_{S(\cdot)\geq L^\star})}{\check{\eta}_n^N(\mathbf{1}_{S(\cdot)\geq L^\star})} = \frac{\sum_{i=1}^{N} f(X_n^i)\mathbf{1}_{S(X_n^i)\geq L^\star}}{\sum_{i=1}^{N} \mathbf{1}_{S(X_n^i)\geq L^\star}}.$$

These particle models associated with a collection of deterministic potential functions G_p and Markov transitions M_p belong to the class of Feynman-Kac particle models. This class of mean field particle models has been extensively studied in a very general context, including the asymptotic behavior as the number N of particles goes to infinity. We refer the reader to [13] and the more recent research monograph [14], with references therein. We will recall some of these results in Section 3.2.

In our specific context, the obvious drawback of these Feynman-Kac particle approximations is the impossibility to fix in advance the successive levels $L_0, \ldots, L_n$, hence the use of adaptive methods that we describe in the following section.

2.3 The adaptive method

An efficient way to estimate the quantities E, P and C is to use Adaptive Multilevel Splitting methods. To describe with some precision these particle

splitting models, it is convenient to consider a collection of potential functions and Markov transitions indexed by $\mathbb{R}$. Thus, for any real number L, we set

$$G_L = \mathbf{1}_{\mathcal{A}_L} \quad \text{with} \quad \mathcal{A}_L = \{x \in \mathbb{R}^d : S(x) \geq L\}.$$

We also consider the collection of Markov transitions from $\mathcal{A}_L$ into itself defined for any $x \in \mathcal{A}_L$ by

$$M_{p,L}(x, dx') = K_p(x, dx')\mathbf{1}_{\mathcal{A}_L}(x') + K_p(x, \bar{\mathcal{A}}_L)\delta_x(dx').$$

As before, we extend $M_{p,L}$ into a transition kernel on $\mathbb{R}^d$ by setting $M_{p,L}(x, dx') = \delta_x(dx')$ whenever $x \notin \mathcal{A}_L$, and we set

$$Q_{p,L}(x, dx') = G_L(x)M_{p,L}(x, dx').$$

In this slight abuse of notation, we have

$$L = L_{p-1} \quad \Longrightarrow \quad (G_L, \mathcal{A}_L) = (G_{p-1}, \mathcal{A}_{p-1}) \quad \text{and} \quad (M_{p,L}, Q_{p,L}) = (M_p, Q_p).$$

Of special interest will be the case where L is a given quantile. We distinguish two cases:

- Firstly, for any positive and finite measure ν on $\mathbb{R}^d$ with a density with respect to Lebesgue's measure, the level L_ν is defined as the $(1 - \alpha)$ quantile of the probability measure $(S_*\nu)/\nu(\mathbb{R}^d)$, that is

$$L_\nu = L(\nu) = F_\nu^{-1}(1 - \alpha) \quad \text{where} \quad F_\nu(y) = \frac{\nu(S^{-1}((-\infty, y]))}{\nu(\mathbb{R}^d)}. \quad (2.7)$$

In order to lighten the notations a bit, we will write

$$G_\nu := G_{L_\nu} \qquad \mathcal{A}_\nu := \mathcal{A}_{L_\nu} \qquad M_{p,\nu} := M_{p,L_\nu}$$

and

$$Q_{p,\nu}(x, dx') = G_\nu(x)M_{p,\nu}(x, dx').$$

- Secondly, given a sample of vectors $(X_i)_{1 \leq i \leq N}$ in $\mathbb{R}^d$, we consider an auxiliary sequence of i.i.d. uniformly distributed random variables $(U_1, \ldots, U_N)$ and the following total order on the couples $(X_i, U_i)_{1 \leq i \leq N}$:

$$(X_i, U_i) < (X_j, U_j) \Leftrightarrow S(X_i) < S(X_j) \text{ or } S(X_i) = S(X_j) \text{ and } U_i < U_j. \quad (2.8)$$

Obviously, since the U_i's are uniformly distributed, equality between two couples almost surely never happens. Hence we can consider the associated order statistics

$$(X_{(1)}, U_{(1)}) < \cdots < (X_{(n)}, U_{(n)}),$$

and we define the empirical $(1 - \alpha)$ quantile L^N as

$$L^N := (S(X_{(\lfloor N(1-\alpha)\rfloor)}), U_{(\lfloor N(1-\alpha)\rfloor)}). \tag{2.9}$$

In particular, one can notice that the number of couples strictly above L^N (with respect to the previous order) is equal to $\lceil N\alpha \rceil$.

Remarks:

- In order to lighten the writings, we will usually not mention that some auxiliary uniform random variables $(U_1, \ldots, U_N)$ are **always** attached to a sample $(X_1, \ldots, X_N)$. In particular, this will be implicit to define the associated empirical quantile L^N, and the relation $S(X_i) > L^N$ must be understood accordingly. Otherwise, the notation L^N refers only to its first component in definition (2.9). However, considering the context, there should be no ambiguity. Note also that when considering the convergence of empirical quantiles, only the first component will be considered, as there is no reason why the uniform random variable would converge.

- In our situation, it turns out that equality between several X_i's will have no influence on the CLT type result we want to establish. Indeed, ties come from the multinomial step of the algorithm below, but as mentioned in the proof of Lemma 6.3, one can control these events very precisely (see for example Chapter 4 of [18]). As we will see, the maximum number of particles on the same level set for S, at stage q, is typically $o_p((\log N)^q)$, making the mass at a single location $o_p((\log N)^q/N)$, while the error between the conditional measure η_q and its particle approximation will be of order $1/\sqrt{N}$ as expected.

In this context, the **adaptive** particle approximation of the flow (2.5) is defined in terms of an $(\mathbb{R}^d)^N$-valued Markov chain $(X_p^1, \ldots, X_p^N)_{p \geq 0}$ with initial distribution $\eta_0^{\otimes N}$. We start with $p = 0$ and a $\eta_0^{\otimes N}$ sample $(X_0^1, \ldots, X_0^N)$. The elementary transitions $X_p^i \rightsquigarrow X_{p+1}^i$ are decomposed into the following separate mechanisms:

1. Quantile step: compute the empirical $(1-\alpha)$ quantile L_p^N of the sample $(X_p^1, \ldots, X_p^N)$ in the sense of (2.9). If $L_p^N \geq L^\star$, then stop the algorithm, else go to step 2 (multinomial step).

2. Multinomial step: draw an N-sample $(X_{p+1/2}^1, \ldots, X_{p+1/2}^N)$ with common distribution

$$\tilde{\eta}_p^N(dx) = \frac{1}{\lceil N\alpha \rceil} \sum_{i:\ S(X_p^i) > L_p^N} \delta_{X_p^i}(dx) = \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} \delta_{\tilde{X}_p^i}(dx).$$

11

3. Exploration step: each $X^i_{p+1/2}$ evolves independently to a new site X^i_{p+1} randomly chosen with distribution $M_{p+1,L^N_p}(X^i_{p+1/2}, dx')$.

4. Incrementation step: $p = p + 1$. Go to step 1.

Denote $\hat{n}$ the last index p such that $L^N_p < L^\star$. This algorithm provides the following estimates:

(i) The estimate of the expectation $E = \mathbb{E}[f(X)] = \mathbb{E}[f(X)\mathbf{1}_{S(X)\geq L^\star}]$ considered in (2.2) is

$$\hat{E} = \alpha^{\hat{n}} \times \eta^N_{\hat{n}}(f) = \alpha^{\hat{n}} \times \frac{1}{N} \sum_{i=1}^N f(X^i_{\hat{n}})\mathbf{1}_{S(X^i_{\hat{n}})\geq L^\star}.$$

(ii) The rare event probability $P = \mathbb{P}(S(X) \geq L^\star)$ considered in (2.3) is estimated by the quantity

$$\hat{P} = \alpha^{\hat{n}} \times \eta^N_{\hat{n}}(\mathbf{1}_{S(\cdot)\geq L^\star}) = \alpha^{\hat{n}} \times \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{S(X^i_{\hat{n}})\geq L^\star}.$$

(iii) For the conditional expectation $C = \mathbb{E}[f(X)|S(X) \geq L^\star]$ considered in (2.6), still with $f = f \times \mathbf{1}_{S(\cdot)\geq L^\star}$, the estimate is

$$\hat{C} = \frac{\eta^N_{\hat{n}}(f)}{\eta^N_{\hat{n}}(\mathbf{1}_{S(\cdot)\geq L^\star})} = \frac{\sum_{i=1}^N f(X^i_{\hat{n}})\mathbf{1}_{S(X^i_{\hat{n}})\geq L^\star}}{\sum_{i=1}^N \mathbf{1}_{S(X^i_{\hat{n}})\geq L^\star}}.$$

The purpose of Section 3.1 is to expose some asymptotic results on these estimators.

2.4 Metropolis-Hastings kernels

Let us briefly recall Metropolis-Hastings algorithm [29, 26], which is a possible way to obtain a collection K_p of η-reversible Markov transitions. We emphasize that, from a practical viewpoint, the kernels K_p are a key ingredient of the previous algorithms, for fixed levels as well as for adaptive ones. Hereafter we follow the presentation of [37].

Let $\mathbf{k}_p$ be a Markov transition kernel of the form

$$\mathbf{k}_p(x, dx') = \mathbf{k}_p(x, x')dx'.$$

Let $E^+ = \{x \in \mathbb{R}^d,\ \eta(x) > 0\}$ and, for the sake of simplicity, assume that $\mathbf{k}_p(x, E^+) = 1$ for any $x \notin E^+$. Next define the acceptance ratio

$$a_p(x, x') = \begin{cases} \min\left\{\dfrac{\eta(x')\mathbf{k}_p(x', x)}{\eta(x)\mathbf{k}_p(x, x')}, 1\right\} & \text{if } \eta(x)\mathbf{k}_p(x, x') > 0, \\ 1 & \text{if } \eta(x)\mathbf{k}_p(x, x') = 0. \end{cases}$$

The success of Metropolis-Hastings algorithm comes from the fact that it only depends on η through ratios of the form $\eta(x')/\eta(x)$, hence η only needs to be known up to a normalizing constant. If we define the off-diagonal density of a Metropolis kernel as

$$\mathbf{k}_p^a(x, x') = a_p(x, x')\mathbf{k}_p(x, x'),$$

and set

$$\mathbf{r}_p^a(x) = 1 - \int_{\mathbb{R}^d} \mathbf{k}_p^a(x, x')dx',$$

then the Metropolis kernel K_p can be written as

$$K_p(x, dx') = \mathbf{k}_p^a(x, x')dx' + \mathbf{r}_p^a(x)\delta_x(dx').$$

Since $\mathbf{k}_p^a$ satisfies the detailed balance equation

$$\eta(x)\mathbf{k}_p^a(x, x') = \eta(x')\mathbf{k}_p^a(x', x), \tag{2.10}$$

it follows that K_p is an η-reversible Markov transition kernel. Finally, let us mention that, for any function $\varphi \in \mathcal{B}(\mathbb{R}^d)$, we will denote $\mathbf{k}_p^a(\varphi)$ the function defined by

$$\mathbf{k}_p^a(\varphi)(x) = \int_{\mathbb{R}^d} \mathbf{k}_p^a(x, x')\varphi(x')dx' = \int_{\mathbb{R}^d} a_p(x, x')\mathbf{k}_p(x, x')\varphi(x')dx',$$

so that

$$K_p(\varphi)(x) = \mathbf{k}_p^a(\varphi)(x) + \mathbf{r}_p^a(x) \times \varphi(x). \tag{2.11}$$

This expression will be useful in the proof of Proposition 5.3.

3　Consistency and fluctuation analysis

3.1　Adaptive Multilevel Splitting

We prove in Theorem 3.1 the almost sure convergence of L_p^N to L_p. As a byproduct, we deduce that the probability that the algorithm does not stop after the right number of steps (i.e., that $\hat{n} \neq n$) goes to zero when N goes to infinity. Then, in Theorem 3.2, we focus our attention on the fluctuations of $\eta_n^N(f)$ around $\eta_n(f)$.

13

Theorem 3.1 *For all $p \in \{0, \ldots, n\}$,*

$$L_p^N \xrightarrow[N \to \infty]{a.s.} L_p.$$

Besides, for all $f \in L^2(\eta)$,

$$\eta_p^N(f) \xrightarrow[N \to \infty]{\mathbb{P}} \eta_p(f),$$

and for all $f \in \mathcal{B}(\mathbb{R}^d)$,

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

Note that a consequence of Theorem 3.1 is that the couple (L_{n-1}^N, L_n^N) converges almost surely to (L_{n-1}, L_n). As claimed before, this ensures that, almost surely for N large enough, $L_{n-1}^N < L^\star < L_n^N$, which means that $\hat{n} = n$.

The fluctuations of η_n^N around the limiting measure η_n are expressed in terms of the normalized Feynman-Kac semigroups $\overline{Q}_{q,p}$ defined by

$$\forall 0 \leq q \leq p \leq n, \qquad \overline{Q}_{q,p} = \frac{Q_{q,p}}{\eta_q(Q_{q,p}(1))} = \alpha^{q-p} \times Q_{q,p}.$$

We also need to specify some regularity assumptions on the score function S and the transition kernels K_q for which our CLT type result is valid. For any $q > 0$, we first introduce the set of functions

$$\mathcal{B}_q = \left\{ g : \mathbb{R}^d \to \mathbb{R}, \ \exists (g_0 \ldots, g_{q-1}) \in \mathcal{B}(\mathbb{R}^d)^q, \ g = K_1(g_0) \cdots K_q(g_{q-1}) \right\}.$$

Notice in particular that any g in $\mathcal{B}_q$ is bounded and inherits the regularity properties of the kernels K_j. Then, for $g \in \mathcal{B}_q$, $x \in \mathbb{R}^d$ and $L \in \mathbb{R}$, let us denote

$$H_q^g(x, L) = \int_{S(x')=L} K_{q+1}(x, x') g(x') \frac{đx'}{|DS(x')|}.$$

Assumption [$\mathcal{H}$]

(i) For any $q \geq 0$, the mapping $x \mapsto H_q^1(x, L_q)$ belongs to $L^2(\eta)$, that is

$$\int \eta(dx) \left(\int_{S(x')=L_q} K_{q+1}(x, x') \frac{đx'}{|DS(x')|} \right)^2 < \infty.$$

14

(*ii*) For any $q > 0$, for any $g \in \mathcal{B}_q$, there exists $h \in L^2(\eta)$ such that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $L \in [L_q - \delta, L_q + \delta]$ and for almost every $x \in \mathbb{R}^d$,

$$\left| H_q^g(x, L) - H_q^g(x, L_q) \right| \leq \varepsilon h(x).$$

We will comment on this assumption in Section 4. In particular, we will see that it is not restrictive and is verified by most models of interest, for example when the level sets $\{S(x') = L\}$ have finite Hausdorff measure or when η and the kernels K_q have light tails.

If the kernels K_j are based on Metropolis-Hastings algorithm as detailed in Section 2.4, then one has to consider the set of functions

$$\mathcal{B}_q^a = \left\{ g : \mathbb{R}^d \to \mathbb{R}, \ \exists (g_0 \ldots, g_{q-1}) \in \mathcal{B}(\mathbb{R}^d)^q, \ g = \mathbf{k}_1^a(g_0) \cdots \mathbf{k}_q^a(g_{q-1}) \right\},$$

as well as the mapping

$$H_q^{g,a}(x, L) = \int_{S(x')=L} \mathbf{k}_{q+1}^a(x, x') g(x') \frac{\mathit{d}x'}{|DS(x')|},$$

and assumption $[\mathcal{H}^a]$ defined as follows.

Assumption $[\mathcal{H}^a]$

(*i*) For any $q \geq 0$, the mapping $x \mapsto H_q^{1,a}(x, L_q)$ belongs to $L^2(\eta)$, that is

$$\int \eta(dx) \left(\int_{S(x')=L_q} \mathbf{k}_{q+1}^a(x, x') \frac{\mathit{d}x'}{|DS(x')|} \right)^2 < \infty,$$

and $\eta(H_q^{1,a}(., L_q)) > 0$.

(*ii*) For any $q > 0$, for any $g \in \mathcal{B}_q^a$, there exists $h \in L^2(\eta)$ such that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $L \in [L_q - \delta, L_q + \delta]$ and for almost every $x \in \mathbb{R}^d$,

$$\left| H_q^{g,a}(x, L) - H_q^{g,a}(x, L_q) \right| \leq \varepsilon h(x).$$

The main result of this paper is the following central limit type theorem.

Theorem 3.2 *Under Assumption $[\mathcal{H}]$ or $[\mathcal{H}^a]$, for any $f \in \mathcal{B}(\mathbb{R}^d)$ such that $f = f \times \mathbf{1}_{S(\cdot) \geq L^\star}$, we have*

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

15

with the variance functional

$$\Gamma(f) := \sum_{p=0}^{n} \eta_p(\overline{Q}_{p,n}(f)^2 - \eta_n(f)^2). \qquad (3.1)$$

Theorems 3.1 and 3.2 allow us to specify the fluctuations of the estimates $\hat{E}$, $\hat{P}$ and $\hat{C}$.

Corollary 3.1 *Under the same assumptions as in Theorem 3.2, we have:*

(i) *for the estimate of the expectation $E = \mathbb{E}[f(X)] = \mathbb{E}[f(X)\mathbf{1}_{S(X) \geq L^\star}]$,*

$$\sqrt{N}\left(\hat{E} - E\right) \xrightarrow[N \to \infty]{\mathcal{D}} \mathcal{N}(0, \alpha^{2n}\Gamma(f)).$$

(ii) *for the rare event probability $P = \mathbb{P}(Y \geq L^\star)$,*

$$\sqrt{N}\left(\hat{P} - P\right) \xrightarrow[N \to \infty]{\mathcal{D}} \mathcal{N}(0, \alpha^{2n}\Gamma(\mathbf{1}_{S(\cdot) \geq L^\star})).$$

(iii) *for the conditional expectation $C = \mathbb{E}[f(X)|S(X) \geq L^\star]$, still with $f = f \times \mathbf{1}_{S(\cdot) \geq L^\star}$,*

$$\sqrt{N}\left(\hat{C} - C\right) \xrightarrow[N \to \infty]{\mathcal{D}} \mathcal{N}(0, \Gamma(g)),$$

where

$$g := \frac{\mathbf{1}_{S(\cdot) \geq L^\star}}{r}\left(f - \frac{\eta_n(f)}{r}\right).$$

In the next section, we compare these results with the ones obtained for the fixed-levels version of Multilevel Splitting, which was initially proposed in [10]. The analysis of this method in the specific context of the present article was done by some of the authors in [8].

3.2 Comparison with the fixed-levels method

In what follows, we return to the optimal Feynman-Kac particle approximation (fixed-levels method) that was presented in Section 2.2.

Theorem 3.3 *For any $f \in \mathcal{B}(\mathbb{R}^d)$, we have the almost sure convergences $\lim_{N \to \infty} \check{\gamma}_n^N(f) = \gamma_n(f)$, and $\lim_{N \to \infty} \check{\eta}_n^N(f) = \eta_n(f)$, as well as the convergences in distribution*

$$\sqrt{N}\left(\check{\gamma}_n^N(f) - \gamma_n(f)\right) \xrightarrow[N \to \infty]{\mathcal{D}} \mathcal{N}(0, \alpha^{2n}\Gamma(f))$$

$$\sqrt{N}\left(\check{\eta}_n^N(f) - \eta_n(f)\right) \xrightarrow[N \to \infty]{\mathcal{D}} \mathcal{N}(0, \Gamma(f - \eta_n(f)))$$

with the variance functional Γ defined in (3.1).

For the proof of this theorem, we report the interested reader to Propositions 9.4.1 and 9.4.2 in [13]. Just note that straightforward computations give that

$$\alpha^{2n}\Gamma(f) \;=\; \alpha^{2n}\sum_{p=0}^{n}\eta_p(\overline{Q}_{p,n}(f)^2 - \eta_n(f)^2)$$

$$= \sum_{p=0}^{n}\gamma_p(1)^2\eta_p((Q_{p,n}(f) - \eta_p(Q_{p,n}(f)))^2),$$

which is exactly the variance given in Propositions 9.4.1 in [13] for the case of a multinomial resampling. In this paper we prefer using the first expression because it is how it will appear naturally in the proofs. In Theorem 3.2, the asymptotic variance has the same form as the one for unnormalized measures in Theorem 3.3 . Actually, in the adaptive case, we have $\gamma_n^N = \alpha^n\eta_n^N$, so that both give the same asymptotic variance, up to a deterministic multiplicative constant.

Note also that $\Gamma(f - \eta_n(f))$ can be rewritten as

$$\Gamma(f - \eta_n(f)) = \sum_{p=0}^{n}\eta_p(\overline{Q}_{p,n}(f - \eta_n(f))^2),$$

which is exactly formula (9.13) in [13].

In the normalized case, it may seem that Theorems 3.2 and 3.3 give different asymptotic variances, and they do in all generality. But we need to carefully consider what we do at the last step and how we construct our estimates. The specificity of the last step is that the level is always $L^\star$ and thus is **not** random.

For the normalized measures, the asymptotic variances clearly coincide for functions f such that $\eta_n(f) = 0$. As we see in the proof of Corollary 3.1, we can write

$$\hat{C} - C = \frac{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}{\eta_n^N(\mathbf{1}_{S(\cdot)\geq L^\star})} \times \left(\eta_n^N(g) - \eta_n(g)\right),$$

where the prefactor $\frac{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}{\eta_n^N(\mathbf{1}_{S(\cdot)\geq L^\star})}$ converge to 1 in probability, and

$$g = \frac{\mathbf{1}_{S(\cdot)\geq L^\star}}{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}\left(f - \frac{\eta_n(f)}{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}\right)$$

is such that $\eta_n(g) = 0$. As the same trick can be done for $\check{C} - C$ (non adaptive case), we have the same asymptotic variance, because g is centered for η_n.

The next corollary, which is a direct consequence of Corollary 3.1 and the above discussion, constitutes the main message of the present article.

Corollary 3.2 *Under Assumption* $[\mathcal{H}]$ *or* $[\mathcal{H}^a]$, *for any* $f \in \mathcal{B}(\mathbb{R}^d)$ *such that* $f = f \times \mathbf{1}_{S(\cdot) \geq L^\star}$, *the estimates* $\hat{E}$ *and* $\check{E}$ *have the same asymptotic variances. The same result holds for the estimates* $\hat{P}$ *and* $\check{P}$ *of the probability* P, *and for the estimates* $\hat{C}$ *and* $\check{C}$ *of the conditional expectation* C.

Interestingly, as detailed in Proposition 3 of [8], there exists another expression for the asymptotic variance of the estimator $\check{P}$. By Corollary 3.2, this expression holds for the estimator $\hat{P}$ as well. We recall it hereafter for the sake of completeness.

Corollary 3.3 *Under Assumption* $[\mathcal{H}]$ *or* $[\mathcal{H}^a]$, *we have*

$$\sqrt{N} \, \frac{\hat{P} - P}{P} \xrightarrow[N \to +\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2) \quad and \quad \sqrt{N} \, \frac{\check{P} - P}{P} \xrightarrow[N \to +\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2),$$

where $\sigma^2 = \frac{\alpha^{2n}}{P^2} \Gamma(\mathbf{1}_{S(\cdot) \geq L})$ *admits the alternative expression*

$$\sigma^2 = (n-1) \times \frac{1 - \alpha}{\alpha} + \frac{1 - r}{r}$$
$$+ \frac{1}{\alpha} \sum_{p=0}^{n-2} \mathbb{E}\left[\left(\frac{\mathbb{P}(S(X_n) \geq L^\star | X_{p+1})}{r \times \alpha^{n-(p+1)}} - 1 \right)^2 \middle| S(X_p) \geq L_p \right]$$
$$+ \frac{1}{r} \times \mathbb{E}\left[\left(\frac{\mathbb{P}(S(X_n) \geq L^\star | X_n)}{r} - 1 \right)^2 \middle| S(X_{n-1}) \geq L_{n-1} \right]. \qquad (3.2)$$

This expression emphasizes that, when using Multilevel Splitting, the relative variance σ^2 is always lower bounded by an incompressible variance term, namely that

$$\sigma^2 \geq (n-1) \times \frac{1 - \alpha}{\alpha} + \frac{1 - r}{r}.$$

The additive terms in (3.2) depend on the mixing properties of the transition kernels M_p. In particular, if at each step we have an "ideal" kernel, meaning that, knowing that $S(X_p) \geq L_p$, X_{p+1} is independent of X_p, then these additive terms vanish. This is the so-called "idealized" version of Adaptive Multilevel Splitting, studied for example in [8, 25, 6, 36].

Finally, let us mention that our results also apply directly to the case of general multilevel splitting. Specifically, let us consider a fixed and known final level $L^\star$ and a sequence of prescribed success probabilities $(\alpha_p)_{p \geq 0}$ corresponding to the (unknown) sequence of levels $-\infty = L_{-1} < L_0 < \cdots < L_{n-1} < L^\star < L_n$, with

$$\alpha_p = \mathbb{P}(S(X) > L_p | S(X) > L_{p-1}) \implies \mathbb{P}(S(X) > L^\star) = r \prod_{p=0}^{n-1} \alpha_p,$$

18

with

$$r = \mathbb{P}(S(X) > L^\star | S(X) > L_{n-1}) \in (\alpha_n, 1).$$

Then, a quick inspection of the proofs ensures that the Adaptive Multilevel Splitting algorithm with a sequence of adaptive levels $L_0^N < \cdots < L_{n-1}^N$ will have the same asymptotic variance as the Multilevel Splitting algorithm with the levels $L_0 < \cdots < L_{n-1}$. Compared to its fixed-levels counterpart, the cost of the adaptive version is just a higher complexity by a factor $\log N$, due to the quicksort of the sample at each step.

4   Discussion on Assumption $[\mathcal{H}]$

In this section we return to Assumption $[\mathcal{H}]$, and show that it is satisfied on several models of interest. For the sake of simplicity, we focus our attention on $[\mathcal{H}]$ and we will not comment on $[\mathcal{H}^a]$, but the following arguments may be repeated *mutatis mutandis* by replacing K_q with $\mathbf{k}_q^a$.

4.1   An alternative formulation of $[\mathcal{H}]$

First we explain how Assumption $[\mathcal{H}](ii)$ can be verified via a condition on the kernels K_q. Specifically, we get an assumption which is easier to check than $[\mathcal{H}](ii)$, while only a bit more restrictive. The proof is given in Section 6.5.

Proposition 4.1 *Assumption $[\mathcal{H}](ii)$ is fulfilled if there exists a function h in $L^2(\eta)$ and a real number $\delta > 0$, such that for any $q > 0$, for all $L \in (L_q - \delta, L_q + \delta)$ and for almost every $x \in \mathbb{R}^d$,*

$$\int_{S(x')=L} \left| \operatorname{div}_{x'} \left(K_{q+1}(x, x') \frac{DS(x')}{|DS(x')|^2} \right) \right| \frac{đx'}{|DS(x')|}$$
$$+ \int_{S(x')=L} \left[\sum_{m=1}^{q} \sum_{j=1}^{d} \int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'' \right] K_{q+1}(x, x') \frac{đx'}{|DS(x')|^2} \leq h(x).$$
$$(4.1)$$

One may think at first sight that this condition is not much easier to handle than $[\mathcal{H}](ii)$ but, considering our framework, we stress the fact that it is much more natural since it involves only the measure η, the score function S and the transition kernels K_q, at the cost of only a very slight restriction. We can also remark that when $K_m = K$ for all m, then the sum in m in equation (4.1) can be omitted.

4.2 Examples

This section exhibits two typical situations where Assumption $[\mathcal{H}]$ is satisfied. In order to verify $[\mathcal{H}](ii)$, we will make use of condition (4.1) given in Proposition 4.1.

4.2.1 The compact case

If the levels sets $S^{-1}(L)$ are compact, then under mild regularity conditions on S and the kernels K_q, it is not difficult to see that Assumption $[\mathcal{H}]$ is satisfied. The remainder of this section details some sufficient conditions.

Let us assume that, for any real number L, the set $\{S(x) \leq L\}$ is bounded, with Hausdorff measure bounded by C_L, and that for all $x \in S^{-1}(L)$, we have $|DS(x)| \geq c_L > 0$. Besides, assume that for all q, $K_q \leq C$. From these assumptions, it is clear that

$$\int \eta(dx) \left(\int_{S(x')=L_q} K_{q+1}(x,x') \frac{đx'}{|DS(x')|} \right)^2 \leq \left(\frac{C_q \times C}{c_q} \right)^2,$$

and condition $[\mathcal{H}](i)$ is fulfilled.

Now we consider $[\mathcal{H}](ii)$, and use condition (4.1) of Proposition 4.1. Denote respectively by $C_{q,\delta}$ and $c_{q,\delta}$ the upper and lower bounds of C_L and c_L for $L \in (L_q - \delta, L_q + \delta)$. We reformulate the first term in (4.1) as the integral on the level set $\{S(x') = L\}$ of the quantity

$$\mathrm{div}_{x'} \left[\frac{K_{q+1}(x,x')}{|DS(x')|^2} DS(x') \right]$$

$$= \frac{K_{q+1}(x,x')}{|DS(x')|^2} \Delta S(x') + \frac{1}{|DS(x')|^2} D_{x'} K_{q+1}(x,x') DS(x')^T$$

$$-2 \frac{K_{q+1}(x,x')}{|DS(x')|^4} DS(x') H_S(x') DS(x')^T,$$

where ΔS and H_S are respectively the Laplacian and the Hessian of S. From this we see that if K_q has bounded first derivatives (in the second variable), and if S is two times continuously differentiable, then this term is bounded by a constant M and, returning to (4.1), we get for all $L \in (L_q - \delta, L_q + \delta)$,

$$\int_{S(x')=L} \left| \mathrm{div}_{x'} \left(K_{q+1}(x,x') \frac{DS(x')}{|DS(x')|^2} \right) \right| \frac{đx'}{|DS(x')|} \leq M \times \frac{C_{q,\delta}}{c_{q,\delta}}.$$

The second term in (4.1), namely

$$\int_{S(x')=L} \left[\sum_{m=1}^{q} \sum_{j=1}^{d} \int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'' \right] K_{q+1}(x, x') \frac{dx'}{|DS(x')|^2},$$

is slightly more challenging because of the inner integral on the whole space. An obvious sufficient condition is that the kernels K_m have bounded first derivatives in the first variable, say by M, and that their supports have uniformly bounded Lebesgue measures, say by ρ. Then we have, for all $L \in (L_q - \delta, L_q + \delta)$,

$$\int_{S(x')=L} \left[\sum_{m=1}^{q} \sum_{j=1}^{d} \int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'' \right] K_{q+1}(x, x') \frac{dx'}{|DS(x')|^2} \leq q d \rho M C \frac{C_{q,\delta}}{c_{q,\delta}^2}$$

and condition (4.1) of Proposition 4.1 is satisfied in this so-called compact case (compact level sets for S plus compact supports for the transition kernels).

4.2.2 The Gaussian case

Outside this compact framework, there are of course other situations where Assumption [$\mathcal{H}$] is satisfied. Indeed, in many cases, K_q and η have exponential decay at infinity (i.e. light tails). If K_q has the form

$$K_q(x, x') \propto \exp(-V_q(x - x')),$$

with V_q equivalent to a polynomial at infinity, then basically its derivatives with respect to x' take the form $P_q(x, x') K_q(x, x')$, where P_q itself is bounded by another polynomial at infinity. Then, roughly speaking, its integral is a moment of the density $K_q(x, \cdot)$, which typically will be bounded by another polynomial in x. This polynomial in x will in turn be integrable by η if η has light tails. The upcoming example is going to make this more precise in the Gaussian case.

Specifically, we will detail the computations on the zero-bit watermarking example of [8, section 5.1]. In this case, the score function is defined for any $x \in \mathbb{R}^d$ by $S(x) = x_1/|x|$, and η is the standard Gaussian distribution on $\mathbb{R}^d$. Thus it is readily seen that, for any $\sigma > 0$, the transition kernel K defined by

$$K(x, x') = \frac{1 + \sigma^2}{2\pi\sigma^2} \exp\left(-\frac{1 + \sigma^2}{2\sigma^2} \left| x' - \frac{x}{\sqrt{1 + \sigma^2}} \right|^2 \right)$$

is η-reversible. We explain in Section 6.6 that in this situation, Assumption [$\mathcal{H}$] is satisfied.

5 Proofs

5.1 Some preliminary notations

We let $\mathcal{F}_{-1}^N := \{\emptyset, \Omega\}$ be the trivial sigma-field and, for $q \geq 0$, we denote by $\mathcal{F}_q^N$ the sigma-field generated

$$\mathcal{F}_q^N := \mathcal{F}_{q-1}^N \vee \sigma\left((X_q^1, U_q^1), \ldots, (X_q^N, U_q^N)\right).$$

Also, according to the definition of the empirical quantile given by (2.9), we set

$$\mathcal{G}_{-1}^N := \sigma\left(L_0^N\right) = \sigma\left(X_0^{(\lfloor N(1-\alpha)\rfloor)}, U_0^{(\lfloor N(1-\alpha)\rfloor)}\right),$$

and, for $q \geq 0$,

$$\mathcal{G}_q^N := \mathcal{F}_q^N \vee \sigma\left(L_{q+1}^N\right) = \mathcal{F}_q^N \vee \sigma\left(X_{q+1}^{(\lfloor N(1-\alpha)\rfloor)}, U_{q+1}^{(\lfloor N(1-\alpha)\rfloor)}\right).$$

Then, given $\mathcal{F}_{q-1}^N$,

$$\eta_q^N := \frac{1}{N} \sum_{1 \leq i \leq N} \delta_{X_q^i}$$

is the empirical measure associated with N conditionally independent random vectors with common distribution

$$\Phi_q(\eta_{q-1}^N) := \frac{1}{\lceil N\alpha \rceil} \sum_{i: \ S(X_{q-1}^i) > L_{q-1}^N} M_{q, L_{q-1}^N}(X_{q-1}^i, \cdot).$$

Next, given $\mathcal{G}_{q-1}^N$ and adapting for instance Theorem 2.1 in [4] to our context, it can be shown that the subsample of the vectors X_q^i above L_q^N are conditionally independent random vectors denoted by $(\widetilde{X}_q^i)_{1 \leq i \leq \lceil N\alpha \rceil}$ and with common distribution

$$\text{Law}\left(\left(\widetilde{X}_q^1, \ldots, \widetilde{X}_q^{\lceil N\alpha \rceil}\right) \ \Big| \ \mathcal{G}_{q-1}^N\right) = \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)^{\otimes \lceil N\alpha \rceil}, \tag{5.1}$$

where, if $L_q^N = (L, u)$, we have for any $x \in \mathbb{R}^d$

$$G_{\eta_q^N}(x) = \mathbf{1}_{S(x)>L} + (1-u)\mathbf{1}_{S(x)=L},$$

and, accordingly,

$$\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right) = \frac{G_{\eta_q^N} \times \Phi_q(\eta_{q-1}^N)}{\Phi_q(\eta_{q-1}^N)(G_{\eta_q^N})}. \tag{5.2}$$

In summary, we have that

$$\widetilde{\eta}_q^N := \Psi_{G_{\eta_q^N}}\left(\eta_q^N\right) = \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} \delta_{\widetilde{X}_q^i}, \tag{5.3}$$

and

$$\Phi_{q+1}\left(\eta_q^N\right) = \Psi_{G_{\eta_q^N}}\left(\eta_q^N\right) M_{q+1,L_{q-1}^N} = \widetilde{\eta}_q^N M_{q+1,L_q^N}. \tag{5.4}$$

Let us also define

$$\Pi_q(\eta_{q-1}^N) := \frac{\lceil N\alpha \rceil}{N} \times \Phi_q(\eta_{q-1}^N) = \frac{\lceil N\alpha \rceil}{N} \times \Psi_{G_{\eta_{q-1}^N}}\left(\eta_{q-1}^N\right) M_{q,L_{q-1}^N}. \tag{5.5}$$

Alternatively, if ν is absolutely continuous, we define the operator Π_q (see also (2.7)) as

$$\Pi_q(\nu) := \alpha \; \Phi_q(\nu) = \alpha \; \Psi_{G_\nu}(\nu) M_{q,\nu}. \tag{5.6}$$

Besides, for any $q < p$ and $\mu = \nu$ or $\mu = \eta_q^N$, we set

$$\Pi_{q,p}(\mu) = \Pi_{q+1}(\mu) Q_{q+1,p} \quad \text{and} \quad \Phi_{q,p}(\mu) = \frac{\Phi_{q+1}(\mu) Q_{q+1,p}}{(\Phi_{q+1}(\mu) Q_{q+1,p})(1)}, \tag{5.7}$$

with the conventions that $\Pi_{q,p} = I_d = \Phi_{q,p}$ whenever $q \geq p$. This yields

$$\eta_p = \alpha^{q-p} \times \Pi_{q,p}(\eta_q) = \Phi_{q,p}(\eta_q). \tag{5.8}$$

Hence, for any $f \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\Pi_{q,p}(\mu) = \mu Q_{q,p,\mu} \quad \text{and} \quad \Phi_{q,p}(\mu)(f) = \Pi_{q,p}(\mu)(f)/\Pi_{q,p}(\mu)(1),$$

with the collection of integral operators $Q_{q,p,\mu}$ defined by

$$Q_{q,p,\mu} \; := \; Q_{q+1,\mu} Q_{q+2} \dots Q_p = Q_{q+1,\mu} Q_{q+1,p}.$$

In addition, using (5.3), we prove

$$\Pi_{q+1}(\eta_q^N) = \eta_q^N Q_{q+1,\eta_q^N} = \eta_q^N(G_{\eta_q^N}) \; \Phi_{q+1}(\eta_q^N) = \frac{\lceil N\alpha \rceil}{N} \; \widetilde{\eta}_q^N M_{q+1,L_q^N}, \tag{5.9}$$

which implies that

$$\Pi_{q,p}(\eta_q^N) = \frac{\lceil N\alpha \rceil}{N} \; \widetilde{\eta}_q^N \widetilde{Q}_{q,p,\eta_q^N} \tag{5.10}$$

whence, thanks to (5.1),

$$\mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \mid \mathcal{G}_{q-1}^N\right] = \frac{\lceil N\alpha \rceil}{N} \; \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right) \widetilde{Q}_{q,p,\eta_q^N}(f), \tag{5.11}$$

23

with the collection of integral operators

$$\widetilde{Q}_{q,p,\mu} := M_{q+1,\mu}Q_{q+1,p}.$$

Note that by construction, we have

$$\widetilde{Q}_{q,p} = M_{q+1}Q_{q+1,p} =: \widetilde{Q}_{q,p} \quad \text{and} \quad Q_{q,p,\eta_q} = Q_{q,p}.$$

We also observe that, according to (5.5),

$$\mathbb{E}\left[\eta_q^N(f) \mid \mathcal{F}_{q-1}^N\right] = \Phi_q\left(\eta_{q-1}^N\right)(f) = \frac{N}{\lceil N\alpha \rceil}\Pi_q(\eta_{q-1}^N)(f), \qquad (5.12)$$

or, said differently,

$$\alpha^{-1} \, \Pi_q(\eta_{q-1}^N) = \rho_N \, \Phi_q\left(\eta_{q-1}^N\right) \quad \text{with} \quad \rho_N := \frac{\lceil N\alpha \rceil}{N\alpha} \qquad (5.13)$$

and

$$\alpha^{-1} \, \Pi_{q-1,p}\left(\eta_{q-1}^N\right) = \alpha^{-1} \, \Pi_{q,p}\left(\Pi_q(\eta_{q-1}^N)\right) = \rho_N \, \Pi_{q,p}\left(\Phi_q\left(\eta_{q-1}^N\right)\right). \qquad (5.14)$$

We note, once and for all, that

$$0 \le \rho_N - 1 < \frac{1}{N\alpha}.$$

Finally, we consider the $\mathcal{G}_{q-1}^N$ measurable random variable ϵ_q^N defined by

$$\epsilon_q^N = 1 - \rho_N\Phi_q\left(\eta_{q-1}^N\right)(G_{\eta_q^N})/\alpha \iff \Phi_q\left(\eta_{q-1}^N\right)(G_{\eta_q^N}) = \rho_N^{-1}\alpha(1-\epsilon_q^N). \quad (5.15)$$

5.2 Proof of Theorem 3.1

We will prove the almost sure convergences, and explain at the end how to get the convergence in probability. We proceed by induction with respect to the time parameter p, as is done for example in [12, 20].

Denoting $X_0^1, \ldots, X_0^N$ an i.i.d. sample with common law $\eta = \eta_0$, the strong law of large numbers tells us that, by definition of η_0^N and η_0, for any $f \in L^2(\eta)$, we have

$$\eta_0^N(f) = \frac{1}{N}\sum_{i=1}^N f(X_0^i) \xrightarrow[N\to\infty]{a.s.} \eta(f) = \eta_0(f).$$

Then, since the cdf F_Y is one-to-one and $L_0 = F_Y^{-1}(1 - \alpha)$, the theory of order statistics ensures that

$$L_0^N = L_{\eta_0^N} \xrightarrow[N \to \infty]{a.s.} L_{\eta_0} = L_0.$$

Next, let us assume that the property is satisfied for $p \geq 0$ and recall that $\mathcal{F}_p^N$ is the sigma-field generated by the random couples (X_p^i, U_p^i) for $i = 1, \ldots, N$. We begin with the following decomposition

$$\left| \eta_{p+1}^N(f) - \eta_{p+1}(f) \right|$$
$$\leq \left| \eta_{p+1}^N(f) - \mathbb{E}\left[\eta_{p+1}^N(f) \big| \mathcal{F}_p^N \right] \right| + \left| \mathbb{E}\left[\eta_{p+1}^N(f) \big| \mathcal{F}_p^N \right] - \eta_{p+1}(f) \right|. \quad (5.16)$$

Concerning the second term, (5.12) implies

$$\mathbb{E}\left[\eta_{p+1}^N(f) \big| \mathcal{F}_p^N \right] = \Phi_{p+1}(\eta_p^N)(f) = \frac{N}{\lceil N\alpha \rceil} \Pi_{p+1}(\eta_p^N)(f),$$

and by Proposition 6.2 page 33, the induction assumption and (5.8), we get

$$\frac{N}{\lceil N\alpha \rceil} \Pi_{p+1}(\eta_p^N)(f) \xrightarrow[N \to \infty]{a.s.} \frac{1}{\alpha} \Pi_{p+1}(\eta_p)(f) = \eta_{p+1}(f).$$

Hence the second term of (5.16) goes almost surely to 0. For the first term of (5.16), recall that given $\mathcal{F}_p^N$, the random variables $f(X_{p+1}^1), \ldots, f(X_{p+1}^N)$ are i.i.d. with mean $\mathbb{E}\left[\eta_{p+1}^N(f) \big| \mathcal{F}_p^N \right]$. Hence, for any $\varepsilon > 0$, Hoeffding's inequality gives

$$\mathbb{P}\left(\left| \eta_{p+1}^N(f) - \mathbb{E}\left[\eta_{p+1}^N(f) \big| \mathcal{F}_p^N \right] \right| > \varepsilon \big| \mathcal{F}_p^N \right) \leq 2 \exp\left\{ -\frac{N\varepsilon^2}{2\|f\|^2} \right\}, \quad (5.17)$$

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

$$\mathbb{P}\left(\left| \eta_{p+1}^N(f) - \mathbb{E}\left[\eta_{p+1}^N(f) \big| \mathcal{F}_p^N \right] \right| > \varepsilon \right) \leq 2 \exp\left\{ -\frac{N\varepsilon^2}{2\|f\|^2} \right\}.$$

Consequently, the choice $\varepsilon_N = N^{-1/4}$ and Borel-Cantelli Lemma show that the first term of (5.16) goes almost surely to 0 as well.

It remains to show the convergence of L_{p+1}^N to L_{p+1}. To achieve this aim, let us denote F_{p+1} the following cdf

$$F_{p+1}(y) = \mathbb{P}(S(X) \leq y \mid S(X) \geq L_p).$$

In this respect, by definition, we have $F_{p+1}(L_{p+1}) = 1 - \alpha$. This being done, one has just to mimic the reasoning of the proof of point (i) in Proposition 6.2 to obtain the desired result.

To get the convergences in probability for functions $f \in L^2(\eta)$, the same arguments apply to the second term of (5.16). About the first one, one may just replace Hoeffding's inequality with Chebyshev's inequality in (5.17) to obtain

$$\mathbb{P}\left(\left|\eta_{p+1}^N(f) - \mathbb{E}\left[\eta_{p+1}^N(f)\big|\mathcal{F}_p^N\right]\right| > \varepsilon\big|\mathcal{F}_p^N\right) \leq \frac{\sigma_N^2}{\varepsilon^2}, \qquad (5.18)$$

where

$$\sigma_N^2 = \mathbb{E}\left[\left(\eta_{p+1}^N(f) - \mathbb{E}\left[\eta_{p+1}^N(f)\big|\mathcal{F}_p^N\right]\right)^2\Big|\mathcal{F}_p^N\right].$$

Given $\mathcal{F}_p^N$, the random variables $X_{p+1}^1, \ldots, X_{p+1}^N$ are i.i.d. with law $\Phi_{p+1}(\eta_p^N)$, so

$$\sigma_N^2 = \frac{1}{N}\left\{\Phi_{p+1}(\eta_p^N)(f^2) - \Phi_{p+1}(\eta_p^N)(f)^2\right\}.$$

Obviously, by (5.12), the induction assumption and Proposition 6.2,

$$N\sigma_N^2 \leq \Phi_{p+1}(\eta_p^N)(f^2) = \frac{N}{\lceil N\alpha\rceil}\,\Pi_{p+1}(\eta_p^N)(f^2) \xrightarrow[N\to\infty]{\mathbb{P}} \eta_{p+1}(f^2).$$

This proves that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\eta_{p+1}^N(f) - \mathbb{E}\left[\eta_{p+1}^N(f)\big|\mathcal{F}_p^N\right]\right| > \varepsilon\big|\mathcal{F}_p^N\right) \xrightarrow[N\to\infty]{\mathbb{P}} 0,$$

and Lebesgue's dominated convergence ensures that

$$\eta_{p+1}^N(f) - \mathbb{E}\left[\eta_{p+1}^N(f)\big|\mathcal{F}_p^N\right] \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

This concludes the proof of Theorem 3.1. ■

5.3 Proof of Theorem 3.2

We use the symbols $\mathbb{V}(.)$ and $\mathbb{V}(.\mid\mathcal{G}_q^N)$ to denote respectively the variance and the conditional variance operators. We start the analysis with a decomposition which is equivalent to the one given (for example) in [13] page 216. Specifically, for any $p \geq 0$, we have the standard following telescoping sum

$$\eta_p^N - \eta_p = \sum_{q=0}^{p}\alpha^{q-p}\left\{\Pi_{q,p}(\eta_q^N) - \alpha^{-1}\,\Pi_{q-1,p}(\eta_{q-1}^N)\right\},$$

with the conventions $\eta_{-1}^N = \eta_0 = \eta$ and $\Pi_0 = \alpha I_d$. By (5.14), this implies that

$$[\eta_p^N - \eta_p](f) = \mathcal{M}_p^N + \mathcal{R}_p^N, \qquad (5.19)$$

where

$$\mathcal{M}_p^N = \sum_{q=0}^{p} \alpha^{q-p} \left\{ \Pi_{q,p}(\eta_q^N)(f) - \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \right\} \tag{5.20}$$

$$+ \sum_{q=0}^{p} \alpha^{q-p} \left\{ \epsilon_q^N \, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] - \mathbb{E}\left[\epsilon_q^N \, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \big| \mathcal{F}_{q-1}^N \right] \right\}$$

is a martingale that will be discussed below, and

$$\mathcal{R}_p^N \;=\; \sum_{q=0}^{p} \alpha^{q-p} \mathbb{E}\left[\epsilon_q^N \, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \big| \mathcal{F}_{q-1}^N \right] \tag{5.21}$$

$$+ \sum_{q=0}^{p} \alpha^{q-p} \left\{ (1 - \epsilon_q^N) \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] - \alpha^{-1} \, \Pi_{q-1,p}(\eta_{q-1}^N) \right\},$$

is a rest that will be negligible. We recall that $\rho_N = \lceil N\alpha \rceil / N$ and that ϵ_q^N was defined in equation (5.15) by

$$\epsilon_q^N = 1 - \frac{\rho_N}{\alpha} \times \Phi_q\left(\eta_{q-1}^N \right) (G_{\eta_q^N}).$$

The analysis of (5.19) is based on a series of technical results.

Proposition 5.1 *For any $q \leq p$ and any $f \in \mathcal{B}(\mathbb{R}^d)$, we have*

$$N\alpha^{2(q-p)} \, \mathbb{V}\left(\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right) \xrightarrow[N\to\infty]{a.s.} \eta_q(\overline{Q}_{q,p}(f)^2) - \alpha^{-1} \, \eta_p(f)^2.$$

Proposition 5.2 *For any $q \leq p$ and any $f \in \mathcal{B}(\mathbb{R}^d)$, we have*

$$\mathbb{E}\left[\sqrt{N} \; \epsilon_q^N \, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \Big| \mathcal{F}_{q-1}^N \right] \xrightarrow[N\to\infty]{\mathbb{P}} 0,$$

and

$$\mathbb{V}\left[\sqrt{N} \; \epsilon_q^N \, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \Big| \mathcal{F}_{q-1}^N \right] \xrightarrow[N\to\infty]{\mathbb{P}} \frac{1-\alpha}{\alpha} \eta_p(f)^2.$$

Proposition 5.3 *Under Assumption $[\mathcal{H}]$, for any $q \leq p$ and any $f \in \mathcal{B}(\mathbb{R}^d)$ such that $f = f \times \mathbf{1}_{S(\cdot) \geq L^\star}$, we have*

$$\sqrt{N} \, \left(\alpha \, (1 - \epsilon_q^N) \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] - \Pi_{q-1,p}(\eta_{q-1}^N)(f) \right) \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

The proofs of these propositions are detailed in Section 6. Now we return to the proof of Theorem 3.2 by considering the decomposition (5.19). By Propositions 5.2 and 5.3, we have that

$$\sqrt{N}\,\mathcal{R}_p^N \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

By (5.10), (5.3) and (5.13), we may write $\mathcal{M}_p^N = \mathcal{M}_p^{N,1} + \mathcal{M}_p^{N,2}$ with

$$
\begin{aligned}
\mathcal{M}_p^{N,1} &= \sum_{q=0}^{p} \alpha^{q-p} \left(\Pi_{q,p}(\eta_q^N)(f) - \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \right) \\
&= \frac{\lceil N\alpha \rceil}{N} \sum_{q=0}^{p} \alpha^{q-p} \left(\widetilde{\eta}_q^N \widetilde{Q}_{q,p,\eta_q^N}(f) - \mathbb{E}\left[\widetilde{\eta}_q^N \widetilde{Q}_{q,p,\eta_q^N}(f) \big| \mathcal{G}_{q-1}^N \right] \right) \\
&= \frac{\rho_N}{\lceil N\alpha \rceil} \sum_{q=0}^{p} \sum_{i=1}^{\lceil N\alpha \rceil} \alpha^{q-p+1} \left(\widetilde{Q}_{q,p,\eta_q^N}(f)(\widetilde{X}_q^i) - \mathbb{E}\left[\widetilde{Q}_{q,p,\eta_q^N}(f)(\widetilde{X}_q^i) \big| \mathcal{G}_{q-1}^N \right] \right)
\end{aligned}
$$

and

$$
\mathcal{M}_p^{N,2} =
$$
$$
\sum_{q=0}^{p} \alpha^{q-p} \left(\epsilon_q^N\, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] - \mathbb{E}\left[\epsilon_q^N\, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \big| \mathcal{F}_{q-1}^N \right] \right).
$$

Now, remember the role of the auxiliary variables $(U_q^1, \ldots, U_q^N)$ as mentioned on pages 10 and 22, and consider the filtration $\mathcal{J} = (\mathcal{J}_j)_{0 \le j \le (p+1)(\lceil N\alpha \rceil+1)-1}$ constructed as follows: for $q \in \{0, \ldots, p\}$,

$$\mathcal{J}_{q(\lceil N\alpha \rceil+1)}^N = \mathcal{G}_{q-1}^N$$

and for $q \in \{0, \ldots, p\}$ and $i \in \{1, \ldots, \lceil N\alpha \rceil\}$,

$$\mathcal{J}_{q(\lceil N\alpha \rceil+1)+i}^N = \mathcal{G}_{q-1}^N \vee \sigma((\widetilde{X}_q^1, \widetilde{U}_q^1), \ldots, (\widetilde{X}_q^i, \widetilde{U}_q^i)) \vee \Sigma_q^N$$

with

$$\Sigma_q^N = \sigma((X_q^j, U_q^j),\ \text{for the indices } j \text{ such that } S(X_q^j) < L_q^N\ ).$$

In particular, note that

$$\mathcal{J}_{q(\lceil N\alpha \rceil+1)+\lceil N\alpha \rceil}^N = \mathcal{J}_{(q+1)(\lceil N\alpha \rceil+1)-1}^N = \mathcal{F}_q^N.$$

Let us define the sequence of random variables $(Z_j^N)_{0 \le j \le (p+1)(\lceil N\alpha \rceil + 1) - 1}$ where the term of rank $q(\lceil N\alpha \rceil + 1)$ is

$$Z_{q(\lceil N\alpha \rceil + 1)}^N$$

$$= \alpha^{q-p} \left\{ \epsilon_q^N \, \mathbb{E} \left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] - \mathbb{E} \left[\epsilon_q^N \, \mathbb{E} \left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N \right] \big| \mathcal{F}_{q-1}^N \right] \right\}$$

$$= \alpha^{q-p} \left\{ \epsilon_q^N \, \mathbb{E} \left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{J}_{q(\lceil N\alpha \rceil + 1)}^N \right] - \right.$$

$$\left. \mathbb{E} \left[\epsilon_q^N \, \mathbb{E} \left[\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{J}_{q(\lceil N\alpha \rceil + 1)}^N \right] \Big| \mathcal{J}_{q(\lceil N\alpha \rceil + 1) - 1}^N \right] \right\},$$

while the term of rank $q(\lceil N\alpha \rceil + 1) + i$, with $1 \le i \le \lceil N\alpha \rceil$, is

$$Z_{q(\lceil N\alpha \rceil + 1) + i}^N$$

$$= \frac{\rho_N}{\lceil N\alpha \rceil} \, \alpha^{q-p+1} \times \left\{ \widetilde{Q}_{q,p,\eta_q^N}(f)(\tilde{X}_q^i) - \mathbb{E} \left[\widetilde{Q}_{q,p,\eta_q^N}(f)(\tilde{X}_q^i) \big| \mathcal{J}_{q(\lceil N\alpha \rceil + 1) + i - 1}^N \right] \right\}.$$

Using the fact that, given $\mathcal{G}_{q-1}^N$, the $\tilde{X}_q^i$'s, $i \in \{1, \ldots, \lceil N\alpha \rceil\}$ are i.i.d. random vectors (see equation (5.1) page 22), and that with similar arguments they are independent of the subsample strictly below L_q^N, it is clear that $(Z_j^N)_{0 \le j \le (p+1)(\lceil N\alpha \rceil + 1) - 1}$ is a triangular array of martingale increments adapted to the filtration $\mathcal{J}$. It is then straightforward to check that

$$\mathcal{M}_p^N = \mathcal{M}_p^{N,1} + \mathcal{M}_p^{N,2} = \sum_{j=0}^{(p+1)(\lceil N\alpha \rceil + 1) - 1} Z_j^N,$$

which is indeed a $\mathcal{J}$-martingale.

Multiplying this large martingale by $\sqrt{N}$, we can use the CLT theorem for martingales page 171 of [30]. The Lindeberg condition is obviously satisfied since f is assumed bounded, and the limits of the conditional variances are specified by Propositions 5.1 and 5.2. This terminates the proof of Theorem 3.2. ∎

Remark This martingale decomposition may be found far from intuitive, but it highlights the contributions to the global error of both the empirical quantile, and the sample error. Moreover, it allows us to have a conditionally i.i.d. sample, and to use well known statistical properties of empirical quantiles.

5.4 Proof of Corollary 3.1

Concerning the proof of (i), we just notice that

$$\sqrt{N}\left(\hat{E} - E\right) = \sqrt{N}\left(\hat{E} - E\right)\mathbf{1}_{\hat{n}=n} + \sqrt{N}\left(\hat{E} - E\right)\mathbf{1}_{\hat{n}\neq n}.$$

Then, for any $\varepsilon > 0$, we have

$$\mathbb{P}\left(\left|\sqrt{N}\left(\hat{E} - E\right)\mathbf{1}_{\hat{n}\neq n}\right| > \varepsilon\right) \leq \mathbb{P}(\hat{n} \neq n).$$

Now, recall that, by Theorem 3.1, L_{n-1}^N and L_n^N converge almost surely to L_{n-1} and L_n, which ensures that $\hat{n}$ converges almost surely to n. As a consequence,

$$\sqrt{N}\left(\hat{E} - E\right)\mathbf{1}_{\hat{n}\neq n} \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

Next, we have

$$\sqrt{N}\left(\hat{E} - E\right)\mathbf{1}_{\hat{n}=n} = \alpha^n \mathbf{1}_{\hat{n}=n} \times \sqrt{N}\left(\eta_n^N(f) - \eta_n(f)\right).$$

The first term on the right hand side converges in probability to α^n and, according to Theorem 3.2, the second one converges in distribution to a Gaussian variable with variance $\Gamma(f)$. Putting all pieces together, we have shown that

$$\sqrt{N}\left(\hat{E} - E\right) \xrightarrow[N\to\infty]{\mathcal{D}} \mathcal{N}(0, \alpha^{2n}\Gamma(f)).$$

Obviously, (ii) is a direct application of this result with $f = \mathbf{1}_{S(\cdot)\geq L^\star}$. For (iii), we have

$$\sqrt{N}\left(\hat{C} - C\right) = \sqrt{N}\left(\frac{\eta_n^N(f)}{\eta_n^N(\mathbf{1}_{S(\cdot)\geq L^\star})} - \frac{\eta_n(f)}{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}\right)$$

$$= \frac{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}{\eta_n^N(\mathbf{1}_{S(\cdot)\geq L^\star})} \times \sqrt{N}\left(\eta_n^N(g) - \eta_n(g)\right),$$

where

$$g = \frac{\mathbf{1}_{S(\cdot)\geq L^\star}}{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}\left(f - \frac{\eta_n(f)}{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})}\right).$$

Since $f = f \times \mathbf{1}_{S(\cdot)\geq L^\star}$, it is clear that $\eta_n(g) = 0$. Taking into account that $\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star}) = r$, we get

$$\Gamma(g) = \sum_{p=0}^{n} \eta_p(\overline{Q}_{p,n}(g)^2) \quad \text{with} \quad g = \frac{\mathbf{1}_{S(\cdot)\geq L^\star}}{r}\left(f - \frac{\eta_n(f)}{r}\right).$$

Moreover, we know from Theorem 3.1 that

$$\frac{\eta_n^N(\mathbf{1}_{S(\cdot)\geq L^\star})}{\eta_n(\mathbf{1}_{S(\cdot)\geq L^\star})} \xrightarrow[N\to\infty]{\mathbb{P}} 1.$$

This concludes the proof of Corollary 3.1.

∎

6 Technical results

This section gathers some general results which are used for establishing the proofs of Theorem 3.1 and Theorem 3.2.

6.1 Some regularity results

For μ an empirical or absolutely continuous probability distribution (like in Section 5.1), and K a transition kernel, we define the transition kernel M_μ as the truncated version of K with respect to μ, that is

$$M_\mu(x,dy) = G_\mu(x)(K(x,dy)G_\mu(y)+K(1-G_\mu)(x)\delta_x(dy))+(1-G_\mu(x))\delta_x(dy).$$

Our first result is quite general but will be of constant use in the other proofs.

Proposition 6.1 *Assume that $\nu(S^{-1}(\{L_\nu\})) = 0$ and that $|L_\mu - L_\nu| \leq \delta$, then there exist two transition kernels $M^{\delta,-}$ and $M^{\delta,+}$ such that*

(i) $M^{\delta,-} \leq M_\mu \leq M^{\delta,+}$,

(ii) for all $f \in L^1(\nu) \cap L^1(\nu K)$, $\lim_{\delta\to 0} |\nu(M^{\delta,+} - M^{\delta,-})(f)| = 0$.

Moreover, the same result holds if we replace respectively M_μ with $R_\mu = G_\mu M_\mu$, as well as $M^{\delta,-}$ with $R^{\delta,-}$, and $M^{\delta,+}$ with $R^{\delta,+}$.

Before proving this result, let us say something about the way we are going to apply it. Typically, we will consider the case where $\nu = \eta_p$ and $K = K_{p+1}$. Since $\eta_p \leq \alpha^{-p}\eta$ and recalling that K_{p+1} is η invariant, it is clear that if f belongs to $L^1(\eta)$, then f is in $L^1(\eta_p) \cap L^1(\eta_p K_{p+1})$ as well. Moreover, the absolute continuity of η ensures that $\eta_p(S^{-1}(\{L_{\eta_p}\})) = 0$.

Proof We will first prove the result for M_μ, the other case is similar, just a bit simpler. We can decompose $M_\mu = M^0 + M^1 + M^2$ with

$$\begin{cases} M^0(x,dy) = G_\mu(x)G_\mu(y)K(x,dy) \\ M^1(x,dy) = G_\mu(x)K(1 - G_\mu)(x)\delta_x(dy) \\ M^2(x,dy) = (1 - G_\mu(x))\delta_x(dy). \end{cases}$$

By construction, $G_{L_\nu+\delta} \le G_\mu \le G_{L_\nu-\delta}$. So we can take

$$M^{\delta,+} = M^{0,\delta,+} + M^{1,\delta,+} + M^{2,\delta,+}$$

with

$$\begin{cases} M^{0,\delta,+}(x,dy) = G_{L_\mu-\delta}(x)G_{L_\mu-\delta}(y)K(x,dy) \\ M^{1,\delta,+}(x,dy) = G_{L_\mu-\delta}(x)K(1-G_{L_\mu+\delta})(x)\delta_x(dy) \\ M^{2,\delta,+}(x,dy) = (1-G_{L_\mu+\delta}(x))\delta_x(dy) \end{cases}$$

and similarly, we can take $M^{\delta,-} = M^{0,\delta,-} + M^{1,\delta,-} + M^{2,\delta,-}$ with

$$\begin{cases} M^{0,\delta,-}(x,dy) = G_{L_\mu+\delta}(x)G_{L_\mu(y)+\delta}(y)K(x,dy) \\ M^{1,\delta,-}(x,dy) = G_{L_\mu+\delta}(x)K(1-G_{L_\mu-\delta})(x)\delta_x(dy) \\ M^{2,\delta,-}(x,dy) = (1-G_{L_\mu-\delta}(x))\delta_x(dy). \end{cases}$$

Then (i) is obviously satisfied. For (ii), we clearly have for all $x \notin S^{-1}(\{L_\nu\})$,

$$(M^{\delta,+} - M^{\delta,-})(f)(x) \xrightarrow[\delta \to 0]{} 0.$$

Moreover, a straightforward computation reveals that

$$\left|(M^{\delta,+} - M^{\delta,-})(f)\right| \le K(|f|) + 2|f|,$$

which belongs to $L^1(\nu)$ by assumption on f. We conclude using Lebesgue's dominated convergence theorem. For the other case, we can apply the same reasoning, by noticing that $R_\mu = G_\mu M_\mu = M^0 + M^1$, so that one can take

$$R^{\delta,+} = M^{0,\delta,+} + M^{1,\delta,+} \qquad \text{and} \qquad R^{\delta,-} = M^{0,\delta,-} + M^{1,\delta,-}.$$

∎

In the upcoming result, (ν_N) is a sequence of empirical probability measures on $\mathbb{R}^d$. We do not need to make further assumptions on its points for now. Moreover, let ν be a fixed and absolutely continuous probability measure on $\mathbb{R}^d$. Denote respectively by L and L_N the $(1-\alpha)$ quantiles of ν and ν_N with respect to the mapping S as defined in Section 2.3, by $\mathcal{A} = \{x \in \mathbb{R}^d : S(x) \ge L\}$ and $\mathcal{A}_N = \{x \in \mathbb{R}^d : S(x) \ge L_N\}$ the associated level sets, and by $G(x) = \mathbf{1}_{\mathcal{A}}(x)$ and $G_N(x) = \mathbf{1}_{\mathcal{A}_N}(x)$ the related potential functions. We will also assume that the probability measure $\nu \circ S^{-1}$ has a density, and that this density is continuous and strictly positive at L.

Moreover, if K is a transition kernel on $\mathbb{R}^d$, we denote respectively by M and M_N its truncated versions according to L and L_N, meaning that

$$M(x,dx') = \mathbf{1}_{\bar{\mathcal{A}}}(x)\delta_x(dx') + \mathbf{1}_{\mathcal{A}}(x)(K(x,\bar{\mathcal{A}})\delta_x(dx') + K(x,dx')\mathbf{1}_{\mathcal{A}}(x')),$$

and M_N accordingly. The action of the mapping Π on ν and ν_N is then defined as $\Pi(\nu) = \nu GM$ and $\Pi(\nu_N) = \nu_N G_N M_N$. The following result exhibits the continuity of Π.

Proposition 6.2 *With the previous notation , if for any $f \in L^1(\nu) \cap L^1(\nu K)$, one has*

$$\nu_N(f) \xrightarrow[N \to \infty]{} \nu(f) \ \text{a.s. (resp. in probability)}$$

then

(i) $L_N \xrightarrow[N \to \infty]{} L$ *a.s. (resp. in probability).*

(ii) $\Pi(\nu_N)(f) \xrightarrow[N \to \infty]{} \Pi(\nu)(f)$ *a.s. (resp. in probability).*

Proof We prove only the convergence a.s., the convergence in probability will follow using a.s. convergence of subsequences.

To prove (i), let us fix $\varepsilon > 0$ and let us denote by F the cdf of the absolutely continuous probability measure $\nu \circ S^{-1}$. By assumption on F, there exist two strictly positive real numbers δ^- and δ^+ such that

$$F(L - \varepsilon) = 1 - \alpha - \delta^- \qquad \text{and} \qquad F(L + \varepsilon) = 1 - \alpha + \delta^+.$$

Applying the almost sure convergence of $\nu_N(f)$ to $\nu(f)$ respectively with $f = \mathbf{1}_{S(\cdot) \leq L - \varepsilon}$ and $f = \mathbf{1}_{S(\cdot) \leq L + \varepsilon}$, we get that for N large enough,

$$\nu_N(\mathbf{1}_{S(\cdot) \leq L - \varepsilon}) \leq 1 - \alpha - \frac{\delta^-}{2} \qquad \text{and} \qquad \nu_N(\mathbf{1}_{S(\cdot) \leq L + \varepsilon}) \geq 1 - \alpha + \frac{\delta^+}{2}.$$

This ensures that, for N large enough, $|L_N - L| \leq \varepsilon$. Since ε is arbitrary, point (i) is proved.

Now we prove (ii). From (i), for any $\delta > 0$, for N larger than some random N_0, we have that $|L_{\nu_N} - L_\nu| \leq \delta$ and we are in a position to apply Proposition 6.1. Moreover, the triangular inequality gives

$$
\begin{aligned}
|(\Pi(\nu_N) - \Pi(\nu))(f)| &= |(\nu_N G_N M_N - \nu G M)(f)| \\
&\leq |\nu_N(G_N M_N - G M)(f)| + |(\nu_N - \nu)(G M(f))|
\end{aligned}
$$

where the second term can be made arbitrarily small by assumption. For the first term, we have

$$|\nu_N(GM - GM)(f)| \leq \nu_N(R^{\delta,+} - R^{\delta,-})(|f|),$$

which converges to $|\nu(R^{\delta,+} - R^{\delta,-})(f)|$ by assumption. We conclude by choosing δ such that the limit is arbitrarily small. ∎

Our next result will be used in the proof of Proposition 5.3.

Corollary 6.1 *For any $q \in \{0, \ldots, n-1\}$, for all $f \in L^2(\eta)$,*

$$\tilde{\eta}_q^N(f) \xrightarrow[N\to\infty]{\mathbb{P}} \eta_{q+1}(f),$$

and for all $f \in \mathcal{B}(\mathbb{R}^d)$,

$$\tilde{\eta}_q^N(f) \xrightarrow[N\to\infty]{a.s.} \eta_{q+1}(f).$$

Proof We only treat the case where f belongs to $L^2(\eta)$. By (5.3), we have

$$\tilde{\eta}_q^N(f) = \Psi_{G_{\eta_q^N}}(\eta_q^N)(f) = \frac{N}{\lceil N\alpha \rceil} \eta_q^N(G_{\eta_q^N} \times f).$$

Assume that the transition kernel K_{q+1} is the identity, that is $K_{q+1}(x, \cdot) = \delta_x$, then by (5.9) and the definition of Π_{q+1}, we may write

$$\tilde{\eta}_q^N(f) = \frac{N}{\lceil N\alpha \rceil} \Pi_{q+1}(\eta_q^N)(f).$$

From Theorem 3.1, we know that

$$\eta_q^N(f) \xrightarrow[N\to\infty]{\mathbb{P}} \eta_q(f).$$

Thus, since

$$L^2(\eta) \subset L^1(\eta_q) = L^1(\eta_q) \cap L^1(\eta_q K_{q+1}),$$

Proposition 6.2 yields

$$\tilde{\eta}_q^N(f) = \frac{N}{\lceil N\alpha \rceil} \Pi_{q+1}(\eta_q^N)(f) \xrightarrow[N\to\infty]{\mathbb{P}} \frac{1}{\alpha} \Pi_{q+1}(\eta_q)(f) = \frac{1}{\alpha} \eta_q(G_q \times f) = \eta_{q+1}(f).$$

$$\blacksquare$$

The upcoming corollary is at the core of the proofs of Propositions 5.1 and 5.2.

Corollary 6.2 *For any $1 \le q \le p < n$, any $f \in \mathcal{B}(\mathbb{R}^d)$, we have*

$$\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)(f) \xrightarrow[N\to\infty]{a.s.} \eta_{q+1}(f),$$

and for any $\beta > 0$,

$$\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left\{\left(\left[\tilde{Q}_{q,p,\eta_q^N} - \tilde{Q}_{q,p}\right](f)\right)^\beta\right\} \xrightarrow[N\to\infty]{a.s.} 0,$$

Proof By Theorem 3.1, we know that for all $f \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\eta_{q-1}^N(f) \xrightarrow[N\to\infty]{a.s.} \eta_{q-1}(f).$$

Hence, by Proposition 6.2, we deduce that for all $f \in \mathcal{B}(\mathbb{R}^d)$,

$$\Pi_q(\eta_{q-1}^N)(f) \xrightarrow[N\to\infty]{a.s.} \Pi_q(\eta_{q-1})(f).$$

Next, by (5.2), we may write

$$\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)(f) = \frac{\Pi_q(\eta_{q-1}^N)(G_{\eta_q^N} f)}{\Pi_q(\eta_{q-1}^N)(G_{\eta_q^N})}.$$

Still by Theorem 3.1, we know that

$$L_{\eta_q^N} = L_q^N \xrightarrow[N\to\infty]{a.s.} L_q.$$

Thus, for any $\delta > 0$, almost surely for N large enough, one has

$$G_{L_q+\delta} \leq G_{\eta_q^N}, G_q \leq G_{L_q-\delta}$$

and the same reasoning as in the proof of Proposition 6.2 shows that for all $f \in \mathcal{B}(\mathbb{R}^d)$,

$$\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)(f) = \frac{\Pi_q(\eta_{q-1}^N)(G_{\eta_q^N} f)}{\Pi_q(\eta_{q-1}^N)(G_{\eta_q^N}).} \xrightarrow[N\to\infty]{a.s.} \frac{\Pi_q(\eta_{q-1})(G_q f)}{\Pi_q(\eta_{q-1})(G_q)} = \eta_{q+1}(f).$$

For the second point, first notice that

$$\left[\widetilde{Q}_{q,p,\eta_q^N} - \widetilde{Q}_{q,p}\right](f) = [M_{q+1,\eta_q^N} - M_{q+1}](Q_{q+1,p}(f)).$$

Then, by the first point of Proposition 6.1, we deduce that almost surely for N large enough,

$$\left|[M_{q+1,\eta_q^N} - M_{q+1}](Q_{q+1,p}(f))\right| \leq \left|[M_{q+1}^{\delta,+} - M_{q+1}^{\delta,-}](Q_{q+1,p}(f))\right|.$$

Therefore, by the previous point,

$$\limsup_{N\to\infty}\left|\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left\{\left(\left[\widetilde{Q}_{q,p,\eta_q^N} - \widetilde{Q}_{q,p}\right](f)\right)^\beta\right\}\right|$$
$$\leq \eta_{q+1}\left(\left|[M_{q+1}^{\delta,+} - M_{q+1}^{\delta,-}](Q_{q+1,p}(f))\right|^\beta\right).$$

Finally, the desired result is just a consequence of the second point of Proposition 6.1. ∎

Basically, the previous results focused on the continuity of the operator Π. In the remainder of this subsection, we go one step further as we are interested in asymptotic expansions. We recall that

$$\mathcal{B}_q^a = \left\{ g : \mathbb{R}^d \to \mathbb{R}, \ \exists (g_0 \ldots, g_{q-1}) \in \mathcal{B}(\mathbb{R}^d)^q, \ g = \mathbf{k}_1^a(g_0) \cdots \mathbf{k}_q^a(g_{q-1}) \right\},$$

and for $g \in \mathcal{B}_q^a$, $x \in \mathbb{R}^d$ and $L \in \mathbb{R}$, we denote

$$H_q^{g,a}(x, L) = \int_{S(x')=L} \mathbf{k}_{q+1}^a(x, x') g(x') \frac{dx'}{|DS(x')|}.$$

Let us first generalize the notations of Assumption $[\mathcal{H}^a]$ to any probability measure ν. As before, we typically have in mind the case where $\nu = \eta_q$ is the restriction of η above level L_{q-1}, in which case Assumption $[\mathcal{H}_\nu^a]$ will be equivalent to Assumption $[\mathcal{H}^a]$. If we consider the kernel K_q instead of $\mathbf{k}_q^a$, we will have exactly the same results, as it is a special case for which $a(x, x')$ is constant equal to 1.

Assumption $[\mathcal{H}_\nu^a]$

(i) For any $q \geq 0$, the mapping $x \mapsto H_q^{1,a}(x, L_q)$ belongs to $L^2(\nu)$, that is

$$\int \nu(dx) \left(\int_{S(x')=L_q} \mathbf{k}_{q+1}^a(x, x') \frac{dx'}{|DS(x')|} \right)^2 < \infty,$$

and $\nu(H_q^{1,a}(., L_q)) > 0$.

(ii) For any $q > 0$, for any $g \in \mathcal{B}_q^a$, there exists $h \in L^2(\nu)$ such that for any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $L \in [L_q - \delta, L_q + \delta]$ and for almost every $x \in \mathbb{R}^d$,

$$\left| H_q^{g,a}(x, L) - H_q^{g,a}(x, L_q) \right| \leq \varepsilon h(x).$$

The following result will be of constant use in the proof of Proposition 5.3.

Lemma 6.1 *Assume that for any $f \in L^2(\nu)$, one has*

$$\nu_N(f) \xrightarrow[N \to \infty]{\mathbb{P}} \nu(f).$$

Then, for any $g \in \mathcal{B}_q^a$ and any $\varphi \in \mathcal{B}(\mathbb{R}^d)$, under Assumption $[\mathcal{H}_\nu^a]$, one has

$$\nu_N \left(\varphi \int_{L_q}^{L_q^N} H_q^{g,a}(\cdot, L) dL \right) = (L_q^N - L_q) \nu(\varphi H_q^{g,a}(\cdot, L_q)) + o_p(L_q^N - L_q).$$

36

Proof We first choose $\varepsilon > 0$. By point (ii) of Assumption $[\mathcal{H}_\nu^a]$, the mapping $L \mapsto H_q^g(x, L)$ is continuous in the neighborhood of L_q for ν almost every x. We consider N large enough such that $L_q^N \in (L_q - \delta, L_q + \delta)$ with arbitrarily large probability, say $1 - \gamma$. Hence, by the mean value theorem, there exists $\tilde{L}$ between L_q and L_q^N such that

$$\int_{L_q}^{L_q^N} H_q^{g,a}(x, L)dL = (L_q^N - L_q) \times H_q^{g,a}(x, \tilde{L}).$$

As a consequence,

$$\frac{\nu_N\left(\varphi \int_{L_q}^{L_q^N} H_q^{g,a}(\cdot, L)dL\right)}{L_q^N - L_q} = \nu_N(\varphi H_q^{g,a}(\cdot, L_q)) + \nu_N(\varphi(H_q^{g,a}(\cdot, \tilde{L}) - H_q^{g,a}(\cdot, L_q))).$$

Since φ and g are both bounded, point (i) of Assumption $[\mathcal{H}_\nu^a]$ ensures that the function $\varphi H_q^{g,a}(\cdot, L_q)$ is in $L^2(\nu)$, so that by the hypothesis of Lemma 6.1,

$$\nu_N(\varphi H_q^{g,a}(\cdot, L_q)) \xrightarrow[N\to\infty]{\mathbb{P}} \nu(\varphi H_q^{g,a}(\cdot, L_q)).$$

Furthermore, by point (ii) of Assumption $[\mathcal{H}_\nu^a]$, we have

$$\left|\nu_N(\varphi(H_q^{g,a}(\cdot, \tilde{L}) - H_q^{g,a}(\cdot, L_q)))\right| \leq (\|\varphi\| \times \nu_N(h)) \times \varepsilon,$$

where, since h belongs to $L^2(\nu)$,

$$\nu_N(h) \xrightarrow[N\to\infty]{\mathbb{P}} \nu(h) < \infty.$$

Since ε and γ are arbitrary, the proof is complete. ∎

6.2   Proof of Proposition 5.1

Our goal is to prove that, for any $q \leq p$ and any $f \in \mathcal{B}(\mathbb{R}^d)$, we have

$$N\alpha^{2(q-p)} \, \mathbb{V}\left(\Pi_{q,p}(\eta_q^N)(f) \big| \mathcal{G}_{q-1}^N\right) \xrightarrow[N\to\infty]{a.s.} \eta_q(\overline{Q}_{q,p}(f)^2) - \alpha^{-1} \, \eta_p(f)^2.$$

By (5.10), we have

$$\Pi_{q,p}(\eta_q^N) = \frac{[N\alpha]}{N} \, \widetilde{\eta}_q^N \widetilde{Q}_{q,p,\eta_q^N}$$

with the measure $\widetilde{\eta}_q^N$ defined in (5.3). This shows that

$$N\alpha^{2(q-p)}\mathbb{V}\left(\Pi_{q,p}(\eta_q^N)(f)\big|\mathcal{G}_{q-1}^N\right) = \left(\frac{\lceil N\alpha\rceil}{N}\right)^2 N\alpha^{2(q-p)}\mathbb{V}\left(\widetilde{\eta}_q^N\widetilde{Q}_{q,p,\eta_q^N}(f)\big|\mathcal{G}_{q-1}^N\right)$$

$$= \alpha^{2(q-p)}\frac{\lceil N\alpha\rceil}{N}\,\mathbb{V}\left(\widetilde{Q}_{q,p,\eta_q^N}(f)(\widetilde{X}_q^1)\big|\mathcal{G}_{q-1}^N\right).$$
(6.1)

On the other hand, we have, thanks to (5.1),

$$\mathbb{V}\left(\left[\widetilde{Q}_{q,p,\eta_q^N}-\widetilde{Q}_{q,p}\right](f)(\widetilde{X}_q^1)\big|\mathcal{G}_{q-1}^N\right)$$
$$= \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left\{\left(\left[\widetilde{Q}_{q,p,\eta_q^N}-\widetilde{Q}_{q,p}\right](f)\right)^2\right\}$$
$$- \left(\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\left[\widetilde{Q}_{q,p,\eta_q^N}-\widetilde{Q}_{q,p}\right](f)\right)\right)^2.$$

By the second point of Corollary 6.2, we deduce that

$$\mathbb{V}\left(\left[\widetilde{Q}_{q,p,\eta_q^N}-\widetilde{Q}_{q,p}\right](f)(\widetilde{X}_q^1)\big|\mathcal{G}_{q-1}^N\right)\xrightarrow[N\to\infty]{a.s.}0.$$

In other words, coming back to (6.1) and applying the first point of Corollary 6.2, we have obtained

$$N\alpha^{2(q-p)}\,\mathbb{V}\left(\Pi_{q,p}(\eta_q^N)(f)\big|\mathcal{G}_{q-1}^N\right)$$
$$\xrightarrow[N\to\infty]{a.s.}\alpha^{2(q-p)+1}\left\{\eta_{q+1}([M_{q+1}Q_{q+1,p}(f)]^2)-(\eta_{q+1}M_{q+1}Q_{q+1,p}(f))^2\right\}.$$

Using elementary computations, it is easy to check that

$$\alpha^{2(q-p)+1}\left\{\eta_{q+1}([M_{q+1}Q_{q+1,p}(f)]^2)-(\eta_{q+1}Q_{q+1,p}(f))^2\right\}$$
$$= \eta_q(\overline{Q}_{q,p}(f)^2)-\alpha^{-1}\,\eta_p(f)^2,$$

which terminates the proof of Proposition 5.1. ∎

6.3 Proof of Proposition 5.2

We intend to show that, for any $q\le p$ and any $f\in\mathcal{B}(\mathbb{R}^d)$, we have

$$\mathbb{E}\left[\sqrt{N}\alpha^{q-p}\,\epsilon_q^N\,\mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f)\big|\mathcal{G}_{q-1}^N\right]\Big|\mathcal{F}_{q-1}^N\right]\xrightarrow[N\to\infty]{\mathbb{P}}0,$$

and

$$\mathbb{V}\left[\sqrt{N}\alpha^{q-p}\,\epsilon_q^N\,\mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f)\big|\mathcal{G}_{q-1}^N\right]\Big|\mathcal{F}_{q-1}^N\right]\xrightarrow[N\to\infty]{\mathbb{P}}\frac{1-\alpha}{\alpha}\,\eta_p(f)^2.$$

38

The proof is carried out given $\mathcal{F}_{q-1}^N$. We begin like in the proof of Proposition 5.1. From (5.11) and the definition of ρ_N, recall that

$$\mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f)|\mathcal{G}_{q-1}^N\right] = \alpha\rho_N\ \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p,\eta_q^N}(f)\right).$$

Hence, the quantity of interest in Proposition 5.2 may be rewritten as follows

$$\sqrt{N}\alpha^{q-p}\ \epsilon_q^N\ \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f)\big|\mathcal{G}_{q-1}^N\right]$$
$$= \left[\sqrt{N}\ \alpha\ \epsilon_q^N\right]\ \rho_N\ \alpha^{q-p}\ \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p,\eta_q^N}(f)\right).$$

We are going to prove first that

$$\rho_N\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p,\eta_q^N}(f)\right) \xrightarrow[N\to\infty]{a.s.} \alpha^{p-q-1}\ \eta_p(f).$$

Because $|\rho_N - 1| \le 1/(N\alpha)$, the factor ρ_N is unimportant. As in the proof of Proposition 5.1, we consider the decomposition

$$\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p,\eta_q^N}(f)\right)$$

$$= \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p,\eta_q^N}(f) - \widetilde{Q}_{q,p}(f)\right) + \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p}(f)\right).$$

The second point of Corollary 6.2 implies that

$$\left|\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\left[\widetilde{Q}_{q,p,\eta_q^N} - \widetilde{Q}_{q,p}\right](f)\right)\right| \xrightarrow[N\to\infty]{a.s.} 0,$$

while the first point of Corollary 6.2 ensures that

$$\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p}(f)\right) \xrightarrow[N\to\infty]{a.s.} \eta_{q+1}\left(\widetilde{Q}_{q,p}(f)\right) = \alpha^{p-q-1}\eta_p(f).$$

As f is bounded, by Lebesgue's dominated convergence theorem, the above convergence also holds in L^2.

Now we prove the first assertion of Proposition 5.2. Let us denote for a moment

$$Z_N = \rho_N\Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N)\right)\left(\widetilde{Q}_{q,p,\eta_q^N}(f)\right)$$

and by $z = \alpha^{p-q-1}\ \eta_p(f)$ its deterministic limit. We have just shown that $Z_N - z$ converges to 0 in L^2. This implies that

$$\mathbb{E}\left[(Z_N - z)^2 \mid \mathcal{F}_{q-1}^N\right] \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

By very similar arguments, we can also see that

$$\mathbb{E}\left[(Z_N^2 - z^2)^2 \mid \mathcal{F}_{q-1}^N\right] \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

We have

$$\sqrt{N}\mathbb{E}\left[\epsilon_q^N Z_N \mid \mathcal{F}_{q-1}^N\right] = \sqrt{N}\mathbb{E}\left[\epsilon_q^N(Z_N - z) \mid \mathcal{F}_{q-1}^N\right] + \sqrt{N}z\mathbb{E}\left[\epsilon_q^N \mid \mathcal{F}_{q-1}^N\right].$$

Since $|\rho_N - 1| \leq 1/(N\alpha)$, we have

$$\epsilon_q^N = \frac{1}{\alpha}\left(\alpha - \Phi_q(\eta_{q-1}^N)(G_{\eta_q^N})\right) + o(1/\sqrt{N}),$$

and the convergence of the second term is a direct consequence of the first result of Lemma 6.2 below. For the first term, we apply Cauchy-Schwarz inequality

$$\left|\sqrt{N}\mathbb{E}\left[\epsilon_q^N(Z_N - z) \mid \mathcal{F}_{q-1}^N\right]\right| \leq \sqrt{N\mathbb{E}\left[(\epsilon_q^N)^2 \mid \mathcal{F}_{q-1}^N\right]}\sqrt{\mathbb{E}\left[(Z_N - z)^2 \mid \mathcal{F}_{q-1}^N\right]},$$

which converges in probability to 0 by the second result of Lemma 6.2, and the L^2 convergence of $Z_N - z$.

For the second assertion of Proposition 5.2, we write

$$N\mathbb{E}\left[(\epsilon_q^N Z_N)^2 \mid \mathcal{F}_{q-1}^N\right] = N\mathbb{E}\left[(\epsilon_q^N)^2(Z_N^2 - z^2) \mid \mathcal{F}_{q-1}^N\right] + Nz^2\mathbb{E}\left[(\epsilon_q^N)^2 \mid \mathcal{F}_{q-1}^N\right].$$

The convergence of the second term is a direct consequence of the second result of Lemma 6.2. For the first term we use Cauchy–Schwarz again

$$N\mathbb{E}\left[(\epsilon_q^N)^2(Z_N^2 - z^2) \mid \mathcal{F}_{q-1}^N\right] \leq \sqrt{N^2\mathbb{E}\left[(\epsilon_q^N)^4 \mid \mathcal{F}_{q-1}^N\right]}\sqrt{\mathbb{E}\left[(Z_N^2 - z^2)^2 \mid \mathcal{F}_{q-1}^N\right]}$$

and we conclude similarly, using the third result of Lemma 6.2. ∎

Lemma 6.2 *For any integer q, we have*

$$\sqrt{N}\ \mathbb{E}\left[\Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}) - \alpha \ \Big| \ \mathcal{F}_{q-1}^N\right] \xrightarrow[N\to\infty]{\mathbb{P}} 0,$$

$$N\ \mathbb{E}\left[\left(\Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}) - \alpha\right)^2 \ \Big| \ \mathcal{F}_{q-1}^N\right] \xrightarrow[N\to\infty]{\mathbb{P}} \alpha(1-\alpha),$$

and

$$N^2\ \mathbb{E}\left[\left(\Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}) - \alpha\right)^4 \ \Big| \ \mathcal{F}_{q-1}^N\right] = \mathcal{O}_p(1).$$

Proof Here again, the reasoning is made given $\mathcal{F}_{q-1}^N$. Recall that $(X_q^i)_{1 \leq i \leq N}$ is an i.i.d. sample with common law $\Phi_q\left(\eta_{q-1}^N\right)$. Accordingly, let us denote $(Y_q^i)_{1 \leq i \leq N} = (S(X_q^i))_{1 \leq i \leq N}$. Also, for any real number L, define the function

$$F_N(L) = 1 - \Phi_q(\eta_{q-1}^N)(G_L),$$

which is more or less a cumulative distribution function. The function F_N is continuous except at a finite number of values, namely at most the $\lceil \alpha N \rceil$ largest values among the Y_{q-1}^i's.

Starting from the sample $(Y_q^i)_{1 \leq i \leq N}$, we also construct a new sample $\mathbf{U} = (U_q^i)_{1 \leq i \leq N}$ as follows. If Y_q^i is a point of continuity of F_N, then $U_q^i = F_N(Y_q^i)$, otherwise we draw U_k^i uniformly in the interval

$$(F_N(Y_q^i), \lim_{h \to 0^+} F_N(Y_q^i + h)).$$

It is then a simple exercise (see for example [35], page 102) to check that $\mathbf{U} = (U_q^1, \ldots, U_q^N)$ is an i.i.d. sample with distribution $\mathcal{U}(0, 1)$. Denoting

$$U_{1-\alpha}^N := U_{\left(\frac{k_N}{N}\right)}^N = U_{\left(\frac{\lceil (1-\alpha)N \rceil}{N}\right)}^N,$$

we may write

$$\Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}) - \alpha = (U_{1-\alpha}^N - (1 - \Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}))) + ((1-\alpha) - U_{1-\alpha}^N)$$
$$= (U_{1-\alpha}^N - F_N(L_q^N)) + ((1-\alpha) - U_{1-\alpha}^N). \tag{6.2}$$

The first term can easily be bounded in absolute value thanks to the following lemma, whose proof is detailed in Section 6.7.

Lemma 6.3 *For any integer q and any $\ell \in \{1, 2, 4\}$, we have*

$$N^{\ell/2} \, \mathbb{E}\left[\left(U_{1-\alpha}^N - F_N(L_q^N)\right)^\ell \, \Big| \, \mathcal{F}_{q-1}^N \right] \xrightarrow[N \to \infty]{\mathbb{P}} 0.$$

For the second term in (6.2), we have

$$(1-\alpha) - U_{1-\alpha}^N = \left((1-\alpha) - \frac{k_N}{N} \right) + \left(\frac{k_N}{N} - U_{\left(\frac{k_N}{N}\right)}^N \right). \tag{6.3}$$

The first term is deterministic and goes to 0. For the second term, it is well known that (see e.g. [35] page 97)

$$U_{\left(\frac{k_N}{N}\right)}^N \sim \text{Beta}(k_N, N - k_N + 1).$$

Therefore

$$\mathbb{E}\left[U^N_{\left(\frac{k_N}{N}\right)}\middle|\ \mathcal{F}^N_{q-1}\right] = \frac{k_N}{N} \qquad \text{and} \qquad \mathbb{V}\left(U^N_{\left(\frac{k_N}{N}\right)}\middle|\ \mathcal{F}^N_{q-1}\right) = \frac{k_N(N - k_N + 1)}{(N+1)^2(N+2)},$$

so that

$$N \times \mathbb{V}\left(U^N_{\left(\frac{k_N}{N}\right)}\middle|\ \mathcal{F}^N_{q-1}\right) = N \times \frac{k_N(N - k_N + 1)}{(N+1)^2(N+2)} \xrightarrow[N\to\infty]{} \alpha(1 - \alpha).$$

We also have

$$N^2 \times \mathbb{E}\left[\left(U^N_{\left(\frac{k_N}{N}\right)} - \frac{k_N}{N}\right)^4\middle|\ \mathcal{F}^N_{q-1}\right]$$

$$= N^2 \times \frac{3k_N(N - k_N + 1)(2(N+1)^2 + k_N(N - k_N + 1)(N - 5))}{(N+1)^4(N+2)(N+3)(N+4)},$$

which is obviously bounded.

Let us prove the first assertion of Lemma 6.2. From (6.2) and (6.3) we deduce

$$\sqrt{N}\ \mathbb{E}\left[\Phi_q(\eta^N_{q-1})(G_{\eta^N_q}) - \alpha\ \middle|\ \mathcal{F}^N_{q-1}\right]$$

$$= \sqrt{N}\ \mathbb{E}\left[U^N_{1-\alpha} - F_N(L^N_q)\middle|\ \mathcal{F}^N_{q-1}\right] + \sqrt{N}\left((1 - \alpha) - \frac{k_N}{N}\right).$$

The first term goes to 0 in probability thanks to Lemma 6.3, the second one is deterministic and goes to 0 since $|(1 - \alpha) - k_N/N| \le 1/N$.

For the second assertion of Lemma 6.2, relation (6.2) gives

$$N\ \mathbb{E}\left[\left(\Phi_q(\eta^N_{q-1})(G_{\eta^N_q}) - \alpha\right)^2\middle|\ \mathcal{F}^N_{q-1}\right]$$

$$= N\ \mathbb{E}\left[\left(U^N_{1-\alpha} - F_N(L^N_q)\right)^2\middle|\ \mathcal{F}^N_{q-1}\right] + N\ \mathbb{E}\left[\left((1 - \alpha) - U^N_{1-\alpha}\right)^2\middle|\ \mathcal{F}^N_{q-1}\right]$$

$$+ N\ \mathbb{E}\left[\left(U^N_{1-\alpha} - F_N(L^N_q)\right)\left((1 - \alpha) - U^N_{1-\alpha}\right)\middle|\ \mathcal{F}^N_{q-1}\right]. \tag{6.4}$$

Here again, the first term goes to 0 in probability thanks to Lemma 6.3. For the second one, just notice that

$$\mathbb{E}\left[\left((1 - \alpha) - U^N_{1-\alpha}\right)^2\middle|\ \mathcal{F}^N_{q-1}\right] = \mathbb{V}\left(U^N_{\left(\frac{k_N}{N}\right)}\middle|\ \mathcal{F}^N_{q-1}\right) + \left((1 - \alpha) - \frac{k_N}{N}\right)^2,$$

which implies that

$$N\ \mathbb{E}\left[\left((1 - \alpha) - U^N_{1-\alpha}\right)^2\middle|\ \mathcal{F}^N_{q-1}\right] \xrightarrow[N\to\infty]{} \alpha(1 - \alpha).$$

Finally, Cauchy-Schwarz shows that the last term in (6.4) goes to 0 in probability, and the second assertion of Lemma 6.2 is established.

Concerning the third assertion of Lemma 6.2, it suffices to remark that

$$N^2 \, \mathbb{E}\left[\left(\Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}) - \alpha \right)^4 \, \middle| \, \mathcal{F}_{q-1}^N \right]$$

$$\leq 4N^2 \, \mathbb{E}\left[\left(U_{1-\alpha}^N - F_N(L_q^N) \right)^4 \, \middle| \, \mathcal{F}_{q-1}^N \right] + 4N^2 \, \mathbb{E}\left[\left((1-\alpha) - U_{1-\alpha}^N \right)^4 \, \middle| \, \mathcal{F}_{q-1}^N \right].$$

Then Lemma 6.3 and the fourth moment of the Beta distribution of interest ensure that this quantity is bounded in probability. This completes the proof of Lemma 6.2. ∎

6.4 Proof of Proposition 5.3

We have to show that, under Assumption $[\mathcal{H}]$, for any $q \leq p$ and any $f \in \mathcal{B}(\mathbb{R}^d)$ such that $f = f \times \mathbf{1}_{S(\cdot) \geq L^\star}$,

$$\sqrt{N} \, \left(\alpha \, (1 - \epsilon_q^N) \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right] - \Pi_{q-1,p}(\eta_{q-1}^N)(f) \right) \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

It turns out that the proof is quite technical and requires several auxiliary results whose proofs are postponed to the end of the present section. Here again, the reasoning is carried out given $\mathcal{F}_{q-1}^N$. By (5.7), (5.9) and the definition of ρ_N, we have

$$\begin{aligned}
\Pi_{q-1,p}(\eta_{q-1}^N)(f) &= \eta_{q-1}^N(G_{\eta_{q-1}^N})\Phi_q(\eta_{q-1}^N)Q_{q,p}(f) \\
&= \alpha\rho_N \, \Phi_q(\eta_{q-1}^N)Q_{q,p}(f)
\end{aligned}$$

and by (5.15) and (5.11),

$$\alpha \, (1 - \epsilon_q^N)\mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right]$$

$$= \rho_N \, \Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}) \, \mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \mid \mathcal{G}_{q-1}^N \right]$$

$$= \alpha\rho_N^2 \, \Phi_q(\eta_{q-1}^N)(G_{\eta_q^N}) \, \Psi_{G_{\eta_q^N}}\left(\Phi_q(\eta_{q-1}^N) \right) \widetilde{Q}_{q,p,\eta_q^N}(f)$$

$$= \alpha\rho_N^2 \, \Phi_q(\eta_{q-1}^N) \left(Q_{q,p,\eta_q^N}(f) \right).$$

Since f is bounded and $\rho_N - 1 = \mathcal{O}(N^{-1})$, this implies that

$$\alpha \, (1 - \epsilon_q^N)\mathbb{E}\left[\Pi_{q,p}(\eta_q^N)(f) \middle| \mathcal{G}_{q-1}^N \right] - \Pi_{q-1,p}(\eta_{q-1}^N)(f)$$

$$= \alpha\rho_N \, \Phi_q\left(\eta_{q-1}^N\right) \left(\left[Q_{q,p,\eta_q^N} - Q_{q,p} \right](f) \right) + \mathcal{O}(N^{-1}).$$

43

Thus, introducing the probability measure $\nu_q^N = \Phi_q\left(\eta_{q-1}^N\right)$ and the bounded function $\varphi = Q_{q+1,p}(f)$, our objective is to show that

$$\sqrt{N}\; \nu_q^N\left(\left[Q_{q+1,\eta_q^N} - Q_{q+1}\right](\varphi)\right) \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

Before going further, let us recall that if $G = \mathbf{1}_{S(\cdot)\geq L}$ is a potential function, K a transition kernel and M its truncated version defined by

$$M(x,dy) = K(x,dy)\; G(y)\; +\; K(1-G)(x)\; \delta_x(dy),$$

then for any finite measure μ and any bounded and measurable function φ, we have the following general formula

$$\mu(GM(\varphi)) = \iint \mu(dy)G(y)K(y,x)G(x)\varphi(x)dx$$
$$+ \iint \mu(dx)G(x)K(x,y)(1-G(y))\varphi(x)dy.$$
$$= \mu(G \times K[G\varphi]) + \mu(K[1-G] \times (G\varphi)). \qquad (6.5)$$

Thus, we get

$$\nu_q^N\left(\left[Q_{q+1,\eta_q^N} - Q_{q+1}\right](\varphi)\right)$$
$$= \nu_q^N\left(G_{\eta_q^N} K_{q+1}[G_{\eta_q^N}\varphi]\right) - \nu_q^N\left(G_q K_{q+1}[G_q\varphi]\right)$$
$$+ \nu_q^N\left(K_{q+1}[1-G_{\eta_q^N}]\,(G_{\eta_q^N}\varphi)\right) - \nu_q^N\left(K_{q+1}[1-G_q](G_q\varphi)\right). \qquad (6.6)$$

We may simplify a bit the latter by noticing that

$$\varphi = Q_{q+1,p}(f) = G_{q+1} \times \widetilde{Q}_{q+1,p}(f) = G_{\eta_{q+1}} \times \widetilde{Q}_{q+1,p}(f).$$

Indeed, we know from Theorem 3.1 that

$$L_{\eta_q^N} \xrightarrow[N\to\infty]{a.s.} L_{\eta_q} = L_q < L_{q+1} = L_{\eta_{q+1}}.$$

Therefore, almost surely for $N > N_0$, we have $G_{\eta_q^N}\varphi = G_q\varphi = \varphi$, and (6.6) reduces to

$$\nu_q^N([Q_{q+1,\eta_q^N} - Q_{q+1}](\varphi)) = \nu_q^N((G_{\eta_q^N} - G_q)K_{q+1}[\varphi]) - \nu_q^N(K_{q+1}[G_{\eta_q^N} - G_q]\varphi).$$
$$(6.7)$$

In the remainder of the proof, we will only treat the more difficult case where the kernels K_p are obtained by the Metropolis-Hastings procedure (see

Section 2.4) and we will suppose that Assumption $[\mathcal{H}^a]$ is satisfied. According to equation (2.11), we have

$$K_p[\varphi](x) = \mathbf{k}_p^a[\varphi](x) + \mathbf{r}_p^a(x) \times \varphi(x).$$

In this expression, recall that $\mathbf{k}_p^a$ has density

$$\mathbf{k}_p^a(x, x') = a_p(x, x')\mathbf{k}_p(x, x').$$

All the upcoming arguments remain valid in the easier case where K_p itself has a density since it suffices to take $a = 1$, so that $\mathbf{r}_p^a = 0$ and $K_p = \mathbf{k}_p$.

In the Metropolis-Hastings situation, combining (6.7) and (2.11), we are led to

$$\nu_q^N([Q_{q+1,\eta_q^N} - Q_{q+1}](\varphi)) = \nu_q^N((G_{\eta_q^N} - G_q)\mathbf{k}_{q+1}^a[\varphi]) - \nu_q^N(\mathbf{k}_{q+1}^a[G_{\eta_q^N} - G_q]\varphi)$$
$$= A_q^N - B_q^N. \tag{6.8}$$

Thanks to the coarea formula, B_q^N rewrites

$$B_q^N = \int \nu_q^N(dx')\varphi(x') \int_{L_q}^{L_q^N} \left(\int_{S(x)=\ell} \mathbf{k}_{q+1}^a(x', x)\frac{dx}{|DS(x)|} \right) d\ell$$
$$= \nu_q^N \left(\varphi \int_{L_q}^{L_q^N} H_q^{1,a}(\cdot, \ell)d\ell \right).$$

Next, since

$$\nu_q^N = \Phi_q\left(\eta_{q-1}^N\right) = \frac{N}{\lceil N\alpha \rceil}\Pi_q(\eta_{q-1}^N),$$

we deduce from Assumption $[\mathcal{H}^a]$, Theorem 3.1, Proposition 6.2 and Lemma 6.1 that

$$B_q^N = (L_q^N - L_q) \times \eta_q \left(\varphi H_q^{1,a}(\cdot, L_q)\right) + o_p(L_q^N - L_q),$$
$$= (L_q^N - L_q) \iint_{S(x)=L_q} \eta_q(dx')\varphi(x')\mathbf{k}_{q+1}^a(x', x)\frac{dx}{|DS(x)|} + o_p(L_q^N - L_q).$$

Concerning A_q^N, coming back to (6.8) and decomposing ν_q^N in absolutely continuous and discrete parts, we may write

$$A_q^N = \nu_q^{N,(0)}((G_{\eta_q^N} - G_q)\mathbf{k}_{q+1}^a[\varphi]) + \nu_q^{N,(1)}((G_{\eta_q^N} - G_q)\mathbf{k}_{q+1}^a[\varphi])$$
$$= A_q^{N,(0)} + A_q^{N,(1)},$$

where

$$\nu_q^{N,(0)}(dx) = \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} \mathbf{k}_q^a(\widetilde{X}_{q-1}^i, dx) G_{\eta_{q-1}^N}(x), \qquad (6.9)$$

and

$$\nu_q^{N,(1)}(dx) = \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} (\mathbf{k}_q^a[1 - G_{\eta_{q-1}^N}](\widetilde{X}_{q-1}^i) + \mathbf{r}_q^a(\widetilde{X}_{q-1}^i))\delta_{\widetilde{X}_{q-1}^i}(dx). \quad (6.10)$$

As previously, since almost surely for $N > N_0$,

$$G_{\eta_{q-1}^N}(x)(G_{\eta_q^N}(x) - G_q(x)) = G_{\eta_q^N}(x) - G_q(x),$$

we get

$$A_q^{N,(0)} = \int \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} \mathbf{k}_q^a(\widetilde{X}_{q-1}^i, dx)(\mathbf{k}_{q+1}^a[\varphi] G_{\eta_{q-1}^N}(G_{\eta_q^N} - G_q))(x)$$

$$= \int \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} \mathbf{k}_q^a(\widetilde{X}_{q-1}^i, dx)(\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q))(x)$$

$$= \tilde{\eta}_{q-1}^N \left(\int_{L_q}^{L_q^N} H_q^{\mathbf{k}_{q+1}^a[\varphi],a}(\cdot, \ell) d\ell \right), \qquad (6.11)$$

the last equation consisting in the application of the coarea formula. Then, Assumption $[\mathcal{H}]$, Corollary 6.1 and Lemma 6.1 yield

$$A_q^{N,(0)} = (L_q^N - L_q) \iint_{S(x)=L_q} \eta_q(dx') \mathbf{k}_q^a(x', x) \mathbf{k}_{q+1}^a[\varphi](x) \frac{dx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q).$$

Using equation (2.10), it is clear that for any pair (x, x'),

$$\eta_q(dx') \, \mathbf{k}_q^a(x', x) \mathbf{1}_{S(x) \geq L_{q-1}} = \alpha^{-q} \, \eta(x') \mathbf{1}_{S(x') \geq L_{q-1}} dx' \mathbf{k}_q^a(x', x) \mathbf{1}_{S(x) \geq L_{q-1}}$$
$$= \alpha^{-q} \, \eta(x) \mathbf{1}_{S(x) \geq L_{q-1}} \mathbf{k}_q^a(x, dx') \mathbf{1}_{S(x') \geq L_{q-1}}.$$

Accordingly, denoting $w_{q-1} = \mathbf{k}_q^a[1 - G_{q-1}]$, this leads to

$$A_q^{N,(0)} = (L_q^N - L_q) \int_{S(x)=L_q} (1 - \mathbf{r}_q^a(x) - w_{q-1}(x))\alpha^{-q}\eta(x)\mathbf{k}_{q+1}^a[\varphi](x)\frac{dx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q). \qquad (6.12)$$

46

By applying again (2.10), and taking into account that $\varphi(x')\mathbf{1}_{S(x')\geq L_{q-1}} = \varphi(x')$, we have

$$\alpha^{-q}\,\eta(x)\mathbf{k}_{q+1}^a[\varphi](x) = \int \eta_q(dx')\mathbf{k}_{q+1}^a(x',x)\varphi(x'),$$

and finally

$$A_q^{N,(0)}$$
$$= (L_q^N - L_q)\iint_{S(x)=L_q} \eta_q(dx')\varphi(x')\mathbf{k}_{q+1}^a(x',x)(1 - \mathbf{r}_q^a(x) - w_{q-1}(x))\frac{dx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q).$$

Next, we come back to $A_q^{N,(1)}$, defined as

$$A_q^{N,(1)} = \frac{1}{\lceil N\alpha\rceil}\sum_{i=1}^{\lceil N\alpha\rceil}\left(\left(\mathbf{k}_q^a[1 - G_{\eta_{q-1}^N}] + \mathbf{r}_q^a\right)\left(\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q)\right)\right)(\widetilde{X}_{q-1}^i)$$
$$= \tilde{\eta}_{q-1}^N\left(\left(\mathbf{k}_q^a[1 - G_{\eta_{q-1}^N}] + \mathbf{r}_q^a\right)\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q)\right). \tag{6.13}$$

Then, if we denote

$$w_{q-1}^N(x) = \mathbf{k}_q^a[1 - G_{\eta_{q-1}^N}](x) = 1 - \mathbf{r}_q^a(x) - \mathbf{k}_q^a[G_{\eta_{q-1}^N}](x),$$

we have

$$A_q^{N,(1)} = \tilde{\eta}_{q-1}^N\left(\left(w_{q-1}^N + \mathbf{r}_q^a\right)\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q)\right).$$

At this step, it is quite natural to consider the deterministic functions $w_{q-1}^{\delta^-} \leq w_{q-1}^{\delta^+}$ defined by

$$w_{q-1}^{\delta^\pm}(x) = \mathbf{k}_q^a[1 - G_{L_{q-1}\pm\delta}](x) = \mathbf{k}_q^a[1 - G_{L_{q-1}\pm\delta}](x).$$

Accordingly, let us also introduce the random variable

$$\hat{A}_q^{N,(1)} = \tilde{\eta}_{q-1}^N\left(\left(w_{q-1} + \mathbf{r}_q^a\right)\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q)\right). \tag{6.14}$$

In what follows, we assume that f is non-negative, otherwise we decompose $f = f^+ - f^-$ and the same reasoning applies to both parts. If $f \geq 0$, then the same is true for $\varphi = Q_{q+1,p}(f)$ and we have $0 \leq \mathbf{k}_{q+1}^a[\varphi] \leq 1$. Besides, we remark that the sign of $w_{q-1}^N(x) - w_{q-1}(x)$ is independent of x, which is also true for $G_{\eta_q^N}(x) - G_q(x)$. As a consequence, since L_{q-1}^N tends almost surely to L_{q-1}, we have that, almost surely for $N > N_0$,

$$\left|A_q^{N,(1)} - \hat{A}_q^{N,(1)}\right| \leq \left|\Delta_{q-1}^N\right|,$$

where
$$\Delta_{q-1}^N = \tilde{\eta}_{q-1}^N \left((w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(G_{\eta_q^N} - G_q) \right).$$

We will first focus our attention on Δ_{q-1}^N and then exhibit the limit of $\hat{A}_q^{N,(1)}$. Concerning Δ_{q-1}^N, we may reformulate it as

$$\Delta_{q-1}^N = \frac{N}{\lceil N\alpha \rceil} \; \eta_{q-1}^N \left((w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(G_{\eta_q^N} - G_q) \right),$$

and Corollary 6.3 implies that

$$\Delta_{q-1}^N = \frac{1}{\alpha} \; \nu_{q-1}^N \left((w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(G_{\eta_q^N} - G_q) \right) + o_p(1/\sqrt{N}). \qquad (6.15)$$

As before, given $\mathcal{F}_{q-2}^N$, we split

$$\nu_{q-1}^N = \Phi_{q-1}(\eta_{q-2}^N) = \nu_{q-1}^{N,(0)} + \nu_{q-1}^{N,(1)}$$

in absolutely continuous and discrete parts, see equations (6.9) and (6.10) with $(q-1)$ instead of q, $\mathbf{k}_{q-1}^a$ instead of $\mathbf{k}_q^a$ and $\mathbf{r}_{q-1}^a$ instead of $\mathbf{r}_q^a$, leading to

$$\Delta_{q-1}^N = \frac{1}{\alpha} \; \left(\Delta_{q-1}^{N,(0)} + \Delta_{q-1}^{N,(1)} \right) + o_p(1/\sqrt{N}),$$

where

$$\Delta_{q-1}^{N,(0)} = \int \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} \mathbf{k}_{q-1}^a(\widetilde{X}_{q-2}^i, dx)((w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(G_{\eta_q^N} - G_q))(x),$$

and

$$\Delta_{q-1}^{N,(1)} = \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} ((w_{q-2}^N + \mathbf{r}_{q-1}^a)(w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(G_{\eta_q^N} - G_q))(\widetilde{X}_{q-2}^i).$$

Clearly, $\Delta_{q-1}^{N,(0)}$ shares some resemblance with $A_q^{N,(0)}$ as given in (6.11). Therefore, mutatis mutandis, we get an equivalent expression as (6.12), namely

$$\Delta_{q-1}^{N,(0)}$$
$$= (L_q^N - L_q) \int_{S(x)=L_q} \eta_{q-1}(x)((w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(1 - \mathbf{r}_{q-1}^a - w_{q-2}))(x) \; \frac{dx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q).$$

Since $0 \leq 1 - \mathbf{r}_{q-1}^a - w_{q-2} = \mathbf{k}_{q-1}^a[G_{q-2}] \leq 1$, we deduce in particular that

$$\left|\Delta_{q-1}^{N,(0)}\right| \leq |L_q^N - L_q| \int_{S(x)=L_q} \eta_{q-1}(x)(w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(x) \, \frac{đx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q). \tag{6.16}$$

Regarding $\Delta_{q-1}^{N,(1)}$, since $0 \leq w_{q-2}^N \leq 1$, we get $\left|\Delta_{q-1}^{N,(1)}\right| \leq \left|\Delta_{q-2}^N\right|$, with

$$\Delta_{q-2}^N = \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} ((w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(G_{\eta_q^N} - G_q))(\widetilde{X}_{q-2}^i).$$

Putting all pieces together yields

$$\left|A_q^{N,(1)} - \hat{A}_q^{N,(1)}\right| \leq \frac{1}{\alpha}\left(\left|\Delta_{q-1}^{N,(0)}\right| + \left|\Delta_{q-2}^N\right|\right) + o_p(1/\sqrt{N}),$$

and finally

$$\left|A_q^{N,(1)} - \hat{A}_q^{N,(1)}\right| \leq \alpha^{-1}\left|\Delta_{q-1}^{N,(0)}\right| + \cdots + \alpha^{1-q}\left|\Delta_1^{N,(0)}\right| + \alpha^{1-q}\left|\Delta_0^N\right| + o_p(1/\sqrt{N}).$$

By (6.16), for every $k \in \{1, \ldots, q-1\}$, we have the upper-bound

$$\left|\Delta_k^{N,(0)}\right| \leq |L_q^N - L_q| \int_{S(x)=L_q} \eta_k(x)(w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(x) \, \frac{đx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q),$$

and, by (6.15), we have

$$\Delta_0^N = \frac{1}{\alpha} \nu_0^N \left((w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(G_{\eta_q^N} - G_q)\right) + o_p(1/\sqrt{N}).$$

Since $\nu_0^N = \eta_0 = \eta$, the coarea formula yields

$$\left|\Delta_0^N\right| \leq \frac{1}{\alpha} |L_q^N - L_q| \int_{S(x)=L_q} \eta_0(x)(w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(x) \, \frac{đx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q) + o_p(1/\sqrt{N}).$$

Lebesgue's dominated convergence theorem ensures that

$$\int_{S(x)=L_q} \eta_k(x)(w_{q-1}^{\delta^+} - w_{q-1}^{\delta^-})(x) \, \frac{đx}{|DS(x)|} \xrightarrow[\delta \to 0]{} 0,$$

49

and Lemma 6.3 says that $L_q^N - L_q = \mathcal{O}_p(1/\sqrt{N})$, so we conclude that

$$A_q^{N,(1)} - \hat{A}_q^{N,(1)} = o_p(1/\sqrt{N}).$$

Now we turn to the estimation of $\hat{A}_q^{N,(1)}$ as defined in (6.14). The analysis is roughly the same as for Δ_{q-1}^N except that we have to be a bit more precise since this time we want an estimate and not an upper-bound. However, we can reformulate it as

$$\hat{A}_q^{N,(1)} = \frac{N}{\lceil N\alpha \rceil} \, \eta_{q-1}^N \left((\mathbf{r}_q^a + w_{q-1})\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q) \right),$$

and Corollary 6.3 implies that

$$\hat{A}_q^{N,(1)} = \frac{1}{\alpha} \, \nu_{q-1}^N \left((\mathbf{r}_q^a + w_{q-1})\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q) \right) + o_p(1/\sqrt{N}).$$

Again, given $\mathcal{F}_{q-2}^N$, we split $\nu_{q-1}^N = \nu_{q-1}^{N,(0)} + \nu_{q-1}^{N,(1)}$ into its absolutely continuous and discrete parts to get

$$\hat{A}_q^{N,(1)} = \frac{1}{\alpha} \left(A_{q-1}^{N,(0)} + A_{q-1}^{N,(1)} \right) + o_p(1/\sqrt{N}),$$

where, as in (6.11) and (6.13),

$$A_{q-1}^{N,(0)} = \int \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} \mathbf{k}_{q-1}^a(\widetilde{X}_{q-2}^i, dx)((\mathbf{r}_q^a + w_{q-1})\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q))(x),$$

and

$$A_{q-1}^{N,(1)} = \frac{1}{\lceil N\alpha \rceil} \sum_{i=1}^{\lceil N\alpha \rceil} (\mathbf{k}_{q-1}^a(1 - G_{\eta_{q-2}^N}) + \mathbf{r}_{q-1}^a)((\mathbf{r}_q^a + w_{q-1})\mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q))(\widetilde{X}_{q-2}^i)$$

By the same arguments as above, under Assumption $[\mathcal{H}^a]$, it is readily seen that

$$A_{q-1}^{N,(0)} = (L_q^N - L_q)$$
$$\iint_{S(x)=L_q} \eta_{q-1}(dx')\varphi(x')\mathbf{k}_{q+1}^a(x', x)((\mathbf{r}_q^a + w_{q-1})(1 - \mathbf{r}_{q-1}^a - w_{q-2}))(x)\frac{dx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q).$$

Moreover, by the same machinery as for the majorization of Δ_{q-1}^N, we get

$$A_{q-1}^{N,(1)} - \hat{A}_{q-1}^{N,(1)} = o_p(1/\sqrt{N}).$$

Consequently, we have

$$A_q^N = A_q^{N,(0)} + \frac{1}{\alpha} A_{q-1}^{N,(0)} + \frac{1}{\alpha} \hat{A}_{q-1}^{N,(1)} + o_p(1/\sqrt{N}).$$

At this point, it remains to notice that

$$\eta_q(dx')\varphi(x') = \frac{1}{\alpha} \eta_{q-1}(dx')\varphi(x'),$$

which implies that

$$A_q^{N,(0)} + \frac{1}{\alpha} A_{q-1}^{N,(0)} = (L_q^N - L_q)$$
$$\iint_{S(x)=L_q} \eta_q(dx') \mathbf{k}_{q+1}^a(x', x)((1 - \mathbf{r}_q^a - w_{q-1})(\mathbf{r}_{q-1}^a + w_{q-2}))(x) \frac{dx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q),$$

and a straightforward recursion gives

$$A_q^N = (L_q^N - L_q)$$
$$\iint_{S(x)=L_q} \eta_q(dx')\varphi(x') \mathbf{k}_{q+1}^a(x', x)((1 - \mathbf{r}_q^a - w_{q-1}) \ldots (\mathbf{r}_1^a + w_0))(x) \frac{dx}{|DS(x)|}$$
$$+ \alpha^{1-q} \hat{A}_1^{N,(1)} + o_p(L_q^N - L_q) + o_p(1/\sqrt{N}),$$

where

$$\hat{A}_1^{N,(1)} = \frac{1}{\alpha^q} \nu_0^N \left((\mathbf{r}_q^a + w_{q-1}) \ldots (\mathbf{r}_1^a + w_0) \mathbf{k}_{q+1}^a[\varphi](G_{\eta_q^N} - G_q) \right) + o_p(1/\sqrt{N}).$$

Since $\nu_0^N = \eta$, we finally get

$$\hat{A}_1^{N,(1)} = (L_q^N - L_q)$$
$$\iint_{S(x)=L_q} \eta_q(dx')\varphi(x') \mathbf{k}_{q+1}^a(x', x)((\mathbf{r}_q^a + w_{q-1}) \ldots (\mathbf{r}_1^a + w_0))(x) \frac{dx}{|DS(x)|}$$
$$+ o_p(L_q^N - L_q) + o_p(1/\sqrt{N}),$$

so that, coming back to (6.8) and thanks to Proposition 6.3, we have eventually shown that

$$\nu_q^N([Q_{q+1,\eta_q^N} - Q_{q+1}](\varphi)) = o_p(L_q^N - L_q) + o_p(1/\sqrt{N}) = o_p(1/\sqrt{N}).$$

This terminates the proof of Proposition 5.3. ∎

The following lemma is a key tool to prove Proposition 6.3 and its Corollary 6.3, which were useful in the previous proof.

Lemma 6.4 *For any $C > 0$, for any integer $0 \le q < n$ and for any $L \in \{L_q, \ldots, L_{n-1}\}$, consider the class of sets*

$$\mathbf{A}_{N,C} = \left\{ S^{-1}\left(\left[L - \frac{c_1}{\sqrt{N}}, L + \frac{c_2}{\sqrt{N}}\right]\right), 0 < c_1 < C, 0 < c_2 < C \right\}.$$

Then, for any $\phi \in \mathcal{B}(\mathbb{R}^d)$, we have that

$$\sup_{A \in \mathbf{A}_{N,C}} \sqrt{N} \left| \nu_q^N(\phi \mathbf{1}_A) - \eta_q^N(\phi \mathbf{1}_A) \right| \xrightarrow[N \to \infty]{\mathbb{P}} 0.$$

Proof Here again, the proof is made given $\mathcal{F}_{q-1}^N$. Let $A_{N,C}$ denote the largest set in $\mathbf{A}_{N,C}$, i.e.

$$A_{N,C} = S^{-1}\left(\left[L - \frac{C}{\sqrt{N}}, L + \frac{C}{\sqrt{N}}\right]\right).$$

Let us write some preliminary algebra. In the following, k_N stands for the number of sample points belonging to $A_{N,C}$, meaning that

$$k_N = N \times \eta_q^N(A_{N,C}) = \sum_{i=1}^{N} \mathbf{1}_{A_{N,C}}(X_q^i).$$

We start from the decomposition

$$\sup_{A \in \mathbf{A}_{N,C}} \sqrt{N} \left| \nu_q^N(\phi \mathbf{1}_A) - \eta_q^N(\phi \mathbf{1}_A) \right|$$

$$\le \sqrt{N} \, \nu_q^N(A_{N,C}) \sup_{A \in \mathbf{A}_{N,C}} \left| \frac{\nu_q^N(\phi \mathbf{1}_A)}{\nu_q^N(A_{N,C})} - \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_A(X_q^i) \phi(X_q^i) \right| \quad (6.17)$$

$$+ \sup_{A \in \mathbf{A}_{N,C}} \left| \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_A(X_q^i) \phi(X_q^i) \right| \times \left| \sqrt{N} \, \nu_q^N(A_{N,C}) - \frac{k_N}{\sqrt{N}} \right|. (6.18)$$

Consider first expression (6.17). We study the class $\mathbf{A}_{N,C}$ from the viewpoint of Vapnik-Chervonenkis theory (see for example Chapters 12 and 13 in [19]). We denote by $s(\mathbf{A}_{N,C}, N)$ the shattering coefficient of $\mathbf{A}_{N,C}$. Very elementary reasoning gives that $s(\mathbf{A}_{N,C}, N) \le N^2$.

As ϕ is bounded, for any $\varepsilon > 0$ we can find a simple function $\phi^\varepsilon = \sum_{j=1}^{n_\varepsilon} b_j \mathbf{1}_{B_j}$ such that $\|\phi - \phi^\varepsilon\| < \varepsilon$. Let us denote by $\mathbf{B}_\varepsilon$ the finite collection of Borelian sets in the expression of ϕ^ε. If we consider now

$$\mathbf{A}_{N,C}^\varepsilon = \{A = A_1 \cap A_2, A_1 \in \mathbf{A}_{N,C}, A_2 \in \mathbf{B}_\varepsilon\},$$

then it is clear that its shatter coefficient verifies $s(\mathbf{A}_{N,C}^{\varepsilon}, N) \leq 2^{n_\varepsilon} N^2$.

Now, in (6.17), we show that the supremum factor goes to 0 in probability. We first have

$$
\sup_{A \in \mathbf{A}_{N,C}} \left| \frac{\nu_q^N(\phi \mathbf{1}_A)}{\nu_q^N(A_{N,C})} - \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_A(X_q^i) \phi(X_q^i) \right|
$$

$$
\leq \sup_{A \in \mathbf{A}_{N,C}} \left| \frac{\nu_q^N((\phi - \phi^\varepsilon)\mathbf{1}_A)}{\nu_q^N(A_{N,C})} \right| + \sup_{A \in \mathbf{A}_{N,C}} \left| \frac{\nu_q^N(\phi^\varepsilon \mathbf{1}_A)}{\nu_q^N(A_{N,C})} - \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_A(X_q^i) \phi^\varepsilon(X_q^i) \right|
$$

$$
+ \sup_{A \in \mathbf{A}_{N,C}} \left| \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_A(X_q^i)(\phi^\varepsilon - \phi)(X_q^i) \right|,
$$

hence

$$
\sup_{A \in \mathbf{A}_{N,C}} \left| \frac{\nu_q^N(\phi \mathbf{1}_A)}{\nu_q^N(A_{N,C})} - \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_A(X_q^i) \phi(X_q^i) \right|
$$

$$
\leq 2\|\phi - \phi^\varepsilon\| + \sup_{A \in \mathbf{A}_{N,C}} \left| \sum_{j=1}^{n_\varepsilon} b_j \left(\frac{\nu_q^N(\mathbf{1}_{A \cap B_j})}{\nu_q^N(A_{N,C})} - \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_{A \cap B_j}(X_q^i) \right) \right|
$$

$$
\leq 2\varepsilon + \left(\sum_{j=1}^{n_\varepsilon} |b_j| \right) \times \sup_{A \in \mathbf{A}_{N,C}^\varepsilon} \left| \frac{\nu_q^N(\mathbf{1}_A)}{\nu_q^N(A_{N,C})} - \frac{1}{k_N} \sum_{i=1}^{N} \mathbf{1}_A(X_q^i) \right|
$$

$$
\leq 2\varepsilon + \left(\sum_{j=1}^{n_\varepsilon} |b_j| \right) \times \varepsilon'
$$

$$
\leq 3\varepsilon,
$$

for ε' chosen small enough, with probability at least

$$
1 - 8s(\mathbf{A}_{N,C}^\varepsilon, N)e^{-N\varepsilon'^2/32} \geq 1 - 2^{n_\varepsilon+3}N^2 e^{-N\varepsilon'^2/32},
$$

which can be made arbitrarily close to 1 for N large enough. We notice that here we have used Theorem 12.5 in [19], and the fact that, given $\mathcal{F}_{q-1}^N$, the X_q^i's are i.i.d. with distribution ν_q^N, and thus the k_N ones in $A_{N,C}$ are i.i.d. with distribution $\nu_q^N . \mathbf{1}_{A_{N,C}} / \nu_q^N(A_{N,C})$.

Now, to complete the proof of the lemma, it suffices to show that the pre-factor $\sqrt{N}\nu_q^N(A_{N,C})$ in (6.17) can be bounded with arbitrarily large probability. In this aim, we proceed by induction on q. Consider first $q = 0$. In that case $\nu_q^N = \eta$, and it is clear using the coarea formula and the law of large numbers that

$$
\nu_q^N(A_{N,C}) = \mathcal{O}_p(1/\sqrt{N}).
$$

For the general case $q > 0$, we have the decomposition $\nu_q^N = \nu_q^{N,(0)} + \nu_q^{N,(1)}$ where the first term is absolutely continuous with respect to Lebesgue's measure, and the second term is a discrete one. A quick inspection reveals that

$$\nu_q^{N,(0)} \leq \frac{1}{\alpha} \eta_{q-1}^N \mathbf{k}_q^a \qquad \text{and} \qquad \nu_q^{N,(1)} \leq \frac{1}{\alpha} \eta_{q-1}^N. \tag{6.19}$$

When applied to $A_{N,C}$ both are $\mathcal{O}_p(1/\sqrt{N})$. For the first one we simply apply the coarea formula and the law of large numbers. For the second one, we notice that $N\eta_{q-1}^N(A_{N,C})$ is a Binomial r.v. with parameters N and $\nu_{q-1}^N(A_{N,C})$. The mean $\nu_{q-1}^N(A_{N,C})$ is $\mathcal{O}_p(1/\sqrt{N})$ by the induction assumption. For the distance to the mean we use Hoeffding's inequality

$$\mathbb{P}\left(\left| \nu_{q-1}^N(A_{N,C}) - \eta_{q-1}^N(A_{N,C}) \right| \geq \frac{A}{\sqrt{N}} \,\middle|\, \mathcal{F}_{q-2}^N \right) \leq 2e^{-2A^2},$$

which can be made arbitrarily small by choosing A large enough. This shows that $|\nu_{q-1}^N(A_{N,C}) - \eta_{q-1}^N(A_{N,C})|$ is also $\mathcal{O}_p(1/\sqrt{N})$.

Consider now expression (6.18). It is clear that the supremum is less than $\|\phi\|$. For the factor $|\sqrt{N}\nu_q^N(A_{N,C}) - \frac{k_N}{\sqrt{N}}|$, let us denote $I_q^N = \nu_q^N(A_{N,C})$. From usual considerations on the X_q^i's, we see that k_N is Binomial $\mathcal{B}(N, I_q^N)$ distributed, thus we have

$$\mathbb{E}\left[\frac{k_N}{N} \,\middle|\, \mathcal{F}_{q-1}^N \right] = I_q^N \qquad \text{and} \qquad \mathbb{V}\left(\frac{k_N}{N} \,\middle|\, \mathcal{F}_{q-1}^N \right) = \frac{I_q^N(1 - I_q^N)}{N}.$$

By Chebyshev's inequality we deduce that, for any $\varepsilon > 0$,

$$\mathbb{P}\left(\sqrt{N} \left| \frac{k_N}{N} - I_q^N \right| > \varepsilon \,\middle|\, \mathcal{F}_{q-1}^N \right) \leq \frac{I_q^N(1 - I_q^N)}{\varepsilon^2} \xrightarrow[N\to\infty]{\mathbb{P}} 0,$$

since, as justified above, $I_q^N = \nu_q^N(A_{N,C}) = \mathcal{O}_P(1/\sqrt{N})$. ∎

Proposition 6.3 *For all $q \in \{0, \dots, n-1\}$,*

$$L_q^N - L_q = \mathcal{O}_p(1/\sqrt{N}).$$

Proof The proof is done by induction on q. We will actually make the induction on the following double property: for all $\delta > 0$, for all measurable function ϕ such that $0 \leq \phi \leq 1$ and with support above L_q (i.e. $\phi = G_q\phi$), there exist $C > 0$ and N_0 such that for all $N > N_0$, with probability at least $(1 - \delta)$, we have

$$\left| L_q^N - L_q \right| \leq \frac{C}{\sqrt{N}} \qquad \text{and} \qquad \left| \left(\eta_q - \nu_q^N \right)(\phi) \right| \leq \frac{C}{\sqrt{N}}.$$

First note that for $q = 0$, since $\nu_0^N = \eta_0$, the second assertion is trivial, and the first one is obtained by very standard properties of empirical quantiles (e.g. CLT) when the i.i.d. sample is drawn from a distribution with a strictly positive density at point L_0.

Now, assume the property is true up to step $(q-1)$. Then, by we have

$$\alpha \left(\nu_q^N - \eta_q \right) (\phi) = \eta_{q-1}^N \left(G_{L_{q-1}^N} M_{q,\eta_{q-1}^N} \phi \right) - \nu_{q-1}^N \left(G_{q-1} M_q \phi \right) \qquad (6.20)$$

$$+ \left(\nu_{q-1}^N - \eta_{q-1} \right) \left(G_{q-1} M_q \phi \right) + o(1/\sqrt{N}). \qquad (6.21)$$

The second term (6.21) is easy as $\| G_{q-1} M_q \phi \| \leq 1$ and, from the recurrence assumption, its absolute value is less than $C/\sqrt{N}$ with probability at least $(1-\delta)$.

For the first term, namely (6.20), let us write

$$\left| \eta_{q-1}^N \left(G_{L_{q-1}^N} M_{q,\eta_{q-1}^N} \phi \right) - \nu_{q-1}^N \left(G_{q-1} M_q \phi \right) \right|$$

$$\leq \left| \eta_{q-1}^N \left(\left(G_{L_{q-1}^N} M_{q,\eta_{q-1}^N} - G_{q-1} M_q \right) \phi \right) \right| \qquad (6.22)$$

$$+ \left| \nu_{q-1}^N \left(G_{q-1} M_q \phi \right) - \eta_{q-1}^N \left(G_{q-1} M_q \phi \right) \right|. \qquad (6.23)$$

Let us first consider (6.23). Since η_{q-1}^N is an empirical measure of an i.i.d. sample drawn with ν_{q-1}^N, Chebyshev's inequality implies that, for all $t > 0$,

$$\mathbb{P} \left(\left| \nu_{q-1}^N \left(G_{q-1} M_q \phi \right) - \eta_{q-1}^N \left(G_{q-1} M_q \phi \right) \right| \geq t \sigma_N \right) \leq \frac{1}{t^2},$$

with

$$\sigma_N \leq \frac{1}{\sqrt{N}} \sqrt{\nu_{q-1}^N \left[(G_{q-1} M_q \phi)^2 \right]} \leq \frac{1}{\sqrt{N}}.$$

Thus, if we take

$$t = 1/\sqrt{\delta} \qquad \text{and} \qquad C > \frac{1}{\sqrt{\delta}},$$

it turns out that, for N large enough, we have with probability at least $(1-\delta)$,

$$\left| \nu_{q-1}^N \left(G_{q-1} M_q \phi \right) - \eta_{q-1}^N \left(G_{q-1} M_q \phi \right) \right| \leq \frac{C}{\sqrt{N}}.$$

Now we decompose (6.22) in a similar way as (6.8) and taking into account that $G_q \phi = \phi$, which gives

$$\eta_{q-1}^N \left(\left(G_{L_{q-1}^N} M_{q,L_{q-1}^N} - G_{q-1} M_q \right) \phi \right)$$

$$= \eta_{q-1}^N \left(\left(G_{L_{q-1}^N} - G_{q-1} \right) \mathbf{k}_q^a[\phi] \right) - \eta_{q-1}^N \left(\phi \, \mathbf{k}_q^a[G_{L_{q-1}^N} - G_{q-1}] \right). \qquad (6.24)$$

With probability at least $(1 - \delta)$, for N large enough, we have for the second term, using the recurrence assumption and the coarea formula,

$$
\begin{aligned}
&\left| \eta_{q-1}^N \left(\phi \, \mathbf{k}_q^a [G_{L_{q-1}^N} - G_{q-1}] \right) \right| \\
&\leq \left| \eta_{q-1}^N \left(\phi \, \mathbf{k}_q^a [G_{L_{q-1} - \frac{C}{\sqrt{N}}} - G_{L_{q-1} + \frac{C}{\sqrt{N}}}] \right) \right| \\
&\leq \left| \eta_{q-1}^N \left(\phi \int_{\{S(y) = L_{q-1}\}} \mathbf{k}_q^a(\cdot, y) \frac{dy}{|DS(y)|} \right) \right| \times \frac{2C}{\sqrt{N}} + o_p(1/\sqrt{N}),
\end{aligned}
$$

with the main factor converging in probability to

$$
\eta_{q-1} \left(\phi \int_{\{S(y) = L_{q-1}\}} \mathbf{k}_q^a(\cdot, y) \frac{dy}{|DS(y)|} \right).
$$

For the first term in (6.24) we have, thanks to Lemma 6.4,

$$
\eta_{q-1}^N \left(\left(G_{L_{q-1}^N} - G_{q-1} \right) \mathbf{k}_q^a[\phi] \right) = \nu_{q-1}^N \left(\left(G_{L_{q-1}^N} - G_{q-1} \right) \mathbf{k}_q^a[\phi] \right) + o_p(1/\sqrt{N}).
$$

We then upper-bound ν_{q-1}^N like in (6.19) in order to write

$$
\begin{aligned}
&\left| \nu_{q-1}^N \left(\left(G_{L_{q-1}^N} - G_{q-1} \right) \mathbf{k}_q^a[\phi] \right) \right| \\
&\leq \frac{1}{\alpha} \left| \eta_{q-2}^N \left(\left(G_{L_{q-1}^N} - G_{q-1} \right) \mathbf{k}_q^a[\phi] \right) \right| \\
&\quad + \frac{1}{\alpha} \left| \eta_{q-2}^N \left(\int \mathbf{k}_{q-1}^a(\cdot, y)(G_{L_{q-1}^N} - G_{q-1})(y) \mathbf{k}_q^a[\phi](y) dy \right) \right|.
\end{aligned}
$$

For the second term, we use the coarea formula and the recurrence assumption just as above, and for the first term, we replace η_{q-2}^N with ν_{q-2}^N by virtue of Lemma 6.4. We iterate the reasoning until we get terms with $\nu_0^N = \eta$, which can be dealt by applying the coarea formula again.

Now we consider the other part of the recurrence assumption. Let us define the function $F_N(\ell) = 1 - \nu_q^N(G_\ell)$ and

$$
L_{\nu_q^N} = \inf\{t \text{ such that } (1 - F_N(t)) \geq 1 - \alpha\}.
$$

Following the same arguments as in the proof of Theorem 3.1, we can easily see that $L_{\nu_q^N}$ a.s. converges to L_q. We obviously have

$$
\left| L_q^N - L_q \right| \leq \left| L_q^N - L_{\nu_q^N} \right| + \left| L_{\nu_q^N} - L_q \right|. \tag{6.25}
$$

We first deal with $|L_q^N - L_{\nu_q^N}|$. From the proof of Lemma 6.2 we see that

$$F_N(L_{\nu_q^N}) = \alpha + o_{L^2}(1/\sqrt{N}),$$

so that

$$\mathbb{V}(F_N(L_q^N) - F_N(L_{\nu_q^N})| \mathcal{F}_{q-1}^N) \leq 2\,\mathbb{E}\left[\left(F_N(L_q^N) - \alpha\right)^2 \bigm| \mathcal{F}_{q-1}^N\right] + o_p(1/N).$$

Moreover, from Lemma 6.2 we have

$$N\,\mathbb{E}\left[\left(F_N(L_q^N) - \alpha\right)^2 \bigm| \mathcal{F}_{q-1}^N\right] \xrightarrow[N\to\infty]{\mathbb{P}} \alpha(1-\alpha), \qquad (6.26)$$

Hence, using Chebyshev's inequality we see that, given $\mathcal{F}_{q-1}^N$, the random variable $\sqrt{N}(F_N(L_q^N) - F_N(L_{\nu_q^N}))$ is bounded with arbitrarily large probability, and so it is unconditionally, for in (6.26) the limit is deterministic.

As mentioned before, the function F_N is absolutely continuous except at a finite number of points, namely at most the $\lceil N\alpha \rceil$ largest Y_{q-1}^i's. Denoting f_N the density of the absolutely continuous part of F_N, and J_i's the heights of the jumps, we may write

$$F_N(L_{\nu_q^N}) - F_N(L_q^N) = \int_{L_q^N}^{L_{\nu_q^N}} f_N(\ell)d\ell + \sum_{i:Y_{q-1}^i \in [L_q^N, L_{\nu_q^N}]} J_i,$$

where $[L_q^N, L_{\nu_q^N}]$ stands for $[L_q^N, L_{\nu_q^N}]$ or $[L_{\nu_q^N}, L_q^N]$. We want to show that, with large probability,

$$|F_N(L_{\nu_q^N}) - F_N(L_q^N)| \geq \left|\int_{L_q^N}^{L_{\nu_q^N}} f_N(\ell)d\ell\right| \geq C_q|L_q^N - L_{\nu_q^N}|,$$

where $C_q > 0$ is some deterministic constant. We have

$$\int_{L_q^N}^{L_{\nu_q^N}} f_N(\ell)d\ell = \tilde{\eta}_{q-1}^N \int_{L_q^N}^{L_{\nu_q^N}} H_q^{1,a}(.,\ell)\,d\ell,$$

Therefore, using assumption $[\mathcal{H}^a]$, as for N large both L_q^N and $L_{\nu_q^N}$ are close to L_q, we can write

$$-\varepsilon h(.) \leq H_q^{1,a}(.,\ell) - H_q^{1,a}(.,L_q) \leq \varepsilon h(.),$$

uniformly for ℓ between L_q^N and $L_{\nu_q^N}$. From all that we get

$$|F_N(L_{\nu_q^N}) - F_N(L_q^N)| \geq |L_q^N - L_{\nu_q^N}| \times \tilde{\eta}_{q-1}^N (H_q^{1,a}(.,L_q) - 2\varepsilon h).$$

57

By the law of large numbers, the last factor on the right can be made larger than $C_q = \eta_q(H_q^{1,a}(., L_q))/4$ with large probability. Notice that $C_q > 0$ by assumption $[\mathcal{H}^a]$. We conclude by reminding that we have just proved that $\sqrt{N}(F_N(L_q^N) - F_N(L_{\nu_q^N}))$ is bounded with arbitrarily large probability.

Now, for the last term $|L_{\nu_q^N} - L_q|$ of (6.25), the technique is quite similar. From the first part of the recurrence, taking $\phi = G_q$, we have with arbitrarily large probability for N large enough,

$$\left| \nu_q^N(G_q) - \alpha \right| = \left| \nu_q^N(G_q) - \eta_q(G_q) \right| \leq \frac{C}{\sqrt{N}}.$$

But we also may write

$$\left| \nu_q^N(G_q) - \alpha \right| = \left| F_N(L_q) - F_N(L_{\nu_q^N}) \right| + o_{L^2}(1/\sqrt{N}).$$

Using the same reasoning as above, we get that for some deterministic constant $C_q' > 0$,

$$C_q' \left| L_q - L_{\nu_q^N} \right| \leq \left| F_N(L_q) - F_N(L_{\nu_q^N}) \right|,$$

and we conclude following the same line. ∎

Our last result is then a direct application of Lemma 6.4 and Proposition 6.3.

Corollary 6.3 *For any integer $0 \leq q < n$ and for any bounded and measurable function ϕ, we have*

$$\eta_q^N(\phi(G_{\eta_q^N} - G_q)) = \nu_q^N(\phi(G_{\eta_q^N} - G_q)) + o_p(1/\sqrt{N}).$$

6.5   Proof of Proposition 4.1

We will use the following auxiliary result, which corresponds to Lemma 2.2 in Legoll and Lelièvre [28].

Lemma 6.5 *Let f denote a mapping from $\mathbb{R}^d$ to $\mathbb{R}$, then the function $F : \mathbb{R} \to \mathbb{R}$ defined by*

$$F(L) = \int_{S(x)=L} f(x) \frac{dx}{|DS(x)|}$$

is differentiable with derivative

$$\begin{aligned} F'(L) &= \int_{S(x)=L} \text{div}\left(f(x) \frac{DS(x)}{|DS(x)|^2} \right) \frac{dx}{|DS(x)|} \\ &= \int_{S(x)=L} \left[\frac{DS(x) \cdot Df(x)}{|DS(x)|^2} + f(x) \, \text{div}\left(\frac{DS(x)}{|DS(x)|^2} \right) \right] \frac{dx}{|DS(x)|}, \end{aligned}$$

provided that the right-hand side is well defined.

Let us apply this result to the context of Proposition 4.1. We remind the reader that

$$H_q^g(x, L) = \int_{S(x')=L} g(x') K_{q+1}(x, x') \frac{đx'}{|DS(x')|}.$$

By the first expression of the derivative in Lemma 6.5, we have

$$\frac{\partial}{\partial s} H_q^g(x, s) = \int_{S(x')=L} \text{div}_{x'} \left[g(x') \frac{DS(x') K_{q+1}(x, x')}{|DS(x')|^2} \right] \frac{đx'}{|DS(x')|},$$

provided that the right-hand term is well defined. To prove this, notice that

$$\text{div}_{x'} \left[g(x') \frac{DS(x') K_{q+1}(x, x')}{|DS(x')|^2} \right] = g(x') \times \text{div}_{x'} \left[\frac{DS(x') K_{q+1}(x, x')}{|DS(x')|^2} \right]$$
$$+ (Dg(x') \cdot DS(x')) \times \frac{K_{q+1}(x, x')}{|DS(x')|^2},$$

where '$\cdot$' stand for the usual scalar product in $\mathbb{R}^d$. For the first term, we use the fact that g is bounded, while for the second one, we apply Cauchy-Schwarz inequality and the inequality between the Euclidean norm $|\cdot|$ and the L_1 norm $|\cdot|_1$ to obtain

$$\left| \text{div}_{x'} \left[g(x') \frac{DS(x') K_{q+1}(x, x')}{|DS(x')|^2} \right] \right|$$
$$\leq C \times \left| \text{div}_{x'} \left[\frac{DS(x') K_{q+1}(x, x')}{|DS(x')|^2} \right] \right| + |\text{div}[g(x')]|_1 \times \frac{K_{q+1}(x, x')}{|DS(x')|}.$$

Concerning the second term, recall that g belongs to

$$\mathcal{B}_q = \left\{ g : \mathbb{R}^d \to \mathbb{R}, \ \exists (g_0 \ldots, g_{q-1}) \in \mathcal{B}(\mathbb{R}^d)^q, \ g = K_1(g_0) \cdots K_q(g_{q-1}) \right\},$$

so that

$$\frac{\partial g}{\partial x'_j}(x') = \sum_{m=1}^{q} K_1(g_0)(x') \ldots K_{m-1}(g_{m-2})(x') \left(\int \frac{\partial}{\partial x'_j} K_m(x', x'') g_{m-1}(x'') dx'' \right)$$
$$K_{m+1}(g_m)(x') \ldots K_q(g_{q-1})(x'),$$

and since all the mappings g_m's are assumed bounded, we get

$$\left| \frac{\partial g}{\partial x'_j}(x') \right| \leq C \sum_{m=1}^{q} \int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'',$$

and finally

$$|\mathrm{div}[g(x')]|_1 \le C \left[\sum_{m=1}^{q} \sum_{j=1}^{d} \int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'' \right].$$

By the assumption of Proposition 4.1, we deduce that $s \mapsto H_q^g(x, s)$ is differentiable. Moreover, using the mean value theorem, we deduce that

$$\left| H_q^g(x, L) - H_q^g(x, L_q) \right| \le |L - L_q| \times \sup_{s \in (L, L_q)} \left| \frac{\partial}{\partial s} H_q^g(x, s) \right| \le C|L - L_q| \times h(x),$$

with $h \in L^2(\eta)$, so that $[\mathcal{H}](ii)$ is satisfied. ∎

6.6   Proof of the Gaussian case

In order to keep the notation as simple as possible, we will explain what happens in dimension $d = 2$ only. Thus, the score function is defined, for any $x = (x_1, x_2) \in \mathbb{R}^2$, by $S(x) = x_1/|x| = \cos x$, so that $-1 \le S(x) \le 1$ and

$$DS(x) = \left[\frac{x_2^2}{|x|^3}, \frac{-x_1 x_2}{|x|^3} \right] \implies |DS(x)| = \frac{|x_2|}{|x|^2}.$$

Hence, denoting $r = |x|$, one has for any $L \in (-1, +1)$

$$S(x) = L \iff (x_1, x_2) = (|x|L, \pm|x|\sqrt{1 - L^2}) = (rL, \pm r\sqrt{1 - L^2})$$

so that

$$|DS(x)| = \frac{\sqrt{1 - L^2}}{|x|} \xrightarrow[|x|\to\infty]{} 0,$$

and, whatever L, $|DS(x)|$ is clearly not bounded from below on the level set $\{S(x) = L\}$. However, for any test function f and any $L \in (-1, +1)$, the coarea formula gives

$$\int_{S(x)=L} f(x) \frac{dx}{|DS(x)|} = \frac{1}{\sqrt{1 - L^2}} \int_0^\infty f(rL, r\sqrt{1 - L^2}) r \, dr$$
$$+ \frac{1}{\sqrt{1 - L^2}} \int_0^\infty f(rL, -r\sqrt{1 - L^2}) r \, dr. \qquad (6.27)$$

In particular, since in this example X is a centered standard Gaussian random vector in $\mathbb{R}^2$, equation (2.1) shows that the random variable $Y = S(X)$ has density

$$f_Y(s) = \frac{1}{\pi\sqrt{1 - s^2}} \mathbf{1}_{|s|<1}.$$

This is not surprising since the point $X/|X|$ is uniformly distributed on the unit circle so that $Y = X_1/|X|$ is just the cosine of such a point. Moreover, the transition kernel $K = K_{q+1}$ is a Gaussian transition kernel defined, for the tuning parameter $\sigma > 0$, by

$$K(x, x') = \frac{1 + \sigma^2}{2\pi\sigma^2} \exp\left(-\frac{1 + \sigma^2}{2\sigma^2}\left|x' - \frac{x}{\sqrt{1 + \sigma^2}}\right|^2\right).$$

Let us recall that point (i) of Assumption $[\mathcal{H}]$ requires that

$$\int \eta(dx) \left(\int_{S(x')=L_q} K_{q+1}(x, x') \frac{dx'}{|DS(x')|}\right)^2 < \infty.$$

In our context, setting

$$I = \int_{S(x')=L} K_{q+1}(x, x') \frac{dx'}{|DS(x')|},$$

denoting $\alpha = \sqrt{(1 + \sigma^2)/\sigma^2}$ and, for any $x = (x_1, x_2)$,

$$A_+ = \frac{x_1 L + x_2\sqrt{1 - L^2}}{\sqrt{1 + \sigma^2}} \qquad \text{and} \qquad A_- = \frac{x_1\sqrt{1 - L^2} - x_2 L}{\sqrt{1 + \sigma^2}},$$

as well as

$$B_+ = \frac{x_1\sqrt{1 - L^2} + x_2 L}{\sqrt{1 + \sigma^2}} \qquad \text{and} \qquad B_- = \frac{x_1 L - x_2\sqrt{1 - L^2}}{\sqrt{1 + \sigma^2}},$$

a straightforward computation reveals that

$$I = \frac{\phi(\alpha A_-)(\phi(\alpha A_+) + \alpha A_+\Phi(\alpha A_+)) + \phi(\alpha B_+)(\phi(\alpha B_-) + \alpha B_-\Phi(\alpha B_-))}{\sqrt{1 - L^2}},$$

where ϕ and Φ are respectively the pdf and the cdf of a standard Gaussian random variable. Since $\max(|\alpha A_+|, |\alpha B_-|) \leq |x|/\sigma$, we deduce that

$$I \leq \frac{2}{\sqrt{1 - L^2}}\left(1 + \frac{|x|}{\sigma}\right),$$

and

$$\int \eta(dx) \left(\int_{S(x')=L_q} K_{q+1}(x, x') \frac{dx'}{|DS(x')|}\right)^2 \leq \frac{4}{1 - L_q^2} \int_{\mathbb{R}^2} \left(1 + \frac{|x|}{\sigma}\right)^2 \frac{e^{-\frac{|x|^2}{2}}}{2\pi} dx$$

which is obviously finite, and therefore $[\mathcal{H}](i)$ is satisfied.

In order to prove that $[\mathcal{H}](ii)$ is fulfilled as well, we will make use of Proposition 4.1. Consider first the integral in the sum. From the expression of K_q, we have for any $m > 0$,

$$\int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'' \leq \int C(|x'_j| + |x''_j|) K_m(x', x'') dx'' \leq C_1 |x'|^{\alpha_1} + C_2,$$

for C_1, C_2 and α_1 large enough. Consequently, we have the same type of upper-bound for the whole expression in brackets, meaning that

$$\sum_{m=1}^{q} \sum_{j=1}^{d} \int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'' \leq C_1 |x'|^{\alpha_1} + C_2.$$

Then, remembering that on the level set $\{S(x') = L\}$, one has $|DS(x')| = \sqrt{1 - L^2}/|x'|$, and since $L_q - \delta \leq L \leq L_q + \delta$, we are led to

$$\int_{S(x')=L} |x'|^{\alpha_1} \frac{K_{q+1}(x, x')}{|DS(x')|^2} \, dx' \leq \int_{S(x')=L} K_{q+1}(x, x') \frac{|x'|^{2+\alpha_1}}{1 - L^2} \, dx'$$

$$\leq \frac{1}{1 - (L_q + \delta)^2} \int_{S(x')=L} K_{q+1}(x_L, x') |x'|^{2+\alpha_1} \, dx'.$$

where $x_L = (|x|L, \pm|x|\sqrt{1 - L^2})$ when $x' = (|x'|L, \pm|x'|\sqrt{1 - L^2})$. Simple geometric facts indeed show that $|x - x'| \geq |x_L - x'|$ and thus $K_{q+1}(x_L, x') \geq K_{q+1}(x, x')$. Now, by using the same formulation as in (6.27), the last integral is in fact one dimensional, and is up to a constant a moment of a Gaussian random variable, which is polynomial in its mean :

$$\int_{S(x')=L} |x'|^{\alpha_1} \frac{K_{q+1}(x, x')}{|DS(x')|^2} \, dx' \leq \frac{1}{1 - (L_q + \delta)^2} (C_1 |x_L|^{\alpha_2} + C_2).$$

Since $|x_L| = |x|$, we have

$$\int_{S(x')=L} |x'|^{\alpha_1} \frac{K_{q+1}(x, x')}{|DS(x')|^2} \leq \frac{1}{1 - (L_q + \delta)^2} (C_1 |x|^{\alpha_2} + C_2),$$

and more generally,

$$\int_{S(x')=L} \left[\sum_{m=1}^{q} \sum_{j=1}^{d} \int \left| \frac{\partial}{\partial x'_j} K_m(x', x'') \right| dx'' \right] K_{q+1}(x, x') \frac{dx'}{|DS(x')|^2}$$

$$\leq \frac{1}{1 - (L_q + \delta)^2} (C_1 |x|^{\alpha_2} + C_2). \tag{6.28}$$

Hence the second tem in equation (4.1) is upper bounded by a polynomial in $|x|$, which is of course integrable with respect to the Gaussian measure η.

Now we consider the first term in (4.1). Observe first that

$$\frac{DS(x')}{|DS(x')|^2} = \left[|x'|, -|x'|\frac{x'_1}{x'_2}\right] \qquad \text{and} \qquad \text{div}\left(\frac{DS(x')}{|DS(x')|^2}\right) = |x'|\frac{x'_1}{(x'_2)^2},$$

we get, when $S(x') = L$ and setting $r = |x'|$ as before,

$$\frac{DS(x')}{|DS(x')|^2} = \left[r, \mp r\frac{L}{\sqrt{1-L^2}}\right] \qquad \text{and} \qquad \text{div}\left(\frac{DS(x')}{|DS(x')|^2}\right) = \frac{L}{1-L^2}.$$

Then

$$\left|\text{div}_{x'}\left[\frac{DS(x')K_{q+1}(x,x')}{|DS(x')|^2}\right]\right|$$

$$\leq \left|\text{div}\left[\frac{DS(x')}{|DS(x')|^2}\right]\right| K_{q+1}(x,x') + \left|\frac{DS(x')}{|DS(x')|^2}\right| \times |D_{x'}K_{q+1}(x,x')|_1$$

$$\leq \frac{L}{1-L^2}K_{q+1}(x,x') + \frac{|x'|}{\sqrt{1-L^2}} \times |D_{x'}K_{q+1}(x,x')|_1.$$

As before, we have

$$|D_{x'}K_{q+1}(x,x')|_1 \leq C(|x| + |x'|)K_{q+1}(x,x'),$$

and, for any $L \in [L_q - \delta, L_q + \delta]$,

$$\left|\text{div}_{x'}\left[\frac{DS(x')K_{q+1}(x,x')}{|DS(x')|^2}\right]\right|$$

$$\leq \left(\frac{L_q+\delta}{1-(L_q+\delta)^2} + \frac{C|x'|(|x|+|x'|)}{\sqrt{1-(L_q+\delta)^2}}\right)K_{q+1}(x,x').$$

which ensures that, for the term

$$\int_{S(x')=L} \left|\text{div}_{x'}\left[\frac{DS(x')K_{q+1}(x,x')}{|DS(x')|^2}\right]\right| \frac{đx'}{|DS(x')|},$$

we get the same type of upper-bound as in (6.28). Putting all things together, we have shown inequality (4.1) of Proposition 4.1, which means that Assumption $[\mathcal{H}](ii)$ is satisfied. $\blacksquare$

6.7 Proof of Lemma 6.3

Our goal is to prove that, for any integer q and any $\ell \in \{1, 2, 4\}$, we have

$$N^{\ell/2} \, \mathbb{E}\left[\left(U_{1-\alpha}^N - F_N(L_q^N) \right)^\ell \, \Big| \, \mathcal{F}_{q-1}^N \right] \xrightarrow[N\to\infty]{\mathbb{P}} 0.$$

The principle is to sequentially upper-bound the left-hand side. Set $q > 0$ and for any $x \in \mathbb{R}^d$ and any $i = 1, \ldots, N$, let us define the random variables

$$W_q^x := \frac{1}{N} \sum_{j=1}^N \mathbf{1}_{X_q^j = x} \qquad \text{and for } i = 1, \ldots, N, \qquad W_q^i := W_q^{X_q^i}.$$

Then, by definition of $F_N(L)$, it is readily seen that

$$\left| U_{1-\alpha}^N - F_N(L_q^N) \right| \leq \sup_{x \in \mathbb{R}^d} W_q^x = \max_{1 \leq i \leq N} W_q^i.$$

First note that, by the assumption on the gradient of S, this supremum can only be reached at a sample point X_{q-1}^i. Indeed, since the level sets of S have zero Lebesgue measure, then as soon as a transition by the kernel K is accepted, it will give almost surely a unique value of S. Hence, the accumulation of the particles X_q^i on a same point X_{q-1}^j can only be caused by resampling.

Specifically, recall that the multinomial step as described in Section 2.3 consists in drawing an N-sample $(\hat{X}_{q-1}^1, \ldots, \hat{X}_{q-1}^N)$ with common distribution

$$\frac{1}{\lceil N\alpha \rceil} \sum_{j:\ X_{q-1}^j \geq L_{q-1}^N} \delta_{X_{q-1}^i}(dx).$$

Let us denote

$$\{x_1, \ldots, x_{\lceil N\alpha \rceil}\} := \{X_{q-1}^j : \ X_{q-1}^j \geq L_{q-1}^N\}$$

the set of the $\lceil N\alpha \rceil$ particles which are cloned at the multinomial step and, for $1 \leq j \leq \lceil N\alpha \rceil$, N_q^j stands for the random number of clones of x_j. Said differently, we have

$$(N_q^1, \ldots, N_q^{\lceil N\alpha \rceil}) \sim \mathcal{M}\left(N, \left(\frac{1}{\lceil N\alpha \rceil}, \ldots, \frac{1}{\lceil N\alpha \rceil} \right) \right),$$

where $\mathcal{M}(n, (p_1, \ldots, p_m))$ is the multinomial law with parameters n and $(p_1, \ldots, p_m)$. Then a moment's thought reveals that

$$\max_{1 \leq i \leq N} W_q^i \leq \max_{1 \leq i \leq N} W_{q-1}^i \times \max_{1 \leq j \leq \lceil N\alpha \rceil} N_q^j$$

and since the N_q^j's are independent of $\mathcal{F}_{q-1}^N$, we are led to

$$\mathbb{E}\left[\left|U_{1-\alpha}^N - F_N(L_q^N)\right|^\ell \;\Big|\; \mathcal{F}_{q-1}^N\right] \leq \left(\max_{1\leq i\leq N} W_{q-1}^i\right)^\ell \mathbb{E}\left[\left(\max_{1\leq j\leq\lceil N\alpha\rceil} N_q^j\right)^\ell\right].$$

Next, Theorem 4.4 in [18] ensures that

$$\mathbb{E}\left[\left(\max_{1\leq j\leq\lceil N\alpha\rceil} N_q^j\right)^\ell\right] \underset{N\to\infty}{\sim} \left(\frac{\log N}{\log\log N}\right)^\ell.$$

In particular, one has

$$\mathbb{E}\left[\left(\max_{1\leq j\leq\lceil N\alpha\rceil} N_q^j\right)^\ell\right] \leq C_\ell(\log N)^\ell,$$

and a straightforward induction gives

$$\mathbb{E}\left[\left|U_{1-\alpha}^N - F_N(L_q^N)\right|^\ell\right] \leq C_\ell^q(\log N)^{\ell q}\,\mathbb{E}\left[\left(\max_{1\leq i\leq N} W_0^i\right)^\ell\right].$$

Finally, as η_0 is absolutely continuous, $\max_{1\leq i\leq N} W_0^i = 1/N$ and we get

$$N^{\ell/2}\,\mathbb{E}\left[\left|U_{1-\alpha}^N - F_N(L_q^N)\right|^\ell\right] \leq C_\ell^q(\log N)^{\ell q}N^{-\ell/2} \xrightarrow[N\to\infty]{} 0,$$

which concludes the proof of Lemma 6.3. ∎

References

[1] 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.

[2] 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.

[3] A. Beskos, A. Jasra, N. Kantas, and A. Thiéry. On the Convergence of Adaptive Sequential Monte Carlo Methods. *Annals of Apllied Probability*, 2015, To appear,.

[4] G. Biau, F. Cérou, and A. Guyader. New insights into approximate Bayesian computation. *Ann. Inst. Henri Poincaré Probab. Stat.*, 51(1):376–403, 2015.

[5] Z.I. Botev and D.P. Kroese. An efficient algorithm for rare-event probability estimation, combinatorial optimization, and counting. *Methodology and Computing in Applied Probability*, 10(4):471–505, 2008.

[6] C.-E. Bréhier, T. Lelièvre, and M. Rousset. Analysis of Adaptive Multilevel Splitting algorithms in an idealized case. *ArXiv e-prints*, 2014.

[7] J.A. Bucklew. *Introduction to rare event simulation*. Springer Series in Statistics. Springer-Verlag, New York, 2004.

[8] F. Cérou, P. Del Moral, T. Furon, and A. Guyader. Sequential Monte Carlo for rare event estimation. *Stat. Comput.*, 22(3):795–808, 2012.

[9] F. Cérou, P. Del Moral, and A. Guyader. A non asymptotic theorem for unnormalized Feynman-Kac particle models. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(3):629–649, 2011.

[10] F. Cérou, P. Del Moral, F. Le Gland, and P. Lezaud. Genetic genealogical models in rare event analysis. *ALEA Lat. Am. J. Probab. Math. Stat.*, 1:181–203, 2006.

[11] F. Cérou and A. Guyader. Adaptive multilevel splitting for rare event analysis. *Stoch. Anal. Appl.*, 25(2):417–443, 2007.

[12] N. Chopin. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. *Ann. Statist.*, 32(6):2385–2411, 2004.

[13] P. Del Moral. *Feynman-Kac formulae, Genealogical and interacting particle systems with applications*. Probability and its Applications. Springer-Verlag, New York, 2004.

[14] P. Del Moral. *Mean field simulation for Monte Carlo integration*. CRC Press, 2013.

[15] P. Del Moral, A. Doucet, and A. Jasra. Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B*, 68(3):411–436, 2006.

[16] P. Del Moral, A. Doucet, and A. Jasra. On adaptive resampling strategies for sequential Monte Carlo methods. *Bernoulli*, 18(1):252–278, 2012.

[17] P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In Jacques Azéma, Michel Ledoux, Michel Émery, and Marc Yor, editors, *Séminaire de Probabilités XXXIV*, volume 1729 of *Lecture Notes in Mathematics*, pages 1–145. Springer Berlin Heidelberg, 2000.

[18] L. Devroye. *Lecture notes on bucket algorithms*, volume 6 of *Progress in Computer Science*. Birkhäuser Boston, Inc., Boston, MA, 1986.

[19] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New York, 1996.

[20] R. Douc and E. Moulines. Limit theorems for weighted samples with applications to sequential Monte Carlo methods. *Ann. Statist.*, 36(5):2344–2376, 2008.

[21] L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[22] F. Giraud and P. Del Moral. Non-asymptotic analysis of adaptive and annealed feynman-kac particle models. *Bernoulli*, 2015, To appear.

[23] P. Glasserman and Y. Wang. Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.*, 7(3):731–746, 1997.

[24] E. Gobet and G. Liu. Rare event simulation using reversible shaking transformations. *SIAM Journal on Scientific Computing*, 2015, To appear.

[25] A. Guyader, N. Hengartner, and E. Matzner-Løber. Simulation and estimation of extreme quantiles and extreme probabilities. *Applied Mathematics and Optimization*, 64:171–196, 2011. 10.1007/s00245-011-9135-z.

[26] W.K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, April 1970.

[27] H. Kahn and T.E. Harris. Estimation of particle transmission by random sampling. *National Bureau of Standards Appl. Math. Series*, 12:27–30, 1951.

[28] F. Legoll and T. Lelièvre. Effective dynamics using conditional expectations. *Nonlinearity*, 23(9):2131–2163, 2010.

[29] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[30] D. Pollard. *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York, 1984.

[31] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 2004.

[32] M.N. Rosenbluth and A.W. Rosenbluth. Monte Carlo calculation of the average extension of molecular chains. *Journal of Chemical Physics*, 23(2):356–359, 1955.

[33] G. Rubino and B. Tuffin. *Rare Event Simulation Using Monte Carlo Methods*. Wiley Publishing, 2009.

[34] R. Rubinstein. The Gibbs cloner for combinatorial optimization, counting and sampling. *Methodol. Comput. Appl. Probab.*, 11(4):491–549, 2009.

[35] G.R. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[36] E. Simonnet. Combinatorial analysis of the adaptive last particle method. *Statistics and Computing*, 2014.

[37] L. Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4):1701–1762, 1994. With discussion and a rejoinder by the author.