The Annals of Applied Probability
2016, Vol. 26, No. 6, 3559–3601
DOI: 10.1214/16-AAP1185

UNBIASEDNESS OF SOME GENERALIZED ADAPTIVE MULTILEVEL SPLITTING ALGORITHMS

BY CHARLES-EDOUARD BRÉHIER[*,1] MAXIME GAZEAU[†], LUDOVIC GOUDENÈGE[‡], TONY LELIÈVRE[§,2] AND MATHIAS ROUSSET[§,2]

CNRS and Institut Camille Jordan, Université Lyon 1[*], University of Toronto[†], Fédération de Mathématiques de l'École Centrale Paris[‡] and Université Paris-Est, CERMICS (ENPC), INRIA[§]

We introduce a generalization of the Adaptive Multilevel Splitting algorithm in the discrete time dynamic setting, namely when it is applied to sample rare events associated with paths of Markov chains. We build an estimator of the rare event probability (and of any nonnormalized quantity associated with this event) which is unbiased, whatever the choice of the importance function and the number of replicas. This has practical consequences on the use of this algorithm, which are illustrated through various numerical experiments.

1. Introduction. The efficient sampling of rare events is a very important topic in various application fields such as reliability analysis, computational statistics or molecular dynamics. Let us start with describing the typical problem of interest in the context of molecular dynamics.

1.1. *Motivation and mathematical setting.* Let us consider the Markov chain $(X_t)_{t\in\mathbb{N}}$ defined as the time discretization of the overdamped Langevin dynamics:

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

Typically, $X_t \in \mathbb{R}^{3N}$ is a high-dimensional vector giving the positions of N particles in $\mathbb{R}^3$ at time th ($h > 0$ being the time step size), $V : \mathbb{R}^{3N} \to \mathbb{R}$ is the potential function [for any set of positions $x \in \mathbb{R}^{3N}$, $V(x)$ is the energy of the configuration], $\beta = (k_B T)^{-1}$ is the inverse temperature and W_t is a standard Brownian motion (so that $W_{(t+1)h} - W_{th}$ is a vector of $3N$ i.i.d. centered Gaussian random variables with variance h). In many cases of interest, the dynamics (1) is metastable: the N particles remain trapped for very long times in some so-called metastable states. These

Received June 2015; revised November 2015.

[1]Supported by SNF Grant 200020-149871/1 for having funded his postdoctoral position at University of Neuchâtel (January 2015–August 2015).

[2]Supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement Number 614492.

MSC2010 subject classifications. 65C05, 65C35.

Key words and phrases. Rare event, adaptive multilevel splitting algorithms, unbiased estimator.

are for instance regions located around local minima of V. This actually corresponds to a physical reality: the timescale at the molecular level (given by h, which is typically chosen at the limit of stability for the stochastic differential equation) is much smaller than the timescales of interest, which correspond to hopping events between metastable states. Let us denote by $A \subset \mathbb{R}^{3N}$ and $B \subset \mathbb{R}^{3N}$ two (disjoint) metastable states. One problem of interest is then the following: for some initial condition outside A and B and close to A, how to efficiently sample paths which reach B before A. In the context of molecular dynamics, such paths are called reactive paths. The efficient sampling of reactive paths is a very important subject in many applications since it is a way to understand the mechanism of the transition between metastable states. In mathematical terms, one is interested in computing, for a given test function $\varphi : (\mathbb{R}^{3N})^{\mathbb{N}} \to \mathbb{R}$ depending on the path $(X_t)_{t\in\mathbb{N}}$ of the Markov chain, the expectation

$$\mathbb{E}\big(\varphi\big((X_t)_{t\in\mathbb{N}}\big)\mathbb{1}_{\tau_B<\tau_A}\big), \tag{2}$$

where $\tau_A = \inf\{t \in \mathbb{N} : X_t \in A\}$, $\tau_B = \inf\{t \in \mathbb{N} : X_t \in B\}$ and $X_0 = x_0 \notin (A \cup B)$ is assumed to be a deterministic initial position close to A: most trajectories starting from x_0 hit A before B. Generalization to a random initial condition is straightforward, by a conditioning argument. If $\varphi = 1$, the above expectation is $\mathbb{P}(\tau_B < \tau_A)$, namely the probability that the Markov chain reaches B before A. This is typically a very small probability: since A is metastable and x_0 is close to A, for most of the realizations, τ_A is smaller than τ_B. This is why naive Monte Carlo methods will not give reliable estimates of (2). We refer for instance to [8, 17] for some examples in the context of molecular simulation. The problem we would like to address in this article is thus the following: how to build "good" estimators of (2), where $(X_t)_{t\in\mathbb{N}}$ is a general Markov chain, and τ_A and τ_B are two stopping times.

1.2. *A short review of the literature on rare event simulation.* A complete review of rare event simulation techniques is out of the scope of this article; we instead refer the interested reader to [1, 11, 12, 38], for instance. Our aim in this section is mainly to explain the interest of splitting techniques in general, and Adaptive Multilevel Splitting (AMS) in particular, to simulate rare events in some specific contexts.

Two main families of algorithms for the efficient sampling of rare events have been studied and applied successfully in many contexts since the pioneering works on Monte Carlo methods in the 1950s [28, 30, 37].

The first family is known as importance sampling: the probability distribution of interest is modified using an importance function, in order to enhance the realization of the rare events; unbiased estimators are recovered thanks to the use of appropriate likelihood ratios.

It may happen, for example for industrial applications in engineering or chemistry, that the stochastic model is only given as a black-box and cannot be modified.

In such situations, importance sampling algorithms are unpractical, and one may rely on splitting strategies as described below, for which by construction the model does not need to be modified and is used as a black-box.

The second family of methods is given by splitting algorithms, which are iterative procedures based on interacting systems of replicas. These are selected using an importance function, and then possibly duplicated and weighted accordingly. A common interpretation is that the state-space is decomposed in a nested sequence of subsets (which are level-sets for the importance function), such that the rare event probability can be written as a (telescoping) product of conditional probabilities, which are easier to compute.

Note that in the last 20 years, splitting algorithms have been studied extensively, and many variants appeared: to cite a few, Subset Simulation [2], (Multilevel) Splitting [25, 32, 33], Nested Sampling [41] and RESTART [43]. Note also the relations with Genealogical Particle models [14, 19, 21] and Sequential Monte Carlo methods [13, 20]. In the nonadaptive versions of the splitting method, the nested sequence of subsets is fixed a priori and it is easy to build unbiased estimators. However, to the best of our knowledge, similar results do not exist for adaptive versions such as the Adaptive Multilevel Splitting (AMS) algorithm, where the sequence of subsets is built on-the-fly using the ensemble of replicas. This is the focus of this article.

Adaptive versions, as proposed in [15], of multilevel splitting algorithms are needed in practice: the performance of the estimation indeed depends on the choice of the levels (i.e., of the nested susbets), which is not trivial if no additional information on the system is available. In previous theoretical studies [5, 9, 10, 15, 27, 40, 44], the consistency (estimators are unbiased) and the efficiency (the algorithm outperforms crude Monte Carlo methods in the rare event regime, when the size of the system of replica increases) have been analyzed under very restrictive conditions, in so-called idealized settings. Especially, these conditions are not satisfied for processes in dimension larger than one, or for discrete-time dynamics.

Note that the sensitivity of the performance with respect to the choice of the importance function is a well-known fact for both importance sampling and splitting algorithms: the variance deteriorates for inappropriate choices; see [24, 26], and numerical simulations in Section 5. Note that a general strategy to design importance functions for diffusions in small noise regimes is by approximating solutions of Hamilton–Jacobi–Bellman equations; these equations are obtained by asymptotic analysis (often related to large deviations behavior); see, for instance, [4, 22, 42] for importance sampling and [18] for splitting algorithms. In the absence of a small parameter (typically small noise) in order to perform asymptotic analysis, such a technique breaks down.

1.3. *The adaptive multilevel splitting algorithm.* In this article, we focus on the Adaptive Multilevel Splitting (AMS) method which has been proposed in [15].

Let us roughly describe the principle of the method; see Section 4.5 for the precise definition of the algorithm. The crucial ingredient we need is an importance function:

$$\xi : \mathbb{R}^{3N} \to \mathbb{R} \tag{3}$$

which will be used to measure the advance of the paths towards B. This function is known as a reaction coordinate in the molecular dynamics community, and this is the terminology we will use here. In this paper, we also call $\xi(X_t)$ the level of the process X_t at time t, and [see (25)]

$$\Xi(X) = \sup\{\xi(X_{t\wedge\tau_A}) : t \in \mathbb{N}\}$$

the maximum level of the Markov chain path X. A useful requirement on ξ is the existence of a level $z_{\max} \in \mathbb{R}$ such that

$$B \subset \{x \in \mathbb{R}^{3N} : \xi(x) \in]z_{\max}, \infty[\}.$$

Then, starting from a system of n_{rep} replicas (all starting from the same initial condition x_0 and stopped at time τ_A), the idea is to remove the *least fit* paths and to duplicate the remaining paths while keeping a fixed number of replicas. The least fit paths are those with the smallest maximum levels $\Xi(X)$. As soon as one of the least fit paths is removed, one of the remaining path is duplicated and then partially resampled: the new path is a copy of the path up to the maximum level of the removed least fit paths, and the end of the trajectory is then sampled using independent random numbers. The algorithm thus goes through three steps: (i) a level computation step (to determine the level under which paths will be removed: this level is computed as an empirical quantile over the maximum levels among the replicas); (ii) a splitting step (to determine which paths will be removed and which ones of the remaining paths will be duplicated); (iii) a partial resampling step (to generate new paths from the selected paths). By iterating these three steps, one obtains successively systems of n_{rep} paths with an increasing minimum of the maximum levels among the replicas. The algorithm is stopped when the current level is larger than $z_{\max}$, and an estimator of (2) is then built using a weighted empirical average over the replicas. The adaptive feature of the algorithm is in the first step (the level computation step): indeed, at each iteration, paths are removed if their maximum level is below some threshold, and these thresholds are determined iteratively using empirical quantiles, rather than by fixing a priori a deterministic sequence of levels (as it would be the case in nonadaptive splitting, or more generally in standard sequential Monte Carlo algorithms; see [13, 19]). All the details of the algorithm will be given in Section 4.5.

In this work, we focus on the application of the AMS algorithm to sample Markov chains, namely discrete time stochastic dynamics, and not continuous time stochastic dynamics as in [15] for example. The reason is mainly practical: in most cases of interest, even if the original model is continuous in time, it is discretized

in time when numerical approximations are needed. There are actually also many cases where the original model is discrete in time (e.g. kinetic Monte Carlo or Markov state models in the context of molecular dynamics).

The discrete time setting, which is thus of practical interest, raises specific questions in the context of the AMS algorithm. First, in the partial resampling step, a natural question which is answered in this article is whether the path should be copied up to the last time before or first time after it reaches the level of the removed paths. Second, in the discrete time context, it may happen that several paths have exactly the same maximum level. This implies some subtleties in the implementation of the splitting step which have a large influence on the quality of the estimators. We refer to [6] and [7], Section 5.1 for concrete examples where bad implementations lead to strongly biased results. One objective of the present article is actually to elucidate a correct implementation of the AMS algorithm in such a discrete time setting.

1.4. *Main results and outline.* The main results and outline of this work are the following:

- In Section 2, we introduce the Generalized Adaptive Multilevel Splitting (GAMS) framework, which encompasses the AMS algorithm presented above. The interest of this generalized setting is twofold. First, it is very useful to write variants of the classical AMS algorithm (see in particular [7], Section 3.5). Second, it highlights the essential mathematical properties that are required to produce unbiased estimators of quantities such as (2).
- In Section 3, we state and prove the main theoretical result (Theorem 3.2) of this article: algorithms which fit in the GAMS framework yield unbiased estimators of the rare event probability, and more generally of any nonnormalized expectation related to the rare event, of the form (2).
- Section 4 is devoted to the detailed presentation of the AMS algorithm discussed above; appropriate implementations of the level computation and splitting steps are made explicit in Section 4.5. In particular, in Section 4.6 we prove that the algorithm fits in the GAMS framework of Section 2.
- Section 5 is entirely devoted to some numerical experiments which illustrate the unbiasedness result, and discuss the efficiency of the AMS algorithm to sample rare events. We discuss through various numerical experiments the influence of the choice of the reaction coordinate ξ on the variance of the estimators and we end up with some practical recommendations in order to get reliable estimates using the AMS algorithm (see Section 5.2). In particular, using the unbiasedness property proven in this paper, it is possible to compare the results obtained using different parameters (in particular different reaction coordinates) in order to assess the quality of the numerical results. In addition, it is easy to build an unbiased estimator with smaller variance by parallelizing computations using many independent realizations of AMS with a fixed number of replicas.

Let us mention that we very often refer to [7], which is an extended version of the present work with in particular additional numerical experiments, and detailed extensions of the AMS algorithm which enter into the GAMS framework.

From a theoretical perspective, this article presents extensions of previous results on splitting algorithms in two directions. On the one hand, it is well known that unbiased estimators of quantities such as (2) can be built for nonadaptive splitting algorithms (i.e., with a fixed sequence of levels), and one of the main contributions of this article is to extend this highly desirable property, under appropriate assumptions, for adaptive algorithms (when levels are computed on-the-fly). On the other hand, compared to previous results in the literature concerning the AMS algorithm, the main novelty of this work is the proof of the unbiasedness in a general setting and whatever the parameters: the number of replicas, the (minimum) number of replicas sampled at each iteration and the reaction coordinate ξ. In previous works (see, for instance, [10, 27, 40]), unbiasedness is proved in an idealized setting, namely when the reaction coordinate is given by $\xi(x) = \mathbb{P}_x(\tau_B < \tau_A)$ (known as the committor function; here, the subscript $x \in \mathbb{R}^{3N}$ indicates that the Markov chain X_t has x as an initial condition), and for a different partial resampling step, where new replicas are sampled according to the conditional distribution of paths conditioned to reach the level of the removed replicas. In many cases of practical interest, in particular for sampling paths of discrete time processes, these two conditions are not met, and the AMS algorithm of Section 4 is a suitable generalization of existing algorithms to deal with such situations.

The proof of unbiasedness is inspired by the interpretation of the AMS algorithm as a sequential Monte Carlo algorithm in path space, in the spirit of [29] (the selection and mutation steps respectively corresponds to the branching step and the partial resampling step in the AMS algorithm). In this interpretation, the iteration index (or "time" index) of the sequential algorithm is given by the increasing levels defined by the reaction coordinate. We refer the interested reader to [7], Section 3.4, where this analogy is made precise.

As explained above, the bias is only one part of the error when using the AMS algorithm: the statistical error (namely the variance) also plays a crucial role in the quality of the estimator as will be shown numerically in Section 5. There are unfortunately very few theoretical results concerning the influence of the choice of ξ on the statistical error. We refer to [9, 16] for an analysis of the statistical error. For discussions about the role of ξ on the statistical error, we also refer to [23, 24, 36]. In particular, in the numerical experiments, we discuss situations for which the confidence intervals of the estimators associated with different reaction coordinates do not overlap if the number of independent realizations of the algorithm is not sufficiently large. We relate this observation to the well-known phenomenon of "apparent bias" for splitting algorithms; see [24].

We would like to stress that our results hold in the setting where a family of partial resampling kernels indexed by the levels is available (see Section 2.1.2 for a precise definition). This is particularly well suited to the sampling of trajectories

of Markov dynamics (see [7], Section 3.4, for another possible setting). In the terminology of [29], we have in mind the dynamic setting (considered, e.g., in [15]), and not the static setting (considered, e.g., in [13, 16]). Nested sampling, [41], is one instance of an adaptive multilevel splitting algorithm devised in the static setting, and which does not enter into our framework; indeed, the partial resampling kernels used in practice (based on few steps of a Metropolis–Hastings algorithm) do not satisfy our assumptions (see Assumption 2 below).

1.5. *Notation*. Before going into the details, let us provide some general notation which is useful in the following:

- The underlying probability space is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. For m σ-fields $\mathcal{F}^1, \dots, \mathcal{F}^m \subset \mathcal{F}$, $\mathcal{F}^1 \vee \cdots \vee \mathcal{F}^m$ denotes the smallest σ-field on Ω containing all the σ-fields $\mathcal{F}^1, \dots, \mathcal{F}^m$. For any $t, s \in \mathbb{N} = \{0, 1, \dots\}$, $t \wedge s = \min\{t, s\}$ and $t \vee s = \max\{t, s\}$. We use the convention $\inf \varnothing = +\infty$. For two sets A and B which are disjoint, $A \sqcup B$ denotes the disjoint set union.
- We work in the following standard setting: random variables take values in state spaces $\mathcal{E}$ which are Polish (namely metrizable, complete for some distance $d_{\mathcal{E}}$ and separable). The associated Borel σ-field is denoted by $\mathcal{B}(\mathcal{E})$. We will give precise examples below (see, e.g., Section 4.1 for the space of trajectories for Markov chains).

  Then $\mathrm{Proba}(\mathcal{E})$ denotes the set of probability distributions on $\mathcal{E}$. It is endowed with the standard Polish structure associated with the Prohorov–Levy metric which metrizes convergence in distribution, that is, weak convergence of probabilities tested against continuous and bounded test functions (see, e.g., [3]). The distribution of a $\mathcal{E}$-valued random variable X will be denoted by $\mathrm{Law}(X)$.
- If $\mathcal{E}_1$ and $\mathcal{E}_2$ are two Polish state spaces, a Markov kernel (or transition probability kernel) $\Pi(x_1, dx_2)$ from $\mathcal{E}_1$ to $\mathcal{E}_2$ is a measurable map from states in $x_1 \in \mathcal{E}_1$, to probability measures in $\mathrm{Proba}(\mathcal{E}_2)$.
- We use the following standard notation associated with probability transitions: for $\varphi : \mathcal{E}_2 \to \mathbb{R}$ a bounded and measurable test function,

$$\Pi(\varphi)(x_1) = \int_{x_2 \in \mathcal{E}_2} \varphi(x_2) \Pi(x_1, dx_2), \qquad \forall x_1 \in \mathcal{E}_1. \tag{4}$$

  Similarly, we use the notation $\pi(\varphi) = \int_{x \in \mathcal{E}_2} \varphi(x) \pi(dx)$ for $\pi \in \mathrm{Proba}(\mathcal{E}_2)$.
- Let X_1 and X_2 be random variables with values respectively in $\mathcal{E}_1$ and $\mathcal{E}_2$, and Π a Markov kernel from $\mathcal{E}_1$ to $\mathcal{E}_2$. We say that X_2 is sampled according to $\Pi(X_1, \cdot)$ [and we denote $X_2 \sim \Pi(X_1, \cdot)$] if $X_2 = f(X_1, U)$ a.s. where, on the one hand, U is a random variable independent of X_1 and of all the random variables introduced before (namely at previous iterations of the algorithm) and, on the other hand, f is a measurable function which is such that $\Pi(x_1, \cdot) = \mathrm{Law}(f(x_1, U))$, for $\mathrm{Law}(X_1)$-almost every $x_1 \in \mathcal{E}_1$.

- Let X be a random variable with values in $\mathcal{E}$, and $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. We say that X is sampled conditionally on $\mathcal{G}$ if there exists a random variable $X_{\mathcal{G}}$ which is $\mathcal{G}$-measurable, a random variable U independent of $X_{\mathcal{G}}$ and all random variables introduced before, and a measurable function f such that $X = f(X_{\mathcal{G}}, U)$.
- A random system of replicas in $\mathcal{E}$ is denoted by

$$\mathcal{X} = \left(X^{(n)}\right)_{n\in I} \in \mathcal{E}^{\mathrm{rep}}, \qquad \operatorname{card} I < +\infty, \tag{5}$$

where $I \subset \mathbb{N}^* = \mathbb{N} \setminus \{0\}$ is a random finite subset of labels and $(X^{(n)})_{n\in I}$ are elements of $\mathcal{E}$. The space $\mathcal{E}^{\mathrm{rep}}$ is endowed with the following distance: for $\mathcal{X}^1 = (x^{(1,n)})_{n\in I^1}$ and $\mathcal{X}^2 = (x^{(2,n)})_{n\in I^2}$ in $\mathcal{E}^{\mathrm{rep}}$, we set

$$d(\mathcal{X}^1, \mathcal{X}^2) = \begin{cases} 2, & \text{if } I^1 \neq I^2, \\ \min\left\{ \displaystyle\sum_{n\in I^1} d_{\mathcal{E}}\left(x^{(1,n)}, x^{(2,n)}\right), 1 \right\}, & \text{if } I^1 = I^2. \end{cases}$$

Endowed with this distance, the set $\mathcal{E}^{\mathrm{rep}}$ is Polish and we denote by $\mathcal{B}(\mathcal{E}^{\mathrm{rep}})$ the Borel σ-field. This σ-field can also be written as follows:

$$\mathcal{B}(\mathcal{E}^{\mathrm{rep}}) = \bigsqcup_{I\in\mathcal{I}} \mathcal{B}(\mathcal{E})^{\otimes I},$$

where $\mathcal{I}$ denotes the ensemble of finite subsets of $\mathbb{N}^*$ (which is a countable set).
- When we consider systems of weighted replicas, to each replica $X^{(n)}$ of the system $\mathcal{X}$ with label $n \in I$ is attached a weight $G^{(n)} \in \mathbb{R}_+$, and we use the notation $\mathcal{X} = (X^{(n)}, G^{(n)})_{n\in I}$. The topological setting is the same as in the previous item, $\mathcal{E}$ being replaced by the augmented state space $\mathcal{E} \times \mathbb{R}$.

2. Generalized adaptative multilevel splitting. In this section, we introduce a general framework for adaptive multilevel splitting algorithms, which contains in particular the AMS algorithm which is described in Section 4 and applied on a concrete example in Section 5. We refer to this framework as the Generalized Adaptive Multilevel Splitting (GAMS) framework.

The interest of this abstract presentation is twofold. First, it highlights the essential mathematical properties that are required to produce unbiased estimators of quantities such as (2), and more generally (6) below. As will be proven in Section 3, any algorithm which enters into the GAMS framework yields unbiased estimators of such quantities. Second, it is very useful to propose variants of the classical AMS algorithm which still yield unbiased estimators; see [7], Section 3.5.

In Section 2.1, we introduce in a general setting the quantities we are interested in computing, and the main ingredients we need to state the GAMS framework, which is then presented in Section 2.2.

2.1. *The setting.* In this section, we introduce the main ingredients and assumptions we need in order to define the GAMS framework.

Let us introduce the state space $\mathcal{P}$ (typically the path space of a random process); it is assumed to be a Polish space and let us denote $\mathcal{B}(\mathcal{P})$ its Borel σ-field. Let X be a random variable with values in $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ and with probability distribution

$$\pi = \mathbb{P} \circ X^{-1} \in \mathrm{Proba}(\mathcal{P}).$$

The aim of the algorithms we present is to estimate

$$\pi(\varphi) = \int_{\mathcal{P}} \varphi(x)\pi(dx) = \mathbb{E}\big(\varphi(X)\big) \tag{6}$$

for a given bounded measurable observable $\varphi : \mathcal{P} \to \mathbb{R}$. Typically, φ has strong variations within a specific set of states occurring with small probability (rare event). Devising an algorithm which is able to efficiently sample these states [so as to obtain an estimator of $\pi(\varphi)$ with a small variance] is the main goal of the presented algorithms.

We need to introduce two ingredients: a filtration on $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ denoted by $(\mathrm{filt}_z)_{z\in\mathbb{R}}$ and a transition probability kernel $\pi_z(x, \cdot)$ from $\mathbb{R} \times \mathcal{P}$ to $\mathcal{P}$. Concrete examples (mainly for sampling paths of Markov chains) of the objects and procedures considered in the present section are given in Section 4.

2.1.1. *The filtration.* We need an additional structure on $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$, namely a filtration indexed by a real number $z \in \mathbb{R}$ (that we refer to as "level" in the following). It is denoted by

$$(\mathrm{filt}_z)_{z\in\mathbb{R}}. \tag{7}$$

This filtration is a nondecreasing family of sub-σ-fields of $\mathcal{B}(\mathcal{P})$: $\mathrm{filt}_z \subset \mathrm{filt}_{z'} \subset \mathcal{B}(\mathcal{P})$ for any $z < z'$. For a given $z \in \mathbb{R}$, the σ-field filt_z is interpreted below as containing the information required to partially resample replicas, conditionally on the knowledge of states up to level z. By convention, we set

$$\mathrm{filt}_{-\infty} = \{\varnothing, \mathcal{P}\}, \qquad \mathrm{filt}_{+\infty} = \mathcal{B}(\mathcal{P}).$$

For any random variable $X : (\Omega, \mathcal{F}) \to (\mathcal{P}, \mathcal{B}(\mathcal{P}))$, we define a filtration $(\mathrm{filt}_z^X)_{z\in\mathbb{R}}$ on the probability space by pulling-back the filtration $(\mathrm{filt}_z)_{z\in\mathbb{R}}$:

$$\mathrm{filt}_z^X = X^{-1}(\mathrm{filt}_z). \tag{8}$$

2.1.2. *The partial resampling kernel $\pi_z(x, \cdot)$.* We introduce a transition probability kernel from $\mathbb{R} \times \mathcal{P}$ to $\mathcal{P}$:

$$(z, x) \in \mathbb{R} \times \mathcal{P} \mapsto \pi_z(x, \cdot) \in \mathrm{Proba}(\mathcal{P}).$$

By convention, for any $x \in \mathcal{P}$, we set

$$\pi_{-\infty}(x, \cdot) = \pi, \qquad \pi_{+\infty}(x, \cdot) = \delta_x,$$

which is consistent with Assumption 1 below.

This kernel is used to perform the partial resampling in the algorithm: for a given level $z \in \mathbb{R}$ and a given state $x \in \mathcal{P}$, $\pi_z(x, dx')$ is the probability distribution of the offspring which is resampled from x knowing the state x up to level z.

In the following, we will refer to this transition probability kernel as a partial resampling kernel.

Let us emphasize that it is implicitly assumed that one knows a practical procedure to sample according to the probability measure π [step (ii) of the initialization step below] and according to the probability distribution $\pi_z(x, \cdot)$, for any $x \in \mathcal{P}$ and $z \in \mathbb{R}$ [step (ii) of the partial resampling step below].

2.1.3. *Assumptions on* $(\mathrm{filt}_z)_{z\in\mathbb{R}}$ *and* $(\pi_z)_{z\in\mathbb{R}}$. We will need two assumptions on $(\mathrm{filt}_z)_{z\in\mathbb{R}}$ and $(\pi_z)_{z\in\mathbb{R}}$. The first one states a right continuity property of the mapping $z \mapsto \pi_z(\phi)(x)$ and is required to apply the Doob's optional stopping theorem in the proof of Lemma 3.6.

ASSUMPTION 1. For any $x \in \mathcal{P}$, and any continuous bounded test function $\varphi : \mathcal{P} \to \mathbb{R}$,

$$\begin{cases} \mathbb{R} \to \mathbb{R}, \\ z \mapsto \pi_z(\varphi)(x) = \displaystyle\int_{y\in\mathcal{P}} \varphi(y)\pi_z(x, dy) \end{cases}$$

is right-continuous. Moreover, $\lim_{z\to-\infty} \pi_z(\varphi)(x) = \pi_{-\infty}(\varphi)(x) = \pi(\varphi)$.

Second, we require a consistency relation between the filtration $(\mathrm{filt}_z)_{z\in\mathbb{R}}$ and the transition probability kernel $(\pi_z)_{z\in\mathbb{R}}$.

ASSUMPTION 2. Let us consider a random variable X, $(\mathrm{filt}_z^X)_{z\in\mathbb{R}}$ and $(\pi_z)_{z\in\mathbb{R}}$ as introduced above. We assume the following consistency relation: if X is distributed according to $\pi_z(x, \cdot)$ for some $(z, x) \in \mathbb{R} \times \mathcal{P}$, then for any $z' \geq z$ and for any bounded measurable test function $\varphi : \mathcal{P} \to \mathbb{R}$,

$$\mathbb{E}\big(\varphi(X)|\mathrm{filt}_{z'}^X\big) = \pi_{z'}(\varphi)(X) \qquad \text{a.s.}$$

As a consequence (by letting $z \to -\infty$ in the previous assumption), if X is distributed according to π, then, for any $z' \in \mathbb{R}$, $\pi_{z'}(X, \cdot)$ is a version of the law of X conditional on $\mathrm{filt}_{z'}^X$. Therefore, the σ-field $\mathrm{filt}_{z'}^X$ defines $\pi_{z'}(x, \cdot)$ for π-almost all x, and can be interpreted as containing all the information on a replica X required to partially resample its state from level z'.

2.1.4. *Filtrations generated by systems of replicas.* We can construct a filtration $(\mathrm{filt}_z^{\mathrm{rep}})_{z\in\mathbb{R}}$ on the space of replicas $(\mathcal{P}^{\mathrm{rep}}, \mathcal{B}(\mathcal{P}^{\mathrm{rep}}))$ [defined by (5)] by considering for all $z \in \mathbb{R}$ the σ-field

$$\mathrm{filt}_z^{\mathrm{rep}} = \bigsqcup_{I\in\mathcal{I}} (\mathrm{filt}_z)^{\otimes I},$$

where, we recall, $\mathcal{I}$ denotes the ensemble of finite subsets of $\mathbb{N}^*$.

Then, if $\mathcal{X} = (X^{(n)})_{n\in I} \in \mathcal{P}^{\mathrm{rep}}$ denotes a random system of replicas—that is, a random variable $\mathcal{X} : (\Omega, \mathcal{F}) \to (\mathcal{P}^{\mathrm{rep}}, \mathcal{B}(\mathcal{P}^{\mathrm{rep}}))$—we also define the filtration $(\mathrm{filt}_z^{\mathcal{X}})_{z\in\mathbb{R}}$ by the pulling-back procedure:

$$\mathrm{filt}_z^{\mathcal{X}} = \mathcal{X}^{-1}\big(\mathrm{filt}_z^{\mathrm{rep}}\big).$$

We also consistently set

$$\mathrm{filt}_{-\infty}^{\mathcal{X}} = \mathcal{X}^{-1}\big(\mathrm{filt}_{-\infty}^{\mathrm{rep}}\big) = \sigma(I) \quad \text{and} \quad \mathrm{filt}_{+\infty}^{\mathcal{X}} = \mathcal{X}^{-1}\big(\mathrm{filt}_{+\infty}^{\mathrm{rep}}\big) = \sigma(\mathcal{X}),$$

where $\sigma(I)$ is the σ-field generated by the random set of labels I.

REMARK 2.1 (On the definition of the filtrations). In many cases of practical interest, for any $z \in \mathbb{R}$, filt_z is defined as the smallest filtration which makes an application $F_z : \mathcal{P} \to (\mathcal{E}, \mathcal{B}(\mathcal{E}))$ measurable, for some Polish space $\mathcal{E}$. Then $\mathrm{filt}_z^{\mathrm{rep}}$ is the smallest σ-field which makes the application $G_z : \mathcal{P}^{\mathrm{rep}} \to \mathcal{E}^{\mathrm{rep}}$ measurable, with $G_z((x^{(n)})_{n\in I}) = (F_z(x^{(n)}))_{n\in I}$.

2.1.5. *Stopping levels.* We finally introduce the notion of stopping level, which is simply a reformulation of the notion of stopping time in our context where the filtrations are indexed by levels instead of times.

DEFINITION 2.2 (Stopping level, Stopped σ-field). Let $(\mathcal{F}_z)_{z\in\mathbb{R}}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. A stopping level Z with respect to $(\mathcal{F}_z)_{z\in\mathbb{R}}$ is a random variable with values in $\mathbb{R}$ such that $\{Z \le z\} \in \mathcal{F}_z$ for any $z \in \mathbb{R} \cup \{-\infty, +\infty\}$. The stopped σ-field, denoted by $\mathcal{F}_Z$, is characterized as follows:

$$A \in \mathcal{F}_Z \quad \text{if and only if} \quad \forall z \in \mathbb{R},\ A \cap \{Z \le z\} \in \mathcal{F}_z.$$

In particular, Z is a $\mathcal{F}_Z$-measurable random variable.

We are now in position to introduce the GAMS framework in the following section.

2.2. *The generalized adaptive multilevel splitting framework.* The aim of this section is to introduce a general framework for splitting algorithms [which we refer to as the Generalized Adaptive Multilevel Splitting (GAMS) framework in the sequel]. It is a convenient, general procedure which allows to define practically implementable algorithms.

It iterates over three successive steps: (1) a branching or splitting step, (2) a partial resampling step and (3) a level computation step. These steps are performed until a suitable stopping criterion is satisfied. We denote by Q_{iter} the number of iterations, which in general is a random variable.

At each iteration step $q \geq 0$ of the algorithm, the distribution π is approximated by an empirical distribution

$$\hat{\pi}^{(q)} = \sum_{n \in I^{(q)}} G^{(n,q)} \delta_{X^{(n,q)}}, \tag{9}$$

over a system of weighted replicas $\mathcal{X}^{(q)} := (X^{(n,q)}, G^{(n,q)})_{n \in I^{(q)}} \in \mathcal{P}^{\mathrm{rep}}$, where $I^{(q)} \subset \mathbb{N}^*$ is the (random) finite set of labels at step q of the algorithm and $G^{(n,q)} \in \mathbb{R}_+$ is the (random) weight attached to the replica $X^{(n,q)}$.

As it will be proven in Section 3, any algorithm which enters into the GAMS framework is such that, for any bounded measurable function $\varphi : \mathcal{P} \to \mathbb{R}$, $\hat{\pi}^{(q)}(\varphi)$ is an unbiased estimator of $\pi(\varphi)$: for any $q \geq 0$, $\mathbb{E}(\hat{\pi}^{(q)}(\varphi)) = \pi(\varphi)$. Moreover, under appropriate assumptions (see Theorem 3.2), this statement can be generalized when q is replaced by the random number of iterations Q_{iter} of the algorithm:

$$\mathbb{E}\Bigg(\sum_{n \in I^{(Q_{\mathrm{iter}})}} G^{(n, Q_{\mathrm{iter}})} \varphi\big(X^{(n, Q_{\mathrm{iter}})}\big) \Bigg) = \pi(\varphi).$$

The proof of this result (Theorem 3.2) is given in Sections 3.2 and 3.3.

As it will become clear, in order to obtain a fully implementable algorithm from the GAMS framework, three procedures need to be made precise (i) the stopping criterion, (ii) the computation rule of the branching numbers and (iii) the computation of the stopping levels. These procedures require to define three sets of random variables that are used in the GAMS framework presented in the next Section 2.2.1: $(S^{(q)})_{q \geq 0}$ (the exit rule defining Q_{iter}), $(B^{(n,q+1)})_{q \geq 0, n \in I^{(q)}}$ (the branching numbers defining duplication of replicas) and $(Z^{(q)})_{q \geq 0}$ (the random levels defining the partial resampling of replicas states). The precise assumptions on these random variables required to have unbiasedness for (9) (Theorem 3.2) are stated in Assumption 3 below (Section 2.2.2). A concrete example of sets of random variables will be given in Section 4, where the AMS algorithm in a Markov chain context is presented.

2.2.1. *Precise definition of the GAMS framework.* We now introduce the Generalized Adaptive Multilevel Splitting (GAMS) framework, which is an iterative procedure on an integer index $q \geq 0$.

The initialization step ($q = 0$):

(i) Define the initial set of labels $I^{(0)} = \{1, \dots, \operatorname{card} I^{(0)}\} \subset \mathbb{N}^*$, where $\operatorname{card} I^{(0)}$ is assumed to be positive and finite.

(ii) Let $(X^{(n,0)})_{n \in I^{(0)}}$ be a sequence of $\mathcal{P}$-valued i.i.d. random variables, distributed according to the probability measure π.

(iii) Initialize uniformly the weights: for any $n \in I^{(0)}$ set $G^{(n,0)} = 1/\operatorname{card} I^{(0)}$.

(iv) Define the system of weighted replicas $\mathcal{X}^{(0)} = (G^{(n,0)}, X^{(n,0)})_{n \in I^{(0)}}$ and for any $z \in \mathbb{R}$, define the σ-field of events $\mathcal{F}_z^{(0)} = \operatorname{filt}_z^{\mathcal{X}^{(0)}}$.

(v) Sample the initial level $Z^{(0)}$.

(vi) Define the σ-field of events[3] $\mathcal{F}^{(0)} = \mathcal{F}_{Z^{(0)}}^{(0)}$.

Iteration. Iterate on $q \geq 0$, while the stopping criterion is not satisfied.

The stopping criterion. Sample the random variable $S^{(q)} \in \{0, 1\}$.
If $S^{(q)} = 0$, then the algorithm stops and we set $Q_{\text{iter}} = q$.
Otherwise, if $S^{(q)} = 1$, the three following steps are performed.

The splitting (*branching*) *step*:

(i) Conditionally on $\mathcal{F}^{(q)}$, sample the $\mathbb{N}$-valued random branching numbers $(B^{(n,q+1)})_{n \in I^{(q)}}$.
Introduce the set of labels of replicas which are removed from the system: $I_{\text{killed}}^{(q+1)} = \{n \in I^{(q)} : B^{(n,q+1)} = 0\}$.

(ii) Compute $K^{(q+1)} = \sum_{n \in I^{(q)}} \max\{B^{(n,q+1)} - 1, 0\}$ the total number of new replicas.

(iii) Introduce the set $I_{\text{new}}^{(q+1)} = \{\max I^{(q)} + 1, \dots, \max I^{(q)} + K^{(q+1)}\} \subset \mathbb{N}^* \setminus I^{(q)}$ for new labels and update the current set of labels

$$I^{(q+1)} = \big(I^{(q)} \setminus I_{\text{killed}}^{(q+1)}\big) \sqcup I_{\text{new}}^{(q+1)}. \tag{10}$$

(iv) Set a children-parent map $P^{(q+1)} : I_{\text{new}}^{(q+1)} \to I^{(q)} \setminus I_{\text{killed}}^{(q+1)}$ such that for any $n \in I^{(q)} \setminus I_{\text{killed}}^{(q+1)}$ we have

$$\operatorname{card}\big\{n' \in I_{\text{new}}^{(q+1)} : P^{(q+1)}\big(n'\big) = n\big\} = B^{(n,q+1)} - 1.$$

This map associates to the label of a new replica the label of its parent. The map is extended to $I^{(q+1)}$ (i.e., to all remaining replicas) as follows: $P^{(q+1)}(n) = n$ for any $n \in I^{(q)} \setminus I_{\text{killed}}^{(q+1)}$.

(v) Update the weights as follows: for all $n' \in I^{(q+1)}$ and $n \in I^{(q)} \setminus I_{\text{killed}}^{(q+1)}$ such that $P^{(q+1)}(n') = n$,

$$G^{(n',q+1)} = \frac{G^{(n,q)}}{\mathbb{E}(B^{(n,q+1)} | \mathcal{F}^{(q)})}. \tag{11}$$

[3]Assumption 3 ensures that $Z^{(0)}$ is a $(\mathcal{F}_z^{(0)})_{z \in \mathbb{R}}$ - stopping level, so that the σ-field $\mathcal{F}_{Z^{(0)}}^{(0)}$ is well defined.

The partial resampling step:

(i) Replicas in $I^{(q)} \setminus I^{(q+1)}_{\mathrm{killed}}$ are not modified, that is, for any $n \in I^{(q)} \setminus I^{(q+1)}_{\mathrm{killed}}$, $X^{(n,q+1)} = X^{(n,q)}$.

(ii) For each $n' \in I^{(q+1)}_{\mathrm{new}}$, $X^{(n',q+1)}$ is sampled independently according to the distribution $\pi_{Z^{(q)}}(X^{(P^{(q+1)}(n'),q)}, dx)$, that is, by partially resampling the state of its parent replica $X^{(P^{(q+1)}(n'),q)}$ at level $Z^{(q)}$.

Then set $\mathcal{X}^{(q+1)} = (X^{(n,q+1)}, G^{(n,q+1)})_{n \in I^{(q+1)}}$.

The level computation step:

(i) For any $z \in \mathbb{R}$, define the σ-field of events

$$\mathcal{F}^{(q+1)}_z = \mathcal{F}^{(q)} \vee \sigma\big(P^{(q+1)}\big) \vee \mathrm{filt}^{\mathcal{X}^{(q+1)}}_z. \tag{12}$$

Note that the σ-field $\sigma(P^{(q+1)})$ generated by $P^{(q+1)}$ contains in particular the σ-field generated by $(B^{(n,q+1)})_{n \in I^{(q)}}$.

(ii) Sample the next level $Z^{(q+1)} \geq Z^{(q)}$.

(iii) Define the σ-field of events[4] $\mathcal{F}^{(q+1)} = \mathcal{F}^{(q+1)}_{Z^{(q+1)}}$.

Increment. Increment $q \leftarrow q + 1$ and go back to the stopping criterion.

The branching number $B^{(n,q+1)}$ introduced in the splitting step (i) represents the number of offsprings of the replica $X^{(n,q)}$. If $B^{(n,q+1)} \geq 1$, the replica $X^{(n,q)}$ will be split into $B^{(n,q+1)}$ replicas: the old one (parent) $X^{(n,q)}$ with label $n \in I^{(q)}$ and, if $B^{(n,q+1)} > 1$, $B^{(n,q+1)} - 1$ new ones (children) that are defined in the partial resampling step. If $B^{(n,q+1)} = 0$, the replica $X^{(n,q)}$ is removed from the system.

Note that the system of weighted replicas $(\mathcal{X}^{(q)})_{q \geq 0}$ and the associated filtration $(\mathcal{F}^{(q)})_{q \geq 0}$ are actually defined for all $q \geq 0$ (and not only up to the iteration Q_{iter}). This is simply verified by considering the iterative procedure above with $S^{(q)} = 1$ for all $q \geq 0$.

REMARK 2.3. The choice of the new weights in (11) can be generalized in the following way. In the splitting step (v), sample the new weights $G^{(n',q+1)}$ conditionally on $\mathcal{F}^{(q)}$ and $P^{(q+1)}$ and assume that the weights satisfy: for all $n \in I^{(q)}$,

$$\mathbb{E}\left(\mathbb{1}_{B^{(n,q+1)} \geq 1} \sum_{n',\, P^{(q+1)}(n') = n} G^{(n',q+1)} \,\middle|\, \mathcal{F}^{(q)}\right) = G^{(n,q)}.$$

Note that the weights defined by (11) satisfy this generalized requirement, but other strategies are also allowed, such as: for all $n' \in I^{(q+1)}$ and $n \in I^{(q)} \setminus I^{(q+1)}_{\mathrm{killed}}$

[4] Assumption 3 ensures that $Z^{(q+1)}$ is a $(\mathcal{F}^{(q+1)}_z)_{z \in \mathbb{R}}$-stopping level, so that the σ-field $\mathcal{F}^{(q+1)}_{Z^{(q+1)}}$ is well defined.

such that $P^{(q+1)}(n') = n$,

$$G^{(n',q+1)} = \frac{G^{(n,q)}}{B^{(n,q+1)}\mathbb{P}(B^{(n,q+1)} \geq 1 | \mathcal{F}^{(q)})}.$$

2.2.2. *From the GAMS framework to a practical algorithm.* In the GAMS framework, we have defined [see (12)] a family of σ-fields $(\mathcal{F}_z^{(q)})_{q\geq 0, z\in\mathbb{R}}$, which is indexed both by the level $z \in \mathbb{R}$ and by the iteration index $q \geq 0$. At the end of the qth iteration of the algorithm ($q \geq 0$), one can think of the σ-field $\mathcal{F}^{(q+1)} = \mathcal{F}^{(q+1)}_{Z^{(q+1)}}$ as containing all the necessary information required to perform the next step of the algorithm.

To make a practical splitting algorithm which enters into the GAMS framework, three sets of random variables need to be defined: $(S^{(q)})_{q\geq 0}$, $(B^{(n,q+1)})_{q\geq 0, n\in I^{(q)}})$ and $(Z^{(q)})_{q\geq 0}$. From now on, we assume the following on these random variables.

ASSUMPTION 3. The random variables $(S^{(q)})_{q\geq 0}$, $(B^{(n,q+1)})_{q\geq 0, n\in I^{(q)}}$, and $(Z^{(q)})_{q\geq 0})$ satisfy the following properties:

- the random variables $(S^{(q)})_{q\geq 0}$ needed for defining the stopping criterion are such that $S^{(q)}$ is $\mathcal{F}^{(q)}$-measurable;
- for each $q \in \mathbb{N}$, the branching numbers $(B^{(n,q+1)})_{n\in I^{(q)}}$ are with values in $\mathbb{N}$, and are assumed to be sampled conditionally on $\mathcal{F}^{(q)}$ (see Section 1.5 for a precise definition), such that $\mathbb{E}(B^{(n,q+1)}|\mathcal{F}^{(q)}) > 0$ a.s.;
- the stopping levels $(Z^{(q)})_{q\geq 0}$ are with values in $\mathbb{R}$, satisfy $Z^{(q+1)} \geq Z^{(q)}$ and are such that $Z^{(q)}$ is a stopping level with respect to $(\mathcal{F}_z^{(q)})_{z\in\mathbb{R}}$ (see Definition 2.2).

Once these three sets of random variables have been defined, the GAMS framework becomes a practical splitting algorithm which yields an unbiased estimator of (6) (this is the claim of Theorem 3.2 proved in Section 3). A concrete example of such an algorithm is given in Section 4.5.

Let us emphasize that the requirement that $Z^{(q)}$ is a $(\mathcal{F}_z^{(q)})_{z\in\mathbb{R}}$-stopping level will be instrumental to apply Doob's optimal stopping Theorem for martingales indexed by levels z, in order to prove Theorem 3.2.

Note the following result, which is a straightforward consequence of the hypothesis on $(S^{(q)})_{q\geq 0}$ in Assumption 3.

PROPOSITION 2.4. *The random variable Q_{iter} is a stopping time with respect to the filtration $(\mathcal{F}^{(q)})_{q\geq 0}$.*

3. The unbiasedness theorem. In the present section, the unbiasedness of the empirical distribution (9) which estimates π is proven. The main result, Theorem 3.2, is stated in Section 3.1. The last two Sections 3.2 and 3.3 are devoted to the proof of Theorem 3.2.

3.1. *Statement of the main result.* We start with the definition of a useful property.

DEFINITION 3.1. A splitting algorithm which enters into the GAMS framework satisfies the almost sure mass conservation property if

$$\forall q \geq 0, \qquad \sum_{n \in I^{(q)}} G^{(n,q)} = 1 \qquad \text{a.s.} \tag{13}$$

The main theoretical result of this paper is the following.

THEOREM 3.2. *Let $(\mathcal{X}^{(q)})_{0 \leq q \leq Q_{\mathrm{iter}}}$ be the sequence of random systems of weighted replicas generated by an algorithm which enters into the GAMS framework of Section 2. In particular, the Assumptions 1 and 2 on the general setting (see Section 2.1) as well as the Assumption 3 on the stopping criterion, branching numbers and level computations (see Section 2.2.2) are supposed to hold.*

Assume moreover that the number of iterations Q_{iter} is almost surely finite [this condition writes $\mathbb{P}(Q_{\mathrm{iter}} < +\infty) = 1$] and that one of the following conditions is satisfied:

- *Q_{iter} is bounded from above by a deterministic constant,*
- *or the almost sure mass conservation (13) is satisfied.*

Then, for any bounded measurable test function $\varphi : \mathcal{P} \to \mathbb{R}$,

$$\mathbb{E}\big(\hat{\pi}^{(Q_{\mathrm{iter}})}(\varphi)\big) = \pi(\varphi).$$

In particular,[5]

$$\forall q \in \mathbb{N}, \qquad \mathbb{E}\big(\hat{\pi}^{(q)}(\varphi)\big) = \pi(\varphi).$$

Note that an algorithm which enters into the GAMS framework does not necessarily satisfy the almost sure mass conservation property of Definition 3.1. Actually, Theorem 3.2 states that the mass conservation holds on average: $\forall q \geq 0$, $\mathbb{E}(\sum_{n' \in I^{(q)}} G^{(n',q)}) = 1$, by taking $\varphi(x) = 1$ and $Q_{\mathrm{iter}} = q$.

Notice that by choosing a given deterministic sequence of levels $Z^{(q)} = z_q$ in the GAMS framework, one actually recovers the well-known unbiasedness result for nonadaptive splitting algorithms, where the levels are fixed a priori.

The strategy we follow to prove Theorem 3.2 is to introduce the sequence of random variables

$$M^{(q)}(\varphi) = \mathbb{E}\big(\hat{\pi}^{(q)}(\varphi) | \mathcal{F}^{(q)}\big) \tag{14}$$

[5]To obtain $Q_{\mathrm{iter}} = q_0$, one simply has to choose

$$S^{(q)} = \begin{cases} 1, & \text{if } q < q_0, \\ 0, & \text{if } q \geq q_0. \end{cases}$$

for a given bounded measurable test function $\varphi : \mathcal{P} \to \mathbb{R}$, and to show that the process $(M^{(q)}(\varphi))_{q\in\mathbb{N}}$ indexed by q is a martingale with respect to the filtration $(\mathcal{F}^{(q)})_{q\in\mathbb{N}}$. Since, by Proposition 2.4, Q_{iter} is a stopping time for this filtration, Doob's stopping theorem for discrete-time martingales can then be applied to obtain Theorem 3.2. The next two sections are devoted to the proof of Theorem 3.2.

3.2. *Proof of Theorem 3.2.* The following definition of conditionally independent replicas will be useful in the proof.

DEFINITION 3.3. Let Z be a random level, $I \subset \mathbb{N}^*$ a finite random set of indices and $\mathcal{G}$ a σ-field of events. We assume that $\sigma(I) \vee \sigma(Z) \subset \mathcal{G}$. We say that the random system of replicas $(X^{(n)})_{n\in I}$ is independently distributed with distribution $(\pi_Z(X^{(n)}, \cdot))_{n\in I}$ conditionally on $\mathcal{G}$, if for any sequence of bounded measurable functions $(\varphi_n)_{n\geq 1}$ from $\mathcal{P}$ to $\mathbb{R}$, we have

$$\mathbb{E}\Big(\prod_{n\in I} \varphi_n\big(X^{(n)}\big)\Big|\mathcal{G}\Big) = \prod_{n\in I} \pi_Z(\varphi_n)\big(X^{(n)}\big).$$

Let us now state two intermediate propositions before proving Theorem 3.2. The first proposition states that, in the sense of Definition 3.3, the set of replicas with indices in $I^{(q)}$ (resp., $I^{(q+1)}$) are $\mathcal{F}^{(q)}$-conditionally independent [resp., $\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})$-conditionally independent] with explicit distributions.

PROPOSITION 3.4. *Let us consider the setting of Theorem 3.2. For any integer* $q \geq 0$:

(i)$_q$ *The replicas* $(X^{(n,q)})_{n\in I^{(q)}}$ *are independent conditionally on the* σ*-field* $\mathcal{F}^{(q)}$, *with distribution* $(\pi_{Z^{(q)}}(X^{(n,q)}, \cdot))_{n\in I^{(q)}}$.

(ii)$_q$ *The replicas* $(X^{(n',q+1)})_{n'\in I^{(q+1)}}$ *are independent conditionally on the* σ*-field* $\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})$, *with distribution* $(\pi_{Z^{(q)}}(X^{(n',q+1)}, \cdot))_{n'\in I^{(q+1)}}$.

The second proposition states intermediate equalities between conditional averages of the empirical distributions, required to obtain the desired martingale property of $(M^{(q)}(\varphi))_{q\geq 0}$, and which are easily obtained from Proposition 3.4. We introduce the following notation for the weighted empirical distribution at the end of the splitting step (i.e., before the partial resampling step is performed):

$$\hat{\pi}^{(q+1/2)} = \sum_{n'\in I^{(q+1)}} G^{(n',q+1)} \delta_{X^{(P^{(q+1)}(n'),q)}}. \tag{15}$$

PROPOSITION 3.5. *Let us consider the setting of Theorem 3.2. For any integer* $q \geq 0$ *and for any bounded measurable test function* $\varphi : \mathcal{P} \to \mathbb{R}$:

(iii)$_q$ $\mathbb{E}(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}) = \mathbb{E}(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)})$.

(iv)$_q$ $\mathbb{E}(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})) = \mathbb{E}(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma(P^{(q+1)}))$.

The proofs of both Proposition 3.4 and Proposition 3.5 are postponed to Section 3.3. We are now in position to prove Theorem 3.2.

PROOF OF THEOREM 3.2. The proof consists of first proving that the process $(M^{(q)}(\varphi))_{q\geq 0}$ defined by (14) is a $(\mathcal{F}^{(q)})_{q\geq 0}$-martingale and then applying the Doob's optional stopping theorem.

Notice that $\mathbb{E}(M^{(q+1)}(\varphi)|\mathcal{F}^{(q)}) = \mathbb{E}(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)})$ and let us compute the right-hand side. First, from point $(\mathrm{iv})_q$ of Proposition 3.5 and since $\mathcal{F}^{(q)} \subset \mathcal{F}^{(q)} \vee \sigma(P^{(q+1)})$, we get

$$\begin{aligned}\mathbb{E}\big(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)}\big) &= \mathbb{E}\big(\mathbb{E}\big(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma\big(P^{(q+1)}\big)\big)|\mathcal{F}^{(q)}\big)\\ &= \mathbb{E}\big(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}\big).\end{aligned}$$

Second, from point $(\mathrm{iii})_q$ of Proposition 3.5 we have

$$\mathbb{E}\big(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}\big) = \mathbb{E}\big(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\big).$$

We thus have for any $q \geq 0$,

$$\mathbb{E}\big(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)}\big) = \mathbb{E}\big(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\big), \tag{16}$$

and $(M^{(q)}(\varphi))_{q\in\mathbb{N}}$ is therefore a $(\mathcal{F}^{(q)})_{q\in\mathbb{N}}$-martingale.

We now focus on stopping the latter martingale at the random iteration Q_{iter}. By assumption, either the almost sure mass conservation property (13) is satisfied, in which case $(M^{(q)}(\varphi))_{q\in\mathbb{N}}$ is a bounded martingale [since $|M^{(q)}(\varphi)| \leq \|\varphi\|_\infty$], or $Q_{\mathrm{iter}} \leq q_{\max}$ for some deterministic integer $q_{\max} \in \mathbb{N}$. In both cases, we apply the Doob's optional stopping theorem (see, for instance, [39], Chapter 7, Section 2, Theorem 1 and Corollaries 1 and 2) to the martingale $(M^{(q)}(\varphi))_{q\in\mathbb{N}}$ and with the stopping time Q_{iter} with respect to the filtration $(\mathcal{F}^{(q)})_{q\in\mathbb{N}}$. We obtain

$$\mathbb{E}\big(\hat{\pi}^{(Q_{\mathrm{iter}})}(\varphi)\big) = \mathbb{E}\big(M^{(0)}(\varphi)\big) = \pi(\varphi)$$

which completes the proof of Theorem 3.2. □

3.3. *Proofs of Propositions 3.4 and 3.5.* Proposition 3.4 requires an additional intermediate result, namely the propagation Lemma 3.6 below. This lemma gives rigorous conditions under which the property on a system of replicas $(X^{(n)})_{n\in I}$ of being independently distributed with distribution $(\pi_Z(X^{(n)},\cdot))_{n\in I}$ conditionally on $\mathcal{F}$ can be transported from the σ-field $\mathcal{F}$ to a larger σ-field. It is based on Doob's optional stopping theorem for martingales indexed by the level variable z. Notice that it is the only result where the right continuity property of Assumption 1 is used.

LEMMA 3.6. *Let us assume that Assumptions 1 and 2 hold. Let* $Z \in \mathbb{R} \cup \{-\infty, +\infty\}$ *be a random level,* $\mathcal{G}$ *a* σ*-field, and* $I \subset \mathbb{N}^*$ *a finite random set of*

labels. Assume that $\sigma(I) \vee \sigma(Z) \subset \mathcal{G}$. *Consider a random system of replicas* $\mathcal{X} = (X^{(n)})_{n\in I}$, *which is independently distributed with distribution* $(\pi_Z(X^{(n)}, \cdot))_{n\in I}$ *conditionally on* $\mathcal{G}$ (*in the sense of Definition* 3.3). *Set*

$$\forall z \in \mathbb{R}, \qquad \mathcal{G}_z = \mathcal{G} \vee \mathrm{filt}_z^{\mathcal{X}}, \tag{17}$$

and assume that $Z' \in \mathbb{R} \cup \{-\infty, +\infty\}$ *is a stopping level for the filtration* $(\mathcal{G}_z)_{z\in\mathbb{R}}$ *such that, almost surely,* $Z' \geq Z$.

Then the replicas $(X^{(n)})_{n\in I}$ *are independently distributed conditionally on* $\mathcal{G}_{Z'}$, *with distribution* $(\pi_{Z'}(X^{(n)}, \cdot))_{n\in I}$.

PROOF. *Step* 1. The first step consists in proving that for any fixed $z \in \mathbb{R}$, the system of replicas is independently distributed with distribution $(\pi_{Z\vee z}(X^{(n)}, \cdot))_{n\in I}$ conditionally on $\mathcal{G} \vee \mathrm{filt}_z^{\mathcal{X}}$. By a standard monotone class argument, it is sufficient to show that

$$\mathbb{E}\Big(\prod_{n\in I} \varphi_n\big(X^{(n)}\big)\psi_n\big(X^{(n)}\big)Y\Big) = \mathbb{E}\Big(\prod_{n\in I} \pi_{Z\vee z}(\varphi_n)\big(X^{(n)}\big)\psi_n\big(X^{(n)}\big)Y\Big),$$

where $(\varphi_n)_{n\geq 1}$ ranges over bounded measurable test functions from $\mathcal{P}$ to $\mathbb{R}$, $(\psi_n)_{n\geq 1}$ ranges over filt_z-measurable test functions from $\mathcal{P}$ to $\mathbb{R}$, and Y over bounded $\mathcal{G}$-measurable random variables.

Let us denote $\mathcal{I} = \mathbb{E}(\prod_{n\in I} \varphi_n(X^{(n)})\psi_n(X^{(n)})Y)$ the left-hand side. Since Y is $\mathcal{G}$-measurable, by Definition 3.3 of the conditional independence we get that

$$\mathcal{I} = \mathbb{E}\Big(\prod_{n\in I} \pi_Z(\varphi_n\psi_n)\big(X^{(n)}\big)Y\Big).$$

The functions $(\psi_n)_{n\geq 1}$ being filt_z-measurable, they are a fortiori $\mathrm{filt}_{z\vee z'}$-measurable for any $z' \in \mathbb{R}$. Assumption 2 on the partial resampling kernel $(\pi_z)_{z\in\mathbb{R}}$ then yields

$$\pi_{z'}(\varphi_n\psi_n)(x) = \pi_{z'}\big(\pi_{z'\vee z}(\psi_n\varphi_n)\big)(x) = \pi_{z'}\big(\psi_n\pi_{z'\vee z}(\varphi_n)\big)(x).$$

As a consequence, using again that the system of replicas $(X^{(n)})_{n\in I}$ is independently distributed with distribution $(\pi_Z(X^{(n)}, \cdot))_{n\in I}$ conditionally on $\mathcal{G}$ and that Y is $\mathcal{G}$-measurable, we get the following identity:

$$\mathcal{I} = \mathbb{E}\Big(\prod_{n\in I} \pi_Z\big(\psi_n\pi_{Z\vee z}(\varphi_n)\big)\big(X^{(n)}\big)Y\Big) = \mathbb{E}\Big(\prod_{n\in I} \pi_{Z\vee z}(\varphi_n)\big(X^{(n)}\big)\psi_n\big(X^{(n)}\big)Y\Big),$$

and this concludes the first step.

Step 2. We now prove the main claim of this lemma, namely the fact that the replicas $(X^{(n)})_{n\in I}$ are independent with distribution $(\pi_{Z'}(X^{(n)}, \cdot))_{n\in I}$ conditionally on $\mathcal{G}_{Z'}$. Let us first assume that the test functions $(\varphi_n)_{n\in I}$ are continuous from $\mathcal{P}$ to $\mathbb{R}$.

In order to come back to a classical setting to apply Doob's optional stopping theorem, let us introduce a continuous, one-to-one and strictly increasing change of level parametrization $\mathcal{Z} : [0, 1] \to \mathbb{R} \cup \{-\infty, +\infty\}$. Let us consider the following stochastic process indexed by $t \in [0, 1]$:

$$N_t = \mathbb{E}\Big(\prod_{n\in I} \varphi_n\big(X^{(n)}\big) \Big| \mathcal{G}_{\mathcal{Z}(t)} \Big).$$

It is a bounded (since I is $\mathcal{G}_{\mathcal{Z}(t)}$-measurable for all t) and thus uniformly integrable martingale with respect to the filtration $(\mathcal{G}_{\mathcal{Z}(t)})_{t\in[0,1]}$. In addition, $N_1 = \mathbb{E}(\prod_{n\in I} \varphi_n(X^{(n)})|\mathcal{G}_{+\infty})$ where $\mathcal{G}_{+\infty} = \mathcal{G} \vee \mathrm{filt}^{\chi}_{+\infty}$.

Thanks to Step 1 above, we get: almost surely, for all $t \in [0, 1]$,

$$N_t = \prod_{n\in I} \pi_{Z\vee\mathcal{Z}(t)}(\varphi_n)\big(X^{(n)}\big).$$

Therefore, N_t is almost surely a right-continuous bounded martingale from Assumption 1 on $(\pi_z)_{z\in\mathbb{R}}$. By assumption, $T' = \mathcal{Z}^{-1}(Z')$ is a $(\mathcal{G}_{\mathcal{Z}(t)})_{t\in[0,1]}$-stopping level, and we can use a Doob's optional stopping argument for right continuous bounded martingales (see, for instance, [31], Theorem 3.22) to get

$$\mathbb{E}(N_1|\mathcal{G}_{\mathcal{Z}(T')}) = N_{T'}$$

which can be rewritten as (since $Z' \geq Z$ a.s.)

$$\mathbb{E}\Big(\prod_{n\in I} \varphi_n\big(X^{(n)}\big) \Big| \mathcal{G}_{Z'} \Big) = \prod_{n\in I} \pi_{Z'}(\varphi_n)\big(X^{(n)}\big).$$

This equality actually holds for any sequence of bounded measurable functions $(\varphi_n)_{n\in I}$ since continuous bounded functions are separating. This completes the proof of Lemma 3.6. □

Thanks to Lemma 3.6 we can now prove Proposition 3.4.

PROOF OF PROPOSITION 3.4. We proceed by induction on the iteration index $q \geq 0$. We first prove directly the statement $(\mathrm{i})_q \Rightarrow (\mathrm{ii})_q$ and then $(\mathrm{ii})_q \Rightarrow (\mathrm{i})_{q+1}$ using Lemma 3.6. The initialization step consists in proving $(\mathrm{i})_0$ using Lemma 3.6. In this proof, $(\varphi_n)_{n\geq 1}$ denotes a sequence of bounded measurable test functions from $\mathcal{P}$ to $\mathbb{R}$. □

PROOF OF $(\mathrm{i})_0$. The statement $(\mathrm{i})_0$ reads

$$\mathbb{E}\Big(\prod_{n\in I^{(0)}} \varphi_n\big(X^{(n,0)}\big) \Big| \mathcal{F}^{(0)} \Big) = \prod_{n\in I^{(0)}} \pi_{Z^{(0)}}(\varphi_n)\big(X^{(n,0)}\big),$$

where $\mathcal{F}^{(0)} = \mathrm{filt}^{\mathcal{X}^{(0)}}_{Z^{(0)}}$. This is exactly the result of Lemma 3.6, taking $Z = -\infty$, $Z' = Z^{(0)}$, $\mathcal{G} = \sigma(I^{(0)})$, and recalling that the replicas are initially independent and distributed according to π. □

PROOF OF $(\mathrm{i})_q \Rightarrow (\mathrm{ii})_q$. Assume that $(\mathrm{i})_q$ holds. We rewrite property $(\mathrm{ii})_q$ as follows:

$$\begin{aligned}(18)\qquad &\mathbb{E}\Big(\prod_{n'\in I^{(q+1)}} \varphi_{n'}\big(X^{(n',q+1)}\big)\Big|\mathcal{F}^{(q)}\vee\sigma\big(P^{(q+1)}\big)\Big)\\ &\quad=\prod_{n'\in I^{(q+1)}} \pi_{Z^{(q)}}(\varphi_{n'})\big(X^{(n',q+1)}\big),\end{aligned}$$

and we now prove this identity.

Let us recall that in the partial resampling step, the replicas with labels in $I^{(q)}$ are not modified and the replicas $(X^{(n',q+1)})_{n'\in I_{\mathrm{new}}^{(q+1)}}$ are sampled in such a way that they are independently distributed conditionally on $\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$, with distribution $(\pi_{Z^{(q)}}(X^{(P^{(q+1)}(n'),q)},\cdot))_{n'\in I_{\mathrm{new}}^{(q+1)}}$. Therefore, by definition of the total set of labels $I^{(q+1)}$ given in (10), one obtains

$$\begin{aligned}&\mathbb{E}\Big(\prod_{n'\in I^{(q+1)}} \varphi_{n'}\big(X^{(n',q+1)}\big)\Big|\mathcal{F}^{(q)}\vee\sigma\big(P^{(q+1)}\big)\Big)\\ &\quad=\mathbb{E}\Big(\prod_{n\in I^{(q)}\setminus I_{\mathrm{killed}}^{(q+1)}} \varphi_n\big(X^{(n,q)}\big)\prod_{n'\in I_{\mathrm{new}}^{(q+1)}} \varphi_{n'}\big(X^{(n',q+1)}\big)\Big|\mathcal{F}^{(q)}\vee\sigma\big(P^{(q+1)}\big)\Big)\\ &\quad=\prod_{n'\in I_{\mathrm{new}}^{(q+1)}} \pi_{Z^{(q)}}(\varphi_{n'})\big(X^{(P^{(q+1)}(n'),q)}\big)\\ &\qquad\times\mathbb{E}\Big(\prod_{n\in I^{(q)}\setminus I_{\mathrm{killed}}^{(q+1)}} \varphi_n\big(X^{(n,q)}\big)\Big|\mathcal{F}^{(q)}\vee\sigma\big(P^{(q+1)}\big)\Big).\end{aligned}$$

Next, from the induction hypothesis $(\mathrm{i})_q$, the replicas $(X^{(n,q)})_{n\in I^{(q)}}$ are independent with distribution $(\pi_{Z^{(q)}}(X^{(n,q)},\cdot))_{n\in I^{(q)}}$ conditionally on $\mathcal{F}^{(q)}$. Since $P^{(q+1)}$ is sampled conditionally on $\mathcal{F}^{(q)}$, the replicas $(X^{(n,q)})_{n\in I^{(q)}}$ are also independent conditionally on $\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$, with the same distributions. Therefore [notice that $I^{(q)}$ and $I_{\mathrm{killed}}^{(q+1)}$ are $\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$-measurable],

$$\begin{aligned}&\mathbb{E}\Big(\prod_{n\in I^{(q)}\setminus I_{\mathrm{killed}}^{(q+1)}} \varphi_n\big(X^{(n,q)}\big)\Big|\mathcal{F}^{(q)}\vee\sigma\big(P^{(q+1)}\big)\Big)\\ &\quad=\prod_{n\in I^{(q)}\setminus I_{\mathrm{killed}}^{(q+1)}} \pi_{Z^{(q)}}(\varphi_n)\big(X^{(n,q)}\big)\\ &\quad=\prod_{n'\in I^{(q)}\setminus I_{\mathrm{killed}}^{(q+1)}} \pi_{Z^{(q)}}(\varphi_{n'})\big(X^{(P^{(q+1)}(n'),q)}\big).\end{aligned}$$

Gathering the results leads to

$$\mathbb{E}\Big(\prod_{n'\in I^{(q+1)}}\varphi_{n'}\big(X^{(n',q+1)}\big)\Big|\mathcal{F}^{(q)}\vee\sigma\big(P^{(q+1)}\big)\Big)$$
$$=\prod_{n'\in I^{(q+1)}}\pi_{Z^{(q)}}(\varphi_{n'})\big(X^{(P^{(q+1)}(n'),q)}\big).$$

From the partial resampling step and Assumption 2, the following identity holds:

$$\forall q\geq 0,\forall n'\in I^{(q+1)},\qquad \pi_{Z^{(q)}}\big(X^{(n',q+1)},\cdot\big)=\pi_{Z^{(q)}}\big(X^{(P^{(q+1)}(n'),q)},\cdot\big). \tag{19}$$

This completes the proof of (18). □

PROOF OF (ii)$_q\Rightarrow$(i)$_{q+1}$. Let us now assume that (ii)$_q$ holds. To prove that (i)$_{q+1}$ holds, it is sufficient to check that

$$\mathbb{E}\Big(\prod_{n\in I^{(q+1)}}\varphi_n\big(X^{(n,q+1)}\big)\Big|\mathcal{F}^{(q+1)}\Big)=\prod_{n\in I^{(q+1)}}\pi_{Z^{(q+1)}}(\varphi_n)\big(X^{(n,q+1)}\big).$$

This is again exactly the result of Lemma 3.6 applied to $\mathcal{X}^{(q+1)}$, taking $Z=Z^{(q)}$, $Z'=Z^{(q+1)}$ and $\mathcal{G}=\mathcal{F}^{(q)}\vee\sigma(P^{(q+1)})$ so that $\mathcal{G}_z=\mathcal{F}_z^{(q+1)}$ [where, we recall, $\mathcal{F}_z^{(q+1)}$ is defined by equation (12)]. □

Finally, let us prove Proposition 3.5.

PROOF OF PROPOSITION 3.5. The first equality (iii)$_q$ is a direct consequence of the definition of the branching numbers. The second equality (iv)$_q$ is obtained as a consequence of Proposition 3.4 by combining (i)$_q$ and (ii)$_q$. □

PROOF OF (iii)$_q$. The proof of this assertion is a direct application of the branching rule. Indeed, by definition of the weights $G^{(n',q+1)}$ given in (11), by definition of the branching numbers $(B^{(n,q+1)})_{n\in I^{(q)}}$ as the number of offsprings of the nth replica, and because these branching numbers are independent of $(G^{(n,q)},X^{(n,q)})_{n\in I^{(q)}}$ conditionally on $\mathcal{F}^{(q)}$, we get

$$\mathbb{E}\big(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)}\big)$$
$$=\mathbb{E}\Big(\sum_{n'\in I^{(q+1)}}\frac{G^{(P^{(q+1)}(n'),q)}}{\mathbb{E}(B^{(P^{(q+1)}(n'),q+1)}|\mathcal{F}^{(q)})}\varphi\big(X^{(P^{(q+1)}(n'),q)}\big)\Big|\mathcal{F}^{(q)}\Big)$$
$$=\mathbb{E}\Big(\sum_{n\in I^{(q)}}\frac{G^{(n,q)}}{\mathbb{E}(B^{(n,q+1)}|\mathcal{F}^{(q)})}B^{(n,q+1)}\varphi\big(X^{(n,q)}\big)\Big|\mathcal{F}^{(q)}\Big)$$
$$=\mathbb{E}\big(\hat{\pi}^{(q)}(\varphi)|\mathcal{F}^{(q)}\big).$$

□

PROOF OF $(\mathrm{i})_q + (\mathrm{ii})_q \Rightarrow (\mathrm{iv})_q$. Using successively $(\mathrm{i})_q$, the identity (19) and $(\mathrm{ii})_q$, we have

$$\begin{aligned}\mathbb{E}\big(\hat{\pi}^{(q+1/2)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma\big(P^{(q+1)}\big)\big) &= \sum_{n' \in I^{(q+1)}} G^{(n',q+1)} \pi_{Z^{(q)}}(\varphi)\big(X^{(P^{(q+1)}(n'),q)}\big) \\ &= \sum_{n' \in I^{(q+1)}} G^{(n',q+1)} \pi_{Z^{(q)}}(\varphi)\big(X^{(n',q+1)}\big) \\ &= \mathbb{E}\big(\hat{\pi}^{(q+1)}(\varphi)|\mathcal{F}^{(q)} \vee \sigma\big(P^{(q+1)}\big)\big). \qquad \square\end{aligned}$$

4. The AMS algorithm for Markov chains. The goal of this section is to define an Adaptive Multilevel Splitting algorithm based on the GAMS framework of Section 2.2.1, when applied to paths of a Markov chain (namely a discrete time stochastic process). In particular, we provide explicit examples of filtrations $(\mathrm{filt}_z)_{z\in\mathbb{R}}$ and partial resampling kernels $(\pi_z)_{z\in\mathbb{R}}$ introduced in Section 2.1, as well as examples of level computation and branching rules in the algorithm.

In order to satisfy Assumptions 1, 2 and 3, and thus to obtain unbiased estimators in this setting, a special care is required to treat the situations when many replicas have the same maximum level, or the situations when there is extinction of the population of replicas. These aspects which are specific to the discrete time setting were not treated in details in many previous works where continuous time diffusions were considered.

4.1. *The Markov chain setting.* Let $\tilde{X} = (\tilde{X}_t)_{t\in\mathbb{N}}$ be a Markov chain, with probability transition P, which takes values in a Polish state space $\mathcal{S}$. Without loss of generality, we assume that $\tilde{X}_0 = x_0$ where $x_0 \in \mathcal{S}$ is a deterministic initial condition.

The state space is the path space

$$\mathcal{P} = \big\{x = (x_t)_{t\in\mathbb{N}} : x_t \in \mathcal{S} \text{ for all } t \in \mathbb{N}\big\}. \tag{20}$$

It is well known that, by introducing the distance $d_{\mathcal{P}}(x, y) = \sum_{t\in\mathbb{N}} \frac{1}{2^t}(1 \wedge \sup_{s\le t} d_{\mathcal{S}}(x_s, y_s))$ (which is a metric for the product topology), the space $(\mathcal{P}, d_{\mathcal{P}})$ is complete and separable. We thus see $\tilde{X}$ as a random variable with values in $\mathcal{P}$.

4.2. *The rare event of interest.* Given two disjoint Borel subsets A and B of $\mathcal{S}$, our objective is the efficient sampling of events such as $\{\tau_B < \tau_A\}$ where

$$\tau_A = \inf\{t \in \mathbb{N} : \tilde{X}_t \in A\} \quad \text{and} \quad \tau_B = \inf\{t \in \mathbb{N} : \tilde{X}_t \in B\}$$

are respectively the first entrance times in A and B. Both τ_A and τ_B are stopping times with respect to the natural filtration of the process $\tilde{X}$.

We are mainly interested in the estimation of the probability $\mathbb{P}(\tau_B < \tau_A)$ in the rare event regime, namely when this probability is very small (typically less

than 10^{-8}). As explained in the Introduction, this occurs for example if the initial condition $x_0 \in A^c \cap B^c$ is such that x_0 is close to A, and A and B are metastable regions for the dynamics. The Markov chain starting from (a neighborhood of) A (resp., B) remains for a very long time near A (resp. B) before going to B (resp., A), and thus, the Markov chain starting from x_0 reaches A before B with a probability close to one. A specific example will be studied in Section 5.

The $\mathcal{P}$-valued random variable X considered to apply the GAMS framework and the associated results is the Markov chain stopped at time τ_A: $X = (X_t)_{t\in\mathbb{N}}$ where

$$X_t = \tilde{X}_{t\wedge\tau_A} \qquad \text{for any } t \in \mathbb{N}. \tag{21}$$

The probability distribution on $(\mathcal{P}, \mathcal{B}(\mathcal{P}))$ is $\pi = \mathrm{Law}(X)$, namely the law of the stopped Markov chain X. The aim of the AMS algorithm is to estimate the (small) probability

$$p = \mathbb{P}(\tau_B < \tau_A) = \mathbb{E}(\mathbb{1}_{\mathrm{T}_B(X)<\mathrm{T}_A(X)}) = \pi(\mathbb{1}_{\mathrm{T}_B(\cdot)<\mathrm{T}_A(\cdot)}), \tag{22}$$

where we denote for any path $x \in \mathcal{P}$

$$\mathrm{T}_A(x) = \inf\{t \in \mathbb{N} : x_t \in A\} \quad \text{and} \quad \mathrm{T}_B(x) = \inf\{t \in \mathbb{N} : x_t \in B\}. \tag{23}$$

More generally, we build unbiased estimators of $\mathbb{E}(\varphi((X_t)_{t\in\mathbb{N}})\mathbb{1}_{\tau_B<\tau_A})$, for bounded measurable function $\varphi : \mathcal{S} \to \mathbb{R}$ [see equation (2)].

REMARK 4.1 (On the stopping times τ_A and τ_B). We defined above the stopping times as first entrance times in some sets A and B. As will become clear below, the definition of the algorithm and the unbiasedness result only require τ_A and τ_B to be stopping times with respect to the natural filtration of the chain $\tilde{X}$.

4.3. *Reaction coordinate*. The crucial ingredient we need to introduce the AMS algorithm is a reaction coordinate, to play the role of an importance function. This is a measurable $\mathbb{R}$-valued mapping defined on the state space $\mathcal{S}$:

$$\xi : \mathcal{S} \to \mathbb{R}.$$

The choice of a good function ξ for given sets A and B is a difficult problem in general. One of the main aims of this paper is to show that whatever the choice of ξ, it is possible to define an unbiased estimator of (22), (2) and more generally (6). The only requirement we impose on ξ is that there exists a constant $z_{\max} \in \mathbb{R}$ such that

$$B \subset \xi^{-1}(]z_{\max}, +\infty[). \tag{24}$$

REMARK 4.2 [On assumption (24)]. Assumption (24) is extremely useful in practice when computing estimators, of (22) and (2); see Section 4.7: it allows to remove from memory the replicas which are declared "retired" in the splitting step at each iteration in the AMS algorithm described in the sequel, since by construction we know in advance that they will not contribute to the computation of the associated estimator. The algorithm thus only requires to retain a fixed number of replicas, denoted by n_{rep} below.

In what follows, the values of ξ are called levels, and they precisely allow us to specify the levels $Z^{(q)}$ computed at each iteration of the GAMS framework. We will very often refer to the maximum level of a path, defined as follows.

DEFINITION 4.3. For any path $x \in \mathcal{P}$, the *maximum level* of x is defined as the supremum of ξ along the path x stopped at $\mathrm{T}_A(x)$:

$$\Xi(x) = \sup\{\xi(x_{t\wedge \mathrm{T}_A(x)}) : t \in \mathbb{N}\} \in \mathbb{R} \cup \{+\infty\}. \tag{25}$$

The function Ξ can be seen as a reaction coordinate on the path space $\mathcal{P}$.

We also introduce for any level $z \in \mathbb{R}$ and any path $x \in \mathcal{P}$

$$\mathrm{T}_z(x) = \inf\{t \in \{0, \dots, \mathrm{T}_A(x)\} : \xi(x_t) > z\}, \tag{26}$$

which is the first entrance time of the path x stopped at $\mathrm{T}_A(x)$ in the set $\xi^{-1}(]z, +\infty[)$. We emphasize on the strict inequality in the above definition of the entrance times $\mathrm{T}_z(x)$: it is one of the important ingredients which allows us to apply the GAMS framework in the Markov chain setting, and thus to obtain unbiased estimators of (6). Notice that the above assumption (24) on B is equivalent to the inequality

$$\forall x \in \mathcal{P}, \qquad \mathrm{T}_{z_{\max}}(x) \le \inf\{t \in \{0, \dots, \mathrm{T}_A(x)\} : x_t \in B\}.$$

We denote by $\tau_z = \mathrm{T}_z(X)$ the entrance time associated with the (stopped) Markov chain X. It is a stopping time for the natural filtration of the Markov chain.

4.4. *Filtration and partial resampling kernel.* We are now in position to define the filtration $(\mathrm{filt}_z)_{z\in\mathbb{R}}$ and the partial resampling kernel $(\pi_z)_{z\in\mathbb{R}}$ of Section 2.1 in the Markov chain setting.

For all $z \in \mathbb{R}$, filt_z is the smallest σ-field which makes the application $x \in \mathcal{P} \mapsto x_{t\wedge(\mathrm{T}_z(x))} \in \mathcal{P}$ measurable, where T_z is defined by (26):

$$\mathrm{filt}_z = \sigma\big(x \mapsto (x_{t\wedge(\mathrm{T}_z(x))})_{t\ge 0}\big). \tag{27}$$

For all $z \in \mathbb{R}$ the partial resampling kernel π_z is defined as follows: for any $x \in \mathcal{P}$, $\pi_z(x, dx') \in \mathrm{Proba}(\mathcal{P})$ is the law of the $\mathcal{P}$-valued random variable Y such that

$$\begin{cases} Y_t = x_t, & \text{if } t \le \mathrm{T}_z(x), \\ \mathrm{Law}(Y_t | Y_s, 0 \le s \le t-1) = P(Y_{t-1}, \cdot), & \text{if } t > \mathrm{T}_z(x) \end{cases} \tag{28}$$

and is stopped at $\mathrm{T}_A(Y)$, when Y hits A. We recall that P is the transition kernel of the Markov chain X. In other words, for $t \le \mathrm{T}_z(x)$, Y is identically x, while for $t > \mathrm{T}_z(x)$, Y_t is generated according to the Markov dynamics on $\mathcal{S}$, with probability transition P, and stopped when reaching A. The partial resampling kernel thus performs a branching of the path x at time $\mathrm{T}_z(x)$ and position $x_{\mathrm{T}_z(x)}$.

Notice from the definition of the partial resampling kernel that if $\Xi(x) \le z$, then this kernel does not modify x: $\mathrm{T}_z(x) = +\infty$ and $\pi_z(x, dx')$ is a Dirac mass: $Y_t = x_{t\wedge \mathrm{T}_A(x)}$ for any $t \in \mathbb{N}$.

Let us now check that Assumptions 1 and 2 are satisfied. The conditions of Assumption 2 are direct consequences of the strong Markov property applied to the chain $t \mapsto X_t \in \mathcal{S}$ defined by (21) at the stopping time τ_z (the strong Markov property always holds true for discrete-time Markov processes).

The right-continuity property of Assumption 1 crucially relies on the definition (26) of $\mathrm{T}_z(x)$ as the entrance time of the path $t \mapsto x_t$ in the level set $\xi^{-1}(]z, +\infty[)$: the fact that $]z, \infty[$ is an open set implies $z \mapsto \mathrm{T}_z(x)$ is right continuous. Indeed, we have the following lemma.

LEMMA 4.4. *Assumption 1 is satisfied for the partial resampling kernel defined by (28). More precisely, for any $x \in \mathcal{P}$, the function $z \in \mathbb{R} \mapsto \pi_z(x, \cdot) \in \mathrm{Proba}(\mathcal{P})$ is piecewise constant and right continuous.*

PROOF. First, assume that $\mathrm{T}_z(x) = +\infty$, which means that $\Xi(x) \le z$. Then, for any $\varepsilon \ge 0$ we still have $\mathrm{T}_{z+\varepsilon}(x) = +\infty$. In that case, $\pi_z(x, \cdot)$ is a Dirac mass: $\pi_z(x, \cdot) = \pi_{z+\varepsilon}(x, \cdot) = \delta_{(x_{t\wedge \mathrm{T}_A(x)})_{t\ge 0}}$.

Now, assume that $\mathrm{T}_z(x) < +\infty$. Then, for $\varepsilon \in]0, \xi(x_{\mathrm{T}_z(x)}) - z[$, $\mathrm{T}_z(x) = \mathrm{T}_{z+\varepsilon}(x)$, and by the definition of the partial resampling kernel, $\pi_z(x, \cdot) = \pi_{z+\varepsilon}(x, \cdot)$. □

4.5. *The AMS algorithm.* In this section, we introduce the AMS algorithm for the sampling of Markov chain trajectories. It is based on the GAMS framework of Section 2.2.1; in addition to the framework, as explained in Section 2.2.2, we make precise the following rules: the stopping criterion, the computation of branching numbers and the computation of the stopping levels. We check below that they satisfy Assumption 3. As a consequence, the GAMS framework encompasses the AMS algorithm, and unbiased estimators of (6) can be defined.

The AMS algorithm iteratively generates a system of weighted replicas in the state space $\mathcal{P}^{\mathrm{rep}}$, using selection and partial resampling steps. The set of all the labels of the replicas generated by the algorithm up to iteration q is denoted by $I^{(q)} = I^{(q)}_{\mathrm{on}} \sqcup I^{(q)}_{\mathrm{off}}$, where $I^{(q)}_{\mathrm{on}}$ is the set of labels of "working" replicas, $I^{(q)}_{\mathrm{off}}$ is the set of labels of replicas which have been declared "retired." We recall that $\sqcup$ denotes the disjoint set union.

The cardinal of $I^{(q)}$ is increasing, while the number of "working" replicas is kept fixed: $\mathrm{card}\, I^{(q)}_{\mathrm{on}} = n_{\mathrm{rep}}$ for any q, where n_{rep} is specified by the user of the algorithm.

An additional parameter $k \in \{1, \dots, n_{\mathrm{rep}} - 1\}$ finally is needed: it is the (minimum) number of replicas sampled at each step of the algorithm. The levels $Z^{(q)}$ are computed as kth order statistics of the maximum levels of "working" replicas.

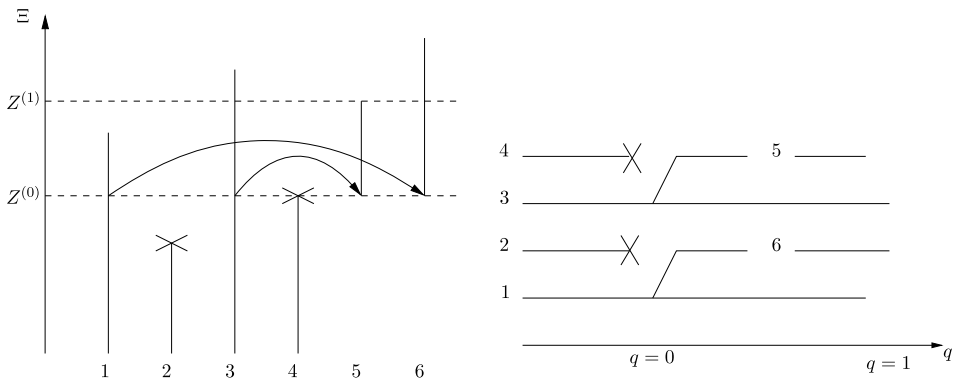

FIG. 1. *Schematic representation of the first iteration of the AMS algorithm, with* $n_{\mathrm{rep}} = 4$ *and* $k = 2$. *The replicas numbered* 2 *and* 4 *are declared retired at the first iteration, and are replaced by the replicas with label* 5 *and* 6, *which are respectively partially resampled from the replicas with labels* 3 *and* 1.

At iteration q, all replicas with maximum levels lower or equal to $Z^{(q)}$ are declared "retired", and new replicas are sampled in order to keep a fixed number n_{rep} of replicas with maximum level strictly larger than $Z^{(q)}$.

We are now in position to introduce the AMS algorithm in full detail (see Figure 1 for a schematic representation of one iteration of the algorithm).

The initialization step ($q = 0$):

(i) Let $(X^{(n,0)})_{1\le n\le n_{\mathrm{rep}}}$ be i.i.d. replicas of the stopped Markov chain in $\mathcal{P}$, distributed according to π. Initialize the sets of labels of working and retired replicas: $I^{(0)} = I^{(0)}_{\mathrm{on}} = \{1, \dots, n_{\mathrm{rep}}\}$ and $I^{(0)}_{\mathrm{off}} = \varnothing$.

(ii) Initialize the weights: $G^{(n,0)} = 1/n_{\mathrm{rep}}$ for $n \in \{1, \dots, n_{\mathrm{rep}}\}$.

(iii) Compute a permutation $\Sigma^{(0)}$ of $I^{(0)}_{\mathrm{on}} = \{1, \dots, n_{\mathrm{rep}}\}$ such that

$$\Xi\big(X^{(\Sigma^{(0)}(1),0)}\big) \le \cdots \le \Xi\big(X^{(\Sigma^{(0)}(n_{\mathrm{rep}}),0)}\big)$$

and set the initial level as the kth order statistics:[6]

$$Z^{(0)} = \Xi\big(X^{(\Sigma^{(0)}(k),0)}\big).$$

(iv) If $\mathrm{card}\{n \in I^{(0)}_{\mathrm{on}} : \Xi(X^{(n,0)}) \le Z^{(0)}\} = n_{\mathrm{rep}}$, then set $Z^{(0)} = +\infty$.

Iterations. Iterate on $q \ge 0$, while the following stopping criterion is not satisfied.

[6]Notice that $\Sigma^{(0)}$ is not necessarily unique since several replicas may have the same maximum level. Nevertheless, the level $Z^{(0)}$ does not depend on the choice of $\Sigma^{(0)}$. The same remark applies to the definition of the level $Z^{(q+1)}$ at iteration $q \ge 0$.

The stopping criterion. If $Z^{(q)} > z_{\max}$, then the algorithm stops and set $Q_{\mathrm{iter}} = q$. Else perform the following four steps.

The splitting (*branching*) *step*:

(i) Set

$$I^{(q)}_{\mathrm{on},>Z^{(q)}} = \left\{ n \in I^{(q)}_{\mathrm{on}},\, \Xi\left(X^{(n,q)}\right) > Z^{(q)} \right\}$$

and $I^{(q)}_{\mathrm{on},\leq Z^{(q)}} = I^{(q)}_{\mathrm{on}} \setminus I^{(q)}_{\mathrm{on},>Z^{(q)}}$.

Set $K^{(q+1)} = \operatorname{card} I^{(q)}_{\mathrm{on},\leq Z^{(q)}} = n_{\mathrm{rep}} - \operatorname{card} I^{(q)}_{\mathrm{on},>Z^{(q)}} \geq k$.

(ii) Introduce a new set $I^{(q+1)}_{\mathrm{new}} = \{\operatorname{card} I^{(q)} + 1, \ldots, \operatorname{card} I^{(q)} + K^{(q+1)}\} \in \mathbb{N}^* \setminus I^{(q)}$ of labels for the new replicas sampled at iteration q.

(iii) Sample the children-parent mapping $P^{(q+1)} : I^{(q+1)}_{\mathrm{new}} \to I^{(q)}_{\mathrm{on},>Z^{(q)}}$, by picking the $K^{(q+1)}$ labels $(P^{(q+1)}(n'))_{n' \in I^{(q+1)}_{\mathrm{new}}}$ independently and uniformly in $I^{(q)}_{\mathrm{on},>Z^{(q)}}$.

Extend the map by setting $P^{(q+1)}(n) = n$ for all $n \in I^{(q)}$.

(iv) Update the sets of labels:

$$I^{(q+1)}_{\mathrm{on}} = I^{(q)}_{\mathrm{on},>Z^{(q)}} \sqcup I^{(q+1)}_{\mathrm{new}}, \qquad I^{(q+1)}_{\mathrm{off}} = I^{(q)}_{\mathrm{off}} \sqcup I^{(q)}_{\mathrm{on},\leq Z^{(q)}},$$

$$I^{(q+1)} = I^{(q+1)}_{\mathrm{on}} \sqcup I^{(q+1)}_{\mathrm{off}}.$$

Update the weights:

$$\begin{cases} G^{(n',q+1)} = G^{(P^{(q+1)}(n'),q)} & n' \in I^{(q+1)}_{\mathrm{off}}, \\ G^{(n',q+1)} = \dfrac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} G^{(P^{(q+1)}(n'),q)} & n' \in I^{(q+1)}_{\mathrm{on}}. \end{cases} \tag{29}$$

The partial resampling step:

(i) Replicas in $I^{(q)}$ are not modified: for $n \in I^{(q)}$, $X^{(n,q+1)} = X^{(n,q)}$.

(ii) For each $n' \in I^{(q+1)}_{\mathrm{new}}$, sample independently $X^{(n',q+1)}$ according to the partial resampling distribution: $\pi_{Z^{(q)}}(X^{(P^{(q+1)}(n'),q)}, dx')$.

The level computation step. Compute a bijective mapping $\Sigma^{(q+1)} : \{1, \ldots, n_{\mathrm{rep}}\} \to I^{(q+1)}_{\mathrm{on}}$ such that

$$\Xi\left(X^{(\Sigma^{(q+1)}(1),q+1)}\right) \leq \cdots \leq \Xi\left(X^{(\Sigma^{(q+1)}(n_{\mathrm{rep}}),q+1)}\right)$$

and set the new level as the kth order statistics:

$$Z^{(q+1)} = \Xi\left(X^{(\Sigma^{(q+1)}(k),q+1)}\right). \tag{30}$$

If $\operatorname{card}\{n \in I^{(q+1)}_{\mathrm{on}} : \Xi(X^{(n,q+1)}) \leq Z^{(q+1)}\} = n_{\mathrm{rep}}$, then set $Z^{(q+1)} = +\infty$.

Increment. Increment $q \leftarrow q+1$, and go back to the stopping criterion step.

This algorithm follows the steps of the GAMS framework, as introduced in Section 2.2. The filtrations $(\mathcal{F}_z^{(q)})_{z\in\mathbb{R}}$ and the σ-fields $\mathcal{F}^{(q)}$, for $q \in \mathbb{N}$, are defined for the AMS algorithm as explained in Section 2.2.

Important remarks on the algorithm need to be made.

Notice that necessarily, in the splitting step (i), $\operatorname{card} I^{(q)}_{\mathrm{on},>Z^{(q)}} \geq 1$ (otherwise $Z^{(q)} = +\infty$ and the stopping criterion has been fulfilled before entering the splitting step of iteration q). As a consequence, the sampling in the splitting step (iii) is well defined.

Note that by construction $\operatorname{card} I^{(q+1)}_{\mathrm{on}} = n_{\mathrm{rep}}$.

Notice that the number of times the loop consisting of the three steps (splitting / partial resampling / level computation) is performed is

$$Q_{\mathrm{iter}} = \inf\{q \geq 0 : Z^{(q)} > z_{\max}\}.$$

If $Z^{(Q_{\mathrm{iter}})} = +\infty$, none of the working replicas at the iteration $Q_{\mathrm{iter}} - 1$ is above the new level $\Xi(X^{(\Sigma^{(Q_{\mathrm{iter}})}(k),\, Q_{\mathrm{iter}})})$ and thus, all of them would have been declared retired at the iteration Q_{iter}: this situation is referred to as extinction. We refer to [7], Remark 2.4, for a discussion of this phenomenon.

Finally, it is very important to notice that the number of replicas sampled at a given iteration q is at least k, but may be larger than k: $K^{(q+1)}$ is not necessarily equal to k. This requires at least two replicas to have $Z^{(q)}$ as the maximum level at the beginning of iteration q, see [7], Figure 2. Actually, it may even happen that, in the level computation step, all the replicas in $I^{(q+1)}_{\mathrm{on}}$ have $Z^{(q+1)}$ as the maximum level, which implies extinction: $\operatorname{card}\{n \in I^{(q+1)}_{\mathrm{on}} : \Xi(X^{n,q+1}) \leq Z^{(q+1)}\} = n_{\mathrm{rep}}$, $Z^{(q+1)} = +\infty$ and the algorithm stops, as explained above.

REMARK 4.5 (Variants and extensions). Using the Generalized Adaptive Multilevel Splitting framework, it is possible to propose variants of the AMS algorithm presented in this section. We refer to [7], Section 3.5, for examples which allow to remove extinction, to reduce the computational cost associated with sorting procedures (computing the level $Z^{(q)}$ at each iteration q using a subset of the ensemble of replicas) or to apply additional selection.

The GAMS framework also allows for different general settings: splitting algorithms can be used to sample other random variables than paths of Markov chains with levels defined as $\sup\{\xi(X_{t\wedge\tau_A})_{t\geq 0}\}$ for some stopping time τ_A and some reaction coordinate function ξ. Actually, under appropriate assumptions, the following cases also enter into the setting of the GAMS framework: path-dependent reaction coordinates (duration of the path, integral over the path), sampling of continuous time stochastic processes (diffusions, jump processes, branching processes), sampling of nonhomogeneous stochastic processes, etc. Finally, the GAMS framework can for instance be applied for the sampling of a Gaussian bridge distribution; see [7], Section 3.5.

4.6. *Verification of Assumption 3.* We prove that the three procedures (the stopping criterion, the computation rule of the branching numbers and the computation of the stopping levels) which are implemented in the AMS algorithm above satisfy the requirements of Assumption 3.

The stopping criterion. In the AMS algorithm, we set $S^{(q)} = \mathbb{1}_{Z^{(q)} \leq z_{\max}}$ which is indeed a $\mathcal{F}^{(q)}$-measurable random variable, since $Z^{(q)}$ is a $(\mathcal{F}_z^{(q)})_{z\in\mathbb{R}}$-stopping level; see Lemma 4.6 below.

The computation rule of the branching numbers. The branching numbers $B^{(n,q+1)}$ are defined by

$$B^{(n,q+1)} = 1 + \operatorname{card}\bigl\{n' \in I_{\text{new}}^{(q+1)} : P^{(q+1)}(n') = n\bigr\}$$

for any $n \in I^{(q)}_{\text{on},>Z^{(q)}}$. We extend the definition for $n \in I^{(q)} \setminus I^{(q)}_{\text{on},>Z^{(q)}}$ by simply setting $B^{(n,q+1)} = 1$. It is then easy to check that they satisfy the requirements of Assumption 3. Notice that in the AMS algorithm, the total number of new replicas $K^{(q+1)} = \sum_{n\in I^{(q)}} \max\{B^{(n,q+1)} - 1, 0\}$ is given by $K^{(q+1)} = \operatorname{card} I^{(q)}_{\text{on},\leq Z^{(q)}}$. Moreover, all branching numbers are positive, so that $I^{(q+1)}_{\text{killed}} = \varnothing$. Another particular feature of the AMS algorithm is that the map $P^{(q+1)}$ takes values in the strict subset $I^{(q)}_{\text{on},>Z^{(q)}}$ of $I^{(q)}$.

Let us check that the computation rule (29) for the weights in the AMS algorithm is indeed consistent with the formula (11) given in the GAMS framework. First, for $n \in I^{(q+1)}_{\text{off}} = I^{(q)}_{\text{on},\leq Z^{(q)}} \sqcup I^{(q)}_{\text{off}}$, $B^{(n,q+1)} = 1$, $P^{(q+1)}(n) = n$ and, consistently, $G^{(n,q+1)} = G^{(n,q)}$. Second, for $n \in I^{(q)}_{\text{on},>Z^{(q)}}$, it is clear that $\mathbb{E}(B^{(n,q+1)}|\mathcal{F}^{(q)})$ does not depend on n (since the random variables are exchangeable in $n \in I^{(q)}_{\text{on},>Z^{(q)}}$). In addition, by construction, $\sum_{n'\in I^{(q)}_{\text{on},>Z^{(q)}}} B^{(n',q+1)} = n_{\text{rep}}$. Thus, we have by a simple counting argument: for any $n \in I^{(q)}_{\text{on},>Z^{(q)}}$,

$$\begin{aligned}\mathbb{E}\bigl(B^{(n,q+1)}|\mathcal{F}^{(q)}\bigr) &= \frac{1}{\operatorname{card} I^{(q)}_{\text{on},>Z^{(q)}}} \sum_{n'\in I^{(q)}_{\text{on},>Z^{(q)}}} \mathbb{E}\bigl(B^{(n',q+1)}|\mathcal{F}^{(q)}\bigr)\\ &= \frac{\mathbb{E}(\sum_{n'\in I^{(q)}_{\text{on},>Z^{(q)}}} B^{(n',q+1)}|\mathcal{F}^{(q)})}{\operatorname{card} I^{(q)}_{\text{on},>Z^{(q)}}} = \frac{n_{\text{rep}}}{n_{\text{rep}} - K^{(q+1)}}.\end{aligned}$$

Thus, for $n \in I^{(q)}_{\text{on},>Z^{(q)}}$ (and since $P^{(q+1)}(n) = n$) the formula $G^{(n,q+1)} = \frac{n_{\text{rep}}-K^{(q+1)}}{n_{\text{rep}}} G^{(n,q)}$ in (29) for the AMS algorithm is indeed consistent with the updating formula (11) for the weights in the GAMS framework.

Third, for $n \in I_{\mathrm{new}}^{(q+1)}$, $G^{(n,q+1)} = G^{(P^{(q+1)}(n),q+1)} = \frac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} G^{(P^{(q+1)}(n),q)}$ which is again consistent with the updating formula (11) for the weights in the GAMS framework since $\frac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} = 1/\mathbb{E}(B^{(P^{(q+1)}(n),q+1)} | \mathcal{F}^{(q)})$.

Computation of the stopping levels. Let us now check that the requirements on $Z^{(q)}$ in Assumption 3 are satisfied. By definition of $Z^{(q+1)}$ [see the level computation step of the AMS algorithm, with (30)], it is clear that $Z^{(q+1)} \geq Z^{(q)}$ (actually, the strict inequality $Z^{(q+1)} > Z^{(q)}$ holds). It remains to prove that $Z^{(q)}$ is a stopping level for the filtration $(\mathcal{F}_z^{(q)})_{z \in \mathbb{R}}$.

We start with an elementary result, which again highlights the importance of the strict inequality $> z$ in the definitions (26) of $\mathrm{T}_z(x)$ and of $\tau_z = \mathrm{T}_z(X)$.

LEMMA 4.6. *Let $X : \Omega \to \mathcal{P}$ be a Markov chain over the state space $\mathcal{S}$ [see equation (20)]. Then the random variable $\Xi(X)$ [where, we recall, the maximum level mapping Ξ is defined by (25)] is a $(\mathrm{filt}_z^X)_{z \in \mathbb{R}}$-stopping level: for any $z \in \mathbb{R}$, $\{\Xi(X) \leq z\} \in \mathrm{filt}_z^X$.*

PROOF. On one hand, we clearly have the equality of subsets of $\mathcal{P}$:

$$\{x \in \mathcal{P} : \Xi(x) \leq z\} = \{x \in \mathcal{P} : \mathrm{T}_z(x) = +\infty\}.$$

On the other hand, $\tau_z = \mathrm{T}_z(X)$ is a filt_z^X-measurable random variable. The result is then a consequence of these two facts. □

We are now in position to prove the last result which is needed for Assumption 3 to hold.

LEMMA 4.7. *For any $q \geq 0$, $Z^{(q)}$ is a stopping level with respect to the filtration $(\mathcal{F}_z^{(q)})_{z \in \mathbb{R}}$: for any $z \in \mathbb{R}$, $\{Z^{(q)} \leq z\} \in \mathcal{F}_z^{(q)}$.*

PROOF. Set by convention $Z^{(-1)} = -\infty$ and let us consider $q \geq 0$. Let us introduce the kth order statistics over the maximum levels at iteration q: $L_k^{(q+1)} = \Xi(X^{(\Sigma^{(q+1)}(k),q+1)})$. Let us also introduce $L_{\max}^{(q+1)} = \max\{\Xi(X^{(n,q+1)}) : n \in I_{\mathrm{on}}^{(q+1)}\}$. By definition of $Z^{(q+1)}$ (see the level computation step of the AMS algorithm),

$$Z^{(q+1)} = L_k^{(q+1)} \mathbb{1}_{\{L_k^{(q+1)} < L_{\max}^{(q+1)}\}} + (+\infty) \mathbb{1}_{\{L_k^{(q+1)} = L_{\max}^{(q+1)}\}}.$$

Therefore, for any $z \in \mathbb{R}$, (using the partition $\Omega = \{L_{\max}^{(q+1)} \leq z\} \sqcup \{L_{\max}^{(q+1)} > z\}$)

$$\begin{aligned}\{Z^{(q+1)} \leq z\} &= \{L_k^{(q+1)} \leq z\} \cap \{L_k^{(q+1)} < L_{\max}^{(q+1)}\} \\ &= \{L_k^{(q+1)} < L_{\max}^{(q+1)} \leq z\} \sqcup \{L_k^{(q+1)} \leq z < L_{\max}^{(q+1)}\}.\end{aligned}$$

These events belong to $\sigma(\{\Xi(X^{(n,q+1)}) \le z\}, \{\Xi(X^{(n,q+1)}) \le Z^{(q)}\}, n \in I^{(q+1)})$ (in particular, the set of labels $I_{\mathrm{on}}^{(q+1)}$ is measurable with respect to the sigma-field $\{\Xi(X^{(n,q+1)}) \le Z^{(q)}\}$). To conclude, note that by construction [level computation step, (i)] and thanks to Lemma 4.6: for any $z \in \mathbb{R}$,

$$\sigma\big(\{\Xi\big(X^{(n,q+1)}\big) \le z\}, \{\Xi\big(X^{(n,q+1)}\big) \le Z^{(q)}\}, n \in I^{(q+1)}\big) \subset \mathcal{F}_z^{(q+1)}. \qquad \square$$

4.7. *The AMS estimator for the probability* (22). An immediate corollary of the results of Sections 4.4 and 4.6 above is that the GAMS framework encompasses the AMS algorithm, and that the unbiasedness result, Theorem 3.2, proven in Section 3 can be applied. We detail the particular example of the estimation of the probability $p = \mathbb{P}(\tau_B < \tau_A)$; see (22).

Observe that the weights are easily computed: for $n \in I_{\mathrm{on}}^{(q+1)}$,

$$G^{(n,q+1)} = \frac{n_{\mathrm{rep}} - K^{(q+1)}}{n_{\mathrm{rep}}} \frac{n_{\mathrm{rep}} - K^{(q)}}{n_{\mathrm{rep}}} \cdots \frac{n_{\mathrm{rep}} - K^{(1)}}{n_{\mathrm{rep}}} \frac{1}{n_{\mathrm{rep}}}.$$

Moreover, the weight of a replica remains constant as soon as it is declared retired (namely from the first iteration q such that its label is in $I_{\mathrm{off}}^{(q+1)}$).

Finally, due to the assumption (24) on B, only replicas with labels in $I_{\mathrm{on}}^{(Q_{\mathrm{iter}})}$ contribute to the estimation of p, and thus, from one iteration to the other, only replicas with labels in $I_{\mathrm{on}}^{(q)}$ have to be retained, namely a system of n_{rep} replicas.

Then, for the specific observable $\varphi(x) = \mathbb{1}_{\mathrm{T}_B(x) < \mathrm{T}_A(x)}$, we obtain the following unbiased estimator of $p = \mathbb{P}(\tau_B < \tau_A)$:

$$\begin{aligned} \hat{p} &= \sum_{n \in I_{\mathrm{on}}^{(Q_{\mathrm{iter}})}} G^{(n,Q_{\mathrm{iter}})} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\mathrm{iter}})}) < \mathrm{T}_A(X^{(n,Q_{\mathrm{iter}})})} \\ &= \frac{n_{\mathrm{rep}} - K^{(Q_{\mathrm{iter}})}}{n_{\mathrm{rep}}} \cdots \frac{n_{\mathrm{rep}} - K^{(1)}}{n_{\mathrm{rep}}} P_{\mathrm{corr}}, \end{aligned} \tag{31}$$

where the so-called "corrector term" is given by

$$P_{\mathrm{corr}} = \frac{1}{n_{\mathrm{rep}}} \sum_{n \in I_{\mathrm{on}}^{(Q_{\mathrm{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\mathrm{iter}})}) < \mathrm{T}_A(X^{(n,Q_{\mathrm{iter}})})} \tag{32}$$

namely the proportion of working replicas that have reached B before A at the final iteration. The properties of this estimator will be numerically investigated in Section 5.

Note that the AMS algorithm presented in this section satisfies the almost sure mass conservation property, Definition 3.1. Applying Theorem 3.2, we obtain the following result.

COROLLARY 4.8. *For any choice of the number of replicas* n_{rep}, *of* k, *and of the reaction coordinate* ξ [*provided it satisfies condition* (24)], $\hat{p}$ *defined by* (31) *is an unbiased estimator of the probability* $p = \mathbb{P}(\tau_B < \tau_A)$ *defined by* (22):

$$\mathbb{E}(\hat{p}) = p = \mathbb{P}(\tau_B < \tau_A).$$

5. Numerical illustration. The aim of this section is to illustrate the behavior of the AMS algorithm as defined in Section 4, in various situations involving discrete-time approximations (1) of the overdamped Langevin dynamics in dimension 2.

We would like to discuss in particular the unbiasedness (Corollary 4.8) of the AMS estimator $\hat{p}$ of $p = \mathbb{P}(\tau_B < \tau_A)$ [defined by the formula (31)] whatever the choice of the reaction coordinate ξ, the number of replicas n_{rep} and the minimal number k of replicas which are declared retired and resampled at each iteration of the AMS algorithm.

In the following, $(\hat{p}_m)_{1 \leqslant m \leqslant N}$ refers to i.i.d. random variables distributed like the estimator $\hat{p}$, obtained by N independent realizations of the algorithm. The associated empirical mean is denoted by

$$\overline{p}_N = \frac{1}{N} \sum_{m=1}^{N} \hat{p}_m. \tag{33}$$

The variance of the estimator $\hat{p}$ is also investigated numerically, and it is shown for a two-dimensional process that the variance heavily depends on the choice of the reaction coordinate. In the following, we will denote by

$$\delta_N = 2 \times \frac{1.96}{\sqrt{N}} \times \sqrt{\frac{1}{N} \sum_{m=1}^{N} (\hat{p}_m)^2 - (\overline{p}_N)^2} \tag{34}$$

the size of the 95% empirical confidence interval computed using the empirical variance obtained over N independent runs of the algorithm.

The section is organized as follows. In Section 5.1, we give an example on which we discuss the efficiency of the AMS algorithm by studying how the convergence of the estimator depends on the reaction coordinate ξ. In Section 5.2, we draw some conclusions and practical recommendations from these numerical experiments. We refer to [7], Section 5, for additional simulations.

5.1. *The bi-channel problem.* The aim of this two-dimensional example is to investigate the importance of the choice of the reaction coordinate on the efficiency of the algorithm, on a typical example which has been used in previous numerical studies; see [17, 34, 35].

5.1.1. *The model.* For an initial condition $X_0 = x_0 \in \mathbb{R}^2$, a time step size $h > 0$, and for $t \in \mathbb{N}$, the Markov process in $\mathbb{R}^2$ we consider is defined by (1) (discretization of the overdamped Langevin dynamics).

In the simulations below, the initial condition is $X_0 = x_0 = (-0.9, 0)$ and the time step is $h = 0.05$. The potential function $V : \mathbb{R}^2 \to \mathbb{R}$ is given by

$$V(x, y) = 0.2x^4 + 0.2\left(y - \frac{1}{3}\right)^4 + 3e^{-x^2-(y-\frac{1}{3})^2} - 3e^{-x^2-(y-\frac{5}{3})^2} - 5e^{-(x-1)^2-y^2} - 5e^{-(x+1)^2-y^2}.$$

This function is plotted on Figure 2. It has two global minima connected one to another by two channels: the upper channel [which goes through the shallow minimum around $(0, 1.5)$] and the lower channel [which goes through the saddle point around $(0, -0.5)$]. The two global minima are close to $m_A = (x_A, y_A) = (-1, 0)$ and $m_B = (x_B, y_B) = (1, 0)$. For some $\rho \in]0, 1[$ (in the numerical applications, we take $\rho = 0.05$), we consider the sets A and B defined as the Euclidean open balls of radius ρ around the two minima m_A and m_B, namely

$$\begin{cases} A = \mathcal{B}(m_A, \rho) = \{(x, y) \in \mathbb{R}^2 : \sqrt{(x - x_A)^2 + (y - y_A)^2} < \rho\}, \\ B = \mathcal{B}(m_B, \rho) = \{(x, y) \in \mathbb{R}^2 : \sqrt{(x - x_B)^2 + (y - y_B)^2} < \rho\}, \end{cases}$$

so that $x_0 \in \mathbb{R}^2 \setminus (A \cup B)$.

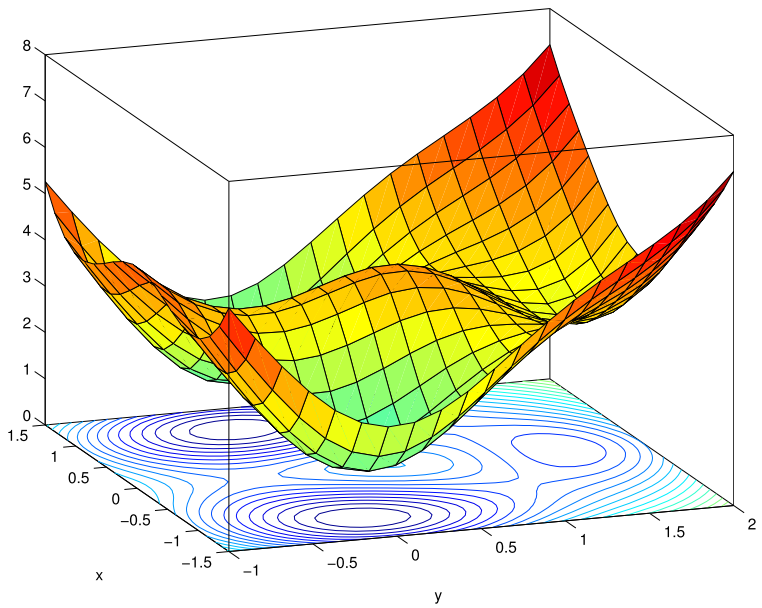

FIG. 2. *Plot of the potential function V for the bi-channel problem.*

Most of the trajectories starting from x_0 hit A before B. Moreover, A and B are metastable states: in the small temperature regime, starting from A (resp. B), it takes a lot of time to leave A (resp. B).

We are interested in the estimation of the probability $p = \mathbb{P}(\tau_B < \tau_A)$, where $\tau_A = \mathrm{T}_A(X)$ and $\tau_B = \mathrm{T}_B(X)$ are the first hitting times of sets A and B by the process X; see equation (23).

We will consider the results of the AMS algorithm for the three reaction coordinates ξ^i with $i \in \{1, 2, 3\}$:

1. The *norm to the initial point* m_A:

$$\xi^1(x, y) = \sqrt{(x - x_A)^2 + (y - y_A)^2}.$$

2. The *norm to the final point* m_B:

$$\xi^2(x, y) = \xi^1(x_B, y_B) - \sqrt{(x - x_B)^2 + (y - y_B)^2}.$$

3. The *abscissa*: $\xi^3(x, y) = x$.

The maximum levels used in the stopping criterion of the algorithm are $z^1_{\max} = z^2_{\max} = 1.9$ and $z^3_{\max} = 0.9$. Notice that for $i \in \{1, 2, 3\}$, we have $B \subset (\xi^i)^{-1}(]z^i_{\max}, +\infty[)$, and thus (24) is satisfied.

In this section, we take $k = 1$ and the number of replicas is $n_{\mathrm{rep}} = 100$. The values of β belong to the set $\{8.67, 9.33, 10\}$ which are associated with probabilities p ranging approximately from 2.10^{-9} to 10^{-10}.

5.1.2. *Evolution of the empirical mean.* Let us first perform simulations with N independent runs of the algorithm, N varying between 1 and 6.10^6. We represent on Figure 3 the evolution as a function of N of the empirical mean $\overline{p}_N$ [defined by (33)] and of the associated 95% confidence intervals $[\overline{p}_n - \delta_N/2, \overline{p}_n + \delta_N/2]$ computed using the empirical variance; see (34).

The colors in the figures are as follows: green (solid line) for ξ^1, red (line with crosses) for ξ^2 and blue (line with circles) for ξ^3. The full lines represent the evolution of the upper and lower bounds of the confidence intervals, while dashed lines represent the evolution of the empirical means.

From these simulations, we observe that:

- When N is sufficiently large, the confidence intervals overlap. This is in agreement with the fact that $\hat{p}$ is an unbiased estimator of p whatever the choice of the reaction coordinate.
- The statistical fluctuations depend a lot on the reaction coordinate. In particular, the results obtained with ξ^1 seem much better than with ξ^2 or ξ^3. We will come back to this in Section 5.1.3.
- The confidence interval being computed empirically, one may conclude that the algorithm is biased by considering the results for N too small (see, e.g., the

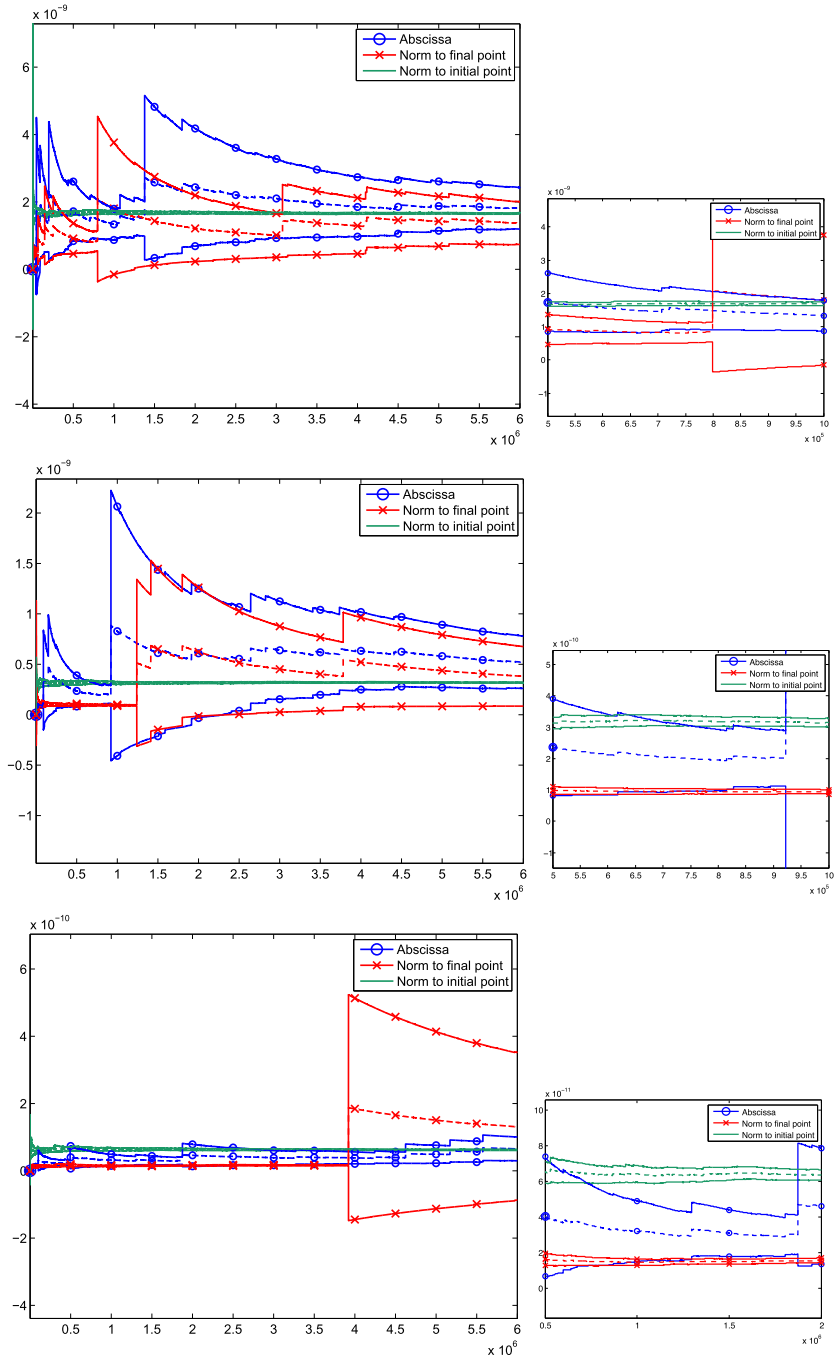

FIG. 3. *Evolution as a function of N of the empirical mean $\overline{p}_N$ and of the associated 95% confidence intervals $[\overline{p}_N - \delta_N/2, \overline{p}_N + \delta_N/2]$. Upper to lower $\beta = 8.67, 9.33, 10$. The right inserts are zooms of the left graphs on smaller values of N, in order to illustrate the "apparent bias" phenomenon.*

graphs in the right column in Figure 3). This is due to the fact that the empirical variance dramatically underestimates the real variance if N is too small. This is a well-known phenomenon for splitting algorithms in general called "apparent bias"; see [24]. As β gets larger (namely as the temperature gets smaller), the number of independent runs N required to observe overlapping confidence intervals gets larger.

We observe that there are some realizations for which the estimator of the probability is very large. These realizations have small probability but they dramatically increase the value of the empirical mean and of the empirical variance. This explains the large variations which are observed on the empirical average and confidence interval as a function of the number of realizations; see Figure 3. As it is usually the case with Monte Carlo simulations for rare event simulations, it is impossible to decide a priori if the sample size N is sufficiently large to give an accurate estimation. However, using the unbiasedness property (Theorem 3.2), a sensible way to choose a number of realizations N is to set it sufficiently large so that the confidence intervals obtained with different reaction coordinates overlap.

5.1.3. *Fluctuations induced by the two channels.* In this section, we compare the results when using two reaction coordinates: ξ^1 (norm to the initial point) and ξ^3 (abscissa). Since the typical behavior we observe in Figure 3 and in Table 1 is the same for ξ^2 and ξ^3, we do not repeat the analysis for ξ^2. Our aim in this section is to relate the large variations observed in Figure 3 with the fact that two channels connect A to B.

As explained above, there are two possible channels for the reactive trajectories going from A to B: the upper channel and the lower channel. For each realization m, one can distinguish the contributions to the estimator $\hat{p}_m$ of the replicas going through the upper channel and the ones going through the lower channel. In the following, for a given path, the trajectory is associated to the upper (resp., lower) channel if the first hitting point of the y-axis is such that

TABLE 1

The bi-channel case. Proportion and conditional probabilities for two reaction coordinates: the norm to the initial point (ξ^1) *and the abscissa* (ξ^3). *e-n stands for* 10^{-n}

	β	N	R_N	ρ_N^{upper}	ρ_N^{mix}	ρ_N^{lower}	$\tilde{p}_N^{\text{upper}}$	$\tilde{p}_N^{\text{mix}}$	$\tilde{p}_N^{\text{lower}}$	$\overline{p}_N$
ξ^1	8.67	2.10^6	0.81	0.45	0.03	0.52	2.7e–09	3.0e–09	2.3e–09	1.7e–09
ξ^3	8.67	2.10^6	0.99	0.0008	0.02	0.98	2.3e–06	5.9e–10	5.5e–10	2.4e–09
ξ^1	9.33	4.10^6	0.72	0.51	0.02	0.47	6.2e–10	6.3e–10	2.5e–10	3.2e–10
ξ^3	9.33	4.10^6	0.97	0.0005	0.02	0.98	1.0e–06	5.6e–11	9.7e–11	6.0e–10
ξ^1	10	6.10^6	0.62	0.51	0.01	0.48	1.5e–10	1.4e–10	5.2e–11	6.2e–11
ξ^3	10	6.10^6	0.92	0.0004	0.01	0.99	1.4e–07	1.5e–11	1.8e–11	6.8e–11

$y > 0.5$ (resp., such that $y \leq 0.5$). More precisely, let us define $\Pi_1(x, y) = x$ and $\Pi_2(x, y) = y$ for any $(x, y) \in \mathbb{R}^2$. For a replica $X = (X_t)_{t\in\mathbb{N}}$ such that $\tau = \inf\{t \in \mathbb{N} : \Pi_1(X_t) > 0\} < \infty$, $X \in \text{Upper}$ if $\Pi_2(X_\tau) > 0.5$ and $X \in \text{Lower}$ if $\Pi_2(X_\tau) \leq 0.5$.

For each realization of the algorithm, we compute the three following quantities:

- the number of replicas which reach B before A:

$$\mathrm{M}^B = \sum_{n\in I_{\text{on}}^{(Q_{\text{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})})<\mathrm{T}_A(X^{(n,Q_{\text{iter}})})};$$

- the number of replicas which reach B before A and go through the upper channel:

$$\mathrm{M}^{B,\text{upper}} = \sum_{n\in I_{\text{on}}^{(Q_{\text{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})})<\mathrm{T}_A(X^{(n,Q_{\text{iter}})})} \mathbb{1}_{X^{(n,Q_{\text{iter}})}\in\text{Upper}};$$

- the number of replicas which reach B before A and go through the lower channel:

$$\mathrm{M}^{B,\text{lower}} = \sum_{n\in I_{\text{on}}^{(Q_{\text{iter}})}} \mathbb{1}_{\mathrm{T}_B(X^{(n,Q_{\text{iter}})})<\mathrm{T}_A(X^{(n,Q_{\text{iter}})})} \mathbb{1}_{X^{(n,Q_{\text{iter}})}\in\text{Lower}}.$$

Notice that $\mathrm{M}^B = \mathrm{M}^{B,\text{upper}} + \mathrm{M}^{B,\text{lower}}$ and that $\mathrm{M}^B \neq 0$ is equivalent to $\hat{p} \neq 0$. When needed, we explicitly indicate the dependence of these quantities on the realization by a lower script m: for $m \in \{1, \ldots, N\}$, we thus denote M_m^B, $\mathrm{M}_m^{B,\text{upper}}$ and $\mathrm{M}_m^{B,\text{lower}}$ the mth realization of M^B, $\mathrm{M}^{B,\text{upper}}$ and $\mathrm{M}^{B,\text{lower}}$.

Let us introduce the set $\mathcal{E}_N = \{m : \hat{p}_m \neq 0\}$ of realizations which lead to a nonzero $\hat{p}$ and the proportion $R_N = \operatorname{card}\mathcal{E}_N/N$ of such realizations. We now divide the realizations in $\mathcal{E}_N$ into three disjoint subsets, with associated proportions.

- All replicas reaching B before A go through the upper channel:

$$\mathcal{E}_N^{\text{upper}} = \{m \in \mathcal{E}_N : \mathrm{M}_m^{B,\text{lower}} = 0\} \text{ and } \rho_N^{\text{upper}} = \frac{\operatorname{card}\mathcal{E}_N^{\text{upper}}}{\operatorname{card}\mathcal{E}_N}.$$

- All replicas reaching B before A go through the lower channel:

$$\mathcal{E}_N^{\text{lower}} = \{m \in \mathcal{E}_N : \mathrm{M}_m^{B,\text{upper}} = 0\} \text{ and } \rho_N^{\text{lower}} = \frac{\operatorname{card}\mathcal{E}_N^{\text{lower}}}{\operatorname{card}\mathcal{E}_N}.$$

- Both channels are used by the replicas reaching B before A:

$$\mathcal{E}_N^{\text{mix}} = \mathcal{E}_N \setminus (\mathcal{E}_N^{\text{upper}} \cup \mathcal{E}_N^{\text{lower}}) \text{ and } \rho_N^{\text{mix}} = \frac{\operatorname{card}\mathcal{E}_N^{\text{mix}}}{\operatorname{card}\mathcal{E}_N}.$$

Obviously, $\rho_N^{\text{upper}} + \rho_N^{\text{lower}} + \rho_N^{\text{mix}} = 1$. Finally, we define conditional estimators for $\hat{p}$ associated with the partition of $\mathcal{E}_N$ defined above:

$$\tilde{p}_N^{\text{upper}} = \frac{\sum_{m \in \mathcal{E}_N^{\text{upper}}} \hat{p}_m}{\operatorname{card} \mathcal{E}_N^{\text{upper}}}, \qquad \tilde{p}_N^{\text{lower}} = \frac{\sum_{m \in \mathcal{E}_N^{\text{lower}}} \hat{p}_m}{\operatorname{card} \mathcal{E}_N^{\text{lower}}} \quad \text{and} \quad \tilde{p}_N^{\text{mix}} = \frac{\sum_{m \in \mathcal{E}_N^{\text{mix}}} \hat{p}_m}{\operatorname{card} \mathcal{E}_N^{\text{mix}}}.$$

Notice that

$$\overline{p}_N = R_N \big(\rho_N^{\text{upper}} \tilde{p}_N^{\text{upper}} + \rho_N^{\text{lower}} \tilde{p}_N^{\text{lower}} + \rho_N^{\text{mix}} \tilde{p}_N^{\text{mix}} \big).$$

In other words, we have separated the nonzero contributions to $\overline{p}_N$ into (i) realizations for which all the replicas go through the upper channel (first term in the parenthesis), (ii) realizations for which all the replicas go through the lower channel (second term in the parenthesis) and finally (iii) realizations for which the two channels are used by the replicas (third term in the parenthesis).

Let us emphasize that contrary to $\overline{p}_N$, the limit when $N \to \infty$ of the estimators R_N, ρ_N^{upper}, ρ_N^{lower}, ρ_N^{mix}, $\tilde{p}_N^{\text{upper}}$, $\tilde{p}_N^{\text{mix}}$ or $\tilde{p}_N^{\text{lower}}$ (for a given value of n_{rep}) depends on the choice of the reaction coordinate ξ; see [7], Remark 5.5.

From Table 1, we observe that for ξ^1, approximately half of the realizations use exclusively the upper channel and the other half use the lower channel. The associated conditional estimators $\tilde{p}_N^{\text{upper}}$ and $\tilde{p}_N^{\text{lower}}$ are very close. This is not the case for ξ^3: only very few realizations go through the upper channel while the associated probability $\tilde{p}_N^{\text{upper}}$ is much larger than the two other ones $\tilde{p}_N^{\text{lower}}$ and $\tilde{p}_N^{\text{mix}}$. This means that a few realizations contribute a lot to the empirical average $\overline{p}_N$. This explains the very large confidence intervals observed with ξ^3 (in comparison with those observed for ξ^1) on Figure 3.

5.2. *Conclusions and practical recommendations.* Let us summarize our findings on these numerical simulations.

- We always observe that for sufficiently large values of N (number of independent Monte Carlo simulations), the confidence intervals of the estimator $\hat{p}$ overlap, whatever n_{rep}, k or ξ. This is in accordance with our theoretical result on the unbiasedness of this estimator.
- In multiple channel cases (namely when multiple pathways exist from A to B), one may observe nonoverlapping empirical confidence intervals of the estimator for different reaction coordinates if the number of independent realizations N is too small. This is related to the fact that very large contributions to the average of the estimator are associated with trajectories going through very unlikely (for the considered reaction coordinate and value of n_{rep}) channels. This is a known phenomenon for splitting algorithms in general; see [24], where it is referred to as "apparent bias". In [7], Section 5.3, we report on additional experiments which indeed show that when a single channel connects A to B (again in a two-dimensional setting), the apparent bias phenomenon disappears. As explained in [24], a good reaction coordinate in a multiple channel case is such

that, conditionally to reach a certain maximum level z, the relative likelihood of the channels used by the paths to reach this maximum level does not depend too much on z. For example, a reaction coordinate close to the committor function is a good candidate to achieve this purpose. This opens the route to adaptive algorithms, where the reaction coordinate would be updated in order to get closer and closer to the committor function as long as successive AMS algorithms are launched (see [17]). We intend to investigate this direction in future works.

Let us also mention that we observed in our simulations that the estimator $\hat{p}$ has a heavy tail so that very few realizations contribute a lot to the empirical average (which is also consistent with the results presented in Section 5.1.3). We refer to [7], Section 5.2.3, for a discussion on this aspect.

As a conclusion to these numerical results, we thus recommend the following in order to get reliable estimates of the probability $\mathbb{P}(\tau_B < \tau_A)$ with the AMS algorithm.

- Thanks to the unbiasedness property, one should check the independence of the computed probability on the choice of the parameters: the number of replicas n_{rep}, the minimum number of sampled replicas per iteration k and, more importantly, the reaction coordinate ξ. In particular, we recommend to perform simulations with various reaction coordinates and to set the minimal number of independent realizations such that the empirical confidence intervals overlap.
- Thanks to the unbiasedness property, one can perform many independent realizations of the algorithm with a relatively small number of replicas, instead of using a few independent realizations with a large number of replicas. Indeed, assume that we are in a regime where the variance scales like $\frac{1}{n_{\mathrm{rep}} \times N}$; this is the case for instance in the so-called ideal case for sufficiently large n_{rep} and N; see [10]. Since the parallelization of independent runs of the algorithm is trivial, for a fixed product $n_{\mathrm{rep}} \times N$ (namely for a fixed CPU cost), the strategy with less replicas is thus much more interesting in terms of wall-clock time (which scales like n_{rep}) than the strategy with more replicas.

Finally, let us recall that one should be careful in the implementation of the splitting and branching steps, in particular in the treatment of replicas which have the same maximum level and in the definition of the branching point in the partial resampling procedure. One output of this work is to identify correct implementations in such cases. For incorrect implementations, strong biases may be observed; see [7], Section 5.1.

Acknowledgments. We would like to thank the referees for their comments which helped us improving the presentation of the paper.

C.-E. Bréhier is grateful to INRIA Rocquencourt for having funded his postdoctoral position (September 2013–December 2014). M. Gazeau is grateful to INRIA Lille Nord Europe where a part of this research was conducted and to Labex

CEMPI (ANR-11-LABX-0007-01). The authors are grateful to the Labex Bezout (ANR-10-LABX-58-01) which supported this project at an early stage, during the CEMRACS 2013. T. Lelièvre would like to acknowledge the Laboratoire International Associé between the Centre National de la Recherche Scientifique (CNRS) and the University of Illinois at Urbana-Champaign (UIUC). Finally, the authors would like to thank F. Bouchet, F. Cérou, A. Guyader, J. Rolland and E. Simonnet for many fruitful discussions.

REFERENCES

[1] ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation*: *Algorithms and Analysis*. *Stochastic Modelling and Applied Probability* **57**. Springer, New York. MR2331321

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

[3] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics*: *Probability and Statistics*. Wiley, New York. MR1700749

[4] BLANCHET, J., GLYNN, P. and LIU, J. C. (2006). State-dependent importance sampling and large deviations. In *Proceedings of the* 1*st International Conference on Performance Evaluation Methodolgies and Tools*, *Valuetools* '06 ACM, New York.

[5] BRÉHIER, C.-E. (2015). Large deviations principle for the adaptive multilevel splitting algorithm in an idealized setting. *ALEA Lat. Am. J. Probab. Math. Stat.* **12** 717–742. MR3446035

[6] BRÉHIER, C. E., CHAUDRU DE RAYNAL, P. E., LEMAIRE, V., PANLOUP, F. and REY, C. (2015). Recent advances in various fields of numerical probability. *ESAIM Proceedings and Reviews* **51** 272–292.

[7] BRÉHIER, C.-E., GAZEAU, M., GOUDENÈGE, L., LELIÈVRE, T. and ROUSSET, M. (2015). Unbiasedness of some generalized adaptive multilevel splitting algorithms. Preprint. Available at arXiv:1505.02674.

[8] BRÉHIER, C.-E., GAZEAU, M., GOUDENÈGE, L. and ROUSSET, M. (2015). Analysis and simulation of rare events for SPDEs. *ESAIM Proceedings and Reviews* **48** 364–384.

[9] BRÉHIER, C.-E., GOUDENÈGE, L. and TUDELA, L. (2014). Central limit theorem for adaptive multilevel splitting estimators in an idealized setting. In *Monte Carlo and Quasi Monte Carlo Methods*. Springer, MCQMC, Leuven, Belgium.

[10] BRÉHIER, C.-E., LELIÈVRE, T. and ROUSSET, M. (2015). Analysis of adaptive multilevel splitting algorithms in an idealized case. *ESAIM Probab. Stat.* **19** 361–394. MR3417480

[11] BUCKLEW, J. A. (2004). *Introduction to Rare Event Simulation*. Springer, New York. MR2045385

[12] CARON, V., GUYADER, A., MUNOZ ZUNIGA, M. and TUFFIN, B. (2014). Some recent results in rare event estimation. In *Journées MAS* 2012. *ESAIM Proc.* **44** 239–259. EDP Sci., Les Ulis. MR3178620

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

[14] CÉROU, F., DEL MORAL, P., LEGLAND, F. and LEZAUD, P. (2006). Genetic genealogical models in rare event analysis. *ALEA Lat. Am. J. Probab. Math. Stat.* **1** 181–203. MR2249654

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

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

[17] CÉROU, F., GUYADER, A., LELIÈVRE, T. and POMMIER, D. (2011). A multiple replica approach to simulate reactive trajectories. *J. Chem. Phys.* **134** 054108 (16 p.).
[18] DEAN, T. and DUPUIS, P. (2009). Splitting for rare event simulation: A large deviation approach to design and analysis. *Stochastic Process. Appl.* **119** 562–587. MR2494004
[19] DEL MORAL, P. (2004). *Feynman–Kac Formulae*. Springer, New York. MR2044973
[20] DEL MORAL, P., DOUCET, A. and JASRA, A. (2006). Sequential Monte Carlo samplers. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 411–436. MR2278333
[21] DEL MORAL, P. and GARNIER, J. (2005). Genealogical particle analysis of rare events. *Ann. Appl. Probab.* **15** 2496–2534. MR2187302
[22] DUPUIS, P. and WANG, H. (2004). Importance sampling, large deviations, and differential games. *Stoch. Stoch. Rep.* **76** 481–508. MR2100018
[23] GARVELS, M. J. J., KROESE, D. P. and VAN OMMEREN, J. C. W. (2002). On the importance function in splitting simulation. *European Transactions on Telecom-Munications* **13** 363–371.
[24] GLASSERMAN, P., HEIDELBERGER, P., SHAHABUDDIN, P. and ZAJIC, T. (1998). A large deviations perspective on the efficiency of multilevel splitting. *IEEE Trans. Automat. Control* **43** 1666–1679. MR1658685
[25] GLASSERMAN, P., HEIDELBERGER, P., SHAHABUDDIN, P. and ZAJIC, T. (1999). Multilevel splitting for estimating rare event probabilities. *Oper. Res.* **47** 585–600. MR1710951
[26] GLASSERMAN, P. and WANG, Y. (1997). Counterexamples in importance sampling for large deviations probabilities. *Ann. Appl. Probab.* **7** 731–746. MR1459268
[27] GUYADER, A., HENGARTNER, N. and MATZNER-LØBER, E. (2011). Simulation and estimation of extreme quantiles and extreme probabilities. *Appl. Math. Optim.* **64** 171–196. MR2822407
[28] HAMMERSLEY, J. M. and HANDSCOMB, D. C. (1965). *Monte Carlo Methods*. Methuen, London. MR0223065
[29] JOHANSEN, A. M., DEL MORAL, P. and DOUCET, A. (2006). Sequential Monte Carlo samplers for rare events. In *Proceedings of the* 6*th International Workshop on Rare Event Simulation* (*RESIM* 2006) 256–267, Bamberg.
[30] KAHN, H. and HARRIS, T. E. (1951). Estimation of particle transmission by random sampling. *National Bureau of Standards* **12** 27–30.
[31] KARATZAS, I. and SHREVE, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. *Graduate Texts in Mathematics* **113**. Springer, New York. MR1121940
[32] LAGNOUX-RENAUDIE, A. (2008). Effective branching splitting method under cost constraint. *Stochastic Process. Appl.* **118** 1820–1851. MR2454466
[33] LAGNOUX-RENAUDIE, A. (2009). A two-step branching splitting model under cost constraint for rare event analysis. *J. Appl. Probab.* **46** 429–452. MR2535824
[34] METZNER, P., SCHÜTTE, C. and VANDEN-EIJNDEN, E. (2006). Illustration of transition path theory on a collection of simple examples. *J. Chem. Phys.* **125** 084110. Available at http://dx.doi.org/10.1063/1.2335447.
[35] PARK, S., SENER, M. K., LU, D. and SCHULTEN, K. (2003). Reaction paths based on mean first-passage times. *J. Chem. Phys.* **119** 1313–1319.
[36] ROLLAND, J. and SIMONNET, E. (2015). Statistical behaviour of adaptive multilevel splitting algorithms in simple models. *J. Comput. Phys.* **283** 541–558. MR3294688
[37] ROSENBLUTH, M. N. and ROSENBLUTH, A. W. (1955). Monte Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.* **23** 356–359.
[38] RUBINO, G. and TUFFIN, B. (2009). Introduction to rare event simulation. In *Rare Event Simulation Using Monte Carlo Methods* 1–13. Wiley, Chichester. MR2730759
[39] SHIRYAYEV, A. N. (1984). *Probability. Graduate Texts in Mathematics* **95**. Springer, New York. MR0737192

[40] SIMONNET, E. (2014). Combinatorial analysis of the adaptive last particle method. *Stat. Comput.* **26** 211–230.

[41] SKILLING, J. (2006). Nested sampling for general Bayesian computation. *Bayesian Anal.* **1** 833–859 (electronic). MR2282208

[42] VANDEN-EIJNDEN, E. and WEARE, J. (2012). Rare event simulation of small noise diffusions. *Comm. Pure Appl. Math.* **65** 1770–1803. MR2982641

[43] VILLÉN-ALTAMIRANO, M. and VILLÉN-ALTAMIRANO RESTART, J. (1991). A method for accelerating rare events simulations. In *Proceeding of the Thirteenth International Teletraffic Congress*, *Volume Copenhagen*, *Denmark*, *June* 19–26 *of Queueing*, *Performance and Control in ATM*: *ITC*-13 *Workshops* 71–76. North-Holland, Amsterdam.

[44] WALTER, C. Moving particles: A parallel optimal multilevel splitting method with applications in quantiles estimation and meta-model-based algorithms. *Struct. Saf.* **55** 10–25.

C.-E. BRÉHIER
CNRS UMR 5208, INSTITUT CAMILLE JORDAN
UNIV LYON, UNIVERSITÉ CLAUDE BERNARD LYON 1
43 BD. DU 11 NOVEMBRE 1918
F-69622 VILLEURBANNE CEDEX
FRANCE
E-MAIL: brehier@math.univ-lyon1.fr

L. GOUDENÈGE
FÉDÉRATION DE MATHÉMATIQUES
DE L'ÉCOLE CENTRALE PARIS
CNRS
GRANDE VOIE DES VIGNES
92295 CHÂTENAY-MALABRY
FRANCE
E-MAIL: goudenege@math.cnrs.fr

M. GAZEAU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF TORONTO
40 ST. GEORGE ST.
TORONTO M5S 2E4
CANADA
E-MAIL: gazeauma@math.toronto.edu

T. LELIÈVRE
M. ROUSSET
UNIVERSITÉ PARIS-EST
CERMICS (ENPC), INRIA
6-8 AVENUE BLAISE PASCAL
CITÉ DESCARTES
F-77455 MARNE-LA-VALLÉE
FRANCE
E-MAIL: lelievre@cermics.enpc.fr
roussetm@cermics.enpc.fr